

Ver. 2. 4/9/2009

## All abstracts, phraseology 2009 in Japan

# Lexicography

František Čermák

Charles University, Prague, Frantisek.Cermak@ff.cuni.cz

## Types of Language Nomination: Universals, Typology and Lexicographical Relevance

### 1. Introduction: Dictionaries and Their Lemmata

- Language nomination draws on morphemes and their combinations or words and their combinations: single- word or multi-word lexemes, stable sentences - the Prague Circle linguistic tradition
- Dictionaries' approach is rather uneven: multi-word nominations are often underrepresented

### 2. Types of Nomination

- S. Ullmann (Ullmann 1966, 224, 232): search for semantic universals
- Four major types of nomination formation or word formation:
  - (A) *derivation*, (B) *compounding*, (C) *polysemy*, (D) *collocation*
- Treatment of *polysemy*: no consensus in dictionaries
- Background Prague typology: (Vladimír Skalička 1979)
- Types and languages: *isolating* or analytical type (English), *agglutinative* type (Finnish) and *inflectional* type (Czech); moreover: *intoflectional* type and *polysynthetic* type
- Analysis of ten typical nouns in large dictionaries of three languages:
  - New Shorter Oxford Concise Dictionary on Historical Principles I-II (NSCD)* for English, *Nykysuomen sanakirja I-III (NS)* for Finnish and *Slovník spisovného jazyka českého I-IV (SSJČ)* for Czech, namely names of (a) body parts, (b) animals, and (c) nature objects

|   |                               |                                    |
|---|-------------------------------|------------------------------------|
| <b>a</b> <i>hlava, nos, oko, jazyk, srdce;</i>  | <b>b</b> <i>pes, kočka;</i>   | <b>c</b> <i>strom, kámen, voda</i> |
| <b>a</b> <i>head, nose, eye, tongue, heart;</i> | <b>b</b> <i>dog, cat;</i>     | <b>c</b> <i>tree, stone, water</i> |
| <b>a</b> <i>pää, nenä, silmä, kieli, sydän;</i> | <b>b</b> <i>koira, kissa;</i> | <b>c</b> <i>puu, kivi, vesi</i>    |

### 3. Polysemy and Its Dictionary Representation in Three Language Types

English (isolating)

Finnish (agglutinative)

Czech (inflectional)

|          |               |                 |              |             |              |            |
|----------|---------------|-----------------|--------------|-------------|--------------|------------|
| <b>a</b> | <i>head</i>   | <i>50 + 22V</i> | <i>pää</i>   | <i>33</i>   | <i>hlava</i> | <i>14</i>  |
|          | <i>hear</i>   | <i>32 + 4V</i>  | <i>sydän</i> | <i>29</i>   | <i>srdce</i> | <i>8</i>   |
|          | <i>tongue</i> | <i>27 + 12V</i> | <i>kieli</i> | <i>6</i>    | <i>jazyk</i> | <i>3</i>   |
|          | <i>eye</i>    | <i>13 + 1V</i>  | <i>silmä</i> | <i>14</i>   | <i>oko</i>   | <i>5</i>   |
|          | <i>nose</i>   | <i>11 + 14V</i> | <i>nenä</i>  | <i>5</i>    | <i>nos</i>   | <i>2</i>   |
| <b>b</b> | <i>dog</i>    | <i>29 + 9V</i>  | <i>koira</i> | <i>7</i>    | <i>pes</i>   | <i>5</i>   |
|          | <i>cat</i>    | <i>13</i>       | <i>kissa</i> | <i>7</i>    | <i>kočka</i> | <i>7</i>   |
| <b>c</b> | <i>tree</i>   | <i>18 + 5V</i>  | <i>puu</i>   | <i>10</i>   | <i>strom</i> | <i>1</i>   |
|          | <i>stone</i>  | <i>33 + 8V</i>  | <i>kivi</i>  | <i>9</i>    | <i>kámen</i> | <i>4</i>   |
|          | <i>water</i>  | <i>22 + 16V</i> | <i>vesi</i>  | <i>21</i>   | <i>voda</i>  | <i>2</i>   |
| Average: |               | <b>34,9</b>     |              | <b>13,6</b> |              | <b>5,1</b> |

#### 4. Derivation and Its Dictionary Representation in Three Language Types

|          | English (isolating) |            | Finnish (agglutinative) |             | Czech (inflectional) |           |
|----------|---------------------|------------|-------------------------|-------------|----------------------|-----------|
| <b>a</b> | <i>head</i>         | <i>7</i>   | <i>pää</i>              | <i>52</i>   | <i>hlava</i>         | <i>37</i> |
|          | <i>heart</i>        | <i>10</i>  | <i>sydän</i>            | <i>8</i>    | <i>srdce</i>         | <i>24</i> |
|          | <i>tongue</i>       | <i>7</i>   | <i>kieli</i>            | <i>11</i>   | <i>jazyk</i>         | <i>5</i>  |
|          | <i>eye</i>          | <i>8</i>   | <i>silmä</i>            | <i>19</i>   | <i>oko</i>           | <i>13</i> |
|          | <i>nose</i>         | <i>10</i>  | <i>nenä</i>             | <i>3</i>    | <i>nos</i>           | <i>20</i> |
| <b>b</b> | <i>dog</i>          | <i>17</i>  | <i>koira</i>            | <i>5</i>    | <i>pes</i>           | <i>6</i>  |
|          | <i>cat</i>          | <i>2</i>   | <i>kissa</i>            | <i>1</i>    | <i>kočka</i>         | <i>20</i> |
| <b>c</b> | <i>tree</i>         | <i>6</i>   | <i>puu</i>              | <i>7</i>    | <i>strom</i>         | <i>23</i> |
|          | <i>stone</i>        | <i>6</i>   | <i>kivi</i>             | <i>14</i>   | <i>kámen</i>         | <i>43</i> |
|          | <i>water</i>        | <i>6</i>   | <i>ves</i>              | <i>18</i>   | <i>voda</i>          | <i>54</i> |
| Average: |                     | <b>7,9</b> |                         | <b>13,8</b> |                      | <b>27</b> |

#### 5. Compounds and Their Representation in Three Language Types

|          |                | ENGL. (isolating) |              | FIN. (agglutinative) |              | CZE.        |
|----------|----------------|-------------------|--------------|----------------------|--------------|-------------|
|          | (inflectional) |                   |              |                      |              |             |
| <b>a</b> | <i>head</i>    | 17                | <i>pää</i>   | 914                  | <i>hlava</i> | 6           |
| <b>b</b> | <i>dog</i>     | 12                | <i>koira</i> | 97                   | <i>pes</i>   | 6           |
| <b>c</b> | <i>wate</i>    | 49                | <i>vesi</i>  | 660                  | <i>voda</i>  | 61          |
| Average: |                | <b>26</b>         |              | <b>557</b>           |              | <b>24,3</b> |

## 6. Conclusions: General and Lexicographic

-*English*: strongest in polysemy and weakest in derivation

*Finnish*: strongest in compounding and weakest in derivation

*Czech*: strongest in derivation, while polysemy is weakest

-Two tentative *implicational universals*:

(1) *The shorter an average lexeme is, the more polysemy the language has (and vice versa).*

(2) *The longer an average lexeme is, the more derivation (in inflectional languages) or compounding (in agglutinative languages) the language has (and vice versa).*

### References

- Čermák 1978 Čermák, F., 1978. Typology of the Germanic Languages with Special Reference to Dutch, in: *Folia Linguistica* XII: 1/2, 1978, 65-106.
- Čermák 1990 Čermák F., 1990. *Syntagmatika a paradigmatika českého slova 2: Morfologie a tvoření slov*. Univerzita Karlova Praha (Syntagmatics and Paradigmatics of the Czech Word: Morphology and Word Formation).
- Čermák 2000a Čermák, F., 2000a. Combination, Collocation and Multi-Word Units, in: *Proceedings of The Ninth Euralex International Congress EURALEX 2000*, eds. U. Heid, S. Evert, E. Lehmann, C. Rohrer, Institut für Maschinelle Sprachverarbeitung Universität Stuttgart, Stuttgart, pp. 489-495.
- Čermák 2000b Čermák, F., 2000b. Jazyková nominace: povaha a souvislost některých univerzálií v sémantice a morfologii (polysémie, derivace a kompozice), in: *Slovo a Slovesnost* 61, 249-256 (Language Nomination: Character and Relationship of Some Universals in Semantics and Morphology).
- Čermák 2001 Čermák, F., 2001a. Substance of Idioms: Perennial Problems, Lack of Data or Theory?, in: *International Journal of Lexicography* 14, 2001, 1-20.
- Čermák 2001b Čermák F., 2001b. *Jazyk a jazykověda. Přehled a slovníky*. 3. vyd. Karolinum Praha (Language and Linguistics. A Survey and Dictionaries, 3rd ed.).

- Drossard 1997 Drossard W., 1997. Polysynthesis and polysynthetic languages in comparative perspective, in: *Typology: prototypes, item orderings and universals*. Proceedings of LP'96, ed. B. Palek, Karolinum Prague, 251-264
- Filipec-Čermák 1985 Filipec J., F. Čermák, 1985. *Česká lexikologie*. Academia Praha (Czech Lexicology).
- Kilgarriff 1992 Kilgarriff A., 1992. *Polysemy*. University of of Sussex (unpublished doctoral dissertation, <http://www.itri.bton.ac.uk/>).
- Mathesius 1975 Mathesius V., 1975. *A Functional Analysis of Present Day English on a General Linguistic Basis*. Academia Prague.
- Nykysuomen sanakirja 1967 *Nykysuomen sanakirja* I-III, 1967. Werner Söderström, Porvoo Helsinki 1967 (cca 201 000 entries, 2381p., = NS, Dictionary of New Finnish).
- Skalička 1951 Skalička V., 1951. *Typ češtiny*. Slovanské nakladatelství Praha (Type of Czech).
- Skalička 1979 Skalička V., 1979. *Typologische Studien*. Viewegh Braunschweig.
- Slovník spisovného jazyka českého 1957-1961 *Slovník spisovného jazyka českého* I-IV, 1957-1961. Naklad. Čs. akademie věd Praha (cca 182 000 entries, 4593p., = SSJČ, Dictionary of the Literary Czech Language).
- New Shorter Oxford Concise Dictionary 1993 *The New Shorter Oxford Concise Dictionary on Historical Principles* I-II, 1993. Clarendon Press Oxford (cca 125 000 entries, 3767p., =NSOCD).
- Ullmann 1962 Ullmann S., Semantic Universals. in: J. Greenberg, ed., 1966 (2. vyd.), *Universals of Language*. M.I.T. Press Cambridge, 217-262.
- Vesikansaa 1977 Vesikansa J., 1977. *Johdokset*. W. Söderström Porvoo-Jelsinki-Juva.

## **English lexicography in Japan—state of the art**

Katsumasa Yagi

Kwansei Gakuin University

English has been one of the most extensively and intensively studied foreign languages in Japan since the “modernization” in the mid-19<sup>th</sup> century. During the past one and a half century, the research results of English have been recorded in numerous articles, books and, most significantly, in English-Japanese dictionaries. Since English-Japanese dictionaries are the handiest tools for English learners, their influence on learners and teachers is extensive and long lasting.

I will first look back on the history of the development of English-Japanese dictionaries in Japan; then I will refer to some of the most influential English-Japanese dictionaries; and then I will glean some interesting descriptions provided in those influential English-Japanese dictionaries; and lastly, how English-Japanese dictionaries have contributed to the description, analysis and explanation of present-day English.

The main focus of this discussion is on the last topic. The problem to be addressed is twofold: one is lexicography as a branch of applied linguistics. Lexicography is a branch of “applied” linguistics and we need to consider how the linguistic facts described and explained in terms of purely linguistic means should be incorporated in dictionaries. The other is lexicography as a means of storage of linguistically analyzed facts of the language. Its contributions to linguistic analyses tend to be kept in dictionaries and never to be openly discussed. I will try to focus on some topics of interest thus stored and hidden in dictionaries with special reference to how “invisible meanings” are reflected in “visible forms.”

**Loss in Clarifying Meaning:  
Analysis of a few problems in an English-Chinese dictionary**

Ma Zhuang-huan

Many studies have proved that most dictionary users “consult a dictionary to check the meanings of a word”. To provide meanings of words is undoubtedly one of the major functions of a dictionary. How to provide meanings, however, is always a difficult task. In a language A-language B dictionary, defining a word is usually replaced with translating it from language A to language B, which is based on an assumption that the user knows the word in language B or the user is able to figure out the meaning of the word by looking it up in a language B-language B dictionary or encyclopedia. In an English-Chinese dictionary, for example, usually the headword is translated into Chinese, with little or no Chinese definition or explanation.

On the one hand, this kind of dictionaries are characterized with some evident merits such as being easy for beginners of English learning, convenient for translation, free from certain semantic difficulties in defining words, space saving and therefore likely to contain more entries. In addition, the users of dictionaries of this kind as English learners tend to correspond the headword with its Chinese equivalent subconsciously since they already have mature command of Chinese. And English-Chinese dictionaries can cater for their desire. On the other hand, dictionaries of this kind also possess certain demerits. From both semantic and lexicographical points of view, the paper attempts to analyze A New English-Chinese Dictionary (Century Edition) (NECD), a popular and best-selling dictionary in China, with an emphasis on how to deal with meaning. The analysis is carried out mainly around the following points through comparison with English-English and English-English & Chinese dictionaries.

In English-English dictionaries, the definition of a headword consists of both linguistic and non-linguistics knowledge. For example, the definition of “Thanksgiving day” in Longman Dictionary of Contemporary English (New Edition) (LDOCE) presents useful though simple information about this day; in NECD, however, there is no such information, even in Chinese, but only a translation of the headword. Besides, the definition of a headword also reveals certain sense relations among the headword and the defining words, and displays the value of each term involved. As Saussure believes, the whole mechanism of a language is realized through syntagmatic and associative relations and in a language system “the value of each term results solely from the simultaneous presence of the others”. When “alive” is defined in English, for example,

it's antonym "dead" is probably mentioned, and some other associated words such as "living", "life", and "existence" may also be included. This kind of helpful linguistic knowledge, however, is absent in the dictionaries like NECD: the English headwords are placed among Chinese words, separated from the system of English, though there may be some illustrations.

In the dictionaries like NECD, when some common and/or simple nouns such as "water", "air", are translated into Chinese without being defined or explained, their Chinese equivalents are not likely to cause any serious comprehension problems for the Chinese dictionary users because it is taken for granted that the users already know the meanings of the Chinese words. But when some jargons in modern science or technology or some terms originated from other cultures and probably exotic or unfamiliar to Chinese people are treated in the same way, their Chinese equivalents may result in certain difficulties for the dictionary users to understand. On the contrary, an English-English & Chinese dictionary offers both English definitions and Chinese equivalents of the headwords and therefore appears advantageous in this respect. In NECD, while most headwords are not followed by definitions, there are a small number of exceptions, e.g. "gene", which is explained briefly in Chinese. However, it should be noted that NECD seems not to be governed by a consistent principle how to handle this issue.

The paper suggests that the dictionaries like NECD, while maintaining their merits, should adopt a more flexible and practical means to deal with meanings of the English headwords: though some words do not have definitions, some other words are to be defined or explained, either in Chinese or English. And there should be a relatively stable or consistent guideline for determining what words to be defined, to what extent, whether in Chinese or English, etc.

# **A study of the annotation of frequent words in four English learner's dictionaries**

Lang Jianguo

English Department, Beijing International Studies University

**Abstract:** With access to several well-known English learner's dictionaries, learners are enabled to obtain better data about high frequency words. However, those dictionary users who intend to improve their vocabulary learning or teaching by making use of such annotated words, are also faced with the difficult choice of these word lists which differ from each other greatly as far as frequency annotation is concerned. This paper compares the annotation of such words in terms of their banding, frequency and quantity in four English learner's dictionaries. It suggests that the list of frequent words should vary in accordance with learners' needs at different stages of their learning. It is also argued that future learner's dictionary compilation should aim for a more accurate and practical annotation of frequent words.

**Key Words:** frequent English words; word frequency annotation; English learner's dictionary

## **四部英语学习词典的词频标注比较**

北京第二外国语学院英语系 郎建国

**提要：**20 世纪末以来，英国几大英语学习词典相继开始提供词频标注，为我们选择英语高频词提供了更为可靠的依据；但面对标注数量、分级、标注方式差异极大的几本词典，词典使用者如何利用这些信息提高英语教与学的效率，就成为令人关注的问题。本文通过比较四部英语学习词典所标注高频词的标注方式、分级依据和数量来探讨这些标注的优缺点，指出对不同阶段、不同目的的英语学习者，所需要的常用词词表要有所不同；英语学习词典的词频标注应朝精细化、实用化方向发展。

**关键词：**英语常用词、词频标注、英语学习词典

# Corpus Linguistics

**František Čermák**

Institute of the Czech National Corpus, Charles University, [frantisek.cermak@ff.cuni.cz](mailto:frantisek.cermak@ff.cuni.cz)

## **The Case of The Czech National Corpus: Its Design and Development**

Against a general background of needs for data and situation of the nineties of the 20th century the story of foundation of the Czech National Corpus, one of the largest in Europe today, is given. Types of data, written, spoken and dialectal, are distinguished, the organisation of the whole project and some of its processing briefly described. The idea of representativeness, rarely to be found elsewhere, is discussed and its criteria, based on research are discussed that tell, among other things, what the proportions of a corpus should be. The basic distinction of imaginative and informative texts is further split up into numerous multi-layered subclasses, each with a percentage slot allocated to it. A counterpart to this written strategy, allowing to plan a written corpus, is formed by a strategy how to collect data for a spoken corpus whose prototypical core is delimited by a number of criteria, too. Next, some of technical and linguistic problems, such as lemmatisation and tagging are discussed and the need to combine statistics with (thousands of) rules is stressed. Finally a survey of various corpora that have been completed and are now offered freely for use on the Internet is given. The survey will be briefly commented on:

| Type                        | Corpus (+year)             | Size                                |
|-----------------------------|----------------------------|-------------------------------------|
| <i>Written</i> synchronic:  | SYN2000                    | 120 mil                             |
|                             | SYN2005                    | 120 mil                             |
|                             | PUB2006                    | 300 mil (i.e. newspapers only)      |
|                             | PUB2008                    | 200-300 mil                         |
| <i>Spoken</i> synchronic:   | PMK (Prague Spoken Corpus) | 0,8 mil                             |
|                             | BMK (Brno Spoken Corpus)   | 0,6 mil                             |
|                             | ORAL 2006                  | 1,3 mil                             |
|                             | ORAL 2008                  | 1 mil                               |
| <i>Historical</i>           | DIA                        | 2 mil (7 centuries from 13th cent.) |
| <i>Specialized</i> Corpora: |                            |                                     |
|                             | Karel Čapek' Corpus        | 3,2 mil                             |

|                              |                         |
|------------------------------|-------------------------|
| Bohumil Hrabal's Corpus      | 2 mil                   |
| Corpus of Totalitarian times | 1,15 mil                |
| Hand-written letters KSK     | 0,94 mil (2000 letters) |

*Parallel* Corpora (Czech vs 20 languages)

## **Corpus-based Collocation Studies in English: With special reference to discourse functions**

Atsuko Furuta Umesaki (Ritsumeikan University)

Corpus search has made it possible to explore features of lexis, grammar and discourse by focusing on the relationship of words with their surroundings, drawing on a large number of examples and with statistical data on frequency. Since Sinclair (1991), remarkable progress has been made in collocation studies on lexis and grammar, and a wide range of lexico-grammatical phenomena have been clarified. This increased understanding has contributed to advances in the compilation of dictionaries and in English language teaching. Recently, however, as can be seen in Biber et al. (1999), Hoey et al. (2007), and others, the attention of researchers has come to include the discourse function of phrases and patterns of usage as well as lexico-grammatical characteristics. In this presentation, after a brief review of definitions of 'collocation' and a quick survey of investigations of lexical and grammatical features of collocation carried out in Japan, some findings will be presented on the discourse function and lexico-grammatical characteristics of standard phrases and patterns of usage.

The term 'collocation' is used in its narrower sense to refer to the co-occurrence of words. In its broader sense, as Partington (1996) shows, citing Firth's idea, 'collocation' refers 'to the co-occurrence of items at all grammatical levels'. Specific collocations in the broader sense indicates patterns and phrases referred to by such terms as lexical phrases (Nattinger & DeCarrico, 1992); fixed expressions (Moon, 1998); lexical bundles (Biber et al. 1999); formulaic sequences (Wray, 2002); multi-word sequences (Stubbs, 2007); phraseological units (Gläser, 1998, etc.) and so on. Each term has a slightly different bearing. A corpus search indicates that *perfectly* collocates with *right* but not usually with *wrong*. An overall picture of adjective collocates of *perfectly* shows that it co-occurs with adjectives with good connotations. Such a phenomenon is called by Sinclair (1987) 'semantic prosody'. Comparison of frequent collocates among synonyms enables us to have an overall view of the core meaning of synonyms. Patterns can be identified in concordance lines obtained from corpus search. Finding patterns and accounting for different patterns will illuminate the function of a grammatical category (e.g. a *to*-infinitive) and the semantic type of the words collocating with the category. An example is predicative noun patterning (Yagi, 1999): *It's a shame* can be followed by or collocate with a *that*-clause (e.g. *It's a shame that you are late.*) whereas *it's a pleasure* can collocate with *to*-infinitive (e.g. *It is a pleasure to teach those children.*) Yagi has disclosed relationships between the content of predicative nouns and the patterns those nouns take.

A phrase has a lexico-grammatical meaning and at the same time fulfills a discourse

function. In other words, a pattern has not only a propositional but also an interpersonal meaning. Taking the pattern '*be* + past-participle + *to*-infinitive' as an example, discourse functions of phrases will be considered in relation to their lexico-grammatical characteristics. Some phrases such as *is said to*, *is expected to*, *is shown to*, *is forced to* appear to perform different discourse functions. For the examination of discourse function, 'metadiscourse' categories presented by Hyland (2005) will be employed.

**Vocabulary in Interlanguage:  
A Study on Corpus of English Essays Written by Asian University Students (CEEAAUS)**

Shin'ichiro ISHIKAWA  
Kobe University

Interlanguage (IL) is a particular type of L2 system used by L2 learners. It is defined as a transitional language form existing between L1 and L2 (Ellis, 1994). Learners who start studying L2 after the so-called critical period need to build up the L2 rules based on their L1 rules, as they cannot directly access the Universal Grammar. Consequently, IL is often strongly influenced by L1, which is called L1 transfer.

Although IL is expected to smoothly shift toward L2, some of the errors occurring in IL will be stabilized or "fossilized." The fossilization is largely caused by (i) language transfer, (ii) transfer of training, (iii) overgeneralization, (iv) strategy of L2, and (v) strategy of L2 communication (Selinker, 1972).

IL has been studied theoretically, mainly from the viewpoint of second language acquisition (SLA) (Skehan, 2008), but since the 1990s the development of large learner's corpora has led to the descriptive analysis of IL. Learners' corpora, which make the NS/NNS gap explicit, are expected to contribute to a closer understanding of IL.

Various learners' corpora have been compiled to date, of which the most influential is presumably the International Corpus of Learners' English (ICLE) (Granger, Dagneaux, & Meunier, 2002). ICLE has already collected more than two million words of English essays, written by learners from different L1 backgrounds. ICLE enables us to conduct two kinds of contrastive interlanguage analysis (CIA) (Granger, 1998), namely, NS/NNS comparison and NNS/NNS comparison. ICLE, however, has several problems requiring careful consideration. The first is the fact that it does not necessarily control topics which may seriously influence the distribution of vocabulary. The second is that it does not include essays written in the writers' L1, which should also be analyzed for probing the L1 transfer.

This is why the author is currently engaged in compiling a new type of learners' corpus, The Corpus of English Essays Written by Asian University Students (CEEAAUS) (Ishikawa, in press). Rigidly controlling writing conditions such as topics, time, length, and reference use, CEEAAUS aims to collect highly homogeneous language data, which is a prerequisite to a reliable discussion concerning the difference between varied writers' groups. CEEAAUS, which consists of five modules: CEEJUS (the collection of English essays by Japanese university students), CEECUS (that by Chinese university

students), CEEKUS (that by Korean university students), CEENAS (that by English native speakers), and CJEJUS (the collection of Japanese essays by Japanese university students), is also intended for use for an enlarged multi-layered CIA (MCIA).

In this presentation, I would like to introduce the details of CEEAUS and discuss several noteworthy findings about the IL vocabulary used by Japanese learners of English.

## Collocations and phraseology of present-day Japanese

Naohiro Takizawa (Nagoya University, Japan)

As far as research on the English language is concerned, corpus-based methodology has gained a strong impetus during the past decade or so. With the advanced technological developments of computers and computer programs, large-sized corpora of English such as the 100-million-word British National Corpus have been compiled and made available to the public.

The trend of using corpora has brought us an entirely new way of doing research in linguistics. Using these gigantic databases has allowed researchers to shed light on areas where the introspection or intuition of native speakers does not always produce satisfactory answers. This is especially true when peripheral linguistic phenomena and conventional aspects of language use such as collocations, among others, are being discussed.

Unfortunately, there is currently no large-sized, well-balanced corpus of the Japanese language comparable to the British National Corpus. The five-year project for compiling a balanced corpus of the Japanese written language—the Balanced Corpus of Contemporary Written Japanese (BCCWJ)—is now under way, and will be completed in another year or two. Until it is completed, for a grammatical/lexical study of the Japanese language, we need to rely on newspaper corpora and other databases like *Aozora Bunko*, which archives more than 8,000 Japanese novels and is accessible through the internet.

In this lecture, I will first touch on the BCCWJ project and then go on to discuss how to retrieve collocations, lexical bundles (Biber et al. (1999)) and, more widely, phraseological units, from Japanese texts found in newspapers, novels and the partially completed BCCWJ corpus. I will then show that frequency alone is not sufficient for identifying important units and that other criteria are necessary. Finally, I will discuss the idea that language production by native speakers is more conventionalized than we usually imagine.

# Phraseology

**František Čermák**

Faculty of Philosophy, Charles University Prague, Czechia, e-mail: [frantisek.cermak@ff.cuni.cz](mailto:frantisek.cermak@ff.cuni.cz)

## Identification of Idioms

### 1. Idioms and Idiomatics (Phraseology)

#### 1.1 Criteria for Idioms:

non-additiveness of constituent meanings  
functional approach

- (1) *beyond my ken* : **ken**  
*at somebody's behest* : **behest**
- (2) *as to (sb/sth)*

transformational deficiency: not specific enough  
Principal methodological questions:

1: Is it possible to define this discipline by a configuration of familiar criteria, valid elsewhere in language ? Or is it necessary and possible to look for some other specific criteria ? (*autonomy of criteria*).

2: Is the set of criteria, applied here, specific for this field only ?

In other words, does it not cover other language areas as well? (*specificity and non-substitutability of criteria*).

3: Do the proposed criteria cover the field as a whole and not some segments of it only, however large they might be? (*adequacy of criteria*).

#### 1.2 Syntagmatic and Paradigmatic Relations

Attempts at the idiom definition such as

(I) discrete, analytical form, unique concatenation of components, syntactic frozenness, non-additiveness of constituent meanings, integral meaning, bound or reevaluated

component meaning (e.g. Coates, Černyševa, Archangelskij, Fraser, Kunin etc.), or (II) formal and semantic stability, nominative integrity, transformational deficiency (e.g. Rojzenzon, Achmanova, Machač, Mokijenko, Chafe, Fraser etc.),

have clearly a predominant syntagmatic (I) or paradigmatic (II) character.

F. de Saussure:

-*locutions toute faites* are not "generated"

-*La loi tout a fait finale du langage est... qu'il n'y a jamais rien qui puisse résider dans un terme.*

-*Il y a entre autres toute une série de phrases qui sont toute faites pour la langue.*

## 2. Idioms: A Structural Approach

### 2.1 Paradigmatic Aspects of Idioms

distributional or membership restriction of some constituents

- |     |   |                  |
|-----|---|------------------|
| (3) | <i>k i t h</i> and <i>kin</i> "friends and relations"               | : <b>kith</b>    |
|     | <i>s p i c k</i> and <i>span</i> "clean and tidy"                   | : <b>spick</b>   |
| (4) | <i>give u m b r a g e</i> , <i>take u m b r a g e</i> at <i>sth</i> | : <b>umbrage</b> |

### 2.2 Collocational Paradigms: a Syntagmatic Aspect

*Collocational Paradigm* (a class of items with which an external item usually and meaningfully collocates) is specific for every word, represents its distribution and determines its meaning.

**DOG** {*bark, growl, whine,...*} x *mutter, whistle, mew...*  
           {*lie, run, stand, jump, dig...*} x *wave, rock,...*  
           {*sniff, track, nose about,...*} etc.

### 2.3 Virtual Paradigms and the System

Virtual paradigm (both intuitively and taxonomically given class where an item belongs) is typically characterized by a substitution possibility between its members ->

analogy.

One-member classes

### 2.3 Idiomatic Syntagmas

Restrictive combinations:

(5) *hit the nail on the head*

(6) *set the Thames on fire*

(7) *be born with a silver spoon in one's mouth*

Although these examples may, for a moment perhaps, give the impression that the criterion of the semantic non-additiveness can be applicable here, examples (4-8) show that is not the case.

### 2.4 Restricted Collocability of the Unique Constituents

*kith, umbrage etc.* : verboids

### 2.6 A Summary and a Definition

virtual paradigm, collocational paradigm

paradigmaticity, function

A definition: the idiom is

*such a unique and fixed combination of at least two elements for which it holds that at least some of these do not function, in the same way, in any other combination or combinations of the kind, or occur in a highly restricted number of them, or in a single one only.*

the concept of **anomaly**

## 3. **Idiom: Identification and Test**

Procedure (and a commutation test):

- (1) find a stable/fixed combination (=a candidate for inclusion)
- (2) find out if at least one of its constituents is not severely limited in its substitution possibility rendering other similar and meaningful combinations (see the definition)
- (3) check presence of a metaphor (not obligatory, though frequent)

#### 4. A Conclusion.

*a fixed and stable syntagma of elements where at least one element is, as seen from the viewpoint of (at least) one other element, a member of a severely restricted and closed paradigm (always in form, though often semantically, too).*

The answers to questions in 1.1 are yes for No. 1 and 3, but no for No. 2

Two general types of anomaly: -in transformational behaviour

-in                   degrees                   of

frozenness/generative aspects

These and other types seem to have their solution and explanation in the nature of complex paradigmatic and syntagmatic aspects.

#### Literature

Arxangelskij, Vladimir L., 1964, *Ustojčivyye frazy v sovremennom russkom jazyke. Osnovy teorii ustojčivyx fraz*

*i problemy obščej frazeologii.* Rostov-na-Donu, Izd. Rost. univ.

Axmanova, Ol'ga S. 1967. *Očerki po obščej i russkoj leksikologii.* Moscow, Uèpedgiz.

Babkin Aleksandr M., 1964, *Leksikografičeskaja razrabotka russkoj frazeologii.* Moscow-Leningrad: Izd.

Nauka. Bally, Charles, 1909, *Traité de stylistique française.* Heidelberg-Paris: Winter.

Carnap, Rudolf, 1966, *Meaning and Necessity.* Chicago U. P. Chicago.

Chafe, Wallace L., 1968, *Idiomaticity as an Anomaly in the Chomskyan Paradigm.* Foundations of Language,

4, 109-125

Chomsky, Noam, 1980, *Rules and Representations.* Oxford, Blackwell.

Coates, William A., 1964, *Meaning in Morphemes and Compound Lexical Units.* In: Proceedings of the 9th

International Congress of Linguists. The Hague, Mouton 1046 - 61.

- Cowie, Anthony P., Ronald Mackin, 1976, *Oxford DicCionary of Current Idiomatic EngLish*. Vol 1: Verbs with  
Prepositions and Particles. London: Oxford U. P.
- Crystal, David, 1980, *A First Dictionary of Linguistics and Phonetics*. London: Deutsch.
- Čermák, František, 1982, *Idiomatika a frazeologie češtiny*. Prague, Universita Karlova (=Idiomatics and Phraseology of Czech).
- Čermák, František, Jiří Hronek, Jaroslav Machač (eds.), *Slovník české frazeologie a idiomatiky*. Prague,  
Academia. *Přirovnání* (1983), *Výrazy neslovesné* (1988), *Výrazy slovesné* (1994)  
(=Dictionary of the Czech  
Phraseology and idiomatics: Similes. Non-Verbal Expressions. Verbal Expressions).
- Čermák František, 1988, *On the Substance of Idioms*, *Folia Linguistica* XXII/3-4, 413-438.
- Čermák, František, 1994, *Idiomatics*. In *The Prague School of Structural and Functional Linguistics*, ed. P.A.  
Luelsdorff, John Benjamins Publ. Comp. Amsterdam, 185-196.
- Čermák, František, 1994, *Czech Idiom Dictionary*. In EURALEX '94, W. Martin, W. Meijs et al. (eds.), Euralex  
Amsterdam, 426-432.
- Čermák, František, 2007, *Frazeologie a idiomatika česká a obecná. Czech and General Phraseology*. Karolinum  
Praha 2007
- Černyševa, Irina I., 1970, *Frazeologija sovremennogo nemeckogo jazyka*. Moscow, Izd. Vysšaja škola.
- Dictionnaire de linguistique*, 1973. J. Dubois, M. Giacomo, L. Guespin, C. Marcellesi, J.-B. Marcellesi, J.-P.  
Mevel (eds). Paris, Librairie Larousse.
- Engler, Rudolf (ed.), 1967. Ferdinand de Saussure, *Cours de linguistique générale. Édition critique* I, fasc. 2  
Wiesbaden, Harrassowitz.
- Filipec Josef, František Čermák, 1985. *Česká lexikologie*. Prague, Academia (=Czech Lexicology).
- Firth, John R. 1957. *Papers in Linguistics 1934-1951*. Oxford-London: Oxford U. P.
- Fleischer, Wolfgang, 1982, *Phraseologie der deutschen Gegenwartssprache*. Leipzig: Bibl. Inst.
- Fraser, Bruce, 1970, *Idioms within a transformational grammar*. Foundations of

Language 6, 22 -42.

Gak, Vladimir. G., 1972, "*K probleme semantičeskoj sintagmatiki*." In: Problemy strukturnoj lingvistiki 1971,

ed. S. K. Šaumjan. Moscow: Izd. Nauka, 373 -395.

Halliday, Michael A. K., 1976, *System and Function in Language*, ed. G. Kress. London, Oxford U. P.

Heupel, Carl, 1975, *Taschenwörterbuch der Linguistik*. München: List.

Hjelmslev, Louis, 1970, *Pour une sémantique structurelle*. In: Essais linguistiques, Travaux du Cercle

linguistique de Copenhague. Vol. XII, 96-112, Copenhague, Nordisk Sprog-og Kulturforlag.

Hjelmslev, Louis, 1972, *O základech teorie jazyka*. Prague, Academia. (=a Czech translation of *Omkring*

*sprogteoriens grundlaeggelse* by F. Čermák, Kobenhavn 1966 = *Prolegomena to a Theory of Language*,

Baltimore 1969).

Kunin, Aleksandr V., 1972, *Frazeologija sovremennogo anglijskogo jazyka*. Moscow, Izd. Meždunarodnyje

otnošenija.

Lyons, John, 1977, *Semantics*. Vol. I, II, Cambridge, Cambridge U. P.

Machač, Jaroslav, 1967, *K lexikologické problematice slovních spojení*. Slovo a slovesnost, 28, 137-49 (=On the

Lexicological Problems of Collocations).

Makkai, Adam, 1972, *Idiom Structure in English*. The Hague-Paris, Mouton.

Melčuk, Igor A. 1960. "*O terminax ustojčivost" i "idiomatičnost"*. Voprosy jazykoznanija 4, 73 - 80.

Mokijenko, Valerij M., 1980, *Slavjanskaja frazeologija*. Moscow, Izd. Vysšaja škola.

Molotkov, Aleksandr I., 1977, *Osnovy frazeologii russkogo jazyka*. Leningrad, Izd. Nauka.

Moon, Rosamund, 1998, *Fixed Expressions and Idioms in English*. Oxford: Clarendon Press.

Ries, John T., 1928, *Zur Wortgruppenlehre*. Prag: Taussig & Taussig.

Rojzenzon, Leonid I., 1973, *Lekcii po obščeej i russkoj frazeologii*. Samarkand, Univ. im. A. Navoi.

de Saussure, Ferdinand. 1916. *Cours de linguistique générale*. Paris: Payot.

Šmelev, Dmitrij N., 1973, *Problemy semantičeskogo analiza leksiki*. Moscow: AN

SSSR, - Inst. russkogo jaz.

Weinreich, Uriel, 1968, *Explorations in Semantic Theory*, In: Current Trends in Linguistics, Vol. III; Theoretical Foundations, ed. T. A. Sebeok. The Hague, Mouton, 395-477.

Weinreich, Uriel, 1969, *Problems in the Analysis of Idioms*. In: Substance and Structure of Language, ed. J. Puhvel. Berkeley-Los Angeles, U. of California Press, 23-81.

Wittgenstein, Ludwig, 1953, *Philosophical Investigations*. Oxford-New York: Blackwell.

*Wörter und Wendungen. Wörterbuch zum deutschen Sprachgebrauch*, 1979, E. Agricola et al. (eds). Leipzig:

Bibliographisches Institut

Joanna Szerszunowicz

Uniwersytet w Białymstoku (Białystok University) Poland

**Linguo-cultural analyses of European phraseological units in a confrontative  
perspective**

The relation between language and culture is of complex nature and it manifests itself in a variety of language phenomena. Phraseological units tend to reflect the culture of a given ethnic community; therefore, in a confrontative perspective the cultural information stored in idioms is elicited to observe differences and similarities. Conducting such analyses renders it possible to reconstruct the naïve pictures of the world contained in particular languages, which represents cultural and spiritual experience of a given ethnic community. The analyses at issue are situated in a field called *cultural linguistics*, anthropologically oriented branch of cognitive linguistics. Cognitive linguistics and conceptual studies provide theoretical grounds for linguo-cultural analyses, which include research studies on phraseological units. According to cognitive linguists, generally speaking, language is part of thought and by conducting analyses of language the underlying conceptual structure can be discovered.

The origins and motivations of phraseological units are of importance in such analyses. First of all, a number of units in European languages originate from the common cultural heritage, e.g. the Bible, Greek mythology, masterpieces of literature, important historic events etc. They tend to be present in many languages in similar forms and have identical meanings. Yet, some observations of this kind are culture-specific. Secondly, many units originally referred to the nation's culture:

national history, literature, customs. Such units either possess a formally different phraseological equivalent in another language, or they have a non-idiomatic equivalent. Other units are carriers of specific concepts, not present in other cultures, so they have only descriptive equivalents. There are also some schemata realized internationally, in which the variable constituent, rich in connotations, is of national character. Moreover, numerous phraseological units of pragmatic character are culture-specific. Furthermore, some units contain culture-bound constituents, for instance, proper names of national importance, names of institution. Thirdly, a number of phraseological units function as carriers of local culture. Some of them are realizations of a schema (which can function nationally or even internationally), in which one constituent is a locally dependent variable.

It is worth stressing that the results of linguo-cultural analyses are implemented in practice. For instance, foreign language teaching benefits from such research, since phraseological false friends are traced. The focus on the cultural character of phraseological units facilitates developing cross-cultural competence of L2 users. The results of analyses at issue can also be implemented in phraseography, since the phraseographic description can be more detailed and can include information which is very important for advanced language learners, so that they could have an insight into another nation's culture.

## English Phraseology in the Japanese context

Katsumasa Yagi

Kwansei Gakuin University

Current prevalence of phraseology in Europe is impressive in its variety and depth. Phraseology is now definitely emerging as one of the important areas of lingo-cultural research. This is amply evidenced by numerous works incorporated in such latest anthologies as *Phraseology in Motion I* (A. H. Buhofer and H. Burger (eds.), Schneider Verlag Hohengehren, 2006), *Phraseology: An International Handbook of Contemporary Research* 2 vols. (H. Burger et al (eds.), Walter de Gruyter, 2007), *Phraseology: An Interdisciplinary Perspective* (S. Granger and F. Meunier (eds.), John Benjamins, 2008).

These works cover wide areas of language studies: idioms, morphology, collocations, set phrases, proverbs, language acquisition and language teaching, etc. Phraseology, however, seems to be quite new to most Japanese linguists and language teachers. In that sense, we are in a position to learn many things from phrasiological studies in Europe. But when we look back on the history of linguistic research in this country and think of the “neo-traditionalism” in English linguistics, it is not surprising to see that we have accumulated plenty of works in linguistic research of English which are now called “phraseological” studies.

The linguistic research in Japan has long been under the strong influence of Anglo-American linguistics especially after world war II : structural linguistics, transformational generative grammar, cognitive linguistics from the United States on the one hand, and systemic functional grammar from the United Kingdom on the other. We have been busy pursuing the fast changing trends in Anglo-American linguistics, and as a result, we tend to leave in the darkness of oblivion the trend in linguistic research and the great achievements thereof in this country by the forerunners since the early stage of modernization in the mid-19<sup>th</sup> century.

The purpose of this discussion is to review those remarkable achievements in linguistic research on the English language here in this country and reconsider their significance in the context of phraseology.

## Phraseological Studies on Present-Day Spoken English

Ai Inoue (National Defense Academy, Japan)

The province of phraseology is extensive and each phraseological researcher carries out their unique research (see Burger (ed.)(2006), Granger and Meunier (2008) etc.). The minimal common ground in phraseological studies is that all studies deal with phrases in its broad sense. I have been working on the phraseology of spoken English by using BNC, WordBanks *Online* and a present-day spoken corpus, which I refer to as LKL Corpus. My basic view on the analysis of spoken English is a theory which was inherited from the tradition of the linguistics analysis of English in Japan and a continuation of the tradition revised and developed under the influence of various linguistics theories developed in the United States and elsewhere.

The central contention of my lecture focuses on introducing phraseological research on present-day spoken English. Its research is rudimentary since there are many problems to be tackled. Before directly entering into the main topic of my lecture, I will discuss what is spoken English, how to obtain data of spoken English in machine readable forms, the significance of using corpora in the empirical study on spoken English, etc. I will argue that corpora for specific purposes are indispensable for the descriptive phraseological studies of present-day spoken English. I will discuss the following four points: (1) the relationship between the functions of phrases like “you know what” and their syntactic features; (2) the relationship between the functions of the phrases and their phonological characteristics; (3) the relationship between expressions like “go and do,” “go to do,” “go do” and grammatical categories; (4) the introduction of newly emerging expressions like *pirate version* (as compared with the usual “pirated version) and the mechanism which underlies in the production of these phenomena.

## A Phraseological Approach to *That*-taking Verbs in English

Makoto Sumiyoshi (Setsunan University)

A. S. Hornby, who came to Japan as an English teacher in 1924, realized that it was necessary to develop a new way of teaching English to Japanese students, and published *the Idiomatic and Syntactic English Dictionary* (1942) and *A Guide to Patterns and Usage in English* (1954, 1975). These books, aimed at Japanese students of English, were written on the basis of his own principle that learning what syntactic patterns a particular word takes is enormously important for students of English.

These two books have bloomed into something special for foreign learners of English in the 21<sup>st</sup> century. *The ISED* has evolved, with revisions, into *The Oxford Advanced Learner's Dictionary* 7<sup>th</sup> edition (2005). In addition, Hornby's 1975 book inspired corpus linguists to invent a totally new approach to language research. Hunston and Francis (1999), advocating a "pattern-based approach to grammar", argues that "a pattern is a phraseology frequently associated with (a sense) of a word, particularly in terms of the prepositions, groups, and clauses that follow the word". This shows that a new dimension had been added to English research when Hornby's method of English teaching was synthesized with the findings highlighted by corpus linguistics.

It is extremely interesting to note that a method of teaching English that was originally developed for Japanese students of English has made a great contribution to language teaching and research in the 21<sup>st</sup> century. The aim of my presentation is to clarify that this phraseological approach to words is so useful and promising in regards to investigating the facts of present-day English, especially focusing on *that*-taking verbs. A phraseological approach to *that*-taking verbs will reveal that a larger number of verbs are more flexibly used with this type of syntactic pattern, that is, *that*-clauses, than has been expected so far in other linguistic approaches to the English language. For example, the verb *apologize*, which has been thought to be incompatible with a *that*-clause by pragmaticians, can take this form in present-day English. Moreover, *cannot bear* has been changing its patterns, resulting in a remote possibility of taking *that*-clauses. These facts can be clarified by in-depth investigations into the patterns in which the verbs are used. A wide variety of verbs will be mentioned in my presentation along with authentic data gleaned from corpora and different data sources. My presentation is, under the influence of newly developed ways of language research such as corpus linguistics, to show one of the aspects of the current state of "traditional" Japanese ways of English research.