PBML

The Prague Bulletin of Mathematical Linguistics
NUMBER 110 APRIL 2018 7–42

Modelling Morphographemic Alternations in Derivation of Czech

Magda Ševčíková

Charles University, Faculty of Mathematics and Physics, Institute of Formal and Applied Linguistics
Prague, Czechia

Abstract

The present paper deals with morphographemic alternations in Czech derivation with regard to the build-up of a large-coverage lexical resource specialized in derivational morphology of contemporary Czech (DeriNet database). After a summary of available descriptions in the Czech linguistic literature and Natural Language Processing, an extensive list of alternations is provided in the first part of the paper with a focus on their manifestation in writing. Due to the significant frequency and limited predictability of alternations in Czech derivation, several bottom-up methods were used in order to adequately model the alternations in DeriNet. Suffix-substitution rules proved to be efficient for alternations in the final position of the stem, whereas a specialized approach of extracting alternations from inflectional paradigms was used for modelling alternations within the roots. Alternations connected with derivation of verbs were handled as a separate task. DeriNet data are expected to be helpful in developing a tool for morphemic segmentation and, once the segmentation is available, to become a reliable resource for data-based description of word formation including alternations in Czech.

1. Introduction

Concerning the internal structure of complex words in the Czech lexicon, derivation is the dominant process of word formation, highly prevailing over compounding in Czech (Dokulil, 1962; Dokulil et al., 1986). All types of derivation (esp. prefixation and suffixation) in Czech may be accompanied by vowel or consonant alternations in the root and/or in affixes.[1] Morphographemic alternations are the major source

[1] In the paper, the term "root" refers to a morpheme that cannot be further analysed while "stem" is used, less specifically, for the part of a word without inflectional affixes (Haspelmath and Sims, 2010; Aronoff,

 Corresponding author: sevcikova@ufal.mff.cuni.cz
Cite as: Magda Ševčíková. Modelling Morphographemic Alternations in Derivation of Czech. The Prague Bulletin of Mathematical Linguistics No. 110, 2018, pp. 7–42. doi: 10.2478/pralin-2018-0001.

of allomorphy in Czech. They diversify the formal shape of a base word and the particular derived word; cf. palatalization of the final consonant of the root morpheme by adding a diminutive suffix in ex. (1) and an analogous alternation of the final consonant in the first diminutive suffix during the subsequent formation of a double diminutive in (2). Several alternations in a single derivational step are documented in ex. (3), namely a vowel alternation in the prefix and a consonant alternation in the final position of the root, or in (4) with a vowel alternation, a vowel insertion, and a consonant alternation in the root.

(1) *hroch*$_{\mathrm{N}}$ 'hippo' $\xrightarrow{\text{ch>š}}$ *hroš-ík*$_{\mathrm{N}}$ (dimin.)[2]

(2) *hroš-ík*$_{\mathrm{N}}$ 'hippo' (dimin.) $\xrightarrow{\text{k>č}}$ *hroš-íč-ek*$_{\mathrm{N}}$ (double dimin.)

(3) *vy-skoč-i-t*$_{\mathrm{V}}$ 'to leap' $\xrightarrow{\text{y>ý, č>k}}$ *vý-skok*$_{\mathrm{N}}$ 'leap'

(4) *vejc-e*$_{\mathrm{N}}$ 'egg' $\xrightarrow{\text{e>a, 0>e, c>č}}$ *vaječ-ný*$_{\mathrm{A}}$ 'made from eggs'

The paper is organized as follows. Starting with a note on terminology, Section 2 provides an overview of linguistic descriptions of morphographemic alternations and available approaches in Natural Language Processing (NLP) of Czech, including the derivational database DeriNet which is in focus of the paper. A detailed classification of alternations in the contemporary Czech lexicon follows in Section 3. Attached to the section, we provide a complete list of alternations supported with examples.

Due to the size of the DeriNet database (exceeding 1 million words), derivational relations, including all types of alternations, have been identified semi-automatically (Sect. 4). Suffix-substitution rules proved to be efficient for alternations in the final position of the stem, whereas a specialized approach of extracting alternations from inflectional paradigms was used for modelling alternations within the roots. Alternations connected with derivation of verbs were handled as a separate task.

Section 5 concludes with an analysis of main types of alternations not yet covered in DeriNet and provides a perspective of using the DeriNet data in the development of a tool for morphemic segmentation as well as in the linguistic research into word formation in general and into morphographemic alternations in particular.

1994). Roots and stems are not together referred to as "bases" (cf. Bauer, 1983, pp. 20f) since we reserve the term "base" for the opposition of a base word vs. a derived word (target word, or derivative). These pairs are referred to as "pairs of base-target words" or "base-target pairs", too.

[2]In the examples, the base word is written first followed by the derivative, the derivational relation is represented by an arrow. The alternations that accompany the derivation are listed above the arrow. The grapheme in the base is written first followed by "$>$" and the corresponding grapheme in the derivative. Boundaries between morphemes are indicated with the hyphens (the morphemic structure is not marked in Sect. 4 since the data are not segmented in the DeriNet network).
In examples on diminutive derivation, we use "dimin." (diminutive) and "double dimin." (double diminutive) instead of the full English translation (e.g. 'small hippo' and 'very small hippo', respectively).

2. Related work

2.1. Note on terminology

Unlike the (mainstream) phonemic and phonological approach of alternations (e.g. Haspelmath and Sims, 2010), the present paper deals with this issue in relation to written Standard Czech.[3] The term "morphographemic alternations" is thus preferred to that of "morphophonemic alternations" or similar terms used in the linguistic literature (cf. morphonological alternations / morphophonological alternations in Matthews, 2007, p. 253; Štekauer et al., 2012; Osolsobě, 2014, pp. 198ff; Ziková, 2015, 2016a,b; Šefčík, 2016b, or phoneme alternations / phonemic alternations / phonological alternations / alternations of phonemes in Dokulil, 1962; Daneš et al., 1967; Dokulil et al., 1986; Osolsobě, 2002; Aronoff, 1976). We neither use the term "ablaut" nor "apophony" (e.g. Lieber and Štekauer, 2014, pp. 125f, Baerman, 2015), since the former term is delimited inconsistently in the description of Czech and the latter term is not anchored in the Czech terminology; see Šefčík (2016a) for details.

Our approach is rooted in Dokulil's onomasiological theory (Dokulil, 1962) and uses common terminology on general aspects of word formation. Lexemes that share the root are called a derivational family; if members of a derivational family are organized according to the direct derivational relations, we speak about derivational trees with regard to the derivational data, rather than using Dokulil's term "word-formation nest".[4] A comment is required on the term "word-formation type" which is defined as a set of words that share a certain word-formation meaning and were derived from bases of the same part-of-speech category by using the same affix (Dokulil, 1962, pp. 68ff); cf. the word-formation type of agentive nouns derived from verbs with the suffix *-tel* in Czech (*učitel* 'teacher', *pozorovatel* 'observer').[5]

2.2. Descriptions of alternations in linguistic literature on Czech derivation

Morphographemic alternations in Czech originate in systemic as well as accidental diachronic changes that emerged during differentiation of Czech from other Slavic languages and are thus subject to historical grammars of the Czech language (Gebauer, 1984–1929; Lamprecht et al., 1986). Consonant alternations were described as changes of non-palatal consonants into palatalized ones (i.e. palatalization; or vice versa as

[3] Alternations that are not mirrored in writing are omitted in the paper; esp. palatalization of consonants is often recorded by the letters *i* or *ě* following the consonant instead of changing the grapheme itself (cf. *t–t(i)* in *bota* 'shoe' → *botička* (dimin.) instead of *t>ť* according to the pronounciation).

[4] The term "word-formation nest" was substituted for the term "derivational paradigm" by Dokulil et al. (1986, p. 207); the latter term has recently established as the core concept of the paradigmatical approach to word formation (Lieber and Štekauer, 2014, pp. 354ff; Booij, 2008; Pounder, 2000; Bauer, 1997 etc.).

[5] Dokulil's definition of the word-formation type ("slovotvorný typ") is thus different from that by Hansen (1985, pp. 28ff) ("Wortbildungstyp").

depalatalization), mostly due to the contact with a front or iotified vowel in order to allow for a more comfortable pronunciation. Three rounds of palatalization of velar consonants in Proto-Slavic were reconstructed, for each round several irregularities and exceptions were stated (Lamprecht et al., 1986; Večerka, 2016). The source of a part of vowel-zero alternations in contemporary Czech are both systematic and accidental changes of the yer-vowels in Proto-Slavic (Lamprecht et al., 1986; Ziková, 2016b). However, due to different counter-tendencies, such as the trend to preserve the vowel quantity of the base word; cf. the *e*>*é* alternation in ex. (5) vs. its lack in (6), and *o*>*ů* in (7) vs. (8) (Dokulil, 1962, p. 170), the resulting synchronic picture of alternations in the Czech derivation seems to be highly irregular (similarly to other languages; cf. Bybee and Brewer, 1980).

(5) *oheň*$_{\text{N}}$ 'fire' $\xrightarrow{\text{e>é, ň>n}}$ *ohén-ek*$_{\text{N}}$ (dimin.)

(6) *účes*$_{\text{N}}$ 'hairstyle' → *účes-ek*$_{\text{N}}$ (dimin.)

(7) *krok*$_{\text{N}}$ 'step' $\xrightarrow{\text{o>ů, k>č}}$ *krůč-ek*$_{\text{N}}$ (dimin.)

(8) *blok*$_{\text{N}}$ 'block' $\xrightarrow{\text{k>č}}$ *bloč-ek*$_{\text{N}}$ (dimin.)

Diachronic changes and the synchronic distribution of vowel-zero alternations in Czech were treated within the framework of generative phonology (Scheer et al., 2011; Scheer and Ziková, 2010). The diachronic perspective is also taken by Pognan and Panevová (2013) who examine common Slavic roots as a basis for research into Slavic intercomprehension. Less recent studies (Stankiewicz, 1986, 1960; Rubenstein, 1950) placed alternations in Czech in the cross-linguistic context of other Slavic languages.

A synchronic description of alternations in derivation was included in Dokulil's fundamental study on Czech derivation (Dokulil, 1962, esp. pp. 159–178), which has become a widely respected and, in fact, the only common ground of word-formation descriptions in Czech grammars and specialized studies since then (Daneš et al., 1967; Dokulil et al., 1986; Čermák, 2012; Štícha, 2013 etc.).[6] However, description of alternations is usually spread over the chapters on word formation and inflectional morphology with only sporadic mutual links. The most complex and elaborate description so far is by Ziková (2015), which is still a pilot study for an intended grammar of Czech and is limited to quantitative alternations and vowel-zero alternations.

Two existing morphemic dictionaries might be relevant for the topic of alternations in Czech. In Šiška's dictionary (Šiška, 2005), root morphemes of a part of the Czech lexicon are grouped together according to their lexical meaning; each n-tuple of the root allomorphs is supplemented with a selective list of lexemes. In the dictionary by

[6]Dokulil's approach, based on differentiation of four onomasiological categories, has influenced approaches to derivation in Czech as well as in other, particularly (but not exclusively) Slavic languages; cf. works on Slovak (e.g. Buzássyová, 1974; Horecký et al., 1989; Furdík, 2004), Polish (Grzegorczykowa et al., 1998), Russian (Švedova, 1980), or Štekauer's application to English (Štekauer, 1998).

Slavíčková (1975), lexemes are analysed into morphemes and listed retrogradely without mutual connections. None of the dictionaries is machine tractable, their usability for our task was very limited.

A formalized description of selected types of alternations in Czech inflection was a part of the inflectional dictionary by Osolsobě (1996); it focused on alternations of consonants in the final position of the stem. The dictionary was used in automatic morphological analysis by the Ajka (later on, Majka) analyser and in other tasks in NLP of Czech (Osolsobě, 2015); see Sect. 2.3.

2.3. Alternations in Natural Language Processing and language resources for Czech

In NLP of Czech, alternations were handled in both large-coverage inflectional analysers used for Czech, namely in the Ajka analyser (Sedláček and Smrž, 2001; Sedláček, 2004; Šmerk, 2007) and in the analyser developed by Hajič (2004).

The dictionary of the analyser Ajka can be searched for derivationally related pairs (or n-tuples) by the tool Deriv (Osolsobě et al., 2009) using regular expressions. When searching for pairs of words with alternations, each alternation must be specified with a separate regular expression. A similar tool, Morfio (Cvrček and Vondřička, 2013), searches for pairs with a common base and different affixes in the Czech National Corpus; the words need not to be in a derivational relation. The tool makes it possible to include several tens of pairs of alternations into the queries (44 pairs without respect to which of the graphemes is in the base and in the derivative). However, both tools suffer from massive overgeneration.

In a close relation to Ajka, a derivational analyser for Czech called Derivancze was developed (Pala and Šmerk, 2015). The data of Derivancze are not available for a free download, but can be queried by a web tool. For a word filled in into the web form, the tool gives a base word and a direct derivative if found in the underlying dictionary data. It was not explicitly addressed by Pala and Šmerk (2015) whether and to which extent alternations were handled in Derivancze. Nevertheless, a random search for several examples containing alternations showed a rather unsystematic approach to this phenomenon. For instance, the diminutive *domek* is correctly linked with the base noun *dům* 'house' in Derivancze whereas the diminutive *stolek* is connected incorrectly with a non-existing string *stol* (instead of *stůl* 'table'), *hrošík* 'hippo' (dimin.) was not found by the tool, no parent was found for *chirurzka* 'woman surgeon'.

The morphological analyser by Hajič is connected with the inflectional dictionary MorfFlex CZ (Hajič and Hlaváčová, 2013). From MorfFlex CZ, the set of lexemes for the DeriNet database was extracted and, moreover, the dictionary has turned out to be an important source of information on morphographemic alternations in derivation; see Sect. 4.[7]

[7] MorfFlex CZ (and thus DeriNet) covers the entire lexicon of contemporary Czech including proper nouns, archaic words, low-frequency words and regular, automatically generated coinages without respect to whether they are attested in a corpus.

Derivational relations are included in other language resources, too, though rather marginally. In Czech WordNet a set of 14 relations was implemented (Pala and Smrž, 2004; Pala and Hlaváčková, 2007). In the Prague Dependency Treebank (Hajič et al., 2006), selected types of derivatives were represented by the lemma of their base word within the deep-syntactic annotation (Razímová and Žabokrtský, 2006).

2.4. DeriNet database as a resource specialized in Czech derivation

A decision that we had to make at the start of the DeriNet project was whether pairs of base and target words with alternating graphemes will be linked together in the database, or whether they stay unconnected. The insufficient attention paid to alternations in Czech linguistics and in NLP of Czech in combination with the complicated nature of alternations were strong arguments against the inclusion of this issue into the semi-automatically constructed resource. On the other side, massive presence of alternations was the main argument in favour of including them into the database.

DeriNet is a large-coverage lexical resource specialized in derivational morphology of Czech; neither composition nor combined word-formation processes have been included so far. It is the only one freely available derivational resource for Czech and, in a broader context of European linguistics, it is in line with recent research in word formation; e.g. word-formation database for Latin (Litta et al., 2016), Démonette for French (Hathout and Namer, 2014), the language-independent approach by (Baranes and Sagot, 2014), DerivBase.Hr for Croatian (Šnajder, 2014), DerivBase for German (Zeller et al., 2013), or CELEX for English, German and Dutch (Baayen et al., 1995).[8]

The design of DeriNet was based on Dokulil's understanding of word-formation nests as internally structured groups of all words based both formally and semantically[9] on the same base in contemporary language without regard to their real etymology (Dokulil, 1962, p. 14, Dokulil et al., 1986, p. 207). Words (represented as nodes in DeriNet) are connected with a link (edge) if they are derivationally related; the edge is oriented from the base to the derivative. At most one base word may be identified for a derived word. Words that are directly and indirectly derived from a particular base word thus form an oriented graph (called derivational tree in the paper).

[8] Approach to alternations was mostly not addressed in the respective publications. Alternations are explicitly referred to by Šnajder (2014), whereas they were not included e.g. by Baranes and Sagot (2014).

[9] The formal and semantic relations of a derived word to its base are discussed as foundation and motivation, respectively, in the onomasiological theory of word-formation (Dokulil, 1962; Dokulil, 1994, pp. 131ff; Štekauer, 1998). If foundation is not in accordance with motivation, priority is given to formal relations (foundation).

The current version of the database, DeriNet 1.4,[10], contains nearly 1,012 thousand lexemes which were extracted from the MorfFlex CZ dictionary. The lexemes are interconnected with more than 774 thousand derivational links.[11] All types of alternations described in Sect. 3 have been included into DeriNet; the methods used are described in Sect. 4.

3. Morphographemic alternations in derivation of Czech

3.1. Delimitation of alternations, basic classification

An alternation is understood as a substitution of a grapheme by another one that occurred during derivation in addition to the proper affixation; the term is used both for the process of replacing a grapheme with another one in a particular morphosyntactic context and for the pair of graphemes occurring in a particular position of the base word and the target word, i.e. for the result of this process. The alternations are identified in a morpheme that is shared by the base and the derivative; cf. ex. (1) to (8) above. On the contrary, examples in (9) and (10) are not considered to contain alternations, the difference *í* vs. *i* in (9) being interpreted as a result of replacing the masculine suffix by the feminine one (resuffixation; Šimandl, 2016), and *a* vs. *á* in (10) as resulting from the substitution of the inflectional ending for a suffix.

(9) *tanečn-ík*$_{\mathrm{N}}$ ‘dancer’ → *tanečn-ice*$_{\mathrm{N}}$ ‘female dancer’

(10) *brank-a*$_{\mathrm{N}}$ ‘goal’ → *brank-ář*$_{\mathrm{N}}$ ‘goalkeeper’

A grapheme alternates with another grapheme or with one of a closed set of graphemes; e.g. *c* changes into *k* in (11) or into *č* in (12). Due to this feature, Osolsobě (2002) describes alternations as a “regular” substitution. It is stressed, however, that the alternations are regular neither in the sense that a given grapheme is always affected by alternation in the given morphographemic context (see (5) vs. (6), and (7) vs. (8)), nor that they are related to a particular type of derivation (e.g. defined by the part-of-speech categories of the base and target word) or even to a particular word-formation type. For instance, the *c*>*č* alternation occurs in derivation of deverbal nouns (12) and in derivation of adjectives from nouns (13). In (14) the *a*>*á* alternation must be applied, otherwise the adjective *vratný* ‘returnable’ might be connected incorrectly with the noun *vrata* ‘gate’ (but it belongs to *vrátit* ‘to return’ with the reverse alternation *á*>*a* in (15)). In (16), the alternation is not present – if applied, the adjective *slávistický* ‘belonging to supporters of Slávie’ in (17) would be derived incorrectly.

[10] `http://ufal.mff.cuni.cz/derinet`
DeriNet 1.0 and 1.2 were published in the Lindat/Clarin repository (Vidra et al., 2015, 2016). The data are freely available for non-commercial purposes under the Creative Commons (CC-BY-NC-SA) licence.

[11] For 238 thousand (23.5 % out of all nodes) no base word has been identified so far. However, more than a half of the parentless nodes is capitalized nouns (more than 124 thousand). Capitalization concerns proper nouns only, which have a limited derivational potential.

(11) *péc-t*$_{\mathrm{V}}$ ‘to bake’ $\xrightarrow{\text{é>e, c>k}}$ *pek-ař*$_{\mathrm{N}}$ ‘baker’

(12) *péc-t*$_{\mathrm{V}}$ ‘to bake’ $\xrightarrow{\text{é>e, c>č}}$ *peč-en-í*$_{\mathrm{N}}$ ‘baking’

(13) *ovc-e*$_{\mathrm{N}}$ ‘sheep’ $\xrightarrow{\text{c>č}}$ *ovč-í*$_{\mathrm{N}}$ ‘belonging to/got from sheep’

(14) *vrat-a*$_{\mathrm{N}}$ ‘gate’ $\xrightarrow{\text{a>á}}$ *vrát-ný*$_{\mathrm{N}}$ ‘porter’

(15) *vrát-i-t*$_{\mathrm{V}}$ ‘to return’ $\xrightarrow{\text{á>a}}$ *vrat-ný*$_{\mathrm{A}}$ ‘returnable’

(16) *slav-ist-a*$_{\mathrm{N}}$ ‘Slavist’ → *slav-is-tický*$_{\mathrm{A}}$ ‘Slavic’

(17) *sláv-ist-a*$_{\mathrm{N}}$ ‘supporter of the sport club Slávie’ → *sláv-ist-ický*$_{\mathrm{A}}$ ‘belonging to the supporters of Slávie’

There are nearly 90 pairs of alternating graphemes in Czech. Since we model derivational relations as oriented from the base word to the derived one, the pairs of alternating graphemes are described as being oriented, too. The “base grapheme” (in the base) vs. the “target grapheme” (in the derivative) are discerned. Pairs of alternating graphemes differ in whether one of the them is always to be found as the base grapheme while the other one as the target grapheme across the lexicon, or if they are found in reverse order in other pairs of lexemes as well (so-called one-directional vs. bidirectional alternations, respectively, according to Osolsobě, 2002; Ziková, 2015, does not take orientation of the alternating graphemes into consideration). The *h*>*z* alternation in (18) is an example of the one-directional alternation in Czech. The graphemes *ch* and *š* enter the alternation *ch*>*š* on the one hand, and *š*>*ch* on the other ((19) vs. (20)).

(18) *drah-ý*$_{\mathrm{A}}$ ‘expensive’ $\xrightarrow{\text{h>z}}$ *draz-e*$_{\mathrm{D}}$ ‘at a high price’

(19) *tich-ý*$_{\mathrm{A}}$ ‘silent’ $\xrightarrow{\text{ch>š}}$ *tiš-e*$_{\mathrm{D}}$ ‘silently’

(20) *po-těš-i-t*$_{\mathrm{V}}$ ‘to please’ $\xrightarrow{\text{š>ch}}$ *po-těch-a*$_{\mathrm{N}}$ ‘pleasure’

The following classification differentiates five types of vowel alternations (A to E), three types of consonant alternations (F to H), and a type of mixed alternations (I; Dokulil, 1962, pp. 162ff, Osolsobě, 2002). The vowel (i.e. vowel-to-vowel) alternations are classified according to the quantity and quality of the base and target graphemes:

A) in quantitative alternations, a vowel is substituted for the same vowel with opposite quantity (short vowels are lengthened (21), long vowels shortened (22)):

   (21) *vy-jet*$_{\mathrm{V}}$ ‘to leave’ $\xrightarrow{\text{y>ý}}$ *vý-jezd*$_{\mathrm{N}}$ ‘leaving’

   (22) *tráv-a*$_{\mathrm{N}}$ ‘grass’ $\xrightarrow{\text{á>a}}$ *trav-natý*$_{\mathrm{A}}$ ‘grassy’

B) in qualitative alternations, a vowel is replaced by a different vowel with the same quantity:

(23) *hrab-a-t*$_{V}$ 'to dig' $\xrightarrow{a>o}$ *hrob*$_{N}$ 'grave'

C) in quantitative-qualitative alternations, a vowel in the base word is replaced by a qualitatively different vowel with opposite quantity in the target word:

(24) *říd-i-t*$_{V}$ 'to direct' $\xrightarrow{í>e}$ *řed-i-tel*$_{N}$ 'director'

(25) *ostrov*$_{N}$ 'island' $\xrightarrow{o>ů}$ *ostrův-ek*$_{N}$ (dimin.)

D) vowel deletion can be described as a type of vowel alternations, too; a vowel (mostly *e* in Czech derivation) is substituted by a zero (vowel-zero alternation):

(26) *pes*$_{N}$ 'dog' $\xrightarrow{e>0}$ *ps-í*$_{A}$ 'belonging to dog'

(27) *such-ý*$_{A}$ 'dry' $\xrightarrow{u>0}$ *sch-nou-t*$_{V}$ 'to become dry'

E) vowel insertion is described as a replacement of a zero by a vowel (zero-vowel alternation):

(28) *hr-á-t*$_{V}$ 'to play' $\xrightarrow{0>e}$ *her-n-a*$_{N}$ 'playroom'

The following types of consonant-to-consonant alternations are applied in Czech:

F) individual alternations when a single consonant is substituted by another one:

(29) *čern-ý*$_{A}$ 'black' $\xrightarrow{n>ň}$ *čerň*$_{N}$ 'black (colour)'

(30) *čern-och*$_{N}$ 'black man' $\xrightarrow{ch>š}$ *čern-oš-ka*$_{N}$ 'black woman'

G) consonant deletion and insertion is peripheral in contemporary Czech, cf. deletions in verb-to-verb derivation (31) and in derivation from proper nouns of foreign origin (32), and insertion of the initial *j* (which is not a prefix) in (33) :

(31) *top-i-t*$_{V}$ 'to drawn'' $\xrightarrow{p>0}$ *to-nou-t*$_{V}$ 'to be drawning'

(32) *Hamburk*$_{N}$ 'Hamburg' $\xrightarrow{k>0}$ *hambur-ský*$_{A}$ 'from Hamburg'

(33) *mí-t*$_{V}$ 'to have' $\xrightarrow{0>j}$ *jmě-ní*$_{N}$ 'property'

H) a substitution of a pair of consonants by a particular pair of consonants is called a group alternation:

(34) *měst-ský*$_{A}$ 'urban' $\xrightarrow{st>šť}$ *měšť-an*$_{N}$ 'burgher'

(35) *čes-ký*$_{A}$ 'Czech' $\xrightarrow{sk>št}$ *češ-tina*$_{N}$ 'Czech language'

I) In addition, in so-called mixed alternations, a vowel is replaced by a combination of a vowel and constant; this type is mostly found in deverbal derivation:

(36) *stá-t*$_{V}$ 'to stand' $\xrightarrow{á>oj}$ *stoj-ící*$_{A}$ 'standing'

(37) *stá-t*$_{\text{V}}$ 'to stand' $\xrightarrow{\text{á>av}}$ *po-stav-i-t*$_{\text{V}}$ 'to set up'

(38) *bí-t* $_{\text{V}}$ 'to beat' $\xrightarrow{\text{í>ij}}$ *bij-ící*$_{\text{A}}$ 'beating'

An alternative classification (into vowel-zero alternations, quantitative alternations, and palatalization alternations) was proposed by Ziková (2015).

3.2. Distribution of morphographemic alternations

In Czech derivation, alternations affect almost the entire repertory of graphemes and all types of morphemes (and, assumably, a considerable part of the Czech lexicon).[12] Considering the repertory of graphemes in Czech, all vowels and consonants, except for *p*, *b*, *f*, *v*, *m*, and *l*, enter alternations. Both vowel and consonant alternations can occur at any position in a word, even at the first one ((39) to (41)).

(39) *úz-ký*$_{\text{A}}$ 'narrow' $\xrightarrow{\text{ú>u}}$ *uz-oučký*$_{\text{A}}$ 'very narrow'

(40) *hn-á-t*$_{\text{V}}$ 'to drive' $\xrightarrow{\text{h>ž, 0>e}}$ *žen-oucí*$_{\text{A}}$ 'driving'

(41) *hr-á-t*$_{\text{V}}$ 'to play' $\xrightarrow{\text{r>ř}}$ *hř-iště*$_{\text{N}}$ 'playground'

Individual alternating pairs differ in frequency.[13] According to an overall estimate provided by Osolsobě (2002), *a* and *á* out of the vowels enter alternations most frequently (*a* changes into *á*, *e*, *ě*, and *o*, the long *á* into *a*, *e*, *i*, and *í*). The pairs *s*>*š*, *k*>*c*, and *c*>*č* are the most frequent consonant alternations. The vowel *o* and the consonant *g* alternate least frequently. Nevertheless, neither the quality nor the frequency of alternating graphemes allow for estimating the productivity of particular alternations (cf. Ziková, 2016a).

Alternations affect all types of morphemes, namely prefixes, roots, and suffixes during derivation (and roots and suffixes during inflection, see the next subsection). Vowel lengthening (plus the alternations *o*>*ů*) occurs in prefixes, roots as well as suffixes, whereas other vowel alternations (shortening, qualitative alternations, and

[12]The amount of words affected by alternations was preliminarily estimated in our study including 500 nouns, adjectives, verbs, and adverbs (consisting of at least two characters, only the first of which was allowed to be uppercased) with the highest token frequency in the representative corpus of Czech (SYN2015, 120 million tokens; Křen et al., 2015). 100 (20 %) out of the examined lemmas involved alternations with respect to their particular base words. For 271 (54.2 %) out of 500 lemmas, it was possible to find at least one derived word that was affected by alternations.
The aim of the study was not to estimate the alternation frequency in the overall data collection. As "phonetic change often progresses often more quickly in items with high token frequency" (Bybee, 2001, p. 11; cf. also Bybee, 2007, p. 270), less alternations are expected in words with lower frequency. We believe that a more precise picture of how alternations are distributed over the lexicon could be inferred from the DeriNet data.

[13]Here and elsewhere in the paper, type frequency in the Czech lexicon is meant if we do not refer to a particular corpus or another data resource.

quantitative-qualitative alternations) are limited to roots and suffixes. Mixed alternations are limited to derivation from verbs and affect final vowels of the root morpheme (these alternations originate in inflection; see Sect. 4.3). Alternations with zero (in both directions) are prototypically found in roots, or less frequently, in suffixes. Consonants alternate mostly in the final position of the stem, forced by the added suffix. Group alternations affect either two final consonants of the stem, or the final consonant of the stem and the first one of the suffix.

In Appendix, we provide an exhaustive list of alternations as observed in the lexicon of contemporary Czech, specifically as manifested in writing. Neither the origin of the alternation,[14] nor the frequency or productivity in the lexicon were taken into consideration. Nearly 90 alternation pairs are listed in alphabetical order according to the form of the alternating grapheme in the base word. Each pair is given in a separate line, the direction of the alternation is of significance. If a pair of graphemes alternates in both directions, it is listed twice in the list (indicated with the note "bidir." with each of the directions). Each pair of alternating graphemes is followed by a set of examples with the particular alternation in prefix, root and suffix (if available). In the rightmost column, we tried to find counter-examples, documenting that a particular grapheme even in a close morphosyntactic context does not necessarily undergo the same change.

3.3. Alternations in derivation vs. in inflection

Most of the morphographemic alternations are found in both derivation and inflection.[15] There are only few pairs limited either to the former, or to the latter area; e.g. the *é>í* and *é>ý* alternations are found in derivation only (42), the *g>z* alternation exclusively in inflection (43). Apart from the distribution (alternations in inflection do not occur in prefixes), the alternations exhibit the same features in inflection as in derivation, esp. massive presence and irregularity.

(42) *polévka*$_{\mathrm{N}}$ 'soup' $\xrightarrow{\text{é>í}}$ *polívka*$_{\mathrm{N}}$ 'soup'

(43) *filolo***g**$_{\mathrm{N}}$ 'philologist', *filolo***z***-ích*$_{\mathrm{loc.sg.masc.anim}}$

Dokulil (1962, p. 112) pointed out the complicated relations between alternations in a particular word and in a word derived from it. He examined the inflectional

[14] We thus omit the difference (pointed out by Dokulil, 1962, pp. 11f) between alternations that are required by a certain word-formation type (they depend on the graphemic structure of the affix and are obligatory, or accompany a certain word-formation type) and alternations that are not – from the synchronic point of view – related to the particular word-formation type "and are thus not considered word-formation alternations" [translated by the author of the paper] (in spite of being systematic in diachrony; e.g. *a>ě* in *svatý* 'holy' → *světec* 'holy man').

[15] The question to which linguistic subdiscipline alternations belong to has been discussed across different approaches (see Bybee and Brewer, 1980, or Bermúdez-Otero and McMahon, 2006, for summaries).

paradigms of both the base and the derivative whether they share an alternation. Here is a simplified list of types based on Dokulil's findings:

1. the derivative (its lemma and all inflected forms) exhibits an alternation with respect to the lemma and all inflected forms of the base word (the particular alternation is not present in the inflectional paradigm of the base word):

   (44) inflection of the base word: *čáp*$_{\mathrm{N}}$ 'stork', *čáp-a*$_{\mathrm{gen.sg}}$, *čáp-ovi*$_{\mathrm{dat.sg}}$ etc.
   derivation: *čáp*$_{\mathrm{N}}$ 'stork' $\xrightarrow{\text{á>a}}$ *čap-í*$_{\mathrm{A}}$ 'belonging to stork' (inflection of the derived word: *čap-ího*$_{\mathrm{gen.sg}}$, *čap-ímu*$_{\mathrm{dat.sg}}$ etc.)

   (45) infl. of the base word: *sprav-ova-t*$_{\mathrm{V}}$ 'administrate', *sprav-uj-i*$_{\mathrm{1.sg.pres.act}}$ etc.
   derivation: *sprav-ova-t*$_{\mathrm{V}}$ 'administrate' $\xrightarrow{\text{a>á}}$ *správ-a*$_{\mathrm{N}}$ 'administration' (inflection of the derived word: *správ-y*$_{\mathrm{gen.sg}}$, *správ-ě*$_{\mathrm{dat.sg}}$ etc.)

2. the derived word exhibits an alternation in its entire inflectional paradigm with respect to the lemma of the base word; however, the alternation is involved in some inflectional forms of the base word:

   (46) inflection of the base word: *d**ů**m*$_{\mathrm{N}}$ 'house', *d**o**m-u*$_{\mathrm{gen.sg}}$, *d**o**m-u*$_{\mathrm{dat.sg}}$, *dům*$_{\mathrm{acc.sg}}$ etc.
   derivation: *dům*$_{\mathrm{N}}$ 'house' $\xrightarrow{\text{ů>o}}$ *dom-ek*$_{\mathrm{N}}$ (dimin.) (inflection of the derived word: *dom-k-u*$_{\mathrm{gen.sg}}$, *dom-k-u*$_{\mathrm{dat.sg}}$, *dom-ek*$_{\mathrm{acc.sg}}$ etc.)

   (47) inflection of the base word: *b**ůh***$_{\mathrm{N}}$ 'god', *b**o**h-a*$_{\mathrm{gen.sg}}$, *b**o**h-u*$_{\mathrm{dat.sg}}$, *b**o**h-a*$_{\mathrm{acc.sg}}$, *b**o**že*$_{\mathrm{voc.sg}}$, *b**o**z-i*$_{\mathrm{nom.pl}}$ etc.
   derivation: *bůh*$_{\mathrm{N}}$ 'god' $\xrightarrow{\text{h>ž}}$ *bůž-ek*$_{\mathrm{N}}$ (dimin.) (inflection of the derived word: *bůž-k-a*$_{\mathrm{gen.sg}}$, *bůž-k-ovi*$_{\mathrm{dat.sg}}$ etc.)
   derivation: *bůh*$_{\mathrm{N}}$ 'god' $\xrightarrow{\text{ů>o, h>ž}}$ *bož-í*$_{\mathrm{A}}$ 'god's' (inflection of the derived word: *bož-ího*$_{\mathrm{gen.sg}}$, *bož-ímu*$_{\mathrm{dat.sg}}$ etc.)

3. the alternation that exhibits the lemma of the derivative with respect to the lemma of the base occurs in the inflected forms of the base (cf. type 1) but, moreover, inflectional forms of the derivative include an alternation with respect to the lemma of the derivative but not to the lemma of the base:

   (48) inflection of the base word: *sta**r**-ý*$_{\mathrm{A}}$ 'old', *star-ého*$_{\mathrm{gen.sg.masc.anim}}$, *sta**ř**-í*$_{\mathrm{nom.pl.masc.anim}}$ etc.
   derivation: *starý*$_{\mathrm{A}}$ 'old' $\xrightarrow{\text{r>ř}}$ *stař-ec*$_{\mathrm{N}}$ 'old man' (inflection of the derived word: *sta**r**-c-e*$_{\mathrm{gen.sg}}$, *sta**r**-c-i*$_{\mathrm{dat.sg}}$ etc.)

The fact that the alternation observed between a base lemma and the lemma of the derivative can be found in inflectional forms of the base (as in the type 2 and 3) were employed in order to find base words for words with alternations in the root

morpheme in DeriNet (see Sect. 4.4). The type 1 above and the relations between the inflectional forms of the derived word and of the base in 3 were not relevant for our purpose.

4. Alternations in the DeriNet database of derivational relations

In this section, the methods used for the establishment of derivational links in DeriNet are described; the main focus is on which type of morphographemic alternations was modelled by the individual method (for general aspects of the build-up of the database see Ševčíková and Žabokrtský, 2014b; Žabokrtský et al., 2016). String-substitution rules, which constitute the methodological core of our approach (Sect. 4.1 and 4.2), were efficient for modelling frequent alternations in the final grapheme of the stem. A significant portion of derivational relations, often with multiple alternations connected with deverbal derivation, was extracted from the inflectional dictionary MorfFlex CZ (Sect. 4.3). In order to cover alternations in roots that emerged for small groups of words or even for individual words only, inflectional paradigms were exploited for alternations and used for the search of the base-target pairs in DeriNet (Sect. 4.4). Alternations connected with prefixation of verbs were handled separately (Sect. 4.5).

4.1. Searching base adjectives for selected groups of derived words

The DeriNet database was initialised in 2013 to underpin the linguistic research project on deadjectival derivation in Czech with a solid data resource. In a set of lexemes extracted from a large corpus of Czech (Bojar et al., 2012), base adjectives were searched for selected groups of derived words. Deadjectival nouns and adverbs were linked to the base adjectives using heuristics that were manually compiled as regular expressions substituting the final string of the derived word for an adjectival string.[16] For instance, the derivational rule in (49), based on the respective regular expression, was used to identify pairs of an adjective (A) ending in *-ý* and a noun (N) consisting of the same grapheme string except for the final *-ost* instead of the adjectival *-ý*. Only few analogous rules were sufficient to cover all nouns in *-ost* (cf. (50)) and to link most of the adverbs with their base adjectives (51).

(49) A-ý>N-ost: *závislý*$_{\mathrm{A}}$ ‘dependent’ → *závislost*$_{\mathrm{N}}$ ‘dependency’

(50) A-í>N-ost: *revoluční*$_{\mathrm{A}}$ ‘revolutionary’ → *revolučnost*$_{\mathrm{N}}$ ‘revolutionarity’
A-í>N-nost: *budoucí*$_{\mathrm{A}}$ ‘future’ → *budoucnost*$_{\mathrm{N}}$ ‘future’

(51) A-ý>D-e: *bílý*$_{\mathrm{A}}$ ‘white’ → *bíle*$_{\mathrm{D}}$ ‘white(ly)’
A-ý>D-ě: *krutý*$_{\mathrm{A}}$ ‘cruel’ → *krutě*$_{\mathrm{D}}$ ‘cruelly’

[16] The strings corresponded either to suffixes, or to inflectional endings, or were longer (and included one or even more characters of the root morpheme). In the paper, rules based on these strings are therefore generally called “string-substitution rules”.

A-í>D-ě: *revoluční*$_{\mathrm{A}}$ 'revolutionary' → *revolučně*$_{\mathrm{D}}$ 'revolutionary'
A-ý>D-y: *přátelský*$_{\mathrm{A}}$ 'friendly' → *přátelsky*$_{\mathrm{D}}$ 'in a friendly way'

At this phase, alternations did not seem to be a significant issue since they were not frequent in our sample of deadjectival derivation or, more precisely, many of the alternations were not mirrored in writing. For instance, in the most frequent group of deadjectival adverbs (with the suffix *-ě*), or in adjectives in *-ičký* and *-inký*, the final consonant of the root is palatalized in pronunciation but stays unpalatalized in writing as the palatalization is represented by the initial vowel of the suffix (52), (53).

(52) *pěkný*$_{\mathrm{A}}$ 'nice' → *pěkně*$_{\mathrm{D}}$ 'nicely'

(53) *chudý*$_{\mathrm{A}}$ 'poor' → *chudičký*$_{\mathrm{A}}$ 'dirt-poor'

There were only several hundreds of derived words with alternations in the data set in total. Alternations in the final graphemes of the stem were encoded in specific derivational rules such as (54) and (55). The entire word-formation type of deadjectival names of languages in *-ina* which includes a group alternation (*sk*>*št* or *ck*>*čt*) was possible to be covered only by two rules in (56).

(54) A-cký>N-čnost: *praktický*$_{\mathrm{A}}$ 'practical' → *praktičnost*$_{\mathrm{N}}$ 'practicality'

(55) A-ký>D-ce: *blízký*$_{\mathrm{A}}$ 'close' → *blízce*$_{\mathrm{D}}$ 'closely'
A-chý>D-še: *jednoduchý*$_{\mathrm{A}}$ 'simple' → *jednoduše*$_{\mathrm{D}}$ 'simply'
A-rý>D-ře: *dobrý*$_{\mathrm{A}}$ 'good' → *dobře*$_{\mathrm{D}}$ 'well'

(56) A-ský>N-ština: *arabský*$_{\mathrm{A}}$ 'Arabic' → *arabština*$_{\mathrm{N}}$ 'Arabic language'
A-cký>N-čtina: *anglický*$_{\mathrm{A}}$ 'English' → *angličtina*$_{\mathrm{N}}$ 'English language'

In our data sample, only few base adjectives underwent alternations in the root. However, since there was a varied spectrum of vowel alternations in the roots and they occurred selectively with individual affixes (see (57) and (58)), base-target pairs with root alternations were identified individually and linked manually in the data.

(57) *mladý*$_{\mathrm{A}}$ 'young' $\xrightarrow{\text{a>á}}$ *mládí*$_{\mathrm{N}}$ 'youth'
mladý$_{\mathrm{A}}$ 'young' $\xrightarrow{\text{a>á}}$ *mládě*$_{\mathrm{N}}$ 'baby animal'
mladý$_{\mathrm{A}}$ 'young' → *mladě*$_{\mathrm{D}}$ 'in a young manner'
mladý$_{\mathrm{A}}$ 'young' $\xrightarrow{\text{a>á}}$ *mládnout*$_{\mathrm{V}}$ 'to become younger'

(58) *bílý*$_{\mathrm{A}}$ 'white' $\xrightarrow{\text{í>ě}}$ *běloučký*$_{\mathrm{A}}$ 'purely white'
bílý$_{\mathrm{A}}$ 'white' $\xrightarrow{\text{í>ě}}$ *běloušˇ*$_{\mathrm{N}}$ 'white horse'
bílý$_{\mathrm{A}}$ 'white' $\xrightarrow{\text{í>ě}}$ *bělit*$_{\mathrm{V}}$ 'to bleach'
bílý$_{\mathrm{A}}$ 'white' → *bílit*$_{\mathrm{V}}$ 'to paint white'

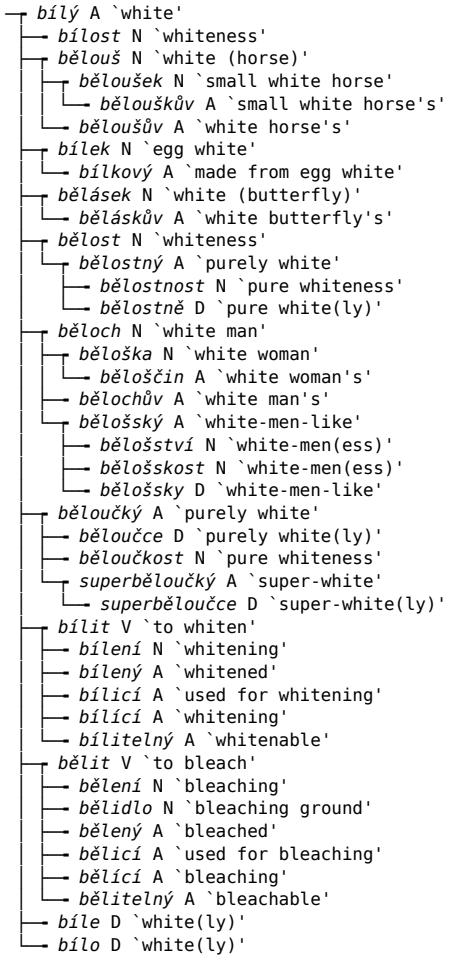

Figure 1. Derivational tree with the root adjective bílý

A derivational tree with the root adjective *bílý* 'white' is displayed in Fig. 1.[17] In the tree, each lexeme is connected by an edge with its direct base word; each edge thus corresponds to a single derivational step. There are five nouns derived directly from *bílý*, namely *bílost*, *běloušˇ*, *bílek*, *bělásek*, *bělost*, and *běloch*, as listed from the top to the bottom of the tree. The noun *běloch* is the base word for the noun *běloška* and for two adjectives (*bělochův*, *bělošský*); on the latter adjective, three other lexemes (*bělošskost*, *bělošství*, *bělošsky*) are based (see the tree for English equivalents).

The adjectival data, under the title AdjDeriNet, were published in 2014 (consisting of app. 18 thousand adjectives with more than 26 thousand nouns, adjectives, verbs and adverbs derived from them; Ševčíková and Žabokrtský, 2014a).

4.2. Alternations in string-substitution rules

The decision to overcome the limitation to deadjectival derivation and to extend the repertory of derivational relations (with the ambition of identifying as many derivational relations in the data as possible) was connected with an attempt to automatize the process of identification of the candidate base-target pairs.

Based on the assumption that lemmas that share a sufficiently long sequence of characters are likely to be derivationally related, pairs of lemmas with a high string similarity (from the left-most character) were identified automatically and, subsequently, grouped according to the strings in which the pair members differed. The differing suffix strings were formalized as string-substitution rules for app. 400 most frequent groups. Out of the list of rules thus obtained, 35 rules were manually selected that reliably corresponded to derivational relations of a base word and a word immediately derived from it.[18] As the next step, the direction of the relation was determined in the rules. The string of the base word was mostly shorter than the corresponding string in the target word; see (59). A single rule often matched several word-formation types; e.g. the first rule in (59) covers derivation of feminine profession nouns from masculine counterparts as well as diminutivization of feminine nouns, the third rule deverbal derivation of both agentive nouns and instrument nouns. Only three out of 35 rules involved alternations, namely an alternation in the final consonant of the stem (60).

[17] In the paper, a simple tree representation was preferred to the graphical output provided by the tools DeriNet Search and DeriNet Viewer since the simple tree requires less space. The tools can be used online for searching the DeriNet data; see `http://ufal.mff.cuni.cz/derinet/search` and `http://ufal.mff.cuni.cz/derinet/viewer`.

[18] The remaining rules either matched less frequent relations, or corresponded to relations between pairs of words that are related only indirectly. For instance, the candidate rule A-ův>A-čin corresponds to the relation between a masculine and a feminine possessive adjective (*manželův* 'husband's' and *manželčin* 'wife's') that both belong into the same derivational family and, thus, into the same derivational tree but are not in the direct base-target relation.

(59) N-a>N-ka: *policista*$_{\mathrm{N}}$ 'policeman' → *policistka*$_{\mathrm{N}}$ 'policewoman', *škola*$_{\mathrm{N}}$ 'school' → *školka*$_{\mathrm{N}}$ 'kindergarden'
A-ý>N-ec: *báz-liv-ý*$_{\mathrm{A}}$ 'timid' → *báz-liv-ec*$_{\mathrm{N}}$ 'coward'
V-t>N-č: *bav-i-t*$_{\mathrm{V}}$ 'entertain' → *bav-i-č*$_{\mathrm{N}}$ 'entertainer', *vypín-a-t*$_{\mathrm{V}}$ 'to switch off' → *vypín-a-č*$_{\mathrm{N}}$ 'switch'

(60) N-ce>A-ční: *akc-e*$_{\mathrm{N}}$ 'action' → *akč-ní*$_{\mathrm{A}}$ 'action'
N-k>N-ček: *zob-ák*$_{\mathrm{N}}$ 'beak' → *zob-áč-ek*$_{\mathrm{N}}$ (dimin.)
N-ce>N-čka: *had-ic-e*$_{\mathrm{N}}$ 'hose' → *had-ič-ka*$_{\mathrm{N}}$ (dimin.)

The set of 35 rules was applied to the data in order to find candidate pairs of base-target words. Incorrect candidate pairs were excluded manually; e.g. in (61) both the suggested base and target nouns are directly derived from the verb *rýt* 'to engrave' (in examples, the correct base word follows in parentheses). A significant portion of the excluded pairs was defective due to the inappropriate approach to alternations (62).

(61) *rytý*$_{\mathrm{A}}$ 'engraved' $\not\rightarrow$ *rytec*$_{\mathrm{N}}$ 'engraver' (*rýt*$_{\mathrm{V}}$ 'to engrave' $\xrightarrow{\text{ý>y}}$ *rytec*$_{\mathrm{N}}$)

(62) *kareta*$_{\mathrm{N}}$ 'Caretta turtle' $\not\rightarrow$ *karetní*$_{\mathrm{A}}$ 'card (game)' (*karta*$_{\mathrm{N}}$ 'card' $\xrightarrow{\text{0>e}}$ *karetní*$_{\mathrm{A}}$)
let$_{\mathrm{N}}$ 'flight' $\not\rightarrow$ *letní*$_{\mathrm{A}}$ 'summer (time)' (*léto*$_{\mathrm{N}}$ 'summer' $\xrightarrow{\text{é>e}}$ *letní*$_{\mathrm{A}}$)

A significantly larger list of (app. 450) string-substitution rules was compiled manually from a representative grammar of Czech (Karlík et al., 2000). As compared to the automatically extracted rules used in the previous step, the manually compiled rules concerned less frequent word-formation types (for instance, deverbal nouns denoting actions, or collective nouns derived from nouns (63)) and, moreover, some of them included highly frequent morphographemic alternations of all types (vowel insertion and deletion, quantitative vowel alternations, and palatalization). Thus, for instance, the rules in (64) match derivation of feminine profession nouns from masculines without an alternation and with the alternation *c*>*č*, which is very frequent with this word-formation type. Similarly, the second rule in (65) includes a frequent vowel deletion that is associated with the derivation of adjectives from nouns. The application of the rules on the data was followed by manual annotation of incorrect base-target pairs, similarly as with the automatically extracted rules.

(63) V-it>N-ba: *léčit*$_{\mathrm{V}}$ 'to treat' → *léčba*$_{\mathrm{N}}$ 'treatment'
N->N-stvo: *člen*$_{\mathrm{N}}$ 'member' → *členstvo*$_{\mathrm{N}}$ 'members'

(64) N->N-ka: *učitel*$_{\mathrm{N}}$ 'teacher' → *učitelka*$_{\mathrm{N}}$ 'female teacher'
N-c>N-čka: *herec*$_{\mathrm{N}}$ 'actor' → *herečka*$_{\mathrm{N}}$ 'actress'

(65) N->A-ový: *achát*$_{\mathrm{N}}$ 'agate' → *achátový*$_{\mathrm{A}}$ 'agate'
N-ek>A-kový: *bílek*$_{\mathrm{N}}$ 'egg white' → *bílkový*$_{\mathrm{A}}$ 'made from egg white' (see Fig. 1)

Table 1. Frequency list of 28 alternation pairs applied in the string-substitution rules

altern.	freq.	altern.	freq.	altern.	freq.	altern.	freq.	altern.	freq.
i>*í*	594	*á*>*a*	104	*h*>*ž*	31	*n*>*ň*	13	*r*>*ř*	4
í>*i*	314	*ě*>*í*	77	*o*>*u*	26	*u*>*ou*	9	*é*>*í*	2
í>*ě*	191	*z*>*ž*	53	*ý*>*y*	21	*o*>*ů*	8	*e*>*é*	2
ou>*u*	178	*a*>*á*	50	*é*>*e*	21	*d*>*ď*	8	*u*>*ú*	1
s>*š*	116	*y*>*ý*	36	*ch*>*š*	17	*ů*>*o*	7		
e>*í*	114	*k*>*č*	36	*c*>*č*	15	*t*>*ť*	7		

Since so far we were able to cover only alterations explicitly encoded in the string-substitution rules the amount of which was still rather limited, an experiment was carried out that allowed alternations in stems during application of both automatically extracted and manually compiled rules. A total of 18 vowel alternations and 10 consonant alternations were selected in advance and applied mechanically together with each string-substitution rule. No more than one alternation was allowed in each pair in order to prevent unmanageable overgeneration of base-target candidate pairs. Nevertheless, if the alternation was applied together with a string-substitution rule that encoded an alternation too, derivations with up to two alternations might be covered for the first time (66).

(66) N-k>A-čí: *ptákN* ‘bird’ $\xrightarrow{\text{á>a, k>č}}$ *ptačíA* ‘bird’s’

Examples of incorrectly suggested pairs were rejected by a human annotator (67). In total, alternations were applied with more than 1,600 derivational links confirmed within the manual annotation; see the frequency list in Table 1.

(67) *kuřeN* ‘chicken’ $\xrightarrow{\text{u>ou}}$ *kouřovýA* ‘smoky’ (*kouřN* ‘smoke’ → *kouřovýA*)
ženaN ‘woman’ $\xrightarrow{\text{e>í}}$ *žíněnýA* ‘made of horsehair’ (*žíněN* ‘horsehair’ → *žíněnýA*)
cenaN ‘price’ $\xrightarrow{\text{e>í}}$ *cínovýA* ‘made of tin’ (*cínN* ‘tin’ → *cínovýA*)
celaN ‘cell’ $\xrightarrow{\text{e>í}}$ *cílovýA* ‘finishing’ (*cílN* ‘finish’ → *cílovýA*)

Application of string-substitution rules was extremely efficient. The steps described in this subsection yielded 350 thousand derivational relations in DeriNet 1.4.

4.3. Alternations extracted from the MorfFlex CZ dictionary

Another considerable portion of derivational links in the DeriNet database was extracted from the MorfFlex CZ dictionary. In MorfFlex CZ, derivational information was encoded as a part of the so-called technical suffix of the lemma (Hajič, 2004;

Hana et al., 2005). The technical suffix **2t* in (68) means that by substituting two final graphemes of the lemma for the grapheme *t*, the base of the adjective is reconstructed (the verb *hubnout* 'to lose weight').

(68) *hubnoucí_^(*2t)* 'losing weight'

MorfFlex CZ was exploited to identify base words for high-frequency groups of words derived mostly from verbs. Derivation from verbs in Czech is specific in that inflected verbal forms rather than the infinitive itself often serve as the base word in derivation. However, since individual verb forms are not involved in DeriNet and are all represented by the infinitive, words derived from different verbal forms had to be linked up to the particular infinitive in the database. Radical changes in the formal shape of deverbal adjectives with respect to the base infinitive are demonstrated in (69) and (70), the changes though do not in fact relate to derivation, but are to be traced back to the inflection (the respective verbal form that entered the derivation is given in square brackets after the infinitive). The formation of deverbal adjectives from transgressives (69) and participles (70) is thus very close to inflection and has been discussed as a transition zone between inflection and derivation in Czech (Dokulil, 1962, pp. 44; Karlík et al., 2000, pp. 172f).

(69) *hnát*$_{\mathrm{V}}$ 'to drive' [*ženouc*$_{\mathrm{transgr.fem.sg.pres.act.impf}}$] → *ženoucí*$_{\mathrm{A}}$ 'driving'
setřít$_{\mathrm{V}}$ 'to wipe' [*setřevši*$_{\mathrm{transgr.fem.sg.past.act.pf}}$] → *setřevší*$_{\mathrm{A}}$ 'wiped'

(70) *hnát*$_{\mathrm{V}}$ 'to drive' [*hnán*$_{\mathrm{3.sg.masc.ptcp.pass.impf}}$] → *hnaný*$_{\mathrm{A}}$ 'driven'
projít$_{\mathrm{V}}$ 'to expire' [*prošel*$_{\mathrm{3.sg.masc.ptcp.past.pf}}$] → *prošlý*$_{\mathrm{A}}$ 'expired'

In addition to the deverbal derivation, the technical suffixes were related to possessive adjectives derived from nouns, which are another word-formation type from the transition zone between derivation and inflection. Some of the most frequent technical suffixes exploited in DeriNet are listed in Table 2. For each suffix, an example lemma is provided and the suffix information is reformulated as a base-target pair.

As for morphographemic alternations, the technical suffixes **3at* and **3it* are examples of vowel alternations in the final grapheme of the stem; vowel deletion is encoded in the last technical suffix in Table 2.

Derivational information from the technical suffixes in MorfFlex CZ was used to establish 399 thousand derivational links in DeriNet 1.4.

4.4. Exploiting inflectional paradigms for description of alternations in derivation

Although alternations were included in each of the methods reported on so far, we were still not satisfied with the coverage of lexemes with alternations. Therefore, a method was proposed that targeted specifically at connecting derived words containing alternations with the correct base word. It focused on alternations that were difficult to cover with the methods described above, especially on changes occurring

Table 2. Technical suffixes of lemmas in MorfFlex used for the establishment of derivational links in DeriNet

technical suffix	lemma with the technical suffix	corresponding derivational relation
**2t*	*hubnoucí_^(*2t)*	*hubnout*$_{\mathrm{V}}$ 'to slim down' → *hubnoucí*$_{\mathrm{A}}$ 'slimming'
	*marinovaný_^(*2t)*	*marinovat*$_{\mathrm{V}}$ 'marinate' → *marinovaný*$_{\mathrm{A}}$ 'marinated'
**4*	*popsatelný_^(*4)*	*popsat*$_{\mathrm{V}}$ 'to describe' → *popsatelný*$_{\mathrm{A}}$ 'describable'
**3at*	*dělání_^(*3at)*	*dělat*$_{\mathrm{V}}$ 'to do' → *dělání*$_{\mathrm{N}}$ 'doing'
**3it*	*bílení_^(*3it)*	*bílit*$_{\mathrm{V}}$ 'to whiten' → *bílení*$_{\mathrm{N}}$ 'whitening' (see Fig. 1)
**2*	*manželův_^(*2)*	*manžel*$_{\mathrm{N}}$ 'husband' → *manželův*$_{\mathrm{A}}$ 'husband's'
**3ec*	*otcův_^(*3ec)*	*otec*$_{\mathrm{N}}$ 'father' → *otcův*$_{\mathrm{A}}$ 'father's'

"deeper" in the root morpheme and on vowel deletion in suffixes. We exploited the fact that alternations which are identified in derived words with respect to their base words might be identical with those observed in the inflectional paradigm of the respective base word (cf. an analogous approach for French by Bonami et al., 2009).

Since lemmas in DeriNet were taken from the inflectional dictionary MorfFlex CZ and both resources are interconnected, information on inflection of DeriNet lemmas is easily accessible. The core issue that alternations are not marked in MorfFlex CZ has been overcome by a provisional, rather technical solution. Each lemma was compared letter-wise from left to right against each of its inflected forms. At least one final grapheme of the inflected form was supposed to be the inflectional ending and thus was not included into the comparison. The inflectional string was marked as containing an alternation, if at any position the character in the lemma differed from that in the inflected form and the pair of differing characters was found in the list of 30 alternations pairs.[19] Due to the alternations *e*>0 and 0>*e*, the lemma might be longer then the inflected substring (71), or the other way round (72). Inflectional strings with one to three alternations with respect to the lemma ("alternated strings") were identified. For a single lemma, more formally different alternated strings could be listed (72).

(71)

lemma	*k*	*r*	*k*	*a*	*v*	***e***	***c***	*krkavec* 'raven'
alternated string	*k*	*r*	*k*	*a*	*v*	**0**	***č***	*krkavč*

(72)

lemma	*d*	*v*	*ů*	***r***	**0**	*dvůr* 'yard'
alternated string 1	*d*	*v*	*o*	*r*	0	*dvor*
alternated string 2	*d*	*v*	*o*	*r*	***e***	*dvore*
alternated string 3	*d*	*v*	*o*	***ř***	0	*dvoř*

[19]The list in Table 1 was enriched with the alternations *e*>0 and 0>*e* for this purpose.

The list of lemmas and corresponding alternated strings was used as input data for the string-substitution rules compiled in the previous steps of the annotation process. The string-substitution (esp. string-adding) rules were applied to the alternated strings instead of to the lemma and the existence of output string suggested by the rule was attested in the data. Manual annotation confirmed app. 2,300 derivational relations that have not been identified so far.

For instance, the rule in (73) was applied on the alternated string from (71) in order to create a link to a derived adjective; the first one out of the alternated strings in (72) turned out to be most effective for creating links between *dvůr* and its derivatives (74).

(73) N->A-í: *krkavec*$_{\mathrm{N}}$ 'raven' / *krkavč* → *krkavčí*$_{\mathrm{A}}$ 'belonging to raven'

(74) N->N-ec: *dvůr*$_{\mathrm{N}}$ 'yard'/ *dvor* → *dvorec*$_{\mathrm{N}}$ 'court'
N->N-ek: *dvůr*$_{\mathrm{N}}$ 'yard'/ *dvor* → *dvorek*$_{\mathrm{N}}$ (dimin.)
N->A-ský: *dvůr*$_{\mathrm{N}}$ 'yard'/ *dvor* → *dvorský*$_{\mathrm{A}}$ 'court (etiquette)'
N->A-ní: *dvůr*$_{\mathrm{N}}$ 'yard'/ *dvor* → *dvorní*$_{\mathrm{A}}$ 'court (lady)'

4.5. Alternations in derivation of verbs from verbs

Derivation of verbs from verbs was addressed separately. In Czech, verbs are derived from verbs by suffixation and prefixation. Prefixation is even prevailing in formation of verbs whereas suffixes predominate over prefixes in derivation of other part-of-speech categories in Czech. Deverbal derivation of verbs is connected with a significant amount of alternations.Deverbal prefixation and suffixation of verbs are both closely interconnected with the category of aspect.[20]

Prefixation either changes imperfective verbs into perfective ones (see the pure aspectual pair of verbs in (75)), or modifies the lexical meaning of an imperfective ((77) and (76)) or perfective verb (resulting in another perfective; (78)). In monosyllabic verbs with a long vowel, the vowel is shortened during prefixation systematically (in addition to (75) and (76), verbs *znát* 'to know', *brát* 'to take', *hnát* 'to ride' belong to this group). Suffixation is used especially to form imperfective counterparts from perfective verbs (79), to derive iterative verbs from imperfectives (80), or secondary imperfectives from prefixed perfectives (81).[21] Suffixation is connected mostly with

[20] In spite of a long-term discussion on the category of aspect, the status of this category is far from clear in Czech and other Slavic languages (e.g. Vey, 1952; Comrie, 1976; Mel'čuk, 1976; Kopečný, 1962; Komárek, 2006). In DeriNet, and thus in the present paper, derivation of verbs is treated with the primary focus on formal features, without respect to whether the affix changes just the aspect of the verb (e.g. "pure perfectivizing" prefixes in Czech) or whether it modifies the lexical meaning of the base verb. For a linguistically rooted discussion on the representation of derivational relations in verbal families with regard to the aspect see (Ševčíková et al., 2017, in press).

[21] The possibility to form a secondary imperfective is used to distinguish pure perfectivizing prefixes (cf. the prefixed derivative in ex. (75) from which the secondary imperfective cannot be derived) from other prefixes (cf. ex. (81) derived from the prefixed verb in ex. (76)).

Table 3. Prefixes used in the verb-to-verb derivation

ad-	*do-*	*na-*	*o-*	*ot-*	*pode-*	*při-*	*roze-*	*u-*	*vý-*	*zá-*
bez-	*dů-*	*ná-*	*ob-*	*ote-*	*pro-*	*před-*	*s-*	*ú-*	*vz-*	*ze-*
de-	*in-*	*nad-*	*obe-*	*pa-*	*přes-*	*půl-*	*se-*	*v-*	*vze-*	*zne-*
des-	*ko-*	*nade-*	*od-*	*po-*	*před-*	*re-*	*sou-*	*ve-*	*z-*	*zu-*
dez-	*kon-*	*ne-*	*ode-*	*pod-*	*předes-*	*roz-*	*sub-*	*vy-*	*za-*	*zů-*

alternations in the root morpheme that are often specific for the particular pair of verbs or are limited to small groups of verbs.

(75) *ps-á-t*$_{\texttt{Vimpf}}$ 'to write' $\xrightarrow{\text{á>a}}$ *na-ps-a-t*$_{\texttt{Vpf}}$ 'to write down'

(76) *ps-á-t*$_{\texttt{Vimpf}}$ 'to write' $\xrightarrow{\text{á>a}}$ *za-ps-a-t*$_{\texttt{Vpf}}$ 'to record'

(77) *skák-a-t*$_{\texttt{Vimpf}}$ 'to jump' $\rightarrow$ *vy-skák-a-t*$_{\texttt{Vpf}}$ 'to jump out'

(78) *skoč-i-t*$_{\texttt{Vpf}}$ 'to jump' $\rightarrow$ *vy-skoč-i-t*$_{\texttt{Vpf}}$ 'to jump out'

(79) *skoč-i-t*$_{\texttt{Vpf}}$ 'to jump' $\xrightarrow{\text{o>á, č>k}}$ *skák-a-t*$_{\texttt{Vimpf}}$ 'to jump'

(80) *skák-a-t*$_{\texttt{Vimpf}}$ 'to jump' $\rightarrow$ *skák-áva-t*$_{\texttt{Vimpf.iter}}$ 'to jump'

(81) *za-ps-a-t*$_{\texttt{Vpf}}$ 'to record' $\xrightarrow{\text{0>i}}$ *za-pis-ova-t*$_{\texttt{Vimpf}}$ 'to record'

As in the existing valency lexicon of Czech verbs Vallex (Lopatková et al., 2015) relations between aspectual pairs of verbs derived by suffixation are explicitly marked, we decided to extract these pairs as the first step in our task of creating derivational relations between verbs in DeriNet. Pairs of verbs that are not in a derivational relation (e.g. suppletive aspectual pairs such as *brát*$_{\texttt{Vimpf}}$ 'to take' – *vzít*$_{\texttt{Vpf}}$ 'to take') were excluded from the list. The usage of an existing, reliable lexical resource was preferred to the above presented methods (particularly string-substitution rules) precisely because of the heterogeneous nature of alternations in this type of suffixation. Second, a list of app. 50 prefixes used in deverbal derivation of verbs (vocalized variants listed as separate items; Table 3) was compiled and used to search Vallex for verbs that are derivationally related to the verbs in the extracted list of aspectual pairs. In these two steps, more than 3,100 verbs were found and preliminarily organized into 660 derivational families with a scope reaching from several tens of verbs (cf. families with the verbs *psát*$_{\texttt{Vimpf}}$ 'to write' or *skočit*$_{\texttt{Vpf}}$ 'to jump') up to pairs of verbs such as *šít*$_{\texttt{Vimpf}}$ 'to sew', *ušít*$_{\texttt{Vpf}}$ 'to finish sewing'.

The inner organization of the derivational families into a derivational tree consisting of oriented binary relations could not be inferred unambiguously from the data itself since there are complicated interconnections between the verbs with respect to

Figure 2. Alternative derivational trees with the root skočit *'to jump'. The derivational tree on the left is preferred in the presented approach (cf. ex. (77) to (80)).*

the form and aspectual characteristics that allow to organize the verbs in several competing ways. For instance, when modelling relations between the verbs *skočit*$_{\text{Vpf}}$ 'to jump', *skákat*$_{\text{Vimpf}}$ 'to jump', *vyskočit*$_{\text{Vpf}}$ 'to jump out', and *vyskákat*$_{\text{Vpf}}$ 'to jump out', the last verb can be interpreted either as the perfective counterpart of *vyskočit* formed through suffixation, or as a prefixed perfective derived from *skákat*. We preferred the prefixation (to creation of aspectual pairs by suffixation) to be a more important organizational principle in DeriNet, therefore, the latter interpretation was chosen in (77) and is mirrored in the tree structure on the left-hand side in Fig. 2; the former, refused interpretation corresponds to the tree on the right. The compared trees differ from the point of view of alternations; the preferred organization is connected with alternations along the single edge *skočit*$_{\text{Vpf}}$ 'to jump' → *skákat*$_{\text{Vimpf}}$ 'to jump'.

The following general guidelines for the inner organization of the derivational families into trees were specified:

- if an unprefixed aspectual pair is available in the derivational family (i.e. the aspectual pair differs in suffixes), the perfective verb is the root of the tree:
  e.g. *skočit*$_{\text{Vpf}}$ 'to jump' → *skákat*$_{\text{Vimpf}}$ 'to jump'
- if only an unprefixed imperfective is available with a prefixed perfective counterpart, the imperfective verb is the root of the tree:
  e.g. *šít*$_{\text{Vimpf}}$ 'to sew' → *ušít*$_{\text{Vpf}}$ 'to finish sewing'
- all prefixed perfectives are derived from the unprefixed counterpart;
  the counterpart is either perfective, e.g. *skočit*$_{\text{Vpf}}$ 'to jump' → *naskočit*$_{\text{Vpf}}$ 'to hop on' | *odskočit*$_{\text{Vpf}}$ 'to jump aside' | *poskočit*$_{\text{Vpf}}$ 'to jump up', ...,
  or imperfective, e.g. *skákat*$_{\text{Vimpf}}$ 'to jump' → *přeskákat*$_{\text{Vpf}}$ 'to jump over' | *vyskákat*$_{\text{Vpf}}$ 'to jump out' ...
- secondary imperfectives were linked to particular prefixed perfectives:
  e.g. *naskočit*$_{\text{Vpf}}$ 'to hop on' → *naskakovat*$_{\text{Vimpf}}$ 'to hop on', *poskočit*$_{\text{Vpf}}$ 'to jump up' → *poskakovat*$_{\text{Vimpf}}$ 'to jump up'
- iterative imperfectives as derived from the imperfective:
  e.g. *skákat*$_{\text{Vimpf}}$ 'to jump' → *skákávat*$_{\text{Vimpf.iter}}$ 'to jump' .

Having the derivational families organized into the tree structures (see Fig. 3), further verbs were searched for in the DeriNet data, using all pieces of information avail-

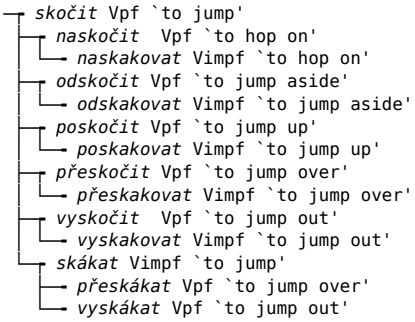

Figure 3. Derivational tree with the root verb skočit *'to jump' consisting of derivationally related verbs from the Vallex dictionary organized according to the adopted guidelines.*

able so far (in particular, string-substitution rules based on the tree structures and list of prefixes). The items were added into the derivational trees according to the guidelines. Compare the simplified derivational trees of the adjective *bílý*$_{\mathrm{A}}$ 'white' in Fig. 4; the simplification consists in displaying verbal nodes only (nodes of other part-of-speech categories were omitted).

The procedure described in Sect. 4.5 resulted in nearly 23 thousand new derivational relations in total. Since the newly connected verbs were mostly roots of subtrees consisting of direct and indirect deverbal derivatives, the new links led to connection of a number of trees into a structure with an extremely high number of nodes.[22]

5. Discussion and conclusions

In the paper, morphographemic alternations were approached from the perspective of semi-automatic modelling of derivational relations in the language resource specialized in derivational morphology of Czech. Methods of creating derivational links in DeriNet were presented with a focus on alternations covered by each of the methods. The method of exploiting inflectional paradigms developed specifically for dealing with alternations with respect to individual lexemes (Sect. 4.4) was less efficient (in terms of absolute frequency of created derivational links) than the string-substitution rules and derivational information from MorfFlex CZ, but it confirmed the feasibility and, also, usefulness of integrating inflectional information into description of derivation. As inflectional resources are elaborated for Czech more comprehensively than derivational data, which seems to be the case for other languages as well, the possible profits should be further explored.

[22]There are nearly 80 trees with more than 500 nodes each in DeriNet 1.4.

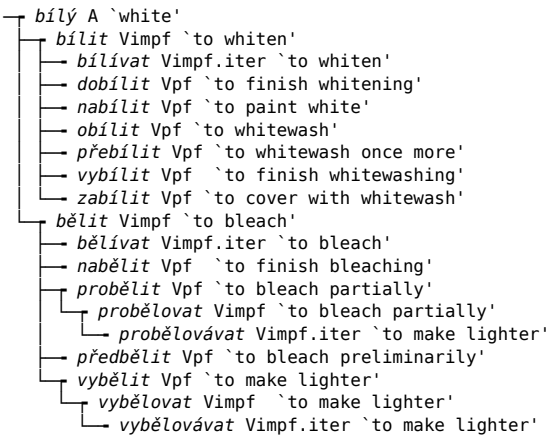

Figure 4. A simplified derivational tree with the root bílý *'white' involving verbs derived directly and indirectly from the adjective, as organized in DeriNet 1.4.*

However, there are still words with alternations that we have not been able to treat so far. The following examples of some significant groups indicate the diversity of problems encountered when extending the coverage of the annotation:

- words with more alternations occurring in a single derivation step; one or more of the alternations usually correlate with alternations in inflection, the other one is in the final grapheme of the stem:

  (82) *vejc-e*$_{\mathrm{N}}$ 'egg' $\xrightarrow{\text{e>a, 0>e, c>č}}$ *vaječ-ný*$_{\mathrm{A}}$ 'made from eggs'
  sníh$_{\mathrm{N}}$ 'snow' $\xrightarrow{\text{í>ě, h>ž}}$ *sněž-ný*$_{\mathrm{A}}$ 'snowy'
  louk-a$_{\mathrm{N}}$ 'meadow' $\xrightarrow{\text{ou>u, k>č}}$ *luč-ní*$_{\mathrm{A}}$ 'meadow'

- words derived from part-of-speech categories that are not contained in DeriNet, particularly from pronouns and numerals:

  (83) *devět*$_{\mathrm{NUM}}$ 'nine' $\xrightarrow{\text{ě>í}}$ *devít-ina*$_{\mathrm{N}}$ '(one) ninth'

- compounds with alternations; a substantial change of the architecture of the database is required in the near future in order to make it possible to represent composition:

  (84) *Bílá*$_{\mathrm{A}}$ *Hora*$_{\mathrm{N}}$ 'White Mountain' (geographical name) $\xrightarrow{\text{í>ě}}$ *běl-o-hor-ský*$_{\mathrm{A}}$ 'from Bílá Hora'

- deverbal nouns are often both formally and semantically based on the whole aspectual pair of verbs (*vysloužit*$_{\mathrm{Vpf}}$ 'to earn', *vysluhovat*$_{\mathrm{Vimpf}}$ 'to earn'); a lin-

guistically adequate solution is to be developed that would enable to connect a word with more than one parent though it is not a compound (without being fused with compounds), etc.

(85) *vy-slouž-i-t*$_{\mathrm{Vpf}}$ 'to earn' $\xrightarrow{\text{y>ý, ou>u}}$ *vý-služ-ba*$_{\mathrm{N}}$ 'retirement' and/or
vy-sluh-ova-t$_{\mathrm{Vimpf}}$ 'to earn' $\xrightarrow{\text{y>ý, h>ž}}$ *vý-služ-ba*$_{\mathrm{N}}$ 'retirement'

The approach to alternations in DeriNet is to be interpreted as the first step in the data-based description of alternations in Czech derivation. The next step is the automatic morphemic segmentation, which makes it possible to look at alternations in connection with particular morphemes. The DeriNet data are expected to be helpful in developing the tools for morphemic segmentation, which is still missing for Czech. For instance, consonant alternations can be detected as an important formal feature indicating the root-suffix (ex. (86)) or suffix-suffix boundary (ex. (87)) while vowel lengthening typically at the second position in nouns derived from verbs (ex. (88)) delimits the prefix-root boundary.

(86) *such-ý*$_{\mathrm{A}}$ 'dry' $\xrightarrow{\text{ch>š}}$ *suš-i-t*$_{\mathrm{V}}$ 'to dry'

(87) *čern-och*$_{\mathrm{N}}$ 'black person' $\xrightarrow{\text{ch>š}}$ *čern-ouš-ek*$_{\mathrm{N}}$ 'black person (demin.)'

(88) *vy-robit*$_{\mathrm{V}}$ 'to product' $\xrightarrow{\text{y>ý}}$ *vý-roba*$_{\mathrm{N}}$ 'production'

Acknowledgements

The research reported on in the present paper was supported by the grant No. GA16-18177S of the Czech Science Foundation.

Bibliography

Aronoff, M. *Word Formation in Generative Grammar*. The MIT Press, Cambridge, 1976.

Aronoff, M. *Morphology by itself: stems and inflectional classes*. The MIT Press, Cambridge, 1994.

Baayen, R., R. Piepenbrock, and L. Gulikers. CELEX2 lexical database. LDC96L14. Philadelphia: Linguistic Data Consortium, 1995.

Baerman, M. *The Oxford Handbook of Inflection*. Oxford University Press, Oxford, 2015.

Baranes, M. and B. Sagot. A Language-independent Approach to Extracting Derivational Relations from an Inflectional Lexicon. In *Proceedings of the 9th International Conference on Language Resources and Evaluation*, pages 2793–2799, Reykjavik, Iceland, May 2014. ISBN 978-2-9517408-8-4.

Bauer, L. *English Word-formation*. Cambridge University Press, Cambridge, 1983.

Bauer, L. Derivational paradigms. In Booij, G. and J. van Marle, editors, *Yearbook of Morphology 1996*, pages 234–256. Kluwer, Dordrecht, 1997.

Bermúdez-Otero, R. and A. McMahon. English Phonology and Morphology. In Aarts, B. and A. McMahon, editors, *The Handbook of English Linguistics*, pages 382–410. Wiley-Blackwell, 2006.

Bojar, O., Z. Žabokrtský, O. Dušek, P. Galuščáková, M. Majliš, D. Mareček, J. Maršík, M. Novák, M. Popel, and A. Tamchyna. The Joy of Parallelism with CzEng 1.0. In *Proceedings of LREC 2012*, pages 3921–3928, Istanbul, Turkey, 2012. ELRA.

Bonami, O., G. Boyé, and F. Kerleroux. L'allomorphie radicale et la relation flexion-construction. In Fradin, Bernard, Françoise Kerleroux, and Marc Plénat, editors, *Aperçus de morphologie du français*, pages 103–125. Presses universitaires de Vincennes, Saint-Denis, 2009.

Booij, G. E. Paradigmatic morphology. In Fradin, Bernard, editor, *La raison morphologique. Hommage á la mémoire de Danielle Corbin*, pages 29–38. Benjamins, Amsterdam, 2008.

Buzássyová, K. *Sémantika slovenských deverbatív*. Veda, Bratislava, 1974.

Bybee, J. L. *Phonology and Language Use*. Cambridge University Press, Cambridge, 2001.

Bybee, J. L. *Frequency of use and the organization of language*. Oxford University Press, New York, 2007.

Bybee, J. L. and M. A. Brewer. Explanation in morphophonemics: Changes in Provençal and Spanish preterite forms. *Lingua*, 52:201–242, 1980.

Comrie, B. *Aspect*. Cambridge University Press, Cambridge, 1976.

Cvrček, V. and P. Vondřička. Nástroj pro slovotvornou analýzu jazykového korpusu. In *Gramatika a korpus*, Hradec Králové, 2013. Gaudeamus.

Čermák, F. *Morfématika a slovotvorba češtiny*. NLN, Prague, 2012.

Daneš, F., M. Dokulil, and J. Kuchař. *Tvoření slov v češtině 2: Odvozování podstatných jmen*. Academia, Prague, 1967.

Dokulil, M. *Tvoření slov v češtině 1: Teorie odvozování slov*. Academia, Prague, 1962.

Dokulil, M. The Prague School's Theoretical and Methodological Contribution to "Word Formation" (Derivology). In Luelsdorff, P. A., editor, *The Prague School of Structural and Functional Linguistics*, pages 123–162. Benjamins, Amsterdam, 1994.

Dokulil, M., K. Horálek, J. Hůrková, M. Knappová, and J. Petr. *Mluvnice češtiny 1*. Academia, Prague, 1986.

Furdík, J. *Slovenská slovotvorba*. Náuka, Prešov, 2004.

Gebauer, Jan. *Historická mluvnice jazyka českého*. Academia, Praha, 1984–1929.

Grzegorczykowa, R., R. Laskowski, and H. Wróble. *Gramatyka współczesnego jezyka polskiego. Morfologia*. WN PWN, Warszawa, 1998.

Hajič, J. *Disambiguation of Rich Inflection: Computational Morphology of Czech*. Karolinum Press, Prague, 2004. ISBN 9788024602820.

Hajič, J. and J. Hlaváčová. MorfFlex CZ. LINDAT/CLARIN digital library at Institute of Formal and Applied Linguistics, Charles University in Prague, 2013. `http://hdl.handle.net/11858/00-097C-0000-0015-A780-9`.

Hajič, J., J. Panevová, E. Hajičová, P. Sgall, P. Pajas, J. Štěpánek, J.í Havelka, M. Mikulová, Z. Žabokrtský, and M. Ševčíková-Razímová. Prague Dependency Treebank 2.0. LDC2006T01. Linguistic Data Consortium, Philadelphia, USA, 2006.

Hana, J., D. Zeman, J. Hajič, H. Hanová, B. Hladká, and E. Jeřábek. *Manual for Morphological Annotation, Revision for the Prague Dependency Treebank 2.0*. UFAL MFF UK, Prague, 2005.

Hansen, K. Wortbildung. In Hansen, B., K. Hansen, A. Neubert, and M. Schentke, editors, *Englische Lexikologie*, pages 27–152. VEB Verlag, Leipzig, 1985.

Haspelmath, M. and A. D. Sims. *Understanding Morphology*. Hodder Education, London, 2010.

Hathout, N. and F. Namer. Démonette, a French derivational morpho-semantic network. *Linguistic Issues in Language Technology*, 11(5):125–168, 2014.

Horecký, J., K. Buzássyová, and J. Bosák. *Dynamika slovnej zásoby súčasnej slovenčiny*. Veda, Bratislava, 1989.

Karlík, P., M. Nekula, and Z. Rusínová. *Příruční mluvnice češtiny*. NLN, Prague, 2000.

Komárek, M. *Příspěvky k české morfologii*. Periplum, Olomouc, 2006.

Kopečný, F. *Slovesný vid v češtině*. Nakladatelství ČSAV, Praha, 1962.

Křen, M., V. Cvrček, T. Čapka, A. Čermáková, M. Hnátková, L. Chlumská, T. Jelínek, D. Kováříková, V. Petkevič, P. Procházka, H. Skoumalová, M. Škrabal, P. Truneček, P. Vondřička, and A. Zasina. SYN2015: A Representative Corpus of Written Czech. UCNK FF UK, Prague, 2015. URL `http://www.korpus.cz`.

Lamprecht, A., D. Šlosar, and J. Bauer. *Historická mluvnice češtiny*. SPN, Praha, 1986.

Lieber, R. and P. Štekauer. *The Oxford Handbook of Derivational Morphology*. Oxford University Press, Oxford, 2014.

Litta, E., M. Passarotti, and C. Culy. Formatio formosa est. Building a Word Formation Based Lexicon for Latin. In *Proceedings of the Third Italian Conference on Computational Linguistics (CLiC-it 2016)*, pages 185–189, Naples, 2016.

Lopatková, M., V. Kettnerová, E. Bejček, A. Vernerová, and Z. Žabokrtský. VALLEX 3.0 – Valency Lexicon of Czech Verbs. Institute of Formal and Applied Linguistics, Charles University in Prague, 2015.

Matthews, P. H. *The Concise Oxford Dictionary of Linguistics*. Oxford University Press, Oxford, 2007.

Mel'čuk, I. A. *Cours de morphologie générale. Vol. 2*. Presses de l'Université de Montréal, Montréal, 1976.

Osolsobě, K. *Algoritmický popis české formální morfologie a strojový slovník češtiny*. PhD thesis, Masarykova univerzita, Brno, 1996.

Osolsobě, K. Alternace hlásková. In Karlík, P., M. Nekula, and J. Pleskalová, editors, *Encyklopedický slovník češtiny*, pages 35–36. NLN, Praha, 2002.

Osolsobě, K. *Česká morfologie a korpusy*. Karolinum, Praha, 2014.

Osolsobě, K. Korpusy jako zdroje dat pro úpravy nástrojů automatické morfologické analýzy. *Časopis pro moderní filologii*, 97:136–145, 2015.

Osolsobě, K., D. Hlaváčková, K. Pala, and P. Šmerk. Exploring Derivational Relations in Czech with the Deriv Tool. In *NLP, Corpus Linguistics, Corpus Based Grammar Research*, pages 152–161, Bratislava, Slovakia, 2009. Tribun. ISBN 978-80-7399-875-2.

Pala, K. and D. Hlaváčková. Derivational Relations in Czech WordNet. In *Proceedings of the 2007 ACL Workshop on Balto-Slavonic Natural Language Processing*, pages 75–81, Prague, 2007.

Pala, K. and P. Smrž. Building Czech Wordnet. *Romanian Journal of Information Science and Technology*, 7:79–88, 2004.

Pala, K. and P. Šmerk. Derivancze – Derivational Analyzer of Czech. In *Text, Speech, and Dialogue*, volume 9302 of *Lecture Notes in Computer Science*, pages 515–523. Springer International Publishing, 2015. doi: 10.1007/978-3-319-24033-6_58.

Pognan, P. and J. Panevová. Génération automatique de lexèmes slaves à partir de leurs racines historiques. Une des bases de l'enseignement multilingue des langues slaves de l'Ouest (Nord et Sud). *Linguistica*, 52:59–75, 2013.

Pounder, A. *Processes and Paradigms in Word-Formation Morphology*. Mouton de Gruyter, Berlin, 2000.

Razímová, M. and Z. Žabokrtský. Annotation of Grammatemes in the Prague Dependency Treebank 2.0. In *Proceedings of the LREC 2006 Workshop on Annotation Science*, pages 12–19, 2006.

Rubenstein, H. *A Comparative Study of Morphophonemic Alternations in Standard Serbo-Croatian, Czech and Russian*. Columbia University, 1950.

Scheer, T. and M. Ziková. The Havlík Pattern and Directional Lower. In *Formal Approaches to Slavic Linguistics. The Second Cornell Meeting*, pages 471–486, Michigan Slavic Publications, 2010.

Scheer, T., A. Starčević, and M. Ziková. Not all zeros are the same and not all alternating vowels are the same. In *Formal Approaches to Slavic Linguistics (FASL 20)*, pages 305–319, Cambridge, Mass., 2011.

Sedláček, R. *Morphematic analyser for Czech*. PhD thesis, Masarykova univerzita, Brno, 2004.

Sedláček, R. and P. Smrž. A New Czech Morphological Analyzer ajka. In *LNCS / Lecture Notes in Artificial Intelligence. Proceedings of the 4th International Conference Text, Speech and Dialogue (TSD 2001)*, pages 100–107, Berlin, 2001. Springer.

Slavíčková, E. *Retrográdní morfematický slovník češtiny*. Academia, Praha, 1975.

Stankiewicz, E. The Consonantal Alternations in the Slavic Declensions. *Word*, 16:183–203, 1960.

Stankiewicz, E. *The Slavic Languages. Unity in Diversity*. Mouton de Gruyter, Berlin, 1986.

Šefčík, O. Ablaut. In Karlík, P., M. Nekula, and J. Pleskalová, editors, *Nový encyklopedický slovník češtiny*, pages 22–24. NLN, Praha, 2016a.

Šefčík, O. Alternace. In Karlík, P., M. Nekula, and J. Pleskalová, editors, *Nový encyklopedický slovník češtiny*, pages 87–88. NLN, Praha, 2016b.

Ševčíková, M. and Z. Žabokrtský. AdjDeriNet. LINDAT/CLARIN digital library at Institute of Formal and Applied Linguistics, Charles University in Prague, 2014a. URL `http://hdl.handle.net/11234/1-1467`.

Ševčíková, M. and Z. Žabokrtský. Word-Formation Network for Czech. In *Proceedings of the 9th International Conference on Language Resources and Evaluation*, pages 1087–1093, Reykjavik, Iceland, May 2014b. ISBN 978-2-9517408-8-4.

Ševčíková, M., J. Panevová, and P. Pognan. Inflectional and derivational paradigm of verbs in Czech: the role of the category of aspect. In *Abstracts from the First Workshop on Paradigmatic Word Formation Modeling*. Université Toulouse, 2017, in press.

Šimandl, J. *Slovník afixů užívaných v češtině*. Karolinum, Praha, 2016.

Šiška, Z. *Bázový morfematický slovník češtiny*. UPOL, Olomouc, 2005.

Šmerk, P. Fast Morphological Analysis of Czech. In *Proceedings of Third Workshop on Recent Advances in Slavonic Natural Language Processing (RASLAN 2007)*, pages 13–16, Brno, 2007. Masaryk University.

Šnajder, J. DerivBase.hr: A High-Coverage Derivational Morphology Resource for Croatian. In *Proceedings of the 9th International Conference on Language Resources and Evaluation*, pages 3371–3377, Reykjavik, Iceland, May 2014. ISBN 978-2-9517408-8-4.

Štekauer, P. *An Onomasiological Theory of English Word-Formation*. Benjamins, Amsterdam, 1998.

Štekauer, P., S. Valera, and L. Körtvélyessy. *Word-Formation in the World's Languages*. Cambridge University Press, Cambridge, 2012.

Štícha, F. *Akademická gramatika spisovné češtiny*. Academia, Praha, 2013.

Švedova, N. J. *Russkaja gramatika. Vol. 1*. Nauka, Moskva, 1980.

Večerka, R. Palatalizace. In Karlík, P., M. Nekula, and J. Pleskalová, editors, *Nový encyklopedický slovník češtiny*, pages 1287–1288. NLN, Praha, 2016.

Vey, M. Les préverbes vides en tchèque moderne. *Revue des Études de Slave*, 29:82–107, 1952.

Vidra, J., Z. Žabokrtský, M. Ševčíková, and M. Straka. DeriNet 1.0. LINDAT/CLARIN digital library at Institute of Formal and Applied Linguistics, Charles University in Prague, 2015. URL `http://hdl.handle.net/11234/1-1520`.

Vidra, J., Z. Žabokrtský, M. Ševčíková, and M. Straka. DeriNet 1.2. LINDAT/CLARIN digital library at Institute of Formal and Applied Linguistics, Charles University in Prague, 2016. URL `http://hdl.handle.net/11234/1-1807`.

Zeller, B., J. Šnajder, and S. Padó. DErivBase: Inducing and Evaluating a Derivational Morphology Resource for German. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics*, pages 1201–1211, Sofia, Bulgaria, 2013.

Ziková, M. Morfonologické alternace v současné češtině. In Ulilčný, O., editor, *Preliminária k moderní mluvnici češtiny*, pages 177–201. UPOL, Olomouc, 2015.

Ziková, M. Alternace kvantity. In Karlík, P., M. Nekula, and J. Pleskalová, editors, *Nový encyklopedický slovník češtiny*, pages 88–90. NLN, Praha, 2016a.

Ziková, M. Alternace vokálů s nulou. In Karlík, P., M. Nekula, and J. Pleskalová, editors, *Nový encyklopedický slovník češtiny*, pages 90–92. NLN, Praha, 2016b.

Žabokrtský, Z., M. Ševčíková, M. Straka, J. Vidra, and A. Limburská. Merging Data Resources for Inflectional and Derivational Morphology in Czech. In *Proceedings of the 10th International Conference on Language Resources and Evaluation*, pages 1307–1314, Portoroz, Slovenia, 2016.

Appendix: Morphographemic alternations in contemporary Czech

	altern.		example	counter-example
(A, B, C) vowel alternations	a>á (bidir.)	in pref.	*za-bal-i-t → zá-bal* 'to pack' 'wet pack'	*na-hr-á-t → na-hr-á-vka* 'to record' 'recording'
		in root	*vrat-a → vrát-ka* 'gate' (dimin.)	*pat-a → pat-ka* 'heel' (dimin.)
		in suf.	*hled-a-t → hled-á-ní* 'to search' 'search'	*hled-a-t → hled-a-ný* 'to search' 'searched'
	a>e (bidir.)	in root	*úřad → úřed-ník* 'office' 'officer'	*ná-klad → ná-klad-ní* 'load' 'cargo'
	a>ě	in root	*svat-ý → svět-ec* 'holy' 'holy man'	
	a>o	in root	*hrab-a-t → hrob* 'to dig' 'grave'	*s-pad-a-t → s-pad* 'to fall down' 'fallout'
	á>a (bidir.)	in root	*kámen → kamen-ný* 'stone' 'stony'	*památ-ka → památ-ný* 'memory' 'memorable'
		in suf.	*ps-á-t → ps-a-ní* 'to write' 'letter'	
	á>e (bidir.)	in root	*o-třás-t → o-třes* 'to shake' 'shake'	*krás-t → krád-ež* 'to steal' 'theft'
	á>i	in root	*ďábel → dibl-ík* 'devil' 'imp'	*ďábel → ďáblík* 'devil' (dimin.)
	á>í (bidir.)	in root	*přá-t → pří-tel* 'to wish' 'friend'	*hrá-t → hrá-č* 'to play' 'player'
	e>a (bidir.)	in root	*vejc-e → vaječ-ný* 'egg' 'made from eggs'	*strejc → strejc-ův* 'uncle' 'uncle's'
	e>á (bidir.)	in root	*deset → desát-ý* 'ten' 'tenth'	
	e>é (bidir.)	in root	*oheň → ohén-ek* 'fire' (dimin.)	*ú-čes → ú-čes-ek* 'hairstyle' (dimin.)
		in suf.	*prst-en → prst-én-ek* 'ring' (dimin.)	
	e>o	in root	*lež-e-t → po-lož-i-t* 'to lie' 'to lay down'	*žel-e-t → o-žel-e-t* 'to regret' 'to do without'
	e>í (bidir.)	in root	*deset → desít-ka* 'ten' '(number) ten'	*raket-a → raket-ka* 'rocket' (dimin.)
	e>ý	in root	*postel → postýl-ka* 'bed' (dimin.)	*činel → činel-ek* 'cymbal' (dimin.)
		in suf.	*uči-tel → uči-týl-ek* 'teacher' (dimin.)	
	ě>í	in root	*květ → kvít-ek*	*paměť → pamět-ník*

altern.		example	counter-example
(bidir.)		'blossom' (dimin.)	'memory' 'survivor'
ě>á (bidir.)	in root	*paměť → památ-ka* 'memory' 'souvenir'	*Bět-a → Bět-ka* fem. name (dimin.)
é>a	in root	*vléc-t → vlak* 'to pull' 'train'	*vléc-t → vlek* 'to pull' 'ski lift'
é>e (bidir.)	in root	*lét-o → let-ní* 'summer' 'summer-adj'	*bazén → bazén-ek* 'pool' (dimin.)
é>í	in root	*mléko → mlíko* 'milk' '(non-stand.)'	*mléko → mléčný* 'milk' 'milk'
	in suf.	*svíc-en → svíc-ín-ek* 'candlestick' (dimin.)	
é>ý	in root	*okén-ko → okýn-ko* 'window' '(non-stand.)'	
	in suf.	*prst-en → prst-ýn-ek* 'ring' (dimin.)	*prst-en → prst-en-ec* 'ring' 'big ring'
i>e	insuf.	*zlob-i-t → zlob-e-ní* 'to misbehave' 'misbehavior'	*zlob-i-t → zlob-i-vý* 'to misbehave' 'naughty'
i>í (bidir.)	in pref.	*při-děl-i-t → pří-děl* 'to assign' 'ration'	*při-hr-á-t → při-hr-á-vka* 'to pass' 'pass'
	in root	*list → líst-ek* 'leaf' (dimin.)	*sešit → sešit-ek* 'block' (dimin.)
	in suf.	*čum-il → čum-íl-ek* 'gaper' (dimin.)	*text-il → text-il-ka* 'textile' 'textile factory'
í>i (bidir.)	in root	*líp-a → lip-ka* 'linden' (dimin.)	*píst → píst-ek* 'piston' (dimin.)
í>a	in root	*žít → žat-va* 'to mow' 'mowing'	
í>á (bidir.)	in root	*přítel → přátel-ský* 'friend' 'friendly'	
í>e (bidir.)	in root	*říd-i-t → řed-i-tel* 'to lead' 'director'	*z-říd-i-t → z-říz-en-ec* 'to establish' 'attendant'
	in suf.	*zaj-íc → zaj-eč-í* 'hare' 'hare's'	*měs-íc → měs-íč-ní* 'moon' 'lunar'
í>ě (bidir.)	in root	*vítr → větr-ný* 'wind' 'windy'	*mír-a → mír-ný* 'degree' 'moderate'
o>á	in root	*s-klon-i-t → s-klán-ě-t* 'to incline$_{\mathrm{pf}}$' 'to incline$_{\mathrm{impf}}$'	
o>ó (bidir.)	in root	*Božen-a → Bóž-a* (fem. name) (familiar)	*Božen-a → Bož-ka* (fem. name) (familiar)
o>ů (bidir.)	in pref.	*pro-střel-i-t → prů-střel* 'to shoot through' 'shot through'	*pro-slov-i-t → pro-slov* 'to give a speech' 'speech'
	in root	*cop → cůp-ek* 'plait' (dimin.)	*strom → stromek* 'tree' (dimin.)

	altern.		example	counter-example
		in suf.	*lib-ost* → *lib-ůst-ka* 'liking' (dimin.)	
	o>ou	in root	*boř-i-t* → *bour-a-t* 'to destroy$_{\mathrm{pf}}$' 'to destroy$_{\mathrm{impf}}$'	*po-noř-i-t* → *po-noř-ova-t* 'to dip$_{\mathrm{pf}}$' 'to dip$_{\mathrm{impf}}$'
		in suf.	*čern-och* → *čern-ouš-ek* 'black man' (dimin.)	*let-os* → *let-oš-ek* 'this year' 'this year'
	ó>o (bidir.)	in root	*próz-a* → *proz-aický* 'prose' 'prosaic'	
	ou>u (bidir.)	in root	*kouř-i-t* → *kuř-ák* 'to smoke' 'smoker'	*bouř-i-t* → *bouř-e* 'to storm' 'storm'
		in suf.	*ln-ou-t* → *ln-u-tí* 'to adhere' 'adhering'	
	u>ou (bidir.)	in root	*dub* → *doub-ek* 'oak' (dimin.)	*stuh-a* → *stuž-ka* 'ribbon' (dimin.)
	u>ú (bidir.)	in pref.	*u-lovit* → *ú-lovek* 'to catch' 'catch'	*ú-toč-i-t* → *ú-tok* 'to attack' 'attack'
	ú>u (bidir.)	in pref.	*ú-cta* → *u-ctivý* 'respect' 'respectful'	*ú-nav-a* → *ú-nav-ný* 'fatique' 'tiring'
		in root	*úz-ký* → *uz-oučký* narrow' (dimin.)	*útl-ý* → *útl-oučký* 'thin' (dimin.)
	ů>o (bidir.)	in root	*kůž-e* → *kož-ený* 'leather' 'leather-adj'	*kůr-a* → *kůr-ový* 'bark' 'bark-adj.'
	y>ý (bidir.)	in pref.	*vy-br-a-t* → *vý-bor* 'to choose' 'board'	*vy-hlás-i-t* → *vy-hláš-ka* 'to declare' 'notice'
		in root	*vys-oký* → *výš-ka* 'high' 'height'	*ryb-a* → *ryb-ka* 'fish' (dimin.)
	ý>y (bidir.)	in root	*hýb-a-t* → *hyb-ný* 'to move' 'movable'	*hýb-a-t* → *hýb-ací* 'to move' 'moving'
(D) vowel deletion	e>0 (bidir.)	in root	*kart-a* → *karet-ní* 'card' 'card-adj'	*nárt* → *nárt-ní* 'instep' 'instep-adj.'
		in suf.	*dár-ek* → *dár-k-ový* 'gift' 'gift-adj'	*do-tek* → *do-tek-ový* 'touch' 'touch-adj.'
	é>0 (bidir.)	in root	*déšť* → *dšt-í-t* 'rain' 'to rain'	
	u>0	in root	*such-ý* → *sch-nout* 'dry' 'to dry'	
(E) vowel insertion	0>e (bidir.)	in root	*hr-a* → *her-ní* 'play' 'playing'	*hr-á-t* → *hr-a* 'to play' 'play'
		in suf.	*služ-b-a* → *služ-eb-ní* 'service' 'business-adj'	
	0>é (bidir.)	in root	*okn-o* → *okén-ko* 'window' (dimin.)	
	0 >o	in root	*hřm-ít* → *hrom*	

	altern.		example	counter-example
			‘to thunder’ ‘thunder’	
	0>i	in root	*na-ps-a-t* → *ná-pis* ‘to write’ ‘sign’	
	0>í	in root	*ps-á-t* → *pís-ař* ‘to write’ ‘writer’	
	0>y	in root	*za-mk-nou-t* → *za-myk-a-t* ‘to lock$_{pf}$’ ‘to lock$_{impf}$’	
	0>ý	in root	*na-zv-a-t* → *na-zýv-a-t* ‘to call$_{pf}$’ ‘to call$_{impf}$’	
(F) consonant alternations	c>č	in root	*ovc-e* → *ovč-í* ‘sheep’ ‘sheep’s’	
		in suf.	*chlap-ec* → *chlap-ečˇ-ek* ‘boy’ (dimin.)	
	c>k (bidir.)	in root	*péc-t* → *pek-ař* ‘to bake’ ‘baker’	*pec* → *pec-ař* ‘oven’ ‘oven builder’
	č>k (bidir.)	in root	*breč-e-t* → *brek* ‘to cry’ ‘cry’	
	d>ď (bidir.)	in root	*hněd-ý* → *hněď* ‘brown’ ‘brown (colour)’	*sled-ova-t* → *sled* ‘to follow’ ‘sequence’
	d>z	in root	*tvrd-ý* → *tvrz* ‘hard’ ‘fort’	*hod-i-t* → *hod* ‘to throw’ ‘throw’
	ď>d (bidir.)	in root	*loď* → *lod-ní* ‘ship’ ‘shipping’	
	g>ž	in root	*chirurg* → *chirurž-ka* ‘surgeon’ ‘woman surgeon’	
	h>z	in root	*tuh-ý* → *tuz-e* ‘solid’ ‘solid(ly)’	
	h>ž (bidir.)	in root	*sníh* → *sníž-ek* ‘snow’ (dimin.)	
	ch>š (bidir.)	in root	*živočich* → *živočiš-ný* ‘animal’ ‘animal-adj’	*všechen* → *po-všech-ný* ‘all’ ‘general’
		in suf.	*čern-och* → *čern-oš-ka* ‘black man’ ‘black woman’	
	k>c (bidir.)	in root	*trpk-ý* → *trpc-e* ‘bitter’ ‘bitterly’	
		in suf.	*blíz-k-ý* → *blíz-c-e* ‘close’ ‘closely’	
	k>č (bidir.)	in root	*ruk-a* → *ruč-ní* ‘hand’ ‘manual’	
		in suf.	*balet-k-a* → *balet-č-in* ‘ballerina’ ‘ballerina’s’	
	k>t	in root	*hrušk-a* → *hrušt-ička*	

	altern.		example	counter-example
			‘pear’ (dimin.)	
		in suf.	*služ-k-a → služ-t-ička*	
			‘housemaid’ (dimin.)	
	n>ň	in root	*čern-ý → čerň*	*u-hrn-ou-t → ú-hrn*
	(bidir.)		‘black’ ‘black (colour)’	‘to sum up’ ‘summary’
		in suf.	*železn-ý → železň-ák*	
			‘iron’ ‘basalt’	
	ň>n	in root	*skříň → skřín-ka*	
	(bidir.)		‘closet’ (dimin.)	
	r>ř	in root	*star-ý → stař-ík*	
	(bidir.)		‘old’ ‘old man’	
	ř>r	in suf.	*truhl-ář → truhl-ár-na*	
	(bidir.)		‘joiner’ ‘joiner’s shop’	
	s>š	in root	*mysl-e-t → myšl-e-ní*	
			‘to think’ ‘thinking’	
	š>ch	in root	*prš-e-t → s-prch-a*	*srš-e-t → srš-atý*
	(bidir.)		‘to rain’ ‘shower’	‘to fume’ ‘furious’
	t>ť	in root	*žlut-ý → žluť*	
	(bidir.)		‘yellow’ ‘yellow (colour)’	
	t>c	in root	*svít-i-t → svíc-e*	
			‘to shine’ ‘candle’	
		in suf.	*o-boh-at-i-t → o-boh-ac-ova-t*	*boh-at-ý → boh-at-ec*
			‘to enrich$_{\mathrm{pf}}$’ ‘to enrich$_{\mathrm{impf}}$’	‘rich’ ‘rich man’
	ť>t	in root	*řiť → řit-ní*	
	(bidir.)		‘anus’ ‘anal’	
	z>ž	in root	*řez-a-t → řež*	*řez-a-t → řez*
			‘to cut’ ‘scuffle’	‘to cut’ ‘section’
	ž>h	in root	*slouž-i-t → sluh-a*	*těž-i-t → těž-ba*
	(bidir.)		‘to serve’ ‘servant’	‘to mine’ ‘mining’
(G) cons. del. and ins.	k>0	in root	*Hamburk → hamburs-ký*	
			‘Hamburg’ ‘from Hamburg’	
	g>0	in root	*Peking → pekin-ský*	
			‘Beijing’ ‘from Beijing’	
	p>0	in root	*kyp-ě-t → ky-nou-t*	*máv-a-t → máv-nou-t*
			‘to brim’ ‘to rise’	‘to wave$_{\mathrm{impf}}$’ ‘to wave$_{\mathrm{pf}}$’
	v>0	in root	*kýv-a-t → ky-nou-t*	
			‘to nod’ ‘to wave’	
	0>j	in root	*mí-t → jmě-ní*	
			‘to have’ ‘property’	
(H) group alternations	ck>čť	root/suf.	*řec-k-ý → řeč-t-ina*	
			‘Greek’ ‘Greek lang.’	
	sk>šť	root/suf.	*rus-k-ý → ruš-t-ina*	

	altern.		example	counter-example
		in suf.	'Russian' 'Russian lang.' *arab-sk-ý* → *arab-št-ina* 'Arabic' 'Arabic lang.'	
	st>šť	in root	*měst-o* → *měšť-an* 'town' 'burgher'	*chvost* → *chvost-an* 'tail' 'saki monkey'
(I) mixed alternations	á>av	in root	*stá-t* → *stav* 'to stand' 'to state'	*stá-t* → *stá-va-t* 'to stand$_{\mathrm{impf}}$' 'to stand$_{\mathrm{iter}}$'
	á>ěj	in root	*vá-t* → *věj-íř* 'to blow' 'fan'	*vá-t* → *vá-nice* 'to blow' 'blizzard'
	á>oj	in root	*stá-t* → *stoj-ící* 'to stand' 'standing'	*stá-t* → *stá-va-t* 'to stand$_{\mathrm{impf}}$' 'to stand$_{\mathrm{iter}}$'
	á>av	in root	*stá-t* → *stav-ba* 'to stand' 'building'	*stá-t* → *stá-va-t* 'to stand$_{\mathrm{impf}}$' 'to stand$_{\mathrm{iter}}$'
	í>ij	in root	*bí-t* → *bij-ící* 'to beat' 'beating'	
	í>oj	in root	*bí-t* → *boj* 'to beat' 'fight'	
	ý>ov	in root	*krý-t* → *krov* 'to cover' 'roof frame'	

Address for correspondence:
Magda Ševčíková
sevcikova@ufal.mff.cuni.cz
Institute of Formal and Applied Linguistics
Faculty of Mathematics and Physics,
Charles University
Malostranské náměstí 25
118 00 Praha 1, Czech Republic