

Lexico-grammatical Compound Units and their Elaboration in Dictionaries

Abstract

The paper studies the lexico-grammatical compound units functioning in the Czech language as prepositions and conjunctions: *v případě* 'in case (of)', *vzhledem k(e)* 'with respect/regard to' and *na základě* 'on the ground/basis (of)'. The author summarizes descriptions of such units in grammars and dictionaries and compares the dictionary elaboration in Czech and English dictionary entries. On the basis of the corpus instances observed and the frequency data the author gives precision to the previous dictionary descriptions and proposes the optimal structure of elaborating such compound units in the future corpus based Czech dictionary.

Keywords: compound units, collocation, idiom, corpus, dictionary

1. Introduction

This contribution comes from the Czech National Corpus (CNK) which is being built at the Faculty of Arts in Prague, Czech Republic. At present it contains about forty million word forms taken from various Czech newspapers. The plan is to compile a representative corpus of about 100 million lexical forms comprising all types of texts and genres.

The newspaper texts included in the present corpus provide a unique opportunity to study lexical and grammatical units, whose frequency is very high in texts such as newspaper articles. One of the most appropriate topics for this research is to study lexico-grammatical phrases, especially secondary compound prepositions and conjunctions, such as *vzhledem k(e) + dat. prep.* and *vzhledem k tomu, že conj.* 'with respect/regard (to)', *na základě + gen.* and *na základě toho, že conj.* 'on the ground/basis (of)' and *v případě + gen* and *v případě, že/kdy conj.* 'in case (of)'. The prepositions are composed of a noun in a specific case form, i. e. *vhled* nom. 'look, appearance', 'respect' - *vzhledem* instr., *případ* "case", *případě* - loc., *základ* 'foundation, basis', *základě* - loc. and a primary preposition *na, v* and *k*.

As G. Francis (1993) states, the corpus linguistics shows a rather great impact of idiomatics on grammar. To a certain degree, syntactic structure always restricts the choice of lexical units, e. g. the structure *not to have the ...est idea (of)* is characterized by the superlative degree of the adjective and the presence of the definite article. The most frequent adjectives used in this structure are the following: *faintest, slightest, least, foggiest, remotest*, the most frequent noun is *idea*, but *conception* and *notion* are possible as well. Francis also gives an example of the fixed compound expression *be a case of*. The expression on the right-hand side of this fixed compound is very often a negative comparison, a parallel between two unfortunate events.

According to the corpus data, it is evident that there is a continuous transition from the phenomenon which was traditionally labeled "grammar" to that which we used to call "vocabulary". As M.A.K. Halliday (1992) mentions, we can treat the whole language continuum either grammatically or lexically. As synsemantic parts of speech, prepositions and

conjunctions can be elaborated in a different way in the grammar and in the dictionary. While grammarians explain the usage and basic meaning of a group of units with similar properties, lexicographers treat each unit separately, trying to point out the basic differences between two or more similarly elaborated units.

2. Description in Czech and English dictionaries and grammars

Let us briefly summarize the elaboration of secondary prepositions and conjunctions in Czech grammars and monolingual dictionaries and then compare Czech and English (monolingual) dictionary entries. Secondary prepositions in the Czech grammar "Mluvnice češtiny" (1986:207) are described from the point of view of their propositionalization according to syntactic, formal and semantic criteria. The most important criteria for the lexicographer are the semantic ones, i.e. a) the position of the preposition in front of a noun, adjective, pronoun or numeral (e.g. *doprostřed* 'into the middle'), b) the original abstract meaning of the prepositionalized word, and c) the possibility of synonymy of the expression with some primary preposition. This grammar "Mluvnice češtiny" was obviously not based on corpus, as the criterion of frequency is relevant as well, cf. the form *vzhledem* which has 5,364 instances in the CNK out of which

75 instances = the instrumental form of the noun *vzhled*
 5,289 instances = the compound *vzhledem k(e)*.

There is a strong tendency in grammars to treat prepositions as a part of speech with the only syntactic meaning reflecting such relations as temporal (*na prahu* 'at the beginning of'), local (*v čele s* 'at the head of'), causal (*v důsledku* 'in consequence of'), conditional and concessive (*v případě*), of respect (*vzhledem k*) and of manner (*na základě*), and several other more detailed relations, e. g. inclusion and exclusion, etc. In spite of the statement in "Mluvnice češtiny" that the meanings of secondary prepositions can be partly mixed (especially the meaning of the original noun and the preposition as a whole), the given examples do not lead to any clear conclusion. As far as the conjunctions are concerned, two basic groups exist: paratactic and hypotactic conjunctions. Within the hypotactic conjunctions, the relations expressed correspond essentially to those of prepositions with the exception of the local relation. The other Czech grammar "Příruční mluvnice češtiny" (1995) does not give such a detailed analysis. Neither of these publications, however, is based on a corpus, so both grammars lack complete lists of given phrases and frequency data.

Czech dictionaries do not usually use multi-word lemmas. Compound expressions and idiomatic units are thus located under the entry of the one-word lemma. The compound prepositions and conjunctions are therefore included under the entry of the noun, i. e. *vzhledem k(e)* can be found under the entry *vzhled* '1. appearance, 2. respect'. In the largest Czech monolingual dictionary "Slovník spisovného jazyka českého" (most recent edition in 1989), the prepositions and conjunctions studied are treated only by synonymy, not by metalanguage definition. This approach is insufficient and may mislead the user. Such a method of elaboration is more typical of grammars than dictionaries and, in the end, produces a list of expressions without any differentiation. On the other hand, the compound expressions examined were elaborated as grammatical idioms in the dictionary of Czech phraseology and idiomatics "Slovník české frazeologie a idiomatiky II" /SČFI/ (1989) in the volume "Non-verbal expressions." Although not even this dictionary was based on a corpus, the two lists of entries contain 397 compound prepositions and 174 conjunctions. Each of the entries includes a metalanguage definition, synonymous expressions, and examples of usage and information on style.

If the Czech monolingual dictionaries and the English dictionaries (which are based exclusively on the various corpora, especially on The Bank of English corpus) are compared, the elaboration of compound expressions is a bit different. First, these expressions are treated as phrases in the English dictionaries. Second, they are included in the one-word lemma, but the arrangement of the entry easily leads the user to the phrase with its definition, examples and information on style, cf. the entry **respect** and the tenth meaning defining the phrase **with respect to** in Collins COBUILD English Dictionary (1995):

10 You use **with respect to** or in British English **in respect of**, to say what something relates to:,, a formal use. *Parents often have little choice with respect to the way their child is medically treated... Where Dr. Shapland feels the system is not working most effectively is in respect of professional training and development.*

The entry **vzhledem k** in the dictionary of Czech phraseology and idiomatics /SČFI/ (translated into English):

(V) **vzhledem k** (S dat) - "with respect to"

(neutr) *delimitation of an activity or some other action from the aspect of (either objective or subjective) relation to another relevant action, state, especially interest of participants, circumstances etc.;*, *s ohledem na, se zřetelem na, se zřetelem k*. The decision was postponed w. r. t. the given circumstances - It is necessary to take the presence of interpreters into account w. r. t. the participation of foreign guests. - The preposition expresses the relation of respect and partly the causal relation.

Cf. - vzhledem k tomu, že/...

The elaboration of Czech material based on the Czech National Corpus will be more precise and detailed, especially considering a) the differences between prepositions and conjunctions and b) number of meanings (polysemy). As the chief editor of SČFI, F. Čermák states, the prepositional idioms enrich the abstract meanings of the prepositions, such as representation - *ve jménu* 'in the name of', exclusion - *s výjimkou* 'with the exception of', inclusion - *zahrnuje v to* 'including', field of action - *v oblasti* 'in the area of' etc.

The corpus analysis of the compound expressions functioning as prepositions and conjunctions at the same time gives the following frequency data:

V PŘÍPADĚ	7,170 instances
(preposition and conjunction)	
(loc.) form <i>případě</i>	14,216 instances
lemma <i>případ</i> (noun)	16,135 instances
V PŘÍPADĚ + gen.	5,378 instances
(preposition: prep+noun)	

V PŘÍPADĚ, ŽE/KDY (conjunction: prep+noun+con)	1,792 instances
<i>v případě, že</i>	1,675 instances
<i>v případě, kdy</i>	117 instances
<hr/>	
VZHLEDEM K(E) (preposition and conjunction)	5,289 instances
instr. form <i>vhledem</i>	5,368 instances
lemma <i>vhled</i> (noun)	409 instances
VZHLEDEM K(E) + dat. (preposition: noun+prep)	3,951 instances
VZHLEDEM K TOMU, ŽE	1,338 instances
<hr/>	
NA ZÁKLADĚ (preposition and conjunction)	5,192 instances
(loc.) form <i>základě</i>	5,938 instances
lemma <i>základ</i> (noun)	746 instances
NA ZÁKLADĚ + gen. (preposition: prep+noun)	5,095 instances
NA ZÁKLADĚ TOHO, ŽE (conjunction: prep+noun+conj)	97 instances

All three compound expressions are primarily prepositions and their conjunctive function is derived from the prepositional one by substituting a noun with a pronoun, cf. *vhledem k* + noun - *vhledem k tomu* (pronoun), *že*. This situation accounts for the fact that, among all expressions consisting of a noun and a primary preposition examined in the corpus, the three compound expressions studied exhibited the highest frequency. The proportions of the instances of a noun and those of a compound preposition containing this noun are, however, different. The most frequent compound form is *v případě* with 7,170 instances. This number corresponds to approximately 50% of the noun instances. On the other hand, the frequency of the noun *vhled* (409 instances) is approximately one tenth the frequency of the expression *vhledem k(e)* (5,289 instances). The frequencies of the noun *základ* and of the compound *na základě* are approximately the same.

3. Discussion

According to J. Sinclair (1991:170) collocation is "the occurrence of two or more words within a short space of each other in a text" while idiom is "a group of two or more words which are chosen together in order to produce a specific meaning or effect in speech and writing" (1991:172). When comparing the usage of the compound units in question, an obvious difference in the idiomatization of each compound is seen. The compounds *v případě* and *na základě*, the structure of which is prep. na + noun-loc. + gen., differ considerably from the compound *vzhledem k*. There are groups of adjectives and pronouns which can be inserted into the structure prep. + [...] + noun-loc. The most frequent adjectives in the case of *v [...] případě* (7,023 instances) are *v každém, žádném, dalším, jiném, opačném, prvním/druhém, horším/lepším případě* (literally 'any, no, following, another, opposite, first/second, worse/better'), and the most frequent pronouns are demonstratives such as *v tomto, takovém, tom případě* ('this, that') and possessive pronouns such as *v jeho, jejím etc. případě* ('his, her' etc.). The adjective and pronoun "input" in the case of the compound *na [...] základě* (681 instances) is quite different. The most frequent adjectives are *na stejném, oboustranném, rovnoprávném, společném základě* (literally 'same, bilateral, equal, common') with possessive and relative pronouns such as *na jeho/ž, jejím/ž, jejich/ž základě* ('his, her, their / whose, which'). The third compound *vzhledem k/e* is fixed and no adjectives or pronouns can be inserted. This compound, it seems, is much older than the previous ones. It is evident that the compounds *v případě* and *na základě* are rather collocations than idioms and the compound *vzhledem k(e)* is the only grammatical idiom among the three compound units studied. These differences, esp. the level of grammaticalization, should be taken into account in the grammatical and lexicographical description.

4. Suggestions for future dictionary elaboration based on the Czech National Corpus

The corpus-based approach can help in specifying and defining the meaning of the phrases studied. The compound preposition *v případě* delimits one activity or action as conditioned on another. The conditional relation can be proved on the corpus concordances.

Typical collocates of the preposition:

<i>potřeba</i> "need"	205 instances
<i>nutnost</i> "necessity"	79 instances

The conjunction *v případě, že/kdy* expresses a conditional meaning as well; nevertheless, a partial temporal meaning can be observed when the conjunction *kdy(ž)* "when" is used, especially in connection with some other temporal data, cf. *Náhrada se poskytuje pouze V PŘÍPADĚ, kdy tento oprávněný nárok byl ... uplatněn do 30. června 1993* "Compensation is provided only IN THE CASE that a legitimate claim was ... made by June 30, 1993".

The compound preposition *vzhledem k(e)* has the meaning of aspect and a partial causal meaning (see above), cf. *ministerstva v ČSSR byla přinucena řešit problémy PCB VZHLEDEM K opakovanému odmítání našeho exportu...* "the ministries of the Czechoslovak Socialist Republic were forced to solve the problems of PCB DUE TO repeated rejection of our export...". According to the corpus data the meaning of aspect prevails.

Typical collocates:

<i>situace</i> "situation"	193 instances
<i>vývoj</i> "development"	58 instances
<i>změna</i> "change", <i>změněný</i> "changed"	36 instances

The compound preposition *na základě* has two meanings according to the SČFI dictionary: delimitation of an activity or action 1) which is a result or consequence of some previous action, circumstances etc.; 2) which serves as the basic approach, principle for some future activity.

Typical collocates:

Ad 1) semantic field "data, observation"

<i>výsledky</i> "results"	163 instances
<i>údaje</i> "data"	47 instances
<i>analýza</i> "analysis"	42 instances

Ad 2) semantic field "certificate, (application) form"

<i>usnesení</i> "ruling"	59 instances
<i>žádost</i> "application"	50 instances
<i>doporučení</i> "recommendation"	44 instances

Transitional group - in legal texts

<i>dohoda</i> "agreement"	275 instances
<i>smlouva</i> "contract"	251 instances

The conjunction *na základě toho, že/co/kdo...* expresses the dependency of one action on another and the corpus instances prove this definition.

In conclusion it can be stated that the expressions examined should be elaborated in a future corpus-based monolingual dictionary either separately in a multi-word lemma or within a clearly-arranged one-word lemma. Each expression should be separated into two sections according to its usage as a preposition or conjunction, and the dictionary should list meta-language definitions, examples, frequency data and information on style.

5. Example of an entry in a future Czech dictionary based on the Czech National Corpus:

v případě (in case of, in the event of) prep. and conj. formal freq CNC 7,170
delimits one activity or action as conditioned on another:

1. prep. (freq CNC 5,378)

obě strany chtějí v případě předčasných voleb vystoupit se společným ekonomickým programem

(in the event of early elections both parties want to present a common economic platform);

typical usage: v případě nutnosti, v případě potřeby ('as needed')

2. conj. (freq CNC 1,792)

v případě, že: V případě, že se nedovoláte přes den, vytočte číslo ve večerních hodinách.

(In case you are unable to get through during the day, please call in the evening)

v případě, kdy: Majetek by mohl být v případě, kdy překupníci nejsou schopni vrátit peníze, zabaven.

(In the event that the middlemen are unable to repay the money, the property can be confiscated.)

Note: similar phrases = v každém, žádném, dalším, jiném, opačném, prvním/druhém, horším/lepším případě (literally 'in any, no, following, another, opposite, first/second, worse/better case'); v tomto, takovém, jeho případě ('in this, his case')

6. References

- Collins COBUILD English Dictionary* (1995). HarperCollins Publishers, London.
- Francis, G. (1993): *A Corpus-Driven Approach To Grammar. Principles, Methods and Examples*. In: Text and Technology. In Honour of J. Sinclair (Baker M. et al eds.). John Benjamins Publishing Company. Philadelphia–Amsterdam, pp. 137–156.
- Halliday, M. A. K. (1993): *Language as System and Language as Instance: The Corpus as a Theoretical Construct*. In: Directions in Corpus Linguistics. Proceedings of Nobel Symposium 82 Stockholm, 4–8 August 1991, ed. by Jan Svartvik. Mouton de Gruyter, Berlin–New York, pp. 61–77.
- Mluvnice češtiny II /Grammar of Czech/*. (1986) Academia, Praha.
- Příruční mluvnice češtiny /A Brief Czech Grammar/* (1995). Nakladatelství Lidové noviny, Praha.
- Sinclair, J. (1991): *Corpus Concordance Collocation*. Oxford University Press.
- Slovník spisovného jazyka českého I-VIII /Dictionary of Literary Czech Language/* (1989). Academia, Praha.
- Slovník české frazeologie a idiomatiky II /Dictionary of Czech Phraseology and Idiomatics/* (1988). Academia, Praha.