

František Čermák  
*Charles University, Prague*

## **Czech Idiom Dictionary**

### **Abstract**

The Dictionary of Czech Idioms is a project of many years' duration which has now resulted in three published volumes (the fourth and final volume is under preparation). Both in its size (containing some 16 000 entries) and design it is rather unique: following descriptions of grammatical and semantic as well as pragmatic features it includes sections giving synonyms, antonyms and other related expressions, information on etymology and equivalents in English, German, French and Russian. Moreover, all of the idioms covered in each volume are also grouped semantically into a parallel semantic dictionary (register).

### **1. Introduction: plan and general character**

As a West Slavonic language neighbouring primarily, but not exclusively, with German, Czech offers a field of study of some interest to those who want to go into such matters as cultural influences in this part of Europe. As a rather rich display of this is to be found in Czech idioms, and as those studying the language have always had some difficulty in mastering them, a project to develop a dictionary of contemporary Czech idioms was started some time ago. The intention of the team of authors, led by J. Hronek, the late J. Machač and myself, has been to make the dictionary as exhaustive as possible, adding some new features that would make it easier and more attractive to use, such as synonyms and antonyms, equivalents in four other languages and also a supplementary semantic dictionary which enables one to look up idioms according to their meaning. Because of these, and some other features, this dictionary (Čermák, Hronek, Machač et al.:1983, 1988, 1994) represents a fairly unique, yet also somewhat untypical step in exploring lexicographical possibilities which, so far, have not been used very much. In its overall character, the dictionary aims to be both a reliable academic book and, at the same time, a multi-purpose reference-book for the general public. It seems inevitable, that in dictionaries of languages spoken by a relatively small number of people one has to aim at as large a readership as possible and that is why there is not much room for highly specific single-purpose dictionaries. Both in the techniques devised and used and the scope of the language covered, this dictionary has very few parallels, even abroad. And judging by the reviews of the first two volumes, it seems, on the whole, that it has been well received.

The bulk of this large, long-term academic project, split into three volumes, is now finished and published. In its present form, the dictionary records some 16 000 contemporary Czech idioms excerpted from all sources available, including manual notes of idioms found in spontaneous language, and is thus one of the largest and relatively exhaustive dictionaries of its kind.

As the identification of potential idioms can never be completely reliable and exhaustive if it is based on intuition and semantics (or metaphor) only, a set of simple rules was worked out, based on a theory of idioms (Čermák:1982, Filipec-Čermák:1985, Čermák:1988) whose elaboration coincided with that of the dictionary project itself. Thus any (1) *collocation (combination)* of at least *two lexemes* has been included which is (2) *stable* and in which (3) *at least one component (lexeme) collocates, in the same function or sense, with the other components in a unique or severely limited manner*. This view of the idiom, stressing collocational **restriction** and **anomaly** of its components, is based on two premises, namely that words collocate basically with other words in a rather free way, though limited by the speaker's choice, and that these collocations are rule-governed. In idioms, both of these aspects are limited to a very high degree or, more frequently, absent. As a consequence, idioms display a number of restrictions and anomalies in their textual behaviour and, also, they often seem to be marked by some form of semantic non-additiveness, if we compare their integral meaning with that of their components. These criteria, presented here in a somewhat simplified form, were used to select and isolate items which underwent a lexicographic treatment (for the principles of which, see also Čermák:1994). The preparatory stages included both verification of form, meaning and usage, selection of variants as well as standardization of form, where such basic questions had to be answered as 'Are these forms two variants of the same idiom or two idioms?'

Finally, the lexicographic material was split into a number of **structural types** based on their joint function and the type of components represented, such as Verb - Noun, Adjective - Noun etc. Major types obtained included comparative idioms (comparisons), nominal, grammatical and verbal idioms. These formed a homogeneous basis for the lexicographic treatment proper and, at the same time, the functionally homogeneous content of the actual published volumes.

The notorious problem of where and by what key the idioms described should be represented in the **alphabet** was solved linguistically, using the nominative power of word classes, which is supported by frequency as well; only in cases of serious doubt or variants, then, can more than one form be found in more than one place in the alphabet. Thus the following scale has been employed for alphabet treatment: *noun* → *adjective* → *verb* → *adverb* → *any other word*; in the case of comparisons this priority sequence starts after the comparative formal signal (Czech *jak(o)*, English *like/as*). Accordingly, idioms are always to be found under their first noun and if that is absent they are found under their first adjective; if these two are not present, it is the first verb and its alphabetical place that is used and so on. For large families of idioms, such as those containing the same body-part noun (e.g. *hand, head, foot* etc.), a secondary subclassification based on the same criteria is used to facilitate orientation within each group of, perhaps, a hundred idioms with the same noun etc.

In the final stages, two other major decisions and a subsequent classification of data had to be made, concerning the relative frequency of idioms

and their transformational relation. It was decided that only idioms with reliable high frequency would get the full lexicographic, or, rather 'phraseographic' treatment and description, whereas the information offered for the rest (i.e. about half of them) would be reduced. This was a great help since, for example, it may be quite difficult to determine the scope of usage of some rare idioms; also, this meant a substantial reduction in the number of pages which went to print. Such a reduction has taken place in the treatment of transformationally linked idioms, too, where in such correlated cases as *take it into one's head* and *put it into one's head* the latter, which happens to differ from the former in having a rather transparent causative character, contains a reference to the former as the only information available.

## 2. Idiom dictionary sections

The standard design of dictionary entry includes the following sections:

- A (1) lemma,
- (2) stylistic information,
- (3) grammatical information,
- (4) transformations,
- B (5) context, valency and function,
- (6) meaning + (7) exemplification of usage,
- (8) additional notes,
- C (9) synonyms, opposites and related expressions,
- D (10) equivalents in four languages.

In the following, some of the features of these will be briefly mentioned.

**Lemma** (1) records the neutral, standard form of the idiom with all of its components and parts, which are specifically and explicitly marked if they may undergo grammatical modifications in actual use, such as tense, number, person etc. Paradigmatic variants of the idiom, based on substitution, are indicated by slashes (e.g. *keep a cool/clear head*) while syntagmatic, linear variants are introduced in parentheses (e.g. *play (it) safe*). An important part of the lemma is its valency which is formally given here, too, employing a number of means, the most frequent ones being indefinite pronouns (such as *sth* in *bring sth to a head*). For purely formal reasons, the word under which the idiom is alphabetically registered has a somewhat different typographical appearance.

**Stylistic information** (2) gives the idiom's dominant type of textual usage (most of the Czech idioms are colloquial) and type of functional expressiveness. Here, as well, some general information about its frequency can be found.

**Grammatical information** (3) is generally introduced by a bold zero-sign to signal that the nature of this information is negative, i.e. all of the

grammatical categories that are not, as a rule, realized with the idiom are explicitly given, such as plurals, persons, tenses, passives, moods, negation, questions etc. This explicit enumeration of the idiom's negative markers and restrictions, giving the user information on how to use the idiom in text, represents, in fact, a full explanation of what is being signalled in general by special marks in the lemma.

Finally, these formal sections of the entry are supplemented by information on **transformations** (4), if any. In addition to such (quasi)transformations whose function is the same, i.e. both the base and the transformation have, for instance, a verbal character (see above), heterofunctional transformations are also included, corresponding to such English examples as *hand in hand* and *go hand in hand*. At the same time, any such transformation serves as a specific reference to another entry, possibly across the volume boundary.

The two following major sections, i.e. that of the **context, valency and function** (5) and **meaning** (6) are correlated in that the first is a general framework for the second. This means that first of all the idiom's typical contexts and situation of use are given including main generalized *actants* or participants related to the idiom (such as *man* in general or more specifically *adult/child, man/woman, chief/subordinate* etc., both, for example, as a contextual subject or object) which are set in a typical situation in which the idiom is used. This situational frame specified by participants is completed by an indication of relevant generalized relations, attitudes and motivation of the participants, or, if the idiom is basically a description of someone's quality, by that general aspect of him or her (such as *body, behaviour* etc.) of which the idiom is a description. Only then, separately, is the idiom's meaning given; it is best seen as a particular specification of the valency's general framework or skeleton formed by participants etc. given in the preceding part. Both context and meaning are separated formally as well, by the use of parentheses for the former. Special care has also been given to the indication of ways in which potential shades of meaning in some of the idiom's less frequent senses are to be signalled, and to outstanding pragmatic features of usage including evaluative aspects. Lexicographic treatment of both sections is based on a restricted and partly formalized metalanguage which, among other things, has a syntax of its own.

This double information, i.e. about context and meaning, is specified in an illustration of the idiom's usage in the **exemplification** (7) section, where typical examples are briefly given. Finally, the section called **notes** (8) could represent the end of the entry proper; here, a number of additional notes is given dealing with the idiom's form (e.g. comments on variants), usage and function (e.g. major speech acts) and, also, with its etymology, if this is known, etc.

In addition, the dictionary entry offers two other useful and unusual sections which could easily be published separately, as independent dictionaries. In one of them, a selection of **synonyms, opposites and related expressions** (9) is given to help the user and to give some immediate information about related

idioms existing side by side with the one under scrutiny.

In the second, the idiom's **equivalents** (10) in English, German, French and Russian are given, making it, effectively, a set of four bilingual dictionaries of this sort. Close attention is paid to the status of the equivalent, where several types of equivalents are distinguished and annotated.

### 3. Semantic dictionary

Finally, a dictionary of this size and complexity (2394 pages altogether) could not be written without using a number of additional tools during its compilation. One of these tools has been modified and standardized to such an extent that it now forms an independent **semantic dictionary**. Being organized semantically, it forms an onomasiological counterpart to the main alphabetical section, i.e. a semasiological dictionary or a thesaurus. All of the idioms of the alphabetical part are reclassified and grouped here into common semantic classes where idioms of the same type and meaning are to be found. From these classes the user can make a reliable selection of the idiom he finds most suitable to his purposes, since all possibilities available in the language are given. All of the other lexicographic information, if needed, can be easily found in the alphabetical part of the same book, where a reference is given in each idiom entry pointing to a semantic class. Thus both dictionaries are closely linked.

As the Czech Dictionary of Idioms (*Slovník české frazeologie a idiomatiky*) is intended to serve more than one purpose, a theoretical linguistic description of the idioms covered by each volume, is also given. Currently, the last (i. e. fourth) volume of this dictionary of Czech idioms, which will cover all kinds of sentential idioms and will wind up this long-term project, is under preparation (now with the help of a computer programme).

### 4. An example of an entry

#### **být° (hubený) jako l u n t**

(kol; nepřizn. posm) 0 ot, neg, imp, imp neg, kond, fut, 1.sg a pl Nom někdo j. l., hubený j. l.

(*Dospělý i odrostlejší dítě postavou:*) *být velmi hubený, vychrtlý, popř. podvyživený.* ◊ *Lunt* (doutnák) býval dlouhý konopný provázek. □ **S**

h u b e n ý, být (hubený) j. koza/špejle, být j. za groš kudla, mohl by se svlékat za bičem **A t l u s t ý**, být j. kulička, být širší než delší, mít cejchu, hřeje ho sádlo **Cf h u b e n ý, s l a b ý**, být j. sušinka/pápěrka, z kalhot by ho vyklepal, když se svleče j. když odejde, žebra by mu mohl počítat, být kost a kůže

• **A** be as thin as a rake/lath **N** dürr w. ein Stecken sein **F** être maigre c. un clou **R** chudoj k. ščepka

Sections of the entry:

-Morphologically modified component of the lemma (1):	<b>být*</b>
-Syntagmatic variant of the lemma (1):	<b>(hubený)</b>
-Word indicating place in the alphabet (1):	<b>l u n t</b>
-Stylistic information (2):	(kol; nepřizn, posm)
-Grammatical information (3):	<b>0</b> ot, neg, imp, imp neg, kond, fut, 1. sg a pl
-Transformations (4):	<b>Nom</b> někdo j.l., hubený j. l.
-Context, valency and function (5):	( <i>Dospělý iodrostlejší dítě postavou</i> )
-Meaning (6):	<i>být velmi hubený, vychrtlý, popř. podvyživený</i>
-Exemplification (7):	-
-Additional notes (8):	◇ <i>Lunt</i> (doutnák) býval dlouhý konopný provázek.
-Reference to semantic dictionary (9):	h u b e n ý
-Synonyms (9):	□ <b>S</b> být (hubený) j. koza/špejle, býtj. za groš kudla, mohl byse svlékat za bičem
-Opposites (9):	<b>A</b> být jako kulička, být širší než delší, mít cejchu, hřeje ho sádlo
-Related expressions (9):	<b>Cf</b> být j. sušinka/pápěrka, z kalhot by ho vyklepal, když se svleče j. když odejde, žebra by mu mohl počítat, být kost a kůže
-English equivalent (10):	<b>E</b> be as thin as a rake/lath
-German equivalent (10):	<b>G</b> dürr wie ein Stecken sein
-French equivalent (10):	<b>F</b> être maigre comme un clou
-Russian equivalent (10):	<b>R</b> chudoj kak ščepka

#### References

- Čermák F., Hronek J., Machač J. (eds.) 1983. *Slovník české frazeologie a idiomatiky. Přirovnání*. Praha: Academia.
- Čermák F., Hronek J., Machač J. (eds.) 1988. *Slovník české frazeologie a idiomatiky. Výrazy neslovesné*. Praha: Academia.
- Čermák F., Hronek J., Machač J. (eds.) 1994. *Slovník české frazeologie a idiomatiky. Výrazy slovesné A-P, R-Ž*. Praha: Academia.
- Čermák F. 1982. *Idiomatika a frazeologie češtiny*. Praha: Univerzita Karlova.
- Čermák F. 1988. "On the Substance of Idioms". *Folia Linguistica* XXII: 413-438. Berlin: Mouton/de Gruyter.
- Čermák F. 1994. "Frazeografie" in F. Čermák, R. Blatná (eds.), *Manuál lexikografie*. Praha: H & H (in print).
- Filipec J., Čermák F. 1985. *Česká lexikologie*. Praha: Academia.

