

## Contributions to the terminology of lexicography

Robert Ilson

In this paper I propose several terms for lexicographers to use. The purpose of these proposals is to enable lexicographers to name notions they deal with frequently, in the hope that, once named, the notions can be understood better and their referents dealt with better in real dictionaries. My paper is thus *onomasiological* (going from notion to name, as Roget's THESAURUS does) rather than *semasiological* (going from name to notion, as a typical dictionary does, whose macrostructure is ordered alphabetically by its headwords).

My paper is also the product of a process, which may itself be of interest for the light it sheds on terminological processes generally. The process has three overlapping stages: *identification* of a useful notion, *circumscription* of the notion by assembling its relevant features so that ultimately a dictionary-style definition may be composed for it, and *lexicalisation* of the notion thus circumscribed; that is, naming it.

To my proposals for new lexicographic terms I shall append a critical discussion of some terms that already exist.

### Lemma and sub-lemma

My first terminological proposal is for an extension of the term *lemma*, now often used to mean 'headword' or 'entry word' (see Robinson 1984: 181). I propose that *lemma* should be extended to mean 'everything preceding the first explanation (or sense number) in a dictionary entry' (leaving *headword* and *entry word* to retain their present meaning).<sup>1</sup> One reason for this extension is that lexicographers need to talk about what information should precede the explanations in their entries (i.e. their articles). For example, should etymologies go there (as in WEBSTER'S (Eighth) NEW COLLEGIATE DICTIONARY – W8) or towards the end of the entry (as in the LONGMAN DICTIONARY OF THE ENGLISH LANGUAGE – LDEL, based on W8)?

<sup>1</sup> Zgusta (1971: section 6.5.2) uses *lemma* much as I do, but does not introduce the notion *sub-lemma*. The discussion of *lemma* (roughly 'headword') in Wiegand (1983) is a salutary reminder that the linguistic and lexicographic status of the hyphens in *-o-* is not the same as that of the apostrophe in *o'*, despite the use lexicographers make of both hyphens and apostrophes in ordering homologues.

The argument for postposing the etymology is that it is consulted less frequently than the explanations. And that seems to have been borne out by the empirical research that led Greenbaum et al. to say: "The students showed little interest in consulting the dictionary for etymology . . ." (1984: 38). But this argument has not been pursued systematically. In this study, 92% of the subjects consulted etymologies occasionally, rarely, or never (1984: 38). But 82% of the subjects consulted pronunciations occasionally, rarely, or never (1984: 39). And no one seems to have enquired whether people ever consult irregular inflexions (such as *sang* and *oxen*). Yet LDEL has kept both pronunciations and irregular inflexions in the lemma (unlike CHAMBERS 20TH CENTURY DICTIONARY – CHAMBERS, which postposes irregular inflexions). Systemic thinking requires that if you take a decision about one part of a dictionary you should consider its implications for other parts of the dictionary.

More generally, systemic thinking means considering each part of the dictionary in relation to the dictionary as a whole. Thus W8, like other Merriam-Webster dictionaries, uses historical order of sense-groups within an entry, going from older to newer. It is systemically consistent with this practice to put the etymology in the lemma; that is, near the beginning of the entry: it is, after all, the oldest part of the entry. It would also be systemically consistent to postpose etymologies in a dictionary using reverse historical order. But LDEL has retained W8's historical order, and in that respect is systemically inconsistent in putting the oldest part of the entry after what is in principle the newest sense-group. On the other hand, LDEL has also departed from historical ordering by ingeniously putting labelled senses, which are restricted diasystematically (in the terminology of Hausmann 1977: Chapter 8), after all others. Perhaps if an ordering subsystem has been introduced into the overall structure of the entry, another disruption is less important.

Up to now I have contrasted a position "in the lemma" with a position "towards the end of the entry". But the notion "towards the end of the entry" is itself in need of clarification – and ultimately of circumscription and lexicalisation. Three positions at least can be identified for postposed etymologies: *before all* the sub-entries, including idioms and the undefined derivatives called run-ons (e.g. LDEL); *before some* of the sub-entries; in particular, before the run-ons (e.g. COLLINS ENGLISH DICTIONARY – CED, Reader's Digest GREAT ILLUSTRATED DICTIONARY – GID); *after all* the sub-entries, in a position typically called "the end of the entry" (e.g. CONCISE OXFORD DICTIONARY – COD, CHAMBERS). The factors associated with each of these positions are fascinating indeed, but lie beyond the scope of this paper.

As an important corollary of this use of *lemma*, I propose the term *sub-lemma*. Unfortunately, the term *sub-lemma* must be treated as polysemous in order to cover as much as I know of the diversity of dictionary practice. *Sub-lemma* 1 is thus 'the lemma of a sub-entry'; *sub-lemma* 2 is 'the space immediately adjacent to each explanation in a dictionary entry; that is, immediately before or

after it'; *sub-lemma* 3 (alas!) is 'the space immediately adjacent to a sub-lemma, if that space is not a lemma or an explanation'.

*Sub-lemma* 3 is required in order to remind us that sub-lemmas can themselves be adjacent. Thus in a dictionary entry text of the form 2 *a*: [explanation] *b*: [explanation] there are three preposed sub-lemmas: between 2 and *a*, between *a* and the following explanation, and between *b* and the following explanation, and the first two are adjacent. The domain of information in the first sub-lemma is 2*a* plus 2*b*; of information in the second, 2*a* only; and of information in the third, 2*b* only.

A problem that lexicographers can now discuss is whether the information permitted in a sub-lemma should be the same, and in the same order and type-face, as the information permitted in a lemma. All dictionaries seem to agree (counter-examples gratefully received!) that some types of information should be allowed in either place. This is particularly true of diasystematic information such as labels. Thus, if a whole entry is British English or American English or Slang or whatever, there will be an appropriate label in the lemma; but if a sense or an idiom is Slang its label will go in the relevant sub-lemma. Merriam-Webster seems to extend this principle to all types of information, including sense-bound spellings, pronunciations, inflexions, and etymologies. Thus W8 gives the following information about *goose*:

<sup>1</sup> *goose* . . . *n, pl. geese* . . . 3 *pl geese*: a tailor's smoothing iron with a goose-neck handle . . .

Thus the general inflexion goes in the lemma; the sense-bound inflexion, in the appropriate sub-lemma. LDEL, by contrast, collects all such information in the lemma:

<sup>1</sup> *goose* . . . *n, pl (1+2) geese* . . . , (3) *geeses* . . .

In certain other cases, Merriam puts an inflexion of restricted distribution in the lemma (in bold) and repeats it in one or more appropriate sub-lemmas (in italic). An example is the treatment of *shined* in W8's entry for <sup>1</sup>*shine*. The LDEL treatment of *shined* is essentially the same as its treatment of *gooses*.

Both W8 and LDEL homograph by part of speech. But in dictionaries that do not do so the choice of lemma or sub-lemma becomes even more important. Thus we have the contrast between the HAMLYN ENCYCLOPEDIA WORLD DICTIONARY (HAMLYN):

**cap** . . . *n., v., capped, capping. – *n.* 1. a covering for the head . . .  
– *v.t.* 16. to provide or cover with or as if with a cap . . .*

and the COLLINS ENGLISH DICTIONARY (CED):

**cap** . . . *n.* 1. a covering for the head . . . ~*vb.* **caps, cap+ping, capped**.  
(*tr.*) 16. to cover, as with a cap . . .

For the verbal inflexions, HAMLYN uses the lemma; CED, the verbal sub-lemma.

Two principles seem to be at odds here: the principle of *similarity* and the principle of *relevance*. The examples from LDEL and HAMLYN instantiate the principle that all information of the same type should be given in the same place – though both LDEL and HAMLYN depart from their principle of similarity in the case of diasystematic information. The examples from W8 and CED instantiate the principle that information should be given at the place where it is most relevant. Of course many – though not all – of these problems disappear if a different homographing policy is adopted: W8 and LDEL have separate main entries for *cap n* and *cap v*, and the LONGMAN DICTIONARY OF CONTEMPORARY ENGLISH (LDOCE) would regard sense-bound pronunciations, inflexions, or etymologies as grounds for homographing, too. But when such problems must be faced, and decisions taken, implemented, and explained to users, I hope their discussion will be clarified by the use of the terms *lemma* and *sub-lemma*.

## Homologues

My second proposal is for a generalisation (though not an abandonment) of the notion *homograph*, the generalized notion to be called *homologue*. The lexicographic meaning of *homologue* should be ‘a headword that has the same letters in the same order as another headword, regardless of capitalisation, diacritics, punctuation, or spaces’. *Homograph* therefore becomes a hyponym of *homologue*.

The ordering of homologues is an important problem for lexicographers. For example, if the phrasal verb *run down*, the adjective *run-down*, and the noun *rundown* are regarded as separate entries, how should they be ordered? Like W8, WEBSTER'S NINTH NEW COLLEGIATE DICTIONARY (W9) has *rundown*, *run-down*, *run down* (solid, hyphenated, open). W9 goes, as it were, “from more word-like to less word-like” (Paul Procter, personal communication). Once the notion *homologue* has been named, possible inconsistencies in the treatment of homologues can be identified. For example, W9 uses historical order for homologues that are homographs or for derivationally related groups of homographs (as it does for senses or semantically related groups of senses within an entry). Thus <sup>1</sup>*porter* (13th Century) precedes <sup>2</sup>*porter* (14th Century). But W9's ordering of homologues that are not homographs can produce the exact opposite, reverse historical ordering, as is shown by its own dating at *rundown* (1908), *run-down* (ca. 1892), and *run down* (1578).

An analogous problem is presented by homologues that differ in capitalisation. Here Merriam puts the lower-case form before the upper-case, perhaps on the grounds that lower case is the normal form for core lexical units. But often the lower-case homologue is derived from the upper-case homologue by generalisation, as when a proper noun is converted into a common noun. A related case is that of *creole adj* and *Creole n* in W9. The lower-case adjective precedes the ca-

pitalised noun, but the adjective entry is dated 1748; the noun, 1604. Once again the result is reverse historical order. A very common problem of this sort concerns the generic use of trademarks. I have not yet found any in W9, but LDEL, following Merriam's principles of homologue ordering, gives *hoover vb* before *Hoover trademark*: both, though unlabelled, are as dictionary entries British English rather than American English. It is fairly clear (though LDEL gives no dates) that the trademark preceded the verb historically.

By far the most important problem of homologue ordering in dictionaries, however, concerns word-partials, which are represented in dictionaries of English with hyphens and/or apostrophes. Thus at the beginning of the letter *o* both W8 and LDEL give:

**o, O, o-, -o-, <sup>1</sup>-o-, <sup>2</sup>-o, o'**,

As W9 does not date entries of this sort, it is not clear what the implications of this system are for historical ordering.

Dictionaries give very inadequate information about the way they order homologues, and I hope that the provision of a name for them will be of use in the instructions that dictionaries give their users: I have chosen the name *homologue* to reinforce the connexion between the examples I have been discussing and *homographs*.

### Definition linkages

My third proposal is to develop a system of terms to name the ways in which definitions are linked. In the W9 entry for <sup>1</sup>*frock* we find examples of a simple definition:

**3:** a woman's dress

two simple definitions co-ordinated asyndetically (or in apposition):

**1:** an outer garment worn by monks and friars: HABIT

and two simple definitions with the second subordinated to the first via a sense-divider:

**2 . . . b:** a workman's outer coat; *esp*: SMOCK FROCK

Furthermore, these two last types can be combined, as in W9's *frozen 2 . . . b*:

incapable of being changed, moved, or undone: FIXED; *specif*: debarred by official action from movement or from change in status

I propose a transfer to the naming of definition linkages of a well-known system of naming sentence types based on clause linkages, so that <sup>1</sup>*frock 3* has a *simple definition*, <sup>1</sup>*frock 1* a *compound definition*, <sup>1</sup>*frock 2b* a *complex definition*, and *frozen 2b* a *compound-complex definition*.

2 An important discussion of linked definitions (as "Définitions redoublées") is to be found in Quemada (1967:458-460).

This terminology makes it easier to diagnose and treat certain cases of definitional pathology. For example, a simple definition may be ill-formed through being over-determined. This problem is particularly likely to arise in the definitions of nouns that contain a lot of detail, or 'encyclopaedic information'. And it is particularly likely to affect dictionaries influenced by the efforts of Philip Gove (1961: 6a) to ban non-restrictive structures from definitions. Thus W9 defines *occipital lobe* as follows:

the posterior lobe of each cerebral hemisphere that bears the visual areas and has the form of a 3-sided pyramid

The problem here is that the *that*-clause is in fact functioning non-restrictively, as each cerebral hemisphere has but one "posterior lobe". There are at least two ways to make this definition well-formed. One is to change the restrictive relative clause into a non-restrictive relative clause (the definition, however, remains *simple*):

the posterior lobe of each cerebral hemisphere, which bears the visual areas and has the form of a 3-sided pyramid

Another way, arguably more explicit, is to change the simple definition into a compound definition:

the posterior lobe of each cerebral hemisphere: the cerebral lobe that bears the visual areas and has the form of a 3-sided pyramid

For further discussion, see Ilson 1981.

I should add that another major type of definition linkage is exemplified in W9's <sup>1</sup>*frock* 2:

2: an outer garment worn chiefly by men: a: a long loose mantle  
b: a workman's outer shirt; *esp*: SMOCK FROCK c: a long woolen jersey worn *esp.* by sailors

This chunk of text is analogous to a sentence of the form:

There are three things wrong with your plan: it's immoral; it's illegal; it's impractical.

A sentence of this type, in which clauses are linked by asyndetic hypotaxis, seems to have no name in traditional grammatical terminology (though it may constitute a type of apposition). But if it seems reasonable to consider it a kind of complex sentence, then the analogous definition can be considered a kind of complex definition, differing from the complex definitions already discussed in that each of its components – 2, 2a, 2b, 2c – constitute independent senses that can be cross-referred to from other parts of the dictionary – although there is no way of distinguishing in such a cross-reference between the first element of <sup>1</sup>*frock* 2 (the simple definition 'an outer garment worn chiefly by men') and the complex definition as a whole. Semantically, this first element functions like the shared meaning element common to a set of synonyms to be discriminated. Terminologically, the set of definitions 2 + 2a + 2b + 2c may be called a *separable*

*complex definition* and its component 2b may be called by contrast an *inseparable complex definition* when a more delicate name for it is required.

### Source language and target language

I should like to finish this discussion of lexicographic terminology with a couple of usage notes of special interest to bilingual lexicographers and language-teachers:

(1) The term *target language* is used differently by lexicographers and language-teachers. For example, in an English-to-Icelandic dictionary for Icelanders learning English, a lexicographer would say that Icelandic was the target language (the language into which the translations from English are made). A language-teacher would say that English was the target language (L2: the language to be learnt). In an Icelandic-to-English dictionary for Icelanders, English remains the target language for language-teachers, and becomes the target language for lexicographers. The term *source language* presents no such problems, as it is used only by lexicographers (and translators), not by language-teachers. For further information, see the relevant entries in the LONGMAN DICTIONARY OF APPLIED LINGUISTICS (LDAL).

(2) The English term *source language* has the French equivalents *langue source*, *langue de départ*, and *langue d'entrée*. The English term *target language* has the French equivalents *langue cible*, *langue d'arrivée*, and *langue de sortie* (the last two only in its lexicographic sense). I find these French equivalents hard to keep straight, because *langue de départ* and *langue de sortie* seem synonymous but are really contradictory opposites, while *langue de départ* and *langue d'entrée* seem opposite, but are really equivalent. And so on.

I don't know whether anyone else has trouble with these terms, but I believe that it is the job of terminologists not only to create new terms, but also to prevent the confusion of existing ones. And it would be a great help to have a multi-lingual glossary of lexicographic terminology.

**Acknowledgement.** I acknowledge with gratitude the assistance of the British Academy and of University College London in making possible my participation in this Congress of EURALEX, and the useful suggestions of Mary Snell-Hornby.

### References

#### *Cited dictionaries*

- CHAMBERS 20TH CENTURY DICTIONARY (CHAMBERS)  
E. Kirkpatrick et al. Edinburgh: W. & R. Chambers (1983).  
COLLINS ENGLISH DICTIONARY (CED)  
P. Hanks et al., London & Glasgow: Collins (1979).

- THE CONCISE OXFORD DICTIONARY OF CURRENT ENGLISH (COD)  
J.B. Sykes et al., Oxford: Clarendon (1982).
- HAMLIN ENCYCLOPEDIA WORLD DICTIONARY (HAMLIN).  
P. Hanks et al., London etc: Hamlyn (1971).
- LONGMAN DICTIONARY OF APPLIED LINGUISTICS (LDAL)  
J. Richards, J. Platt, H. Weber et al., Harlow: Longman (1985).
- LONGMAN DICTIONARY OF CONTEMPORARY ENGLISH (LDOCE)  
P. Procter et al., Harlow & London: Longman (1978).
- LONGMAN DICTIONARY OF THE ENGLISH LANGUAGE (LDEL)  
H. Gay, B. O'Kill, K. Seed, J. Whitcut et al., Harlow: Longman (1984).
- READER'S DIGEST GREAT ILLUSTRATED DICTIONARY (GID)  
R. Ilson et al., London: Reader's Digest (1984).
- WEBSTER'S (Eighth) NEW COLLEGIATE DICTIONARY (W8)  
H.B. Woolf et al., Springfield: G & C. Merriam (1973/74).
- WEBSTER'S NINTH NEW COLLEGIATE DICTIONARY (W9)  
F.C. Mish et al., Springfield: Merriam-Webster (1983).

### *Other literature*

- Gove, P.B. (1961), Preface to *Webster's Third New International Dictionary*, Springfield: G. & C. Merriam, 6a-7a.
- Greenbaum, S./Meyer, C.F./Taylor, J. (1984), "The Image of the Dictionary for American College Students", in: *Dictionaries* 6, 31-53.
- Hausmann, Franz Josef (1977), *Einführung in die Benutzung der neufranzösischen Wörterbücher*, Tübingen: Niemeyer.
- Ilson, R. (1981), "Definition Structure and Clause Structure", paper delivered at AILA Lund 1981; abstract in: Sigurd, B./Svartvik, J. (eds.), *AILA 81 Proceedings I*, Lund: University, 341.
- Procter, P. (personal communication).
- Quemada, B. (1967), *Les Dictionnaires du français moderne 1539-1863*, Paris: Didier.
- Robinson, J. (1984), "Glossary of English Lexicographical Terms, A-Z", in: Kipfer, B.A., *Workbook on Lexicography*, Exeter: University of Exeter, 173-188.
- Wiegand, Herbert Ernst (1983), "Was ist eigentlich ein Lemma?", in: *Germanistische Linguistik* 1-4/82, Hildesheim etc.: Olms, 401-474.
- Zgusta, L. (1971), *Manual of Lexicography*, Prague & The Hague: Mouton.