

Statistical Detection of Diminutives in Czech

Věra Schmiedtová¹

1. The aim of the paper

The aim of this paper is to show the strongly pragmatic nature of Czech diminutive nouns by exploring their adjectival collocates. The starting point of the following analysis is the identification of adjectives with pragmatic meaning which occur in the environment of diminutives most frequently. The next task is to find to what extent they participate in the total frequency of diminutives (i.e., to establish how many adjectives are involved, and what is the sum of their frequencies) and to compare the situation in diminutive and non-diminutive nouns.

2. Diminutives in Czech compared with other languages

Compared to English, diminutives in Czech are a relatively frequent phenomenon. They are mostly formed by suffixes which include the consonant *č*, specifically *-ček* (masculine nouns), *-čka* (feminine nouns), and *-čko* (neuters). However, a number of Czech diminutives are formed by a polysemous suffix without the consonant *č*, such as *-ek / -nek*, e.g. in *obrázek, tatínek* (partly *-ík*, as in *vozik, pytlík*), or *-ka, -ko*. These suffixes without the consonant *č* need not primarily signal a diminutive status – they can be abstracts, e.g. *důsledek* (consequence), *otázka* (question), *spánek* (sleep), feminised nouns, e.g. *ředitelka* (directress), names of objects, e.g. *šálek* (cup), *zkumavka* (test tube), etc.

In English these diminutive suffixes have their counterpart in the suffix *-ie, -let* etc., e.g. *birdie, doggie, rivulet*, in German in the suffixes *-chen, -lein*, e.g. *Vögelchen, Mädchen, Väterlein*, in Dutch in *-je, -tje*, e.g. *huisje, vogeltje*, in French in *-ette*, e.g. *maisonette, filette*, etc.

3. The sample of the most frequent diminutives in contemporary Czech

In contemporary Czech the range of the most frequent diminutives is largely stabilised as is shown by the comparison of the most frequent diminutives according to gender in the two 100-million-word Czech corpora – SYN2000 and SYN2005 (for more detail on frequencies in both corpora see Appendix 1):

3.1. Masculine nouns

Among the masculine nouns twelve of the fifteen most frequent diminutives are the same in both corpora: *balíček, dědeček, háček, chlapeček, koníček, krůček, míček*,

¹ Institute of the Czech National Corpus, Charles University, Prague
e-mail: vera.schmiedtova@ff.cuni.cz

ptáček, rámeček, sáček, strýček, žebříček ('packet', 'grandfather', 'boy', 'horse', 'step', 'ball', 'bird', 'frame', 'bag', 'uncle', 'ladder'²).

3.2. Feminine nouns

All ten of the most frequent diminutives of the feminine gender in both corpora coincide: *babička, písnička, lžička, ulička, holčička, svíčka, lavička, sklenička, krabička, hlavička* ('grandmother', 'song', 'spoon', 'street', 'girl', 'candle', 'bench', 'glass', 'box', 'head'³).

3.3. Neuters

In the group of neuters once again all ten of the most frequent diminutive nouns in both corpora are the same: *městečko, vajíčko, kolečko, víčko, tričko, sluníčko, očko, slovíčko, poličko, srdíčko* ('town', 'egg', 'wheel' / 'ring', 'cover'⁴, 't-shirt', 'sun', 'eye', 'word', 'field', 'heart').

4. The relationship between diminutive and non-diminutive nouns

If we compare diminutive nouns and their derivational bases in the corpus SYN2005 (e.g. *městečko* – *město*), i.e. in written language, we find that the more frequent of the two are non-diminutive nouns. We may show this by comparing the most frequent diminutive nouns in the corpus SYN2005 and their non-diminutive counterparts (for a more detailed overview see Tables 4, 5 and 6, Appendix 2). For the sake of simplicity we shall compare the most frequent representatives of each gender, i.e. *městečko* (n.), *písnička* (f.) and *koníček* (m.). For instance, the non-diminutive noun *město* (town) is 15.7 times more frequent than the diminutive *městečko*, the noun *píseň* (song) is 2.2 times more frequent than the diminutive *písnička*, and the noun *kůň* (horse) is 10.6 times more frequent than the corresponding diminutive *koníček*.

In nouns where the frequency ratio is reversed, i.e. the diminutive nouns is the more frequent one in the pair, the diminutive is semantically different from its non-diminutive counterpart. Again we shall focus on the three most frequent representatives: The noun *balík* (packet, package) is used terminologically in one fifth of its occurrences in contemporary Czech – e.g. *balík akcií* (a block of shares), *etc.*, while the diminutive *balíček* is mostly used in the essential meaning (packet). The noun *svíčka* (candle) compared to the underived *svíce* is used in a number of terminological expressions, especially *zapalovací svíčky* (spark plugs), while *svíce* has again only the essential meaning (candle). The nouns *tričko* and *triko* (t-shirt) are interchangeable in contemporary Czech, though the neutral term in this pair tends to be the diminutive *tričko*, whereas the less frequent *triko* is found predominantly in phraseology - *honit si triko* (blow one's own trumpet) a *vzít (si) něco na své/svý triko* (to do sth. on one's own responsibility).

By contrast, in spoken language diminutives often tend to be more frequent than the initial non-diminutive nouns. This applies especially to objects (the

² Special diminutive meaning of the diminutive *žebříček* is 'ranking list'.

³ Special diminutive meaning of the diminutive *hlavička* is 'heading'.

⁴ Special diminutive meaning of the diminutive *víčko* is 'eyelid'.

frequencies found in the one-million-word corpus of spoken Czech ORAL2006 are given in Appendix 2). Once again, we shall focus on the three most frequent representatives in each gender: the diminutive *knižka* (f.) is 3.5 times more frequent than the corresponding noun *kniha* (book), the diminutive *vajíčko* (n.) is 3.4 times more frequent than the noun *vejce* (egg), and the diminutive *dárek* is even 7 times more frequent than the noun *dar* (gift).

As regards names of persons, the situation in spoken language is the opposite, compare the following nouns: ('grandfather', 'girl', 'baby')

Deminutive	Frequency
<i>dědeček</i>	118
<i>holčička</i>	51
<i>miminko</i>	28

Underived noun	Frequency
<i>děda</i>	198
<i>holka</i>	463
<i>mimino</i>	20

Table 1.

In written language (the SYN2005 corpus) the relationship between diminutive and non-diminutive nouns is such that the non-diminutive nouns are usually more frequent:

('grandmother', 'grandfather', 'girl', 'uncle', 'boy', 'aunt'):

Deminutive	Frequency
<i>babička</i>	6505
<i>dědeček</i>	3738
<i>holčička</i>	3722
<i>strýček</i>	2502
<i>chlapeček</i>	1201
<i>tetička</i>	907

Underived noun	Frequency
<i>bába</i>	930
<i>děda</i>	2495
<i>holka</i>	7683
<i>strýc</i>	2996
<i>chlapec</i>	15714
<i>teta</i>	4830

Table 2.

Comparison of the six most frequent diminutive nouns designating persons with their underived counterparts clearly shows that the latter, i.e. non-diminutive nouns, are the more frequent of the pairs. The only exception is the pair *holčička* – *holka* which will be dealt with in section 6.2.

5. The meaning of diminutives

Although the term “diminutives” suggests that they are names for small (or smaller than usual) entities, objects, and persons, and they are indeed interpreted accordingly in Czech grammars and dictionaries, the corpus data show that the use of a diminutive in contemporary Czech is primarily a signal for the designation of entities, objects, and persons to which the speakers have a positive relationship, which means diminutives have primarily positive pragmatic meaning. This is attested to not only by the fact that diminutives have a higher frequency in spoken language, but also by their having positive connotations, hence also positive contexts, cf. *pivečko*, *pivíčko*, or *pivko* – all referring to good beer served in a 0.5 l glasses - as in *miluje pivečko* (he loves beer), *dát si pivečko* (have a glass of beer) in contrast to *malé pivo* (small beer), meaning (1) a 0.3 l glass of beer, (2) a small person.

The evaluative function (for details see Čermák, 1995) may concentrate on (A) sensory values, which can be both positive and negative (*noisy, bright*), or on (B) intellectual values which can be (Ba) moral (*pleasant, good*), (Bb) aesthetic-emotional (*nice, appealing*), (Bc) utilitarian (*suitable, handy, good*) and (Bd) universal (*good, horrible*).

In written texts we find that diminutive nouns tend to have collocates with positive meaning. The present paper will focus on evaluative adjectives such as *krásný* (beautiful), *pěkný* (nice, pretty), *dobrý* (good), *milý* (pleasing), *etc.*, which appear within the span of -3, 3 words from KWIC. The set of positively evaluative adjectives in the environment of diminutives which are clearly marked as such by the consonant *č* in the masculine suffix *-ček*, the feminine suffix *-čka* and the neutral *-čko* found in the SYN2005 are presented in Appendix 3.

The list of the commonest evaluative adjectives occurring before diminutives regardless of gender is the following :

(‘nice’, ‘pleasing’, ‘soft’, ‘beautiful’, ‘dear’, ‘good’, ‘nice’, ‘charming’, ‘sweet’)

<i>pěkný</i>	208
<i>milý</i>	190
<i>jemný</i>	189
<i>krásný</i>	189
<i>drahý</i>	167
<i>dobrý</i>	137
<i>hezký</i>	136
<i>roztomilý</i>	98
<i>sladký</i>	74

Table 3.

6. Comparison of positively evaluative adjectival collocates in the most frequent diminutives and their non-diminutive counterparts

6.1. The set of evaluative adjectives

The following analysis will concentrate on the ten most frequent diminutive nouns found in the SYN2005 corpus in order to establish the percentage of evaluative and other adjectives with positive meaning co-occurring with them. The analysis will be based on a list of 160 evaluative adjectives (see Appendix 4).

6.2. The ten most frequent diminutive nouns in the SYN2005 corpus

The first column in Table 4 gives the most frequent Czech diminutives found in the 2005 corpus with their non-diminutive counterparts below. The second column gives the frequency of these nouns. The third column presents the number of evaluative adjectives which occur around these diminutives with the sum of the frequencies of these adjectives following the slash. For example: if the diminutive noun is preceded by the adjectives *hezký*, *pěkný* and *krásný*, with their respective frequencies 10, 8 and 15, the entry will be 3 / 33. The fourth column shows the proportion of the total frequency of the evaluative adjectives (i.e. the figure 33 in the above example) in the frequency of the diminutive noun – if the noun has the frequency of 1000 occurrences

this proportion would be 3.3 percent. The last two columns describe the numbers of multiword terms and idioms in which the given diminutive, or non-diminutive, noun occurs; the last but one section after the slash once again gives the sum of all frequencies, the last column provides the proportion of terms and idioms in the frequency of the noun. (Compared diminutives and their underived opposites: ‘grandmother’, ‘town’, ‘song’, ‘street’, ‘grandfather’, ‘girl’, ‘packet’, ‘egg’, ‘uncle’, ‘candle’)

Noun	Frequency in SYN2005	Number of eval.adj./freq.	Percent	Number of terms and idioms/freq.	Percent
<i>babička</i>	6505	33 / 127	1.9	1 / 24	0.3
<i>bába</i>	930	5 / 5	0.5	4 / 235	25.0
<i>městečko</i>	4990	16 / 150	3.0	0 / 0	0
<i>město</i>	78470	5 / 516	0.6	1 / 6148	7.8
<i>písnička</i>	4520	14 / 241	5.3	2 / 39	0.8
<i>píseň</i>	10005	10 / 414	4.1	0 / 0	0
<i>ulička</i>	4063	10 / 69	1.7	2 / 546	13.4
<i>ulice</i>	33630	9 / 223	0.6	0 / 0	0
<i>dědeček</i>	3738	8 / 72	1.9	0 / 0	0
<i>děda</i>	2495	3 / 30	1.2	0 / 0	0
<i>holčička</i>	3722	16 / 211	5.6	0 / 0	0
<i>holka</i>	7683	20 / 570	7.4	2 / 40	0.5
<i>balíček</i>	3111	9 / 39	1.25	2 / 42	1.3
<i>balík</i>	2814	3 ⁵ / 12	0.4	2 / 212	7.5
<i>vajíčko</i>	3074	12 / 60	2.0	2 / 492	16.0
<i>vejce</i>	6435	6 / 41	0.6	3 / 139	2.1
<i>strýček</i>	2502	8 / 105	4.1	1 / 17	0.6
<i>strýc</i>	2996	7 / 35	1.1	0 / 0	0
<i>svíčka</i>	2741	4 / 22	0.8	1 / 61	2.2
<i>svíce</i>	880	3 / 6	0.6	0 / 0	0

Table 4.

A few notes on the above table.

Of primary importance is the comparison of the proportional representation of evaluative adjectives among the collocates of the nouns under study. As is clear from the fourth column of the table, the diminutive has usually a higher proportion of evaluative adjective than the non-diminutive noun. The only exception is the pair *holčička* – *holka* (girl) where the noun *holka* has 7.4 percent of evaluative collocates, while *holčička* only 5.6 percent. This is due to the fact that the noun *holka* is used in Czech not only for an adolescent girl, but also for a young woman and for this reason the collocates more frequently accentuate feminine charm, while in the diminutive *holčička* roughly eighteen percent of collocates denote immaturity (*-měsíční*, *-letá holčička* / -month-old, -year-old girl, *novorozená holčička* / new-born girl, etc.).

Now in detail to the three most frequent diminutives and their counterparts in each gender, namely *babička* – *bába* (f.), *městečko* – *město* (n.) and *dědeček* – *děda* (m.):

***babička* – *bába*:**

In the contextual surroundings of the diminutive *babička* (grandma) the search yielded the following evaluative adjectives: *bohatý* (rich, one occurrence), *čilý*

⁵ The adjectives *pěkný*, *slušný* and *pořádný* were not counted (the meaning is ‘big’ or ‘great’ in this case).

(sprightly, two occurrences), *dobrý* (good, twenty-four occurrences), *drahý* (dear, five occurrences), *hezký* (pretty, one occurrence), *hodný* (good, kind, sixteen occurrences), *chytrý* (clever, one occurrence), *ideální* (ideal, one occurrence), *jedinečný* (unique, one occurrence), *kouzelný* (charming, six occurrences), *krásný* (beautiful, three occurrences), *laskavý* (kind, six occurrences), *milý* (pleasing, fourteen occurrences), *moudrý* (wise, one occurrence), *nádherný* (wonderful, five occurrences), *oblíbený* (popular, one occurrence), *pečlivý* (solicitous, one occurrence), *pěkný* (nice, one occurrence), *pravý* (real, seven occurrences), *proslulý* (famous, one occurrence), *prozíravý* (far-sighted, one occurrence), *roztomilý* (charming, one occurrence), *rozumný* (sensible, one occurrence), *skvělý* (perfect, one occurrence), *slavný* (renowned, six occurrences), *starostlivý* (considerate, two occurrences), *šťastný* (happy, eleven occurrences), *štedrý* (open-handed, one occurrence), *vděčný* (grateful, one occurrence), *veselý* (cheerful, one occurrence), *vynikající* (excellent, one occurrence), *zábavný* (entertaining, one occurrence), *zdatný* (fit, one occurrence). The total is thirty-three adjectives with frequency sum of 127. This noun is also part of the multiword term *porodní babička* (midwife), with a frequency of twenty-four occurrences. The non-diminutive noun *bába* (grandmother, old woman) has a strongly negative pragmatic connotation in Czech, cf. *stará bába* (hag, seventy-five occurrences), positively evaluative adjectives tend to have single occurrences: *opatrný* (careful), *skromný* (modest), *šikovný* (handy), *štedrý* (open-handed), *zbožný* (religious). Next, the word *bába* is used in the term *porodní bába* (176 occurrences, midwife) and in the idiom *hra na slepou bábu* (thirty-seven occurrences, blind man's buff), *svíčková bába* (fourteen occurrences, churchy woman) and *bába kořenářka* (eight occurrences, herbwoman).

městečko – město:

Adjectives of a positively evaluative nature which occur in the environment of the diminutive *městečko* include: *hezký* (pretty, eleven occurrences), *klidný* (calm, sixteen occurrences), *kouzelný* (charming, five occurrences), *krásný* (beautiful, eleven occurrences), *milý* (pleasant, eight occurrences), *moderní* (modern, eight occurrences), *nádherný* (marvellous, three occurrences), *nenápadný* (quiet, four occurrences), *pěkný* (nice, sixteen occurrences), *proslulý* (famous, eight occurrences), *půvabný* (lovely, twenty-one occurrences), *skromný* (modest, three occurrences), *slavný* (famous, four occurrences), *tradiční* (traditional, five occurrences), *typický* (typical, fourteen occurrences), *zajímavý* (interesting, thirteen occurrences). The diminutive *městečko* is not part of any terms or idioms. The non-diminutive noun *město* has the following evaluative adjectival collocates: *krásný* (beautiful, 260 occurrences), *moderní* (modern, 173 occurrences), *ideální* (ideal, forty-one occurrences), *půvabný* (lovely, twenty-three occurrences), *význačný* (outstanding, seventeen occurrences). The most frequent collocation of the noun *město* is *hlavní město* (6148 occurrences, capital).

dědeček – děda:

The diminutive *dědeček* (grandpa) co-occurs with the following evaluative adjectives: *dobrý* (good, nineteen occurrences), *hodný* (kind, fourteen occurrences), *slavný* (famous, thirteen occurrences), *milý* (pleasing, eight occurrences), *pravý* (real, six occurrences), *bohatý* (rich, four occurrences), *moudrý* (wise, five occurrences) a *štedrý* (open-handed, three occurrences). The non-diminutive noun *děda* has only three positively evaluative adjective in its contextual surroundings, namely *dobrý* (good, twenty-four occurrences), *milý* (pleasing, three occurrences), and *tichý* (quiet, three occurrences). The diminutive *dědeček* has the same proportion of evaluative adjectives as the diminutive *babička* (grandma) – 1.9 percent, but the non-diminutive noun *děda* has a much more positive meaning than *bába* (1.2 percent and 0.5 percent respectively).

6.3 Comparison of the corpora SYN2005 and SYN2000

The following table gives only adjectives with positive meaning, including their proportion in the overall frequency of the noun, in the above most frequent diminutives and their non-diminutive counterparts, as they occurred in both representative 100-million-word Czech corpora. The procedure was identical for each corpus:

Noun	Corpus	Frequency	Number of eval. adj. / frequency	Percentage
<i>babička</i>	SYN2005	6505	33 / 127	1.9
	SYN2000	4465	17 / 46	1.0
<i>bába</i>	SYN2005	930	5 / 5	0.5
	SYN2000	577	2 / 2	0.3
<i>městečko</i>	SYN2005	4990	16 / 150	3.0
	SYN2000	3613	12 / 34	0.9
<i>město</i>	SYN2005	78470	5 / 516	0.6
	SYN2000	82798	5 / 197	0.2
<i>písnička</i>	SYN2005	4520	14 / 241	5.3
	SYN2000	5154	14 / 129	2.5
<i>píseň</i>	SYN2005	10005	10 / 414	4.1
	SYN2000	8429	10 / 189	2.2
<i>ulička</i>	SYN2005	4063	10 / 69	1.7
	SYN2000	2262	3 / 14	0.6
<i>ulice</i>	SYN2005	33630	9 / 223	0.6
	SYN2000	30749	5 / 61	0.2
<i>dědeček</i>	SYN2005	3738	8 / 72	1.9
	SYN2000	3037	8 / 31	1.0
<i>děda</i>	SYN2005	2495	3 / 30	1.2
	SYN2000	954	7 / 7	0.7
<i>holčička</i>	SYN2005	3722	16 / 211	5.6
	SYN2000	1723	14 / 76	4.4
<i>holka</i>	SYN2005	7683	20 / 570	7.4
	SYN2000	4935	22 / 294	5.9
<i>balíček</i>	SYN2005	3111	9 / 39	1.3
	SYN2000	2267	4 / 6	0.2
<i>balík</i>	SYN2005	2814	3 / 12	0.4
	SYN2000	3236	3 / 5	0.1
<i>vajíčko</i>	SYN2005	3074	12 / 60	2.0
	SYN2000	1908	4 / 17	0.9
<i>vejce</i>	SYN2005	6435	6 / 41	0.6
	SYN2000	4585	6 / 33	0.7
<i>strýček</i>	SYN2005	2502	8 / 105	4.1
	SYN2000	1034	13 / 58	5.6
<i>strýc</i>	SYN2005	2996	7 / 35	1.1
	SYN2000	1869	6 / 28	1.5
<i>svíčka</i>	SYN2005	2741	4 / 22	0.8
	SYN2000	1782	2 / 6	0.3
<i>svíce</i>	SYN2005	880	3 / 6	0.6
	SYN2000	639	1 / 1	0.1

Table 5.

Although the two corpora have a different proportional representation of text types⁶, the basic difference when comparing the diminutive and non-diminutive nouns does not change: the non-diminutive noun always has a much smaller percentage of evaluative and other positive adjectival collocates in its context.

7. Comparison of the positively evaluative adjectival collocates in diminutives in the frequency zone of 1000-900 occurrences with their non-diminutive counterparts

In this table the Czech nouns and their diminutive counterparts corresponding to English nouns ‘river’, ‘while’, ‘key’, ‘step’, ‘lid’, ‘noodle’, ‘hand’, ‘branch’, ‘aunt’ and ‘fish’ will be studied:

Noun	Frequency	Number of eval. adj./freq.	Percentage	Number of terms & idioms/freq.	Percentage
<i>řička</i>	985	3 / 13	1.3	0 / 0	0
<i>řeka</i>	18815	3 / 153	0.8	0 / 0	0
<i>chvilíčka</i>	982	0 / 0 ⁷	0	0 / 0	0
<i>chvíle</i>	54243	15 / 1132	2.0	20 / 6768	12.5
<i>klíček</i>	972	2 / 5	0.5	2 / 198	20.0
<i>klíč</i>	7020	1 / 7	0.1	4 / 1743	24.7
<i>krůček</i>	971	10 / 37	3.8	3 / 215	22.0
<i>krok</i>	21564	13 / 349	1.6	5 / 1353	6.2
<i>poklička</i>	950	6 / 57	6.0	1 / 32	3.3
<i>poklice</i>	234	2 / 2	0.8	0 / 0	0
<i>nudlička</i>	943	5 / 45	4.7	0 / 0	0
<i>nudle</i>	809	4 / 16	1.9	5 / 15	1.8
<i>ručička</i>	937	9 / 29	3.0	1 / 483	51.5⁸
<i>ruka</i>	85181	11 / 809	0.9	45 / 6893	8.1
<i>větrvička</i>	934	2 / 11	1.1	0 / 0	0
<i>větev</i>	5383	6 / 23	0.4	1 / 17	0.3
<i>tetička</i>	907	6 / 16	1.7	0 / 0	0
<i>teta</i>	4830	10 / 74	1.5	0 / 0	0
<i>rybička</i>	898	12 / 28	3.1	2 / 114	12.6
<i>ryba</i>	9202	10 / 84	0.9	11 / 413	4.5

Table 6.

A note on the above table:

It is clear that compared to the previous table the diminutives from this frequency zone are far more often used as parts of terms and idioms – see the last column. Diminutive nouns are also far more semantically differentiated from their non-diminutive counterparts than was the case in the group of the most frequent diminutives.

⁶ Distribution of main text types in the SYN2005 corpus: imaginative literature forty-seven percent, scientific literature twenty-seven percent and newspapers thirty-three percent; distribution of the main text types in the SYN2000 corpus: fifteen percent - twenty-five percent - sixty percent

⁷ The diminutive *chvilíčka* je synonymous with the word *okamžik* (moment); no evaluative adjectives occur in its environment, only particles such as *jenom*, *jen*, *ještě*, *teprve*, e.g. *jen na chvilíčku*, etc.

⁸ It is used in the sense ‘an indicator of time (on the watch)’, or ‘an indicator of physical values (on measuring devices)’

Inasmuch as this frequency zone does not include neuter noun, we shall confine ourselves to only two representatives:

říčka – řeka:

The collocates of the diminutive *říčka* (rivulet) include evaluative adjectives *bystrý* (fast-flowing, three occurrences), *klidný* (calm, eight occurrences) a *kouzelný* (fascinating, two occurrences). The non-diminutive noun *řeka* (river) has similar adjectival collocates, namely *bystrý* (fast-flowing, twenty-two occurrences), *klidný* (calm, thirty-three occurrences), and *čistý* (clear, fifty-nine occurrences).

klíček – klíč:

This pair exhibits a relatively high proportion of terms and idioms, therefore even the number and frequency of evaluative adjectives is much smaller, cf. two evaluative adjectives occurring with the diminutive *klíček* – *jemný* (soft, three occurrences), *lehký* (light, two occurrences), while the non-diminutive noun *klíč* (key) has only one – *přesný* (accurate, seven occurrences). The noun *klíček* most commonly appears in the collocation *klíčky od auta / vozu* (car keys), next in the meaning ‘collarbone’ (or clavicle) and ‘shoot of a plant’ (or sprout), whereas the non-diminutive noun *klíč* is used in musical terminology, as in *houslový klíč* (the G/treble clef), etc., then in the sense ‘a means providing access to or understanding of something’ (cipher key, clue), or as part of the name for a tool, e.g. in the collocation *francouzský klíč* (monkey wrench); in phraseology it is a component of the idioms *být klíčem k něčemu* (be the key to, be the secret of) and *udělat něco na klíč* (deliver for immediate occupancy, provide a turnkey service).

8. Diminutives in the spoken language corpus ORAL2006

As was mentioned above, spoken language prefers diminutives to non-diminutive nouns, especially as regards names describing objects of daily use and food. From the total of 108 diminutive nouns 60 diminutives are formed with a suffix including č, i.e., *-ček, -čka, -čko*, in some of the nouns an abbreviated diminutive form is the more frequent one, cf. *taťka, hod'ka, pivko* instead of *tatínek, hodinka, pivečko / pivíčko* (dad, hour, beer). It is also characteristic of spoken Czech that the pronouns *nějaký* (some) and *ten* (this, that) are used with greater frequency, which is also evident in their collocability with diminutives, for instance, *ňáká / nějaká / ta knížka* (some / this / that book). In some cases even spoken Czech uses evaluative adjectives, as in *krásný miminko, krásný tělíčko, krásný tričko, pěkněj / hezkej dárek* (beautiful baby, beautiful body, beautiful t-shirt, nice present) etc. Some diminutives are used only in a specific sense, e.g. *prstýnek* – ‘snubní prstýnek’ (wedding ring), *rohlíček* – ‘vanilkový rohlíček’ (vanilla Christmas pastry), *stroměček* – ‘vánoční stroměček’ (Christmas tree), *stehýnko* – ‘kuřecí stehýnko’ (chicken leg) etc.

9. Conclusions

The paper strives to show that the collocational range of diminutives in Czech involves a high proportion of evaluative adjectival collocates, i.e. that Czech diminutives exhibit a pronounced pragmatic meaning. The analysis arrived at the following preliminary results:

1) Diminutives are typical especially of spoken language, therefore in the spoken language they are more frequent than their non-diminutive counterparts. The only

exception is kinship terms in which the spoken language prefers abbreviated forms without the diminutive suffix (*cf.* the more frequent *děda* as against *dědeček*).

2) Diminutives in the written language are less frequent than their non-diminutive counterparts – the only exceptions are pairs which have become semantically differentiated, e.g. *balíček*, *svíčka*, *tričko* (packet, candle, t-shirt).

3) The proportion of frequencies of evaluative adjectival collocates in the overall frequency of a diminutive noun is always higher than in the frequency of a non-diminutive noun – exceptions are pairs of the type *holčička* – *holka* (girl), where the higher proportion of the positively evaluative collocates is higher due to the accentuation of feminine charm; this relationship was confirmed in both 100-million-word corpora of Czech.

4) The stock of the most frequent diminutives in the written language is stable in Czech as the search of both the 100-million-word Czech corpora has shown.

5) Diminutives with lower frequencies are far more often used as components of multi-word terms and idioms, *cf.* *klíček* (collarbone, sprout, 20 percent) and *ručička* (second / minute / hour hand, 51.5 percent).

The same testing procedure can be applied to pairs where the diminutive does not include the explicit signal of diminutiveness, as in *stránka* – *strana* (page), *maminka* – *máma* (mum), *obrázek* – *obraz* (picture), *tatínek* – *táta* (dad), *okénko* – *okno* (window), *jezířko* – *jezero* (lake), *etc.*

References

- Čermák, F. (1993) Spoken Czech, in E. Eckert (ed.) Varieties of Czech. Studies in Czech Sociolinguistics, pp. 27–41. Amsterdam/Atlanta: Rodopi.
- Čermák, F. (1995) Functional System and Evaluation, in E. Hajičová, M. Červenka, O. Leška, P. Sgall (eds) Prague Linguistic Circle Papers, pp. 73–84. Amsterdam-Philadelphia: John Benjamins Publishing Company.
- Čermák, F. (2006) Polysémie a kolokace: případ adjektiva *měkký* [Polysemy and collocation: the case of the adjective *měkký/soft*], in F. Čermák, M. Šulc (eds) Studie z korpusové lingvistiky 2. Kolokace. [Studies in Corpus Linguistics 2. Collocations] Prague: Nakladatelství Lidové noviny.
- Čermák, F., J. Holub (2005) Syntagmatika a paradigmatica českého slova. Valence a kolokabilita. [Syntagmatics and paradigmatics of the Czech word. Valency and collocability] Prague: Nakladatelství Karolinum.
- Slovník české frazeologie a idiomatiky II. Výrazy neslovesné (1988) [Dictionary of Czech phraseology and idiomatics. Non-verbal expressions]. F. Čermák, J. Hronek, J. Machač (eds).
- Slovník české frazeologie a idiomatiky III. Výrazy slovesné (1994) [Dictionary of Czech phraseology and idiomatics. Verbal expressions]. F. Čermák, J. Holub, J. Hronek (eds).
- Czech National Corpus – SYN2005 (2005), Institute of the Czech national Corpus, Faculty of Arts, Charles University, Prague. Available on-line from <http://ucnk.ff.cuni.cz>.
- Czech National Corpus – SYN2000 (2000), Institute of the Czech national Corpus, Faculty of Arts, Charles University, Prague. Available on-line from <http://ucnk.ff.cuni.cz>.

Czech National Corpus – ORAL2006 (2006), Institute of the Czech national Corpus,
Faculty of Arts, Charles University, Prague. Available on-line from
<http://ucnk.ff.cuni.cz>.

Appendix 1: The most frequent diminutives

1) masculine diminutives

SYN2005		freq.
<i>dědeček</i>	‘grandfather’	3738
<i>balíček</i>	‘packet’	3111
<i>strýček</i>	‘uncle’	2502
<i>žebříček</i>	‘ladder’ / ‘ranking list’	1929
<i>háček</i>	‘hook’	1452
<i>sáček</i>	‘bag’	1374
<i>koniček</i>	‘horse’ / ‘hobby’	1330
<i>chlapeček</i>	‘boy’	1201
<i>ptáček</i>	‘bird’	1183
<i>rámeček</i>	‘frame’	1179
<i>sameček</i>	‘male’	1019
<i>míček</i>	‘ball’	1013
<i>strejček</i>	‘uncle (inform.)’	993
<i>klíček</i>	‘key’	972
<i>krůček</i>	‘step’	971

SYN2000		freq.
<i>žebříček</i>	‘ladder’ / ‘ranking list’	4234
<i>dědeček</i>	‘grandfather’	3037
<i>balíček</i>	‘packet’	2267
<i>večerníček</i>	‘TV bedtime story’	1804
<i>vlček</i>	‘wolf’	1544
<i>koniček</i>	‘horse’ / ‘hobby’	1233
<i>háček</i>	‘hook’	1094
<i>strýček</i>	‘uncle’	1034
<i>ptáček</i>	‘bird’	1008
<i>míček</i>	‘ball’	1006
<i>rámeček</i>	‘frame’	962
<i>krůček</i>	‘step’	891
<i>sáček</i>	‘bag’	885
<i>chlapeček</i>	‘boy’	669
<i>zámeček</i>	‘castle’	636

Table 1

2) feminine diminutives

SYN2005		freq.
<i>babička</i>	‘grandmother’	6505
<i>písnička</i>	‘song’	4520
<i>lžička</i>	‘spoon’	4230
<i>ulička</i>	‘street’ / ‘lane’	4063
<i>holčička</i>	‘girl’	3722
<i>svíčka</i>	‘candle’	2741
<i>lavička</i>	‘bench’	2501
<i>sklenička</i>	‘glass’ / ‘goblet’	2349
<i>krabička</i>	‘box’	2220
<i>hlavička</i>	‘head’ / ‘heading’	2156

SYN2000		freq.
<i>písnička</i>	‘song’	5154
<i>babička</i>	‘grandmother’	4465
<i>ulička</i>	‘street’ / ‘lane’	2262
<i>lavička</i>	‘bench’	2226
<i>lžička</i>	‘spoon’	2134
<i>hlavička</i>	‘head’ / ‘heading’	2087
<i>svíčka</i>	‘candle’	1782
<i>holčička</i>	‘girl’	1723
<i>krabička</i>	‘box’	1530
<i>sklenička</i>	‘glass’ / ‘goblet’	1215

Table 2

3) neuter diminutives

SYN2005		freq.
<i>městečko</i>	‘town’	4990
<i>vajíčko</i>	‘egg’	3074
<i>kolečko</i>	‘wheel’ / ‘ring’	2437
<i>víčko</i>	‘cover’ / ‘eyelid’	2174
<i>tričko</i>	‘t-shirt’	1900
<i>sluníčko</i>	‘sun’	1439
<i>očko</i>	‘eye’	1068
<i>slovíčko</i>	‘word’	628
<i>políčko</i>	‘field’	586
<i>srdíčko</i>	‘heart’	477

SYN2000		freq.
<i>městečko</i>	‘town’	3613
<i>vajíčko</i>	‘egg’	1908
<i>kolečko</i>	‘wheel’ / ‘ring’	1630
<i>víčko</i>	‘cover’ / ‘eyelid’	1289
<i>tričko</i>	‘t-shirt’	1215
<i>sluníčko</i>	‘sun’	954
<i>očko</i>	‘eye’	808
<i>políčko</i>	‘field’	568
<i>slovíčko</i>	‘word’	480
<i>srdíčko</i>	‘heart’	323

Table 3

Appendix 2

1) the relationship between diminutives and their non-diminutive counterparts in the corpus SYN2005

a) the non-diminutive noun is the more frequent of the two

Diminutive	Frequency		Underived noun	Frequency
<i>městečko</i>	4990	‘town’	<i>město</i>	78470
<i>písnička</i>	4520	‘song’	<i>píseň</i>	10005
<i>lžička</i>	4230	‘spoon’	<i>lžice</i>	7611
<i>ulička</i>	4063	‘street’	<i>ulice</i>	33630
<i>vajíčko</i>	3074	‘egg’	<i>vejce</i>	6435
<i>lavička</i>	2501	‘bench’	<i>lavice</i>	3443
<i>sklenička</i>	2349	‘glass’	<i>sklenice</i>	4836
<i>krabička</i>	2220	‘box’	<i>krabice</i>	3592
<i>hlavička</i>	2156	‘head’	<i>hlava</i>	68689
<i>kulička</i>	1665	‘globe’	<i>koule</i>	2677
<i>kostička</i>	1612	‘bone’	<i>kost</i>	7406
<i>sluníčko</i>	1439	‘sun’	<i>slunce</i>	17747
<i>koníček</i>	1330	‘horse’	<i>kůň</i>	14134
<i>vesnička</i>	1241	‘village’	<i>vesnice</i>	9240

Table 4

b) the diminutive noun is the more frequent one – semantic differentiation of the diminutive

Diminutive	Frequency		Underived noun	Frequency
<i>balíček</i>	3111	‘packet’ / ‘parcel’	<i>balík</i>	2814
<i>háček</i>	1452	‘hook’ / ‘needle’	<i>hák</i>	660
<i>svíčka</i>	2741	‘candle’ / ‘plug’	<i>svíce</i>	880
<i>tričko</i>	1900	‘t-shirt’	<i>triko</i>	590
<i>víčko</i>	2174	‘cover’ / ‘eyelid’	<i>víko</i>	1109
<i>žebříček</i>	1929	‘ladder’ / ‘ranking list’	<i>žebřík</i>	1391

Table 5

2) the relationship between diminutives and their non-diminutive counterparts in the corpus ORAL2006

Diminutive	Frequency		Underived noun	Frequency
<i>knížka</i>	150	‘book’	<i>kniha</i>	42
<i>sluníčko</i>	76	‘sun’	<i>slunce</i>	22
<i>vajíčko</i>	68	‘egg’	<i>vejce</i>	20
<i>tričko</i>	66	‘t-shirt’	<i>triko</i>	15
<i>dárek</i>	64	‘present’ / ‘gift’	<i>dar</i>	9
<i>lístek</i>	56	‘leaf’ / ‘ticket’	<i>list</i>	29
<i>sklenička</i>	45	‘glass’ / ‘goblet’	<i>sklenice</i>	10
<i>krabička</i>	41	‘box’	<i>krabice</i>	40
<i>lednička</i>	38	‘refrigerator’	<i>lednice</i>	27
<i>prstýnek</i>	30	‘ring’	<i>prsten</i>	14

Table 6

Appendix 3 – the most frequent evaluative adjectives in the environment of the diminutive nouns

1) Masculine diminutives with the suffix *-ček*:

<i>milý</i>	‘pleasing’	89
<i>hodný</i>	‘kind’	88
<i>dobrý</i>	‘good’	71
<i>pěkný</i>	‘pretty’	69
<i>drahý</i>	‘dear’	65
<i>krásný</i>	‘beautiful’	60
<i>roztomilý</i>	‘charming’	44
<i>jemný</i>	‘delicate’	42
<i>kouzelný</i>	‘fascinating’	41
<i>pravý</i>	‘real’	39

Table 7

<i>oblíbený</i>	‘favourite’	35
<i>lehký</i>	‘light’	30
<i>měkký</i>	‘soft’	27
<i>hezký</i>	‘nice’	26
<i>teplý</i>	‘warm’	26
<i>bohatý</i>	‘rich’	25
<i>skromný</i>	‘modest’	25
<i>veselý</i>	‘cheerful’	22
<i>rozkošný</i>	‘delightful’	21
<i>sladký</i>	‘sweet’	21

A similar list for nouns with the suffix *-ek (-nek)*, i.e., without the explicit signal of diminutiveness, looks as follows.

<i>jemný</i>	‘delicate’	298
<i>hezký</i>	‘nice’	146
<i>pěkný</i>	‘pretty’	126
<i>krásný</i>	‘beautiful’	111
<i>lehký</i>	‘light’	97
<i>milý</i>	‘pleasing’	87

Table 8

<i>drahý</i>	‘dear’	84
<i>příjemný</i>	‘pleasant’	61
<i>teplý</i>	‘warm’	54
<i>čistý</i>	‘clean’	44
<i>měkký</i>	‘soft’	32
<i>skromný</i>	‘modest’	22

2) Feminine diminutives with the suffix *-čka*

<i>jemný</i>	‘delicate’	118
<i>pěkný</i>	‘pretty’	109
<i>krásný</i>	‘beautiful’	107
<i>hezký</i>	‘nice’	88
<i>drahý</i>	‘dear’	82
<i>milý</i>	‘pleasing’	81
<i>hodný</i>	‘kind’	76
<i>veselý</i>	‘cheerful’	62
<i>dobrý</i>	‘good’	49

Table 9

<i>roztomilý</i>	‘charming’	44
<i>chytrý</i>	‘clever’	43
<i>kouzelný</i>	‘fascinating’	42
<i>sladký</i>	‘sweet’	37
<i>půvabný</i>	‘lovely’	27
<i>čistý</i>	‘clean’	27
<i>slušný</i>	‘decent’	24
<i>nenápadný</i>	‘discreet’	20
<i>rozkošný</i>	‘delightful’	20

3) Neuter diminutives with the suffix *-čko*:

<i>pěkný</i>	‘pretty’	30
<i>jemný</i>	‘delicate’	29
<i>hezký</i>	‘nice’	22
<i>krásný</i>	‘beautiful’	22
<i>drahý</i>	‘dear’	20
<i>milý</i>	‘pleasing’	20
<i>půvabný</i>	‘lovely’	20
<i>klidný</i>	‘calm’	19

Table 10

<i>měkký</i>	‘soft’	18
<i>dobrý</i>	‘good’	17
<i>oblíbený</i>	‘favourite’	16
<i>sladký</i>	‘sweet’	16
<i>příjemný</i>	‘pleasant’	11
<i>zajímavý</i>	‘interesting’	11
<i>roztomilý</i>	‘charming’	10

Appendix 4: The list of 160 positively evaluative adjectives (after Čermák-Holub 2005)

<i>aktivní</i> (active)	<i>moderní</i> (modern)	<i>světlý</i> (light, pale)
<i>báječný</i> (splendid, magnificent)	<i>moudrý</i> (wise)	<i>sympatický</i> (pleasant, likeable)
<i>bdělý</i> (watchful)	<i>nádherný</i> (wonderful)	<i>šikovný</i> (handy)
<i>bezpečný</i> (safe)	<i>nadšený</i> (enthusiastic)	<i>šťastný</i> (happy)
<i>blahodárný</i> (wholesome, beneficial)	<i>náležitý</i> (proper)	<i>štedrý</i> (open-handed)
<i>blahý</i> (blissful)	<i>nenápadný</i> (discreet, quiet)	<i>teplý</i> (warm)
<i>blažený</i> (blissfully happy)	<i>obdivuhodný</i> (admirable)	<i>tichý</i> (quiet, silent)
<i>bodrý</i> (jovial)	<i>obětavý</i> (selfless, devoted)	<i>tradiční</i> (traditional)
<i>bohatý</i> (rich)	<i>oblíbený</i> (popular, favorite)	<i>trpělivý</i> (patient)
<i>bystrý</i> (quick, fast flowing, bright)	<i>obratný</i> (skilful)	<i>typický</i> (typical)
<i>čilý</i> (sprightly)	<i>odvážný</i> (courageous)	<i>upřímný</i> (frank, candid)
<i>čiperný</i> (lively, vivacious)	<i>opatrný</i> (cautious, careful)	<i>urozený</i> (noble, gentle)
<i>čistý</i> (clean, clear)	<i>pečlivý</i> (solicitous)	<i>úspěšný</i> (successful)
<i>dobromyslný</i> (good-natured)	<i>pěkný</i> (nice)	<i>ušlechtilý</i> (noble-minded, magnanimous)
<i>dobrý</i> (good)	<i>pokorný</i> (meek)	<i>užaslý</i> (amazed, astonished)
<i>dojemný</i> (touching, moving)	<i>pořádný</i> (orderly, neat)	<i>užitečný</i> (useful)
<i>drahocenný</i> (valuable)	<i>poslušný</i> (obedient)	<i>vášnivý</i> (passionate)
<i>drahý</i> (dear)	<i>pozitivní</i> (positive)	<i>vážený</i> (respected, reputable)
<i>dramatický</i> (dramatic)	<i>pracovitý</i> (industrious, hard-working)	<i>vážný</i> (serious)
<i>družný</i> (friendly)	<i>praktický</i> (practical)	<i>vděčný</i> (grateful)
<i>důležitý</i> (important)	<i>pravý</i> (real)	<i>velkolepý</i> (grand, magnificent)
<i>důsledný</i> (consistent)	<i>proslulý</i> (famous)	<i>velkorysý</i> (generous, magnanimous)
<i>důstojný</i> (dignified)	<i>prospěšný</i> (beneficial, useful)	<i>věrný</i> (faithful)
<i>dychtivý</i> (eager)	<i>prozíravý</i> (far-sighted)	<i>věřící</i> (religious)
<i>elegantní</i> (elegant)	<i>překrásný</i> (exquisite, lovely)	<i>veselý</i> (cheerful)
<i>geniální</i> (brilliant, genius)	<i>překvapený</i> (surprised)	<i>vhodný</i> (suitable)
<i>harmonický</i> (harmonic)	<i>přesný</i> (exact, accurate)	<i>vlahý</i> (mild)
<i>hebký</i> (smooth)	<i>přesvědčivý</i> (convincing, persuasive)	<i>vlažný</i> (lukewarm., tepid)
<i>hezký</i> (pretty)	<i>příhodný</i> (suitable, appropriate)	<i>vlhký</i> (wet, moist)
<i>hladký</i> (smooth, plain)	<i>příjemný</i> (pleasant)	<i>vídný</i> (approachable)
<i>hlasitý</i> (loud)	<i>příznivý</i> (advantageous, favourable)	<i>vlivný</i> (influential)
<i>hodný</i> (good, kind)	<i>půvabný</i> (lovely, gracious)	<i>vnímavý</i> (perceptive)
<i>horlivý</i> (eager, enthusiastic)	<i>rozkošný</i> (delectable, lovely)	<i>vroucí</i> (affectionate)
<i>hrdinný</i> (heroic)	<i>rozmarňný</i> (moody, capricious)	<i>všestranný</i> (versatile)
<i>hrdý</i> (proud)	<i>roztomilý</i> (charming)	<i>výbojný</i> (aggressive)
<i>charakterní</i> (upright, honest)	<i>rozumný</i> (sensible)	<i>výhodný</i> (advantageous)
<i>chytrý</i> (clever)	<i>sebevědomý</i> (self-confident)	<i>vynikající</i> (excellent)
<i>ideální</i> (ideal)	<i>shovívavý</i> (indulgent)	<i>vytrvalý</i> (persistent)
<i>iniciativní</i> (enterprising, go-ahead)	<i>skromný</i> (modest)	<i>význačný</i> (prominent, outstanding)
<i>jedinečný</i> (unique)	<i>skvělý</i> (perfect)	<i>vzrušený</i> (excited)
<i>jemný</i> (soft, fine, delicate)	<i>sladký</i> (sweet)	<i>záživný</i> (entertaining)
<i>kladný</i> (positive)	<i>slavný</i> (famous, renowned)	<i>zajímavý</i> (interesting)
<i>klidný</i> (calm)	<i>slibný</i> (promising)	<i>zámožný</i> (affluent, opulent)
<i>kouzelný</i> (charming, fascinating)	<i>sličný</i> (comely, lovely)	<i>zasloužený</i> (well-deserved)
<i>krásný</i> (beautiful)	<i>slušný</i> (decent, respectable)	<i>závažný</i> (relevant, weighty)
<i>laskavý</i> (kind)	<i>smělý</i> (fearless)	<i>zbožný</i> (pious, religious)
<i>lehký</i> (light)	<i>snadný</i> (easy, simple)	<i>zdatný</i> (fit)
<i>lepší</i> (better)	<i>soucitný</i> (sympathetic)	<i>zkušený</i> (experienced)
<i>libý</i> (delightful)	<i>spořádaný</i> (well-ordered)	<i>znamenitý</i> (excellent)
<i>měkký</i> (soft)	<i>spravedlivý</i> (rightful, fair)	<i>zručný</i> (skilful, deft)
<i>milosrdný</i> (merciful)	<i>správný</i> (right, correct)	<i>zvědavý</i> (curious)
<i>milý</i> (pleasing, pleasant)	<i>srděčný</i> (cordial, hearty)	<i>žádoucí</i> (desirable)
<i>mimořádný</i> (extraordinary)	<i>starostlivý</i> (considerate)	
<i>mistrovský</i> (masterly)	<i>statečný</i> (brave)	

