

Gabriele Stein

TOWARDS A THEORY OF LEXICOGRAPHY:

PRINCIPLES AND/V.S. PRACTICE IN MODERN ENGLISH DICTIONARIES

In spite of the undeniable progress to be noted in the making of language dictionaries (cf. Dubois and Dubois 1971) lexicography in general still suffers from some major deficiencies, of which I shall mention three:

(1) A language dictionary supplies information on a specific language. This information is based on a close study and thorough analysis of the structure of the language under consideration, language structure here taken in its widest sense. Linguists, sometimes with the help of statisticians and computers, have made available a number of descriptions of particular aspects of that language which differ according to the theoretical framework chosen. Yet whereas linguists are still far from being able to produce anything like a consistent description for one single language, the lexicographer is expected to do so. So far he has tried to acquit himself of this task as best as he can, but as I see it, in a rather randomly selective and often inconsistent way. The question which he has not asked persistently is whether there are any lexicographical interdependencies between these different theoretical descriptions or, more precisely, whether there are any overall lexicographical guiding principles that tell him which of them to select, and in which combination. Such guiding principles that could only get their legitimization from the function of the dictionary itself would give his work the necessary internal consistency.

(2) The dictionary representing a specific type of text, the lexicographer has to present his data in a specific way. From the early beginnings of English lexicography, dictionary makers have for instance used different styles of writing, printing, spacings, punctuations, etc. to single out specific types of information. In our days the presentation of lexicographical information has become rather conventionalized and one often wonders whether lexicographers are not just following suit without asking themselves what these conventions stand for. This is to be regretted, for it largely testifies to the fact that forms and methods of presentation have not been recognized as inherent constituents in lexicography. And the question that has been neglected but that is central to lexicography is whether there are any intrinsic interdependencies between the linguistic data given and the methods used to present them.

(3) Because of the deficiency outlined in (2) the question has not been given enough consideration whether there are hierarchies of dependencies, and if there are, which they are, whether they are obligatory and/or optional ones and which effects any changes of hierarchy would have on the overall structure of a dictionary.

If lexicography is to become a more theoretically based branch

of science these issues will have to be explored in depth. Matters are, however, complicated by the fact that these principles and methods of work are only partially disclosed in the prefatory matter of a dictionary. Introductions to dictionaries are usually very concise guides of how to use the dictionaries in question. Quite understandably they are not full-length accounts of all the underlying principles. Since few dictionary editors did, or maybe were permitted to, publish the theoretical and practical foundations of their work (Philip B. Gove was one of the few exceptions), it is one of the basic tasks of the student of lexicography to scrutinize existing dictionaries and lay these principles and methods of work open. Once they have been fully disclosed, it will be possible to identify those areas in which they could, or better still, have to be interrelated and in which internal systematicity is feasible and a scientific must.

In the present paper I would like to take a few steps towards such a more theoretically based framework of lexicography. Although most of the points that I shall make will be relevant for a general theory of lexicography, the aim of my paper is more modest: it is meant as a contribution to modern English lexicography. Since even the field of English lexicography is vast I shall confine myself to modern English monolingual alphabetical desk dictionaries. Within this field I shall draw attention to a number of very common principles and methods used, demonstrate that most of them constitute rather isolated policy decisions, and finally make some suggestions as to how they could become interrelated in a more systematic way.

The consultation of a dictionary represents a non-verbalized pedagogic discourse (Dubois and Dubois 1971:49ff.). The user of a language dictionary wants to obtain specific information on spelling, pronunciation, meaning, grammar, language variety, etc. and he expects the dictionary to supply him with this very information. The dictionary user is thus a learner and the dictionary the silent discourse partner or teacher. The dictionary is not unique in being such a 'silent teacher'. A handbook, for instance, would be another case in point. There are, however, notable differences between the classroom teaching situation and that in which the handbook or dictionary replaces the teacher. The very characteristic of classroom teaching is person-to-person interaction, oral communication. Teaching/learning by handbook or dictionary, on the other hand, has to rely on the written medium, the text. Other important differences are:

(a) The classroom

The classroom learner is not always motivated. Since the teaching stretches over a period of time the learning matter is built up in progressive sequences. Because of this progression earlier sequences may leave issues aside or unexplained, simplify matters, and be not fully exhaustive. In later sequences the earlier simplified presentation will be taken up and expanded. In the person-to-person interaction the teacher assumes a corrective function. Whenever something has been misunderstood or not understood at all, he will repeat it, present it in a different way and thus ensure that the message will get through.

(b) The handbook

The handbook reader/learner is always motivated. Teaching/learning again spreads over a certain period - the time length available is at least that needed for reading the book from beginning to end. Extensions in time and space allow for sequential teaching/reading. The handbook can proceed step by step, from simpler issues to more complicated ones. Since there is, however, no such institution as the teacher present, the form of presentation assumes a higher communicative function than in oral teaching. This higher communicative function is reflected in the development of specific forms of presentation to outline dependencies, to give prominence, etc., e.g. the use of a specific code of language, the use of different printing types, the sectioning into paragraphs and chapters, the use of diagrams and pictures. Questions and solutions at the end of individual chapters or the book may introduce some of the features of the corrective teacher.

(c) The dictionary

The dictionary user is a motivated learner. The characteristic difference between the oral learner and the handbook student on the one hand and the dictionary user on the other is that the latter is not satisfied with gradually accumulating packaged information; he wants full-scale information and he wants it on the spot. In addition, the information itself cannot rely on contextual features as in the person-to-person interaction which leaves substantial room for implicitness. A dictionary entry has to be self-contained and explicit. Because of the specific form of the dictionary space is limited. The information to be given is reduced to its relevant constituent minimum. At the same time presentation becomes highly significant and communicative. The text type of the dictionary entry has a grammar and vocabulary of its own which so far has not yet been studied with the depth it deserves (but cf. Robert Ilson's contribution in this volume). Both the text of a dictionary entry and its form of presentation complement each other, they constitute an indivisible lexicographical unit. Each taken separately would destroy the unity of information given. Presentation is thus functional. Such presentation elements as typeface, order, mention, non-mention or omission, are all constitutive, are all functional lexicographical elements. This functional character of presentation has to my mind not yet been recognized as fully as it should. Presentation will necessarily differ for pragmatic, sociolinguistic and strictly structural (or systemic) aspects of language. For the latter aspect a weak theoretical requirement would be that lexicographical presentation will not suggest structural properties which contradict language reality, and a strong one would be that lexicographical presentation should reflect language structure.

From the assumption that the language dictionary is a silent language teacher I would like to derive another requirement for lexicography: explicitness. Teaching is providing explicit information on the matter to be taught. The dictionary being limited with respect to space, one might prefer to talk of 'controlled explicitness'. The principle of explicitness is at the same time a lexicographical task and a lexicographical challenge: it calls for continuing investigations into the facts of language to find out more about them and to develop appropriate forms of presentation. I

shall illustrate the points made with some examples taken from modern English dictionaries.

It is true that modern English lexicography, following and absorbing linguistic research (cf. Gove 1962; Hartmann 1972), is becoming more and more explicit, detailed. Yet there are still many areas where improvement is desirable.

In the case of inflectional morphology, for instance, this could be achieved at no cost at all. Irregular inflectional forms such as children, feet, gave, etc. are usually listed as boldface headwords. Dictionaries vary, however, with respect to the lexicographical treatment given to these items. Non-explicit and, correlated with it, non-direct description is for instance a characteristic feature of the CONCISE OXFORD DICTIONARY OF CURRENT ENGLISH (COD), even in its latest edition (1982). The cross-reference item SEE refers children back to child, feet to foot, gave to give. No teacher of English, if asked about these forms, would give such an indirect answer as "children has something to do with child, feet with foot, and gave with give". He/ she would give a direct and explicit answer: "children is the plural of child, feet of foot, and gave the past tense of give". The same holds for COD's cross-references for spelling variants.

Some dictionary makers are already trying to make the application of restrictive labels more explicit. With respect to regional labels there is already something like a tradition of recording lexemic variants side by side for British and American English, e.g. autumn, AmE fall; baggage, BrE luggage, etc. For subject labels COLLINS ENGLISH DICTIONARY is so far the most explicit, by relating such items as jaundice and (Pathol.) icterus, that is, general and technical items. And CHAMBERS UNIVERSAL LEARNERS' DICTIONARY has introduced comparison into the field of usage labels. A verb like commence, for instance, is not simply called 'formal', it is put into relation with begin and start, and we then find explicit usage labels such as the ones in entries E₁ to E₃:

E₁: commence ... (more formal than begin and start)

E₂: altitude ... (more techn than height)

E₃: ample ... (more formal than enough)

For the ordinary lexicographer the recording of irregular inflectional forms and usage restrictions, to stay with the examples, would be unrelated issues and, left with the choice of selecting between the two outlined options, his choice might be random. For the 'explicit lexicographer' explicitness would be the guiding principle and it would determine his choice of form of presentation.

The functional and structural nature of presentation is illustrated by the following example. It is based on present-day English dictionaries and exemplifies what I have called one of the basic tasks of the student of lexicography: to attempt to reveal the lexicographical principles and methods of work which underly actual lexicographical practice but which are not fully disclosed in the prefatory matter of the dictionaries in question.

The type of dictionary under discussion usually distinguishes two types of entries: main entries and secondary entries. The most common secondary entry is the run-on entry where a second entry is literally run on to the first entry. Sub-entries where a second entry starts a new further indented line are less common, they are for instance used in CHAMBERS UNIVERSAL LEARNERS' DICTIONARY. Primary words, that is lexemes which are neither compounds nor derivatives, as well as affixes are generally given headword status. Idioms on the other hand are generally treated as run-ons or sub-entries. For compounds, derivatives and phrasal verbs practice varies. They are either listed as headwords or as run-ons or sub-entries. As run-ons they may either be defined or not defined at all. Lexicographical practice and presentation - if taken to be functional - thus suggests that there are two or even three kinds of lexical items; defined headwords, defined run-ons/sub-entries and undefined run-ons. If we disregard for the moment idioms because they are linguistically quite different from other linguistic elements, this form of presentation poses a number of problems for the practical lexicographer. Which language reality or language structure does this practice reflect? With a strict main entry policy there is no problem of observing alphabetical order, one of the basic lexicographical principles used in the type of dictionary under consideration. A run-on entry/sub-entry policy, however, calls for additional criteria and in doing so also for a hierarchy of principles.

The main and secondary entry distinction is based on the concept of word transparency. Lexicographical practice is thus trying to capture a certain structural aspect of language itself. If the meaning of a complex lexical item is derivable from its constituent elements, e.g. learner from to learn and -er, the item is said to be transparent or self-explanatory and thus a run-on entry candidate. If the meaning or all the senses of a complex lexical item cannot be derived from its constituents, e.g. the sense of reader as a 'university teacher above the rank of a lecturer', the item is regarded as lexicalized and as such qualifies for main entry treatment. A close study of one particular dictionary or a comparison of different desk dictionaries, however, reveals that the distinction between self-explanatory and lexicalized is far from easy, and that word transparency is a matter of degree (cf. Gove 1966; Stein 1976).

The treatment of derivatives is particularly unsatisfactory in this respect. The senses which a dictionary lists for a specific affix are usually those which are abstracted from actual formations with it. That is, because English has, for example, a number of -er derivatives from verbs with the meaning 'someone who does the action denoted by the verb habitually', the dictionary entry for the agential suffix -er will include such a sense after the first general sense 'someone who performs the action, activity denoted by the verb'. The explicit recording of the sense 'habitual' would make derivatives with it, e.g. drinker, smoker, regular formations and thus candidates for undefined run-on entries. The feature 'habitual' does, however, not occur with all deverbal -er agent nouns. It could even be regarded as a lexicalized feature of the affix. The listing of derivatives with affixes that have several senses, above all regular productive ones and lexicalized ones - and this is quite common for -dom, -er, -hood, -ity, -ness, etc. - as undefined run-ons thus perverts the principle itself.

The practice of recording self-explanatory suffixal derivatives as undefined run-on entries has some serious theoretical shortcomings (cf. Stein 1976), but it tries at least to capture a structural property of language. It is difficult to make out such a motivated link between language structure and form of presentation in the run-on entry practice found in the COD. It lists compounds, suffixal derivatives and phrasal verbs as run-on entries. Compounds and phrasal verbs are usually defined, derivatives are usually not. Lexicalized and non-lexicalized items are thus treated alike. If it is not lexicalization that accounts for the distinction between the two types of dictionary entry, which other principle is there? It cannot be the principle of recording word families because the entries in question omit all those combinations in which the headword functions as the head and not the modifier. It cannot be a coding principle either: the decoding user needs defined entries; undefined run-on entries, however, document that the formations in question exist in the language and are available for the encoding user. It cannot be a principle of assembling complex words with the same modifier: the headword of the whole entry rarely functions as a modifier in phrasal verbs; the headword buzz for instance is not a modifier in buzz off. The only alternative would then be a purely formal principle, e.g. that of assembling lexical items with the same initial string of phonemes or graphemes. Yet even this principle is not applicable: if meaning is irrelevant, lexicalized derivatives should have been treated a run-ons as well. The conclusion to be drawn is that the run-on entry practice in the COD has no foundation in actual language structure, and that it is nearly impossible to detect any guiding principle behind it.

Run-on entries obviously interrupt alphabetical order. In dictionaries with such entries the alphabetical principle thus rates second after word transparency. In most modern English desk dictionaries, however, there is an internal inconsistency in this respect. From the assumption that presentation in lexicography is functional and also structural in those areas where it reflects structural properties of language it follows that similar language structures should be presented in the same way. For self-explanatory prefixal derivatives, however, the alphabetical principle is dominant rather than word transparency. The dictionaries under consideration are heavily biased towards listing them as main entries with full definitions. Within the treatment of derivatives two conflicting coding principles are thus to be observed: the decoding one for prefixal formations and the encoding one for suffixal derivatives. Some dictionary makers have tried to remedy this contradictory situation by listing self-explanatory prefixal derivatives not ex-directory, but ex-dictionary, that is, separated off at the bottom of the page which has the prefix entry. I do not think that this solution is satisfactory. It introduces a fourth type of dictionary entry which is not justified by language structure. A better solution to this problem caused by alphabetization is found in the DOUBLEDAY DICTIONARY where the self-explanatory prefixal combinations to be included in the dictionary are listed undefined after the headwork entry of the prefix in question.

One of the dictionaries in which the presentation of undefined prefixal and suffixal derivatives is most consistent is the LONGMAN NEW UNIVERSAL DICTIONARY: undefined prefixal derivatives are listed as boldface headwords.

The conclusion to be drawn from this brief sample is that internal lexicographical consistency can only be achieved if lexicographers become more principle- and practice-conscious.

References

- Dubois, J. and Dubois, C. (1971) Introduction à la lexicographie: le dictionnaire (Collection Langue et Langage). Paris: Larousse
- Gove, P.B. (1962) "Linguistic advances in lexicography" in Dictionaries and THAT Dictionary: A Casebook on the Aims of Lexicographers and the Targets of Reviewers ed. by J. Sledd and W.R. Ebbitt. Chicago: Scott, Foresman & Co.
- Gove, P.B. (1966) "Self-explanatory words" American Speech 41: 182-198
- Hartmann, R.R.K. (1972) "Über den Einfluß der linguistischen Semantik auf die englisch-amerikanische Lexikographie" Linguistik und Didaktik 3: 197-208
- Ilson, R. (1984) "The communicative significance of some lexicographic conventions" in this volume
- Stein, G. (1976) "On some deficiencies in English dictionaries" Anglo-American Forum 5: 1-27