PROBLEMS IN THE ANALYSIS OF CZECH FOR MACHINE TRANSLATION
- A CATEGORIAL GRAMMAR AND
FUNCTIONAL SENTENCE PERSPECTIVE APPROACH

Examensarbete vid Datalingvistlinjen
Göteborgs Universitet
Vårterminen 1994

Sylvana Sofkova

Handledare:
Joakim Nivre, Institutionen för lingvistik
Gudrun Magnusdottir, Institutionen för svenska språket
ABSTRACT

This thesis is about the analysis of Czech for machine translation with emphasis on word order and definiteness marking.

Two problems have been studied: the problem of parsing in a 'free' word order language and the problem of assigning species to nouns in a way that allows a correct choice of articles in translation to English.

The background part includes an introduction to Czech grammar and aspects in the analysis of discourse structure. Different theories of topic, comment and focus are discussed and the relations between noun phrases, species and articles in Czech and English are investigated. The background also includes an introduction to categorial grammar chart parsing and a short discussion of previous implementations.

A solution to the parsing problem is presented which extends the basic categorial grammar formalism with rules of functional composition, associativity and undirected functors. The theory of Functional Sentence Perspective (FSP), as one of the theories concerning topic, comment and focus, has been chosen as the basis for the treatment of species assignment to nouns in Czech and an algorithm that assigns species to all the nouns in a Czech sentence has been developed.

A system for grammatical analysis of Czech has been implemented. The program consists of two parts, a parser based on categorial grammar extended with quasi-logical form (QLF) and a discourse analysis part handling the problem of definiteness marking in Czech. The system takes a small discourse as input, analyses every sentence in the categorial grammar chart parser and feeds the output to the discourse analysis component, where the sentence is examined with respect to the topic and focus distribution as defined in the FSP-theory. The analysis of topic and focus distribution provides the basis for the assignment of species to nouns.

The system returns a two level QLF for every sentence in the input discourse. The first level mainly describes the grammatical functions involved, while the second gives a list of all constituents, ordered as in the surface word order. In this list, every word is represented as a term, where every term includes the word's lexeme and its inflectional features. In the case of a noun, the list of features includes an appropriate species feature obtained in the discourse analysis part, which can be used for the choice of article in automatic translation to English.

The implementation was tested on a small sample and showed satisfactory results for all the sentences tested.

The implementation is carried out in the AAIS-Prolog (Advanced Artificial Intelligence Systems' Prolog) environment.
LIST OF ABBREVIATIONS

The following abbreviations are used in the literal English translation corresponding to the Czech sentences:

AM animate masculine gender
IM inanimate masculine gender
F feminine gender
N neuter gender
Nom nominative case
Gen genitive case
Dat dative case
Acc accusative case
Voc vocative case
Loc locative case
Ins instrumental case
S singular
P plural
1PS 1st person singular
2PS 2nd person singular
3PS 3rd person singular
1PP 1st person plural
2PP 2nd person plural
3PP 3rd person plural
INF infinitive form
IMP imperative form
AUX auxiliary verb
Pres present tense
Fut future tense
Past past tense
Pass passive voice
refl. reflexive pronoun
cp. conditional particle
# TABLE OF CONTENTS

PREFACE ................................................................................................................................. 1

1 INTRODUCTION .................................................................................................................... 2

2 BACKGROUND .......................................................................................................................... 4
   2.1 PARSING WITH CATEGORIAL GRAMMAR ........................................................................... 4
      2.1.1 INTRODUCTION ........................................................................................................... 4
      2.1.2 BASIC CATEGORIAL GRAMMAR FORMALISM ............................................................... 4
      2.1.3 EXTENDING THE SET OF CATEGORIES ..................................................................... 6
      2.1.4 EXTENDING THE SET OF RULES ................................................................................. 7
      2.1.5 CHART PARSING ......................................................................................................... 10
   2.2 TOPIC, COMMENT AND FOCUS .......................................................................................... 12
      2.2.1 INTRODUCTION .......................................................................................................... 12
      2.2.2 GENERAL DEFINITIONS ............................................................................................. 12
      2.2.3 THE THEORY OF 'FUNCTIONAL SENTENCE PERSPECTIVE' ........................................ 14
      2.2.4 SUMMARY .................................................................................................................. 21
   2.3 CZECH GRAMMAR ............................................................................................................. 22
      2.3.1 GENERAL CHARACTERISTICS OF CZECH .................................................................. 22
      2.3.2 NOUNS ........................................................................................................................ 23
      2.3.3 PRONOUNS ................................................................................................................... 27
      2.3.4 WORD ORDER ............................................................................................................ 29
   2.4 NOUN PHRASES AND ARTICLES IN ENGLISH ................................................................... 35
      2.4.1 INTRODUCTION .......................................................................................................... 35
      2.4.2 DEFINITE VS. INDEFINITE ARTICLE ......................................................................... 35
      2.4.3 NOUN PHRASES WITHOUT AN ARTICLE .................................................................... 36
   2.5 PREVIOUS IMPLEMENTATIONS ........................................................................................... 37
      2.5.1 INTRODUCTION .......................................................................................................... 37
      2.5.2 TOPIC-FOCUS TREATMENT ....................................................................................... 37
      2.5.3 ARTICLE CHOICE AND ASPECT ............................................................................... 37

3 METHOD ................................................................................................................................ 39
   3.1 INTRODUCTION ................................................................................................................ 39
   3.2 PARSING ............................................................................................................................ 39
   3.3 DISCOURSE ANALYSIS ..................................................................................................... 39
      3.3.1 INTRODUCTION .......................................................................................................... 39
      3.3.2 TOPIC-FOCUS RECOGNITION .................................................................................... 40
      3.3.3 SPECIES ASSIGNMENT ............................................................................................. 41
   3.4 TESTING .............................................................................................................................. 41

4 ANALYSIS ............................................................................................................................ 42
   4.1 INTRODUCTION ................................................................................................................ 42
   4.2 PARSING IN A 'FREE' WORD ORDER LANGUAGE .............................................................. 42
      4.2.1 INTRODUCTION .......................................................................................................... 42
      4.2.2 THE CATEGORIAL GRAMMAR LEXICON .................................................................... 42
      4.2.3 CATEGORIAL GRAMMAR RULES .............................................................................. 44
      4.2.4 CONCLUSION .............................................................................................................. 50
   4.3 EXTENDING THE CG-FORMALISM ...................................................................................... 51
   4.4 DISCOURSE ANALYSIS .................................................................................................... 52
      4.4.1 INTRODUCTION .......................................................................................................... 52
      4.4.2 TOPIC AND FOCUS RECOGNITION .......................................................................... 53
      4.4.3 SPECIES ASSIGNMENT ............................................................................................. 53
PREFACE

This essay represents the work developed for the degree of Master of Arts in computational linguistics at the University of Göteborg. The work started in November 1993 at the Department of Language and Linguistics at UMIST\textsuperscript{1}, with studies of Czech grammar and the implementation of a categorial grammar chart parser for Czech. The work continued in Gothenburg with further development of the parser and discourse analysis.

I would like to thank my supervisors here at the University of Göteborg:

Joakim Nivre, Department of Linguistics and
Gudrun Magnusdottir, Department of Swedish

for their constructive ideas and comments, help in all ways and support when I needed.

I would also like to thank the staff and students in the Department of Language and Linguistics at UMIST, who supported me during my stay in Manchester, especially:

Bill Black for being my supervisor,
Kristiina Jokinen for her good advice and support, and
Charles Boisvert for his help and ideas.

\textsuperscript{1} The University of Manchester Institute of Science and Technology.
1 INTRODUCTION

The purpose of this thesis is to investigate some problems in the analysis of Czech for machine translation. Czech is a language with "free word order", in the sense that the order of the main grammatical functions in the sentence, such as subject, verb and object, may vary greatly.

Consider the following two sentences:

(1) Žák napsal DOPIS.²
   PUPIL(AM,S,Nom) WRITE(AM,S,Past) LETTER(IM,S,Acc)
   - The pupil wrote a letter.

(2) DOPIS napsal ŽÁK.
    LETTER(IM,S,Acc) WRITE(AM,S,Past) PUPIL(AM,S,Nom)
    - A pupil wrote the letter.

As we can see from these examples, changing the word order does not affect the distribution of grammatical relations or of semantic roles (actor, patient, addressee, etc.), since these are indicated mainly by morphological means. In fact, any of the six logically possible permutations of the three words Žák, napsal, dopis is a grammatical Czech sentence with Žák as the subject and dopis as the object.

On the other hand, although the permutations have the same semantic roles and the same grammatical relations, they are by no means equivalent. In particular, they have a different topic-comment structure. In (1), Žák is interpreted as topic (given information) and dopis as comment (new information). In (2), we find the reverse pattern. Note also that there is no definite or indefinite article in Czech that can be used to indicate whether a noun phrase represents given or new information. If we compare the Czech sentences in (1) and (2) to their English equivalents, we see a striking difference: in the English sentences, the basic word order is the same in both cases (and mainly carries information about grammatical relations), whereas the signalling of given-new information is done by means of articles (the definite article indicating given information and the indefinite article new information).

Thus, we can say (somewhat simplified) that Czech uses word order to indicate the given-new distinction and grammatical morphemes to indicate grammatical functions, while English uses word order to indicate grammatical functions and grammatical morphemes (articles) to indicate the given-new distinction.

² The sentences are followed first by a literal English translation using the abbreviations listed in the beginning of this essay and secondly by an idiomatic English translation. Capital letters indicate the intonation centre of the sentence, by which is meant the element in a sentence that normally carries the main sentential stress. In Czech, like in most languages with a free word order, there is a strong tendency to arrange the sentence so that the intonation centre falls on the last word of a sentence. In other words, normal intonation in Czech appears when the intonation centre is on the last element in a sentence. Other positions of intonation centre are in Czech considered as emotionally marked.
In order to do machine translation from Czech to a language such as English, we must therefore face the following two problems in the analysis of Czech:

(i) We must be able to cope with the grammatically free word order in parsing, i.e. we must be able to assign the correct grammatical functions to different phrases despite the fact that they may occur in virtually any order,

(ii) We must be able to assign species (definite or indefinite) to noun phrases despite the fact that there is no overt marking of this feature by means of articles.

The assignment of grammatical functions is needed in order to get the right word order in the English translation. The assignment of species to Czech nouns is needed for the choice of article in the English translation. In the following I will refer to the first problem as the parsing problem. The second one I will refer to as the problem of species assignment or definiteness marking.

The purpose of the present study is to investigate these two problems in the analysis of Czech for machine translation, as well as propose solutions and implement these solutions in a running system. An important restriction in the scope of this study is that, although the work has been motivated by the perspective of machine translation, I will not be concerned with machine translation as such (i.e. I will not discuss the actual generation of, say, English sentences from the representations outputted by the analysis system). It is my hope that the ideas presented in the thesis, as well as the implementation of these ideas in an analysis module for Czech, can be integrated in a machine translation system, but this is something that falls outside the scope of this study.

The thesis is structured as follows. Chapter 2 provides the theoretical background for the study as well as an orientation about Czech grammar. Chapter 3 is devoted to methodological questions and chapter 4 contains my analysis of the parsing problem and the species assignment problem. Chapter 5 describes the implementation of a system for analysis of Czech, based on the proposals in chapter 4. Chapter 6 contains test results and chapter 7 presents conclusions and suggestions for further work.

Two appendixes are added at the end. Appendix I contains the test runs from the implemented system. Appendix II includes an overview of the declensions of nouns, adjectives and pronouns and the conjugations of verbs in Czech.
2 BACKGROUND

In this background chapter, I will begin (in section 2.1) by discussing categorial grammar and its use in parsing, since this is the grammatical formalism that I have chosen for the parsing system.

In section 2.2 I discuss the notions of topic, comment and focus which are important for the problem of species assignment. In particular, I introduce the theory of Functional Sentence Perspective (FSP), which will be used in the analysis of topic-comment structure and species assignment in chapters 4 and 5.

Section 2.3 gives a brief introduction to Czech grammar, focusing on the aspects that are central to this study, and section 2.4 contains a discussion of definiteness marking in English in order to determine the distinctions that are relevant from the machine translation perspective.

Section 2.5, finally, discusses some previous implementations of relevance to this work.

2.1 PARSING WITH CATEGORIAL GRAMMAR

2.1.1 INTRODUCTION

Categorial grammars (CG) have a long and distinguished history, that is rooted in the mathematical logic of Frege (1879, 1891, 1892) and Ajdukiewicz (1935), where the categories were semantic categories in contrast to later CG, where the categories have a syntactic component. The first attempt to develop CG for the description of a natural language was done by Bar-Hillel (1953), in the way of computational treatments of syntax (i.e. for use in machine translation).

Another important contribution is given by Lambek (1958) with a mathematical perspective on the syntactic structure that includes different algebraic operations in order to combine the syntactic categories. Ades & Steedman (1982), Steedman (1985, 1988) and Bach (1984, 1988), among others, represent recent interest in CG, arising from integration of semantic and syntactic analysis.

The categorial grammar formalism is described in the following subsections, listing the basic formalism and then some extensions. A description of chart parsing in the categorial grammar environment closes this section.

2.1.2 BASIC CATEGORIAL GRAMMAR FORMALISM

The original categorial grammar, used by Ajdukiewicz and Bar-Hillel, is defined in terms of two basic, syntactically and semantically complete, categories: names (n), referring to entities, and sentences (s). All other categories are called functor categories and are defined in terms of how they combine with other categories to form complete expressions. Formally, the full set of categories (given a finite set of basic categories) is defined as follows:

1. Every basic category is a category.

2. If X and Y are categories, then X/Y and XY are categories.
In CG every lexical entity in a language is assigned one or more categories, where the category may be of the basic atomic type or a functor category defined in terms of other categories. In the CG notation, the category of expressions that combine with a preceding expression of category Y to yield an expression of category X is written X/\ Y, where Y is said to be the Argument and X the Result or Value. The backward leaning of the slash ('\') indicates that the argument is to precede the functor syntactically. For example, an intransitive verb in English can be categorised as s/vn, i.e. the verb combines with a preceding name to form a sentence (e.g. 'John runs'). By contrast, the category of expressions that combine with a succeeding expression of category Y to yield an expression of category X is written X\Y. Here, the slash '/' indicates a succeeding argument. In the following I prefer to use the terms Argument and Result.

Two combinatory rules of functional application are used to combine functors with their arguments:

Functional application rules (A):

- Forward application - permits a functor category to combine with an appropriate argument to its right.

  \[\text{Result} / \text{Argument} \rightarrow \text{Argument} \rightarrow \text{Result}\]

- Backward application - permits a functor category to combine with an appropriate argument to its left.

  \[\text{Argument} \rightarrow \text{Result} / \text{Argument} \rightarrow \text{Result}\]

Figure 2.1 shows an example of a structural derivation in categorial grammar. The notation commonly used for derivations includes:

- the string of words,
- the lexical categories placed under the corresponding words,
- a horizontal line indicating rule application, where the extension of the line shows the scope of the rule application and the rule applied is indicated at the end of the line:
  > for forward application
  < for backward application.
- labelling of the line by the result category.

```
Eva reads a book
n  (s\n)n  n/n  n
---------- A
n
---------- A
s\n
----------< A
s
```

Figure 2.1: Structural derivation in categorial grammar.
One of the distinctive features of CG is that, because information about syntactic combination is contained in (functor) categories, very few syntactic rules are needed in the grammar. In fact, the simplest categorial grammar contains only the rules of functional application, given above.

There are several notational systems in use for CG. The system presented here is usually referred to as "Steedman notation" and is used as the standard notation in Wood's (1993) textbook.

2.1.3 EXTENDING THE SET OF CATEGORIES

As Wood (1993) points out, the original, semantically motivated, minimal set of atomic categories (given above) is not sufficiently expressive for any sort of serious linguistic description. It is clearly necessary on both semantic and syntactic grounds to allow both n (noun) and np (noun phrase) as distinct atomic categories, thus extending the basic set of categories to three.

With this extension, restrictions may be put on noun phrases so that any number of adjectives can precede a noun, but only one determiner. The distinction is shown in figure 2.2, where in (a) the minimal set of CG categories is used and in (b) the extended formalism.

(a) that small hungry dog
n/n n/n n/n n

(b) that small hungry dog
n/n n/n n/n n

----------- => A
n
----------- => A
n
----------- => A
n
----------- => A
n
Figure 2.2: Structural derivation:
(a) using the minimal set of categories.
(b) using the extended set of categories.
2.1.4 EXTENDING THE SET OF RULES

Usually the basic functional application rules are also not enough for parsing and additional rules are needed. In this section three additional types of rules will be presented - the associativity rules, the functional composition rules and the 'neutral slash' rules - which will all be used in the following work. Of these rules, the former two derive from the so-called Lambek calculus (Lambek 1958).

**Associativity**

Wood (1993:37), discussing the work of Lambek (1958), defines the associativity rule as a unary rule, "which says that a function with two arguments, one on either side, can combine with them in either order". The rules follow below:

Associativity rules (AS):

- \((\text{Result}/\text{Argument1})/\text{Argument2} \rightarrow (\text{Result}/\text{Argument2})/\text{Argument1}\)
- \((\text{Result}/\text{Argument2})/\text{Argument1} \rightarrow (\text{Result}/\text{Argument1})/\text{Argument2}\)

Figure 2.3 shows a structural derivation, (a) using only the basic functional application rules, compared to (b) a structural derivation using the functional associativity rules defined above.

(a) John likes Jane
n \(\text{(s\text{\textbackslash n})/n}\) n
\(\text{s\textbackslash n}\)
------------------------\(\rightarrow\) A
s\(\text{n}\)
------------------------< A
s

(b) John likes Jane
n \(\text{(s\text{\textbackslash n})/n}\) n
\(\text{(s\text{\textbackslash n})\text{\textbackslash n}}\)
------------------------\(\rightarrow\) AS
s\(\text{n}\)
------------------------< A
s
------------------------\(\rightarrow\) A
s

**Figure 2.3:** Structural derivation:
(a) without the functional associativity rules,
(b) using the functional associativity rules.

**Functional composition**

Functional composition, also introduced by Lambek (1958), may be used for combining a functor category with another functor category, where the result category of one category is the argument category of the other.

Functional composition rules (C):

- **Forward composition**
  \(\text{Result}/\text{Result1} \rightarrow \text{Result}/\text{Argument}\)

- **Backward composition**
  \(\text{Result1}/\text{Argument} \rightarrow \text{Result}/\text{Result1}\)
The functional composition rules apply for example to a noun phrase with multiple adjectives. Compare the two derivations in figure 2.4, where in (a) the derivation is done only by application rules and in (b) by composition rules.

\[
\begin{align*}
\text{(a) the big red balloon} & \quad \text{(b) the big red balloon} \\
\text{np/n} & \quad \text{np/n} \\
\text{n/n} & \quad \text{n/n} \\
\text{n/n} & \quad \text{n/n} \\
\text{n} & \quad \text{n} \\
\text{---------} & \quad \text{---------} \\
\text{A} & \quad \text{C} \\
\text{n} & \quad \text{np} \\
\text{---------} & \quad \text{---------} \\
\text{A} & \quad \text{C} \\
\text{n} & \quad \text{np} \\
\text{---------} & \quad \text{---------} \\
\text{A} & \quad \text{A} \\
\text{np} & \quad \text{np}
\end{align*}
\]

**Figure 2.4:** Structural derivation:
(a) using application rules,
(b) using composition rules.

The difference is that in 2.4(a) application can only work through from right to left, whereas in 2.4(b) composition applies from left to right.

**Neutral slash**

The neutral slash rules introduce another connective - the neutral slash ' | ' - which historically was the first connective used in categorial grammar. In the early works of Ajdukiewicz (1935) categorial grammar was primarily applied to formal languages and no special attention was paid to "word order". The directional connectives were introduced in the 1950s when Bar-Hillel (1953) applied categorial grammar to natural language.

As the name reveals the connective is directionally neutral, i.e. with this connective it is possible to combine functor categories with appropriate arguments regardless of whether the arguments precede or succeed the functor category. That means that we have to introduce new rules for application, associativity and composition, which have the same form as the rules presented earlier, except that the connectives are replaced by the neutral connective.

Notice that the associativity rules now includes only one rule, because of the neutral connective. The rule is listed below, followed by rules for application, and composition.

**Associativity rule (NAS):**

- \((\text{Result} \mid \text{Argument1}) \mid \text{Argument2} \rightarrow (\text{Result} \mid \text{Argument2}) \mid \text{Argument1}\)

**Functional application rules (NA):**

- **Forward application**
  \(\text{Result} \mid \text{Argument} \rightarrow \text{Result}\)

- **Backward application**
  \(\text{Argument} \rightarrow \text{Result} \mid \text{Argument} \rightarrow \text{Result}\)
Functional composition rules (NC):

- **Forward composition**
  \[ \text{Result} | \text{Result1} \quad \text{Result1} | \text{Argument} \rightarrow \text{Result} | \text{Argument} \]

- **Backward composition**
  \[ \text{Result1} | \text{Argument} \quad \text{Result} | \text{Result1} \rightarrow \text{Result} | \text{Argument} \]

In English one can use the neutral slash rules for example to combine sentences with sentence modifiers (see figure 2.5).

\[
\begin{align*}
\textbf{(a)} & \quad \text{Unfortunately} \quad \textit{Mary did not come,} \\
& \quad s \mid s \quad s \quad \longrightarrow \quad \text{NA} \\
& \quad s \\
\end{align*}
\]

\[
\begin{align*}
\textbf{(b)} & \quad \textit{Mary did not come} \quad \text{unfortunately.} \\
& \quad s \quad s \mid s \quad \longrightarrow \quad \text{NA} \\
& \quad s
\end{align*}
\]

**Figure 2.5**: Structural derivation using neutral slash rules.
(a) Forward application.
(b) Backward application.

For Czech, representing a language with a free word order, the neutral slash rules are even more useful, which will be shown in more detail in chapter 4.
2.1.5 CHART PARSING

Parsing involves building syntactic representations of natural language expressions, e.g. in the form of tree structures. Several techniques can be used and chart parsing is one of them. In Gazdar & Mellish (1989), where tables and charts are described using phrase structure grammar, a chart is described as a set of edges. The textbook is very useful as an introduction to the chart parsing technique.

In my work I use a categorial grammar chart parser, which is simply a parser where the parsing is done by building a chart adapted to the categorial grammar formalism.

The characteristics of building a chart are:

1. Numbering of word and phrase positions in a sentence; in other words, numbering of the nodes in the syntactic tree.

2. Drawing of arcs between the nodes which are positions, labelled with the syntactic category of the word or phrase, which simply corresponds to building a syntactic tree.

An example given in the categorial grammar formalism is shown in figure 2.6.

![Figure 2.6: Building a chart.](image)
The chart parsing procedure that takes place can be described as follows (the figure 2.6 is used as example):

1. Look up the words in the lexicon, i.e. assign syntactic categories to them.
   - eva - np
   - reads - s\np\np
   - a - np/n
   - book - n

2. Assign edges to these categories and store the edges in a database:
   - edge(0,1,np).
   - edge(1,2,s\np\np).
   - edge(2,3,np/n).
   - edge(3,4,n).

3. Try to combine the edges to phrases, i.e. apply the combinatory rules:
   - edge(2,4,np).
   - edge(1,4,s\np).
   - edge(0,4,s).

Several different strategies may apply in chart parsing and one of them is the strategy using a bottom-up, breadth-first procedure. There are (at least) two possible ways to use this particular strategy in chart parsing. Either, first looking up all the words in the lexicon and then trying to combine them to phrases - Algorithm 1, or doing the two steps simultaneously - Algorithm 2.3

Algorithm 1:

1. Take a word from the input string and add an edge to the database for every category the word obtains. Continue doing so for every word in the string until the end of string is reached.

2. Try to combine the edges to phrases, by applying the CG rules, starting at the beginning of the chart.

Algorithm 2:

1. Take a word from the input string and add an edge to the database for every category the word obtains.

2. Look after if the asserted edge can combine with an edge to its left in the chart.

3. Take the next word in the string and apply the steps 1 and 2. Continue so until the end of string has been reached.

I use a strategy similar to Algorithm 2 in my implementation of the CG-chart parser. But more about that in chapter 5.

---

3 A note of warning is that the algorithms are not described in every detail.
2.2 TOPIC, COMMENT AND FOCUS

2.2.1 INTRODUCTION

As mentioned in the introduction (chapter 1), the word order in Czech is determined mainly by topic-comment structure. In this section the terms topic, comment and focus will be explained in the form of an overview of present theories. Special emphasis lies on the theory of Functional Sentence Perspective, which closes this section.

Even if the terminological confusion in this area is rather extreme, the concept pairs "topic-comment" and "theme-rheme" can be considered as synonyms. To avoid further confusion in the terminology I prefer to refer to those pairs only as "topic-comment".4

2.2.2 GENERAL DEFINITIONS

Allwood & Andersson (1988) define the topic as "what one talks about" and the comment as "what one says" in a sentence. Dahl (1974:4) gives similar definitions: "the topic of a sentence (or whatever name it is given) is 'what the sentence is about' and the comment is 'what is said about the topic'." Both works also discuss the possibility of explaining the topic and comment in terms of "givenness" and "newness", i.e. that the topic denotes the old information, something already known from the context, while the comment stands for the new information.

According to these definitions, the topic-comment structure in the example sentence below is distributed as follows:

(1) John eats strawberries.

   TOPIC  COMMENT

Dahl (1974:7) further explains the topic-comment structure as a reference relation. "To be able to say something about an entity, we must first pick it out for our listener, in other words we must refer to it." The topic must be a referring expression, which also implies the restriction of definiteness on topics, i.e. that the elements in topic must be grammatically definite. According to this reasoning, the sentence (1) above is analyzed as having two parts, one where we select an individual - John - and one where we say something about him - that he eats strawberries.

Often a question is used to make it easier to decide what element(s) in a sentence belong to the topic and comment respectively, with respect to the "given-new" distinction.


   TOPIC  COMMENT

4 Exception is made in quotations, where the terms used in the original version are retained.
The question as an example of an applicable context is important here. If one changes the context of a sentence, the information structure can change, as shown in the example (3).

    COMMENT TOPIC

Allwood & Andersson (1988) also introduce the terms "focus" and "background" besides the terms "topic" and "comment". Focus is defined as being a part of the comment, in speech often indicated by a strong stress. The term "background" is used as a complement to focus for everything else in a sentence that is not in focus.

The distinction between topic-comment and focus-background levels in a sentence structure is made in that topic and comment concerns the structure of a sentence with respect to "what the sentence is about" and "what is said about the topic", respectively; focus and background applies to sentence structure with regard to what is said or stated and what is presupposed as known. In Allwood & Andersson's words, the thing one states (focus) is naturally a part of "what is said" and "what one talks about" (topic) is naturally a part of that which is not stated (background).

In many sentences, the two concept pairs may coincide:

    COMMENT TOPIC
    FOCUS BACKGROUND

In other cases a tripartite structure is the result:

    TOPIC/COMMENT
    BACKGROUND FOCUS

The concepts are not treated in the same way in all theories. For example in the work of the Prague school the concept pairs "topic-comment" and "background-focus" are not kept apart. Rather we talk about a sentence structure divided into a topic part and a focus part. Thus, one would say that in the example (6) below, 'John eats' is topic and 'strawberries' is focus, whereas in (7) 'John' is topic and 'eats strawberries' is focus.

    TOPIC FOCUS

    TOPIC FOCUS

In Firbas (1974:13) an additional concept is used - "transition", which is described as involving "... elements that actually belong to the rheme, but occur at its periphery and in this way intermediate between theme and rheme."

(8) John eats strawberries.
    TOPIC TRANSITION COMMENT
A conclusion one may draw here is that as mentioned in the beginning of this section, there is some terminological inconsistency in this area. Only the concepts topic-comment and theme-rHEME are used in the same way. Another regularity, according to Dahl (1974), can be found in that in all languages (i.e. so far investigated) the sentence stress constitutes a universal means of signalling topic-comment structure. However, for the theory of the Prague school, focus is defined as a sentence part that is contextually unbound (contextual boundness will be explained later in more detail). The definition in Dahl (1974:2) "... the focus of a sentence is normally the carrier of the main sentential stress." holds only when the focus part contains one sentence element. In other words, in the terminology of the Prague school the focus always includes the carrier of the main sentential stress, but other sentence elements may also be part of the focus.

What I find rather confusing with all these theories, besides the terminology, is the role of context in decisions of a sentence's topic-comment structure. Often the preceding context is concerned only as an example of a context for which the given sentence is "applicable". In other words, a sentence topic-comment structure is described for an isolated sentence. What I mean, is that because the topic-comment distribution in a sentence is dependent of the context in which it occurs, it is rather confusing to have a topic-comment structure for an isolated sentence. In my opinion, deeper analysis, maybe in the form of a sentence used in several different (longer) contexts or an analysis of a longer text, is more appropriate here. I only found one work including an analysis of a text, namely Firbas (1992).

2.2.3 THE THEORY OF 'FUNCTIONAL SENTENCE PERSPECTIVE'

Introduction

A theory developed by the Prague School, in Czech "aktuální členění větné", is known in English as: "functional sentence perspective" (FSP). Some ideas regarding this theory have already been mentioned in the previous section. Here follows a further description.

One of the basic questions concerning FSP is the question if a sentence is divided into two parts (topic-comment) or if the division is into a number of parts (so called 'communicative dynamism').

---

5 Vilém Mathesius (1882 - 1945) began to study these problems in Czechoslovakia in the period between the two wars. He used the term 'aktuální' in reference to the organisation of the sentence as implemented and/or perceived at the moment just described. Like French 'actuel' and German 'aktuell', Czech 'aktuální' conveys the meaning of 'being of immediate interest and concern'. Unfortunately, this meaning does not associate so readily with English 'actual', which primarily suggests the meaning of 'existing in fact', 'real'. This was decisive for the introduction of the term 'functional sentence perspective', which is now regarded as an equivalent of 'aktuální členění větné'.
The dichotomy of topic and comment is defined by Danes et al (1987) as:

- **complementary**, i.e. topic and comment cannot exist without each other and every sentence item can be placed either in the topic or in the comment;

- **systemic**, i.e. FSP is a part of the language system, which implies that it concerns most of the sentences in the language; sentences that are not possible to divide into topic and comment or where the division is less clear are considered as unusual or special;

- **universal**, in that FSP is a characteristics common to all indo-european languages and many other non-indoeuropean languages.\(^6\)

In this theory, focus is defined in terms of contextual boundness as the unconnected part of a sentence. Thus, in this terminology, a sentence structure is divided into topic and focus parts rather than topic and comment. What is meant by contextual boundness and communicative dynamism is explained in the following sections.

**Context and contextual boundness**

Contextual boundness is defined with respect to the given-new distinction (see section 2.2.2), i.e. an element is contextually bound if it is known from the previous context, or contextually unbound if it brings new information to the context. Also, given elements that are not explicitly mentioned in the previous text, but are more or less related to an element mentioned earlier, are considered to be contextually bound. Contextual boundness is also connected with the given situation, not only with the preceding verbal context.

According to this the following types of context, where a certain sentence element is contextually bound, can be distinguished:

(a) **Linguistic context** - where the known element from the previous context is usually not repeated literally. It can for example be further expanded with new information, expressed by a pronoun or similar means. For the decision of the boundary between topic and comment parts it is relevant with one, possibly several sentences occurring before the given sentence. With growing distance, the importance of linguistic context becomes weak, and as a rule it does not extend further than to the limit of a paragraph.

(b) **Situational context** - covers facts that are not mentioned in the linguistic context, but which are nevertheless known to the author and addressee from the situation, under which the utterance arises. We can then distinguish between: - immediate situational context, and - extensive situational context

(c) **Experiential context** - the author assumes that he/she shares together with the addressee common knowledge about the "world" that he/she considers as known in the communication act.

\(^6\) Languages can, however, differ in expressing FSP by different means.
Communicative dynamism

As mentioned in the introduction to this section, FSP does not concern only the dichotomy of topic and focus, but also a more fine division in a scale or hierarchy of "communicative dynamism" (CD). According to Sgall et al (1980), CD can be considered as "deep word order" phenomena, i.e. a scale or a hierarchy of elements ordered in such a way that a point can be determined that separates the topic from the focus in a sentence. Often the verb is considered as the main boundary between topic and focus. Sgall et al (1980) claim that in the surface word order it is not always possible to find such a point, especially because for example the verb in Czech is usually "moved" to the second place in the sentence, even if it is followed by contextually bound elements.

In other words, what communicative dynamism actually means is to which extent sentence elements contribute towards further development of communication. In the scale of communicative dynamism the elements are ordered from left to right, where the rightmost element has the highest degree of communicative dynamism, i.e. contributes most to the development of communication.

The connection between the topic-focus dichotomy and the CD hierarchy makes it possible to distinguish a proper topic (i.e. the least dynamic member of the topic part) and a proper focus (i.e. the most dynamic member of the focus part) in a sentence.

The semantic content of a sentence element is the decisive factor of placement in the hierarchy of communicative dynamism, i.e. the distribution of degrees of CD is especially dependent upon the ordering of the types of verb complements (actants). For a given language it is possible to construct a hierarchy of actants, called "systemic ordering" (SO)\(^7\), that may be generally expressed in the grammar of that language and which is the basis of CD in the sense that the degree of CD of contextually unbound members is directly decided by the systemic ordering, whereas contextually bound members are considered less dynamic, i.e. every topic element has a lower degree of communicative dynamism than any focus element.

In other words, the hierarchy of CD within focus is not qualified by the speaker's choice, but by the type of verb complements. On the other hand, the choice of degrees of CD within topic may be affected by:

- linguistic factors, e.g. actor is more often chosen as the proper topic of a sentence than other complements; or

- factors connected with the hierarchy of elements in the common stock of information, i.e. with higher or lower degree of activation according to if the given object was mentioned in the immediate previous context (e.g. the least dynamic element in a sentence is often the element that in the immediate previous context was contextually unbound), or if it is the object of topical interest of the participants in the conversation, or if its topicalisation, given by a more distant context, is not yet so high; or

- factors concerning the text structure, e.g. contrast, sentences following each other having common topic, etc.

---

\(^7\) The systemic ordering of some of the main complements is identical for many languages, having the form: actor - addressee - patients.
Word order and intonation

Word order and intonation are the basic means of expressing FSP in Czech. As mentioned in the introduction (see chapter 1), normal intonation in Czech appears when the intonation centre\(^8\) is on the last element in a sentence. Other positions of the intonation centre are in Czech considered as emotionally marked.

As explained in the previous section concerning the scale of CD in a sentence, the deep word order within the focus is determined by the systemic ordering, i.e. an ordering of the complements of the verb. Here follows the systemic ordering for Czech from Sgall et al (1980):

\[
\begin{align*}
\text{actor} & \\
\text{time} & \quad (\text{when}) \\
& \quad (\text{since when}) \\
& \quad (\text{till when}) \\
& \quad (\text{how often}) \\
& \quad (\text{how long}) \\
\text{place} & \quad (\text{where}) \\
\text{manner} & \\
\text{measurement} & \\
\text{degree} & \\
\text{instrument} & \quad (\text{means}) \\
\text{direction} & \quad (\text{which way}) \\
\text{addressee} & \\
\text{origin} & \\
\text{direction} & \quad (\text{from where}) \\
\text{patiens} & \\
\text{direction} & \quad (\text{where}) \\
\text{result} & \\
\text{condition} & \\
\text{purpose} & \\
\text{cause} & 
\end{align*}
\]

As mentioned before, the elements in the scale of CD are ordered from left to right, i.e. the less dynamic element stands to the left of an element that is more dynamic. In other words, in the case of systemic ordering, which determines the order of actants within the focus part of a sentence, the ordering of actants is from left to right, e.g. actor before patiens. The actor is then considered as less dynamic, i.e. it contributes less to the development of communication, than the patiens.

In accordance with these definitions, Sgall et al (1980) make a distinction between a primary and marked word order with respect to the systemic ordering within the focus part:\(^9\)

(a) **Primary order** is given by the systemic ordering of the actants, where
- all actants belong to the focus, or
- the least dynamic, i.e. the leftmost of them, is contextually bound.

---

\(^8\) By the intonation centre of a sentence is meant the element in a sentence that carries the main sentential stress.

\(^9\) Notice that this is only valid when a sentence has normal intonation.
(b) **Marked order** occurs when the less dynamic actant(s) is contextually bound, i.e. the actant(s) does not follow the systemic ordering.

In other words, the surface structure of a sentence with a primary word order corresponds directly to the scale of CD, whereas a sentence with a marked word order does not correspond to the scale.

The topic-focus structure of a Czech sentence is analysed by Sgall et al. (1980) in the following rules:

1. If a sentence has normal intonation, i.e. the intonation centre is at the end of the sentence, the following rules apply:

   (i) All complements standing before the verb are connected to the context. An exception is the subject, which can be connected or not.

   (ii) For the complements that follow the verb, the limit between the topic and focus lies in front of such a complement, if that complement and all other following complements in the sentence are ordered in accordance with the systemic ordering.

   (iii) The verb is ambiguous in every sentence with normal intonation. Exceptions are those sentences where the verb is the carrier of the intonation centre. A carrier of the intonation centre is always a part of the focus.

2. Otherwise, i.e. when the intonation centre is not at the end of a sentence, the following is valid:

   (iv) All complements standing after the intonation centre are part of the topic.

   (v) The carrier of the intonation centre is unambiguously part of the focus.

   (vi) For the elements standing before the intonation centre the rules (i) to (iii) are valid. Where we talk about the last element is now meant the carrier of the intonation centre even if it is not at the end of the sentence.

A comment has to be made about rule (ii): the word order of complements that violate this condition are not allowed to belong to the focus. In other words, in a sentence where the complements following the verb are ordered in such a way that there is a complement A followed by a complement B which in the systemic ordering precedes A, then the complement A is always part of the topic. This is also valid for all the complements that are less dynamic than the complement A, i.e. that stand to the left of the complement A. Another thing is that after applying this rule, the focus contains only the potential members (except the last word, which is always in the focus, i.e. in the case of normal intonation).
Consider the following sentences:

(1) Chlapec koupil dívce ZMRZLINU.
     BOY(AM,S,Nom) BUY(AM,S,Past) GIRL(F,S,Dat) ICE-CREAM(F,S,Acc)
(a) - A/The boy bought AN ICE-CREAM to a girl.
(b) - A/The boy bought AN ICE-CREAM to the girl.

(2) Chlapec koupil zmrzlinu DÍVCE.
     BOY(AM,S,Nom) BUY(AM,S,Past) ICE-CREAM(F,S,Acc) GIRL(F,S,Dat)
 - A/The boy bought the ice-cream TO A GIRL.

We assume that the sentences have normal intonation. According to the rules of topic-focus structure in a sentence with normal intonation, the verb complements that precede the verb are always contextually bound, except the subject which is ambiguous. In other words, the sentence subject (actor; 'chlapec') in both examples may be contextually bound or unbound, depending on the previous context.

As regards the verb complements that follow the verb, in example (1) the word order of the complements coincide with the systemic ordering, i.e. addressee ('dívce') before patients ('zmrzlinu'). According to rule (ii), the limit between topic and focus lies then in front of the addressee, but the addressee is still ambiguous. Thus, the two actants may both belong to the focus - (a) - or the addressee is contextually bound (depending on the previous context) - (b). The patients is always contextually unbound, i.e. it is the carrier of the main sentential stress in the sentence, which is always in focus.

In example (2), on the other hand, the actants do not follow the systemic ordering, which means that the patients ('zmrzlinu') is always in the topic, no matter in which context the sentence appears. The boundary between topic and focus lies in front of the addressee ('dívce'), which is the carrier of the intonation centre, i.e. always in the focus.

In other words, the actants in the sentence (1) are ordered according to the primary word order, whereas in the sentence (2) the actants follow the marked word order.

**Initial sentences**

Sentences that function as an introductory sentence in a text, may behave as sentences that occur in the middle of a text, i.e. they can be linked to the title of a text or to the previous text unit (paragraphs, chapters etc.). Or, if we assume that there is no previous linguistic context included, the sentence is considered as contextually unbound, i.e. the sentence does not include any topic element. The relation to the situational or experiential context then becomes more important.

According to Sgall et al. (1980), in Czech sentences without any contextually bound part the first least dynamic element is the verb and the most dynamic is the member appearing on the scene or whose existence is asserted. In other words, the verbs that express "appearance on the scene" or "existence" as a rule have lower degree of CD than their subjects. In a sentence with normal intonation the first actant of these verbs (i.e. actor) follows the verb and is the carrier of the intonation centre, as shown in examples (3) and (4).
(3) **Přišlo** **JARO.**
COME(N,S,Past) SPRING(N,S,Nom)
- Spring came.

(4) **Objevil se nějaký** **MLADÍK.**
APPEAR(AM,S,Past) refl. SOME(AM,S,Nom) YOUNG-MAN(AM,S,Nom)
- Some/A young man appeared.

However, verbs of this type may have more or less concrete lexical meaning and may also be connected with different types of adverbials. The sentences obtain a similar structure, i.e. the subject, as the carrier of the intonation centre, follows the verb and the adverbial is now the less dynamic element (i.e. it is considered contextually bound). Examples are given in sentences (5) and (6).

(5) **V Praze žije má SESTRA.**
IN PRAGUE(F,S,Loc) LIVE(3PS,Pres) MY(F,S,Nom) SISTER(F,S,Nom)
- My sister lives in Prague.

(6) **V parcích by měly kvést RŮŽE.**
IN PARK(IM,P,Loc) cp. SHOULD(F,P,Past) BLOOM(INF) ROSE(F,P,Nom)
- Roses should bloom in the gardens/parks.

Some of the verbs that express appearance on the scene or existence, may have the function of introducing an object, but also other functions. The distinction between the different functions of a verb is made in that in the sentences without any contextually bound part the verb precedes the actor in the scale of CD (as in examples (5) and (6)), whereas the word order actor - verb is considered as directly corresponding to the scale of CD in those sentences where the actor is contextually bound, which is shown in the examples (7) and (8) below.

(7) **Matka žije v PRAZE.**
MOTHER(F,S,Nom) LIVE(3PS,Pres) IN PRAGUE(F,S,Loc)
- Mother lives in Prague.

(8) **Růže by měly růst pod oknem.**
ROSE(F,P,Nom) cp. SHOULD(F,P,Past) GROW(INF)
UNDER WINDOW(N,S,Ins)
- The roses should grow under a window.

In both sentences the actor precedes the verb. According to the definitions given above, the actor is then contextually bound and the verbs do not introduce anything. In the sentence (7) 'mother' specifies 'my mother' or 'our mother' and in the sentence (8) 'roses' concern 'our roses' or 'that roses' (i.e. which were mentioned previously).

A conclusion here is that no problem arises in applying the rules of topic-focus structure (given in the previous subsection) to a sentence that introduces a text.
2.2.4 SUMMARY

The characteristics of the theory of Functional Sentence Perspective can be summarised into the following statements:

- Topic and focus are defined in terms of contextual boundness. The part of a sentence that is contextually unbound, i.e. new to the context, belongs to focus, whereas the contextually bound part, i.e. the part which is already known from the previous context, belongs to topic.

- The distribution of topic and focus in a sentence is determined by the scale of communicative dynamism. The semantic content of a sentence element is the decisive factor of placement in the hierarchy of communicative dynamism, i.e. the distribution of degrees of CD is especially dependent on the ordering of the types of verb complements (actants). For a given language it is possible to construct a hierarchy of actants, called "systemic ordering" (SO).

- In Czech, the word order and intonation are the basic means for determining the topic-focus structure of a sentence. Thus, in a sentence with normal intonation, the verb is considered as being the boundary between the potential topic and focus parts. The complements that precede the verb are in the topic and the complements that follow the verb are in the focus. 'Systemic ordering', i.e. the ordering of verb complements, is the norm for the word order within focus. The complements that violate this ordering are considered as contextually bound in every context of the sentence. An element in a sentence that is the carrier of the main stress is always part of the focus.
2.3 CZECH GRAMMAR

2.3.1 GENERAL CHARACTERISTICS OF CZECH

This section gives a brief introduction to the Czech language, starting with alphabet and phonology, continuing with a short description of the grammatical categories, word order properties and other features.

The Czech alphabet has 31 characters and uses the Latin alphabet extended with additional 'accents' (diacritic signs) put above vowels and certain consonants:

- the čárka "´" and kroužek "˝" are used only with vowels and indicate length;
- the háček "˘" is used for certain consonants - c, d, t, n, r, s, z - and the single vowel - e.

Intonation does not comprise such a great number of tunes and patterns as in English. Stress is invariably on the first syllable of a word. The contrasts between short and long vowels, and voiced and voiceless consonants are phonetically significant, e.g.

```
    rada - piece of advice    rada [ra:da] - glad
    vila - villa             vila [vi:la] - fairy
    ten  - that              den  - day
    peru - I wash            beru - I take
    kosa - scythe            koza - goat
```

There are practically no homophones in Czech, i.e. words sounding alike but meaning different things.

Czech has no definiteness marking on nouns, i.e. no articles. There are three genders and seven cases for the noun. Adjectives and pronouns must agree in gender, number and case with the noun they qualify. Verbs are conjugated by person and number. Pronoun subjects are normally not expressed, since the verb endings show which person and number is meant.

There are three tenses in Czech - present, past and future - and the lack of tenses is compensated for by aspects.

Negation is formed by particles ne- (used with verbs, nouns and adjectives) and ni- (with negative pronouns and adverbs).

Other significant characteristics of Czech is that there are no functional variations permitting verbs to become nouns, or nouns to be used adjectivally as in English. In other words, parts of speech are easily recognisable by their form, they do not overlap and the functions they have to perform in a sentence are clearly defined.

The order of words is much freer than in English, French or German, and also the oriental languages. As mentioned in the introduction (see chapter 1), it is the case-form with its inflections that convey the meaning, not the word order. In other words, the basic marker of grammatical relations (i.e. subject, direct object, indirect object) is not the word order but rather the morphology. The word order is determined by pragmatic factors such as topic-comment structure.

22
The following sections describe in further detail those parts of Czech grammar that are important and relevant to the topics of this study, i.e. nouns, some pronouns and word order. The facts and most of the examples are taken from Harkins (1968), Heim (1982) and Sova (1962).

Every sentence example in Czech is followed by a literal English translation and then an idiomatic English sentence corresponding to the Czech sentence.10

Appendix II contains a survey of the declension of nouns, adjectives and pronouns, and the conjugation of verbs.

2.3.2 Nouns

Nouns in Czech are inflected for number, case and gender. In this section describes these inflectional features of nouns, together with definiteness.

Number

As in most indoeuropean languages, Czech has two numbers, i.e. singular and plural, e.g.

\[
\begin{align*}
\text{pán} & \ [\text{paːn}] \text{ (gentleman)} & \text{pání} & \ [\text{paːni}] \text{ (gentlemen)} \\
\text{hrad} & \ \text{ (castle)} & \text{hrady} & \ \text{ (castles)} \\
\text{kniha} & \ \text{ (book)} & \text{knihy} & \ \text{ (books),} \\
\text{okno} & \ \text{ (window)} & \text{okna} & \ \text{ (windows)}
\end{align*}
\]

A few nouns which denote paired parts of the body - eyes, ears, hands, feet, etc., retain the old dual declension instead of the plural, e.g.

\[
\begin{align*}
\text{oko} & \ \text{ (eye)} & \text{ocho} & \ [\text{uxo}] \ \text{ (ear)} & \text{ruka} & \ \text{ (hand, arm)} \\
\text{ocí} & \ [\text{ɔʃi}] \ \text{ (eyes)} & \text{uši} & \ [\text{uʃi}] \ \text{ (ears)} & \text{ruce} & \ [\text{rutse}] \ \text{ (hands, arms)}
\end{align*}
\]

---

10 Explanations for the appropriate abbreviations used here can be found in the beginning of this essay.
Case

Czech nouns are inflected according to their grammatical functions (i.e. subject, direct object, indirect object) in the sentence. There are, in total, seven cases in Czech, indicating these sentence functions:

1. NOMINATIVE case indicates the subject or predicate nominative of the sentence. It answers the questions - who? or what?  

   (1) Přijela babička.
       ARRIVE(F,S,Past) GRANDMOTHER(F,S,Nom)  
       - Grandmother came.  

   (2) Jablko je zelené.
       APPLE(N,S,Nom) BE(N,S,Pres) GREEN(N,S,Nom)  
       - The apple is green.

2. GENITIVe case indicates possession among other things. It answers the questions - of whom?, of what?, whose?

   (3) Přijela Josefova babička.
       ARRIVE(F,S,Past) JOSEPH(F,S,Gen) GRANDMOTHER(F,S,Nom)  
       - Joseph's grandmother came.

3. DATIVE case indicates the indirect object.
   It answers the questions - to whom? (are you writing, etc.), to what? (are you listening, etc.).

   (4) Syn píše otcí.
       SON(AM,S,Nom) WRITE(3PS,Pres) FATHER(AM,S,Dat)  
       - The son writes to his father.

4. ACCUSATIVE case indicates the direct object.
   It answers the questions - whom? (do you see, love, hear, etc.), what? (do you see, etc.).

   (5) Syn píše dopis otcí
       SON(AM,S,Nom) WRITE(3PS,Pres) LETTER(1M,S,Acc) FATHER(AM,S,Dat)  
       - The son writes a letter to his father.

\[ \text{11} \] The 'questions' are used in Czech in order to help to determine a noun's case.

\[ \text{12} \] With bold type is indicated the particular word that carries the grammatical function indicated by the case (here the subject).
5. **VOCATIVE** case is used for addressing a person or a thing.

(6) Anno! Syn píše dopis otcí.
    ANNA(F,S,Voc) SON(AM,S,Nom) WRITE(3PS,Pres) LETTER(IM,S,Acc)
    FATHER(AM,S,Dat)
    - Anna! The son writes a letter to his father.

6. **LOCATIVE** or prepositional case is always used with prepositions, it cannot stand alone. It indicates position in space or time, among other things. It answers the questions - *about whom?*, *about what?* (are you speaking).

(7) Syn píše dopis otcí o vnukovi.
    SON(AM,S,Nom) WRITE(3PS,Pres) LETTER(IM,S,Acc) FATHER(AM,S,Dat)
    ABOUT GRANDSON(AM,S,Loc)
    - The son writes a letter to his father about his grandson.

The prepositional case is not the only case that follow prepositions, but it is the only one that cannot stand without them.

7. **INSTRUMENTAL** case indicates the means by which an action is performed. It answers the questions - *with whom?* *with what?*

(8) Syn píše dopis otcí o vnukovi černým perem.
    SON(AM,S,Nom) WRITE(3PS,Pres) LETTER(IM,S,Acc) FATHER(AM,S,Dat)
    ABOUT GRANDSON(AM,S,Loc) BLACK(N,S,Ins) PEN(N,S,Ins)
    - The son writes a letter to his father about his grandson with a black pen.

**Gender**

Gender in Czech is a matter of grammar. It is the ending of a noun that gives the clue to its grammatical gender which is the basis of the declensions. The ending of the noun given in the nominative singular usually indicates its gender. All Czech nouns belong to one of three genders:

- **masculine** - often further divided in animate and inanimate; mostly nouns ending in consonants, ex. [pán] (pa:n) (gentleman), [sešit] (seːʃɪt) (exercise book);

- **feminine** - nouns ending in -a, -e or a consonant, ex. [knihá] (kniːhɑ) (book), [ulice] (uːliːʦe) (street), [kolėj] (koːlɛj) (track);

- **neuter** - ends in -o, -e or -ř, ex. [okno] (oknoː) (window), [moře] (moːɾɛ) (sea), [náměstí] (naːmɐʃtɪː) (square).
Definiteness

As mentioned before, there is no overt marking of species, or definiteness, in Czech. The noun kniha corresponds to book, the book, or a book, depending on the position in sentence:

(9) Karel čte knihu.
    KAREL(M,S,Nom)  READ(3PS,Pres)  BOOK(F,S,Acc)
    - Karel reads a book.

(10) Kniha je zajímavá.
     BOOK(F,S,Nom)  BE(3PS,Pres)  INTERESTING(F,S,Nom)
     - The book is interesting.

According to Sgall et al (1980), with regard to the FSP-theory (see section 2.2.3), contextually bound nouns, i.e. nouns included in the topic part of a sentence, are usually treated as definite or generic. If there is (in a specific context) another type of noun in the topic, then it is explicitly indicated by lexical means, i.e. indefinite pronouns. On the other hand in the focus, the nouns are usually connected with so called specific delimitation (= indefiniteness). A demonstrative pronoun may be used to indicate definiteness. Consider the following sentences:

(11) Řeka protéká ROVINOU.
     RIVER(F,S,Nom)  RUN-THROUGH(3PS,Pres)  PLAIN(F,S,Ins)
     - The river runs through a plain.

(12) Rovinou protéká ŘEKA.
     PLAIN(F,S,Ins)  RUN-THROUGH(3PS,Pres)  RIVER(F,S,Nom)
     - A river runs through the plain.

In the sentence (11) the noun 'řeka' (= river) is in the topic of that sentence, i.e. it is contextually bound. In other words the noun is definite, which is also expressed in the corresponding English translation, where the noun has a definite article. The noun 'rovinou' (= plain) on the other hand is placed in focus, which is indicated with the placement in the intonation centre of that sentence. This noun is considered as indefinite in Czech and in the corresponding English translation the noun has the indefinite article. In the sentence (12) the situation is the reverse. Now the noun 'rovinou' has definite species, whereas the noun 'řeka' is in focus and has indefinite species.

A sentence with the same topic-focus distribution as in (1), but with the delimitation properties as in (2), then obtains the following structure:

(13) Nějaká řeka protéká
tou ROVINOU.

SOME(F,S,Nom)  RIVER(F,S,Nom)  RUN-THROUGH(3PS,Pres)

THAT  PLAIN(F,S,Ins)

- Some/A river runs through that/the plain.

Pronouns are described in the following section.
In other words, the noun placed in the topic part of the sentence (13) now has an indefinite species because of the indefinite pronoun, and the noun in focus, which is normally considered as indefinite is now modified by the demonstrative pronoun and has a definite interpretation. The conclusion is that the definiteness distribution in Czech sentences is not only decided by the topic-focus structure of a sentence; pronominal determiners may also play a role.

The demonstrative pronoun ten (this, that)\textsuperscript{14} is sometimes used where the definite article 'the' would be used in English, which is shown in sentence (13). Another example is:

(14) Kde je ten slovník?  
WHERE BE(3PS,Pres) THAT DICTIONARY(IM,S,Nom)  
- Where is the/that/this dictionary?

For an overview of the declension of nouns, see Appendix II.

2.3.3 PRONOUNS

As shown in the previous chapter also pronouns may influence the distribution of definiteness in a Czech sentence. Here follows a survey of the relevant pronouns. Free pronouns (i.e. personal pronouns) are also listed here. As mentioned earlier in section 2.3.1, pronouns have to agree in gender, number and case with the noun they determine.

Personal pronouns

Personal pronouns as the subject of a sentence are normally omitted, they are only used for emphasis or for the sake of clarity, i.e. in apposition and contrast:

(15) My jsme zde, ale on zde není.  
WE(Nom) BE(1PP,Pres) HERE BUT HE(AM,S,Nom) HERE IS-NOT  
- We are here, but he isn’t here.

(16) To jsem já.  
IT(N,S,Nom) BE(1PS,Pres) l(Nom)  
- It’s me.

(17) Kdo jste vy?  
WHO BE(2PP,Pres) YOU(P,Nom)  
- Who are you?

\textsuperscript{14} The form of the pronoun varies depending on the gender, number and case of the qualified noun.
The personal pronouns are:

SINGULAR
1. já
2. ty
3. on (masc.)
ona (fem.)
ono (neut.)

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>SINGULAR</td>
<td>PLURAL</td>
</tr>
<tr>
<td>1. já</td>
<td>1. my</td>
</tr>
<tr>
<td>2. ty</td>
<td>2. vy</td>
</tr>
<tr>
<td>3. on</td>
<td>3. oni</td>
</tr>
<tr>
<td>ona</td>
<td>ony</td>
</tr>
<tr>
<td>ono</td>
<td>ona</td>
</tr>
</tbody>
</table>

Possessive pronouns

Possessives are used less in Czech than in English. The example sentence below has two possible translations in English:

(18) Hledám slovník.  
LOOK-FOR(1PS,Pres) DICTIONARY(IM,S,Acc)  
- I am looking for a dictionary.  
- I am looking for my dictionary.

The possessive pronouns are:

SINGULAR
1. můj [mu:]  my, mine
2. tvůj [tvu:] your, yours
3. jeho  his
   její  her, hers

PLURAL
1. nás [nas]  our, ours
2. vás [va:]  your, yours
3. jejich  their, theirs

Possessives jeho and jejich are indeclinable.

Demonstrative pronouns

The demonstrative pronouns in Czech may be used as emphatic pronouns. With inanimate nouns this is the standard use:

Example: ten - this, that
   masculine - ten, ex. ten den - this day, that day
   feminine - ta, ex. ta kniha - this book, that book
   neuter - to, ex. to slovo - this word, that word.

Adding a suffix -hle or -to to any of ten - forms narrows its meaning:
   tahle kniha - this book (right here), that book (over there)
   tento dům - this house (right here), that house (over there).

15 Notice that in speech the form 'oni' may be used for all genders.
(19) Vidíš ten dům / SEE(2PS,Pres) THAT(IM,S,Acc) HOUSE(IM,S,Acc) to auto / THAT(N,S,Acc) CAR(N,S,Acc) tu knihu? THAT(F,S,Acc) BOOK(F,S,Acc)

- Do you see that house, that car, that book?

Indefinite pronouns

Examples: všechen - all
někdo[někdɔ] - someone, anyone
některý[některiː] - one (of them), some sort (or kind) of
nějaký[nějakːiː] - some, any
kdosi - someone
cosi - somewhat

For an overview of the declension of pronouns, see Appendix II.

2.3.4 WORD ORDER

The definitions of the characteristics of word order in Czech and some examples have already been given (see chapter I and section 2.3.1), here a closer view will be given.

As noted several times already, the order of words in Czech is much less strict than in English. The word order is determined by the topic-comment structure of a sentence, rather than grammatical relations. In other words, the sentence constituents alter their positions in a sentence depending on whether they are new or old in the context. The different positions in a sentence and their functions are described below. The section is closed by giving some discourse examples in Czech.

First and last position

The first and last words (or phrases) in a Czech sentence have greater stress, i.e. those words which are most important, or which receive the strongest emphasis, begin or end the sentence.

As mentioned before (see chapter 1), normal intonation in Czech appears when the intonation centre is on the last element in a sentence. Other positions of intonation centre are in Czech considered as emotionally marked.

The following sentences are examples of sentences expressed in normal intonation (Normal) and stressed intonation (Stressed).  

---

16 A sentence intonation centre, i.e. the carrier of the main stress in a sentence, is indicated by capital letters.
• Normal: Žák jde dnes DO ŠKOLY.
PUPIL(AM,S,Nom) GO(3PS,Pres) TODAY TO SCHOOL(F,S,Gen)
- The pupil is going to school today.

Stressed: DNES žák do školy nejde.
TODAY PUPIL(AM,S,Nom) TO SCHOOL(F,S,Gen) NOT-GO(3PS,Pres)
- Today the pupil is not going to school.
(= emphasis on today and is not going , since an exception
to the pupil's normal habit of going is implied).

• Normal: Pan Novák zná PANA DVOŘÁKA.
MR.(Nom) NOVAK(Nom) KNOW(3PS,Pres) MR.(Acc) DVORAK(Acc)
- Mr. Novák knows Mr. Dvorák.

Stressed: PANA DVOŘÁKA zná pan Novák.
MR.(Acc) DVORAK(Acc) KNOW(3PS,Pres) MR.(Nom) NOVAK(Nom)
- It is Mr. Dvorák whom Mr. Novák knows.
- Mr. Dvorák is known by Mr. Novák.

• Normal: To se mi NELÍBÍ.
IT(N,S,Nom) refl. I(Gen) NOT-LIKE(3PS,Pres)
- I don't like that.

Stressed: MNĚ se to nelíbí.
I(Gen) refl. IT(N,S,Nom) NOT-LIKE(3PS,Pres)
- I don't like that ( whatever anyone else may think).

• Normal: Udělal jsi to DOBŘE.
DO(M,S,Past) BE(2PS,AUX) IT(N,S,Acc) WELL
- You did that well.

Stressed: DOBŘE jsi to udělal.
WELL BE(2PS,AUX) IT(N,S,Acc) DO(M,S,Past)
- You did that well (very well, not at all badly).

• Normal: Pan Novák jel DO PRAHY.
MR.(Nom) NOVAK(Nom) GO(AM,S,Past) TO PRAGUE( Gen)
- Mr. Novák went to Prague.

Stressed: DO PRAHY jel pan Novák.
TO PRAGUE( Gen) GO( AM,S,Past) MR.(Nom) NOVAK(Nom)
- Mr. Novák (not Mr. Dvorák) went to Prague ( not to Brno).
Last position

The last position in a sentence is often reserved for an element of the sentence which is new to the context.

(20) Včera v noci zemřel
    YESTERDAY IN NIGHT(F,S,Loc) DIE(AM,S,Past)
    MŮJ OTEC.
    MY(AM,S,Nom) FATHER(AM,S,Nom)
    - Last night my father died.
    = the new information here is the subject of the sentence
    --> who died? - my father

(21) Můj otec zemřel
    MY(AM,S,Nom) FATHER(AM,S,Nom) DIE(AM,S,Past)
    VČERA V NOCI.
    YESTERDAY IN NIGHT(F,S,Loc)
    - My father died last night.
    = in this sentence it is the time that is new to the context
    --> when? - last night

(22) Můj otec včera v noci
    MY(AM,S,Nom) FATHER(AM,S,Nom) YESTERDAY IN NIGHT(F,S,Loc)
    ZEMŘEL.
    DIE(AM,S,Past)
    - My father last night died.
    = this sentence has the verb as the new element
    --> what happened? - he died

Second position

The second position in a sentence, i.e. the one following the first stressed word or phrase is that of least emphasis (a(and), i(and), ale(but) and avšak(but, however) are not considered stressed). This position is often taken by short, common words on which no special stress is laid. Such words are called enclitics.

Certain words are always or almost always enclitic:
- the auxiliaries of the past tense and the subjunctive mood (být, by),
- the reflexive enclitic pronouns (se, sí),
- several conjunctions and conjunctive adverbs -však, sice(however)
- the enclitic forms of the personal pronouns in the dative and accusative (occasionally in the genitive): mě, tě, mě, ti, etc.

Certain other words are frequently, but not always enclitic. Such are:
- the verb to be, the verb to have,
- most pronouns (except forms which are always stressed, such as já, ry, mne, tebe, etc.),
- certain adverbs and particles - tak(so), už(already), ještě(yet, still, even), asi, snad (perhaps, maybe), skoro(almost), tedy (then, accordingly), pryj(they say), etc.,
- certain conjunctions and conjunctive adverbs - proto(therefore), třeba(if necessary), přece(yet, still), etc.
These words are enclitic at times, but when special stress is placed on them, they assume other positions:

**Normal:** Udělal jste to DOBRÉ.
DO(AM,S,Past) BE(2PP,AUX) IT(N,S,Acc) WELL
- You did that well.

**Stressed:** TO jste dobře udělal.
IT(N,S,Acc) BE(2PP,AUX) WELL DO(AM,S,Past)
- That you did well.

Many other words are capable of assuming enclitic position when they are not stressed.

**Discourse**

The following four sentences are an example of a possible discourse in Czech. The text is taken from a language learning book. The sentences represent a written text, which, as mentioned before (see chapter 1), means that they will be interpreted as having normal intonation, i.e. the carrier of the main sentential stress is the last word in a sentence.17

(23) **V pokoji stál STŮL.**
IN ROOM(IM,S,Loc) STAND(IM,S,Past) TABLE(IM,S,Nom)
- In the room stood a table.

(24) **Na stole ležely knihy**
ON TABLE(IM,S,Loc) LAY(F,P,Past) BOOK(F,P,Nom)
a PAPÍR.
AND PAPER(IM,S,Nom)
- On the table there lay books and paper.

(25) **Za knihami hořela SVÍČKA.**
BEHIND BOOK(F,P,Ins) BURN(F,S,Past) CANDLE(F,S,Nom)
- Behind the books there burned a candle.

(26) **Při svíčce četl neznámý ČLOVĚK.**
BY CANDLE(F,S,Loc) READ(AM,S,Past) UNKNOWN(AM,S,Nom)
MAN(AM,S,Nom)
- By the candle a stranger was reading.

According to the theory of FSP, the observation here is that all the sentences involved have a primary word order, i.e. the complements preceding the verb are in topic, whereas the complements following the verb are in focus. In other words, the information that is new to the context occurs towards the end of a sentence, whereas the information already known from the previous context is placed in the beginning of the sentences.

---

17 Again, the intonation centre in a sentence is indicated with capital letters.
In more detail, in the sentence (23) the word 'stůl' (= table) is mentioned for the first time and appears at the end of the sentence. In the following sentence (24) the noun is expanded with further information, appearing in the beginning of that sentence within a prepositional phrase ('na stole'). In other words, the noun within the prepositional phrase in the beginning of sentence (24) refers back to the previously mentioned 'table' in sentence (23), extended with additional information, i.e. the location. The same happens in the rest of sentences, where nouns introduced at the end of previous sentences are placed at the beginning of the following sentences, often further expanded with additional information. In fact all the sentences may be considered as initial, having the word order adverbial - verb - actor (see section 2.2.3).

The observation made here is that the distribution of topic-focus (as the decisive factor of definiteness distribution in Czech) corresponds to the definiteness distribution in the corresponding English translations, where noun phrases in topic (i.e. in the beginning of a Czech sentence) obtain a definite article in English, whereas newly introduced nouns (i.e. at the end of a Czech sentence) have an indefinite article in English.

Another interesting example is the discourse example below (also taken from a language learning book) which shows even more clearly how the definiteness is distributed in Czech.

(27) Do domu, který jsem střežil, vešel MUŽ.

INTO HOUSE(IM,S,Gen) WHICH(IM,S,Nom) BE(1PS,AUX) WATCH(AM,S,Past) ENTER(AM,S,Past) MAN(AM,S,Nom)

- A man entered a house that I was watching.

(a) Muž vyšel ve tři hodiny.

MAN(AM,S,Nom) GO-OUT(AM,S,Past) AT THREE HOUR(F,P,Nom)

- The man left at three o'clock.

(b) Ve tři hodiny vyšel MUŽ.

AT THREE HOUR(F,P,Nom) GO-OUT(AM,S,Past) MAN(AM,S,Nom)

- A man left at three o'clock.

Again, the sentences represent a written text with normal intonation and a primary word order. The sentence (27) may be followed by sentences (a) or (b) depending on the circumstances. In the initial sentence (27) 'a man' is introduced, which is indicated by the placement of the noun 'myn u ž' (= man) at the end of that sentence (i.e. again the word order adverbial - verb - actor).

If the sentence (27) is followed by sentence (a), the interpretation will be that we are talking about the same man that entered the house, leaving at three o'clock.

In the case of sentence (b) the situation changes. Choosing this sentence would mean that some another man, i.e. not the man introduced in the initial sentence as entering the house, has left the house at three o'clock. In other words, a new individual is introduced here, which is indicated by the placement of the noun at the end of the sentence (b).

Once again, the definiteness distribution in the Czech sentences corresponds to the definiteness distribution in the corresponding English translations.
The conclusion is, that it is very important in Czech to have the sentence elements in the right place, otherwise the interpretation will be wrong.

The last discourse example is taken from a book for children, i.e. no corresponding English translation was available here.

(28) Samoobsluha je za PARKEM.

SUPERMARKET(F,S,Nom) BE(3PS,Pres) BEHIND PARK(IM,S,Ins)
- The supermarket is behind a park.

(29) Tam stojí v zahradách

THERE STAND(3PP,Pres) IN GARDEN(F,P,Loc)
pár starých DOMŮ.

COUPLE(IM,Nom) OLD(M,P,Gen) HOUSE(IM,P,Gen)
- There in the gardens stands a couple of old houses.

(30) Ty domy mají JMÉNA.

THAT(IM,P,Nom) HOUSE(IM,P,Nom) HAVE(3PP,Pres) NAME(N,P,Acc)
- That/The houses have names.

The sentences represent a written text, i.e. they have normal intonation, but the word order is not always primary.

The initial sentence (28) has a primary word order and is of the type that has a topic part, i.e. the word order here is actor - verb, whereas in initial sentences without any topic part the actor follows the verb (see section 2.2.3). In the following sentence (29) the word order is marked, i.e. the actant indicating place (v zahradách) stands before the actor (pár starých domů), whereas within the systemic ordering (see section 2.2.3) the actor is placed before the actant of place. That implies that the actant of place is here interpreted as definite in all contexts. Notice that the actant does not occur in the previous sentence, which may be explained with reference to the knowledge that a house is usually surrounded by a garden, i.e. the experiential context takes over here.

The last sentence (30) again has a primary word order, where the actor is now considered as known partly because of the preceding demonstrative pronoun and partly because of its position in the sentence.

The observation in this text is that the topic-focus structure rules also apply to the experiential context, not only to the linguistic context.
2.4 NOUN PHRASES AND ARTICLES IN ENGLISH

2.4.1 INTRODUCTION

English requires obligatory marking of definiteness on (most) noun phrases. In other words, nouns in English are preceded by a definite or indefinite article, but there are also cases when no article occurs.

This section describes briefly the definite and indefinite article and lists some of the cases where a noun has no article in English. These distinctions are relevant for the analysis of species in Czech from a machine translation perspective.

2.4.2 DEFINITE VS. INDEFINITE ARTICLE

According to Gawronski (1993), the indefinite article is generally treated as:

- signalizing "newness" in discourse (this function is, however, restricted to the non-generic instances of the indefinite article),
- giving an introductory or objective presentation,
- being unmarked with respect to the notions of "familiarity" and "uniqueness";

in contrast to the definite article, that has the characteristics of:

- signalizing "familiarity" in discourse,
- giving a subjective or anaphoric presentation,
- being marked with respect to the notions of "familiarity" and "uniqueness".

Gawronski (1993:122) gives the definition of articles with respect to uniqueness as: "... the definite article can be seen as a special type of existential quantifier, expressing ... that there exists a unique entity or a unique set of entities. The indefinite article is often treated as a neutral existential quantifier with respect to uniqueness." What she means is that noun phrases containing the indefinite article draw attention to a class rather than to a concrete member of the class.

The definitions of definite and indefinite article given above coincide in the FSP theory with the definitions of topic and focus respectively, in that focus is also defined as introducing something new to a discourse, whereas topic is considered as known, i.e. familiar to the context. In other words, this analysis is compatible with the hypotheses that Czech noun phrases in topic should be translated to definite noun phrases in English and those in focus to indefinite noun phrases.
2.4.3 NOUN PHRASES WITHOUT AN ARTICLE

The definiteness marking in English is not just restricted to the definite and indefinite article. In other words, some noun phrases may occur without any article in English, which complicates the problem of article choice in translation to English. Some of the cases are listed below, taken from Gawronska (1993).

In English, articles normally are not used with:

- proper names;
- quasi-proper names, i.e. names used in a similar way as proper names, e.g. God, mummy, duddy, etc.;
- mass names (NB: the article is used when an individuation is required, e.g. 'the wine is served', 'the famous English tea');
- certain lexicalized phrases, e.g. 'play football', 'go to bed', 'be at table', etc.;
- indefinite plurals without article:
  - indefinite plurals are characterized as referring to some number of individuals which belong to the denotation of the noun; their function is reminiscent of that of mass names;
  - definite plurals are seen as referring to a group of individuals taken as whole;
- presence of certain pronominal determiners and modifiers (possessive, demonstrative or indefinite pronouns) excludes the possibility of article use;
- specific rules stating what determiners are in complementary distribution with the articles, e.g. 'play the piano'.

These exceptions will be taken into account in the analysis of Czech noun phrases below.
2.5 PREVIOUS IMPLEMENTATIONS

2.5.1 INTRODUCTION

This chapter presents two implementations that are immediately relevant to the topic of this study. The first is "A Dependency-Based Parser for Topic and Focus", which actually solves the topic-focus recognition problem for languages with free word order.\textsuperscript{18} The second is a machine translation approach to article and aspect choice in translation from English and Swedish to Russian and Polish and vice versa.

2.5.2 TOPIC-FOCUS TREATMENT

"A Dependency-Based Parser for Topic and Focus" in Hajicova (1991) represents an implementation based on the theory of Functional Sentence Perspective.

The parser is based on the assumption that all sentences in a text are analysed into two main parts: topic and focus. The content of a sentence is represented as "... a projective rooted tree with the root labelled by a complex symbol of a verb and its daughter nodes by those of the complementations of the verb, i.e. participants." (in Hajicova 1991:127). In other words, the main verb is the highest node and considered as the boundary between the topic and focus parts. If the verb is not contextually bound, it belongs to the focus part, and so do all contextually unbound daughter nodes of the verb.

The top-down dimension of the tree reflects the structural characteristics of the sentence. The left-to-right dimension represents the deep word order.

The notions of focus, topic and contextual boundness are claimed to provide an effective tool for anaphora resolution.

2.5.3 ARTICLE CHOICE AND ASPECT

Another implementation done by Barbara Gawronska gives some general principles for the choice of articles and aspect in automatic translation from English and Swedish to Russian and Polish and vice versa (see Gawronska 1993). The main goal of this thesis was to investigate the invariant meaning of the Swedish and English articles and the aspect category in Polish and Russian.

The implementation is done within the machine translation project Swetra (Swedish Computer Translation Research), developed at the Department of Linguistics, Lund University, which started in 1987.

It is applicable for isolated sentences and short fragments of newspaper texts. The strategy used here is based on a combination of programming in Flex (an expert system for automatic taxonomy design, translating a standarized variant of 'Machinese English' into Prolog structures), DCG (Definite Clause Grammar) and 'pure' LPA Prolog. The parsing and translation strategy chosen for the

\textsuperscript{18} I want to point out here, that the reason why I did not take this implementation into account when developing my own system, is that I did not know about its existence before I started on my own implementation. I have not been able yet to actually compare this implementation to my own system.
purpose of experiments with article and aspect choice is called FlexRG - flexible, lexicon-oriented referent grammar.

This implementation also considers the rules of FSP with regard to preverbal and postverbal positions of noun phrases, but other rules are also used for example to check coreference.
3 METHOD

3.1 INTRODUCTION

As stated in the introduction (see chapter 1) I have chosen to focus on two problems that arise in the analysis of Czech for machine translation into a language like English. First, Czech is a 'free' word order language, which implies that a parser for handling free word order has to be implemented. Second, English requires marking of definiteness on (most) noun phrases, whereas no such marking occurs in Czech. The solution to the second problem involves using a topic-focus analysis to assign species to Czech noun phrases. In the following, I will use the term discourse analysis to refer to the conjunction of topic-focus analysis and species assignment.

This section describes the methodological decisions made both with regard to parsing and discourse analysis. At the end of the chapter the method used to test the implementation is described and a short description of the available corpus material is given.

3.2 PARSING

A categorial grammar chart parser has been chosen for the grammatical analysis of Czech. The choice of this parsing system was partly arbitrary and influenced by the fact that categorial grammar was used by people at UMIST where I began the work for this thesis. It remains to be seen if this choice was appropriate. The implementation language used is Prolog, in the environment of AAIS-Prolog (Advanced Artificial Intelligence Systems' Prolog).

The basic categorial grammar formalism and the method of chart parsing have already been described in the previous chapter (see section 2.1). In the next chapter I will discuss the analysis of Czech and the implementation is described in chapter 5.

3.3 DISCOURSE ANALYSIS

3.3.1 INTRODUCTION

The second part of the analysis, which I will refer to as discourse analysis, is divided into two parts: topic-focus recognition and species assignment. The theory of Functional Sentence Perspective (FSP), described in section 2.2.3, is the background theory for this part of the analysis. In the following subsections the methods of topic-focus recognition and species assignment are discussed.
3.3.2 TOPIC-FOCUS RECOGNITION

In accordance with the definitions of the topic-focus structure in Czech as defined within the FSP-theory (see section 2.2.3), the following assumptions are crucial for an automatic recognition of topic and focus in a Czech sentence:

(a) Either the input is a spoken discourse and the recognition procedure includes an acoustic analysis, or written (printed) texts are analyzed.

(b) Every sentence includes a focus element, since otherwise it would convey no information relevant for communication. On the other hand, there may be sentences without topic (initial sentences).

(c) The main verb in a sentence is considered to be the main boundary between the topic and focus parts.

(d) The complements in the focus part of a sentence follow the systemic ordering for Czech, given in section 2.2.3.

(e) The focus elements and the subject (if in topic) must be compared to the previous context to completely determine their topic-focus status.

Regarding (a), a restriction to written texts is necessary here, i.e. no acoustic analysis is going to take place. As mentioned before, written texts in Czech are assumed to have normal intonation, which means that the carrier of the main sentential stress is the last word in a sentence. This implies that (b) is also true, i.e. within the FSP theory the element in the intonation centre is always in focus and I assume that all sentences have an intonation centre.

According to the rules of topic-focus structure, a sentence has to be split into these two parts. In order to be able to split a sentence in topic and focus parts, a position has to be defined that functions as the boundary between the parts. In Czech this position is represented by the verb (see point (c)), but the problem is that even though the verb is the limit in the deep word order, in the surface word order the situation may be different. This implies that the word order of the complements that follow the verb, i.e. the potential members of focus, have to be compared to the systemic ordering, which represent the deep word order (see point (d)).

However, for topic and focus recognition in a text, this is still not enough. There is the possibility that some complements included in the focus part are contextually bound, i.e. the focus part may still include just the potential members. Also the subject of a sentence that precedes the verb, i.e. is considered as contextually bound, may be contextually unbound - point (e). In other words, the context also plays a part in the topic and focus recognition for Czech and previous sentences should then be stored in some way.

The question is if the whole sentences should be stored. The study concerns noun phrases in the first place and therefore a restriction to store only the noun phrases has been adopted. Consequently, the problem that a verb in a Czech sentence is always ambiguous according to FSP theory, will not be treated here, except if the verb is the carrier of the intonation centre. In other words, the topic-focus recognition in a text will be incomplete with regard to verbs that are not in the intonation centre.
3.3.3 SPECIES ASSIGNMENT

The second part of the discourse analysis is the assignment of an appropriate species value to all noun phrases in a text. I distinguish three different values - 'definite', 'indefinite' and 'none'. The first two are assigned according to the topic-focus distribution in a sentence, i.e. after the topic-focus recognition, so that the noun phrases in topic obtain a definite species and noun phrases in focus an indefinite species.

The value 'none' is connected to the analysis of English and the machine translation perspective. In section 2.4, different types of noun phrases in English were distinguished, including those where no article occurs. The species value 'none' is therefore reserved for those noun phrases that are translated to English noun phrases without article. In the present implementation, these are the noun phrases involving determiners in the form of possessive, demonstrative and indefinite pronouns, and proper names.

3.4 TESTING

I must say that it was not easy to find appropriate texts in order to check the algorithm for species assignment. The problem was that several Czech texts had no corresponding English translation, which meant that the amount of available corpus material was rather small.

However, analyses have been done of written texts of one car manual, some scientific articles and texts, shorter texts in language learning books and some children books.

Only the shorter texts in language learning books and books for children are used in testing the implemented system. The structure of sentences in the scientific articles and texts is too complex for the parser. In these texts the discourse analysis has been done by hand.

The analysis proceeded by considering first a sentence in the Czech text and then comparing it to the corresponding English translation. The species value (the article) a noun phrase obtained in English was compared to the noun phrase position in the topic-focus structure of a Czech sentence. In the case of texts without any corresponding English translation no such comparison was possible, except with regard to my own intuitions.

The test texts used as the input to the implemented system have already been given in the section about word order in Czech (see section 2.3.4). The test runs are listed in Appendix I.
4 ANALYSIS

4.1 INTRODUCTION

In this chapter, I propose an analysis of the two problems identified in the introduction (see chapter 1). First, the problems arising with parsing in a 'free' word order language will be discussed with the appropriate solutions. Then the problems of species assignment will be treated.

4.2 PARSING IN A 'FREE' WORD ORDER LANGUAGE

4.2.1 INTRODUCTION

As mentioned earlier, the word order in Czech is rather free. Parsing in a 'free' word order language requires some means for combining sentence constituents regardless of the order in which they are introduced. In the categorial grammar environment, the neutral connective, called 'neutral slash', satisfies these restrictions. The rules using this connective have already been introduced in section 2.1.4. This section includes further explanations for the use of the neutral slash rules when parsing in a 'free' word order language.

In the following sections I first discuss the lexicon, including the parts of speech together with their appropriate categorial grammar categories, and then all the categorial grammar rules used for chart parsing in Czech are discussed, together with appropriate examples.

4.2.2 THE CATEGORIAL GRAMMAR LEXICON

Appropriate lexical categories have been defined for every part of speech included. Some parts of speech involve more than one category. In the present implementation, there are eleven different lexical categories.

Nouns

Nouns are divided into two types - common and proper. Common nouns in Czech require no determiners to form a noun phrase, which is why both common and proper nouns are assigned the category noun phrase - 'np' in categorial grammar notation.

Pronouns

Pronouns have two categories depending on whether they are determiners or free pronouns. Free pronouns have the same category as nouns, i.e. 'np'. Determiners also form noun phrases, but have to be followed by another noun phrase, in categorial grammar notation 'np/np'. The lexicon includes personal, possessive, demonstrative and indefinite pronouns.
Adjectives

Adjectives modify a noun, which implies that they should have the same category as determiners, i.e. np/np.

Verbs

Verbs are divided into different types according to the number of complements they take. Every type has two different categories depending on whether the verb takes a subject or not. In Czech, subject pronouns may be omitted, because the verb endings already indicate the grammatical person (see section 2.3.3). Therefore, two lexeme categories are defined for every type of verb:

- The first category is for the case where the subject pronoun or noun is present, e.g. s | np for intransitive verbs.
- The second category is for the case where no overt subject occurs, e.g. s for intransitive verbs.

The lexicon includes intransitive verbs, that take only a subject as complement, transitive verbs, which take two complements (i.e. subject and object) and ditransitive verbs with three complements (subject, direct object and indirect object).

Prepositions

Prepositions have the category - s | s / np - which means that they form a sentence modifier if they are followed by a noun phrase. In other words, prepositional phrases take the first or last position in a sentence according to this category, but by using functional composition rules they may also occur in other positions in a sentence.

Adverbs

Adverbs have the same category as prepositional phrases and may therefore occur in the same positions.

Conjunctions

Conjunctions are defined to bind both sentences and noun phrases together, i.e. they form a sentence (noun phrase) if they are preceded by a sentence (noun phrase) and followed by another sentence (noun phrase). Thus, they have the category - C | C | C.

Figure 4.1 gives a summary of all the parts of speech and their syntactic categories in categorial grammar.
### Figure 4.1: Lexeme categories.

#### 4.2.3 CATEGORIAL GRAMMAR RULES

Here I want to give more explanations in the form of examples to all the categorial grammar rules useful for parsing in a 'free' word order language like Czech. First, examples of the rules of functional application using the directional connectives will be described and then all the rules using the neutral connective.

In order to avoid confusion in the terminology, I prefer to mark all rules using the neutral connective with the capital N. For example the rules for functional application, involving this connective will be called 'functional N-application rules'.

**Application**

The rules for functional application using the directional slashes may be used in Czech for the sentence parts that have a fixed position. Such sentence elements are for example noun modifiers, which always precede the noun they modify. Consider the example (1) below.

(1) ten velký starý dům

THAT(IM,S,Nom) BIG(IM,S,Nom) OLD(IM,S,Nom) HOUSE(IM,S,Nom)

- that big old house

The demonstrative pronoun 'ten' and the adjectives 'velký' and 'starý' precede the noun 'dům'. The structural derivation of this noun phrase is shown in figure 4.3, where the rules of functional application, listed in figure 4.2, are used.
Functional application rules (A):

- Forward application
  Result/Argument → Argument

- Backward application
  Argument → Result\Argument

**Figure 4.2:** Functional application rules.

\[
\begin{array}{lllll}
\text{ten} & \text{velký} & \text{starý} & \text{dům} \\
np/np & np/np & np/np & np \\
\hline
& & & \rightarrow A \\
p & & \rightarrow A \\
p & \rightarrow A \\
p & \rightarrow A \\
\end{array}
\]

**Figure 4.3:** Structural derivation in Czech using functional application rules.

The modifiers within noun phrases in Czech actually have the same word order as the determiners and modifiers in English noun phrases, i.e. they precede the noun. When we compare figure 4.3 above to figure 2.4(a) (section 2.1.4), the only difference between these figures is that, in the English example the noun has the category 'noun' (n), whereas in the Czech figure a noun has the category 'noun phrase' (np). The reason for this different choice of syntactic category is that nouns in Czech build noun phrases on their own, whereas nouns in English often require a determiner to form a noun phrase. One drawback of this choice is that the restriction on noun phrases described in section 2.1.3, where any number of adjectives can precede a noun, but only one determiner, is no longer in force (c.f. figure 2.2 in section 2.1.3).

**Neutral application**

Using functional application rules with the neutral connective is very useful in Czech, where the structures of sentences may differ greatly. Consider example (2) below.

(2) Petr napsal DOPIS.

PETR(AM,S,Nom) WRITE(AM,S,Past) LETTER(IM,S,Acc)
- Peter wrote a letter.

In different contexts the sentence word order changes. The sentence examples below illustrate such alternations. The English questions within parentheses are examples of a possible context for a given sentence.
(a) Petr napsal DOPIS.
   - (What did Peter do?)
     Peter wrote A LETTER.
(b) DOPIS napsal PETR.
   - (Who wrote the letter?)
     PETR wrote the letter.
(c) Petr dopis NAPSAL.
   - (What did Peter do with the letter?)
     Peter WROTE the letter.

The structural derivation of the sentences using the neutral slash rules, given in figure 4.4, is shown in figure 4.5.

Functional N-application rules (NA):

- Forward application
  Result | Argument  Argument --> Result

- Backward application
  Argument  Result | Argument --> Result

**Figure 4.4:** Functional application rules using the neutral connective.

(a) Petr  napsal  dopis
   np(Nom)  s  |  np(Nom)  |  np(Acc)  |  np(Acc)

   ----------------------------------------------------------> NA
   s  |  np(Nom)
   --------------------------------< NA
   s

(b) dopis  napsal  Petr
   np(Acc)  s  |  np(Nom)  |  np(Acc)  |  np(Nom)

   ----------------------------------------------------------< NA
   s  |  np(Nom)
   --------------------------------< NA
   s

(c) Petr  dopis  napsal
   np(Nom)  np(Acc)  s  |  np(Nom)  |  np(Acc)

   ----------------------------------------------------------< NA
   s  |  np(Nom)
   --------------------------------< NA
   s

**Figure 4.5:** Structural derivations of the word order changes in Czech using neutral slash rules.\(^{19}\)

\(^{19}\) Notice that the bracketing within categories is omitted. The arguments applies from right to left. For example the transitive verb "napsal" has the category \((s \mid np(Nom)) \mid np(Acc)\), where the noun phrase in accusative has a higher precedence than the noun phrase in nominative. The same method is applied in all the rules which will be given, which implies that no bracketing will be shown neither with definitions of rules nor structural derivation.
Neutral composition

The functional composition rules are useful in Czech when a sentence includes a prepositional phrase for example. A prepositional phrase in Czech can occur at the beginning of a sentence or at the end of a sentence, but there is also the possibility that it follows the verb in a sentence, and there are other possibilities as well. Consider the example (3) below.

(3) V Perle koupila Anna KOBERECK.
    IN PERLA(F,S,Loc) BUY(F,S,Past) ANNA(F,S,Nom) CARPET(IM,S,Acc)
    - In Perla Anna bought A CARPET.20

The possible word order changes for this sentence are listed below (again, with a question as example context):

(a) V Perle koupila Anna KOBERECK.
    - (What did Anna buy in Perla?)
      In Perla Anna bought a CARPET.
(b) Koberec koupila Anna V PERLE.
    - (Where did Anna buy the carpet?)
      Anna bought the carpet IN PERLA.
(c) Anna koupila v Perle KOBERECK.
    - (What have Anna done?)
      Anna bought a CARPET in Perla.

For the sentences (a) and (b) the usual functional application rules with neutral slash are enough to parse the sentences. The sentence (c) requires other means, because the prepositional phrase interrupts the sentence structure. The rule of functional composition, given in figure 4.6, may then be used, as is shown in figure 4.7.

Functional N-composition rules (NC):

- Forward composition
  Result | Result₁ | Result₁ | Argument --> Result | Argument

- Backward composition
  Result₁ | Argument | Result | Result₁ --> Result | Argument

Figure 4.6: Functional composition rules using the neutral connective.

---

20 "Perla" is the name of a department store; it also means "pearl".

47
Neutral associativity

The functional N-associativity rules (see figure 4.8(b)) are a modified version of the associativity rule given in section 2.1.4, here repeated in figure 4.8(a).

(a) Unary associativity rule:

- (Result | Argument1) | Argument2 --> (Result | Argument2) | Argument1

(b) Functional N-associativity rules (NAS):

- Forward associativity
  (Result | Argument1) | Argument2 Argument1 --> Result | Argument2

- Backward associativity
  Argument1 (Result | Argument1) | Argument2 -->> Result | Argument2

Figure 4.8: Functional associativity rules using the neutral connective.
(a) Unary variant.
(b) Binary variant.

When applying a categorial grammar rule the argument must stand immediately before or after the functor category, otherwise no combination is possible. In a 'free' word order language no definite order exists in which the verb complements are introduced. In other words, the only thing one can do is list the complements, but the order of constituents may differ from sentence to sentence. In other words, these new associativity rules (figure 4.8(b)) make it possible to combine arguments with their functor categories even if the argument does not immediately precede or follow the functor.

The redefined associativity rule is no longer a unary rule, which implies definition of two rules, i.e. forward and backward. The reason for this change is simply efficiency, i.e. using the unary definition of this rule implies changing first the order of arguments and then applying other rules, whereas in the redefined version no change is needed, the appropriate argument is directly combined with its functor category.

For example, according to the categorial grammar, the ditransitive verbs in Czech build a sentence if they are preceded or followed by a noun phrase in nominative case (subject), a noun phrase in accusative case (direct object) and a noun phrase in dative case (indirect object). We cannot decide the order of the noun phrases, because of the possible word order changes. In the categorial
grammar notation, the ditransitive verbs may then obtain a syntactic category of the form

\[ s \mid \text{np(nom)} \mid \text{np(acc)} \mid \text{np(dat)}. \]

The neutral connective in itself indicates that the noun phrases may come either before or after the functor in any word order in a sentence. But functional application, as mentioned above, requires that the dative noun phrase in this case is immediately adjacent to the verb, which is not always the case. Let me explain this by giving an example.

Consider the sentence (4) below.

(4) Adam dal Evě JABLKO.

ADAM(AM,S,Nom) GIVE(AM,S,Past) EVA(F,S,Dat) APPLE(N,S,Acc)
- Adam gave an apple to Eva.

The following possible combinations of word order can be realised for the sentence (4), again, instantiated with a question as a context example:

(a) Adam dal Evě JABLKO. \[ \rightarrow \text{Nom} \oplus \text{Dat} \oplus \text{Acc} \]
- (What happened?)
  Adam gave AN APPLE to Eva.

(b) Adam dal jablko EVĚ. \[ \rightarrow \text{Nom} \oplus \text{Acc} \oplus \text{Dat} \]
- (To whom gave Adam the apple?)
  Adam gave the apple to EVA.

(c) Jablko dal Evě ADAM. \[ \rightarrow \text{Acc} \oplus \text{Dat} \oplus \text{Nom} \]
- (Who gave the apple to Eva?)
  ADAM gave the apple to Eva.

(d) Jablko dal Adarim EVĚ. \[ \rightarrow \text{Acc} \oplus \text{Nom} \oplus \text{Dat} \]
- (What happened to the apple?)
  Adam gave the apple to EVA.

(e) Evě dal jablko ADAM. \[ \rightarrow \text{Dat} \oplus \text{Acc} \oplus \text{Nom} \]
- (Who gave the apple to Eva?)
  ADAM gave the apple to Eva.

(f) Evě dal Adam JABLKO. \[ \rightarrow \text{Dat} \oplus \text{Nom} \oplus \text{Acc} \]
- (What happened to Eva?)
  Adam gave AN APPLE to Eva.

No problems arise with sentences (a), (c) or (e), which are parsed using the functional N-application rules. In the sentences (b) and (d) the noun phrases in dative stand at the end, i.e. they do not immediately follow or precede the verb. In (f) the noun phrase in dative immediately precedes the verb, but now the accusative is at the end. In other words, parsing sentences (b), (d) and (f) gives no result, when using only the functional N-application rules. The functional N-associativity rules defined above in figure 4.8(b) can then be used, which
makes it easy to combine the verb with its complements regardless of the order in which they are introduced.

In figure 4.9 below the structural derivation of the sentences (b), (d) and (f) is shown, using these rules.

Figure 4.9: Structural derivation using functional associativity rules.
1. Sentence (b) above.
2. Sentence (d) above.
3. Sentence (f) above.

4.2.4 CONCLUSION

In this section a list of the included parts of speech has been given and I have tried to show how the combination of rules for functional application, composition and associativity using both the directional and undirectional connectives, may be used for parsing in Czech. The implementation follows in the next chapter.

The adjacency principle, i.e. that "combinatory rules may only apply to entities which are linguistically realized and adjacent" (see Wood 1983:134), is known as one of the major contested points within CG. The principle seems intuitively natural, but causes problems in real language, as seen in the parsing of a free word order language like Czech. In other words, by using the rules of functional composition and associativity, the principle has been given up here.
4.3 EXTENDING THE CG-FORMALISM

The basic categorial grammar formalism includes only a syntactic representation. In other words, parsing a sentence in the categorial grammar environment only gives the information that the sentence is of the category s. Therefore an extension called a quasi-logical form (QLF) has been made, that gives more information about the sentence constituents.

The QLF formalism is based on the formalism developed by John Phillips at UMIST (Black 1993). The basic units are indices, properties and relations. A logical form consists of an unordered set of terms, where each term is either a property predicated of an index, or a relation between two indices. The notation depicts properties and relations as functors and indices as their arguments, i.e. within brackets. A logical form representing 'John saw Mary' might be:

c: john(j) & see(e) & mary(m) & past(e) & actor(e,j) & patient(e,m).

Slight modifications of the QLF formalism described above was needed in order to adjust the formalism to the discourse analysis part of the system. Indices, properties and relations are still the basic units, only the logical form is changed. Two different levels are distinguished:

- **GRAMMATICAL FUNCTIONS**: which includes the syntactic category of the parsed expression together with a list of all the grammatical functions involved. The representational form of this level has the same structure as Phillips' formalism, i.e. it consists of an unordered set of terms.

- **SENTENCE STRUCTURE**: which includes terms with indices indicating properties and relations of the words involved. This level differs from Phillips' formalism in that it is an ordered list of terms, reflecting the surface order of the corresponding expressions.

The sentence corresponding to 'John saw Mary' in Czech, given in (2), obtains the analysis shown in figure 4.10.

(2) Jan viděl Marii.
     JOHN(AM,S,Nom) SEE(AM,S,Past) MARY(F,S,Acc)
     - John saw Mary.

**GRAMMATICAL FUNCTIONS:**
- s#S & pred(S,V) & subj(S,J) & obj(S,M)

**SENTENCE STRUCTURE:**
- [nominal(J, head=jan, modifiers=[], features=(type=proper,pers=3, numb=sing,gend=an_masc, case=nom)),
  verbal(V,videti, features=(type=finite, tense=past),
  nominal(M, head=maria, modifiers=[], features=(type=proper,pers=3, numb=sing,gend=fem, case=acc))]

*Figure 4.10:* Quasi-logical form of the sentence (2) above.
The representation of sentence (2) in figure 4.10 above can be read as follows:

- **Grammatical functions level:**
  "There is a sentence S and the predicate of S is V, the subject of S is J and the object of S is M."

- **Sentence structure level:**
  "J is a nominal with head 'jan' and no modifiers. It is a proper name, in third person singular, animate masculine gender and nominative case. V is a verb 'videti', which is finite and in the present tense. M is a nominal with head 'maria' and no modifiers. It is a proper name, in third person singular, feminine gender and accusative case.

The two levels are related to each other through the indices. In our example, the sentence subject corresponds to the nominal J with the head 'jan' and the object to the nominal M with the head 'maria'. The sentence predicate is realized in the index V which corresponds to the verb 'videti'.

The reason why the sentence level is an ordered list is simply because in the discourse analysis the order in which the sentence constituents are introduced is important and must not be changed.

### 4.4 DISCOURSE ANALYSIS

#### 4.4.1 INTRODUCTION

Czech, as mentioned several times, have no articles. If we want to translate a text in Czech to English, a language that requires articles in one way or the other, how will we solve the problem of choosing the correct article?

According to Gawronksa (1993), for the choice of an appropriate article to a noun, we have to know what she calls "the referential power of an expression". In other words, we are looking for a previously introduced element that is identical vs. non-identical to the given element. In the case of the definite article, the search must then give the same or a related expression.

In the terms of the theory of Functional Sentence Perspective (FSP), discussed in section 2.2.3, what this means is that a noun phrase having the feature of definite species must be contextually bound with respect to the definition of contextual boundness. In other words, the FSP-theory can give rather good clues to the problem of determining which nouns in a sentence should have the definite species and which should have the indefinite species, based on the topic-focus analysis of the sentence.

The following sections specify the solution to the problem of species assignment in more detail. First, the topic and focus recognition is discussed, followed by a section describing the species assignment.
4.4.2 TOPIC AND FOCUS RECOGNITION

In my opinion, the FSP-theory can be used in Czech in order to decide the species of a noun phrase in a text. In other words, the theory may be a means of dividing a sentence into a contextually bound (topic) and a contextually unbound (focus) part, where the noun phrases in the topic are considered as definite and the noun phrases in the focus as indefinite.

Here below the procedure of topic and focus recognition in Czech written texts is defined using the FSP topic-focus structure rules given in section 2.2.3, supplemented with rules concerning the role of context.

For every sentence:

(i) Split a sentence into two parts with respect to the position of the main verb. The verb complements that precede the verb are (tentatively) in the topic and the verb complements that follow the verb are (tentatively) in the focus.

(ii) Compare the word order of the complements in the focus part to the systemic ordering. The complements that do not follow the systemic ordering are contextually bound. The boundary between topic and focus is then placed in front of the first complement that follows the systemic ordering. (Notice that here is no need to check the last complement in a sentence. This complement always belongs to the focus, i.e. it is the carrier of the main sentential stress.)

(iii) So far, only the potential focus has been built, i.e. the focus part may contain verb complements that are contextually bound. Compare the focus to the previous context searching from left to right, i.e. take the less dynamic complement first. The boundary between topic and focus is then placed in front of the first contextually unbound complement. Again, the last complement in the sentence is always contextually unbound, i.e. in the focus.

(iv) If the subject of a sentence is part of the topic, compare it to the previous context. There could be the possibility that it is contextually unbound.

4.4.3 SPECIES ASSIGNMENT

As mentioned in the beginning of the previous section, the assignment of an appropriate species value to a noun phrase may be solved by splitting a sentence in a topic and a focus part, according to the definitions for topic and focus recognition, and then assigning definite species to the noun phrases that are part of the topic and indefinite species to the noun phrases that are members of the focus part of the sentence. In other words, for the species assignment a sentence topic and focus has to be recognized. The only thing that remains to be seen is if this method is useful only in the linguistic context or if it may also apply within the situational and experiential context.

In section 2.4 English noun phrases have been discussed and as we have seen, not all noun phrases in English have an article. Some of these types are included in the implementation, which implies that before assigning a species value to a noun phrase with regard to the topic and focus, a noun phrase should
be checked to see if it is of the type that have no articles in English, because if so, then no species assignment is needed.

The procedure for assigning species (i.e. definite, indefinite or none) to a noun phrase can then be summarised as follows:

For every noun phrase in a sentence:

(i) if the noun phrase is of the type that requires no article in English, no species assignment occurs;

(ii) according to the topic focus recognition in a sentence, assign a definite species to a noun phrase in topic and an indefinite species to a noun phrase in focus.

4.5 CONCLUSION

In this chapter, I have presented an approach to the analysis of Czech for machine translation, combining categorial grammar (extended with QLF) and FSP theory in order to handle both the parsing problem and the species assignment problem. In the next chapter, I will present an implementation of this approach.
5 IMPLEMENTATION

5.1 INTRODUCTION

This chapter presents the implementation of a system for analysis of Czech based on the ideas presented in the preceding chapter. The system consists of two parts: a categorial grammar chart parser and a component for topic-focus analysis and species assignment, called discourse analysis. Everything has been implemented from scratch.

The parser is based on categorial grammar extended with a slight modified version of a quasi logical form (QLF) formalism developed by John Phillips at UMIST. The discourse analysis part is inspired by the theory of Functional Sentence Perspective described in section 2.2.3. The implementation language used is Prolog, in the environment of AAIS-Prolog (Advanced Artificial Intelligence Systems' Prolog).

The system takes a sequence of sentences and returns a two level QLF, where the first level displays the syntactic category and the grammatical functions involved, and the second level represents the surface order of the corresponding expressions, including information about the species of noun phrases. For example, consider the sentence (1) below.

(1) Do domu vešel MUŽ.
    INTO HOUSE(IM,S,Gen) ENTER(AM,S,Past) MAN(AM,S,Nom)
    - A MAN entered the house.

Parsing, topic-focus analysis and species assignment for this sentence results in the output structure shown in figure 5.1.

GRAMMATICAL FUNCTIONS:
* s#S & (pred(S,V) & subj(S,N1)) & adverbial(S,P,N2)

SENTENCE STRUCTURE:
* [preposition(P,do),
  nominal(N2,head=dum,modifiers=[],features=(species=def,type=common,
  pers=3,numb=sing,gender=in_masc,case=gen)),
  verbal(V,vejiti,features= (type=finitate=flex=flex)),
  nominal(N1,head=muz,modifiers=[],features=(species=indef,type=common,
  pers=3,numb=sing,gender=an_masc,case=nom))]

**Figure 5.1:** The two-level QLF of sentence (1) above.

The rest of this chapter is structured as follows: First, I give a short survey of the system components. Secondly, the parsing components are described. The third part is devoted to the discourse analysis component. At the end of the chapter, an overview of the whole program is given. Test runs are given in Appendix 1.

55
5.2 THE SYSTEM STRUCTURE

The analyzer for Czech consists of two main parts: the categorial grammar chart parser and the discourse analysis component. Altogether, ten different modules have been implemented. They are briefly described below:

**GENERAL** includes the predicates that call the parsing and discourse analysis procedures, the functions for reading in a natural language string from the input and the help procedures used in the system. In the beginning of this file a start call is made to all modules involved. In other words, the whole system starts by calling this module.

**PARSER** is the module implementing the parsing algorithm and the categorial grammar rules for combining words to phrases.

**LOOKUPS** contains search procedures for the appropriate lexemes in the lexicon.

**LEXICON** contains all lexemes used and procedures for assigning appropriate categories to the input words.

**NOUNS, PRONOUNS, ADJECTIVES** and **VERBS** are names of modules, that include the inflectional procedures of the different parts of speech.

**DISCOURSE** is the module handling the topic-focus recognition and species assignment.

**TESTS** simply includes test procedures and stored test strings.

The system architecture is depicted in figure 5.2, where the arrows indicate how the modules are connected to each other.
Figure 5.2: The system architecture.
5.3 THE PARSING PROCEDURE

5.3.1 PARSE

Parse

The procedure parse/2 takes a list of words (String) as input and returns a two level quasi-logical form, describing the sentence syntactic category together with the grammatical functions involved (GF) and the sentence structure (SenStr).

\[
\text{parse(String,s\#GF:SenStr) : - abolish(edge,3),}
\]
\[
\text{parse1(String,0,EndPos),}
\]
\[
\text{edge(0,EndPos,s\#GF:Str),}
\]
\[
\text{result(Str,SenStr).}
\]

What happens is that first all previously stored edges in the system data base are abolished by the predicate abolish/2. Then the following procedure parse1/3 is called, where the actual parsing takes place. After all words in a string have been parsed, the last edge is called in order to check if there is an edge that corresponds to the whole input string, i.e. of the category 's' (= sentence). This check has to be done, because of the verb's two different categories. The predicate result/2 ends the parsing procedure. What happens here will be explained in the following example.

Consider the sentence (3) below. Parsing this sentence gives the corresponding two level quasi-logical form shown in figure 5.3.

\[
\text{(3) Ten mladý muž dal}
\]
\[
\text{THAT(AM,S,Nom) YOUNG(AM,S,Nom) MAN(AM,S,Nom) GIVE(AM,S,Past)}
\]
\[
\text{knihu ŽENĚ.}
\]
\[
\text{BOOK(F,S,Acc) WOMAN(F,S,Dat)}
\]
\[- That young man gave the book to a woman.
\]

GRAMMATICAL FUNCTIONS:
\[
s\#_779 \& \text{pred(_779,_780)} \& \text{subj(_779,_781)} \& \text{dir_obj(_779,_782)}
\]
\[
\& \text{indir_obj(_779,_783)}
\]

SENTENCE STRUCTURE:
\[
[\text{pronoun(_781,_784,ten.type=dem)},
\]
\[
\text{adjective(_784,_785,mlady)},
\]
\[
\text{noun(_785,muz),}
\]
\[
\text{features= (type=common,pers=3,numb=sing,gend=an_masc,case=nom)),}
\]
\[
\text{verbal(_780,dati,features= (type=finit,e,tense=pass)),}
\]
\[
\text{noun(_782,kniha),}
\]
\[
\text{features= (type=common,pers=3,numb=sing,gend=fem,case=acc)),}
\]
\[
\text{noun(_783,zena),}
\]
\[
\text{features= (type=common,pers=3,numb=sing,gend=fem,case=dat))]
\]

Figure 5.3: Quasi-logical form of sentence (3) above.
In the "sentence structure" in figure 5.3, the nouns and their modifiers are stored separately. The predicate result/2 rebuilds the form of this list, so that all nouns and their modifiers are brought together to form a new category 'nominal', which actually corresponds to a noun phrase. In other words, the nouns and their modifiers belong to the same category - 'nominal' - where the noun is the head of this category and the modifiers are stored in a list. The result is given in figure 5.4.

**GRAMMATICAL FUNCTIONS:**

\[
\text{s\#}_780 & \text{ pred\(_{-780,-781}\) & subj\(_{-780,-782}\) & dir\_obj\(_{-780,-783}\) & indir\_obj\(_{-780,-784}\)}
\]

**SENTENCE STRUCTURE:**

\[
\text{nominal\(_{-782,head=muz,}\)}
\text{ modifiers=[pronoun\(_{ten,type=dem},\text{adjective\(_{mlady})},\)}
\text{ features=\(_{type=\text{common},\text{pers}=3,\text{numb=\text{sing}, gend=an_masc, case=nom}}\)},
\text{ verbal\(_{-781,dat, features=\(_{type=\text{finite}, tense=past}))\), \}
\text{ nominal\(_{-783, head=kniha,}\)}
\text{ modifiers=[]),}
\text{ features=\(_{type=\text{common},\text{pers}=3,\text{numb=\text{sing}, gend=fem, case=acc}}\)},
\text{ nominal\(_{-784, head=zena,}\)}
\text{ modifiers=[]),}
\text{ features=\(_{type=\text{common},\text{pers}=3,\text{numb=\text{sing}, gend=fem, case=dat}}\)})
\]

**Figure 5.4:** Final quasi-logical form of sentence (1) above.

The remainder of this section describes the parsing procedure in more detail.

**Parse1**

The parsing procedure parse1/3 takes as input a list of words and the initial position - 0 - and returns the end position of the string, which corresponds to the number of words involved.

parse1([],Pos,Pos).

\[\text{parse1([Word|Words],Pos,EndPos) :- Pos1 is Pos + 1,}\]
\[\\text{lookup(Lexeme,Agreement,Word),}\]
\[\text{lexeme(Lexeme,Agreement,GF,Str),}\]
\[\text{add_edges(Pos,Pos1,GF;Str),}\]
\[\text{parse1(Words,Pos1,EndPos).}\]

It is a recursive procedure, which begins by finding the end position for the first word on the list. Then a search is made in the lexicon by the predicate lookup/3, which returns the word's lexeme (Lexeme) and the inflectional features (Agreement), i.e. the type, person, number, gender, case or tense, depending on which part of speech it is. The predicate lexeme/4 takes the Lexeme and Agreement as input and returns a quasi-logical form including the lexeme's syntactic category together with the grammatical functions (GF) and a list (Str) including a term that represents the lexeme and its inflectional features, which will at the end be part of the output's "sentence structure" level. After that, the procedure add_edges/3 adds an edge to the database and applies the categorial grammar rules. Finally, parse1 is called recursively with the tail.
of the list and the end position of the first word as arguments. Hence all these procedures are done for every word in the string till the end has been reached.

5.3.2 LOOKUP

The procedure lookup/3 looks up words in the lexicon. It takes a word as input and returns the word's lexeme and inflectional features if any.

lookup(Lexeme,_,Lexeme) :- adv(Lexeme).
lookup(Lexeme,_,Lexeme) :- conj(Lexeme).
lookup(Lexeme,_,Lexeme) :- prep(Lexeme,_,_).
lookup(Lexeme,Agreement,Word) :- lookup_verb(Lexeme,Agreement,Word).
lookup(Lexeme,Agreement,Word) :- lookup_pron(Lexeme,Agreement,Word).
lookup(Lexeme,Agreement,Word) :- lookup_adj(Lexeme,Agreement,Word).
lookup(Lexeme,Agreement,Word) :- lookup_noun(Lexeme,Agreement,Word).

The result of calling this procedure with an inflected word as argument is the word's lexeme, i.e. its basic form, and the appropriate inflectional values, i.e. the person, number, gender, case or tense depending on whether it is a noun, adjective, pronoun or verb.

Adverbs, conjunctions and prepositions have no inflections, which means that their form is always identical to the basic form representing the lexeme. Some examples are given in figure 5.5.

1. (a) ?-lookup(Lexeme,Agreement,eva). (b)?-lookup(Lexeme,Agreement,eve).
   Lexeme = eva,
   Agreement = agr(sing,fem,ins) ;
   no

2. (a) ?-lookup(Lexeme,Agreement,spati). (b)?-lookup(Lexeme,Agreement,spal).
   Lexeme = spati,
   Agreement = agr(1,sing,_,9,pres) ;
   no

3. ?-lookup(Lexeme,Agreement,a).
   Lexeme = a,
   Agreement = _ ;
   no

**Figure 5.5:** Looking up words.
1. two examples of looking up the proper noun 'Eva',
2. two examples of looking up the verb 'spati' (= to sleep),
3. looking up the conjunction 'a' (= and).
5.3.3 LEXEME

The predicate lexeme/4 assigns appropriate syntactic categories to every lexeme in a string and builds a term for the lexeme, involving the lexeme and its inflectional features. The term is stored as a list, which in the output of the parsing will be part of the "sentence structure". All categories that have sentence (s) as their final result also include a representation of the grammatical functions involved, together with the syntactic category.

Figures 5.6 and 5.7 show the lexeme categories of nouns and transitive verbs, respectively. These lexeme categories differ in that the categories of transitive verbs include a description of the grammatical functions involved in the sentence, whereas no corresponding description is included in the category for nouns. The reason is, as mentioned above, that the transitive verbs build a sentence at the end, or in other words they may be the highest node in a parse tree.

Concerning the two lexeme categories for transitive verbs, both categories involve the same grammatical functions. The explanation for this is that although the second category represents a transitive verb that forms a sentence where the subject is omitted, this does not mean that the sentence has no subject at all. The subject is expressed in the verb, where the value of the subject may be found. In other words, the omitted pronoun is looked up and stored in the output, which is needed in translation to English, which does not permit omission of the subject.

Also agreement constraints are defined in the lexeme predicate. We see that in figure 5.6, the inflectional features of a noun phrase representing the subject of a sentence agrees with the inflectional features of the verb. In other words, the noun's person (P), number (N) and gender (G) have the same values as the verb's person, number and gender.

```
lexeme(Noun, agr(N,G,C), np(agr(3,N,G,C))#I, [X]) :-
    proper(Noun,G),
    X =.. [noun,I,Noun,features=(type=proper,pers=3,numb=N,gend=G,case=C)];
    common(Noun,G),
    X =.. [noun,I,Noun,features=(type=common,pers=3,numb=N,gend=G,case=C)].
```

Figure 5.6: Lexeme category for nouns.
(a) FIRST CATEGORY

\[
\text{lexeme(Verb,agr(P,N,G,T),}
\]

\[
s\#S & \text{pred(S,I) & subj(S,Subj) & obj(S,Obj)}
\]

\[
! \text{np(agr(P,N,G,nom))#Subj}
\]

\[
! \text{np(agr(\_,\_,\_,acc))#Obj},
\]

\[
[X]) :-
\]

\[
\text{verb(Verb,tr),}
\]

\[
X =.. [\text{verbal,I,Verb,features=(type=finite,tense=T)}].
\]

(b) SECOND CATEGORY

\[
\text{lexeme(Verb,agr(P,N,G,T),}
\]

\[
s\#S & \text{pred(S,I) & subj(S,Subj) & obj(S,Obj)}
\]

\[
! \text{np(agr(\_,\_,\_,acc))#Obj},
\]

\[
[X,V]) :-
\]

\[
\text{verb(Verb,tr),}
\]

\[
V =.. [\text{verbal,I,Verb,features=(type=finite,tense=T)}],
\]

\[
\text{lookup_pron(Pronoun,agr(prs,P,N,G,nom),\_)}
\]

\[
X =.. [\text{pronoun,Subj,Pronoun,}
\]

\[
\text{features=(type=prs_noemphatic,pers=P,numb=N,}
\]

\[
gend=G,\text{case=nomi}].
\]

**Figure 5.7:** Lexeme categories for transitive verbs.

(a) "First category" with the subject instantiated.

(b) "Second category" without any subject instantiated.

### 5.3.4 ADD EDGES

The procedure add_edges/3 asserts an edge in the system database for every parsed word, including the word's start and end position and its lexeme category. After that, the application of categorial grammar rules takes place in the procedure apply_rules/3, which is looking for another edge to the left of the current edge and tries to combine the two edges to a phrase. For the resulting category an edge is asserted to the database and a new try to combine this category with its leftmost antecedent is made. If no rules can be used, the next word is parsed.

\[
\text{add_edges(Pos,Pos1,Cat) :- assert(edge(Pos,Pos1,Cat)),}
\]

\[
\text{apply_rules(Pos,Pos1,Cat).}
\]

\[
\text{add_edges(\_,\_,\_).}
\]

\[
\text{apply_rules(Pos1,Pos2,Cat1) :- edge(Pos,Pos1,Cat2),}
\]

\[
\text{rule(Cat1,Cat2,MCat),}
\]

\[
\text{add_edges(Pos,Pos2,MCat).}
\]

62
5.3.5 RULE

The rules are stored in the module rule/3. The parser includes the basic categorial grammar rules for functional application and the neutral slash rules for functional application, associativity and composition. An exclamation mark - '!' - denotes the neutral connective. All rules are listed in figures 5.8 and 5.9 below.

(a) rule(Root/Arg:FunStr,Arg:ArgStr,Root:MotherStr)
    :- append(FunStr,ArgStr,MotherStr).

(b) rule(Arg:ArgStr,Root/Arg:FunStr,Root:MotherStr)
    :- append(ArgStr,FunStr,MotherStr).

Figure 5.8: Categorial grammar rule for functional application:
(a) forward application,
(b) backward application.

1. (a) rule(Root!Arg:FunStr,Arg:ArgStr,Root:MotherStr)
    :- append(FunStr,ArgStr,MotherStr).

(b) rule(Arg:ArgStr,Root!Arg:FunStr,Root:MotherStr)
    :- append(ArgStr,FunStr,MotherStr).

2. (a) rule(Root!Arg!Arg1:FunStr,Arg:ArgStr,Root!Arg1:MotherStr)
    :- append(FunStr,ArgStr,MotherStr).

(b) rule(Arg:ArgStr,Root!Arg!Arg1:FunStr,Root!Arg1:MotherStr)
    :- append(ArgStr,FunStr,MotherStr).

3. (a) rule(s#S!s#S1:FunStr1,s#S1!X:FunStr2,s#S1X:MotherStr)
    :- append(FunStr1,FunStr2,MotherStr).

(b) rule(s#S1!X:FunStr1,s#S1s#S1:FunStr2,s#S1X:MotherStr)
    :- append(FunStr1,FunStr2,MotherStr).

Figure 5.9: The categorial grammar rules using neutral slash:
1. functional application rules,
2. functional associativity rules,
3. functional composition rules;
   (a) and (b) denotes forward resp. backward definitions of the rules.
To comment on the figures 5.8 and 5.9, the predicate rule/3 takes two lexeme categories and returns a lexeme category that is the result of combining these two categories. In other words, two expressions are combined to a phrase.

Every lexeme category includes the syntactic category and the corresponding lexeme stored as a term list. The rules take one functor category (e.g. Root/Arg:FunStr) and its argument (Arg:ArgStr) and combine them to the resulting functor category (Root:MotherStr). The function append/3 concatenates the categories term lists (FunStr and ArgStr) to a resulting list (MotherStr), which at the end of parsing forms the "sentence structure" list.

In figure 5.9, it should be noted that the application of the functional composition rules is restricted to lexeme categories with syntactic category 's', i.e. only functors resulting in a sentence may be involved here.

5.3.6 OVERVIEW OF THE PARSING ALGORITHM

The parsing algorithm may be summarised as follows:

1. Begin with an empty chart.
2. Take the next word from the input string and look it up in the lexicon. Assign a lexeme category to the word.
3. Add an edge from the current node to the next node, labelling it with the information about the word's lexeme category. This is the current edge.
4. Test to see if there are any edges already on the chart which end at the node where the current edge begins.
5. Find out whether there are any rules which can be applied to combine these two edges.
6. If so, add the resulting edge and repeat steps 4 to 6 with this new edge as the current edge.
7. If not, return to step 2.

This continues until the end of the string is reached.

This algorithm resembles Algorithm 2, described in section 2.1.5. The difference is that not all categories are assigned at once to a word. In fact, another search is not made until the parse has failed.

Another thing that has to be clarified here is that a word does not obtain several syntactic categories, but rather inflections. The explanation for this fact is that parts of speech in Czech are clearly distinguished, i.e. there is no ambiguity as to the category of a word, whereas the same word form may have different inflectional features.

Graphically the algorithm can be depicted as in figure 5.10.
Figure 5.10: The categorial grammar chart parsing algorithm.
5.4 DISCOURSE ANALYSIS

The whole procedure for analysing a text with regard to the appropriate assignment of species features to every noun involved contains two main parts. The first has already been described, i.e. parsing. The second part will be described below.

First, the structure will be given and then the different parts - the topic and focus identification and the species assignment - will be described in more detail. The chapter closes with a summary of the whole procedure of analysing a text.

5.4.1 PROGRAM STRUCTURE

The main predicate for discourse analysis - discourse/1 - takes as input a list of sentences. First the initial sentence is analysed and then the rest of the sentences.

\[
\text{discourse([Sentence1|Sentences])} \leftarrow \text{abolish(discourse\_base,3)}, \\
\text{\quad initial\_sentence(Sentence1,GF:Str),} \\
\text{\quad other\_sentences(Sentences).}
\]

The initial sentence has no previous context to be compared to, which means that it is treated differently from the rest of the sentences. Compare the figures 5.11 and 5.12 listing the initial\_sentence/2 predicate and other\_sentences/1 predicate respectively.

\[
\text{initial\_sentence(Sentence1,GF:NewStr) \leftarrow} \\
\text{\quad parse(Sentence1,GF:Str),} \\
\text{\quad build\_tf\_lists(Str,Topic,Focus),} \\
\text{\quad add\_features(Str,Topic,Focus,NewStr),} \\
\text{\quad add\_to\_ddbase(NewStr).}
\]

**Figure 5.11:** Initial sentence predicate.

\[
\text{other\_sentences([Sentence1|Sentences])} \leftarrow \\
\text{\quad parse(Sentence,GF:Str),} \\
\text{\quad build\_tf\_lists(Str,Topic,Focus),} \\
\text{\quad check\_focus(Focus,Topic,NewTopic,NewFocus),} \\
\text{\quad check\_subject(NewTopic,NewFocus,NTopic,NFocus),} \\
\text{\quad add\_features(Str,NTopic,NFocus,NewStr),} \\
\text{\quad add\_to\_ddbase(NTopic,NFocus),} \\
\text{\quad other\_sentences(Sentences).}
\]

**Figure 5.12:** Other sentences predicate.

The predicates differ in that the procedure of other\_sentences/1 includes additional functions for comparing the sentences focus part and subject to the previous context.

66
The following sections describe the predicates involved in the discourse analysis in more detail.

5.4.2 TOPIC AND FOCUS IDENTIFICATION

The topic and focus recognition occurs according to the definitions given in chapter 4. The predicate build_t(topic)l(focus)_lists/3 splits the "sentence structure" list into two parts - the Topic list and the Focus list - containing sentence elements that belong to topic and focus respectively. The verb is treated as the main boundary between topic and focus and is included in the topic list. The only time the verb is added to focus instead of topic is when it is the last element in a sentence, i.e. being the carrier of the intonation centre. For noun phrases that are complements of the verb, the following is valid:

(a) The complements preceding the verb are included in the topic.

(b) The complements following the verb are compared to the systemic ordering. The boundary between topic (to the left) and focus (to the right) may be drawn between any two complements, provided that those belonging to the focus are arranged in the surface word order in accordance with the systemic ordering.

\[
\text{build_t_lists}(	ext{Str}, \text{NewTopic}, \text{NewFocus}) ::
\]
\[
\begin{align*}
\text{build_lists} & (\text{Str}, \text{Topic}, \text{Focus}), \\
\text{assign_f} & (\text{Focus}, \text{FL}), \\
\text{systemic_order} & (\text{SO}), \\
\text{check_order} & (\text{FL}, \text{SO}, \text{Topic}, \text{Focus}, \text{NewTopic}, \text{NewFocus}).
\end{align*}
\]

First the predicate build_lists/3 divides the sentence structure list into two parts, with respect to the position of the verb in the sentence. Every sentence element to the left of the verb is placed in the topic list (Topic) and every sentence element to the right of the verb is placed in the focus list (Focus).

The predicate assign_f(unctions)/2 assigns the appropriate semantic functions to the nouns in the focus list in order to check the systemic ordering of them. The systemic ordering (SO) (see section 2.2.3) implemented so far is as follows:

\[
\begin{align*}
\text{actor} \\
\text{time} \\
\text{place} \\
\text{addressee} \\
\text{from_direction} \\
\text{patiens} \\
\text{where-direction}
\end{align*}
\]

The procedure check_order/6 takes as input the list containing the semantic functions of the nouns in focus (FL), the systemic ordering list (SO), the topic and focus lists (Topic, Focus) and returns new topic and focus lists (NewTopic, NewFocus) obtained after comparison of the systemic ordering to the order of the nouns in focus. In other words, the order of nouns in the focus list is compared to the systemic ordering list. If it differs, the nouns that do not follow the systemic ordering are taken out from the focus list and added to the topic list.
instead. The focus list always contains at least one element - the carrier of the intonation centre. 

After this procedure the focus list still only contains its potential members, i.e. the elements in focus can be contextually bound, except for the last element in the sentence, which is always in focus. In order to check if the members of the focus list are contextually bound or not, we have to have access to the previous context.

In the case of the initial sentence, there is no previous context. In other words, the topic and focus recognition stops here, assuming that the focus list contains only contextually unbound elements. What remains is just to add the appropriate features of species to every noun, a process that will be described in the following subsection.

The rest of the sentences, on the other hand, may be compared to the previous context. This is done by storing all noun phrases from the previous sentences in a discourse database, using the predicates add_to_ddbase/1 and add_to_ddbase1/1. The first predicate is used only for the initial sentence and the second predicate for the rest of the sentences, where it is prevented to add the same nouns twice. For every nominal in the "sentence structure", a clause is asserted in the discourse database (discourse_base), listing the head, modifiers and the number of the nominal.

```
add_to_ddbase([nominal(_,head=Noun,Modifiers,features=(_,_,N,_,_))lS]) :-
  assert(discourse_base(Noun,Modifiers,N)),
  add_to_ddbase(S).
add_to_ddbase([lS]) :- add_to_ddbase(S).
add_to_ddbase1([lS]) :-
  add_to_ddbase1(S).
```

The analysis of topic and focus recognition within the rest of the sentences continues with a control of the focus list, which is done in the procedure check_focus/4.

```
check_focus([X],T,T,[X]).
check_focus([nominal(_,head=Noun,M,features=(_,_,N,_,_))lF],T,NT,NF) :-
  discourse_base(Noun,M,N),
  append(T,[nominal(_,head=Noun,M,features=(_,_,N,_,_))],NT),
  check_focus(F,T,T,NF).
check_focus(F,T,T,F).
```

This procedure compares the noun phrases in the focus list to the discourse database (discourse_base). If a noun in the focus list is already stored in the database, i.e. is contextually bound, then it is taken out from the focus list and added to the topic list instead.

Now, if the topic list contains the sentence subject, then the subject has to be checked, which is done by the predicate check_subject/4. The reason is, that the subject may be contextually unbound, though it is part of the topic.
check_subject(T,F,NT,NF) :-
    member(nominal(I,head=Noun,M,features=(_,N,_,case=nom)),T),
    not discourse_base(Noun,M,N),
    delete_noun(nominal(I,head=Noun,M,features=(_,N,_,case=nom)),T,NT),
    append([nominal(I,head=Noun,M,features=(_,N,_,case=nom))],F,NF).

check_subject(T,F,T,F).

The predicate checks whether the subject is already stored in the discourse database. If it is not, then this means that it is mentioned for the first time, i.e. contextually unbound. The subject is then taken out from the topic list and added to the focus list instead.

After this step, the final topic and focus recognition has been reached and the procedures assigning an appropriate species to every noun phrase may proceed.

5.4.3 SPECIES ASSIGNMENT

The procedures assigning species to a noun phrase strictly follows the topic-focus identification. The predicate add_features/4 simply assigns the species feature to a noun phrase depending on whether it is in the topic or focus list. But before that, the nouns are checked to see if they are of the type that have no articles at all (i.e. in English). In the present implementation there are two such noun types:

- common nouns that are preceded by a possessive, demonstrative or indefinite pronoun, and
- proper nouns

The predicate add_features/4 takes as input the "sentence structure" (Str) and the topic and focus lists (Topic, Focus) and returns a new sentence structure list (NewStr) involving the appropriate features of species for every nominal in the list.

First, the nouns that obtain no article are checked. In other words, the proper nouns, nouns modified by possessive, demonstrative or indefinite pronouns obtain the feature species=no, i.e. which means that no article will be added when these nouns are translated into English.

add_features([X|Sem],T,F,[Y|NewP]) :-
    X = nominal(In,H,modifiers=L,features=(Ty,P,N,G,C)),
    (Ty=(type=proper);
     member(pronoun(_,type=poss),L);
     member(pronoun(_,type=dem),L);
     member(pronoun(_,type=indef),L)),
    Y = nominal(In,H,modifiers=L,features=(species=no,Ty,P,N,G,C)),
    add_features(Sem,T,F,NewP).
Then all common nouns in the topic list obtain the feature species=def (i.e. definite article) and all common nouns in the focus list obtain the feature species=indef (i.e. indefinite article).

\[
\text{add}\_\text{features}([\text{XISem}],T,F,[Y|\text{NewP}]) \ :- \ Y = \text{nominal}(\text{In},H,M,\text{features}=(\text{Ty},P,N,G,C)), \\
\quad \text{add}\_\text{features}(\text{Sem},T,F,\text{NewP}).
\]

\[
\text{add}\_\text{features}([\text{XISem}],T,F,[Y|\text{NewP}]) \ :- \ X = \text{nominal}(\text{In},H,M,\text{features}=(\text{Ty},P,N,G,C)), \\
\quad \text{member}(X,T), \\
\quad Y = \text{nominal}(\text{In},H,M,\text{features}=(\text{species}=\text{def},\text{Ty},P,N,G,C)), \\
\quad \text{add}\_\text{features}(\text{Sem},T,F,\text{NewP}).
\]

Other parts of speech are just added to the new list.

\[
\text{add}\_\text{features}([\text{XISem}],T,F,[\text{XINewP}]) \ :- \ \text{add}\_\text{features}(\text{Sem},T,F,\text{NewP}).
\]

\[
\text{add}\_\text{features}([],\ldots,[]).
\]

### 5.5 SUMMARY

The discourse analysis starts with the predicate discourse/1. What happens is that every sentence is parsed by the categorial grammar chart parser, where in the two-level QLF the "sentence structure" list is used in further analyses. The list is split into two parts - topic and focus - with respect to the 'systemic ordering' and the previous context (when available). After that, species is assigned to every noun depending on whether it belongs to topic or focus.

The procedure of discourse analysis may be summarised in more detail in the following way (the appropriate procedures are listed within parentheses):

For every sentence:

(i) Parse the sentence using the categorial grammar chart parser. In the output QLF, use the "sentence structure" list for further analysis. (parse/2)

(ii) Split the sentence in topic and focus, so that the verb complements that precede the verb together with the verb belong to the topic list and the rest of the sentence, i.e. the verb complements that follow the verb are in the focus list. (build_lists/3)

(iii) Regard the complements in the focus list that do not follow the systemic ordering as contextually bound, i.e. extract them from the focus list and add them to the topic list instead. The boundary between topic and focus then lies in front of the first member in the focus list. (check_order/6)

(iv) Always leave the last word in a sentence, i.e. the carrier of the main sentential stress, in focus.
(v) For the species assignment, check all noun phrases in the sentence to see if they should have a species with value 'no', which means that this noun phrase obtains no article in translation to English. Common nouns modified by possessive, demonstrative or indefinite pronouns obtain this species and the same is true for proper nouns. Otherwise add the species - definite or indefinite - to the feature list of a noun phrase depending on whether it belongs to the topic or focus list. Definite species to nouns in the topic list and indefinite to nouns in the focus list. (add_features/4)

(vi) Add all nouns, together with their modifiers and the feature 'number', to the discourse database, so that the nouns in the following sentence can be compared to the database in order to check if a noun has been mentioned earlier in the discourse. (add_to_ddbase/1)

For every sentence, except the initial sentence:

(vii) Add all the nouns in the focus list that are already in the discourse database, i.e. that are contextually bound, to the topic list instead. Make the search from left to right, i.e. if the leftmost element in the focus list is not in the database, then no other elements have to be checked. On the other hand, if an element is already stored in the database, then extract it from the focus list and add it to the topic list instead, and continue the search with the next leftmost element. If there is only one element left in the focus list, i.e. the last word in a sentence, leave it in the focus list. It is the carrier of the main sentential stress, which as mentioned earlier is always in focus. (check_focus/4)

(viii) If the subject of a sentence is part of the topic list and it is not stored in the database, i.e. it is contextually unbound), extract it from the topic list and add it to the focus list instead. (check_subject/4)

Figure 5.13 depicts the set of procedures involved in the whole system. Test runs are given in Appendix I.
(a) The initial sentence.

(b) Other sentences.

Figure 5.13: The set of procedures involved in the whole program.
6 TESTING

6.1 INTRODUCTION

Appendix I contains test runs, first some runs concerning only parsing and
then for the whole system. In the following, I will first discuss the system tests
divided in accordance to the input corpus they take and then some tests
cocerning only the parsing part will be given.

6.2 THE SYSTEM TESTS

6.2.1 CORPUS 1

The first test (test 1) has as input a text taken from a book on language learning,
that was previously discussed in section 2.3.4. The text is repeated here below.

(1) V pokoji stál STŮL.
IN ROOM(IM,S,Loc) BE(IM,S,Past) TABLE(IM,S,Nom)
- In the room was a table.

(2) Na stole ležely knihy
ON TABLE(IM,S,Loc) LAY(F,P,Past) BOOK(F,P,Nom)
a PAPÍR.
AND PAPER(IM,S,Nom)
- On the table there lay books and a paper.

(3) Za knihami hořela SVÍČKA.
BEHIND BOOK(F,P,Ins) BURN(F,S,Past) CANDLE(F,S,Nom)
- Behind the books there burned a candle.

(4) Při svícce četl neznámý
BY CANDLE(F,S,Loc) READ(AM,S,Past) UNKNOWN(AM,S,Nom)
ČLOVĚK.
MAN(AM,S,Nom)
- By the candle a stranger was reading.

The assignment of species features to the noun phrases in the text is fully
satisfactory, i.e. in the case of translation to English the noun phrases included
would obtain the same species as in the corresponding English translations.
The test with its output is stored in Appendix I.
6.2.2 CORPUS 2

The text used as input to the tests 2 to 5 comes from a book on language learning and has also been previously discussed in section 2.3.4. A small adjustment to fit the system lexicon has been made here. The text is as follows:

(5) Do domu vešel MUŽ.
    INTO HOUSE(IM,S,Gen) ENTER(AM,S,Past) MAN(AM,S,Nom)
    - A man entered the house.

(a) Muž vyšel za jednu HODINU.
    MAN(AM,S,Nom) GO-OUT(AM,S,Past) AFTER ONE HOUR(F,P,Nom)
    - The man left after one hour.

(b) Za jednu hodinu vyšel MUŽ.
    AFTER ONE HOUR(F,P,Nom) GO-OUT(AM,S,Past) MAN(AM,S,Nom)
    - A man left after one hour.

The different tests are discussed below. All of them take the initial sentence (5) as input, but the following sentence is different for every test. The input text is given in the beginning of every test and a comment is made.

The species assignment in sentence (5) done by the implemented system completely corresponds to the English translation, i.e. the noun in the prepositional phrase (domu(Gen) = house) has definite species and the newly introduced noun phrase (muz = man) ending the sentence (i.e. in focus) has the species value 'indefinite'.

Test 2

STRING: [do,domu,vesel,muz] = "into house entered man"
STRING: [muz,vyself,za,jednu,hodinu] = "man left after one hour"

Test 2 is using sentence (a) as second sentence, where the noun phrase 'man' obtains definite species, i.e. is considered as definite. In other words the interpretation given in English would be:

A man entered the house.
The man left after one hour.

which corresponds exactly to the sentences (5) and (5a) given above. In other words, the species assignment succeeds here as well.

Test 3

STRING: [do,domu,vesel,muz] = "into house entered man"
STRING: [za,jednu,hodinu,vyself,muz] = "after one hour left man"

Test 3 uses sentence (b) as the second sentence input and here also the distribution of species features corresponds to the English translation, i.e. the noun phrase 'man' in the second sentence is assigned indefinite species feature, and the whole discourse may then be interpreted in English as:
A man entered the house.
A man left after one hour.

i.e. completely corresponding to sentences (5) and (5b) above.

Although the species assignment succeeded here, the noun phrases 'man' (muz) are not interpreted as different individuals, which is shown in the database, where only one instance of noun 'man' (muz) is stored.

discourse_base(dum, modifiers = [], numb = sing).
discourse_base(muz, modifiers = [], numb = sing).
discourse_base(hodina, modifiers = [pronoun(jeden, type = indef)], numb = sing).

Test 4

STRING: [do, domu, vesel, muz] = "into house entered man"
STRING: [za, jednu, hodinu, vysel, ten, muz] = "after one hour left that man"

Test 4 is using a variant of sentence (b), where the last noun phrase includes a demonstrative pronoun (ten muz = 'that man'). In other words, the 'man' (muz) introduced in the initial sentence and the noun phrase 'that man' (ten muz) in the second sentence should be interpreted as the same individual, i.e. the noun phrase in the second sentence should have definite species. The interpretation should be:

A man entered the house.
That man left after one hour.

My system assigns indefinite species to the 'man' in the initial sentence.
A species feature of the value 'none' is assigned to the man in the second sentence, which means that in translation to English the noun phrase has no article, which actually corresponds to the English interpretation. But by looking at the list of stored nouns we realise that the nouns are not interpreted as denoting the same individual:

discourse_base(dum, modifiers = [], numb = sing).
discourse_base(muz, modifiers = [], numb = sing).
discourse_base(hodina, modifiers = [pronoun(jeden, type = indef)], numb = sing).
discourse_base(muz, modifiers = [pronoun(ten, type = dem)], numb = sing).

Two separate instances of 'man' (muz) are listed here, one without any modifiers and one containing the demonstrative pronoun as modifier.

In the topic-focus recognition part the noun phrase 'that man' is interpreted as belonging to the focus, which is true knowing the fact that the noun phrase is in the intonation centre of that sentence (see section 2.3.2, sentences (11) to (13)).
Test 5

STRING: [do, domu, vesel, muz] = "into house entered man"
STRING: [nejaky, muz, vsel, za, jednu, hodinu] = "some man left after one hour"

Test 5 uses a variant of sentence (a), where 'man' is preceded by an indefinite pronoun. In the system the noun 'man' obtains 'none' as the species value, because of the preceding indefinite pronoun. The interpretation here is:

A man entered the house.
Some man left after one hour.

The individuals are not considered as the same, which is shown in the discourse database listing two different instances of 'man' (muz):

discourse_base(dum, modifiers = [], numb = sing).
discourse_base(muz, modifiers = [], numb = sing).
discourse_base(muz, modifiers = [pronoun(nejaky, type = indef)], numb = sing).
discourse_base(hodina, modifiers = [pronoun(jeden, type = indef)], numb = sing).

Also in the topic-focus recognition, the noun phrase 'some man' (nejaky muz) as the subject of that sentence is stored in the focus list, i.e. considered as contextually unbound.

The tests and their outputs may be seen in Appendix I.

6.2.3 CORPUS 3

Test 6 has as input a text taken from a book for children, where no corresponding English translation was available (i.e. the translation here is my own). The text has also been discussed previously in section 2.3.4. Here below the text is repeated with small changes according to the system lexicon adjustment.

(6) samoobsluha je za parkem.
SUPERMARKET(F, S, Nom) BE(3PS, Pres) BEHIND PARK(IM, S, Ins)
-The supermarket is behind a park.

(7) tam stoji v zahradě domu.
THERE STAND(3PP, Pres) IN GARDEN(F, P, Loc) HOUSE(IM, S, Nom)
-There in the garden stands a house.

(8) dom má jméno.
HOUSE(IM, S, Nom) HAVE(3PS, Pres) NAME(N, S, Acc)
-The house has a name.

No problems arise here, the noun phrases in here obtain the species features corresponding to the English translations. The sentence (7), which has a marked word order is also handled by the system (for details see section 2.3.4).
6.2.4 CORPUS 4

The rest of the test runs (test 7 to 9) use my own texts, which are composed of the following sentences:

(9) V pokoji stál STŮL.
    IN ROOM(IM,S,Loc) STAND(IM,S,Past) TABLE(IM,S,Nom)
    - In the room was a table.

(10) U stolu psal muž
     AT TABLE(IM,S,Gen) WRITE(AM,S,Past) MAN(AM,S,Nom)
     DOPIS.
     LETTER(IM,S.Acc)
     - At the table a/the man wrote a letter.

(11) U stolu seděl MUŽ.
     AT TABLE(IM,S,Gen) SIT(AM,S,Past) MAN(AM,S,Nom)
     - At the table sat a man.

(12) Muž dal dopis ŽENĚ.
     MAN(AM,S,Nom) GIVE(AM,S,Past) LETTER(IM,S,Acc)
     WOMAN(F,S,Dat)
     - A man gave the letter to a woman.

(13) Muž dal ženě DOPIS.
     MAN(AM,S,Nom) GIVE(AM,S,Past) WOMAN(F,S,Dat)
     LETTER(IM,S,Acc)
     - A man gave a letter to a woman.

(14) Do pokoje vešla ŽENA.
     INTO ROOM(IM,S,Gen) ENTER(F,S,Past) WOMAN(F,S,Nom)
     - A woman entered the room.

With these tests I want to show, that the system also can handle species assignment in sentences where more than one verb complement follow the verb, regardless of whether the sentence has a primary or marked word order.

The tests are listed below together with their input text.

Test 7

STRING: [v,pokoji,stat,stu] = "in room stood table"
STRING: [u,stolu,psal,muz,dopis] = "at table wrote man letter"
STRING: [muz,dal,dopis,zena] = "man gave letter(acc) to_woman"

The initial sentence in this test has the word order adverbial - verb - actor, which means that the adverbial should be contextually bound, whereas the actor is new to the context (see section 2.2.3 about initial sentences).

In the second sentence the adverbial in the beginning should also be interpreted as contextually bound, according to its position in the sentence (before verb = topic) and also knowing that the noun phrase table (stul) was previously mentioned (introduced). The verb complements following the verb
are ordered in accordance with the systemic ordering (SO) in Czech (actor before patient), which implies that they belong to the focus (i.e., are contextually unbound). In other words the sentence has a primary word order.

The last sentence in this discourse has marked word order in that the verb complements following the verb have the order patient before addressee which is not in accordance with the SO. What this means is that the patient (dopis = letter) should be interpreted as contextually bound in all contexts. In this context the letter is previously mentioned and I assume that the letter in the second sentence corresponds to the letter in the last sentence. In other words the interpretation of this text in English is:

There stood a table in the room.
A man was writing a letter at the table.
The man gave the letter to a woman.

which also corresponds to the species assignment in the output of this test (see Appendix I).

Test 8

STRING: [v,pokoji,stal,stul] = "in room stood table"
STRING: [u,stolu,sedel,muz] = "at table sat man"
STRING: [muz,dal,zene,dopis] = "man gave to woman letter(acc)"

The initial sentence here is the same as in the previous test and in fact the second sentence has the same structure as the first one, i.e. the word order adverbial - verb - actor.
The last sentence has a primary word order, where both the actants following the verb are contextually unbound. The interpretation here is:

There stood a table in the room.
A man sat at the table.
The man gave a letter to a woman.

The system species assignment corresponds to the interpretation given above, which may be seen in the output of this test in Appendix I.

Test 9

STRING: [v,pokoji,stal,stul] = "in room stood table"
STRING: [u,stolu,sedel,muz] = "at table sat man"
STRING: [do,pokoje,vesla,zena] = "into room enter woman"
STRING: [muz,dal,zene,dopis] = "man gave to woman letter(acc)"

The two first sentences are the same as in the previous test. The third sentence has a primary word order and is introducing an individual (zena = woman) to the context (i.e. with the indefinite species value).
The last sentence was also used in the previous sentence, but the difference here is that the sentence has a primary word order, where the addressee (zena = to woman) is contextually bound, i.e. it is known from the previous context (introduced in the third sentence). The interpretation is:
There stood a table in the room.
A man sat at the table.
A woman entered the room.
The man gave a letter to the woman.

The system handles even this problem, i.e. the species assignment corresponds to the interpretation given above (see Appendix I).

6.3 PARSING

A test predicate for parsing has been implemented and some test runs are given in Appendix I, listed with the run time given in minutes and seconds.

For example the shortest test of the tests listed here takes two seconds (test 8) and the longest is more than four minutes long (test 4).

Another thing that increases the run time is the number of forms a word has. As mentioned in section 2.3.1, parts of speech are clearly distinguished in Czech, i.e. there is no ambiguity as to the category of a word, but as to the inflectional form a word has. In other words, a word does not obtain several syntactic categories, but the same word may have different inflectional features.

For example the tests 5 and 6 differ in the run time (2.20 and 0.25 respectively), though they include the same number of words and the input differs only in the last word, which in test 5 is the plural nominative form of the noun 'book', whereas in test 6 the noun is in singular nominative form. The inputs to the tests are listed here below:

**Test 5**
STRING: [na, stole, lezi, knihy] = "on table lie books"

**Test 6**
STRING: [na, stole, lezi, kniha] = "on table lie book"

The different inflectional values for the words in test 5 and test 6 are given here below, listed in the order their are obtained in the parsing procedure.

'na' is a preposition, i.e. a not inflected word, but in combination with a noun as in a prepositional phrase, the locative case is required for the noun.

'stole' as a noun has the form: (IM,S,Loc)

'leži' is a verb, and it has the forms: (3PS,Pres), (3PP,Pres)

'knihy' as a noun obtains the forms: (F,P,Nom), (F,S,Gen), (F,P,Acc)

'kniha' as a noun has the form: (F,S,Nom)

The problem is, if we consider the sentence (1), that the parser starts with the preposition and then continues with the noun 'stole', which obtains the locative case, thus a prepositional phrase is built. The parser continues with the next word - the verb - that has the value singular. No rules applies and a look is made on the noun 'knihy'. The noun has features plural and nominative case
and the parser tries to combine the verb and the noun, but fails, i.e. the verb has the feature singular. A new try follows, but the parser does not "know" that the problem is in the verb. In other words, the parser is trying to combine the verb with all the other feature possibilities for the noun 'knihy', which happens to be the last word in the sentence, and then searching for new possible feature values of the verb. It is the check of the three different forms (listed above) that the noun has that increases the run time in this particular case.

Parsing sentence (2) is not so complicated, because the noun 'knih a' has the feature singular and may then be combined with the verb that also obtains the feature singular in the first call to the lookup predicate.

6.4 SUMMARY

The conclusion is that the system handles the problem of topic-focus recognition and assigns correct species values to the noun phrases, but the database is not working satisfactory, which was shown in the tests 3 and 4.
7 CONCLUSION

The main goal of this work has been to put forth an appropriate analysis to solve the two problems defined in the introduction (see chapter 1), i.e. the problem of parsing in a free word order language and the problem of assigning species to noun phrases in Czech, and to test these ideas by means of a practical implementation.

A categorial grammar chart parser has been implemented with the additional discourse analysis component treating topic-focus recognition and species assignment. The parser handles the problem of free word order by a combination of categorial grammar rules using the neutral connective. The theory of Functional Sentence Perspective was used as an appropriate basis for the analysis of species distribution in Czech.

The parsing module is able to parse rather long and complex sentences, although the run time increases rapidly with the length of the sentence. The discourse analysis module, consisting of a topic-focus recognition component and a species assignment component, is much faster than the parsing part. If we look at the test runs discussed in the previous chapter, we see that the topic-focus recognition works well and the same holds for the assignment of species. It should be mentioned, however, that the test corpus is rather limited and does not contain very complicated structures.

The parsing procedure can be made more efficient by defining a search procedure in such a way that the parser knows what inflectional form of a word it is looking for and thus the current problem of many alternate inflection forms that are active edges can be decreased. In other words, when we parse a sentence, we first look at the verb, which gives information about the types of complements it requires, and then we check the complements to see if they satisfy the constraints. The question is whether this may be achieved with the chart parsing method or whether another parsing method should be chosen instead.

Another interesting question is whether the implemented system could also be applicable to other Slavic languages and to which extent. In other words, it may be worth while to investigate whether other Slavic languages behave the same with respect to species assignment.

And finally, the implementation should be included in a complete machine translation system with transfer and generation modules into another language, in order to test the performance.
REFERENCES


APPENDIX I

This appendix includes the test runs. First some test runs only for parsing are given and then for the whole system.

In the beginning of parsing and system tests examples are shown using the predicate 'main_p' resp. 'main_d', with the natural language input. The user input is indicated by italics and comments are indicated in bold type. The rest of test runs are used with the test predicates - test/1 (for parsing) and dtest/1 (for the whole system). Every test includes the run time, listed within parentheses in minutes and seconds.
PARSING TESTS

First a test with a natural language input using the predicate main_p.

?- main_p.
PLEASE, WRITE IN THE SENTENCE(S) YOU WANT TO PARSE!
END EVERY SENTENCE WITH A FULL STOP.
PLEASE ENTER AN ASTERISK (*) WHEN YOU FINISHED.

>> HALLO. Adam pise dopis. Dopis lezi nastole. Stole na lezi kniha. *

STRING: [hallo]

Checking the input - SORRY - there is no word "hallo" in the lexicon.

******************************************************************************

STRING: [adam,pise,dopis]

Checking the input - DONE.
Parsing - DONE.

GRAMMATICAL FUNCTIONS:
s#_1524&pred(_1524,_1525)&subj(_1524,_1526)&obj(_1524,_1527)

SENTENCE STRUCTURE:
{nominal(_1526,head=adam,modifiers=[],features=(type=proper,pers=3,numb=sing,
gend=an_masc,case=nom)),verbal(_1525,psati,features=(type=finite,tense=pres)),
nominal(_1527,head=dopis,modifiers=[],features=(type=common,pers=3,numb=sing,
gend=in_masc,case=acc))]

******************************************************************************

STRING: [dopis,lezi,nastole]

Checking the input - SORRY - there is no word "nastole" in the lexicon.

******************************************************************************

STRING: [stole,na,lezi,kniha]

Checking the input - DONE.
Parsing - SORRY - cannot parse this sentence.

******************************************************************************

yes
The rest of tests use the test predicate test/1.

Test 1 (0.14)
?- test(7).
STRING: [svicka, hori] = "candle burns"

GRAMMATICAL FUNCTIONS:
s#320&pred(320, 321)&subj(320, 322)

SENTENCE STRUCTURE:
[nominal(_322, head=svicka, modifiers=[], features=(type=common, pers=3, numb=sing, gend=fem, case=nom)), verbal(_321, horeti, features=(type=finite, tense=pres))]

yes

Test 2 (0.24)
?- test(16).
STRING: [muz, otevrel, okno] = "man opened window"

GRAMMATICAL FUNCTIONS:
s#397&pred(397, 398)&subj(397, 399)&obj(397, 400)

SENTENCE STRUCTURE:
[nominal(_399, head=muz, modifiers=[], features=(type=common, pers=3, numb=sing, gend=an_masc, case=nom)), verbal(_398, otevrel, features=(type=finite, tense=past)), nominal(_400, head=okno, modifiers=[], features=(type=common, pers=3, numb=sing, gend=neut, case=acc))]

yes

Test 3 (0.35)
?- test(20).
STRING: [muz, dal, knihu, zene] = "man gave book(acc) to woman"

GRAMMATICAL FUNCTIONS:
s#592&pred(592, 593)&subj(592, 594)&dir_obj(592, 595)&indir_obj(592, 596)

SENTENCE STRUCTURE:
[nominal(_594, head=muz, modifiers=[], features=(type=common, pers=3, numb=sing, gend=an_masc, case=nom)), verbal(_593, dal, features=(type=finite, tense=past)), nominal(_596, head=knihu, modifiers=[], features=(type=common, pers=3, numb=sing, gend=fem, case=acc)), nominal(_596, head=zena, modifiers=[], features=(type=common, pers=3, numb=sing, gend=fem, case=dat))]

yes

Test 4 (4.30)
?- test(53).
STRING: [ten, rozlobeny, muz, dal, tu, zajimavou, cernou, knihu, te, nezname, zene] = "that angry man gave that interesting black book(acc) to that unknown woman"

GRAMMATICAL FUNCTIONS:
s#1866&pred(1866, 1867)&subj(1866, 1868)&dir_obj(1866, 1869)&indir_obj(1866, 1870)

SENTENCE STRUCTURE:
[nominal(_1868, head=muz, modifiers=[pronoun(ten, type=dem), adjective(rozlobeny)], features=(type=common, pers=3, numb=sing, gend=an_masc, case=nom)), verbal(_1867, dal, features=(type=finite, tense=past)), nominal(_1869, head=knihu, modifiers=[pronoun(ten, type=dem), adjective(zajimavou), adjective(cernou)], features=(type=common, pers=3, numb=sing, gend=fem, case=acc)), nominal(_1870, head=zena, modifiers=[pronoun(ten, type=dem), adjective(nezname)], features=(type=common, pers=3, numb=sing, gend=fem, case=dat))]

yes
Test 5 (2.20)
?- test(61).
STRING: [na, stole, lezi, knihy] = "on table lie books"

GRAMMATICAL FUNCTIONS:
s#_507&(pred(_507,_508)&subj(_507,_509))&adverbial(_507,_510,_511)

SENTENCE STRUCTURE:
[preposition(_510, na), nominal(_511, head=stul, modifiers=[], features=(type=common,
pers=3, numb=sing, gend=in_nasc, case=loc)), verbal(_508, lezeti, features=(type=finite,
tense=pres)), nominal(_509, head=kniha, modifiers=[], features=(type=common, pers=3,
numb=plur, gend=fem, case=nom))]

yes

Test 6 (0.25)
?- test(62).
STRING: [na, stole, lezi, knihy] = "on table lie book"

GRAMMATICAL FUNCTIONS:
s#_507&(pred(_507,_508)&subj(_507,_509))&adverbial(_507,_510,_511)

SENTENCE STRUCTURE:
[preposition(_510, na), nominal(_511, head=stul, modifiers=[], features=(type=common,
pers=3, numb=sing, gend=in_nasc, case=loc)), verbal(_508, lezeti, features=(type=finite,
tense=pres)), nominal(_509, head=kniha, modifiers=[], features=(type=common, pers=3,
numb=sing, gend=fem, case=nom))]

yes

Test 7 (0.26)
?- test(70).
STRING: [eva, spi, a, adam, mluvi] = "eva sleeps and adam talks"

GRAMMATICAL FUNCTIONS:
s#_636&_637&_638&_639&_640&_641

SENTENCE STRUCTURE:
nominal(_639, head=eva, modifiers=[], features=(type=proper, pers=3, numb=sing, gend=fem,
case=nom)), verbal(_638, spati, features=(type=finite, tense=pres)), nominal(_641,
head=adam, modifiers=[], features=(type=proper, pers=3, numb=sing, gend=in_nasc,
case=nom)), verbal(_640, mluvi, features=(type=finite, tense=pres))

yes

Test 8 (0.02)
?- test(82).
STRING: [spim] = "I sleep"

GRAMMATICAL FUNCTIONS:
s#_102&pred(_102,_103)&subj(_102,_104)

SENTENCE STRUCTURE:
[pronoun(_104, ja, features=(type=pra_noemphatic, pers=1, numb=sing, gend=_105, case=nom)),
verbal(_103, spati, features=(type=finite, tense=pres))]

yes
**SYSTEM TESTS**

First some tests with a natural language input using the predicate main_d.

1.
?- main_d.
PLEASE, WRITE IN THE SENTENCE(S) YOU WANT TO ANALYSE!
END EVERY SENTENCE WITH A FULL STOP.
PLEASE ENTER AN ASTERISK (*) WHEN YOU FINISHED.

>> Dodomu vesel muz. Muz vysel.*

Checking the input - SORRY - there is no word "dodomu" in the lexicon.
PLEASE TRY AGAIN!

   yes

2.
?- main_d.
PLEASE, WRITE IN THE SENTENCE(S) YOU WANT TO ANALYSE!
END EVERY SENTENCE WITH A FULL STOP.
PLEASE ENTER AN ASTERISK (*) WHEN YOU FINISHED.

>> stole na lezel adam. Adam cetl knihu.*

Checking the input - DONE.

   STRING: [stole, na, lezel, adam]

Checking the initial sentence - SORRY - cannot continue, because parsing of this sentence did not succeeded.
PLEASE TRY AGAIN!

   yes

3.
?- main_d.
PLEASE, WRITE IN THE SENTENCE(S) YOU WANT TO ANALYSE!
END EVERY SENTENCE WITH A FULL STOP.
PLEASE ENTER AN ASTERISK (*) WHEN YOU FINISHED.

>> Do domu vesel muz. Muz vysel hodinu za.*

Checking the input - DONE.

   STRING: [do, domu, vesel, muz]

Checking the initial sentence - DONE.
Parsing the initial sentence - DONE.
Analysing the initial sentence - DONE.

**GRAMMATICAL FUNCTIONS:**

s#_1159&(pred(_1159,_1160)&subj(_1159,_1161))&adverbial(_1159,_1162,_1163)

**SENTENCE STRUCTURE:**

{preposition(_1162,do),nominal(_1163,head=dom,modifiers=[]),features=(species=def,
type=common,pers=3,numb=sing,gend=in_masc,case=gen)},verbal(_1160,vejiti,features=
(type=finite,tense=past),nominal(_1161,head=muz,modifiers=[]),features=(species=indef,typ
e=common,pers=3,numb=sing,gend=an_masc,case=nom))
The rest of tests use the test predicate dtest/1.
After every test run the discourse database is listed, in order to
check which nouns have been stored.

Test 1 (6.39)

?- dtest([101,102,103,104]).
LISTING ALL SENTENCES:

STRING: [v,pokoji,stal,stu1] = "in room stood table"
STRING: [na,stole,lezely,knihy,a,papir] = "on table lay(past) books and paper"
STRING: [za,knhami,horela,svicka] = "behind books burned candle"
STRING: [pri,svicce,cet1,neznamy,clovek] = "by candle slept unknown human_being"

ANALYSES:

STRING: [v,pokoji,stal,stu1]

Parsing the initial sentence - DONE.
Analysing the initial sentence - DONE.

GRAMMATICAL FUNCTIONS:
S#_421&(pred(_421,_422)&subj(_421,_423))&adverbial(_421,_424,_425)

SENTENCE STRUCTURE:
[preposition(_424,v),nominal(_425,head=pokojo,modifiers=[],features=(species=def,
type=common,pers=3,numb=sing,gend=in masc,case=loc)),verbal(_422,statl,features=
(type=finte,tense=past)),nominal(_423,head=stul,modifiers=[],features=
(species=indef,type=common,pers=3,numb=sing,gend=in masc,case=nom))]

*******************************************************************************

STRING: [na,stole,lezely,knihy,a,papir]

Parsing - DONE.
Analysing - DONE.

GRAMMATICAL FUNCTIONS:
S#_1320&(pred(_1320,_1321)&subj(_1320,_1322&_1323))&adverbial(_1320,_1324,_1325)

SENTENCE STRUCTURE:
[preposition(_1324,na),nominal(_1325,head=stul,modifiers=[],features=(species=def,
type=common,pers=3,numb=sing,gend=in masc,case=loc)),verbal(_1321,lezeti,features=
(type=finte,tense=past)),nominal(_1322,head=kniha,modifiers=[],features=(species=indef,type=common,pers=3,
numb=plur,gend=fem,case=nom)),nominal(_1323,head=papir,modifiers=[],features=(species=indef,type=common,pers=3,
numb=sing,gend=in masc,case=nom))]

*******************************************************************************

89
STRING: [za,knihami,herea,svicka]

Parsing - DONE.
Analysing - DONE.

GRAMMATICAL FUNCTIONS:
  s1#_2177&{pred(_2177,_2178)&subj(_2177,_2179))&adverbial(_2177,_2180,_2181)}

SENTENCE STRUCTURE:
  [preposition(_2180,za),nominal(_2181,head=knihy,modifiers=[],features={species=def,
type=common,pers=3,numb=plur,gend=fem,case=ins}),verbal(_2178,herevuti,features=
  {type=finite,tense=past}),nominal(_2179,head=svicka,modifiers=[],features=
  {species=indef,type=common,pers=3,numb=sing,gend=fem,case=nom})]

*******************************************************************************

STRING: [pri,svicce,ctel,neznamy,clovek]

Parsing - DONE.
Analysing - DONE.

GRAMMATICAL FUNCTIONS:
  s1#_3093&{pred(_3093,_3094)&subj(_3093,_3095))&adverbial(_3093,_3096,_3097)}

SENTENCE STRUCTURE:
  [preposition(_3096,pri),nominal(_3097,head=svicka,modifiers=[],features={species=def,
type=common,pers=3,numb=sing,gend=fem,case=loc}),verbal(_3094,cistiti,features=
  {type=finite,tense=past}),nominal(_3095,head=clovek,modifiers=[adjective(neznamy)],
  features={species=indef,type=common,pers=3,numb=sing,gend=an_masc,case=nom})]

*******************************************************************************

yes

?- listing(discourse_base).

/ *
  * discourse_base/3
  */

  discourse_base(poko), modifiers = [], numb = sing).
  discourse_base(stul, modifiers = [], numb = sing).
  discourse_base(knihy, modifiers = [], numb = plur).
  discourse_base(papir, modifiers = [], numb = sing).
  discourse_base(svicka, modifiers = [], numb = sing).
  discourse_base(clovek, modifiers = [adjective(neznamy)], numb = sing).
  yes

Test 2 (0.54)

?- dtest([[111,112]]).

LISTING ALL SENTENCES:

STRING: [do,domu,vesel,muz] = "into house entered man"
STRING: [muz,vesel,za,jednu,hodinu] = "man left after one hour"

ANALYSES:

STRING: [do,domu,vesel,muz]

Parsing the initial sentence - DONE.
Analysing the initial sentence - DONE.
GRAMMATICAL FUNCTIONS:
s#_34& (pred(_348,_349)&subj(_348,_350))&adverbial(_348,_351,_352)

SENTENCE STRUCTURE:
(preposition(_351,do), nominal(_352,head=dum,modifiers=[], features=(species=def, type=common,pers=3,numb=sing,gend=in masc,case=gen)), verbal(_349,vejiti,features=(type=finite,tense=past)), nominal(_350,head=muz,modifiers=[], features=(species=indef,type=common,pers=3,numb=sing,gend=an masc,case=nom))]

******************************************************************************

STRING: [muz,vysel,za,jednu,hodinu]

Parsing - DONE.
Analysing - DONE.

GRAMMATICAL FUNCTIONS:
s#_1079& (pred(_1079,_1080)&subj(_1079,_1081))&adverbial(_1079,_1082,_1083)

SENTENCE STRUCTURE:
[nominal(_1081,head=muz,modifiers=[],features=(species=def,type=common,pers=3, numb=sing,gend=an masc,case=nom)), verbal(_1080,vejiti,features=(type=finite, tense=past)), preposition(_1082,za), nominal(_1083,head=hodina, modifiers=[pronoun(jeden,type=indef)],features=(species=no,type=common,pers=3, numb=sing,gend=fem,case=acc))]

******************************************************************************

? - listing(discourse_base).

/*
 * discourse_base/3
 */
discourse_base(dum, modifiers=[], numb = sing).
discourse_base(muz, modifiers=[], numb = sing).
discourse_base(hodina,modifiers=[pronoun(jeden,type=indef)], numb = sing).

Test 3 (0.54)

? - dtest([111,113]).
LISTING ALL SENTENCES:

STRING: [do,domu,vysel,muz] = "into house entered man"
STRING: [za,jednu,hodinu,vysel,muz] = "after one hour left man"

ANALYSES:

STRING: [do,domu,vysel,muz]

Parsing the initial sentence - DONE.
Analysing the initial sentence - DONE.

GRAMMATICAL FUNCTIONS:
s#_34& (pred(_348,_349)&subj(_348,_350))&adverbial(_348,_351,_352)

SENTENCE STRUCTURE:
(preposition(_351,do), nominal(_352,head=dum,modifiers=[], features=(species=def, type=common,pers=3,numb=sing,gend=in masc,case=gen)), verbal(_349,vejiti,features=(type=finite,tense=past)), nominal(_350,head=muz,modifiers=[], features=(species=indef,type=common,pers=3,numb=sing,gend=an masc,case=nom)))
STRING:  [za,jednu,hodinu,vysel,muz]

Parsing - DONE.
Analysing - DONE.

GRAMMATICAL FUNCTIONS:
s#_1107&(pred(_1107,_1108)&subj(_1107,_1109))&adverbial(_1107,_1110,_1111)

SENTENCE STRUCTURE:
[preposition(_1110,za),nominal(_1111,head=hodina,modifiers=[pronoun(jeden,
type= indef)],features={species=no,type=common,pers=3,numb=sing,gend=fem,case=acc}),
verbal(_1108,vyjti,features={type=finite,tense=past}),nominal(_1109,head=muz,
modifiers=[]),features={species= indef,type=common,pers=3,numb=sing,gend=an masc,
case=nom})]

******************************************************************************

yes

?- listing(discourse_base).

/*
 * discourse_base/3
*/
discourse_base(dum, modifiers = [], numb = sing).
discourse_base(muz, modifiers = [], numb = sing).
discourse_base(hodina,modifiers = [pronoun(jeden,type = indef)],numb = sing).
    yes

Test 4 (1.00)

?- dtest([111,114]).
LISTING ALL SENTENCES:

STRING:  [do,domu,vesel,muz] = "into house entered man"
STRING:  [za,jednu,hodinu,vysel,ten,muz] = "after one hour left that man"

ANALYSES:

STRING:  [do,domu,vesel,muz]

Parsing the initial sentence - DONE.
Analysing the initial sentence - DONE.

GRAMMATICAL FUNCTIONS:
s#_348&(pred(_348,_349)&subj(_348,_350))&adverbial(_348,_351,_352)

SENTENCE STRUCTURE:
[preposition(_351,do),nominal(_352,head=dum,modifiers=[]),features={species=def,
type=common,pers=3,numb=sing,gend=in masc,case=gen}),verbal(_349,vyjti,
features={type=finite,tense=past}),nominal(_350,head=muz,modifiers=[]),
features={species= indef,type=common,pers=3,numb=sing,gend=an masc,case=nom})]

******************************************************************************

STRING:  [za,jednu,hodinu,vysel,ten,muz]

Parsing - DONE.
Analysing - DONE.

92
GRAMMATICAL FUNCTIONS:
s#_1177 & (pred(_1177, _1178) & subj(_1177, _1179)) & adverbial(_1177, _1180, _1181)

SENTENCE STRUCTURE:
(preposition(_1180, za), nominal(_1181, head=hodina, modifiers=[pronoun(jeden, type=indef)], features=(species=no, type=common, pers=3, numb=sing, gend=fem, case=acc)), verbal(_1178, vyjiti, features=(type=finite, tense=past)), nominal(_1179, head=muz, modifiers=[pronoun(ten, type=dem)], features=(species=no, type=common, pers=3, numb=sing, gend=an masc, case=nom))

********************************************************************************
yes

/*
discourse_base/dum, modifiers = [{}], numb = sing).
discourse_base/muz, modifiers = [{}], numb = sing).
discourse_base/hodina, modifiers = [pronoun(jeden, type = indef)], numb = sing).
discourse_base/muz, modifiers = [pronoun(ten, type = dem)], numb = sing).
yes

Test 5 (1.03)

?- dtest([111,115]).
LISTING ALL SENTENCES:

STRING: [do, domu, vesel, muz] = "into house entered man"
STRING: [nejaky, muz, vesel, za, jednu, hodinu] = "some man left after one hour"

ANALYSES:

STRING: [do, domu, vesel, muz]

Parsing the initial sentence - DONE.
Analyzing the initial sentence - DONE.

GRAMMATICAL FUNCTIONS:
s#_348 & (pred(_348, _349) & subj(_348, _350)) & adverbial(_348, _351, _352)

SENTENCE STRUCTURE:
(preposition(_351, do), nominal(_352, head=dum, modifiers=[], features=(species=def, type=common, pers=3, numb=sing, gend=in masc, case=gen)), verbal(_349, vejiti, features=(type=finite, tense=past)), nominal(_350, head=muz, modifiers=[], features=(species=indef, type=common, pers=3, numb=sing, gend=an masc, case=nom))

********************************************************************************

STRING: [nejaky, muz, vesel, za, jednu, hodinu]

Parsing - DONE.
Analyzing - DONE.

GRAMMATICAL FUNCTIONS:
s#_1195 & (pred(_1195, _1196) & subj(_1195, _1197)) & adverbial(_1195, _1198, _1199)

SENTENCE STRUCTURE:
[nominal(_1197, head=muz, modifiers=[pronoun(nejaky, type=indef)], features=(species=no, type=common, pers=3, numb=sing, gend=an masc, case=nom)), verbal(_1196, vyjiti,
Test 6 (1.42)

?- dtest([121,122,123]).
LISTING ALL SENTENCES:

STRING: [samoobsluha,je,za,parkem] = "supermarket is behind park"
STRING: [tam,stoji,v,zahrade,dum] = "there stands in garden house"
STRING: [dum,ma,jmeno] = "house has name"

ANALYSES:

STRING: [samoobsluha,je,za,parkem]

Parsing the initial sentence - DONE.
Analysing the initial sentence - DONE.

GRAMMATICAL FUNCTIONS:
s#_617&(pred(_617, _618)&subj(_617, _619))&adverbial(_617, _620, _621)

SENTENCE STRUCTURE:
[nominal(_619,head=samoobsluha,modifiers=[]),features=(species=def,type=common,pers=3,numb=sing,gend=fem,case-nom)), verbal(_618,byti,features=(type=finite,tense=pres)), preposition(_620, za), nominal(_621, head=park,modifiers=[]),features=(species=inde, type=common,pers=3,numb=sing,gend=in masc,case=ins))]

***********************************************************************

STRING: [tam,stoji,v,zahrade,dum]

Parsing - DONE.
Analysing - DONE.

GRAMMATICAL FUNCTIONS:
s#_1443&((pred(_1443, _1444)&subj(_1443, _1445))&adverbial(_1443, _1446, _1445))&adverbial(_1443, _1448)

SENTENCE STRUCTURE:
[adverb(_1448,tam), verbal(_1444,stati,features=(type=finite,tense=pres)), preposition(_1446, v), nominal(_1445,head=zahrada,modifiers=[]),features=(species=def, type=common,pers=3,numb=sing,gend=fem,case=loc)), nominal(_1445,head=dum,modifiers=[]),features=(species=inde, type=common,pers=3,numb=sing,gend=in masc,case=nom))]

***********************************************************************
STRING: [dum, ma, jmeno]

Parsing - DONE.
Analysing - DONE.

GRAMMATICAL FUNCTIONS:
s#_2189&pred(_2189, _2190)&subj(_2189, _2191)&obj(_2189, _2192)

SENTENCE STRUCTURE:
[nominal(_2191, head=dum, modifiers=[]), features=(species=def, type=common, pers=3, numb=sing, gend=in masc, case=nom)), verbal(_2190, mti, features=(type=finite, tense=pres)), nominal(_2192, head=jmeno, modifiers=[]), features=(species=indef, type=common, pers=3, numb=sing, gend=neut, case=acc))]

****************************
yes

?- listing(discourse_base).
/*
 * discourse_base/3
*/
discourse_base(samoobluha, modifiers = [], numb = sing).
discourse_base(park, modifiers = [], numb = sing).
discourse_base(zahrada, modifiers = [], numb = sing).
discourse_base(dum, modifiers = [], numb = sing).
discourse_base(jmeno, modifiers = [], numb = sing).
yes

Test 7 (3.02)

?- dtest([101, 105, 108]).
LISTING ALL SENTENCES:
STRING: [v, pokoji, stal, stul] = "in room stood table"
STRING: [u, stolu, psal, muz, dopis] = "at table wrote man letter"
STRING: [muz, dal, dopis, zene] = "man gave letter(acc) to woman"

ANALYSES:

STRING: [v, pokoji, stal, stul]

Parsing the initial sentence - DONE.
Analysing the initial sentence - DONE.

GRAMMATICAL FUNCTIONS:
s#_415&pred(_415, _416)&subj(_415, _417)&adverbial(_415, _418, _419)

SENTENCE STRUCTURE:
[preposition(_418, v), nominal(_419, head=pokoj, modifiers=[]), features=(species=def, type=common, pers=3, numb=sing, gend=in masc, case=loc)), verbal(_416, mti, features=(type=finite, tense=past)), nominal(_417, head=stul, modifiers=[]), features=(species=indef, type=common, pers=3, numb=sing, gend=in masc, case=nom))]

****************************
STRING: [u, stolu, psal, muz, dopis]

Parsing - DONE.
Analysing - DONE.

95
GRAMMATICAL FUNCTIONS:
s#_1066&(pred(_1066,_1067)&subj(_1066,_1068)&obj(_1066,_1069))
&adverbal(_1066,_1070,_1071)

SENTENCE STRUCTURE:
[preposition(_1070,u),nominal(_1071,head=stul,modifiers=[],features=(species=def,
type=common,pers=3,numb=sing,gend=in_masc,case=gen)),verbal(_1067,psel1,
features=(type=finite,tense=past)),nominal(_1068,head=muz,modifiers=[],
features=(species=indef,type=common,pers=3,numb=sing,gend=an_masc,case=nom)),
nominal(_1069,head=dopis,modifiers=[],features=(species=indef,type=common,pers=3,
numb=sing,gend=in_masc,case=acc))]

******************************************************************************

STRING:  [muz,dal,dopis,zene]

Parsing - DONE.
Analysing - Done.

GRAMMATICAL FUNCTIONS:
s#_1836&pred(_1836,_1837)&subj(_1836,_1838)&dir_obj(_1836,_1839)
&indir_obj(_1836,_1839)

SENTENCE STRUCTURE:
[nominal(_1838,head=muz,modifiers=[],features=(species=def,type=common,pers=3,
numb=sing,gend=an_masc,case=nom)),verbal(_1837,dati,features=(type=finite,
tense=past)),nominal(_1839,head=dopis,modifiers=[],features=(species=def,type=common,per
s=3,numb=sing,gend=in_masc,case=acc)),nominal(_1839,head=zena,modifiers=[],
features=(species=indef,type=common,pers=3,numb=sing,gend=fem,case=dat))]

******************************************************************************

? - listing(discourse_base).
/*
 * discourse_base/3
 */
discourse_base(pokoj, modifiers = [], numb = sing).
discourse_base(stul, modifiers = [], numb = sing).
discourse_base(muz, modifiers = [], numb = sing).
discourse_base(dopis, modifiers = [], numb = sing).
discourse_base(zena, modifiers = [], numb = sing).

Test 8 (2.39)

? - dtest([101,106,109]).
LISTING ALL SENTENCES:

STRING:  [v,pokoj,stat,stat] = "in room stood table"
STRING:  [u,stu,sel,sel,muz] = "at table sat man"
STRING:  [muz,dal,zena,dopis] = "man gave to woman letter(acc)"

ANALYSES:

STRING:  [v,pokoj,stat,stat]

Parsing the initial sentence - DONE.
Analysing the initial sentence - DONE.

96
GRAMMATICAL FUNCTIONS:
 s#\_415& (pred\(_415\), _416) & subj\(_415\), _417) & adverbial\(_415\), _418, _419)

SENTENCE STRUCTURE:
 [preposition\(_418\), nominal\(_419\), head=pokoj, modifiers=[], features={species=def, type=common, pers=3, numb=sing, gend=in masc, case=loc}, verbal\(_416\), stati, features={type=finite, tense=past}, nominal\(_417\), head=stul, modifiers=[], features={species=ind, type=common, pers=3, numb=sing, gend=in masc, case=nom})

*******************************************************************************

STRING: [u, stol, sedel, muz]

Parsing - DONE.
Analysing - DONE.

GRAMMATICAL FUNCTIONS:
 s#\_990& (pred\(_990\), _991) & subj\(_990\), _992) & adverbial\(_990\), _993, _994)

SENTENCE STRUCTURE:
 [preposition\(_993\), nominal\(_994\), head=stul, modifiers=[], features={species=def, type=common, pers=3, numb=sing, gend=in masc, case=gen}, verbal\(_991\), sedeti, features={type=finite, tense=past}, nominal\(_992\), head=muz, modifiers=[], features={species=ind, type=common, pers=3, numb=sing, gend=in masc, case=nom})

*******************************************************************************

STRING: [miz, dal, zene, dopis]

Parsing - DONE.
Analysing - DONE.

GRAMMATICAL FUNCTIONS:
 s#\_1647&pred\(_1647\), _1648) & subj\(_1647\), _1649) & dir_obj\(_1647\), _1650) & indir_obj\(_1647\), _1651)

SENTENCE STRUCTURE:
 [nominal\(_1649\), head=muz, modifiers=[], features={species=def, type=common, pers=3, numb=sing, gend=in masc, case=nom}, verbal\(_1648\), dat1, features={type=finite, tense=past}, nominal\(_1651\), head=zena, modifiers=[], features={species=ind, type=common, pers=3, numb=sing, gend=fem, case=dat}, nominal\(_1650\), head=dopis, modifiers=[], features={species=ind, type=common, pers=3, numb=sing, gend=in masc, case=acc})

*******************************************************************************

yes

?- listing(discourse_base).

/
+ discourse_base/3
*/
discourse_base(pokoj, modifiers=[], numb=sing).
discourse_base(stul, modifiers=[], numb=sing).
discourse_base(muz, modifiers=[], numb=sing).
discourse_base(zena, modifiers=[], numb=sing).
discourse_base(dopis, modifiers=[], numb=sing).

yes
Test 9 (4.41)

?- dest(101,106,107,109).
LISTING ALL SENTENCES:

STRING: [v,pokoji,stal,stul] = "in room stood table"
STRING: [u,stolu,sekel,muiz] = "at table sat man"
STRING: [do,pokoje,vesla,zena] = "into room enter woman"
STRING: [muiz,dai,zena,dopis] = "man gave to woman letter(acc)"

ANALYSES:

STRING: [v,pokoji,stal,stul]

Parsing the initial sentence - DONE.
Analysing the initial sentence - DONE.

GRAMMATICAL FUNCTIONS:
s#_421&(pred(_421,_422)&subj(_421,_423))&adverbial(_421,_424,_425)

SENTENCE STRUCTURE:
[preposition(_424,v),nominal(_425,head=pokojo,modifiers=[]),features=(species=def,
type=common,pers=3,numb=sing,gend=in_masc,case=loc),verbal(_422,stat1,
features=(type=finte,tense=past)),nominal(_423,head=stul,modifiers=[],
features=(species=indef,type=common,pers=3,numb=sing,gend=in_masc,case=nom))]

******************************************************************************

STRING: [u,stolu,sekel,muiz]

Parsing - DONE.
Analysing - DONE.

GRAMMATICAL FUNCTIONS:
s#_996&(pred(_996,_997)&subj(_996,_998))&adverbial(_996,_999,_1000)

SENTENCE STRUCTURE:
[preposition(_999,u),nominal(_1000,head=stul,modifiers=[]),features=(species=def,
type=common,pers=3,numb=sing,gend=in_masc,case=gen),verbal(_997,seleti,
features=(type=finte,tense=past)),nominal(_998,head=muiz,modifiers=[],
features=(species=indef,type=common,pers=3,numb=sing,gend=an_masc,case=nom))]

******************************************************************************

STRING: [do,pokoje,vesla,zena]

Parsing - DONE.
Analysing - DONE.

GRAMMATICAL FUNCTIONS:
s#_1672&(pred(_1672,_1673)&subj(_1672,_1674))&adverbial(_1672,_1675,_1676)

SENTENCE STRUCTURE:
[preposition(_1675,do),nominal(_1676,head=pokoj,modifiers=[]),features=(species=def,
type=common,pers=3,numb=sing,gend=in_masc,case=gen),verbal(_1673,vejiti,
features=(type=finte,tense=past)),nominal(_1674,head=zena,modifiers=[],
features=(species=indef,type=common,pers=3,numb=sing,gend=fem,case=nom))]

******************************************************************************
STRING: [muz,dal,zene,dopis]

Parsing - DONE.
Analysing - DONE.

GRAMMATICAL FUNCTIONS:
s#_2329&pred(_2329,_2330)&subj(_2329,_2331)&dir_obj(_2329,_2332)
&indir_obj(_2329,_2333)

SENTENCE STRUCTURE:
[nominal(_2331,head=muz,modifiers=[],features=(species=def,type=common,
pers=3,numb=sg,gend=an_masc,case=ncm)),verbal(_2330,dati,features=(type=finit,
tense=past)),nominal(_2333,head=zene,modifiers=[],features=(species=def,type=common,
pers=3,numb=sg,gend=fem,case=dat)),nominal(_2332,head=dopis,modifiers=[],
features=(species=indef,type=common,pers=3,numb=sg,gend=in_masc,case=acc))]

*****************
yes

?- listing(discourse_base).

/*
 * discourse_base/3
 */
discourse_base(pokoj, modifiers=[], numb=sg).
discourse_base(stul, modifiers=[], numb=sg).
discourse_base(muz, modifiers=[], numb=sg).
discourse_base(zene, modifiers=[], numb=sg).
discourse_base(dopis, modifiers=[], numb=sg).
yes

99
APPENDIX II

This appendix includes the Czech declensions of nouns, adjectives and pronouns, and conjugation of verbs.

Appendix A: Table of noun endings. (Harkins; 1968, 276-277)
Appendix B: Conjugation of verbs. (Harkins; 1968, 278-279)
Appendix C: Adjective declensions. (Heim; 1982, 240-241)
Appendix D: Pronoun declensions. (Heim; 1982, 242-244)
## APPENDIX A: TABLE OF NOUN ENDINGS

Note that the accusative case has been placed after the nominative, and the prepositional after the dative.

### Singular

<table>
<thead>
<tr>
<th></th>
<th>Masculine Animates</th>
<th>Masculine Inanimates</th>
<th>Neuters</th>
<th>Feminines</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nom.</td>
<td>-a</td>
<td>-e</td>
<td>-o</td>
<td>-a</td>
</tr>
<tr>
<td>Acc.</td>
<td>-u</td>
<td>-e</td>
<td>-o (-e)</td>
<td>-u</td>
</tr>
<tr>
<td>Gen.</td>
<td>-y</td>
<td>-a</td>
<td>-i</td>
<td>-y</td>
</tr>
<tr>
<td>Dat.</td>
<td>-ovi</td>
<td>-u</td>
<td>-i</td>
<td>-e</td>
</tr>
<tr>
<td>Pr.</td>
<td></td>
<td>-i</td>
<td>-i</td>
<td>like nominative</td>
</tr>
<tr>
<td>Ins.</td>
<td>-ou</td>
<td>-em</td>
<td>-ím</td>
<td>-i</td>
</tr>
<tr>
<td>Voc.</td>
<td>-o</td>
<td>-e</td>
<td>-i</td>
<td>-o</td>
</tr>
</tbody>
</table>

### Plural

<table>
<thead>
<tr>
<th></th>
<th>Masculine Animates</th>
<th>Masculine Inanimates</th>
<th>Neuters</th>
<th>Feminines</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nom.</td>
<td>-1, -ové, -é</td>
<td>-a</td>
<td>-e (-e)</td>
<td>-u</td>
</tr>
<tr>
<td>Voc.</td>
<td>-y for hard nouns; -e for soft nouns</td>
<td>-a</td>
<td>-i</td>
<td>-y</td>
</tr>
<tr>
<td>Acc.</td>
<td>-y for hard nouns; -e for soft nouns</td>
<td>-a</td>
<td>-i</td>
<td>-y</td>
</tr>
<tr>
<td>Gen.</td>
<td>-1</td>
<td>no ending</td>
<td>no ending</td>
<td>no ending</td>
</tr>
<tr>
<td>Dat.</td>
<td>-ém</td>
<td>-ím</td>
<td>-em</td>
<td>-ám</td>
</tr>
<tr>
<td>Pr.</td>
<td>-ech for hard nouns; -ich for soft nouns</td>
<td>-ím</td>
<td>-em</td>
<td>-ách</td>
</tr>
<tr>
<td>Ins.</td>
<td>-y for hard nouns; -i for soft nouns</td>
<td>-ím</td>
<td>-em</td>
<td>-am</td>
</tr>
</tbody>
</table>

These tables provide the basic structure for noun declensions in the Czech language, showing how different grammatical cases and genders are marked by suffixes on the noun.
APPENDIX A

Notes to Appendix A

An asterisk (*) represents endings which palatalize the preceding consonant.

Masculine animate dative and prepositional singular: The endings -ů and -ově alternate for hard nouns; -í and -ouř for soft nouns. Masculines ending in -a take -ově only.

Masculine vocative singular: The normal ending is -ě, but after k, ch and h the ending is -ů.

Masculine inanimate prepositional singular: The ending -ě (-ě), with palatalization of the preceding consonant, occurs for hard nouns. The ending -ů is regular after final r, k, ch and h; otherwise the ending -ě (-ě) is more frequent. Some nouns take either ending, while with some the endings are semantically differentiated, e.g., v obchodě means in the store; v obchodě, in business. See Czech-English Vocabulary for the correct ending when in doubt.

Masculine animate nominative plural: The endings -í and -ově sometimes are alternants, sometimes are distinguished (see Czech-English Vocabulary). Masculines in -a (bláha) take only -ově, except nouns in -ista (houzlista), which take only -ě.

Masculine prepositional plural: Nouns ending in k, ch and h undergo palatalization and end in -ých, -ých and -ých.

Neuter prepositional singular: After final r, k, ch and h the ending is -ů; otherwise -ě (-ě) with palatalization of the preceding consonant is more common. See Czech-English Vocabulary.

Neuter prepositional plural: After k, ch and h the ending is -ých, or -ých with palatalization of the preceding consonant.

Feminine genitive plural: Nouns ending in -ě and -ě usually have -e, -ě in the genitive plural.

Feminine soft nouns: A number of feminines ending in a consonant have the endings of duše in the genitive singular and the entire plural. See Czech-English Vocabulary.

APPENDIX B: CONJUGATION OF VERBS

FIRST CONJUGATION

dělati, to do (impf.)

děláti, to do (impf.)

Present:
délám délám dávám dávám

délás délátě dávás dávate

dělá délají dává dávají

Imperative:
dělej, -me, -te
dávej, -me, -te

Past:
dělal, -a, etc.
dával, -a, etc.

Past pass. part.: délán, déláný
dáván, dávaný

Verbal noun:
délání dávání

Pres. act. part.: délaje, délaje(j)e
dávejte, dávají(j)e

Past act. part.: délal, -a, -e
dával, -a, -e

SECOND CONJUGATION

Infinitives in -ěti (-ěti)

slyšetí, to hear

házeti, to throw

(impf.) (impf.)

Present:
slyším slyšime házím házime

slyšiš slyšíte hází házite

slyší slyší házej házejí

Imperative:
slyš, -me, -te házej, -me, -te

Past:
slyšel, slyšela, etc.
házel, házela, etc.

Past pass. part.: slyšený já
házený

Verbal noun:
slyšení házení

Pres. act. part.: slyše, slyší(j)e
házejte, házejí(j)e

Past act. part.: slyšev, -ši, -še
házev, -ši, -še

Infinitives in -ěti

prosíti, to ask

myslíti, to think

(impf.) (impf.)

Present:
prosím prosíme myslím myslíme

prosiš prosíte mysliš myslič

prosi prosí

myslí myslí

Imperative:
pros, -me, -te

myslí mysleme, myslete
# APPENDIX B

## Infinitives in -ati

<table>
<thead>
<tr>
<th>Hard Stems</th>
<th>Soft Stems</th>
</tr>
</thead>
<tbody>
<tr>
<td>poslatí, to send</td>
<td>vázati, to tie</td>
</tr>
<tr>
<td>(perf.)</td>
<td>(imperf.)</td>
</tr>
</tbody>
</table>

### Present (future):
- poslu posléme: váží (-u) vážeme
- poslě poslite: vážě vážete
- poslě poslou: váže váží (-ou)

### Imperative:
- poslí, posléme, poslite: važ, -me, -te

### Past:
- poslal, -a, etc.: vázal, -a, etc.

### Past pass. part.:
- poslán, poslaný: vázan, vázaný

### Verbal noun:
- poslání: vázání

### Pres. act. part.:
- posle, poslíc(e): váže, vážíc(e)

### Past act. part.:
- poslav, -ši, -še: vázav, -ši, -še

## Infinitives in -ovati

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>dékovati, to thank</td>
<td>kupovati, to buy</td>
</tr>
<tr>
<td>(imperf.)</td>
<td>(imperf.)</td>
</tr>
</tbody>
</table>

### Present:
- dékují (-u) dékujeme: kupují (-u) kupujeme
- dékujíš dékujete: kupuješ kupujete
- dékují dékují: kupuje kupují

### Imperative:
- dékují, -me, -te: kupují, -me, -te

### Past:
- dékoval, -a, etc.: kupoval, -a, etc.

### Past pass. part.:
- dékovan, dékovaný: kupovan, kupovaný

### Verbal noun:
- dékování: kupování

### Pres. act. part.:
- dékují, dékujíc(e): kupuje, kupujíc(e)

### Past act. part.:
- dékovan, -ši, -še: kupovav, -ši, -še

## IRREGULAR VERBS

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>být, to be (imperf.)</td>
<td>mít, to have (imperf.)</td>
</tr>
</tbody>
</table>

### Present:
- jsem: mám máme
- jsi: máš máte
- je (jest; jsou): má máj

### Future:
- budu budeme
- buděš budete
- bude budou

### Imperative:
- buď, -me -te: měj, -me, -te

# APPENDIX B

## Past:
- byl, byla, etc.

### Verbal noun:
- být (jmění)

### Pres. act. part.:
- jsem, jsou (mají)

### Past act. part.:
- byv, -ši, -še

### Fut. act. part.:
- budu, budou, (mají)

### Past:
- jím, jíme

### Pres. act. part.:
- jde, jde (mají)

### Past act. part.:
- jedou, -ši, -še

### Imperative:
- jde, jde, -te: věděj, -ši, -še

### Past pass. part.:
- jedou

### Verbal noun:
- jedou, -ši, -še

### Pres. act. part.:
- jedou, jedou (mají)

### Past act. part.:
- jedou, -ši, -še

### Past:
- věděl, věděla, etc.

### Pres. act. part.:
- věděl, věděla (mají)

### Past act. part.:
- věděl, věděla, -ši, -še

### Imperative:
- věděj, -ši, -še

### Pres. act. part.:
- vědělo, věděla (mají)

### Past act. part.:
- vědělo, věděla, -ši, -še

### Past:
- vědět, věděla, etc.

### Pres. act. part.:
- vědělo, věděla (mají)

### Past act. part.:
- vědělo, věděla, -ši, -še

### Past:
- spěl, spěla, etc.

### Pres. act. part.:
- spěl, spěla (mají)

### Past act. part.:
- spěl, spěla, -ši, -še

### Imperative:
- spěj, -me, -te: spěj, -me, -te

### Past pass. part.:
- spěl, spěl, -ši, -še

### Verbal noun:
- spěl, spěla, -ši, -še

### Pres. act. part.:
- spěl, spěla (mají)

### Past act. part.:
- spěl, spěla, -ši, -še

### Past:
- spěl, spěla, -ši, -še
APPENDIX B

Past: prosil, -a, etc.
Past pass. part.: prošen(y)
Verbal noun: prošení
Pres. act. part.: prose, prosíč(e)
Past act. part.: prosiv, -ši, -še

THIRD CONJUGATION

Monosyllabic Infinitives, Consonant Stems

Present: čísti, to read (impf.) moći, to be able (impf.)
čtu čtěm mohu můžeme
čte čtete může můžete
čte čtou mohu mohou

Imperative: čti, čtěme, čtěte
Past: četl, četla, etc. mohl, mohla, etc.
Past pass. part.: čten(y)
Verbal noun: čtení
Pres. act. part.: čtá, čtouc(e)
Past act. part.: čet, -ši, -še

dostati, to get (perf.) říci, to say (perf.)
dostanu dostaneme řeknu řekneme
dostaneš dostanete řekne řeknou

dostanou dostanou řekně, řekněme, řekněte
Past: dostal, dostala, etc. řekl, řekla, etc.
Past pass. part.: dostal(y)
Verbal noun: dostání
Pres. act. part.: dostana, dostanouc(e)
Past act. part.: dostav, -ši, -še
Pres. act. part. (impf.): řka, řkouc(e)

Monosyllabic Infinitives, Vocalic Stems

kryti, to cover chtiti, to want (impf.)
kryji (-u) kryjeme chci choce
kryje kryjí (-ou) chce choce
Imperative: kryj, -me, -te
Past: kryl, kryla, etc. chtel, -me, -te
Past pass. part.: kryt(y)
Verbal noun: krytí
Pres. act. part.: kryje, kryjíč(e)
Past act. part.: kryv, -ši, -še

Infinitives in -nouti

Vocalic Stems Consonant Stems

minouti, to pass, miss tahnouti, to pull (impf.)
minu mineme tahnú tahneme
mineš minete tahněš tahnete
mine minou táhne tahnou

Imperative: miní, -me, -te
Past: minul, minula, etc. tahní, tahníme, tahníte
Past pass part.: minut(y) tažený
Verbal noun: minutí
Pres. act. part.: mina, minouc(e)
Past act. part.: minuv, -ši, -še

Past: jít, to go on foot jeti, to ride (impf.)
Present: jdu jdu jde jde jde jde jde jdu
Imperative: jdi, jďeme, jďete jde jďeme jde jďete
Past: šel, šla, šlo, etc. jel, jela, etc.
Verbal noun: jítí jí jeda, jduč(e)
Pres. act. part.: šed, -ši, -še
Past act. part.: jev, -ši, -še

APPENDIX B

jítí, to go on foot jeti, to ride (impf.)

Present: jdu jdu jde jde jde jde jde jdu
Imperative: jdi, jďeme, jďete jde jďeme jde jďete
Past: šel, šla, šlo, etc. jel, jela, etc.
Verbal noun: jítí jí jeda, jduč(e)
Pres. act. part.: šed, -ši, -še
Past act. part.: jev, -ši, -še

Monosyllabic Infinitives, Vocalic Stems

kryti, to cover chtiti, to want (impf.)
kryji (-u) kryjeme chci choce
kryje kryjí (-ou) chce choce
Imperative: kryj, -me, -te
Past: kryl, kryla, etc. chtel, -me, -te
Past pass. part.: kryt(y)
Verbal noun: krytí
Pres. act. part.: kryje, kryjíč(e)
Past act. part.: kryv, -ši, -še

Infinitives in -nouti

Vocalic Stems Consonant Stems

minouti, to pass, miss tahnouti, to pull (impf.)
minu mineme tahnú tahneme
mineš minete tahněš tahnete
mine minou táhne tahnou

Imperative: miní, -me, -te tahní, tahníme, tahníte
Past: minul, minula, etc. tahní, tahníme, tahníte
Past pass part.: minut(y) tažený
Verbal noun: minutí tazení
Pres. act. part.: mina, minouc(e) táhna, tahnouc(e)
Past act. part.: minuv, -ši, -še tahnuv, -ši, -še
## APPENDIX C

### ADJECTIVE DECLENSIONS

<table>
<thead>
<tr>
<th>nový</th>
<th>m.inan.</th>
<th>m.anim.</th>
<th>n.</th>
<th>f.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1&amp;5. nový</td>
<td>nový</td>
<td>nový</td>
<td>nové</td>
<td>nové</td>
</tr>
<tr>
<td>2. nového</td>
<td>nového</td>
<td>nového</td>
<td>nové</td>
<td>nové</td>
</tr>
<tr>
<td>3. novému</td>
<td>novému</td>
<td>novému</td>
<td>nové</td>
<td>novou</td>
</tr>
<tr>
<td>4. nový</td>
<td>nového</td>
<td>nové</td>
<td>nové</td>
<td>novou</td>
</tr>
<tr>
<td>5. novém</td>
<td>novém</td>
<td>novém</td>
<td>nové</td>
<td>novou</td>
</tr>
<tr>
<td>7. novým</td>
<td>novým</td>
<td>novým</td>
<td>nové</td>
<td>novou</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>matčín</th>
<th>m.inan.</th>
<th>m.anim.</th>
<th>n.</th>
<th>f.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1&amp;5. matčín</td>
<td>matčín</td>
<td>matčín</td>
<td>matčino</td>
<td>matčína</td>
</tr>
<tr>
<td>2. matčina</td>
<td>matčina</td>
<td>matčina</td>
<td>matčiny</td>
<td>matčiny</td>
</tr>
<tr>
<td>3. matčinu</td>
<td>matčinu</td>
<td>matčinu</td>
<td>matčiny</td>
<td>matčiny</td>
</tr>
<tr>
<td>4. matčín</td>
<td>matčina</td>
<td>matčina</td>
<td>matčiny</td>
<td>matčiny</td>
</tr>
<tr>
<td>5. matčín</td>
<td>matčina</td>
<td>matčina</td>
<td>matčiny</td>
<td>matčiny</td>
</tr>
<tr>
<td>6. matčiným</td>
<td>matčiným</td>
<td>matčiným</td>
<td>matčinými</td>
<td>matčinými</td>
</tr>
<tr>
<td>7. matčinými</td>
<td>matčinými</td>
<td>matčinými</td>
<td>matčinými</td>
<td>matčinými</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>hlavní</th>
<th>m.inan.</th>
<th>m.anim.</th>
<th>n.</th>
<th>f.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1&amp;5. hlavní</td>
<td>hlavní</td>
<td>hlavní</td>
<td>hlavní</td>
<td>hlavní</td>
</tr>
<tr>
<td>2. hlavního</td>
<td>hlavního</td>
<td>hlavního</td>
<td>hlavní</td>
<td>hlavní</td>
</tr>
<tr>
<td>3. hlavním</td>
<td>hlavním</td>
<td>hlavním</td>
<td>hlavní</td>
<td>hlavní</td>
</tr>
<tr>
<td>4. hlavní</td>
<td>hlavní</td>
<td>hlavní</td>
<td>hlavní</td>
<td>hlavní</td>
</tr>
<tr>
<td>6. hlavním</td>
<td>hlavním</td>
<td>hlavním</td>
<td>hlavní</td>
<td>hlavní</td>
</tr>
<tr>
<td>7. hlavním</td>
<td>hlavním</td>
<td>hlavním</td>
<td>hlavní</td>
<td>hlavní</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>hlavních</th>
<th>m.inan.</th>
<th>m.anim.</th>
<th>n.</th>
<th>f.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1&amp;5. hlavních</td>
<td>hlavních</td>
<td>hlavních</td>
<td>hlavních</td>
<td>hlavních</td>
</tr>
<tr>
<td>2. hlavních</td>
<td>hlavních</td>
<td>hlavních</td>
<td>hlavních</td>
<td>hlavních</td>
</tr>
<tr>
<td>3. hlavním</td>
<td>hlavním</td>
<td>hlavním</td>
<td>hlavních</td>
<td>hlavních</td>
</tr>
<tr>
<td>4. hlavní</td>
<td>hlavní</td>
<td>hlavní</td>
<td>hlavních</td>
<td>hlavních</td>
</tr>
<tr>
<td>6. hlavních</td>
<td>hlavních</td>
<td>hlavních</td>
<td>hlavních</td>
<td>hlavních</td>
</tr>
<tr>
<td>7. hlavními</td>
<td>hlavními</td>
<td>hlavními</td>
<td>hlavních</td>
<td>hlavních</td>
</tr>
</tbody>
</table>
## APPENDIX D

### PRONOUN DECLENSIONS

#### Personal Pronouns

<table>
<thead>
<tr>
<th>1&amp;5. já</th>
<th>2. mne/mě</th>
<th>3. mně/mi</th>
<th>4. mne/mě</th>
<th>6. mně</th>
<th>7. mnou</th>
</tr>
</thead>
<tbody>
<tr>
<td>ty</td>
<td>tebe/tě</td>
<td>sobě/si</td>
<td>tebe/tě</td>
<td>sobě/se</td>
<td>tebou</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>1&amp;5. my</th>
<th>2. nás</th>
<th>3. nám</th>
<th>4. nás</th>
<th>6. nás</th>
<th>7. námi</th>
</tr>
</thead>
<tbody>
<tr>
<td>vý</td>
<td>váš</td>
<td>vám</td>
<td>váš</td>
<td>váš</td>
<td>vámi</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>1&amp;5. on</th>
<th>2. jeho/něj/ho</th>
<th>3. jemu/mu</th>
<th>4. jeho/něj/ho</th>
<th>6. něm</th>
<th>7. jim</th>
</tr>
</thead>
<tbody>
<tr>
<td>ono</td>
<td>jeho/ho</td>
<td>jemu/mu</td>
<td>jeho/ho</td>
<td>něm</td>
<td>jim</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>1&amp;5. oni/ona/ony</th>
<th>2. jich</th>
<th>3. jim</th>
<th>4. je</th>
<th>6. nich</th>
<th>7. jimí</th>
</tr>
</thead>
<tbody>
<tr>
<td>jich</td>
<td>jim</td>
<td>je</td>
<td>nich</td>
<td>jimí</td>
<td></td>
</tr>
</tbody>
</table>

#### Possessive Pronouns

<table>
<thead>
<tr>
<th>můj</th>
<th>m.inan.</th>
<th>m.anim.</th>
<th>n.</th>
<th>f.</th>
</tr>
</thead>
<tbody>
<tr>
<td>můj</td>
<td>mé/moje</td>
<td>mého</td>
<td>mé</td>
<td>mé/moje</td>
</tr>
<tr>
<td>mého</td>
<td>měmu</td>
<td>mě</td>
<td>mě</td>
<td>mě/moje</td>
</tr>
<tr>
<td>měmu</td>
<td>mého</td>
<td>mé/moje</td>
<td>mé</td>
<td>mě/moje</td>
</tr>
<tr>
<td>mé</td>
<td>mém</td>
<td>mé</td>
<td>mém</td>
<td>mém</td>
</tr>
<tr>
<td>mém</td>
<td>mým</td>
<td>mýmu</td>
<td>mým</td>
<td>mým</td>
</tr>
<tr>
<td>mýmu</td>
<td>mí/moje</td>
<td>mými</td>
<td>mými</td>
<td>mými</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>náš</th>
<th>náš</th>
<th>naš</th>
<th>naš</th>
</tr>
</thead>
<tbody>
<tr>
<td>náš</td>
<td>našem</td>
<td>našem</td>
<td>naš</td>
</tr>
<tr>
<td>naš</td>
<td>našem</td>
<td>našem</td>
<td>naš</td>
</tr>
<tr>
<td>naš</td>
<td>naši</td>
<td>naši</td>
<td>naš</td>
</tr>
<tr>
<td>naši</td>
<td>naši</td>
<td>naši</td>
<td>naš</td>
</tr>
<tr>
<td>naši</td>
<td>naš</td>
<td>naš</td>
<td>naš</td>
</tr>
<tr>
<td>naš</td>
<td>naš</td>
<td>naš</td>
<td>naš</td>
</tr>
<tr>
<td>naš</td>
<td>naš</td>
<td>naš</td>
<td>naš</td>
</tr>
</tbody>
</table>

#### Miscellaneous Pronouns

<table>
<thead>
<tr>
<th>ten</th>
</tr>
</thead>
<tbody>
<tr>
<td>1&amp;5.</td>
</tr>
<tr>
<td>ten</td>
</tr>
<tr>
<td>to</td>
</tr>
<tr>
<td>ta</td>
</tr>
</tbody>
</table>
### APPENDIX D

<table>
<thead>
<tr>
<th></th>
<th>m.inan.</th>
<th>m.anim.</th>
<th>n.</th>
<th>f.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.&amp;5.</td>
<td>ty</td>
<td>ti</td>
<td>ta</td>
<td>ty</td>
</tr>
<tr>
<td>2.</td>
<td>těch</td>
<td>těch</td>
<td>těch</td>
<td>těch</td>
</tr>
<tr>
<td>3.</td>
<td>těm</td>
<td>těm</td>
<td>těm</td>
<td>těm</td>
</tr>
<tr>
<td>4.</td>
<td>ty</td>
<td>ty</td>
<td>ta</td>
<td>ty</td>
</tr>
<tr>
<td>6.</td>
<td>těch</td>
<td>těch</td>
<td>těch</td>
<td>těch</td>
</tr>
<tr>
<td>7.</td>
<td>těmi</td>
<td>těmi</td>
<td>těmi</td>
<td>těmi</td>
</tr>
</tbody>
</table>

#### jeden

<table>
<thead>
<tr>
<th></th>
<th>jeden</th>
<th>jeden</th>
<th>jedno</th>
<th>jedna</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.&amp;5.</td>
<td>jeden</td>
<td>jeden</td>
<td>jedno</td>
<td>jedna</td>
</tr>
<tr>
<td>2.</td>
<td>jednoho</td>
<td>jednoho</td>
<td>jednoho</td>
<td>jedné</td>
</tr>
<tr>
<td>3.</td>
<td>jednomu</td>
<td>jednomu</td>
<td>jednomu</td>
<td>jedné</td>
</tr>
<tr>
<td>4.</td>
<td>jeden</td>
<td>jednoho</td>
<td>jedno</td>
<td>jednu</td>
</tr>
<tr>
<td>6.</td>
<td>jednom</td>
<td>jednom</td>
<td>jednom</td>
<td>jedné</td>
</tr>
<tr>
<td>7.</td>
<td>jedním</td>
<td>jedním</td>
<td>jedním</td>
<td>jednou</td>
</tr>
</tbody>
</table>

#### všechny

<table>
<thead>
<tr>
<th></th>
<th>všechny</th>
<th>všechny</th>
<th>všechno</th>
<th>všechna</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.&amp;5.</td>
<td>všechny</td>
<td>všechny</td>
<td>všechno</td>
<td>všechna</td>
</tr>
<tr>
<td>2.</td>
<td>všeho</td>
<td>všeho</td>
<td>všeho</td>
<td>vší</td>
</tr>
<tr>
<td>3.</td>
<td>všemu</td>
<td>všemu</td>
<td>všemu</td>
<td>vší</td>
</tr>
<tr>
<td>4.</td>
<td>všechny</td>
<td>všeho</td>
<td>všechno</td>
<td>všechnu</td>
</tr>
<tr>
<td>6.</td>
<td>všem</td>
<td>všem</td>
<td>všem</td>
<td>vší</td>
</tr>
<tr>
<td>7.</td>
<td>vším</td>
<td>vším</td>
<td>vším</td>
<td>vší</td>
</tr>
</tbody>
</table>

#### všechny

<table>
<thead>
<tr>
<th></th>
<th>všechny</th>
<th>všechny</th>
<th>všechna</th>
<th>všechny</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.&amp;5.</td>
<td>všechny</td>
<td>všechny</td>
<td>všechna</td>
<td>všechny</td>
</tr>
<tr>
<td>2.</td>
<td>všech</td>
<td>všech</td>
<td>všech</td>
<td>všech</td>
</tr>
<tr>
<td>3.</td>
<td>všem</td>
<td>všem</td>
<td>všem</td>
<td>všem</td>
</tr>
<tr>
<td>4.</td>
<td>všechny</td>
<td>všechny</td>
<td>všechna</td>
<td>všechny</td>
</tr>
<tr>
<td>6.</td>
<td>všech</td>
<td>všech</td>
<td>všech</td>
<td>všech</td>
</tr>
<tr>
<td>7.</td>
<td>všemi</td>
<td>všemi</td>
<td>všemi</td>
<td>všemi</td>
</tr>
</tbody>
</table>
APPENDIX A: TABLE OF NOUN ENDINGS

Note that the accusative case has been placed after the nominative, and the prepositional after the dative.

### Singular

<table>
<thead>
<tr>
<th></th>
<th>Masculine Animates</th>
<th>Masculine Inanimates</th>
<th>Neuters</th>
<th>Feminines</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>in -a slouha</td>
<td>in -e vědce</td>
<td>in hard cons. muz</td>
<td>in hard cons. syn</td>
</tr>
<tr>
<td>Nom.</td>
<td>-a</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Acc.</td>
<td>-u</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gen.</td>
<td>-y</td>
<td>-e</td>
<td>-a</td>
<td>-e (-e)</td>
</tr>
<tr>
<td>Dat.</td>
<td>-ovi</td>
<td>-u</td>
<td>-i</td>
<td>-u</td>
</tr>
<tr>
<td>Pr.</td>
<td></td>
<td>-em</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ins.</td>
<td>-ou</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Voc.</td>
<td>-o</td>
<td>-e</td>
<td>-l</td>
<td>-e (-e)</td>
</tr>
</tbody>
</table>

### Plural

<table>
<thead>
<tr>
<th></th>
<th>Masculine Animates</th>
<th>Masculine Inanimates</th>
<th>Neuters</th>
<th>Feminines</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>sluhové vůdce</td>
<td>muži (-ové)</td>
<td>syni</td>
<td>hrady</td>
</tr>
<tr>
<td>Nom.</td>
<td>-i*, -ové, -e</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Voc.</td>
<td>-y for hard nouns; -e for soft nouns</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Acc.</td>
<td>-y</td>
<td>-a</td>
<td>-e (-e)</td>
<td>-i</td>
</tr>
<tr>
<td>Gen.</td>
<td>-l</td>
<td>-l</td>
<td>no</td>
<td>-i ending</td>
</tr>
<tr>
<td>Dat.</td>
<td>-im</td>
<td></td>
<td>-em</td>
<td></td>
</tr>
<tr>
<td>Pr.</td>
<td>-ech for hard nouns; -ich for soft nouns</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ins.</td>
<td>-y for hard nouns; -l for soft nouns</td>
<td></td>
<td>-imi</td>
<td>-em (-emi)</td>
</tr>
</tbody>
</table>
APPENDIX A

Notes to Appendix A

An asterisk (*) represents endings which palatalize the preceding consonant.

Masculine animate dative and prepositional singular: The endings -u and -ovi alternate for hard nouns; -1 and -ovi for soft nouns. Masculines ending in -à take -ovi only.

Masculine vocative singular: The normal ending is -e, but after k, ch and h the ending is -u.

Masculine inanimate prepositional singular: The ending -e (-è), with palatalization of the preceding consonant, often occurs for hard nouns. The ending -u is regular after final r, k, ch and h; otherwise the ending -e (-è) is more frequent. Some nouns take either ending, while some of the endings are semantically differentiated, e.g., v obchodě means in the store; v obchodu, in business. See Czech-English Vocabulary for the correct ending when in doubt.

Masculine animate nominative plural: The endings -i and -ovi sometimes are alternants, sometimes are distinguished (see Czech-English Vocabulary). Masculines in -a (sluha) take only -ovi, except nouns in -ista (houslista), which take only -e.

Masculine prepositional plural: Nouns ending in k, ch and h undergo palatalization and end in -ich, -sich and -ich.

Neuter prepositional singular: After final r, k, ch and h the ending is -u; otherwise -e (-è) with palatalization of the preceding consonant is more common. See Czech-English Vocabulary.

Neuter prepositional plural: After k, ch and h the ending is -ach, or -ich with palatalization of the preceding consonant.

Feminine genitive plural: Nouns ending in -ce and -ně usually have -e, -ë in the genitive plural.

Feminine soft nouns: A number of feminines ending in a consonant have the endings of duše in the genitive singular and the entire plural. See Czech-English Vocabulary.

APPENDIX B: CONJUGATION OF VERBS

FIRST CONJUGATION

delati, to do (impf.) dávatì, to give (impf.)

Present:
delám delámme dávám dáváme
deláš deláste dáváš dáváte
delá delájí dává dávaji

Imperative:
deláj, -me, -te dávej, -me, -te
Past:
delal, -a, etc. dával, -a, etc.
Past pass. part.: delán, delaný dávan, dávaný

Verbal noun:
delání dávání
Pres. act. part.: delaje, delajíc(e) dáveje, dávajíc(e)
Past act. part.: delav, -ši, -še dávav, -ši, -še

SECOND CONJUGATION

Infinitives in -eti (-ëti)
slyšeti, to hear házetì, to throw (impf.) házetì, to throw (impf.)

Present:
slyším slyšíme házím házime
slyšá slyšíte hází házite
slyší slyší házi házejí

Imperative:
slyš, -me, -te házej, -me, -te
Past:
slyšel, slyšela, etc. házel, házela, etc.
Past pass. part.: slyšen(y) házen(y)

Verbal noun:
slyšení házení
Pres. act. part.: slyšì, slyšic(e) hážeje, házejíč(e)
Past act. part.: slyšev, -ši, -še házev, -ši, -še

Infinitives in -iti

prositi, to ask mysli, to think (impf.) mysli, to think (impf.)

Present:
prosím prosíme mysím mysíme
prosiš prosíte myslí myslíte
prosi prosí mysli mysli

Imperative:
pros, -me, -te myslí, myslíme, myslete
APPENDIX B

Infinitives in -ati

Hard Stems

Present (future): pošlu, pošleme, pošlete, pošle, pošlu
Past: poslal, -a, etc.
Past part. poslán, poslaný
Verbal noun: posláni
Pres. act. part.: pošlal, -a, etc.
Past act. part.: poslal, -ši, -še

Soft Stems

Present: váži, vážeme, vážete, váži, váži (-u)
Present (future): váži (-u), vážeme
Past: vážal, -a, etc.
Past part. vázan, vázaný
Verbal noun: vázaní
Pres. act. part.: vážal, -a, etc.
Past act. part.: vážal, -ši, -še

Infinitives in -ovatí

děkovatí, to thank (impf.)
kuropyvatí, to buy (impf.)

Present: děkuji, -u, děkujeme, děkuje, děkuje, děkuje, děkuje
Past: děkova, -a, etc.
Past part. děkovaný
Verbal noun: děkování
Pres. act. part.: děkova, děkuje, děkuje
Past act. part.: děkova, -ši, -še

Imperative:

děkuji, -u, děkujeme, děkuje, děkuje, děkuje, děkuje

Imparfait

býti, to be (impf.)
mítí, to have (impf.)

Present: jsem, jsem
jsi, jste
je (jest; jase)
Future: budu, budeme, budete, bude
Imperative: budu, -me, -te

Past: byl, byla, etc.
Verbal noun: bytí
Pres. act. part.: jsem, jsem
Past act. part.: byv, -ši, -še
Past act. part.: bude, bude, etc.

APPENDIX B

měl, měla, etc.
(jměn)
Pres. act. part. jsem, jsem
Past act. part. byv, -ši, -še
Past act. part. bude, bude, etc.

jísti, to eat (impf.)
věděti, to know (impf.)

Present: jím, jíme
jiš, jíte
jí, jí, -te
Past: jed, jedla, etc.
Past part. jeden
Verbal noun: jeden
Pres. act. part.: jed, jedla, etc.
Past act. part.: jed, -ši, -še

věděti, -še

státi, to stand (impf.)

Present: stojím, stojíme
stojiš, stojíte
stoji, stojí
Past: stál, stála, etc.
Verbal noun: stání
Pres. act. part.: stojí, stojí, stojí
Past act. part.: stáv, -ši, -še

státi se, to become, happen (perf.)

Present (future): stojím se stane
stojiš se stane
stoji se stane
Past: staň se, -me se, -te se
Past part. stalo, stala, etc.
Verbal noun: staní
Pres. act. part.: stojí, stojí, stojí
Past act. part.: став, -ši, -še

sťti se, to sleep

Present: bojím se, bojíme se
bojiš se, bojíte se
bojí se, bojí se
Past: bál se, bála se, etc.
Past part. bál
Verbal noun: bání
Pres. act. part.: bojí, bojí, bojí
Past act. part.: бáv, -ši, -še

spáti, to sleep

Present: bojí se, bojí se
bojiš se, bojíte se
bojí se, bojí se
Past: bál se, bála se, etc.
Past pass. part. ...
Verbal noun: spaní
Pres. act. part.: bojí, bojí, bojí
Past act. part.: báv, -ši, -še

spáti, to sleep
APPENDIX B

Past: prosil, -a, etc. mysil, -a, etc.
Past pass. part.: prošen(y) myšlen(y)
Verbal noun: prošení myšlení
Pres. act. part.: prose, prosíc(e) mysle, myslíč(e)
Past act. part.: prosiv, -ši, -še mysliv, -ši, -še

THIRD CONJUGATION

Monosyllabic Infinitives, Consonant Stems

Čísti, to read (impf.) moči, to be able (impf.)
Present: čtu čteme mohu můžeme
čte čtete může můžete
čet čtou mohou
Imperative: čti, čtěme, čtě
Past: četl, četla, etc. mohl, mohla, etc.
Past pass. part.: čtěn(y)
Verbal noun: čtení možení
Pres. act. part.: čta, čtouc(e) moha, mohouc(e)
Past act. part.: čet, -ší, -še moh, -ši, -še
dostati, to get (perf.) říci, to say (perf.)
Future: dostanu dostaneme řeknu řekneme
dostaněs dostanete řekněs řeknete
dostane dostanou řeknou
Imperative: dostal, dostala, etc. řekl, řekla, etc.
Past: dostal, dostala, etc. řekl, řekla, etc.
Past pass. part.: dostal, dostala, etc. řekl, řekla, etc.
Verbal noun: dostání řečení
Pres. act. part.: dostana, dostanouc(e) řekna, řeknouc(e)
Past act. part.: dostav, -ši, -še řek, -ši, -še
Pres. act. part. (impf.): řka, řkouc(e)

APPENDIX B

jít, to go on foot
jedt, to ride (impf.)

Present: jdu jde jdu jedu jedeme jedě
jde jdu jedu jedeme jedě
Imperative: jdi, jděme, jděte jel, jela, etc.
Past: šel, šla, šlo, etc. jel, jela, etc.
Verbal noun: jítí jítí
Pres. act. part.: jda, jduc(e) jeda, jedouc(e)
Past act. part.: šed, -ší, -še jev, -ši, -še

Monosyllabic Infinitives, Vocalic Stems

krytí, to cover ochtítí, to want

Present: kryjí (-u) kryjeme choji chojeme
kryje kryjete choje chojete
kryjí (-ou) chojí
Imperative: kryj, -me, -te chtěj, -me, -te
Past: kryl, kryla, etc. chtel, chtěla, etc.
Past pass. part.: kryt(y)
Verbal noun: krytí chtění
Pres. act. part.: kryje, kryjíč(e) chtěje, chtějíč(e)
Past act. part.: kryv, -ší, -še chtvět, -ši, -še

Infinitives in -nouti

Vocalic Stems
minnouti, to pass, miss

Consonant Stems
táhnouti, to pull

Present: minu minem minou tahu tánhme
tahně tahněte
mine minete táhne táhnou
míns -me, -te tání, tánhme, tánhnte
Imperative: minul, minula, etc. tání, táhla, etc.
Past: minut(y)
Verbal noun: minutí
tážení
Pres. act. part.: mina, minouc(e) táhna, táhnouc(e)
Past act. part.: minuv, -ši, -še táhnuv, -ši, -še
# APPENDIX C

## ADJECTIVE DECLENSIONS

<table>
<thead>
<tr>
<th>nový</th>
<th>m.inan.</th>
<th>m.anim.</th>
<th>n.</th>
<th>f.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1&amp;5. nový</td>
<td>nový</td>
<td>nové</td>
<td>nová</td>
<td></td>
</tr>
<tr>
<td>2. nového</td>
<td>nového</td>
<td>nového</td>
<td>nové</td>
<td></td>
</tr>
<tr>
<td>3. novému</td>
<td>novému</td>
<td>novému</td>
<td>nové</td>
<td></td>
</tr>
<tr>
<td>4. nový</td>
<td>nového</td>
<td>nové</td>
<td>novou</td>
<td></td>
</tr>
<tr>
<td>6. novém</td>
<td>novém</td>
<td>nové</td>
<td>nové</td>
<td></td>
</tr>
<tr>
<td>7. novým</td>
<td>novým</td>
<td>novým</td>
<td>novou</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>hlavní</th>
<th>m.inan.</th>
<th>m.anim.</th>
<th>n.</th>
<th>f.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1&amp;5. hlavní</td>
<td>hlavní</td>
<td>hlavní</td>
<td>hlavní</td>
<td></td>
</tr>
<tr>
<td>2. hlavního</td>
<td>hlavního</td>
<td>hlavního</td>
<td>hlavní</td>
<td></td>
</tr>
<tr>
<td>3. hlavnímu</td>
<td>hlavnímu</td>
<td>hlavnímu</td>
<td>hlavní</td>
<td></td>
</tr>
<tr>
<td>4. hlavní</td>
<td>hlavního</td>
<td>hlavní</td>
<td>hlavní</td>
<td></td>
</tr>
<tr>
<td>6. hlavním</td>
<td>hlavním</td>
<td>hlavním</td>
<td>hlavní</td>
<td></td>
</tr>
<tr>
<td>7. hlavním</td>
<td>hlavním</td>
<td>hlavní</td>
<td>hlavní</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>matčin</th>
<th>m.inan.</th>
<th>m.anim.</th>
<th>n.</th>
<th>f.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1&amp;5. matčin</td>
<td>matčin</td>
<td>matčino</td>
<td>matčina</td>
<td></td>
</tr>
<tr>
<td>2. matčina</td>
<td>matčina</td>
<td>matčina</td>
<td>matčiny</td>
<td></td>
</tr>
<tr>
<td>3. matčinu</td>
<td>matčinu</td>
<td>matčinu</td>
<td>matčině</td>
<td></td>
</tr>
<tr>
<td>4. matčin</td>
<td>matčina</td>
<td>matčino</td>
<td>matčině</td>
<td></td>
</tr>
<tr>
<td>6. matčině</td>
<td>matčině</td>
<td>matčině</td>
<td>matčině</td>
<td></td>
</tr>
<tr>
<td>7. matčiným</td>
<td>matčiným</td>
<td>matčiným</td>
<td>matčině</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>matčiny</th>
<th>m.inan.</th>
<th>m.anim.</th>
<th>n.</th>
<th>f.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1&amp;5. matčiny</td>
<td>matčiny</td>
<td>matčina</td>
<td>matčiny</td>
<td></td>
</tr>
<tr>
<td>2. matčíných</td>
<td>matčíných</td>
<td>matčíných</td>
<td>matčíných</td>
<td></td>
</tr>
<tr>
<td>3. matčíným</td>
<td>matčíným</td>
<td>matčíným</td>
<td>matčíným</td>
<td></td>
</tr>
<tr>
<td>4. matčiny</td>
<td>matčina</td>
<td>matčina</td>
<td>matčiny</td>
<td></td>
</tr>
<tr>
<td>6. matčíných</td>
<td>matčíných</td>
<td>matčíných</td>
<td>matčíných</td>
<td></td>
</tr>
<tr>
<td>7. matčínými</td>
<td>matčínými</td>
<td>matčínými</td>
<td>matčínými</td>
<td></td>
</tr>
</tbody>
</table>
# APPENDIX D

## PRONOUN DECLENSIONS

### Personal Pronouns

<table>
<thead>
<tr>
<th>1&amp;5. já</th>
<th>ty</th>
<th>sebe/se</th>
</tr>
</thead>
<tbody>
<tr>
<td>2. mne/mě</td>
<td>tebe/tě</td>
<td></td>
</tr>
<tr>
<td>3. mně/mi</td>
<td>tobě/tí</td>
<td>sobě/sí</td>
</tr>
<tr>
<td>4. mne/mě</td>
<td>tebe/tě</td>
<td>sebe/se</td>
</tr>
<tr>
<td>6. mně</td>
<td>toběʼ</td>
<td>sobě</td>
</tr>
<tr>
<td>7. mnou</td>
<td>tebou</td>
<td>sebou</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>1&amp;5. my</th>
<th>vy</th>
<th>sebou</th>
</tr>
</thead>
<tbody>
<tr>
<td>2. nás</td>
<td>vás</td>
<td></td>
</tr>
<tr>
<td>3. nám</td>
<td>vám</td>
<td></td>
</tr>
<tr>
<td>4. nás</td>
<td>vás</td>
<td></td>
</tr>
<tr>
<td>6. nás</td>
<td>vás</td>
<td></td>
</tr>
<tr>
<td>7. námi</td>
<td>vámi</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>1&amp;5. on</th>
<th>ono</th>
<th>ona</th>
</tr>
</thead>
<tbody>
<tr>
<td>2. jeho/něj/ho</td>
<td>jeho/ho</td>
<td>jí</td>
</tr>
<tr>
<td>3. jemu/mu</td>
<td>jemu/mu</td>
<td>jí</td>
</tr>
<tr>
<td>4. jeho/něj/ho</td>
<td>je/ho</td>
<td>jí</td>
</tr>
<tr>
<td>6. něm</td>
<td>něm</td>
<td>jí</td>
</tr>
<tr>
<td>7. jím</td>
<td>jím</td>
<td>jí</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>1&amp;5. oni/ona/ony</th>
<th>ono</th>
<th>ona</th>
</tr>
</thead>
<tbody>
<tr>
<td>2. jich</td>
<td>jeho/ho</td>
<td>jí</td>
</tr>
<tr>
<td>3. jim</td>
<td>jemu/mu</td>
<td>jí</td>
</tr>
<tr>
<td>4. je</td>
<td>je/ho</td>
<td>jí</td>
</tr>
<tr>
<td>6. nich</td>
<td>něm</td>
<td>jí</td>
</tr>
<tr>
<td>7. jimí</td>
<td>jím</td>
<td>jí</td>
</tr>
</tbody>
</table>

### Possessive Pronouns

<table>
<thead>
<tr>
<th>1&amp;5. múj</th>
<th>můj</th>
<th>mé/moje</th>
<th>má/moje</th>
</tr>
</thead>
<tbody>
<tr>
<td>2. mého</td>
<td>mého</td>
<td>mého</td>
<td>mé</td>
</tr>
<tr>
<td>3. mému</td>
<td>mému</td>
<td>mému</td>
<td>mé</td>
</tr>
<tr>
<td>4. můj</td>
<td>mého</td>
<td>mé/moje</td>
<td>mou/moju</td>
</tr>
<tr>
<td>6. měm</td>
<td>měm</td>
<td>měm</td>
<td>mé</td>
</tr>
<tr>
<td>7. mým</td>
<td>mým</td>
<td>mým</td>
<td>mou</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>1&amp;5. mé/moje</th>
<th>mí/moju</th>
<th>má/moje</th>
<th>mé/moje</th>
</tr>
</thead>
<tbody>
<tr>
<td>2. mých</td>
<td>mých</td>
<td>mých</td>
<td>mých</td>
</tr>
<tr>
<td>3. mým</td>
<td>mým</td>
<td>mým</td>
<td>mým</td>
</tr>
<tr>
<td>4. mé/moje</td>
<td>mé/moje</td>
<td>má/moje</td>
<td>mé/moje</td>
</tr>
<tr>
<td>6. mých</td>
<td>mých</td>
<td>mých</td>
<td>mých</td>
</tr>
<tr>
<td>7. mými</td>
<td>mými</td>
<td>mými</td>
<td>mými</td>
</tr>
</tbody>
</table>

### Miscellaneous Pronouns

<table>
<thead>
<tr>
<th>1&amp;5. ten</th>
<th>ten</th>
<th>to</th>
<th>ta</th>
</tr>
</thead>
<tbody>
<tr>
<td>2. toho</td>
<td>toho</td>
<td>toho</td>
<td>té</td>
</tr>
<tr>
<td>3. tomu</td>
<td>tomu</td>
<td>tomu</td>
<td>té</td>
</tr>
<tr>
<td>4. ten</td>
<td>toho</td>
<td>to</td>
<td>tu</td>
</tr>
<tr>
<td>6. tom</td>
<td>tom</td>
<td>tom</td>
<td>té</td>
</tr>
<tr>
<td>7. tím</td>
<td>tím</td>
<td>tím</td>
<td>tou</td>
</tr>
<tr>
<td></td>
<td>m.inan.</td>
<td>m.anim.</td>
<td>n.</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>1&amp;5.</td>
<td>ty</td>
<td>ti</td>
<td>ta</td>
</tr>
<tr>
<td>2.</td>
<td>těch</td>
<td>těch</td>
<td>těch</td>
</tr>
<tr>
<td>3.</td>
<td>těm</td>
<td>těm</td>
<td>těm</td>
</tr>
<tr>
<td>4.</td>
<td>ty</td>
<td>ty</td>
<td>ta</td>
</tr>
<tr>
<td>6.</td>
<td>těch</td>
<td>těch</td>
<td>těch</td>
</tr>
<tr>
<td>7.</td>
<td>těmi</td>
<td>těmi</td>
<td>těmi</td>
</tr>
</tbody>
</table>

### jeden

<table>
<thead>
<tr>
<th></th>
<th>jeden</th>
<th>jeden</th>
<th>jedno</th>
<th>jedna</th>
</tr>
</thead>
<tbody>
<tr>
<td>1&amp;5.</td>
<td>jeden</td>
<td>jeden</td>
<td>jedno</td>
<td>jedna</td>
</tr>
<tr>
<td>2.</td>
<td>jednoho</td>
<td>jednoho</td>
<td>jednoho</td>
<td>jedné</td>
</tr>
<tr>
<td>3.</td>
<td>jednomu</td>
<td>jednomu</td>
<td>jednomu</td>
<td>jedné</td>
</tr>
<tr>
<td>4.</td>
<td>jeden</td>
<td>jednoho</td>
<td>jedno</td>
<td>jednu</td>
</tr>
<tr>
<td>6.</td>
<td>jednom</td>
<td>jednom</td>
<td>jednom</td>
<td>jedné</td>
</tr>
<tr>
<td>7.</td>
<td>jedním</td>
<td>jedním</td>
<td>jedním</td>
<td>jednou</td>
</tr>
</tbody>
</table>

### všechn

<table>
<thead>
<tr>
<th></th>
<th>všechn</th>
<th>všechn</th>
<th>všechno</th>
<th>všechna</th>
</tr>
</thead>
<tbody>
<tr>
<td>1&amp;5.</td>
<td>všechn</td>
<td>všechn</td>
<td>všechno</td>
<td>všechna</td>
</tr>
<tr>
<td>2.</td>
<td>všeho</td>
<td>všeho</td>
<td>všeho</td>
<td>vši</td>
</tr>
<tr>
<td>3.</td>
<td>všemu</td>
<td>všemu</td>
<td>všemu</td>
<td>vši</td>
</tr>
<tr>
<td>4.</td>
<td>všechn</td>
<td>všeho</td>
<td>všechno</td>
<td>všechnu</td>
</tr>
<tr>
<td>6.</td>
<td>všem</td>
<td>všem</td>
<td>všem</td>
<td>vši</td>
</tr>
<tr>
<td>7.</td>
<td>vším</td>
<td>vším</td>
<td>vším</td>
<td>vši</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>všichni</th>
<th>všichni</th>
<th>všechna</th>
<th>všichny</th>
</tr>
</thead>
<tbody>
<tr>
<td>1&amp;5.</td>
<td>všichni</td>
<td>všichni</td>
<td>všechna</td>
<td>všichny</td>
</tr>
<tr>
<td>2.</td>
<td>všech</td>
<td>všech</td>
<td>všech</td>
<td>všech</td>
</tr>
<tr>
<td>3.</td>
<td>všem</td>
<td>všem</td>
<td>všem</td>
<td>všem</td>
</tr>
<tr>
<td>4.</td>
<td>všechny</td>
<td>všechny</td>
<td>všechna</td>
<td>všechny</td>
</tr>
<tr>
<td>6.</td>
<td>všech</td>
<td>všech</td>
<td>všech</td>
<td>všech</td>
</tr>
<tr>
<td>7.</td>
<td>všemi</td>
<td>všemi</td>
<td>všemi</td>
<td>všemi</td>
</tr>
</tbody>
</table>