
Although the phraseology of a small language like Czech may be of relatively limited general interest to speakers of other languages except to a small group of specialists, the qualities of the four-volume dictionary under review are such that it definitely merits a closer look. It is monumental not just in sheer size, with the four volumes totalling over three and a half thousand pages (and more than 35 000 entries), but especially in the breadth and depth of the information provided. Together, the four volumes, using a remarkably uniform format, strive for nothing less than the full, comprehensive description of a language’s contemporary phraseology, unparalleled as the authors claim in any other language. The English counterpart (probably unsurpassed by any other English idiom dictionary so far) comparable to the dictionary under review appears to be the *Oxford Dictionary of Current Idiomatic English* by Cowie et al., which deals with multiword verbs (Vol.1, 1975) and phrase, clause and sentence idioms (Vol.2, 1983). Although it is less than half the size (15 000 entries), the range of expressions included and particularly its theoretical background and detailed linguistic description of the expressions provide a valid basis for comparison.

The aims of the Dictionary of Czech Phraseology and Idiomatics (DCPI in the following) are as ambitious as the range of information it provides. The DCPI is intended to serve as both active and passive mono- and multilingual dictionary, targeting all types of general and specialized users, such as students, teachers, and translators, by offering them a rich source of phraseological data for practical purposes of text comprehension and production, but also researchers by supplementing the data in each volume with a linguistic description of the type of idioms covered. The dictionary thus responds to the regrettable situation in this area of Czech language description: the obsolescent, insufficient and sparse reference books on idioms and the absence of theoretical treatment of the subject.

The dictionary’s publishing history reflects the efforts that have gone into its compilation. The first volume on similes came out nearly three decades ago in 1983; the second volume on nonverbal phrasemes (nominal, adjectival, adverbial and grammatical) was published in 1988; the third volume on verb-based phrasemes appeared in 1994 (originally in two parts, AP, 757 pp., and RŽ,
and finally the fourth volume dealing with propositional phrasemes/idioms appeared in 2009, together with the previous three (revised, updated and supplemented) as a complete set.

As might be expected in a work spanning more than a quarter of a century, the team of collaborators have seen many changes. The linchpin of the whole project has been František Čermák, who has acted as the editor-in-chief (together with Jiří Hronek and Jaroslav Machač in the first three volumes, alone in the last one), editor, compiler and author of the Introduction and The Principles of the Dictionary’s Compilation and Use in the front matter and of the theoretical description of the given type of phraseme/idiom (with Libuše Kroupová in Vol. 2) in the back matter in each of the four volumes. In addition to these three, sixteen more compilers appeared in the team at one time or another (R. Blatná, V. Červená, M. Churavý, K. Filipová, J. Holub, L. Janský, M. Kopřivová, J. Luttererová, V. Mejsťík, J. Sedláková, M. Šára, V. Šustalová, Z. Švamberková, A. Trnková, E. Vejvoda, J. Zima). The team includes linguists from the Faculty of Arts, Charles University in Prague, and partly from the Institute for Czech Language, Czech Academy of Sciences. The phrasemes are provided with equivalents in four European languages (English, German, French and Russian) which were supplied by the respective native speakers—Bohemists (D. Short, J. Ostmeyer, M. Gasnier, P. Pognan, M.-L. Jouannaux, L.N. Beloruss-Beloshevskaya, T.N. Danilova, O. Fedosov, L. Stepanova).

The theoretical framework underlying the selection and classification of the phrasemes/idioms in the dictionary set is presented in the grammatical overview for the given type of phraseme at the end of each volume and in Čermák’s (2007) collection of papers on phraseology (reviewed in IJL by Bozděchová 2008). Lexical items described combinatorially as phrasemes and semantically as idioms (typically non-compositional, lexico-semantically, grammatically and collocationally anomalous) are analyzed by Čermák in terms of their components and their formal, semantic and collocational features (including such aspects as degree of fixedness and variability). On the basis of this they are classified from three aspects: structural (in terms of word-class combinatorics or word-class-clause combinatorics), semantic (thematic, sense-relational) and functional (communicative, onomasiological, pragmatic, stylistic functions whose presence and specific configuration determine the degree of idiomaticity). One other important aspect of phrasemes/idioms examined by Čermák is the possibility of their variation and transformation/transposition in text. He divides the transformations into interstructural and interlevel (word-class or phrase ones, e.g. nominalization), introstructural (e.g. Aktionsart shifts due to the use of lexical variants of the same word-class, e.g., from ‘inchoative’ to ‘durative’) and contextual (adapting the form and function to the context, but not affecting the identity of the phraseme).
For lexicographic purposes the obvious starting point is the author’s formal-structural classification of phrasemes and idioms based on the evidence provided by the input components. (In the following, English equivalents of the Czech phrasemes/idioms will be used for the sake of simplicity.) Čermák uses a tripartite classification distinguishing between lexical, collocational and propositional phrasemes. As the dictionary set under review deals with only the latter two, it is enough to say that lexical phrasemes are complex one-word lexemes (derivatives, compounds) in which the morphemes produce an opaque combination (e.g. nondescript, outpace, cutthroat, trainspotting). Also, recognizing the scalar nature of the lexico-semantic and grammatical anomaly of phrasemes/idioms (manifested by severely restricted collocability and paradigmatic commutability of their components and restricted applicability of grammatical categories and syntactic transformations, such as passivization, etc.), Čermák allows for the category of quasi-phrasemes and quasi-idioms in his classification. They are situated in the transition zone between the regular and irregular language and are typically marked by relative restrictedness in terms of collocability (wide open; verbonominal expressions draw/pay/turn attention to, etc.), semantics (e.g., strictly defined technical terms: wide ball, dead travel) and function (in comparison with) or of a combination of these factors.

Čermák views a prototypical phraseme/idiom basically as a binary syntagma (subordinate, coordinate) involving the combination of the four lexical word-classes, noun N, adjective A, verb V, adverb Adv, bearing the meaning in the lexicon, which produce 16 collocational variations, V-N, A-N, N-A, N-Adv, Adv-N, V-A, V-Adv, Adv-V, A-Adv, Adv-A, N-N, V-V, A-A, Adv-Adv, two of them resulting in a propositional syntagma with the nominal component in subject position, N-V, A-V. The relationship between the components can be explicitly marked by a preposition or conjunction (the noun phrase may be replaced by a prepositional phrase, etc.). However, function words, pronouns, conjunctions, prepositions, particles, etc., may not only participate in collocational phrasemes, they can form them on their own as well. The head of the phrase can be realised also by a paradigmatically related word-class (e.g. the noun by a pronoun or numeral), and the basic syntagma (V-Adv) can be extended through a structural transformation to include more components (V-Adv-Adv: to be neither here nor there). Actually, structures of more than two words are seen as subtypes of the binary ones. In addition, there are interstratal phrasemes/idioms, combining components from different levels (lexeme/phrase + clause: e.g., look as if butter wouldn’t melt in one’s mouth, as a variant of the V-Adv type).

Collocational phrasemes/idioms may be broadly divided into verbal (verb-based, V-N, V-A, V-V, V-Adv; some of them verge on verbonominal quasi-phrase/semes) and nonverbal (noun-, adjective- or adverb-based). In addition, they include some special categories which cut across the verbal-nonverbal division, particularly similes (V-like/as if, N/(as) A/Adv
as/like) and binomials (N-conj-N, alpha and omega; N-prep-N, day after day; N-N, hocus-pocus; A-conj-A, short and sweet; V-conj-V, live or die; Adv-conj-Adv, now and then, etc.). Another category subsumed by Čermák under collocational phrasemes is that of grammatical idioms, or multi-word lexemes functionally equivalent to single-word grammatical word-classes, including complex prepositions (as to, on behalf of), conjunctions (as if, even though, in order that, as long as), pronouns (anyone who) or particles (pragmatic markers, all right, on the contrary). The concept may be extended to the congealed let’s and the invariably following question-tag shall we, etc.

**Propositional phrasemes/idioms** are analyzed by Čermák according to the number of propositions they involve (mono- and polypropositional) and their monologic (monosubjective) or dialogic (intersubjective) nature. Monopropositional phrasemes consist of regular simple (one-clause) sentences, but also of what are sometimes called verbless (or ‘minor’) sentences, capable of functioning on their own (No pain, no gain). They do not require a response from the hearer; their components include both lexical words (and their collocational extensions) and/or (only) function words. Polypropositional phrasemes are subclassified into complex(-compound) sentences not intended to elicit response from the hearer and into formulaic sentences (dyadic propositions) exchanged by interlocutors in conversation (cf. adjacency pairs). In addition Čermák mentions other possible classifications (interjectional, Dead right!, nominal, Heads or tails?, imperative types, Don’t even think about it!, etc.), also functional (vocative, You must be joking!, contact, See you later!, thematic, Once upon a time...), or metalinguistic, It’s on the tip of my tongue) and pragmatic (factual, declarative, voluntative, expressive and emotional).

These categories cover all kinds of phrasemes known by their traditional fuzzy labels such as proverbs, sayings, saws, adages, catch-phrases, catchwords, bywords, winged words, maxims, slogans, mottos, morals, precepts, gnomes, epigrams, aphorisms, apothegms, witticisms, quips, but also clichés, quotations, etc.

These theoretical starting points determine the lexicographic decisions at all levels and unify the dictionary’s four volumes. Although each volume is relatively self-contained, all of them are interrelated through sharing a similar framestructure (macro- and microstructure), dictionary components (design of the dictionary proper and the outside matter) and their features. The latter make up the following general model: 1. Front matter - Introduction, Principles of Compiling and Using this Dictionary (and its three sections, Types of phrasemes described and their selection, Dictionary entries and their features, Semantic index), Abbreviations and Symbols; 2. Dictionary of Phrasemes/Idioms; 3. Back matter - Semantic index (onomasiological dictionary), Theoretical description of the type of phraseme/idiom covered by the volume.
In a dictionary of idioms the macrostructure and the selection of items for the entry list are always a formidable challenge. The authors make use of both possible modes of access to the information, i.e., alphabetical (semasiological) and semantic (onomasiological). The problem of alphabetical arrangement of the phrasemes/idioms following from the fact that by definition they consist of two or more words is resolved by the principle of ordering by the first noun, in the absence of a noun by the first adjective, then by the first verb, and finally the first adverb, and, if the phraseme features no lexical word, by any first word. This simple principle is applied systematically in all four volumes and has made the usual doubling of the alphabetical arrangement (alphabetical index) redundant.

Likewise, the onomasiological supplement, in the form of alphabetically arranged semantic concepts (key words, hyperonyms, general synonyms of the lemmas), is provided in all four volumes (it is omitted only for grammatical idioms) and substantially extends the search possibilities. It started as one of the additional tools during the compilation which was gradually modified and standardized to such an extent that it has become an integral part of the dictionary. For instance, in the 4th volume dealing with sentential proverbs the semantic index forms close to 1/5 of the dictionary proper (176 pp. : 1024 pp.).

The entry list features phrasemes/idioms selected from a wide variety of sources – the available lexicographic sources, conversation books, the phraseological archives of the Institute for Czech Language, excerption from modern novels, spoken discourse recordings, usage questionnaires, association tests, etc. In addition to fixedness and familiarity of the expressions, an important criterion for inclusion was corpus-based frequency (with only some types of syntagmatic and paradigmatic variants being excluded). The overarching general policy for all volumes is to cover phrasemes/idioms current and common in contemporary language. As of the second volume and especially in the last one the main source of data was the Czech National Corpus (see http://ucnk.ff.cuni.cz) with several hundred million word-forms.

The basic (micro)structure of the dictionary entry is again standardized for all volumes and consists of four sections or paragraphs (cf. the simile entry (být) utahaný / ušťaný j. kůň in Figure 1): 1. Lemma section in bold face indicates parts of the phrase which are grammatically (brackets, °) and lexically variant (slash), often shortening the frequently repeated comparator (j. = jak, jako, ‘like’); 2. Indented formal description provides stylistic information in the brackets (kol. = colloquial; důraz = emphasis) followed by grammatical information; boldface 0 indicates that the following grammatical features/transformations (ot = question, imp = imperative, neg = negation) do not obtain with the phrase; boldface Nom indicates the case of the referent in subject position; 3. Unindented semantic description is subdivided into (a) contextual information (in brackets: characterization of agent, causes of state, etc.) followed by (b) definition, (c) additional usage information, here the plural form.
of the key word ( ), (d) synonyms/cross-references to semantic index (■), (e) antonyms (A), (f) parallel terms functionally broadly equivalent (Cf); (4) English/German/Russian equivalents. The entries come in three forms: full (see the example), reduced (typically to lemma and definition only) and entry-status cross-reference.

Regardless of the standardized format each of the volumes in the set displays certain specific features following from the type of phraseme/idiom covered. Volume 1, dealing with similes, is the smallest (with 2050 full and 350 reduced entries, not counting entry-status cross-references) and oldest of the set. The reason to start with similes makes sense as they represent semantically and formally specific and easily demarcated collocational subcategory of manageable size. The typical structure for the simile (V-A-comparator-N) is shown by the above example. The fact that the volume came with a sophisticated, fully fledged format that could be successfully applied to the other volumes basically unchanged is explained by the genesis of the volume which had been in the making from the 1960s, but with theoretical preparation for more than forty years. This also explains why so few additions were necessary in this 2nd edition (58 entries).

By contrast, Volume 2 on nonverbal expressions could fully profit from the use of a corpus in this new edition and was expanded by 400 entries, bringing the number of (full and reduced) entries to slightly more than 4000. The most frequent structures are of the types A-N (see the entry dobré bydlo in Figure 1) and N-(prep/conj)-N (secret of success), supplemented by a number of less frequent types, e.g., A/Adv-A (wishy washy), and their variants and extensions. In addition, the volume contains a list of 743 complex grammatical phrasemes, divided into a section on prepositions (in contrast to) and another one dealing with conjunctions and particles (no matter how, inasmuch as, well, of course). The entries of grammatical phrasemes are reduced to the lemma and a paragraph collapsing stylistic information and semantic description, including

**Figure 1:** The entries (být) utahany / ušťaný j. kûň (be dog-tired; Vol. 1, similes) and dobré bydlo (an easy life; Vol. 2, nonverbal expressions)
contextual and pragmatic delimitation (supplied with a synonym), sentential examples and grammatical specification (♦). No foreign equivalents are provided.

The third volume covering verbal expressions is almost equal in size to the largest 4th volume, but contains the greatest number of entries, close to 10,000 verbal phrasemes (full and reduced), and about the same number of quasi-phrasemes. The expressions include a wide range of syntagmas, V-N/NP (have a nose for), V-A (see the entry **být na obě oči slepý** in Figure 2), V-Adv (know better than to), V-V (let st ride), V-clause (never know when to stop), V-function word (be for) and a number of variants and extensions. Because of the expressions sharing the same key element (cf. over 200 phrasemes with the word ‘head’), the ordering system had to be further refined and specified (hierarchical and level ordering, number ordering, etc.) to deal with such cases. As was mentioned above, the onomasiological dictionary is the most extensive compared to those in the other volumes and takes up 20 per cent of the text. This volume also tackles a rather specific sociolinguistic issue of whether the idioms should be presented in standard Czech (as preferred by the Institute for Czech Language) or in their authentic colloquial form. The authors have decided for the latter approach. Thus the dictionary reflects the on-going heated debate between the guardians of standard Czech and its purity and those who point out that the gap between the codified literary norm and the spoken form of Czech used by the generality of speakers in most situations is widening and argue for a compromise solution.

The greatest challenge for the authors was inevitably the fourth volume compiled exclusively by the Faculty of Arts team, dealing with propositional

![Figure 2](http://ijl.oxfordjournals.org/)

**Figure 2:** The entries **být na obě oči slepý** (cannot see for looking; Vol. 3, verbal expressions) and **Netláč na pilu!** (Don’t push too hard; Vol. 4, sentential expressions)
Propositional phrasemes are enormously varied structurally and functionally (involving verbal and nonverbal clauses, dialogic and non-dialogic, interjectional and adverbal, etc.). Their ordering follows the general principle (N→V→A→Adv→first word) but occasionally requires specific solutions. As they are fully fledged utterances, their entries also include prosodic description (introduced by ~ in a square), a novel feature not found in any other dictionary of idioms I know of. The volume is the largest of the four by page count, presenting some 9000 entries. The onomasiological section (totalling some 14 per cent of the whole text) must have been the most difficult to compile compared to the other volumes due to the complex semantics of the propositions, requiring the construction of an ingenious descriptive apparatus (consider the difficulties posed by expressions such as Now you’re talking!).

In a set of dictionaries of this size, addressing the whole spectrum of phrasemes/idioms, the reader is bound to come across minor inconsistencies or debatable points. For instance, the theoretical definition of collocational phrasemes/idioms, underlying Volume 2, as binary structures, i.e. two-element variations (of open-class items N, V, Adj, Adv) with repetitions, does not explicitly provide for one specific subtype, although it is represented in the dictionary, i.e. an open-class plus closed-class-item structure (e.g., prep-N: v kostce, in a nutshell, na místě, on the spot).

From a lexicographic point of view, it is a pity that foreign-language equivalents are not supplied systematically, as happens with the above common expressions in a nutshell and on the spot. The authors claim that equivalents accompany (only) phrasemes with high frequency. Given the notoriously low frequency of idioms in general, drawing the line between high and low frequencies is problematic, as is, in fact, the very application of the frequency criterion. The best policy would be to provide equivalents wherever possible. Multilingual equivalents are certainly an interesting innovation, though translation of idioms is so tricky, context-sensitive, variety-dependent, etc., that respective bilingual idiom dictionaries would be necessary to cover the ground adequately (and remove discrepancies in description such as the copula being parenthesized in the Czech expression but not in the equivalents in the other languages, see Figure 1).

While the dictionary set is consistent in the use of stylistic labels, the information on the source of the quotation (and the lemma) in the entries is generally missing or rather the user will find it only in the front-matter introduction stating in which dictionaries or corpora the expressions were found. In Cowie et al. (1975, 1983), by contrast, the reference to the literary or newspaper source (together with the date) is provided in the entry wherever possible, which gives the user at least some idea about the temporal and stylistic context in which the item was used. The DCPI mentions the source only in the entries on sentential expressions (Vol. 4) drawn from the Bible or collections of proverbs.
(Čelakovský, etc.). They appear in just a fraction of the Volume 4 entries, though.

Interestingly, the very strength of the DCPI, its meticulous scholarly description, raises the question of what its target users and uses really are. While Cowie et al. is meant for learners of English and English teaching, which clearly defines its target group and above all fully justifies its detailed and comprehensive grammatical and lexical descriptive apparatus, in the case of the DCPI the pedagogical uses are severely limited (with few foreign learners of Czech around). The user group is thus restricted mostly to dedicated native speakers of Czech with genuine interest in this layer of vocabulary (and willing to pay the rather prohibitive price for the dictionary). Also, in purely practical terms, owing to page layout and the quality of paper, the DCPI, particularly Volumes 3 and 4, are quite bulky and heavy, which means that browsing in them is not an easy task. This handicap, however, should be removed by the electronic version under preparation.

None of these reservations, however, detract from the overall achievement of the authors. It is quite impressive, not only considering the scope of the dictionary set, encompassing all types of idiom, but also because of the accuracy and comprehensiveness of the lexicographic and linguistic description of the expressions, the overall amount of information included, or the formal consistency of presentation resulting in a generally easy navigation through the dictionary. The selection of the phrasemes/idioms appears to be balanced, utilizing everything that the state-of-art theory and technical means (corpus linguistics procedures) can offer. The dictionary introduces a number of novel features (paradigmatic and syntagmatic variants; prosodic properties of sentential phrasemes; focus on types of phrasemes previously ignored, such as grammatical idioms; foreign-language equivalents, etc.).

Obviously, the area of phraseology is so extensive as well as fluid and fuzzy that a fully exhaustive coverage even in a four-volume dictionary cannot be realistically expected. However, the DCPI comes as close to it as is possible for a paper dictionary and in the present state of affairs. Once the planned computerization of the dictionary goes ahead, offering the possibility of continuous on-line expansion and updating (and eliminating most of the problematic points mentioned above), the DCPI, exceptional as it is now, will become a phraseological database in a league of its own.

References

A. Dictionaries

B. Other literature


AlesˇKle´gr
Charles University in Prague, Czech Republic
ales.klegr@ff.cuni.cz
doi:10.1093/ijl/ecq021
Advance access publication 30 June 2010


Die vorliegende Rezension ist sowohl aus der Perspektive der deutsch- und tschechischsprachigen Wörterbuchbenutzer als auch aus der metalexikographischen Perspektive geschrieben.¹