

A Statistical Approach to Czech Prepositional Collocations

Renata Blatná

Institute of the Czech National Corpus, Faculty of Arts, Charles University
nám. J. Palacha 2
116 38 Prague
Czech Republic
renata.blatna@ff.cuni.cz

Abstract

The paper deals with prepositional collocations in Czech which consist of a preposition and an autosemantic word, esp. abstract noun, either of paradigmatic type, e. g. *v rámci* (within the framework of), or non-paradigmatic type – *vzhledem k* (with respect to). After the description of their rhythmical and lexical structure, a statistical analysis of five chosen groups follows. The statistical values are frequency of the abstract noun, frequency of the prepositional collocation, ratio between these two frequencies and MI-score. It appears that the high statistical data, i. e. MI-score and the ratio between the collocation frequency and the lemma frequency are fundamental for searching of the basic member of each group as well as of other fully lexicalized members. These collocations should be elaborated in specific lemmas.

1. Topic

The aim of this paper is to study Czech compound prepositions or prepositional collocations, such as *according to*, *with respect to* etc. in English (*s ohledem na* in Czech). These prepositions mostly consist of a preposition and an autosemantic word, esp. noun. Some of these compounds in Czech are fully lexicalized, such as *vzhledem k* (owing to) or *v rámci* (within the framework of), some of them are used rarely or in specific types of texts. In this paper the compounds will be studied from the aspect of their a) frequency, b) frequency of the autosemantic lemma, c) ratio between these two frequencies, d) MI-score and e) structure of the prepositional compound. The open question is: under what conditions do Czech users consider a specific combination as a compound preposition? The representative 100 mil. corpus of the Czech language SYN2000 (Czech National Corpus) which was built at Charles University, Faculty of Arts in Prague, includes 60 per cent of newspaper texts, 15 per cent of fiction and factual literature, and 25 per cent of academic prose. As these prepositions occur mostly in newspapers and academic prose (85 per cent of the corpus, i. e. 85 mil. word forms) there is a strong possibility of exploiting the results given below in lexicography.

2. Structure of the Czech Prepositional Collocations

There are 370 prepositional collocations in Czech, which are described in the appendix of “Czech Idiom Dictionary” (“Slovník české frazeologie a idiomatiky”) devoted to nominal idioms. The structure of these collocations can be either paradigmatic or nonparadigmatic

(Čermák 1996). The paradigmatic type consists of a preposition and an abstract noun, i. e. prep - N^{Abst}, e. g. *v rámci*, or of a preposition, an abstract noun and another preposition, i. e. prep - N^{Abst} - prep, e. g. *na rozdíl od*. Other types are nonparadigmatic, esp. those including other parts of speech, such as verbs, adverbs and synsemantic words. We will examine here 303 paradigmatic collocations and compare them with 10 nonparadigmatic collocations including nouns as well. First nonparadigmatic type consists of a noun in instrumental form and a preposition, i. e. N^c - prep, e. g. *vzhledem k* and the second type consists of a noun, a preposition and a repetition of the same noun, i. e. N - prep - N, e. g. *bok po boku*.

3. Abstract Nouns

There is quite a limited number of the abstract nouns which can stand as components of the prepositional collocations. With the exception of several nouns, such as *život* 'life', *práce* 'work', *svět* 'world', *slovo* 'word', *rok* 'year', *národ* 'nation', *pravda* 'truth' the most frequent abstract nouns are on the list. According to the manuscript of F. Čermák „Grammatical phrasemes and idioms“ the majority of the abstract nouns is of a static, non-dynamic character, derived from verbs, cf. *pomoci - pomoc* 'to help - help', *přát si - přání* 'to wish - wish', *vztahovat se - vztah* 'to relate - relation' etc. Some of them have local and temporal meaning, e. g. *střed* 'centre', *konec* 'end', *oblast* 'area', *doba*, *čas* 'time, days' etc. In addition to the primary abstract nouns the prepositional collocations with secondary abstract nouns, i. e. metaphorically used concrete nouns appear as well, such as somatic ones, e. g. *bok* 'side', *čelo* 'forehead, fore', *ruka* 'hand', *tvář* 'face', and some others, e. g. *vrub* 'score', *půda* 'soil, ground', *zástěrka* 'apron, cover-up, pretext'. Nevertheless, the somatic nouns occur in the non-paradigmatic types in the first place, e.g. *tváří v tvář* 'face to face' (see 6.2). There is a rule that an abstract noun is usually used in one prepositional collocation, while its occurrence in two prepositional collocations is a signal of variants. This can be tested on the basis of their translation into English, e. g. the Czech collocations *s podporou* and *za podpory* have the same English equivalent 'with the support of'. The only exception are the nouns *doba* and *čas* 'time, days' which can occur in several prepositional collocations.

4. Paradigmatic Types

As the structure of collocations will be compared with their MI-score, two paradigmatic types will be treated as a common structure prep - N^{Abst} (within MI-score only two members can be statistically taken into account). In the following list the main paradigmatic structures are given. Explanation of the table: The first word of the structure is the Czech one-word preposition, e. g. *bez*, *do*, *na*, *v*, *s*, *z* (English without, to, on, in, with, from), the second word is the abstract noun, where M,F,N stand for masculine, feminine, neuter in specific case and the number of the case is 2, 3, 4, 6 or 7 for genitive, dative, accusative, local, instrumental. The slash / divides the grammatical information from the rhythmical information, i. e. the number signifies number of syllables of the structure. As an example, the most frequent member of the group is chosen. It can be clarified on one of the structures, e. g. the structure number 12 has five members and consists of the preposition *na* 'on' and a three-syllable feminine noun in accusative form.

NO.	STRUCTURE	EXAMPLE	ENGLISH EQUIVALENT	NUMBER OF MEMBERS
1	bez + M2/3	<i>bez ohledu (k/na)</i>	,without regard to'	4
2	bez + N2/3	<i>bez vědomí</i>	,without s.o.'s knowing'	1
3	do + F2/1	<i>do řad</i>	,into y ranks of'	1
4	do + F2/2	<i>do doby</i>	,till the time of'	2
5	do + M2/2	<i>do středu</i>	,into the centre of'	2
6	do + N2/2	<i>do centra</i>	,into the centre of'	1
7	do + N2/3	<i>do popředí</i>	,to the fore of'	2
8	k + F3/2	<i>k počtě</i>	,in honour of'	2
9	ke + F3/2	<i>ke škodě</i>	,to the detriment of'	1
10	ke + M3/2	<i>ke konci</i>	,towards the end of'	1
11	na + F4/2	<i>na dobu</i>	,for the period of'	7
12	na + F4/3	<i>na adresu</i>	directed at	5
13	na + F6/2	<i>na půdě</i>	,on the grounds of'	3
14	na + F6/3	<i>na úrovni</i>	,on the level of'	4
15	na + M4/2	<i>na rozdíl (od)</i>	,in contrast to'	19
16	na + M4/3	<i>na motivy</i>	,on the leitmotifs of'	1
17	na + M6/2	<i>na konci</i>	,in the end of'	3
18	na + M6/3	<i>na základě</i>	,on the basis of'	10
19	na + N4/2	<i>na přání</i>	,according to wish of'	2
20	na + N4/3	<i>na znamení</i>	,as a sign of'	1
21	na + N4/4	<i>na naléhání</i>	,on the insistence of'	1
22	na + N6/3	<i>na pozadí</i>	,against the background of'	2
23	o + F6/3	<i>o otázce</i>	,about the problem of'	1
24	o + M6/3	<i>o rozsahu</i>	,in the extent of'	1
25	od + F2/1	<i>od dob</i>	,since the days of'	1
26	od + F2/2	<i>od doby</i>	,since the time of'	1
27	po + F4/2	<i>po dobu</i>	,for the duration of'	1
28	po + F6/2	<i>po stránce</i>	,in respect of'	1
29	po + F6/3	<i>po dohodě (s)</i>	,by agreement with'	2
30	po + M6/2	<i>po boku (s)</i>	,side by side with'	2
31	po + M6/3	<i>po způsobu</i>	,after the fashion of'	2
32	pod + F7/2	<i>pod tíhou</i>	,under the weight of'	1
33	pod + F7/3	<i>pod záštitou</i>	,under the auspices of'	8
34	pod + F7/4	<i>pod patronací</i>	,under the auspices of'	1
35	pod + M7/2	<i>pod názvem</i>	,under the title of'	6
36	pod + M7/3	<i>pod titulem</i>	,under the title of'	1
37	pod + M7/4	<i>pod patronátem</i>	,under the auspices of'	1
38	pod + N7/2	<i>pod heslem</i>	,under the slogan of'	2
39	pod + N7/3	<i>pod vedením</i>	,under the leadership of'	1
40	pro + F4/3	<i>pro potřeby</i>	,for uses of'	2

41	pro + M4/2	<i>pro případ</i>	,in case of'	1
42	při + F6/4	<i>při příležitosti</i>	,on the occasion of'	2
43	při + M4/3	<i>pro účely</i>	,for the purposes of'	1
44	při + M6/3	<i>při pokusu</i>	,attempting to'	1
45	s + F7/4	<i>s účinností</i>	,with effect from'	1
46	s + M7/2	<i>s cílem</i>	,with the aim of'	2
47	s + M7/3	<i>s ohledem (k/na)</i>	,with a view to'	7
48	s + N7/3	<i>s vědomím</i>	,with consent of'	2
49	s + N7/4	<i>s odvoláním (na)</i>	,with reference to'	5
50	s+ F7/3	<i>s pomocí</i>	,with the help of'	4
51	se + M7/2	<i>se sklonem (k)</i>	,tending to'	1
52	se + M7/3	<i>se zřetelem (k)</i>	,with a view to'	3
53	se + M7/4	<i>se zaměřením</i>	,with a view to'	1
54	se + N7/2	<i>se zřením</i>	,with a view to'	1
55	u + F2/5	<i>u příležitosti</i>	,on the occasion of'	1
56	u + N2/3	<i>u vědomí</i>	,being aware of'	1
57	v + F4/3	<i>v náhradu (za)</i>	,in compensation for'	2
58	v + F6/2	<i>v době</i>	,during the time of'	9
59	v + F6/3	<i>v oblasti</i>	,in the area of'	15
60	v + F6/4	<i>v souvislosti (s)</i>	,in connection with'	4
61	v + F6/5	<i>v záležitosti</i>	,in the affair of'	1
62	v + M4/3	<i>v neprospěch</i>	,to the disadvantage of'	1
63	v + M6/2	<i>v rámci</i>	,within the framework of'	5
64	v + M6/3	<i>v případě</i>	,in the case of'	14
65	v + M6/4	<i>v protikladu (k)</i>	,in contrast with'	2
66	v + N6/2	<i>v čele</i>	,at the head of'	4
67	v + N6/3	<i>v okolí</i>	,in the environs of'	9
68	v + N6/4	<i>v porovnání (s)</i>	,in comparison with'	4
69	ve + F6/2	<i>ve sféře</i>	,in the sphere of'	10
70	ve + F6/3	<i>ve šlépějích</i>	,in footsteps of'	1
71	ve + F6/4	<i>ve vzdálenosti</i>	,in the distance of'	2
72	ve + M4/2	<i>ve prospěch</i>	,for the benefit of'	1
73	ve + M6/2	<i>ve vztahu (k)</i>	,in relation with'	8
74	ve + N6/2	<i>ve znění</i>	,in the wording of'	3
75	z + F2/1	<i>z řad</i>	,out of the ranks of'	1
76	z + F2/2	<i>z funkce</i>	,in the capacity of'	5
77	z + F2/3	<i>z pozice</i>	,in the position of'	4
78	z + F2/6	<i>z iniciativy</i>	,on the initiative of'	1
79	z + M2/3	<i>z důvodů</i>	,for the reasons of'	11
80	z + M2/4	<i>z nedostatku</i>	,due to lack of'	1
81	z + N2/3	<i>z hlediska</i>	,from the aspect of'	1
82	z + N2/4	<i>z rozhodnutí</i>	,following the decision of'	2

83	za + F2/1	<i>za dob</i>	,in the days of ^c	1
84	za + F2/2	<i>za dobu</i>	,during the time of ^c	2
85	za + F2/3	<i>za účasti</i>	,with the participation of ^c	2
86	za + F2/4	<i>za přítomnosti</i>	,in presence of ^c	4
87	za + F2/4	<i>za cenu</i>	,at the (very) cost of ^c	1
88	za + F7/3	<i>za příčinou</i>	,for the aim of ^c	1
89	za + M2/2	<i>za času</i>	in the days of ^c	2
90	za + M2/3	<i>za souhlasu</i>	,with the agreement of ^c	1
91	za + M2/4	<i>za předpokladu</i>	,on the assumption of ^c	2
92	za + M7/3	<i>za účelem</i>	,with the aim of ^c	1
93	za + N2/3	<i>za přispění</i>	,with the contribution of ^c	1
94	ze + F2/2	<i>ze strany</i>	,from the side of ^c	2
95	ze + F2/4	<i>ze vzdálenosti</i>	,from the distance of ^c	1
96	ze + M2/2	<i>ze středu</i>	,from the centre of ^c	1
97	ze + N2/4	<i>ze stanoviska</i>	,from the aspect of ^c	1

Table 1: Paradigmatic structures of the Czech prepositional collocations

It appears that the groups with 1-2 members include a noun form consisting either of 1 or 4 and more syllables, cf. *na vrub* ,at the expense of^c and *pod patronátem* ,under the auspices of^c etc. On the other hand, the groups consisting of 3 and more members have usually two- or three-syllable noun form, such as *z důvodů* ,for the reasons of^c, *ve sféře* ,in the sphere of^c, *v okolí* ,in the environs of^c etc. However, it must be emphasized that 97 types of prepositional collocations in the table 1 can be reduced if the vocalized prepositions – i.e. *ve*, *se*, *ke*, *ze* are included in one group together with the non-vocalized prepositions *v*, *s*, *k*, cf. two groups *ve prospěch* (72) and *v neprospěch* (62) The prepositional collocation *ve prospěch* includes a vocalized preposition *ve* (1 syllable) and the noun consists of two syllables - *prospěch* ,benefit^c, opposed to the prepositional collocation with the opposite meaning *v neprospěch* ,to the disadvantage of^c which has a non-vocalized preposition *v* (0 syllables) and a three-syllable noun *neprospěch* ,disadvantage^c. Similarly, into one group the groups *v oblasti* ,in the area of^c(59) and *ve sféře* ,in the spere of^c (69), *se sklonem* (*k*) ,tending to^c (51) and *s ohledem* (*k/na*) ,with a view to^c (47) etc. can be incorporated.

5. Statistical Analysis of the Prepositional Collocations

In the following analysis 5 groups will be studied in a more detailed way. The number of the members in each of the following group will include additional 3-5 members, i. e. more than in the previous group, namely 4 members, 8 members, 11 members, 15 members, 20 members. The frequency of the lemma and frequency of the collocation as well as their ratio (collocation frequency/lemma frequency) and mutual information (MI-score) will be taken into account.

5.1 The „bez+M2/3“ Group

In this group collocations with preposition *bez* ‚without‘ and a masculine noun in genitive case the form of which consists of three syllables are studied.

bez + M2/3		bez ohledu (k/na) (without regard to)			4
LEMMA	LEMMA FREQ.	RATIO	MI- SCORE	COLLOCATION	COLLOCATION FREQ.
<i>nárok</i>	10278	0.01	8.026	<i>bez nároku (na)</i>	176
<i>ohled</i>	11071	0.39	9.924	<i>bez ohledu (k/na)</i>	4362
<i>rozdíl</i>	23087	0.02	9.133	<i>bez rozdílu</i>	558
<i>zřetel</i>	1190	0.04	7.838	<i>bez zřetele (k/na)</i>	42

Table 2: Type *bez ohledu (k/na)*

The MI-score values in this group grow from 7.8 to 9.9. The collocation with the highest MI-score value is *bez ohledu k/na* which is the most frequent but which has the highest ratio between collocation and lemma frequency as well = 0.39. The average ratio of the other members of this group is 0.03. It seems that collocation *bez ohledu k/na* is the strongest member of this group and under its influence the rest of this group are taken as prepositional collocations. The MI-score depends on the frequency of the first preposition – frequency of preposition *bez* is 86340.

The nouns used in this group are *nárok* ‚claim‘, *ohled* ‚regard‘, *rozdíl* ‚difference‘ and *zřetel* ‚respect, regard‘.

5.2 The „pod+F7/3“ Group

In this group the collocations with the structure preposition *pod* ‚under‘ and feminine noun in instrumental consisting of three syllables are studied:

pod + F7/3		pod záštitou (under the auspices of)			8
LEMMA	LEMMA FREQ.	RATIO	MI- SCORE	COLLOCATIO N	COLLOCATIO N FREQ.
<i>ochrana</i>	17791	0.01	8.402	<i>pod ochranou</i>	189
<i>pečet'</i>	652	0.01	6.708	<i>pod pečeti</i>	9
<i>podmínka</i>	36458	0.01	7.302	<i>pod podmínkou</i>	229
<i>pohružka</i>	313	0.47	10.467	<i>pod pohružkou</i>	147
<i>taktovka</i>	650	0.66	10.476	<i>pod taktovkou</i>	429
<i>záminka</i>	1163	0.24	9.729	<i>pod záminkou</i>	282
<i>zástěrka</i>	227	0.05	8.077	<i>pod zástěrkou</i>	11
<i>záštita</i>	950	0.59	10.528	<i>pod záštitou</i>	561

Table 3: Type *pod záštitou*

This group is very strong. Two members have the ratio over 0.50 (*pod taktovkou* ,under the baton‘ and *pod záštitou* ,under the auspices of‘) and one is very near to this value (0.47 – *pod pohrůzkou* ,under threat of‘). All these three representatives have MI-scores over 10. The MI-score value depends on the frequency of the preposition *pod* ,under‘ which is rather low – 69374. The average ratio of the other member of this group is 0.06. It can be expected that in future new compound prepositions with this structure will appear. The nouns occurring in this group are *ochrana* ,protection‘, *pečet* ,seal‘, *podmínka* ,condition‘, *pohrůžka* ,threat‘, *taktovka* ,baton‘, *záminka* ,pretext‘, *zástěrka* ,cover-up, pretext‘, *záštita* ,auspice‘.

5.3 The „z+M2/3“ Group

In this group collocations with preposition *z* ,from‘ and a three-syllable form of masculine noun in genitive are studied:

z + M2/3		z důvodů (for reasons of)			11
LEMMA	LEMMA FREQ.	RATIO	MI- SCORE	COLLOCATIO N	COLLOCATIO N FREQ.
<i>aspekt</i>	3570	0.00	2.352	<i>z aspektu</i>	10
<i>dosah</i>	1894	0.07	5.031	<i>z dosahu</i>	125
<i>důvod</i>	35358	0.05	5.446	<i>z důvodu</i>	1750
<i>důvod</i>	35358	0.06	4.859	<i>z důvodů</i>	2061
<i>návod</i>	2321	0.01	2.652	<i>z návodu</i>	21
<i>podnět</i>	3873	0.06	5.974	<i>z podnětu</i>	227
<i>pohled</i>	24309	0.06	4.877	<i>z pohledu</i>	1451
<i>popud</i>	699	0.11	5.899	<i>z popudu</i>	82
<i>příkaz</i>	4250	0.01	3.876	<i>z příkazu</i>	60
<i>rozkaz</i>	2690	0.01	3.364	<i>z rozkazu</i>	25
<i>titul</i>	16510	0.02	4.053	<i>z titulu</i>	405

Table 4: Type *z důvodů*

This group is very specific. There is no member with the ratio value near 0.50. The highest ratio is 0.11 (*z popudu* ,on the initiative of‘). The MI-scores vary from 2 to 6 which depends on the high frequency of the preposition *z* ,from‘ – 861071. These collocations are nevertheless treated in various dictionaries and grammars as prepositions. The reason may be the existence of a similar structure but with a neuter noun, *z hlediska* ,from the aspect of‘ with the ratio 0.55 and MI-score 6.432. The nouns which appear in this group are *aspekt* ,aspect‘, *dosah* ,reach‘, *důvod* ,reason‘, *návod* ,instruction‘, *podnět* ,stimulation‘, *pohled* ,sight‘, *popud* ,initiative‘, *příkaz* ,order‘, *rozkaz* ,command‘, *titul* ,title‘. (*z titulu* ,by virtue of‘).

5.4 The „v+F6/3“ group

In this group collocations with the structure preposition *v* ‚in‘ and three-syllable feminine noun in local are described:

v + F6/3		v oblasti (in the area of)			15
LEMMA	LEMMA FREQ.	RATIO	MI- SCORE	COLLOCATIO N	COLLOCATIO N FREQ.
<i>blížkost</i>	4040	0.41	4.517	<i>v blízkosti</i>	1674
<i>hodnota</i>	32203	0.09	4.783	<i>v hodnotě</i>	3035
<i>intence</i>	403	0.45	5.249	<i>v intencích</i>	185
<i>naděje</i>	12808	0.00	2.261	<i>v naději na</i>	36
<i>oblast</i>	44806	0.27	4.364	<i>v oblasti</i>	12040
<i>osoba</i>	32311	0.01	2.327	<i>v osobě</i>	224
<i>otázka</i>	48635	0.03	4.137	<i>v otázce</i>	1299
<i>otázka</i>	48635	0.01	3.458	<i>v otázkách</i>	420
<i>platnost</i>	7109	0.06	3.421	<i>v platnosti</i>	420
<i>podmínka</i>	36458	0.03	3.039	<i>v podmínkách</i>	978
<i>podoba</i>	17494	0.24	4.539	<i>v podobě</i>	4244
<i>poloha</i>	6494	0.04	3.120	<i>v poloze</i>	248
<i>povaha</i>	6112	0.02	2.440	<i>v povaze</i>	107
<i>rovina</i>	4228	0.15	4.128	<i>v rovině</i>	637
<i>úloha</i>	7399	0.01	3.440	<i>v úloze</i>	158

Table 5: Type *v oblasti*

In this group, two members have the ratio value very close to 0.50 – *v blízkosti* ‚nearby‘ and *v intencích* ‚within the intentions of‘ but it seems that the leaders of this group are *v oblasti* ‚in the area of‘ with ratio 0.27 and *v podobě* ‚in the form of‘ with ratio 0.24 and with higher frequencies of the collocation and quite high MI-scores about 4. As the preposition *v* ‚in‘ is the most frequent preposition in Czech with 2525275 occurrences, MI-scores are rather low, from 3 to 6. The average ratio of the remaining 11 members is 0.04. The nouns in this group are *blížkost* ‚neighbourhood‘, *hodnota* ‚value‘, *intence* ‚intention‘, *naděje* ‚hope‘, *oblast* ‚area‘, *osoba* ‚person‘, *otázka* ‚question‘, *platnost* ‚validity‘, *podmínka* ‚condition‘, *podoba* ‚form‘, *poloha* ‚position‘, *povaha* ‚character‘, *rovina* ‚level‘, *úloha* ‚role‘.

5.5 The „na+M4/2“ Group

In this group the structure the preposition *na* ‚on‘ and a two-syllable masculine noun in accusative is examined:

na + M4/2		na rozdíl (od) (in contrast to)			19
LEMMA	LEMMA FREQ.	RATIO	MI- SCORE	COLLOCATIO N	COLLOCATIO N FREQ.
<i>doklad</i>	5982	0.00	0.166	<i>na doklad</i>	18
<i>důkaz</i>	7465	0.03	2.945	<i>na důkaz</i>	249
<i>konec</i>	57888	0.02	3.235	<i>na konec</i>	1038
<i>motiv</i>	5845	0.00	-0.224	<i>na motiv</i>	20
<i>námět</i>	3059	0.01	1.826	<i>na námět</i>	35
<i>nátlak</i>	1739	0.09	3.851	<i>na nátlak</i>	152
<i>návod</i>	2321	0.00	-1.709	<i>na návod</i>	4
<i>návrh</i>	34017	0.02	1.971	<i>na návrh</i>	796
<i>podnět</i>	3873	0.03	2.960	<i>na podnět</i>	129
<i>pokyn</i>	3262	0.06	3.660	<i>na pokyn</i>	212
<i>popud</i>	699	0.27	5.324	<i>na popud</i>	192
<i>pořad</i>	10843	0.03	2.631	<i>na pořad</i>	280
<i>protest</i>	4593	0.18	5.084	<i>na protest</i>	818
<i>příkaz</i>	4250	0.06	3.447	<i>na příkaz</i>	258
<i>účet</i>	10526	0.09	4.329	<i>na účet</i>	920
<i>úkor</i>	1945	0.95	6.113	<i>na úkor</i>	1849
<i>úvod</i>	4720	0.09	4.858	<i>na úvod</i>	408
<i>závěr</i>	18773	0.14	5.312	<i>na závěr</i>	2771
<i>způsob</i>	44054	0.01	1.875	<i>na způsob</i>	537

Table 6: Type *na rozdíl (od)*

For the whole group only one member is central. The collocation *na úkor*, at the expense of has the ratio 0.95 and MI-score over 6. The reason why a collocation with the same meaning but with one-syllable noun form – *na vrub*, on tick of – is used as a compound preposition is in its comparable statistical data: ratio 0.86 and MI-score 6.019. The other members have sometimes very low MI-scores, even negative values. The average ratio of the remaining 18 members of this group is 0.06. The nouns occurring in this group are *doklad* 'evidence', *důkaz* 'proof', *konec* 'end', *motiv* 'leitmotif', *námět* 'subject', *nátlak* 'pressure', *návod* 'instruction', *návrh* 'suggestion', *podnět* 'stimulation', *pokyn* 'hint', *popud* 'initiative', *pořad* 'agenda', *protest* 'protest', *příkaz* 'order', *účet* 'account', *úkor* 'detriment', *úvod* 'introduction', *závěr* 'conclusion', *způsob* 'manner, fashion'.

6. Nonparadigmatic Types

There are two groups including abstract nouns. The first group is represented by the collocation *vzhledem k*, i. e. abstract noun in instrumental form and a preposition, the second group includes abstract noun – preposition – abstract noun.

6.1 Structures with Instrumental Forms

The first group consists of two-syllable or three-syllable form instrumental form and *k*, 'to':

Noun/2-3 + k		vzhledem k (with respect to)			5
LEMMA	LEMMA FREQ.	RATIO	MI- SCORE	COLLOCATIO N	COLLOCATIO N FREQ.
<i>čelo</i>	13925	0.02	5.953	<i>čelem k</i>	209
<i>směr</i>	21619	0.15	6.355	<i>směrem k</i>	3305
<i>vzhled</i>	17487	0.62	7.824	<i>vzhledem k</i>	14967
<i>náhrada</i>	4772	0.03	6.846	<i>náhradou za</i>	149
<i>odměna</i>	3906	0.03	6.313	<i>odměnou za</i>	120

Table 7: Type *vzhledem k*

It is evident that in the 2-syllable group that the less frequent members are fully fixed to the member with highest ratio 0.62 and MI-score over 7 while in the 3-syllable group the statistical data are comparable. There usage is supported by the masculine group. The nouns used in this nonparadigmatic type are *čelo* 'forehead, fore', *směr* 'direction', *vzhled* 'appearance', *náhrada* 'compensation', *odměna* 'reward'.

6.2 The Structure Noun – Preposition – Noun

This structure is not so evident but the strongest member is *tváří v tvář* because of the ratio value. These prepositions are on the decline.

Noun-Prep-Noun		tváří v tvář (face to face)			3
LEMMA	LEMMA FREQ.	RATIO	MI- SCORE	COLLOCATIO N	COLLOCATIO N FREQ.
<i>bok</i>	4551	0.02	3.353	<i>bok po boku</i>	118
<i>ruka</i>	53446	0.01	1.669	<i>ruku v ruce</i>	498
<i>tvář</i>	20359	0.07	3.680	<i>tváří v tvář</i>	1451

Table 8: Type *tváří v tvář*

7. Conclusion

After this statistical research it is possible to extend the number of units in each studied group and get new lemmas for the dictionary description on the basis of the corpus SYN2000. These observations are relevant for future lexicographic work, esp. on a new Czech dictionary. The high statistical data, i. e. MI-score and the ratio between the collocation frequency and the lemma frequency are fundamental for searching of the basic member of the group as well as of other fully lexicalized members. These collocations should be elaborated in specific lemmas.

8. Acknowledgments

The work on this paper was supported by the Ministry of Education of the Czech Republic and its research project No. MSM 112100002.

References

- Čermák, F. 1996. *Systém, funkce, forma a sémantika českých předložek*. Slovo a slovesnost, 57, pp. 30-46.
- Čermák, F. (manuscript). *Gramatické frazémy a idiomy*.
- Čermák, F., Hronek, J. and Machač, J. (eds.) 1988. *Slovník české frazeologie a idiomatiky. Výrazy neslovesné*. Praha: Academia.
- Czech National Corpus - SYN2000. Institute of the Czech National Corpus, Faculty of Arts, Charles University, Prague, Czech Republic: <<http://ucnk.ff.cuni.cz>>.