

lemma in such an entry is polysemous, the indication of etymology has thus been subjected to right-extrapolation (see further pp. 352f.).

4.3.3 Textual condensation now and in the future

All the examples shown in this section have been taken from general-purpose dictionaries for adults, which is the type of dictionary where textual condensation is most conspicuous. However, what is gained through space-saving devices is often counterbalanced by a loss of clarity, and in dictionaries intended for user categories such as children and students, a greater portion of the information conveyed must be presented as full text. To this may be added that, generally speaking, demands for greater user-friendliness are continuously made on dictionaries today, and, as a result, the degree of textual condensation (the **PROPORTIONAL DENSITY**) has decreased in many general-purpose dictionaries, even in those intended for adults. For example, the presentation of grammatical information has been simplified, the lemma is less often replaced by representative symbols, the presentation of alternative wordings is not so complicated as before, etc.

As mentioned earlier, the main purpose of textual condensation is a space-saving one. However, the need to save space, and thus the need for textual condensation, applies primarily to print dictionaries; in an electronic dictionary, the amount of space available is almost unlimited. The question then arises as to whether or not textual condensation should occur in electronic dictionaries at all. Matters of this kind are discussed in Chapter 26 (pp. 439–441).

Literature

Dictionaries and their components as text: *Lexicographica* 6 (1990) (thematic part).
 Components and structures of general-purpose dictionaries: Hausmann & Wiegand 1989 (monolingual dictionaries); Hausmann & Werner 1991 (bilingual dictionaries).

Different types of dictionary structure and their relationships to different types of lexical information: see the chapters on microstructure (Ch. 20), macrostructure (Ch. 21), megastructure (Ch. 22) and cross-reference structure (Ch. 23), and also the chapters on the individual information types.

Textual condensation: Wolski 1989a, Gouws 2005c (monolingual dictionaries); Wolski 1991 (bilingual dictionaries); Corréard 2002 (relevance of textual condensation in electronic dictionaries); Gouws 2005b (niching).

The role of layout and typography: Almind & Bergenholz 2000; Luna 2004; Gallard 2005.

Svensén (2003) Handbook of Lexicography

5 The lemma

Mais qu'est-ce qu'un mot? Question simple en apparence, sauf pour les linguistes. ALAIN REY

5.1 Functions of the lemma

The functions and theoretical status of the **LEMMA**¹ have been given rather differing descriptions in metalexicographic literature (Wolski 1989b); the way it is treated here implies a certain simplification which aims mainly at meeting the needs of practical lexicography.

The lemma functions as a representative of a linguistic sign; in a dictionary it represents the lexical item described in the individual dictionary entry (the **LEMMA SIGN**)².

In a dictionary having an alphabetical macrostructure (see p. 368), the lemma also has the function of determining the position of the entry in the lemma list. The characteristic of the lemma that consists of a certain sequence of letters and, hence, determines the position of the entry in the lemma list is termed the **GUIDING ELEMENT** of the lemma, and the lemma itself is said to be the **GUIDING-ELEMENT CARRIER** (Wiegand 1989: 372).³

It is through the lemma that the users find the information provided about the lemma sign. In some cases, they will not be given direct access to this information but be cross-referred to another lemma.⁴

The lemma thus forms part of the macrostructure of the dictionary. However, in this book it is regarded as forming part of the dictionary's microstructure

¹ A common synonym of 'lemma' is 'headword'. In this book, the term 'lemma' has been preferred mainly because 'headword' will be somewhat problematic when the lemma consists of a multi-word lexical item (see below, pp. 102–4).

² One and the same lemma may also represent several lemma signs (see p. 106). – For the sake of simplicity, the term 'lemma sign' is used in this book only when there is a special need to distinguish the linguistic sign described in an entry from the representative of that sign heading the entry. Where no such distinction is thought to be necessary, the simple term 'lemma' is used throughout.

³ It is this only when the macrostructure is alphabetical that the lemma functions as a guiding element carrier (see pp. 377f.).

⁴ A lemma whose only function is to cross-refer to another lemma is termed **CROSS-REFERENCE LEMMA**.

(see p. 344) as well, mainly because it is, directly or indirectly, the address of most of the information items appearing in an entry and also functions itself as a carrier of certain types of information about the lemma sign (spelling, word division, pronunciation, etc.; see pp. 109–12, 116–19).

5.2 Establishment of lemmas

'What, actually, is a word?' is the question asked in the motto of this chapter. The matter is not without interest in dictionary-making since it is of decisive importance for the **ESTABLISHMENT OF LEMMAS**. This operation is not simply the same thing as lemmatization (cf. p. 52). Naturally, the establishment of lemmas assumes previous lemmatization, but it also includes deciding how lexical items having identical base forms are to be presented in the dictionary, and to what extent word elements and multi-word lexical items are to be accorded lemma status.

5.2.1 Homonymy and polysemy

In the compilation of a dictionary, uniform principles must be applied in deciding what should be entered under one and the same lemma. Above all, it must be determined how lexical items having identical base forms are to be presented. There are two possibilities: either they should be considered to belong to the same lemma, in which case only one lemma needs to be established in the dictionary, or they should be considered to belong to different lemmas, which requires the establishment of two or more lemmas.

5.2.1.1 Some terms and concepts

The base forms of two given lexical items may be identical in various respects. If the two forms are identical as regards spelling, there is a relationship of **НОМОГРАФИЯ** between them (i.e. they are **НОМОГРАФНЫ**):

lead [led] 'a metal'
lead [li:d] 'go in front'

If the two forms are identical as regards pronunciation, there is a relationship of **НОМОФОНИЯ** between them (i.e. they are **НОМОФОННЫ**):

sight [saɪt] 'seeing'
site [saɪt] 'place'

Homography and homophony can obviously occur together:

ball [bɔ:l] 'round object'
ball [bɔ:l] 'dance'

Two lexical items where the base forms are identical as regards spelling and pronunciation may also be identical or different as regards other characteristics, for example inflection and part-of-speech membership. In the case of *ball*, both are the same:

ball *n.* -s 'round object'
ball *n.* -s 'dance'

In the following case, there is identity as to part of speech but not as to inflection:

weave *vb.* *wove* *woven* 'interlace'
weave *vb.* *wove* *woven* 'dodge'

If the two items differ as regards part of speech, they probably differ as to inflection as well:

bear *n.* -s 'an animal'
bear *vb.* *bore* *borne* 'carry'

A lexical item consisting of a certain limited meaning on the content side and a certain form on the expression side is sometimes called a **ЛЕКСЕМА**. Two lexemes are:

bar *n.* -s 'rod'
bar *n.* -s 'drinking place'

In this case, there is complete formal agreement. The next pair of lexemes show formal differences:

weave *vb.* *wove* *woven* 'interlace'
weave *vb.* *wove* *woven* 'dodge'

In Table 5.1 (p. 96), different types of likeness and difference between lexemes are summarized.

We have thus to take a view about a number of lexemes. They differ in meaning, but they are alike in that their base forms have the same spelling; otherwise they present differences as regards pronunciation, inflection and part-of-speech membership. The problem is to decide whether the lexemes of a given pair should be regarded as belonging to the same lemma or not. In other words:

- Are the lexemes to be regarded as representing different meanings of the 'same' word? If so, there is a relationship of **ПОЛЫСЕМЬ** between them.
- Are the lexemes to be regarded as two 'different' words? If so, there is a relationship of **НОМОНИМ** between them.

Table 5.1 Comparison of lexeme pairs

Pairs of lexemes	Agreement as regards:				Basic meaning
	Pronunciation	Part of speech	Inflection	Eymology	
crown 'headgear of king' 'top branches of tree'	+	+	+	+	+
bar 'rod' 'drinking-place'	+	+	+	+	-
ball 'round object' 'dance'	+	+	+	-	-
weave 'interlace' 'dodge'	+	+	-	-	-
bear 'an animal' 'carry'	+	-	-	-	-
lead 'a metal' 'go in front'	-	-	-	-	-

Normally, polysemy is manifested in the *microstructure* of the dictionary, the lexemes being presented as different senses within one and the same entry. Homonymy is manifested in the *macrostructure* of the dictionary, the lexemes being presented in different entries.

There are several approaches to the problem of polysemy-homonymy, which, in turn, have given rise to different methods in practical lexicography. In what follows, four of them will be presented and discussed.

5.2.1.2 The historical approach

The traditional approach is based on historical criteria. Lexemes with the same form but different origin are treated as homonyms and presented in separate entries:

ball¹ *n.* -s 'round object' [of Germanic origin]
ball² *n.* -s 'dance' [of Romance origin]

However, there is often inconsistency here. For example, *hull* (of a ship) is usually given a separate entry despite its being derived from *hull* in the sense 'outer covering of fruit, pod, husk', so that etymologically there should be only one entry.

Such inconsistencies in applying the historical approach are not unusual, and this is not surprising. There are various degrees of relatedness between words, and it is hard to establish a general rule as to where the line is to be drawn between relatedness and non-relatedness. It is indeed debatable whether or not users derive

any benefit from a historically based language description of the kind implied by this treatment. Above all, it must be regarded as difficult and inconsistent, seeing that the users are probably without any etymological knowledge.

Below is an invented (and possibly over-explicit) example showing how the word form *match*, together with a selection of compounds and derivatives, might be treated applying the historical approach:⁵

- ¹**match** *I n.* (-es) **1** game, competition **2** one who is (somebody's) equal
3 something that combines well **4** possible wife/husband; marriage **II**
vb. (-ed, -ing, -es) **1** cause to compete **2** be (somebody's) equal **3**
combine well [of Germanic origin]
²**match** *n.* (-es) wooden stick for making fire [of Romance origin]
matchbox *n.* box of wooden sticks for making fire
matching *adj.* combining well
matchless *adj.* unequalled
¹**matchmaker** *n.* **1** one who attempts to arrange a marriage **2** one who arranges competitions
²**matchmaker** *n.* manufacturer of wooden sticks for making fire
¹**matchmaking** *n.* **1** attempt to arrange a marriage **2** arrangement of competitions
²**matchmaking** *n.* manufacture of wooden sticks for making fire
match point *n.* decisive situation in game
matchwood *n.* wood used for manufacture of sticks for making fire

As can be seen, ¹**match** *I* (the noun) and ²**match**, although identical as regards part of speech and inflection, have been treated as homonyms owing to their different origin. On the other hand, under ¹**match**, ¹**matchmaker** and ¹**matchmaking** there is polysemy because of the common origin of the lexemes. According to traditional principles of arrangement, ¹**match** *I* and ¹**match** *II*, although differing as regards part of speech and inflection, have been brought together because of their common spelling and origin.

5.2.1.3 The semantic approach

A synchronic contemporary dictionary is not suitably based on historical principles. Attempts have therefore been made to solve the problem by replacing the historical approach with a semantic one, where lexemes having the same form but quite different meanings are treated as homonyms. This implies that *hull* of a ship and *hull* of a strawberry are 'different' words and will appear in separate entries. If,

⁵ The information given about meaning and etymology has as its sole purpose to identify the different lexemes.

however, the meanings can be derived from the same basic meaning, the word is polyssemous. Thus *crowm* 'headgear of king' and *crowm* 'top branches of tree' are two meanings of the 'same' word and will appear in the same entry.

Obviously, however, it is again difficult to set up uniform rules. Where should the line be drawn between 'quite different meanings' and 'the same basic meaning'? Indeed, what is a 'basic meaning'?

If the semantic approach is applied in a synchronic contemporary dictionary, the word form *match*, with its compounds and derivatives, might be treated in the following way:

- ¹match *n.* (-es) game, competition
- ²match *n.* (-es) one who is (somebody's) equal
- ³match *n.* (-es) something that combines well
- ⁴match *n.* (-es) possible wife/husband; marriage
- ⁵match *n.* (-es) wooden stick for making fire
- ⁶match *vb.* (-ed, -ing, -es) cause to compete
- ⁷match *vb.* (-ed, -ing, -es) be (somebody's) equal
- ⁸match *vb.* (-ed, -ing, -es) combine well
- matchbox *n.* box of wooden sticks for making fire
- matching *adj.* combining well
- matchless *adj.* unequalled
- ¹matchmaker *n.* one who attempts to arrange a marriage
- ²matchmaker *n.* one who arranges competitions
- ³matchmaker *n.* manufacturer of wooden sticks for making fire
- ¹matchmaking *n.* attempt to arrange a marriage
- ²matchmaking *n.* arrangement of competitions
- ³matchmaking *n.* manufacture of wooden sticks for making fire
- match point *n.* decisive situation in game
- matchwood *n.* wood used for manufacture of sticks for making fire

Here it should be noted that all the *match* lexemes, as well as all the *matchmaker* and *matchmaking* lexemes, despite their formal likeness, have been presented as homonyms because of their semantic differences, and there are no instances of polysyny.

5.2.1.4 The morphosemantic approach

In French monolingual lexicography, in particular, an approach has been developed that is partly based on semantic criteria but nevertheless differs radically from the one described above. Each entry contains a single 'main lexeme', possibly with extended and specialized sub-senses. Two lexemes having the same formal features (and often the same origin) will thus be regarded as instances of homonymy, not polysyny. In this way, the microstructure is 'relieved' and the macrostructure

made more complicated instead. A word such as the French verb *commander* is therefore treated as follows:

- ¹commander *v.t.* 'influence'
- ²commander *v.t.* 'order (goods)'
- ³commander *v.t.* 'control (a machine, etc.)'

This method, called in French *dégroupement* (in English *NONHOMONYMIZATION*), is thus characterized by the lexemes being treated as homonyms on a semantic basis. However, the morphosemantic approach, as indicated by the term, has other implications as well. The structure of the 'main lexemes' is based not only on the semantic properties of the words but also on their morphological properties, since different lexemes are often accompanied by different sets of derivatives and compounds. The entry, therefore, may contain not only the 'main lemma' but also a number of *sublemmas*, i.e. the derivatives and compounds based on it. Hence, under '*commander*', all the derivatives and compounds of the verb where the meaning 'influence' is involved are brought together and defined. Under '*commander*', the derivatives and compounds involving the meaning 'order' are treated, etc. This implies, in turn, that a given derivative or compound may appear under several 'main lexemes', all depending on what lexeme(s) it is connected to semantically. The procedure of bringing together derivatives and compounds on a semantic basis, which is called *regroupement* (in English *grouping*), is regarded as complementary to *dégroupement*.

The morphosemantic approach, if applied to the word form *match*, could yield a result such as the following:

- ¹match *n.* (-es) game, competition – ¹matchmaker *n.* one who arranges competitions – ¹matchmaking *n.* arrangement of competitions – match point *n.* decisive situation in game
- ²match *n.* (-es) one who is (somebody's) equal – matchless *adj.* unequalled
- ³match *n.* (-es) something that combines well
- ⁴match *n.* (-es) possible wife/husband; marriage – ²matchmaker *n.* one who attempts to arrange a marriage – ²matchmaking *n.* attempt to arrange a marriage
- ⁵match *n.* (-es) wooden stick for making fire – matchbox *n.* box of wooden sticks for making fire – ³matchmaker *n.* manufacturer of wooden sticks for making fire – ³matchmaking *n.* manufacture of wooden sticks for making fire – matchwood *n.* wood used for manufacture of sticks for making fire
- ⁶match *vb.* (-ed, -ing, -es) cause to compete
- ⁷match *vb.* (-ed, -ing, -es) be (somebody's) equal
- ⁸match *vb.* (-ed, -ing, -es) combine well – matching *adj.* combining well

In this example, it should be noted that the five noun lexemes **match**, as well as the three verb lexemes, despite the formal likeness within each group, have been treated as homonyms. The reason is that they are different as regards meaning, but also that they are accompanied by different sets of derivatives and compounds. Furthermore, if one and the same derivative or compound appears under several of the homonymous base words (as is the case with **matchmaker** and **matchmaking**), there will be homonymy at this level as well.

5.2.1.5 The formal-grammatical approach

Another way of resolving the problem of homonymy–polysemy is to base the arrangement on purely formal and grammatical criteria, which has the advantage of seldom causing any problems of demarcation. Lexemes spelt alike but belonging to different parts of speech or otherwise formally different are to be regarded as ‘different’ words (homonyms) and be presented in separate entries. Polysemy means differences, from a semantic point of view, between lexemes with the same spelling, the same part-of-speech membership and otherwise the same formal features, and these lexemes are presented in the same entry.

We now return to the word form *match* for the last time, in order to see how it could be treated applying the formal-grammatical approach:

- ¹ **match** *n.* (-es) 1 game, competition 2 one who is (somebody’s) equal 3 something that combines well 4 possible wife/husband; marriage 5 wooden stick for making fire
- ² **match** *vb.* (-ed, -ing, -es) 1 cause to compete 2 be (somebody’s) equal 3 combine well
- matchbox** *n.* box of wooden sticks for making fire
- matching** *adj.* combining well
- matchless** *adj.* unequalled
- matchmaker** *n.* 1 one who attempts to arrange a marriage 2 one who arranges competitions 3 manufacturer of wooden sticks for making fire
- matchmaking** *n.* 1 attempt to arrange a marriage 2 arrangement of competitions 3 manufacture of wooden sticks for making fire
- match point** *n.* decisive situation in game
- matchwood** *n.* wood used for manufacture of sticks for making fire

An approach based solely on formal criteria is of course exceedingly well suited for data processing. However, when applied to practical lexicography, it can sometimes result in a method of presentation that may seem a bit strange and counter-intuitive to the user. For instance, lexemes that, according to the user’s linguistic instinct, have nothing semantically in common may appear under the same lemma. In the above example, **match** 5 looks out of place alongside the other

lexemes under the lemma concerned, as is the case with **matchmaker** 3 and **matchmaking** 3. Conversely, if the formal approach is applied strictly, there will be cases where semantically related lexemes having the same base form and the same part-of-speech membership, contrary to the linguistic instinct of the users, must be separated because there are inflectional differences:

- ¹ **antenna** *n.* (pl. -ae) sensory organ of insects and crustaceans
- ² **antenna** *n.* (pl. -as) aerial
- ¹ **speed** *n.* rapidity of movement)
- ² **speed** *vb.* (*speed speed*) 1 move quickly 2 cause to move quickly
- ³ **speed** *vb.* (*speeded speeded*) drive faster than is permitted

5.2.1.6 Homonymy and polysemy in practice

The methods of presenting words of this kind differ from dictionary to dictionary, and there are many intermediate and mixed forms of the approaches just presented (see further pp. 364–7). When assessing the advantages and drawbacks of different methods of presentation, focus should be on not only the underlying principles and their capability of providing an adequate description of the language. It is at least as important to form an opinion of their respective value to the users, asking oneself questions such as: How do users behave intuitively when searching for information in a dictionary? What kinds of information do they need? Do these needs vary according to the type of dictionary involved, e.g. general-purpose dictionary v. learner’s dictionary, L1 dictionary v. L2 dictionary, monolingual dictionary v. bilingual dictionary?

The weakness of the historical approach is, as already mentioned, that the users, lacking etymological knowledge, are as a rule completely unable to understand the principles underlying the method of presentation. Therefore, this approach may be considered to be partly counterintuitive as far as the users are concerned.

As for the semantic approach, we have seen that the uncertainty as to what should be regarded as ‘quite different meanings’ may cause difficulties. But these difficulties are first and foremost the concern of the lexicographer; they do not affect the users since they have no preconceived ideas in this respect. Instead, one should consider the result: a dictionary where the microstructure is less complicated and where there is, instead, a richer macrostructure, which in all probability will contribute to greater clarity and also be pedagogically advantageous.

The morphosemantic method has, through its rich macrostructure, partly the same advantages as the semantic one. The fact that each ‘main lexeme’ is accompanied by its derivatives and compounds can, in a learner’s dictionary, be favourable in terms of vocabulary building. However, in a general-purpose dictionary, which is intended solely for consultation, it may cause difficulties since the user

cannot be expected to know under which lexeme (or, rather, lexemes) a certain compound or derivative should be sought. In addition, this approach requires a great number of cross-references, which will further aggravate the search problems.

Finally, the formal-grammatical approach. It is characterized by a high degree of consistency: by applying a few simple rules the users can always be sure of finding the lexeme they are looking for. However, the fact that this approach may lead to the bringing together of formally equivalent but semantically quite different lexemes, and to the separation of semantically related lexemes exhibiting only slight formal differences, must be regarded as a disturbing element.

5.2.2 Multi-word lexical items featuring as lemmas

We will now return for a moment to the question asked in the motto of this chapter: 'What, actually, is a word?' A typographer would probably answer that a 'word' is a group of characters placed together with spaces or punctuation marks before or after. But how, then, should we deal with an expression such as *air mail*? Does it consist of two words, whereas for example *airline* and *air-bed* each consist of one word? On this point, some dictionaries follow the purely graphic criterion and enter the open compounds as examples (with definitions or equivalents) under one of the simplex words.

However, in English a hyphenated form such as *air-mail* is as plausible as the one written separately. In consequence, the user may have to look in two different places in the dictionary before the expression is found. If it comes to the worst, it may be necessary to look in three places, as the user cannot know in advance whether *air mail* is entered under *air* or under *mail*. These three expressions have been formed in an exactly parallel way, and the graphic form cannot be held to justify treating them in different ways, particularly if the aim is to make things easier for the user. The conclusion is that items of this kind, whether written separately, hyphenated or solid, should be accorded the same lemma status.

The treatment of lexical items consisting of several graphic words (**MULTI-WORD LEXICAL ITEMS**) in dictionaries has been, and still is, the subject of a lively discussion, and practice differs between object languages as well as between dictionaries. It is then not only a matter of deciding whether or not the item should be accorded lemma status (i.e. appear as a **MULTI-WORD LEMMA**); it must also be determined where in the dictionary it is to be presented. The latter problem will be treated on pp. 194–8, 353–60, 369f.; here, only the former will be discussed.

The basic prerequisite for accorded lemma status to a multi-word item is that it has undergone some kind of **LEXICALIZATION**, i.e. that it has been stored in our

mental lexicon as a unit. Among the typical cases of lexicalization the following have been mentioned (see, for instance, Zgusta 1971: 144):

- (a) the case exemplified by French *pomme de terre*, on the grounds that the meaning of the expression cannot be deduced from the meanings of its component parts (it has nothing to do with 'apples in the ground');
- (b) a second case, also exemplified by French *pomme de terre*, in this instance on the grounds that the expression refers to a single concept, as can be deduced from the fact that it is possible to make exchanges such as *nous mangeons des pommes de terre* – *nous mangeons des carottes*;
- (c) as in French *bien que*, on the grounds that the expression can be replaced by a synonym consisting of a single graphic word (*quoique*).

An additional possible case is (Rey-Debove 1971: 117):

- (d) as in French *bon enfant*, on the grounds that it can form single-word derivatives (*bonenfantisme, bonenfantise*).

In this context, however, there may be differences between language groups. For example, Schnorr (1991: 2815f.) points out that compound expressions such as *data processing* and *space shuttle* normally are understood as concepts by native speakers of Germanic languages and, therefore, can be accorded lemma status in dictionaries intended for these user groups. On the other hand, it may be more difficult for those not familiar with Romance languages to realize that the same thing is true of *homme de terre* and *femme de chambre*, or to see the difference between the lexicalized *la femme de chambre* and the non-lexicalized *la femme du maître*.

Generally speaking, there is now a tendency to accord, to a greater extent than before, lemma status to lexicalized noun phrases (multi-word items), not least in L2 dictionaries and in dictionaries with a nesting macrostructure (see pp. 374–6). To some extent, this development is also reflected in many general-purpose dictionaries and in dictionaries where nesting does not occur.

The last element of such an item will then be treated in different places in the dictionary, which may imply a certain loss of information. This can be compensated by listing, in the entry for the last element, all the compounds which include that element (Zgusta 1989: 302).

Verb phrases may also be entered in the lemma position. For instance, in English dictionaries this is done with phrasal verbs, either with a particle (e.g. *find out*), with a preposition (e.g. *go to*) or with both (e.g. *check up on*). This method has some pedagogical value, but only in so far as the users find the information they are looking for. It may be necessary to look in two places, which may be a long way apart, as in *do ... door ... do without*. This problem can be avoided if the entries are

arranged according to the word-by-word principle, which takes the word spaces into account (see further p. 369), or if phrasal verbs are entered as sublemmas (see further pp. 232, 361).

Idioms are normally regarded as independent lexical items which should be accorded lemma status. However, a difficult problem is involved here, namely that of deciding where in the lemma list they are to be entered (see further pp. 194–8).

The increasing tendency to accord lemma status to noun phrases and verb phrases has made some contribution towards making these items easier to find in the dictionary. The search problems are even smaller in electronic dictionaries, since the search facilities available there make it easy for the user to access such an item, irrespective of its status in the dictionary text; it hardly matters if it is entered in the lemma position, as a sublemma or merely as a defined/translated example in the entry for one of its component parts.

5.2.3 Word elements featuring as lemmas

The possibility of presenting productive prefixes, suffixes, first elements and last elements in the lemma position has already been touched upon (see p. 66). These types of lemma have a special function in that they do not directly give the meaning of words that may be encountered in a text. Instead, they equip the users with a key to the systematics of word formation and provide a means of working out by analogy the meaning of words that are not entered in their full form (see further pp. 132f.).

Showing word elements in the lemma position is appropriate primarily in dictionaries that are intended mainly for reception. The introduction of this type of information in a dictionary intended for production, on the other hand, involves a certain amount of risk. The same thing can be expressed with different prefixes or suffixes; a given last element does not form a compound with every first element. Also, the user of a production dictionary, who is presumed to have but limited knowledge of the object language, has no means of knowing what to do in a given situation. However, the risk of misunderstanding can be reduced by giving, for every word element, a number of typical examples.

5.2.4 Abbreviations and clippings featuring as lemmas

Abbreviations may be treated in various ways. In the past, it was common to list them separately in an appendix. Nowadays, there is an increasing tendency, above all in bilingual dictionaries, to include them in their alphabetical positions in the

lemma list. In an L2→L1 dictionary, the explanation of an abbreviation most often consists of an (implicit) cross-reference to the unabbreviated form:

ONU *f. abbr. of Organización de las Naciones Unidas* UN

Abbreviations should not be confused with clippings, which are usually more colloquial than the corresponding full forms:

Kripjo *short for* Kriminalpolizei

Ed *short for* Edgar, Edmund, Edward

5.3 Grammatical form of the lemma

The form of the lemma sign that appears as a lemma in the dictionary – the **LEMMA FORM** (or **LOOK-UP FORM**, or **CITATION FORM**) – functions as a representative of the whole paradigm of the lemma sign and is selected from this paradigm according to certain conventions. Provided that the structure of the object language in question allows it, the grammatical forms most used as lemma forms are:

- nouns in the singular (nominative);
- adjectives in the positive singular (nominative, masculine);
- adverbs in the positive;
- verbs in the (present) infinitive (active).

There are some exceptions to this, and practice varies between dictionaries. The important thing here is to give a correct description of the entry language, while at the same time meeting the practical needs of the users. For instance, if a noun is used only in the plural, the plural form should be entered as lemma form, for example *scissors*. However, if the base form is unusual but not quite out of use, it may nevertheless be necessary to include it in the lemma position, indicating that the form is hardly used and cross-referring to the current (inflected) form.

It has also been discussed whether or not the infinitive is really the most suitable lemma form of a verb. In the case of certain languages, for instance Swedish, it has been maintained that from a pedagogical point of view it would be better to enter a verb in the present tense, with this as a basis it is easier to derive correctly the other forms of the verb than doing so from the infinitive. This method has a certain tradition behind it, being the normal one in, for instance, Greek and Latin dictionaries.

In order to facilitate the use of the dictionary, irregular inflected forms are normally entered in the lemma position, with cross-references to the base forms. Practice differs here according to dictionary type: L2→L1 dictionaries will normally provide such cross-references more liberally than L1 dictionaries, and the

same is true of learners' dictionaries in comparison with general-purpose dictionaries. Usually, the irregular forms appearing in the lemma position are restricted to plurals, feminine forms, forms of comparison, principal parts, etc.

Another problem in this context concerns the treatment of inflected forms having other semantic and/or syntagmatic properties than the base form. For instance, should a German-English L2→L1 dictionary give information about *Daten* 'data' under *Datum* or under a separate lemma *Daten*, and should an English-German L2→L1 dictionary explain *arms* 'weapons' under *arm* or under a separate lemma *arms* (Schoor 1991: 281-4)? The meaning of a plural form of this kind is not reflected in the singular form, and it is debatable whether the singular should really be regarded as the lemma form in such a case. The conclusion is that it may often be appropriate to treat certain inflected forms as lemmas in their own right, for instance plurals of nouns (as in the examples above), comparative and superlative forms of determiners/pronouns such as *more/most* and *less/least*, and participles used as adjectives such as *lost*, *missing*, etc.

5.4 Presentation of the lemma

Typographically, the lemma is usually characterized by the structure indicator bold or semi-bold type, sometimes in combination with a larger type size than the rest of the entry, differing typeface or even differing colour.

The lemma form may be subjected to textual condensation. For instance, part of it may be shown between brackets, indicating that there is a shorter form:

glycerine[e]

This notation may cause a loss of information since the user has no means of knowing which is the main form and which is the variant. Therefore, it should be used only where there are no differences as to meaning and usage between the two alternatives.

A further type of textual condensation of the lemma form that may occur in, for instance, bilingual dictionaries where French is the source language involves the microstructure of the dictionary:

blond, -e 1 *adj.* blond II *n.* 1 *m/f* (person) blond(e) 2 *m* (colour) blond 3 *f* (beer) lager

In both these cases a MULTIPLE LEMMA, i.e. a lemma representing several lemma signs, has been created (*glycerin* and *glycerine*, and the two nouns **blond** and **blonde**).

Another type of textual condensation involving the lemma form occurs at macrostructural level (see further pp. 371-6). If every text block in the dictionary contains only one lemma, this must appear as full text (FULL LEMMA):

Gold mmmmmmmmmmmmmmmmm

mm

Goldader mmmmmmmmmmm

mmmmmmmmmmmmmmmm

Goldbrokat mmmmmmmmmmm

mmmm mmmmmmm

On the other hand, if the block contains several lemmas (i.e. it is a niche or a nest), the lemmas may be shown as full text, but they are often subjected to textual condensation. A condensed lemma (a PARTIAL LEMMA) is normally introduced by a representation symbol, e.g. a hyphen or a swung dash:

Gold mmmmmmmmmmmmmmmmm

mmmm -**ader** mmmmmmmmmmmmm

mmmmmmmmmmmm -**brokat** mmmmm

mmmmmmmmmmmmmmmmmmmm

Condensation of lemmas will obviously make searching in the dictionary more difficult. There may be long entries at the start of a nest, and it can often be difficult to find, far along in the nest, a partial lemma that may consist only of a representation symbol and one or two alphabetical characters:

fondlerie ... [1 line] -**eur** ... [2 lines] -**is** ... [1 line] -**oir** ... [1 line] -**re** ...

[21 lines] -**rière** ... [1 line] -**s** ... [23 lines] -**u**

It will probably be hard for the user to spot 's' twenty-seven lines into the nest and 'u' after fifty lines in all. This method of presentation is gradually being abandoned in modern dictionaries, at least in dictionaries where the density is low and the entries, consequently, rather long. In such cases, the small gains made in space are lost many times over in clarity. In dictionaries where the density is high and the entries short, more space will be saved and the loss in clarity not be so great.

However, even in the latter case one must beware of carrying one's space-saving zeal so far as to go beyond what can be regarded as justified from the linguistic and pedagogical point of view. Above all, one should not base the condensation on spelling only; the partial lemma must at least in some sense be a morphologically meaningful unit. An authentic example (Rettig 1985: 110):

gigantesque [...] -**antisne** [...] -**olio** [...] -**ot** [...]

As pointed out earlier, homonymous lemmas are usually marked in order to be distinguishable from each other. This marking normally consists of a numeral

(a **HOMOXYM NUMBER**), which is placed before or after the lemma. It may have the same type size as the lemma or be a small-type superscript number:

1 force <i>n.</i>	¹ force <i>n.</i>
2 force <i>vb.</i>	² force <i>vb.</i>
force (1) <i>n.</i>	force ¹ <i>n.</i>
force (2) <i>vb.</i>	force ² <i>vb.</i>

It is not of any great importance which system is chosen: the main thing is that the marking is made so that there is no risk for it (a) to be interpreted as being part of the spelling, or (b) to be confused with the section marks (see pp. 350f., 360f.).

There are dictionaries which do not use homonym numbers at all, in some cases without putting anything in their place. However, some modern learners' dictionaries applying the semantic approach to the problem of polysemy-homonymy (see pp. 97f.) have replaced homonym numbers by **GUIDE WORDS** (or **SIGNPOSTS**) directing the user to the relevant information (see further p. 351):

mean [EXERCISE] <i>verb</i>
mean [HAVE RESULT] <i>verb</i>
mean [INTEND] <i>verb</i>
mean [NOT GENEROUS] <i>adj.</i>
mean [NOT KIND] <i>adj.</i>
mean [HAVE IMPORTANCE] <i>verb</i>
mean [WOLLENT] <i>adj.</i>
mean [VALUE] <i>noun</i>

(adapted from CALD)

Literature

The 'lemma' (the term and the concept), types of lemma, establishment of lemmas: Wojski 1989b (general); Schuor 1991 (bilingual dictionaries).

Homonymy v. polysemy: Zöfgen 1989 (monolingual dictionaries in general), 1994: 84–104 (learners' dictionaries); Moon 1987 (considerations made in the *COBUILD* project); Herbst 1996: 348–9 (comparison of English L2 dictionaries); Popp 1999: 131–3 (development in English L2 dictionaries); Zwaneburg 1983 (*dégroupement* and *regroupement*); Bornas 1986 (the morphosemantic approach as applied in French dictionaries).

Lemmaic status of multi-word items: van der Meer 1996.

Word elements featuring as lemmas: Müller 1989b (prefixes, prefixoids, suffixes and suffixoids); Alsina & DeCesaris 1998 (suffixes).

6 | Spelling and word division

Dictionaries are more responsive to usage in the matter of pronunciation than they are in spelling. ALLEN WALKER READ

6.1 Spelling

Spelling is one of the most important types of information in a general-purpose dictionary. For instance, user surveys have shown that the need for information about spelling is one of the most frequent reasons for consulting a dictionary or for acquiring a dictionary at all. Also, spelling is one of the information types that are most frequently involved in activities concerning language purity or standardization.

Information about spelling is conveyed by the lemma, for which this is one of the main functions. However, what is shown by the indication of spelling (i.e. by the lemma form) is only a fraction of the orthographic realization of the lemma sign, since it includes only one of the forms of the paradigm. The spelling of the other forms must be inferred from the lemma form, and, in cases where the users cannot be expected to be capable of doing this, the spelling must be inferred from the indications of inflection.

A user who, in a production mode, is unsure of the spelling of a word, or makes an incorrect assumption about the spelling of a word, may need help to find the right entry in the dictionary. Therefore, it will not be completely inappropriate to show frequent misspellings in the lemma position, indicating that the spelling is incorrect and cross-referring to the correct form. The incorrect form should preferably not, like a 'normal' lemma, be shown in bold type:

embarrass incorrect spelling of EMBARRASS

In electronic dictionaries, there is no need to search in various places in order to find the correct spelling, since the dictionary automatically displays a list of alternative suggestions if an incorrect or incomplete spelling is input by the user (see p. 442).

The question of how spelling information is to be presented in the dictionary requires a number of decisions on the part of the lexicographer.

It is not an unimportant matter whether or not a certain word should be written with a capital initial letter. It is therefore inappropriate to show all the lemmas with a capital initial letter (which actually occurs in some dictionaries).