

On the presentation of collocations in monolingual dictionaries

Ulrich Heid

Universität Stuttgart, Institut für maschinelle Sprachverarbeitung
Azenbergstr. 12
D 70174 Stuttgart
heid@ims.uni-stuttgart.de

Abstract

The presentation of collocations in the microstructure of dictionary articles has received comparatively little metalexicographic attention.

Prompted by the recent publication of two new collocation dictionaries, the *Oxford Collocations Dictionary for Students of English* (OCDSE) and the *Dictionnaire Combinatoire du Français* (DCF), we analyse presentational devices for collocations in general and specialized monolingual dictionaries and make proposals towards efficient presentation. We do so by first listing the main types of linguistic facts on collocations that require presentation in dictionaries, and by then discussing the place of collocations in the microstructure, the syntagmatic form of collocation indications, as well as their order and grouping.

1. Introduction

The importance of collocational information and the need for collocation indications in dictionaries are widely recognized.

Monolingual and bilingual dictionaries tend to include collocations in their microstructure; in particular, large monolingual dictionaries and learners' dictionaries pay attention to collocations. In addition, some specialized dictionaries have been published, over the last few years, dealing primarily or exclusively with collocations; among these are Ilgenfritz et al. (1989), the new edition of *Duden Stilwörterbuch* (2001), the *Oxford Collocations Dictionary for Students of English* (OCDSE, 2002) and, most recently, the *Dictionnaire Combinatoire du Français* (DCF, 2003). The latter two will be analyzed and compared with some monolingual definition dictionaries. The comparison will show which presentational devices are available; some of them, present in specialized collocational dictionaries only, would also be useful within general monolingual dictionaries.

We use the term collocation to denote binary word combinations, where one element is semantically autonomous (the 'base', in Hausmann's terms, cf. Hausmann 1989, 2004) and the other (the 'collocate') lexically selected and 'syn-semantic' with the first one (Hausmann). Collocations tend to follow syntactic patterns and have thus been classified, according to the word class of the base and of the collocate, into noun+verb and noun+adjective collocations, as well as verb+adverb, and adjective+adverb collocations (and a few other, less frequent types). Our discussion will be limited to noun+verb and noun+adjective-collocations, but our findings should be applicable to other types of collocations as well.

1.1 Metalexicographic issues on collocations

Metalexicography has discussed the handling of collocations in dictionaries in terms of the selection of items to be treated in a dictionary (cf. Bahns 1995), the headword under which collocations have to be lemmatized (cf. e.g. Hausmann 1989) and, to a minor extent, the way in which collocations are presented in the microstructure (cf. e.g. Cowie 1978).¹ We will now focus on this latter point.

We first discuss the linguistic phenomena which must be kept track of in collocation indications in monolingual dictionaries (section 2), then analyze presentational devices in different types of monolingual dictionaries (section 3), and finally draw a few conclusions for an improved presentation of collocational data in monolingual dictionaries. To allow for comparison, we use entries s.v. *hope* (and its equivalents) throughout this article.

1.2 Description vs. Presentation -- Lexicographie vs. Dictionnaire

We start from Quemada's distinction between *lexicographie* and *dictionnaire*, which emphasizes two different aspects of the work of lexicographers. *Lexicographie* is taken to denote all those activities which lead to a detailed data description: corpus analysis, collection of corpus data for linguistic phenomena, identification of linguistic properties of a word or a collocation etc. *Dictionnaire* is the term used for all activities of data presentation: selecting illustrative examples, deciding on abbreviations to be used to convey information about certain descriptive results, etc. We will use Quemada's distinction, translating his terms into *description* (for *lexicographie*, cf. section 2) vs. *presentation* (for *dictionnaire*, section 3), mainly to avoid potential ambiguities of the term "lexicography".

2. Descriptive aspects of collocations

The types of information which must be given, in a dictionary, for a collocation, depend essentially on the intended use and user group of the dictionary. Learners' dictionaries have to be most detailed on collocations (cf. Bahns 1996, Hollos 2004), but also large monolingual definition dictionaries should give sufficient details (cf., for example, Schafroths criticism of the 10 volume *Duden. Das grosse Wörterbuch der Deutschen Sprache*, Schafroth (2003)).

A "maximalist" view of collocation description is one which assumes that collocations must be described in as much detail as single word items (cf. Heid 1998). This leads to the following list of phenomena, for noun+verb and noun+adjective collocations:

- lexical combinatorics of the collocations components
 - the cooccurrence of two lexemes in a collocation (noun+verb; noun+adjective);
 - for lists of lexemes that cooccur collocationally with a given item, it should be clarified whether these lists are closed (i.e. enumerable sets of lemmas) or open, but e.g. definable in terms of semantic classes; an example of a closed list is DE *etwas/nichts/(nicht) viel/eine Menge/grosse Stücke von jmdm halten* ('estimate sb (much/not)');

- morphosyntactic details
 - the preferences of a noun in a noun+verb or noun+adjective collocation with respect to number: DE *sich Hoffnungen machen* is mostly in the plural, as is EN *high hopes*; the same applies to other morphosyntactic properties, such as case in German (e.g. *guter Hoffnung sein* in the genitive), determination, use of possessives (e.g. DE *sein Veto einlegen*), etc.
 - modifiability of a noun in a noun+verb collocation (e.g. *eine (wichtige) Frage stellen* vs. *Zur { } Sprache bringen*);
- syntactic subcategorization
 - of the verb in a noun+verb collocation (use of prepositions, etc.);
 - of the noun in a noun+verb collocation (e.g. DE *zu dem Schluß kommen, daß*, where the *daß*-clause depends on the noun *Schluß*; ('conclude that...');
 - of the collocation as a whole (e.g. DE *in der Lage sein + zu + INF*, where none of the elements of the combination takes an infinitive, outside the complex expression);
- semantic properties of the collocation:
 - in terms of synonymy with other expressions; in terms of aspectual or aktionsart classifications, etc.
- pragmatic properties:
 - with respect to diasystematic variation (i.e. geographical, domain specific, style,...);
 - with respect to frequency in a given corpus.

Table 1 summarizes the most important phenomena from the list above.

Level	Phenomenon	Examples
Lexical Combinatorics	two lexemes open/closed collocate list	<i>pay+attention</i> <i>{etwas/nichts...} halten von</i>
Morphosyntax	noun: singular/plural modification of N determination of N	<i>high hopes</i> <i>raise a (ADJ) question</i> <i>zur (= definite) Sprache bringen</i>
Syntax	verb valency noun valency	<i>raise+ OBJ (question)</i> <i>cherish the hope that+clause</i>
Semantics	Synonymy	<i>Vorschlag machen/unterbreiten</i>
Pragmatics	diasystematic marks frequency in a corpus	style, geographic use,...

Table 1: Types of information for noun+verb collocations: Summary

In addition, it has been argued that collocations, as any other kind of lexical items, should be described within their syntagmatic context: collocations may be combined, e.g. into lexical triples, DE (*Kritik+ü*)+(scharf+*Kritik*): *scharfe Kritik üben* ('criticize massively'), cf. Heid (1994) and Zinsmeister/Heid (2003). Combinations of collocations and selectional phenomena may lead to longer contexts (cf. Schafroth (2003): *die Polizei führt umfangreiche Ermittlungen gegen jmd* ('the police conduct detailed investigations on sb.')).

Furthermore, a lexical semantic description of collocations is often required, either in a (semi-)formalized notation (as in Mel'čuk's Explanatory and combinatorial dictionaries (ECDs), or implicitly, through meaning explanations, synonyms etc.

3. The presentation of collocations

We distinguish three major aspects of the presentation of collocations in dictionaries, of which we only discuss the last one:

- the selection of collocation candidates described in a dictionary;
- the lemmatization of collocations (under the base or under the collocate, or under both);
- the microstructural presentation devices used to convey information about the linguistic properties of collocations.

3.1 The place of collocations in the microstructure

Where do we find collocations in the microstructure of a dictionary article? Bahns (1996): 41 notices that most learners' dictionaries place collocation indications within the demonstration (or: illustration) part of the microstructure, not in the explanation (or: definition) part (terminology of Zöfgen (1986)). An example is OALD 4 (entry in (Fig. 1)), where collocation indications follow the definitions. As Bahns (1996): 44 notices, Cobuild has collocations within the actual definition texts.

Hope [...]

desire for sth to happen, combined with the expectation that it will:

cherish a/the hope that he will recover° *a ray of hope* ie a slight hope ° *Our hopes for fine weather were not disappointed* ° *We've set/pinned all our hopes on you* ° *She has (high) hopes of winning [...]*

Figure 1: Extracts from the entry s.v. hope in OALD-4

Few dictionaries have a special "collocation zone", e.g. GWDaF (see the entry in (Fig. 5)). The Van Dale series of mono- and bilingual dictionaries also has a separate "collocation zone", which essentially replaces the demonstration part (see the bilingual entry (Fig. 2)).

hoop [...] I < de (m.) >0.1 [opeenhoping] tas [...] II < de > 0.1 [verwachting] *espoir* (m.) => *espérance* 1.1 een sprankje, vleugje ~ [...] un rayon d' *espoir*, une lueur d' *espoir* [...] 3.1 iemands ~ de grond in boren *détruire*, *ruiner les espoirs de qn*; ~ geven *inspirer*, *donner de l' espoir*

Figure 2: Extracts from the entry s.v. *hoop* from the *Van Dale Groot Woordenboek Nederlands-Frans*, VDGW-NF

3.2 The syntagmatic form of collocation indications

Some dictionaries illustrate collocations by means of full example sentences (e.g. OALD4 (Fig. 1)) or definition sentences (*Cobuild*). Others use shortened forms of presentation, such as formulae similar to valency patterns (cf. DCF, (Fig. 3)) or just listings of collocates, as they are already found in Lacroix (1931), (Fig. 4).

The usefulness of these devices depends on the use of further devices:

- a listing of collocates for a given base is sufficient, if the listing is (syntactically and semantically) contextualized (as, e.g. in OCDSE, cf. (Fig. 7));
- example sentences tend to be polyfunctional, especially in large monolingual dictionaries. There, often citations from authors rather serve to show non-standard uses of a word than to indicate collocational usage. This difference should then however be made explicit for the user.

Espoir

aimer sans espoir: phr. aimer sans espérer être aimé [...]
caresser un espoir, une espérance: phr. espérer qqch [...]
garder espoir: phr. continuer d' espérer
garder l'espoir de <inf>: phr. continuer d' espérer que <prop>
gonflé d'espoir: qual. qui espère beaucoup [...]
nourrir l' espoir de <inf>, que <prop>: phr. espérer <inf>, que <prop>
nourrir l'espoir: phr. continuer d' espérer [...]

Figure 3: Extracts from the entry s.v. *espoir* from *Dictionnaire combinatoire du français*, DCF

espoir. Caresser un espoir. Vivre dans l' espoir.

Voir *espérance*.

espérance. Nourrir, fonder, former, exprimer, formuler, garder, conserver, perdre, abandonner, bannir, raviver, autoriser, enlever, lasser, retrouver, réveiller, dépasser, tromper, décevoir, trahir, caresser une espérance. Se bercer d' espérances. Une espérance naît, renait, grandit, se réalise, s' évanouit. - QUAL. vaine, douce, illusoire, chimérique, ferme, vivace, grande.

Figure 4: The entries s.v. *espoir* and *espérance* from LACROIX 1931

Thus, either examples should be marked as giving vs. not giving collocations, or a separation into "zones" should be used to clearly distinguish collocational data from other example sentences.

Hoffnung die; -, -en; 1 e-e H. (auf etw. (Akk)) der starke Wunsch od. Glaube, daß etw. geschehen wird < e-e begründete, berechtigte, falsche, schwache H.; sich/j-m Hoffnung(en) machen; in j-m Hoffnung(en) (er)wecken; H. schöpfen; (keine, wenig) H. haben; j-m e-e/die H. nehmen; die H. aufgeben, verlieren >: *es gibt kaum noch H, daß er gesund wird; sie ging voller H. in die Prüfung; [...] 2 [...]*

Figure 5: Extracts from the entry s.v. *Hoffnung* from *Langenscheidts Großwörterbuch Deutsch als Fremdsprache*, LGDaF

The polyfunctionality of examples is most striking with items that do not participate in collocations. Verbs like DE *erforschen*, EN *investigate, analyse*, etc. may take very many kinds of object nouns, but no significant collocations. It would help users to understand this fact, if the examples given were distinguished from collocational examples, as they are only instances from a potentially large set (the notation "..., etc." is not sufficient).

3.3 The order and grouping of collocation indications

Bahns (1996:114) concludes his study on collocations in learners' dictionaries with a list of guidelines for macro- and microstructural presentation devices for a collocations dictionary. For the order and the grouping of collocations, he proposes the following two devices (summarized in italics and commented):

1. *by structural types: noun+adjective, verb+subject, verb+object, ...*

This grouping is already present in the dictionary by Lacroix (1931, cf. (Fig. 4)) and in most other dictionaries. It is useful for productive and receptive use of the dictionary.

2. *Within a structure type: grouping by alphabet or by frequency.*

An alphabetical ("semasiological") order may facilitate access to collocations in a decoding situation; for an encoding use, it seems unappropriate to us, as its perspective is opposed to the "onomasiological" needs of a learner who knows a base,

but is not sure about the collocates required by that base. An example of this practice can be found in the monolingual definition dictionary by Palazzi/Folena (Fig. 6). Note that DCF has a flat alphabetic list of collocations by collocates: this makes it hard to infer any semantic classification of the collocations, except from the meaning explanations (by reading the whole article) (cf. (Fig. 3))

speranza[...] sf. 1. condizione di chi attende con desiderio fiducioso che si realizzi un bene futuro: *mi sorride la speranza di un meritato riposo; ho una mezza speranza, un filo di speranza [...]* 2. cosa o persona in cui si ripongono le proprie speranze: *i figli sono la speranza degli genitori, [...] N. 1. Sin.* aspettazione, assegnamento, [...] | debole, dolce, falsa, fervida, folle, fondata, [...] viva | abbandonare, carezzare, concepire [...], perdere, realizzare, rinverdire | *Contr.* delusione, disperazione.

Figure 6: Extracts from the entry s.v. *speranza* in PALAZZI/FOLENA

A frequency-based grouping is useful if collocation frequency information is given explicitly in the dictionary. Even though corpus-based lexicography provides such data, we are not aware of any frequency indications in dictionaries, so far.² Furthermore, contrastive frequencies, attached to (roughly) synonymous collocations, would be useful: for example, DE *Vorschlag+machen* ('make+proposal') is more frequent than its synonym *Vorschlag unterbreiten*.

In addition there are other (maybe more successful) grouping and ordering principles; these allow the lexicographer not only to structure the collocation lists, but also to express more of the collocations' linguistic properties:

- syntactic subcategorization of the elements of the collocation (and, consequently, the insertion of a collocation into the sentence), as given in the DCF;
- semantic relatedness of different collocations, (quasi-)synonymy, as also in part given in the DCF;
- relatedness of a collocation with a reading of a polysemic base.

OCDSE has a broad syntactic classification of collocations, for example, into verb+subject and verb+object-collocations (cf. the entry in (Fig. 7)).

The syntactic construction of collocations is however not used as a grouping criterion. The main strength of OCDSE, in our view, is the use of a semantic grouping criterion (by semantic similarity: "onomasiological"), within the broad syntactic classification operated at the first grouping level. A third grouping level captures the relationship with readings of the base³ (cf. numbering 1, 2 in (Fig. 7)).

hope noun**1 belief that sth you want will happen**

- ADJ. considerable, fervent, great, high *a feeling of considerable hope* ◦ *It is my fervent hope that you will be able to take this project forward.* | high (only used with 'hopes') *Hopes are high that a resolution to the conflict can be found.* [...]
- QUANT. flicker, glimmer, ray, spark *I looked at her and felt a glimmer of hope.*
- HOPE + VERB lie, rest *Her only hope lay in escape.* [...]

2 sth you wish for

ADJ. high ~s *They have high hopes for their children.* [...]

Figure 7: Extracts from the entry s.v. *hope* in OCDSE

The relationship between collocations and the readings of a polysemous base may be expressed in different ways. The Van Dale dictionaries have introduced the cijfer-punt-cijfer-code, which relates collocations to numbered readings of the base (second number), and, at the same time, provides a classification by syntactic type (e.g. N/A vs. N/V, first number, cf. 1.1, 3.1 in (Fig. 2)). DCF only has a rudimentary grouping by readings. The entry s.v. *temps*, for example, includes collocations of all readings ("time", "times", "weather", etc.).

3.4 Other presentational issues: morphosyntax and text condensation

From the types of linguistic facts about collocations which need to be presented in a dictionary, the morphosyntactic properties have not yet been mentioned. It is vital for a learner to know about these. Mel'čuk's ECDs are, very explicit on determination in verb+object collocations.

Hoffnung[...] 1. 'starker Wunsch, Glaube, dass etw. Bestimmtes eintritt, geschieht od. dass etw., jmd. etw. Bestimmtes, Positives bewirkt'; ↗ FELD VII.6.1: *eine leise, schwache, vage ~ haben; das ist eine irrite, berechtigte ~; alle ~en waren auf ihn gerichtet; [...] große ~en auf die Kur, das wissenschaftliche Gutachten, den Arzt, Rechtsanwalt setzen; sich, jmdm. falsche, übertriebene ~en machen [...]*

Figure 8: Extracts from the entry s.v. *Hoffnung* in KEMPCKE'S WDaf

WDaf (cf. (Fig. 8)) and OCDSE indicate morphosyntactic preferences implicitly in the form of the collocation indications or in examples: *sich, jmdm falsche, übertriebene -en machen* in (Fig. 8). An exceptional case is the indication s.v. *high hopes* in OCDSE. More such indications would be welcome, and in line with OCDSE's programme to provide collocations in context.

A second, related issue concerns lexicographic text condensation. Examples where collocations are combined (cf. WDaf in (Fig. 8)) are functionally ambiguous: they may

illustrate a frequent triple, i.e. a "naturally occurring" combination of two collocations and thus be a useful element of a description of collocations in context, or they may be the result of a text condensation attempt. The latter is problematic, unless marked explicitly.

4. Conclusion

In this paper, we have examined current practice in the presentation of collocations in monolingual dictionaries (for bilingual dictionaries, see Hollös (2004)).

An analysis of OCDSE and DCF, two recent collocational dictionaries (published 2002 and 2003, respectively), shows major differences between the two. OCDSE is "onomasiological" in its microstructure, insofar as it groups collocations by readings of the base, then by syntax and, within each syntactic group by semantic relatedness (or sometimes quasi-synonymy). Furthermore, its graphical and layout devices support the structure of the entries. Collocations are contextualized by means of example sentences, yet more morphosyntactic details would be welcome.

DCF, on the other hand, is mainly "semasiologically" organized, with alphabetical lists of collocates in each entry. Semantic groups of collocations can at most be inferred from a comparison of the meaning explanations given, and, in cases of quasi-synonymous collocations, from the listing of all potentially synonymous items in the same place. DCF is, however, strong on the syntactic subcategorization side, as it has detailed formalized valency patterns for all collocations.

None of the dictionaries gives frequency data for collocations, and mostly, morphosyntactic properties (e.g. preferences for singular and plural) are indicated only explicitly. Furthermore, the status and form of examples is not always clear in most general definition dictionaries: examples may illustrate collocations or selectional properties, or even the (perhaps idiosyncratic) use of a word by a given author - all of this without visible distinction.

Thus, a few presentational desiderata remain open, also after the publication of OCDSE and DCF. Corpus linguistics is in a position to provide raw material, examples, and frequency data for most of the linguistic phenomena described in section 2. A modified version of the OCDSE model (which would include a simple frequency annotation of collocations, e.g. like the diamonds system for words, in Cobuild, as well as more morphosyntactic and syntactic details) could be a good basis for new collocation dictionaries.

Endnotes

1. Most recently, Hollös (2004) has extensively covered all of these aspects, from the point of view of bilingual learner lexicography. As her book has only appeared at the time of finalizing the present article, it was unfortunately impossible to keep track of it.
2. It would also be interesting to compare the collocations listed in different dictionaries of different size, all dependent on a common source. Do the small dictionaries only contain the most frequent collocations? Thanks to Werner Scholze-Stubenrecht for this suggestion.
3. This is almost as close as one can get to the rigorously semantic structuring advocated, for example, by Mel'čuk in the ECDs, without compromising the ease of use of the dictionary. Similar attempts at semantic grouping have so far mainly been undertaken in specialized lexicography (cf. Cohen 1986, Verlinde et al. 1993).

References

- Bahns, J., 1986. *Kollokationen als lexikographisches Problem – Eine Analyse allgemeiner und spezieller Lernerwörterbücher des Englischen*, Tübingen: Niemeyer.
- Cohen, B. 1986. *Lexique de cooccurrences -- Bourse – conjoncture économique*, Montreal: Linguatech.
- Cowie, A. P. 1978 "The place of illustrative material and collocations in the design of a learner's dictionary", in P. Strevens (ed.): *In Honour of A.S. Hornby*, Oxford: Oxford University Press.
- DCF: Zingle, H., Brobeck-Zingle, M.-L. 2003. *Dictionnaire combinatoire du français -- Expressions, locutions et constructions*, Paris: La Maison di Dictionnaire.
- Duden 2: Drozdowski, D. et al. 7/1988, 8/2001. *DUDEN. Stilwörterbuch de deutschen Sprache -- Die Verwendung der Wörter im Satz*, Mannheim: Dudenverlag.
- ECD: Mel'cuk, I.A. et al. 1984, 2: 1988, 3: 1992. *Dictionnaire explicatif et combinatoire du français contemporain*, Montreal: Presses de l'université de Montréal.
- LGDaF: Götz, D., Haensch, G., Wellmann, H. (eds.) 1993. *Langenscheidts Großwörterbuch Deutsch als Fremdsprache*, Berlin, München: Langenscheidt.
- Hausmann, F. J.: 1989. "Le dictionnaire de collocations", art. 95 in: F.J. Hausmann et al. (eds.): *Wörterbücher -- Dictionaries -- Dictionnaires. Ein internationales Handbuch*, Berlin: De Gruyter).
- Hausmann, F. J. 2003. Was sind eigentlich Kollokationen?" in *Akten der Jahrestagung 2003*, Mannheim: IDS.
- Hollo, Z. 1989 *Lernerlexikographie: syntagmatisch. Konzeption für ein deutsch-ungarisches Lernerwörterbuch*, Tübingen: Niemeyer.)
- Ilgenfritz, P. et al. 1989. *Langenscheidts Kontextwörterbuch Französisch-Deutsch*, Berlin/München: Langenscheidt.
- Lacroix, U.: 1931. *Les mots et les idées -- Dictionnaire des termes cadrant avec les idées*, Paris: Fernand Nathan.
- OALD-4: Hornby, A. S. 1989. *Oxford Advanced Learner's Dictionary of current English, Fourth Edition* (A.P. Cowie), Oxford: Oxford University Press.
- OCDSE: Crowther, J. et al. 2002. *Oxford Collocations Dictionary for students of English*, Oxford: Oxford University Press.
- Palazzi/Folena: Palazzi, F. and Folena, G. 1992. *Dizionario della lingua italiana*, Torino: Loescher Editore.
- Schafroth, E. 2003. "Kollokationen im GWDS", in: H.E. Wiegand (ed.): *Untersuchungen zur kommerziellen Lexikographie der deutschen Gegenwartssprache*, Vol. I, Tübingen: Niemeyer.
- VDGW-NF: Al, B. P. F. et al. 1985. *Van Dale Groot Woordenboek Nederlands-Frans*, Utrecht/Antwerpen: Van Dale.
- Verlinde, S., Binon, J. and van Dyck, Jan 1993. *Dictionnaire contextuel du français économique*, Leuven: Garant.
- WDaF: Kempcke, G. et al. 2000. *Wörterbuch Deutsch als Fremdsprache*, Berlin, New York: Walter de Gruyter.
- Zöfgen, E. 1986. "Kollokation -- Kontextualisierung – (Beleg-)Satz. Anmerkungen zur Theorie und Praxis des lexikographischen Beispiels", in A.Barrera-Vidal et al. (eds.): *Französische Sprachlehre und bon usage. Festschrift für Hans-Wilhelm Klein zum 75. Geburtstag*, Ismaning: Hueber.