Relations between Inflectional and Derivation Patterns

Karel Pala
Faculty of Informatics
Masaryk University Brno
pala@fi.muni.cz

Radek Sedláček
Faculty of Informatics
Masaryk University Brno
rsedlac@fi.muni.cz

Marek Veber
Faculty of Informatics
Masaryk University Brno
mara@fi.muni.cz

Abstract

One of the main goals of this paper is to describe a formal procedure linking inflectional and derivational processes in Czech and to indicate that they can be, if appropriate tools and resources are used, applied to other Slavonic languages. The tools developed at the NLP Laboratory FI MU, have been used, particularly the morphological analyser `ajka` and the program `I_par` for processing and maintaining the morphological database.

1 Introduction

In this paper we report on an exploration of the relations between inflection and word derivation (WD) in Czech. At first the *inflectional morphology* has to be mentioned which deals with formation of word forms by inflection, i. e. by morphological processes like the declension of nouns, adjectives, pronouns and numerals, the conjugation of verbs, and also forming degrees of adjectives and adverbs. The inflectional morphology (in Czech) is formally quite well described and at present we have reliable software tools at our disposal that make it possible both to generate and recognise arbitrary Czech word forms. They are `ajka` a morphological analyser and generator (Sedláček and Smrž, 2001) and `I_par` a program for morphological database (Veber, 2002).

The second area is *derivational morphology*, i. e. word derivation as such – it describes the processes of the derivation of new words (one word expressions) as distinct from basic ones (word bases). These processes operate on the morpheme level whose results follow from the ways of combining morphemes (prefixes, stems or roots, infixes, suffixes) using suffixation, prefixation and stem alterations (infixation). We have to bear in mind that the consequences of these formal processes have a semantic nature.

The relations between the WD processes and inflectional morphology have been extensively studied in Czech linguistic literature, see e. g. (Dokulil, 1962; Karlík et al., 1995; Petr, 1986) where one may find informal descriptions of the WD processes using terms like *"fundace" (basic derivation), as well as mutation, transposition, modification, adaptation*, and others.

The most interesting linguistic analyses link the derivation patterns with the inflectional ones. For example, in Karlík (1995) it is pointed out that the nouns with the suffix *-tel* having agentive meaning belong to the inflectional paradigm *muž (man)*. In other words, it is possible to say that some inflectional patterns determine the sets of derivational suffixes forming semantically consistent groups of nouns.

In this paper we set it as our task to map the relations between inflectional and word derivational patterns. In comparison with previous research we present the exploration of a large collection of data: our stem dictionary for Czech contains 385,066 items. The association of selected inflectional and WD patterns has been performed semi-automatically with the tool `I_par` and the

Czech morphological database which contains all the necessary information about the inflectional paradigms in Czech (2,042 paradigms for all 10 POS).

The WD relations as they are described in the linguistic WD theories can be, after some modifications, appropriately linked to the semantic relations as they are used in contemporary ontologies and then applied within the inference engines that form a necessary part of the natural language processing mechanisms.

However, it is true that algorithmic descriptions of WD relations have been worked out only recently and they do not cover the WD processes in all their complexity but just at the basic level.

2 Inflectional analysis

The inflectional analysis is a part of the complex morphemic decomposition of a word and its first task is to identify in a given word form the stem and ending. If the word form is not a lemma (basic form of the word) it is necessary to associate the stem with its respective ending.
It has to be kept in mind that the derivational analysis deals only with lemmata and not with the individual word forms.

For the purpose of this work we perform inflectional analysis with the tool (program) `ajka` (Sedláček, 1999) which is further developed in the NLP Laboratory (Sedláček and Smrž, 2001) at FI MU. It is also used as a standard tool for lemmatization (identification of the basic word forms) and as a morphological tagger.

3 The Basic WD Relation = *"fundace"*

In the Czech WD theory all words belonging to a given word stock are classified either as *motivated* or *non-motivated*.

For *motivated* words we are able to describe their meaning using another word or words, e. g. *cvičiště (exercising ground, drill square)* is a place where exercising is done. *Non-motivated* words cannot be described in the same way. Their meaning consists only in signifying the thing, which is why they are sometimes called words-signs, as e. g. *stůl (table), tráva (grass)*. Thus it is obvious that when making WD analyses we are more interested in *motivated* words, since only they display the derivational structure from which we are able to reconstruct the process of their formation.

Formation of *motivated* word follows the basic WD relation called *"fundace"* (Dokulil, 1962). When we find the *"fundace"*-relation for all *motivated* words we obtain a complex of relations that form a hierarchical structure. In WD theory this structure is usually called *word derivation nest*, *word derivation lattice* and *word derivation sequence*. Word derivation nests (WDN) will be our principal focus.

WDN can be defined as a set of all related words that gather around a common stem. The core of WDN is a *root* of the non-motivated word and other words in WDN are derived from it either directly as in *led (ice)→led-ový (icy)* or indirectly through another related word. In this way *word derivation sequences* are created, such as *led (ice)→led-ový (icy)→ledov-ec (glacier)→ledovc-ový (glacial)*.

The appropriate formal apparatus for representing these relations and structures are graphs. For this purpose the special types of graphs (see Fig. 1), particularly graph-trees are used which further link up into forests. Strictly speaking we are dealing mainly with forests.

The tree nodes are labelled with lemmata, and the next node is created dependent on its predecessor. The individual trees then represent the WDN, the WD unions have just one level, subtrees and sequence correspond to the paths in the graph.

4 The Semantic Aspects of the Basic WD Relation

The semantic component of the *"fundace"*-relation consists in the fact that meaning of *"funded"* words can be derived from the meaning of the *"funding"* one, that they are semantically linked and that language users know these relations and use them in understanding language and also in enriching their knowledge about the universe. Thus *učitel (teacher) je ten, kdo učí (is the one who teaches), zoubek (small tooth) je malý zub (is a tooth of small size)*. The meaning that follows from the WD relations is usually labelled as *word derivation* or *internal meaning*. This is a rather narrow point of view, typical of linguistic WD theories, since there is no reason to introduce separate semantic relations – they are the same as other semantic relations, for example the Internal Language Relations defined within EuroWordNet 1, 2 (Vossen, 1998).

For the computer processing of the word meanings, it is necessary to bear in mind that semantic relations have a common nature irrespective of the forms by which they are expressed. From this point of view, it is obvious that there is no relevant difference between WD meaning and 'normal' lexical meaning as it is treated within the framework of semantic networks that are nowadays so popular in NLP and Language Engineering. Thus WDN can be understood as semantic networks of a special sort that will soon become relevant in the area of knowledge representation.

5 Word Derivation Analysis

The purpose of WD analysis is to find out the WD structure of the *"funded"* word by applying the basic WD relation, i. e. to find out its WD base and the respective *formant* (see below).

WD base can be defined as the part of the *"funded"* word that is taken over from the *"funding"* word. Typically it is an identical string that occurs in both *"funded"* and *"funding"* word – various phonological alternations can take place here, of course. For example, *pole (field)→pol-ní (field-ADJ)*.

The procedure works in the following way: an element is added to the WD base of the word (typically root or stem) and together they form a new word. The added element is usually called *formant* in Czech WD theory (Dokulil, 1962) and can be formed by one or more morphemes. A *formant* can be

- a suffix *kotel-na (boiler room)*, which classifies a word as belonging to a more general semantic group (here the Czech suffix *-na* with the inflectional morpheme *-a* at its end creates the names of places),
- the ending *zkáz-a (destruction)* where the inflectional morpheme also operates as derivational morpheme or the derivational morpheme can be considered to be empty,
- prefix *s-lepit (glue back together)* which just modifies the meaning of the word but does not change its POS and inflectional paradigm.

Compound formants are also possible, and may consist either of a prefix combined with suffix, e. g. in *pří-ruč-ní (reference)* or prefix combined with an ending *před-měst-í (suburb)*.

6 Relations between Inflection and Derivation

The data that can be found in the existing resources, e. g. (Karlík et al., 1995) are limited in number – they contain selected examples only and show only the main types of WD processes. Information about the functional load of the individual suffixes is either missing or is only outlined by means of expressions like *"very frequent"*, *"frequent"*, *"less frequent"* without giving any numbers.

To explore the situation more adequately we used a more representative collection of data, particularly the morphological database `I_par` together with the stem dictionary which is a part of the morphological analyser `ajka`.

In other words, all the items in the stem dictionary are associated with their respective inflectional paradigm(s), e. g. for nouns there are 746 inflectional paradigms that presently cover 131,188 noun stems. The number of the noun paradigms looks quite large but one should bear in mind that in our algorithmic description of Czech inflection we deal with a detailed hierarchical subclassification of the noun paradigms which, however, is just based on the 14 main paradigms as they are given in standard grammars of Czech.

The present data allow us to find the functional load defined as the number of nouns with the given suffix and particular semantic feature, e. g. agentive, instrument or property etc. First we have to know for a certain suffix how many lemmata occur with the given suffix, Table 1 shows that *-ák* has 1,379 occurrences. However, it is more important to know the specific semantic features that indicate which semantic classes the nouns ending with *-ák* belong to – this is obtained in the process of finding which nouns belong in which inflectional paradigms. If we look at the tags we can see that Czech nouns with the suffix *-ák* fall into two large groups – those denoting agentives and other animate beings (total 733 nouns) and those denoting inanimate things such as instruments and others (total 633, not classified 13). In the case of *-ák* its functional load is distributed evenly.

freq.	%	pattern	gender	sem. feature
641	47.1	vlk	MANIM.	agentives
326	23.6	flok	MINAN.	
263	19.1	krk	MINAN.	
77	5.6	Štěrbák	MANIM.	family names
16	1.2	dupák	MINAN.	
10	0.7	Azték	MANIM.	names of tribes
10	0.7	hřibák	MINAN.	
7	0.5	pulčík	MINAN.	names of fungi
6	0.4	kozák	MINAN.	
5	0.4	dubák	MINAN.	
5	0.4	Batak	MANIM.	ethnic groups
13	1.0	various		not classified
Σ 1,379				

Table 1: SUBST MASK, suffix: *-ák*

freq.	%	pattern	gender	sem. feature
908	93.9	učitel	MANIM.	agentives
15	1.6	bez	MINAN.	not classified
11	1.1	mocnitel	MINAN.	math. expr.
8	0.8	součinitel	MINAN.	math. expr.
7	0.7	hotel	MINAN.	hotels
5	0.5	cíl	MINAN.	not classified
4	0.4	stroj	MINAN.	not classified
4	0.4	soutěž	FEMIN.	not classified
3	0.3	obyvatel	MANIM.	inhabitants
2	0.2	strašpytel	MANIM.	not classified
Σ 967				

Table 2: SUBST, suffix: *-tel*

7 Rules for WD Processes

So far we have been talking about the derivation of word forms in terms of morphemes, stems, roots, lemmata, etc. More formally, WD processes employ strings of letters (morphemes) carrying grammatical and lexical information. It can be observed that the WD processes have quite a regular nature, thus one can express them by means of certain rules (WD patterns)

WD is a hierarchically structured process, which will be reflected in the construction of the rules (WD patterns) – we will build them as cascades going from simpler to more complicated patterns.

We start from a linguist's hypothesis which defines a pattern capturing the changes between the word forms and other constraints on the form of grammatical tags associated with the searched entries. Then we can automatically look up n-tuples of the entries in the existing morphological database (using the `I_par` tool), where

1. all the members (=entries) exist in the database;
2. the members of n-tuple fulfil the required hypothesis, i. e. the changes between the strings belonging to the individual members of the n-tuple are described in given pattern and entries correspond to the additional constraints.

The following n-tuple has to be searched for:

	form	POS-tag condition
1.	*S*-`zace`	SUBST FEM
2.	*S*-`stický`	ADJ
3.	*S*-`sta`	SUBST MAS ANIMAL
4.	*S*-`smus`	SUBST MAS UNANIMAL

where *"S-"* is arbitrary but has to be an identical string for all members of the n-tuple

Table 3: searching the relations, verifying hypotheses

Suppose that there is an algorithm (see Section 7.1) which in the respective morphological database will find the n-tuples (see Table 4) matching the hypothesis formulated by a linguist (see Table 3). The linguist can take the derived list of n-tuples and determine which n-tuples are correct and which not. The list of the positive examples will arise together with the list of exceptions not matching the formulated hypothesis, e. g. *přít* → *přítel*, word form *přítel (friend)* is derived from *"přát (to wish sb well)"*, not *"přít se (to argue with sb)"*.

It is now clear that by using a hierarchical connection of the new pattern to the original one (for the individual members of the n-tuple) it is possible to derive from a single entry not only the neighbouring entries in the n-tuple but also other word forms (generated from the entries included in the n-tuple by means of the respective original patterns).

If the changes in the word form put into effect by the (new) pattern express[1] a well-defined semantic relation, the entries can be *"virtualized"*,

[1] for the newly generated word form it is also possible to derive algorithmically (infer) lexical meaning

	1st member	2nd member	3rd member	4th member
Patterns:	**růže**	**otrocký_P**	**husita_P**	**komunismus**
	realizace	realistický	realista	realismus
	centralizace	centralistický	centralista	centralismus
	humanizace	humanistický	humanista	humanismus
	idealizace	idealistický	idealista	idealismus
	komunizace	komunistický	komunista	komunismus
	…	…	…	…
Patterns:	**růže**	**starořecký_P**	**husita_P**	**komunismus**
	romanizace	romanistický	romanista	romantismus
	spiritualizace	spiritualistický	spiritualista	spiritualismus
	synchronizace	synchronistický	synchronista	synchronismus
	kolektivizace	kolektivistický	kolektivista	kolektivismus
	modernizace	modernistický	modernista	modernismus
	…	…	…	…

Table 4: searching of the relations, verifying hypotheses, **output**

thus we do not need to keep all entries in the lexicon but only those entries which are basic (motivating) word forms for the neighbouring entries of the found n-tuples.

The entries that are thereby eliminated from the lexicon can be constructed according to the new pattern from the basic (motivating) word form. The original word forms can be determined algorithmically and their original lexical meaning can be inferred as well.

We will reduce the lexicon using the description of the WD process which yields the predictable changes in the semantics of the derived entries.

The WD process can be illustrated by the Fig. 1. It can be seen that the sub-entries *humanizace (humanisation)*, *humanisticky (humanistically* `ADV`*)*, *humanistický (humanistic* `ADJ`*)*, *humanistčin (humanist's* `FEM POSS ADJ`*)*, *humanistka (humanist* `FEM`*)*, *humanistův (humanist's* `MAS POSS ADJ`*)*, *humanista (humanist* `MAS`*)*, *humanismus (humanism)*, can be assigned:

1. either to the respective infl. paradigms:
   humanizace:`růže`
   humanisticky:`otrocky`
   humanistický:`otrocký`
   humanistčin:`matčin`
   humanistka:`matka`
   humanistův:`otcův`
   humanista:`husita`
   humanismus:`komunismus`
2. or to:
   humanizace:`růže`
   humanistický:`otrocký_P`
   humanista:`husita_P`
   humanismus:`komunismus`
3. or to a deriv. pattern (meta-pattern):
   humanismus:`komunismus_P`

In the second and third cases, the reduction of the lexicon can be observed. The pattern `komunismus_P` derives the word forms by exchanging the string at the end of the basic (motivating) word form:

$$smus \rightarrow zace, sticky, sta, smus$$

and the corresponding change of the attributes of the constructed tag.

For an implementation of these WD patterns, a parallel with Finite State Automata (FST) is useful. The property of chaining (Roche and Schabes, 1997) is very suitable here – it allows us to build WD patterns as hierarchical modules. This property makes it possible to limit the duplicity of the stored information and increase their lucidity.

7.1 WD Relation Mining

We explained how to extend the morphological database employing the regular changes of word forms that can be observed in the course of the WD processes (Osolsobě et al., 2002). We have shown that if the WD processes are described by the rules it is possible to reduce our stem dictionary and eventually to obtain a dictionary of roots.

To make the process of searching for the discrete description of WD processes simpler we have implemented an algorithm that looks up the relations between the strings corresponding to the individual entries in the morphological database.

The input for the algorithm is a description of the variations of the individual word forms together with conditions placed on the attributes of the respective grammatical tags.

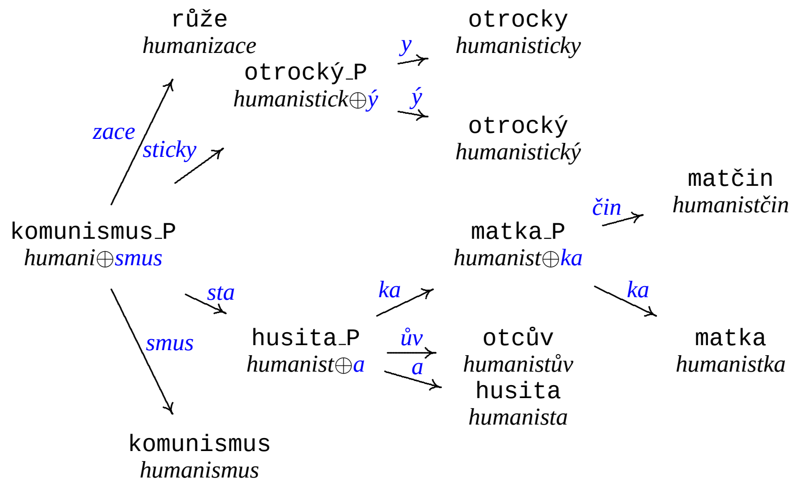

Figure 1: Using the derivational pattern to reduce the stem dictionary

To describe the variations of the word forms we will use:

- variables `$1`, `$2`, ... (values $\in \Sigma^*$),
- constants / *"affixes"* : $A_{1,1}, A_{1,2}, \ldots \in \Sigma^*$
- concatenation operator $\oplus$
- strings $S_i \in \{A_{i,1}, A_{i,2}, \ldots \texttt{\$1}, \texttt{\$2}, \ldots\}^*$
- conditions – constraints on the values of given attributes, eventually a determination whether the given word form has to be present in the database: $C_1, C_2, \ldots$

7.1.1 Input

The task assumes:

- n ... number of the word forms searched for,
- n-tuple: $(S_1, C_1) \ldots (S_n, C_n)$.

The S_i strings should be written in such a way so as not to contain the pairs constant – constant, variable – variable standing adjacent.

- two neighbouring constants can be merged into one
- two variables can be separated by ϵ constant (empty string).
- if the variable at beginning is required, or at the end of the string, then we set $A_{i,1}$, or $A_{i,m}$ $\equiv \epsilon$

Each string S_i thus can be given without loosing any generality in the following way:

$$S_i \equiv \begin{array}{l} A_{i,1} \oplus \texttt{\$1} \oplus A_{i,2} \oplus \texttt{\$2} \oplus \\ A_{i,3} \oplus \ldots \oplus \texttt{\$m} \oplus A_{i,m+1} \end{array} \qquad (1)$$

We know that $A_{i,j}$ are constants and `$j` variables which can take values from Σ^*. For an arbitrary string S_i, a regular grammar can be written (see Eq. 2).

$$\begin{array}{ll} S \rightarrow A_{i,1}\texttt{\$1}N_1 & \texttt{\$1}|\ldots|\texttt{\$m} \rightarrow E \\ N_1 \rightarrow A_{i,2}\texttt{\$2}N_2 & E \rightarrow a|aE|b|bE|\ldots \\ \ldots & \\ N_m \rightarrow A_{i,m+1} & \text{where } E \in \Sigma^* \end{array} \qquad (2)$$

It can be seen that for each string S_i a non-deterministic transducer can be constructed that takes a word form on the input and on the output, it produces a set of all acceptable evaluations of the variables `$1` ... `$m`, i. e. a set (possibly empty) of the m-tuples of members of set Σ^*.

7.1.2 The Algorithm

First we have to select the pairs (S_i, C_i) for which the requirement in the condition states that their corresponding word forms have to occur in the database. Those word forms, strings and pairs will be called *located*. The word forms that we can determine from the located ones after substitution values for the variables in the strings will be labelled as *inferred*.

We can speak here about free and bound occurrences of the variables. Free variables will be determined during the computation of the same automaton in which they take place. Bound variables are dependent on the computation of other automata. The values are instantiated for bound variables before the computing the automaton in which the variables occur. Thus we can work with them in a given automaton as with the constants – this simplifies the automaton.

When a given word form is accepted by the transducer (for string S_i) we obtain the respective evaluation variables included in S_i as an output. If the same variables occur also in other strings they can be substituted (instantiated) by the values.

Thus step by step we will construct the respective FS automata for located strings S_i using the instantiation of the variables. If the automaton does not contain any free variables it is obvious that the respective pair is inferred (it can be located at the same time) – these will be labelled as *inferred+located*).

The order in which the individual automata will be applied can be optimised. A certain part of the state space being searched can be eliminated in advance based on conditions C_i, i. e. it is enough to search/eliminate entries associated with the patterns which guarantee/eliminate some attributes of the tag.

We suppose that by means of *located* strings all the variables used in the *inferred* strings can be instantiated in such a way that we will be able to determine correctly inferred word forms relying only on the knowledge of the located word forms, i. e. in the cases where the inferred strings do not contain free variables. In the opposite case the algorithm has to stop prematurely.

The optimisation will determine the order in which the individual automata containing free variables will be applied.

We will start with the first automaton following the order determined by the optimisation. Step by step we will go through all the entries and then for all possible evaluations we will instantiate the variables and continue with searching the entries acceptable for the next automaton (according to the given ordering), i. e. we look for the next element of the respective n-tuple.

If we succeed in the instantiation of all variables and determine all inferred word forms and if all *inferred+located* word forms are found in the database then the currently determined n-tuple can be sent to the output.

8 The First Results from our Data

Table 5 displays the individual steps taken during forming the respective WD nest. The step *A* (see Table 5) consisted in the derivation of masculine possessives using suffix *-ův*. It is obvious that this derivation is regular, the number of lemmata has not changed – all of them have been assigned to the paradigm for the possessives *otcův (father's)*. In the step *B* the gender of noun is changed from masculine to feminine using suffix *-ka*. Moreover, in this step the paradigms *neumětel* a *Kocáb_nM* have been removed. Also the number of lemmata assigned to the paradigm *učitel* (teacher) has been reduced to half, i. e. from 908 to 454. This means that according to our data (our morphological database) half of the agentives cannot form the feminine counterpart and this result can be expected to be confirmed by examining a larger corpus. The step *C* is again regular – it consists in the derivation of feminine possessives using suffix *-in* with a number of lemmata not being changed. In the step *D* the adjectives are formed by means of suffix *-ský* and this process is less regular. From the possible 454 lemmata belonging to the paradigm *učitel* the adjectives are derived only from 113+21+16=150. Moreover, these adjectives split into three adjective paradigms *pražský (Prague)*, *společenský (social)* and *kremžský (Crems)* depending on whether they form a comparative and adverb or not. The following step is again regular – it involves the derivation of adverbs from adjectives by shortening the last vowel from *ý* to *y*. It can be seen that from the adjectives belonging to the paradigm *kremžský* such adverbs cannot be formed at all. The step *E* is irregular as well and it involves the derivation of the nouns from the respective adjectives by replacing the suffix *-ský* for *-ství*.

9 Conclusions

The purpose of the paper is to show how selected word derivation relations in Czech can be described using the morphological analyser `ajka` and the program `I_par` which works with the Czech morphological database. The Czech data necessary for this description are: stem dictionary used by `ajka` containing 385,066 Czech stems

908	učitel,otcův
3	obyvatel,otcův
2	přítel,otcův
1	neumětel,otcův
1	Kocáb_nM,otcův

A →

454	učitel,otcův,matka
2	obyvatel,otcův,matka
2	přítel,otcův,matka

B →

454	učitel,otcův,matka,matčin
2	obyvatel,otcův,matka,matčin
2	přítel,otcův,matka,matčin

C ↓

113	učitel,otcův,matka,matčin,pražský
21	učitel,otcův,matka,matčin,společenský
16	učitel,otcův,matka,matčin,kremžský
2	obyvatel,otcův,matka,matčin,pražský

D ←

113	učitel,otcův,matka,matčin,pražský,pražsky
21	učitel,otcův,matka,matčin,společenský,společensky
2	obyvatel,otcův,matka,matčin,pražský,pražsky

E ↓

46	učitel,otcův,matka,matčin,pražský,pražsky,stavení
19	učitel,otcův,matka,matčin,společenský,společensky,stavení

Table 5: WD nest for *-tel*

belonging to the 10 parts of speech, the Czech morphological database comprising 2,042 inflectional paradigms and the set of the Czech suffixes mentioned in this paper (*-tel*, *-ák*, *-ův*, *-ka*, *-in*, *-ský*, *-cký*, *-sky*, *-cky*, *-ství*, *-ismus*, *-ista*, *-izace*).

Within this task we also demonstrated, using the selected examples, how the inflectional paradigms can be employed to create more general but less regular word derivation patterns, how both can be linked together and how WD patterns can be used to describe the selected WD processes in an algorithmic way. This we regard as a relevant contribution to the theory. In our opinion these examples are general enough to show that the whole WD system for Czech can be grasped successfully in this way.

In this research we are working with Czech data only, but we firmly believe that if similar data for Russian, Slovak, Croatian or Serbian etc., were supplied similar results would be arrived at. It is, of course, obvious that different WD rules have to be formulated for other Slavonic languages but if a similar system of inflectional paradigms were used and the corresponding groups of suffixes as well, using tools like `ajka` and `I_par` would enable the mapping of word derivation relations accordingly.

The continuation of this research will lead to building a Czech Derivation Dictionary integrated with WD rules and thus later yielding a WD automaton for Czech.

Acknowledgement

This work was supported by Ministry of Education, Research Program CEZ:J07/98:143300003.

References

Miloš Dokulil. 1962. *Tvoření slov v češtině 1 (Word Derivation in Czech)*. Nakladatelství ČSAV, Praha. In Czech.

Petr Karlík, Marek Nekula, and Zdenka Rusínová. 1995. *Příruční mluvnice češtiny (Reference Grammar of Czech)*. Nakladatelství Lidové noviny, Praha. In Czech.

Lauri Karttunen and Kent Wittenburg. 1983. A Two-Level Morphological Analysis of English. In *Texas Linguistic Forum*, volume 22, pages 217–228.

Klára Osolsobě, Karel Pala, Radek Sedláček, and Marek Veber. 2002. A Procedure for Word Derivational Processes Concerning Lexicon Extension in Highly Inflected Languages. In *Proceedings of the Conference LREC2002*, volume 3, pages 998–1003, Las Palmas, May 29-31. ELRA.

Jan Petr. 1986. *Mluvnice češtiny II. (Grammar of Czech)*. Academia, Praha. In Czech.

Emmanuel Roche and Yves Schabes, editors. 1997. *Finite-State Language Processing*. MIT Press.

Radek Sedláček. 1999. Morfologický analyzátor češtiny (Morphological Analyser of Czech). Master's thesis, FI MU, Brno. In Czech.

Radek Sedláček and Pavel Smrž. 2001. A New Czech Morphological Analyser `ajka`. In *Proceedings of TSD 2001*, pages 100–107, Berlin. Springer-Verlag.

Marek Veber. 2002. *Nástroje pro textové korpusy a morfologické databáze (Tools for Text Corpora and Morphological Databases)*. Ph.D. thesis, FI MU, Brno. In Czech.

Piek Vossen. 1998. Set of Common Base Concepts in EuroWordNet-2. Technical Report 2D001, Department of Computational Linguistics, Amsterodam, October.