
DIMINUTIVE FORMS OF FAMILY MEMBERS IN ENGLISH AND SLOVAK LANGUAGE - CORPUS BASED ANALYSIS

Marianna Hudcovičová

University of SS.Cyril and Methodius, Faculty of Arts, Department of English and American
Studies, Trnava, Slovakia marianna.hudcovicova@ucm.sk

Abstract: The study concentrates on the analysis of diminutive forms of the members of the family, i.e. mother, father, daughter, son in English and Slovak language. We are able to detect specific suffixes in order to create diminutive forms. The target is to compare these known suffixes and analyse their occurrence, frequency and usage in the electronic corpora. For the purpose of the research Slovak – English Parallel Corpus (par-sken-all-4.0-sk, par-sken-all-4.0-en) was used. The research will focus on approaches used in traditional linguistics and their possible and effective relationship with the approaches within corpus linguistics. In the corpus linguistics, there are used mathematical, statistical methods that enable us to analyse precisely linguistic items. The article will concentrate on application of electronic corpora tools, i.e. average rate frequency, frequency, occurrence in linguistic analysis. Moreover, it will offer experiences and methodological guidelines towards the corpus-based analysis of linguistics. Specific topics to deal with are e.g. morphological characteristics, analysis of the morphemes, word-formation analysis and lexical analysis. In the field of traditional linguistics, linguistic analyses have predominantly focused on the structural aspects of language, with the objective of identifying the various structural units and classes that constitute the language. Conversely, the corpus-based approach places greater emphasis on language use, specifically how speakers and writers make use of the resources available to them within their language. The corpus-based approach studies the actual language use in naturally occurring texts, which are characterised by an authentic character. The electronic corpus allows for the study of the behaviour of individual words. It is capable of identifying words that occur with greater or lesser frequency. In the field of grammar, we can examine a specific grammatical construction with restrictions on form or context can be examined. This allows us to investigate the factors that determine the choice of one construction over another, which may be influenced by factors such as lexis, grammar, or stylistic considerations. The spectrum of the contexts may enrich the general linguistic picture thanks to identification of the genre of the analysed words in the context of the sentence. The aim of the article is to present approaches how to combine methods of traditional linguistics and corpus linguistics in order to prepare the objective linguistic analysis. Thank to large corpora and data, they enable us to analyse linguistic items with precision, to quantify the typical features of linguistic items and to detect the use of diminutive forms representing the members of the family across genres and registers in both languages.

Keywords: linguistics, corpus linguistics, electronic corpora, family, diminutive form

1. INTRODUCTION

The paper will focus on the approaches used in traditional linguistics and their possible and effective relationship with the approaches used in corpus linguistics. Specific topics to deal with are e.g. morphological characteristics, analysis of the morphemes, word-formation analysis and lexical analysis. The study concentrates on the analysis of diminutive forms of the members of the family, i.e. neutral forms are *mother, father, daughter, son* in English and *matka, otec, dcéra, syn* in Slovak language. In order to create diminutive forms in the Slovak language, various suffixes are used for a particular gender of nouns. Noun of the female gender have a limited number of suffixes as well as nouns of male or neutral gender.

We are able to detect specific suffixes in order to create diminutive forms. The target is to compare selected family members in the diminutive forms and analyse their occurrence, frequency and usage in the electronic corpora.

2. DIMINUTIVE FORMS IN SLOVAK AND IN ENGLISH

Slovak and English belong to the typologically different language types, i.e. synthetic features prevail in Slovak and analytic patterns are preferable in English. Affixation is a very productive way within word-formation processes in synthetic languages. On the other hand, analytic languages use independent free morphemes for the linguistic features created by affixation. Dressler et al. (2024) examines the impact of the typological property of morphological richness on diminutive formation in Italian, Austrian German, English, and Tunisian Arabic. It investigates the priority of the pragmatics over the semantics of diminutives in these languages.

Siegel in Štekauer (2000) distinguishes two classes of affixes. This classification is based on the different phonological and morphological behaviour of the two classes. Class I and Class II affixes.

Class I prefixes (+boundary) include e.g. in-, con-, de-.

Class II prefixes (# boundary) include e.g. anti-, pro-, circum-.

Class I (+boundary) suffixes are Latinate ones –able, -en, -ate, -ion, -ity, -y (noun forming)

Class II (# boundary) suffixes are –ness, -less, -ly, -al (noun forming).

Kvetko (2021, p.34) presents suffixation as the process of expressing diminutive forms in English language. This type of suffixation is based on the sense expressed by a suffix, i.e. diminutiveness (emotional relation): -y/-ie, -let, -ling: *daddy, birdie, doggie, Johnny, starlet, piglet, duckling, etc.* Kvetko (ibid.) postulates that there are relatively fewer diminutive suffixes in English than in Slovak. (*John, Johnny; Ján, Janko, Janičko, Janulienko*). Crystal (p. 138, 2019) enumerates among the suffixes examples such as –ling, -let, -ette that modify the neutral form of the nouns and transform them into diminutive nouns.

In the practical analysis of derived words the following characteristics of derivational affixes should be pointed out by Ginzburg (1966).

1. The origin of the affix in question (native or borrowed)
2. The productivity of the affix (highly productive, semi-productive, non-productive)
3. The stems of the parts of speech it is added to (e.g. a noun-stem, an adjective stem, etc.)
4. The part of speech the affix forms (e.g. nouns, adjectives, verbs, etc.)
5. Its meaning (e.g. agent, appurtenance, diminutiveness, etc.)
6. Its stylistic reference (e.g. neutral, terminological, slang, etc.)

Diminutives are formed in Modern English by the native as well as borrowed suffixes. The former suffixes from the early periods have diminished. Trnka (2014) prepared a list of suffixes in the Modern English Period. The suffix –ling has appeared under the influence of Nordic languages even to modify the nouns denoting to people, e.g. *lordling, dukeling, princeling, etc.*) The suffix –et is productive in the particular words, e.g. *islet, riveret, baronet, tablet, bullet, sonnet, etc.* In the 17th century the suffix –ette was borrowed into the English language, it can be found in the words *chemisette, cigarette, etiquette, etc.* A very productive suffix is the suffix –let, e.g. *anklet, armet, bracelet, chaplet, necklet, ringlet, leglet, etc.* The suffix –kin is not very productive, it can be found in nouns with appellative function, e.g. *devilkin, ladykin, napkin, barmkin, etc.* The suffix –ee is typical for a familiar and child language, e.g. *goatee, bootee, coatee, shirtee, etc.* The Anglo-Irish origin can be identified in the suffix –een, e.g. *girleen, squireen, buckee.* The most productive suffix is –ie.

It has appeared in

- a) the proper nouns in Scotland, e.g. *Lowrie, Perrie and Willie*
- b) the nouns denoting to people, e.g. *daddy, mammy, laddie, lassie, brownie, auntie, sonny, etc.*
- c) other nouns, e.g. *birdie, doggie, dolly*
- d) surnames in the closed-friend or closed-family circles, e.g. *Bozzy (Boswell), Boney (Bonaparte), Busy (Bismarck).*

Lančarič (p. 86, 2016) classifies suffixes according to their thematic sense. Diminutive suffixes are e.g. –let, -ling, -y/ie as in *leaflet, wolfling, ducky, birdie*). Next to the previous type, three other types fall into this category:

1. Agent forming suffixes
2. Appurtenance forming suffixes
3. Collectivity and status forming suffixes

According to Stoffa (p. 33, 2020) diminutives express:

1. Positive relationship of the subject towards the object, e.g. *dieťa → dieťaťko, srdce → srdiečko, víno → vínko, vínečko.* Diminutive first names belong to this category, too (hypocoristic nouns), e.g. *Hermína → Hermínka; Karína → Karínka; Kornel → Kornelko; Ladislav → Lacko.* One first name can form more diminutive forms, e.g. *Brigita → Brigitka, Gitka; Dezider → Dežko, Dežinko; Ján → Janko, Janík, Janičko; Michaela → Michaelka, Miška.*
2. Negative, often ironic relationship of the subject towards the object, e.g. *advokát → advokátik, asistent → asistentík, čeľaď → čeliadka, novinár → novinárík, papaláš → papalášik, redaktor → redaktorík; román → románik (krátkymilostný vzťah), strana (pol.) → stranička, úradník → úradníček, zlodej → zlodejíček.*
3. Small measure or small degree of the naming object, e.g. *ampula → ampulka, bomba → bombička, brožúra → brožúrka, bublina → bublinka, ceruza → ceruzka, clona → clonka, čln → člno, čiapka → čiapočka, dielňa → dielnička, doska → doštička, dutina → dutinka, forma → formička, formula → formulka, hák → háčik, hviezda → hviezdička, chyba → chybička, ikona → ikonka, jadro → jadierko, kryštál → kryštálik, lano → lanko, motor → motorček.*

According to Kumorová (2023) diminutives express the quantity, or a smaller dimension, a smaller measure of something or a certain attitude by means of typical diminutive suffixes (–ček, -ečko, -ík, -ok, -očka), e.g. *koník, gulôčka, kvietok.* The diminutives also express a quality, they can express a positive (*mamulienka, dcéruška*) or negative expressive attitude of the speaker - irony, e.g. *románik.* Some denotations, especially terms, are not felt in

the linguistic consciousness as diminutive or expressive, e.g. *hodinky, pilka, lyžička*. Dolník et al. (2010, pp. 421-6) discuss intensification of Slovak diminutives and link this function to smallness.

3. ELECTRONIC CORPORA IN THE ANALYSIS

For the purpose of the research Slovak – English Parallel Corpus (par-sken-all-4.0-sk, par-sken-all-4.0-en) was used. The research will focus on approaches used in traditional linguistics and their possible and effective relationship with the approaches within corpus linguistics. In the corpus linguistics, there are used mathematical, statistical methods that enable us to analyse precisely linguistic items. The article will concentrate on application of electronic corpora tools, i.e. average rate frequency, frequency, occurrence in linguistic analysis. The application of the electronic corpus in lexicology is as follows: The electronic corpus allows for the study of the behaviour of individual words, which is particularly useful for the compilation of dictionaries (lexicography). The electronic corpus is capable of identifying words that occur with greater or lesser frequency. In the field of grammar, we can examine a specific grammatical construction with restrictions on form or context. This allows us to investigate the factors that determine the choice of one construction over another, which may be influenced by factors such as lexis, grammar, or stylistic considerations. In the case of corpus linguistics, mathematical and statistical methods and tools are employed to analyse linguistic items with precision, to utilise large data sets and to quantify the typical features of linguistic items.

Lexical and grammatical studies frequently contrast usage by mode, domain, register, and so forth. The British National Corpus (BNC) has incorporated this information into header mark-up, for example, the term "lovely" is used more frequently by women than men.

Benko (2016:14) states that it is crucial to be aware of the spectrum of contexts encompassed by the corpus, which can be further illuminated by non-corpus research. A number of linguistic phenomena are revealed in a markedly different light.

The Slovak National Corpus (SNK) is employed for the purpose of conducting contrastive analysis. The electronic database contains Slovak language texts from 1955 onwards and encompasses a wide range of linguistic styles, genres, areas, regions, and so forth. The database comprises authentic Slovak texts, extended by addition of various linguistic information. It should be noted that the corpus is not intended to serve as a replacement for any existing linguistic reference books. The corpus contains a monolingual corpus of written texts, a manually morphologically annotated corpus, r-mak versions, a morphological database of the Slovak language, paradigms of the Slovak nouns, paradigms of the Slovak verbs, parallel corpora, a web corpus, a corpus of Slovak Wikipedia and Neczyklopédia. The corpus comprises specialised corpora, namely corpora of texts prior to 1955, a corpus of texts from 864 to 1843, and a corpus of texts from 1845 to 1954. The spoken part of the electronic corpus comprises the following elements: a corpus of spoken Slovak, a corpus of dialects, a historical Slovak corpus and a Slovak Terminology Database.

Slovak-English Parallel Corpus, the current version par-sken-4.0 contains 556 million tokens (261 million tokens in the Slovak half, 295 million tokens in the English one). The corpus consists of two parts: the subcorpus of fiction (200 million tokens – 92 million tokens in the Slovak part, 108 million tokens in the English part) and the subcorpus of freely available texts. To query the subcorpus of fiction, use NoSketch Engine for the English half, for the Slovak half. To access the whole corpus, use the web interface NoSketchEngine to query the Slovak texts or the English texts. Knowledge of NoSketch Engine and CQL is recommended. Slovak-English Parallel Corpus is a database containing texts for both Slovak and English language. Slovak texts are translated into English or vice versa. Texts are automatically aligned at sentence level. Slovak texts are automatically morphologically annotated by the tagger Morče which has been trained and tuned on tagset developed by the SNK. English texts are part-of-speech tagged with The Penn Treebank Tagset, using the TreeTagger software.

When we analyse language, we may distinguish between two main approaches: studies of structure and studies of use. In the field of traditional linguistics, linguistic analyses have predominantly focused on the structural aspects of language, with the objective of identifying the various structural units and classes that constitute the language.

Conversely, the corpus-based approach places greater emphasis on language use, specifically how speakers and writers make use of the resources available to them within their language. The corpus-based approach studies the actual language use in naturally occurring texts, which are characterised by an authentic character. The principles of traditional linguistics in combination with additional methods of corpus linguistics, such as those derived from mathematical and statistical approaches, have been applied.

The approach to language through corpora, as proposed by Sinclair (1991), has been applied. He approaches language as a probabilistic and statistically measurable system of preferences. The analysis commenced with qualitative, functional interpretations, based on quantitative methods, which comprise the description and characterisation of selected diminutive forms i.e. *mamička, dcérka, otecko, synček* in Slovak. By using Slovak – English parallel corpus, a list of English equivalents for a particular Slovak family member has been prepared.

Afterwards qualitative analysis has been applied. Thanks to data set, it was possible to detect the occurrence and frequency of the individual English equivalents in the corpus. Because the corpus of annotated according to genres, results have been examined across genre variations. This feature signals language in use. All results have been explained, exemplified and interpreted. The analysis concentrated on the authentic language at the level of parole.

4. ANALYSIS OF DIMINUTIVE FORMS IN ELECTRONIC CORPUS

For the analysis, Slovak-English Parallel Corpus (version par-sken-4.0) was used. It is possible to examine a selected entry in the Slovak and English section. The first analysis denotes to diminutive form *otecko*. In the English section of the corpus, the following equivalents are detected, e.g. *father, dad, daddy, papa and pa*. In order to create a diminutive form by morphological means, i.e. suffixation, only the example *daddy* fulfils this criterion if we consider the noun *dad* as a base. The origin of the word *dad* is not stated in the Collins Concise Dictionary concentrating on the history of the words. To express emotional closeness within the family relationships in the English language, even the borrowed words are used, e.g. *papa and pa*. The word *papa* was borrowed from French language in the 17th century, (Collins Concise Dictionary, 1999). Monosyllabic word *pa* expresses the closes and emotional relation by shortening. In Slovak the base form is modified the the suffix, in English specific words express emotional closeness per se. Moreover, in such intimate family relationships, borrowed words penetrated into these field in English.

Table 1 Otecko and Diminutive forms from the Slovak-English Parallel Corpus

| Diminutive form in Slovak | Equivalent of diminutive form in English |
|---------------------------|--|
| otecko | father |
| | dad |
| | daddy |
| | papa |
| | pa |

Source: Slovak-English Parallel Corpus (version par-sken-4.0)

The examined Slovak diminutive form *synček* was represented by seven equivalents, i.e. *son, little son, my son, little boy, little one, baby boy and dear boy*. The equivalent *son* is more appropriate for the base form *syn* in the Slovak language. For the rest of the examples, a masculine gender is transparent in six cases by the nouns *son* or *boy*. A diminutive meaning is not expressed by suffixation at all in the selected corpus but by premodifiers describing emotional closeness, e.g. *little, baby or dear*.

Table 2 Synček and Diminutive forms from the Slovak-English Parallel Corpus

| Diminutive form in Slovak | Equivalent of diminutive form in English |
|---------------------------|--|
| synček | son |
| | little son |
| | my son |
| | little boy |
| | little one |
| | baby boy |
| | dear boy |

Source: Slovak-English Parallel Corpus (version par-sken-4.0)

The Slovak diminutive form *mamička* is represented by six equivalents in the English section of the corpus. The equivalent mother does not express the diminutive form. The suffixes *-ie* and *-y* are productive in the examples *mummie, mommy*. In all diminutive equivalents the initial letter is identical with the base form, letter *m*. The origin is not stated in Collins Concise Dictionary (1999). The shortening is applied in the case of *ma*. No premodifiers are detected to express emotional closeness.

Table 3 Mamička and Diminutive forms from the Slovak-English Parallel Corpus

| Diminutive form in Slovak | Equivalent of diminutive form in English |
|---------------------------|--|
| mamička | mother |
| | mummie |
| | mum |

| | |
|--|-------|
| | mommy |
| | mamma |
| | ma |
| | |

Source: Slovak-English Parallel Corpus (version par-sken-4.0)

English equivalents expressing *dcérka* in English are as follows *daughter*, *little girl*, *baby girl*, *little daughter* and *baby daughter*. As it is seen in the previous analyses, the base form of the word is again recurrent although it does not express the emotional closeness. In the rest of the equivalents premodification is present. There are adjectives with the expressive value, e.g. *little* and *baby*.

Table 4 Dcérka and Diminutive forms from the Slovak-English Parallel Corpus

| Diminutive form in Slovak | Equivalent of diminutive form in English |
|---------------------------|--|
| dcérka | daughter |
| | little girl |
| | baby girl |
| | little daughter |
| | baby daughter |

Source: Slovak-English Parallel Corpus (version par-sken-4.0)

5. CONCLUSION

A large part of every language, the frequently recurring patterns, which we might call socially salient, is conventional and for that very reason, unmemorable. As the lexicon is concerned, social salience (on the form of frequency of use) and cognitive salience (in the form of ease of recall) are independent variables (Hanks 2013).

Slovak language represents predominantly a synthetic type of the languages. It is significant even in the forming of diminutive forms. On the other hand, according to the results of the study it can be seen that the ways in forming diminutive forms are not straightforward in English. It belongs to the analytic type of the languages and the field of diminutive word formation proves this fact. According to the results, independent words were used to create the opposition to the neutral form of the word. i.e. *son*-neutral form, *little son*, *little boy*, *baby boy* – diminutive forms.

English is less expressive in comparison to Slovak language. It uses different ways next to suffixes how to demonstrate expressive features. All in all, the base form of the word is recurrent although it does not express such emotional closeness in comparison to particular suffixes of using premodifiers carrying the emotional value. Theoretical background on the level langue can be found in the theoretical studies on lexicology. Language use is vitally important and that is the reason why it is important to examine the analysed diminutive words in the electronic corpora. Thanks to large data set, the analysis of selected diminutive words in English and Slovak is more valuable.

REFERENCES

- Benko, V. et al. (2016) *Webové korpusy Aranea*, Univerzita Komenského v Bratislave. Bratislava.
- Biber, D. et al. (1998) *Corpus Linguistics*, Cambridge University Press, Cambridge.
- Cruse, D. A. (1986) *Lexical Semantics*, Cambridge: Cambridge University Press. Oxford
- Collins Concise Dictionary. (1999) Harper Collins Publishers, Glasgow
- Crystal, D. (2022) *The Cambridge Encyclopedia of the English Language*, 3rd Edition, Cambridge University Press, Cambridge.
- Cambridge Advanced Learner's Dictionary. (2007) Cambridge University Press, Cambridge.
- Dolník, J., et al. (2010). *Morfologické aspekty súčasnej slovenčiny*. Bratislava. VEDA.
- Dressler, W., Mattiello, E. & Ritt-Benmimoun, V. (2024). 14 Morphological richness and priority of pragmatics over semantics in Italian, Arabic, German and English diminutives. In S. Manova, L. Grestenberger & K. Korecky-
- Kröll (Ed.), *Diminutives across Languages, Theoretical Frameworks and Linguistic Domains* (pp. 335-362). Berlin, Boston: De Gruyter Mouton. <https://doi.org/10.1515/9783110792874-014>.
- Ginzburg. R.S. (1966) *A course in Modern English Lexicology*. Higher School Publishing House. Moscow.
- Hanks, P. (2013) *Lexical analysis: norms and exploitations*, MIT Press. Massachusetts.
- Parallel English – Slovak corpus (par – sken – all – 4.0. en) <https://bonito.korpus.sk/> (accessed: 21.11. 2024).

- Kumorová, Z. (2023) Lexikológia slovenského jazyka v učiteľskej praxi. Univerzita sv. Cyrila a Metoda v Trnave, ISBN 978-80-572-0312-4.
- Kvetko, P. (2021) Translation Studies, UCM, Trnava.
- Lančarič, D. (2016) English Lexicology. Theory and Exercises. Kirsch Verlag. Nümbrecht.
- Slovak National Corpus (bonito.korpus.sk) <https://bonito.korpus.sk/> (accessed: 21.11. 2024).
- SNK Slovenský národný korpus. Slovak <https://korpus.sk/korpusy-a-databazy/korpusy-snk/> (accessed: 21.11. 2024).
- Sinclair, J. (1991) Corpus, Concordance, Collocation. Oxford University Press, Oxford.
- Stoffa, J. (2020) Problémové slovenské deminutíva v edukačnej praxi. DOI: 10.5507/tvv.2020.015 In: Trendy ve vzdělávání 2020, roč.13, číslo 2, Univerzita Palackého, Olomouc.
- Štekauer, P. (2000) Rudiments of the English Language. Slovacontact. Prešov ISBN 80 88876 04 4
- Trnka, B. (2014) Rozbor nynejší spisovné angličtiny. Karolinum. Praha.