Barbara A. Kipfer

METHODS OF ORDERING SENSES WITHIN ENTRIES

Introduction

The arrangement of senses within the dictionary article is one of the most important decisions facing lexicographers, both in terms of meeting users' needs and expectations and in providing appropriate and accurate ordering for particular types of dictionaries. The basic rule for the lexicographer is to prepare the most useful and informative book possible, using common sense and keeping the reader in mind, above all. The purpose of this paper is to clarify the methodologies for ordering senses within entries and suggest new technological aids for preparing arrangements.

In general, there are three different ways in which definitions are ordered: (1) by usage or frequency (the most frequent use first), (2) by clustering the various definitions around several core or basic uses, such as the original use and major metaphorical uses -- as in logical ordering and psychologically-meaningful ordering, and (3) in chronological or historical order. There are arguments for and against each of these ways, and which of them the editor chooses depends on the ultimate purpose and use of the dictionary.

Usage ordering

The arrangement of the senses according to their usage -- the way in which words and phrases are actually used in a language community -- is chosen mainly for its practical utility for the majority of dictionary readers. Putting the most frequently-used senses first seems to be the approach chosen for most general dictionaries, although this can mean that any reader browsing through the entire group of definitions under the entry word can lose something by not seeing its historical development.

How is the central /core /standard /most relevant /most general /most current /most modern / most simple meaning determined? Certainly the lexicographer relies on citational evidence in his files and data in other current dictionaries, but in many cases he must make a judgment; an informed judgment, granted -- but is this accurate? There will always be words for which there is not enough citational evidence or not enough testimony from a variety of sources, and yet the lexicographer must make a decision.

However, it is the opinion of some that an arrangement based on the frequency of usage of the individual senses serves little useful purpose and might, in fact, be misleading. The argument usually offered in favor of usage ordering is that students often stop after the first definition. Even if that is true for a few

poor students, some linguists believe it is not usually the most common meaning that is being sought. There is also no completely reliable semantic count available for much of the lexis on which to base usage ordering.

What semantic counts are available, and do they show what is most common today? Lorge and Thorndike did their statistics in 1938, and no other semantic count as ambitious has been undertaken since. Clarence Barnhart reported in 1967 that probably no proper survey of the OXFORD ENGLISH DICTIONARY (OED) and the CENTURY DICTIONARY has ever been made. New semantic counts could be important evidence for arranging dictionary definitions, but they are so ambitious and time-consuming that it is doubtful any more counts will be hand-produced. Perhaps when natural language processing by computer is even more advanced, the computer may be programmed to automatically compile such counts.

In a quick survey, I found 31 instances of words whose ordering in four dictionaries varies from the usage information in the Brown Corpus of Present-Day American English (Kučera and Francis 1967). For instance, award occurs 60 times as a noun in the Brown Corpus and only 22 times as a verb form. However, neither the RANDOM HOUSE COLLEGE DICTIONARY, the CONCISE OXFORD DICTIONARY, the COLLINS ENGLISH DICTIONARY, nor the AMERICAN HERITAGE DICTIONARY -- the first three of which subscribe to usage ordering and the last to psychologically-meaningful ordering -- arrange their entries that way. All four list the verb first and then the noun. Nineteen words which have a higher frequency in the noun form were ordered oppositely in the four dictionaries; 12 other words which are more common as verbs were listed as nouns first in the four dictionaries.

These differences may be compounded or remedied when new editions are prepared or updated printings are run. The order of senses sometimes needs to be changed between printings; this is especially true at the end of entries in usage ordering dictionaries, as various senses have become more or less common over the years.

Logical ordering

Finngeir Hiorth (1954) studied the concept of logical order and found that though this concept was mentioned often in English and foreign language dictionaries, most lexicographers had only vague notions of what it meant. They arranged meanings according to their alleged logical order without being able to explain what they did in the process.

The concept of logical order studied by Hiorth is the assertion of a hypothesis about the historical development of meanings without any real investigations. It is based on explicit and implicit theories which lexicographers consider tenable -- which differs from lexicographer to lexicographer. The application of the concept of logical order also varies among lexicographers, as the application is based on the knowledge of the lexicographers themselves. This kind of treatment should not be accepted in these scientific times.

What does this mean to the dictionary user? It is one thing to supplement an historical arrangement with some logical ordering. It is quite another to exercise the singular judgment of a set of editors, without full investigation, and call it 'logical ordering'. The latter approach is of no real value to the user.

Dictionary editors who try to group meanings under core sub-meanings attempt to give the meanings of the entry word upon which several metaphorical uses may depend, but of course then lose both the feature of being able to put the most frequently-used meaning first or of showing detailed historical development. The reader should be given either historical or usage ordering, possibly coupled with a logical ordering for clarification only; anything other than that is simply an exercise of inordinate subjective judgment.

Psychologically-meaningful ordering

Psychologically-meaningful ordering is the sequencing of senses not necessarily with the implication that one of them could be derived historically, logically, or semantically from the other.

In the AMERICAN HERITAGE DICTIONARY we read that:

> Numerous English words have a spread of more than three or four distinct meanings or shades of meaning that must be identified and distinguished as separate semantic aspects and presented in a meaningful and useful order. The editors of this Dictionary have taken the position that the most useful order for the general user is neither historical nor by statistical frequency, even if sufficient evidence were available for either of those schemes. The order used here is an effort to arrange a complex word in a psychologically meaningful order, with one subgroup leading into another, so that the word can to some extent be perceived as a structured unit rather than a string of unrelated senses.
>
> In the 1982 edition of the AHD, this is revised as follows:
>
> When a word has more than one sense, those senses are arranged in such a way that a complex word can to some extent be perceived as a structured unit. Senses are not arranged historically or by frequency of use. Rather, they are ordered analytically, according to central meaning clusters from which related subsenses and additional separate senses may evolve. Such a meaningful order is considered to be a most useful presentation for the general reader.

Why do they consider it to be a most useful presentation? Certainly the method is somewhat logical -- but the user does not garner either semantic or historical knowledge about words and the ordering is not easy for the lexicographer to prepare without a great deal of subjective judgment. There is nothing scientific about it, unless you qualify what the lexicographer may know from his own studies.

Most dictionary users do not realize that more than one order even exists. A survey of users would probably reveal that most think definitions are ordered according to how they are currently used. Many are not aware that historical ordering exists, let alone the complicated style of psychologically-meaningful ordering. The psychologically-meaningful ordering is simply not of best service to the user and bears many of the same disadvantages as the sole use of logical ordering.

Historical ordering

Many dictionaries are prepared with an historical focus or character, the result of determinations of the intention of the work and its expected uses. A dictionary having this character will try to indicate the succession of historical developments of words.

Strict historical order has a definite and unambiguous meaning: meaning A precedes in historical order the meaning B only if the hypothesis is justifiable that A appeared earlier in time in the language than B. This hypothesis must be based on a collection of excerpts. If lexicographers disagree on the historical order of meanings, it is due to poor and/or insufficient data. Yet even in an historical dictionary, the historical order must be combined with the logical order: the main divisions of the semantic description are ordered historically, but senses closely related are grouped within them in a sort of semantic genealogical tree.

The compiler of the historical dictionary frequently cannot indicate the single senses of the words, and also not the single quotations, in their real historical sequence because the presentation would be rather chaotic. The editors must very often present their material in logical groups or by semantic connections. For example, the editors of the OXFORD ENGLISH DICTIONARY could not follow strict historical order for the word earth because, they said, "Men's notions of the shape and position of the earth have so greatly changed since Old Teutonic times" (A.W. Read 1974). They were obliged to compromise with a logical order.

A reasonable approach in preparing the historical dictionary, then, is to order the materials to bring out the dynamics of lexical development, with attention toward showing the succession and mutual compatibility of meanings.

Another disadvantage to chronological order is that as one goes through the various definitions there is often less

information at certain periods than at others. For example, some of the dates of the earliest use of the meanings would be very difficult to ascertain, merely because one would realize that certain definitions do not have many written citations, or much evidence of their use, during the earliest stages of their use.

In existing historical dictionaries, then, all too frequently the meaning listed first is not the oldest on record, nor the oldest by reconstruction, but the one most familiar nowadays or thought fittest to have been the semantic origin. Documentation may seldom be sufficient to enable the lexicographer to make more than an educated guess as to the ordering and, also, the history of many languages is still unknown.

Special dictionaries and ordering

The establishment of the single senses and their organization within entries are among the lexicographer's cardinal tasks. Various arrangements are possible, the determining factors being for whom the dictionary is intended and for what use or uses. Lexicographers preparing special dictionaries face this problem in catering to their audiences' needs.

A.S. Hornby, editor of the OXFORD ADVANCED LEARNER'S DICTIONARY, writes that:

> another problem is that of the order in which semantic varieties are to be placed. For those to whom English is a foreign language, those who are learning(not necessarily 'studying') Englishfor present-day needs, the order in which semantic varieties are entered should perhaps be based on frequency rather than on historical principles. Such users of a dictionary are more likely to meet, and to need to use, words in their current senses. These are the senses, therefore, which should be entered first. (Hornby 1965 : 104-110)

Computer assistance in ordering decisions

Computer-produced citations, concordances, and frequency counts may be used in preparing the ordering of senses within dictionaries. This information can be the basis of a more scientific approach to the arrangement of definitions which, up till now, has generally been either by historical development or by usage as determined by lexicographers through their own knowledge and the citational evidence from reading programs.

The lexicographer needs knowledge which may not be readily available from editorially-chosen citations. He needs to know how many different words there are in a representative selection of contemporary texts, what they are, and what the frequency of usage is. These tasks, though elementary, require an inordinate amount of time for a large corpus if done by hand. The computer can be much more than a machine for performing calculations; it can process, organize, and compare textual data for dictionary-making.

Each year it is necessary for dictionary publishers to scan large amounts of printed matter in their reading programs and lexicographers are just now asking for assistance from computer storage and retrieval systems for these tasks. A corpus compiled with computer assistance may be a more accurate and representative collection of data than that prepared by the selection of citations by human readers. The lexicographer's purpose is to present the current state of the language he is planning to describe. The computer can help promote accuracy and is much faster than the work of a traditional reading program. Computer assistance may enable the lexicographer to prepare and revise dictionaries more quickly and give users more valuable and up-to-date books than ever before.

Perhaps the computer-retrieved citations should be used to complement a selective approach; the former can fill important gaps in knowledge of the complete semantic range of lexical items. Based on a combined set of citations, the lexicographer will be better able to formulate judgments about the distribution and the frequency of different word senses -- and, therefore, their arrangement in a usage ordering scheme.

Concordances and frequency counts can also be reliable and valuable tools in arranging definitions according to usage. From a concordance, the lexicographer can sort through the various meanings and may get a better grasp of contemporary usage than can be obtained from citations gathered by traditional techniques. Such a listing could be invaluable in decisions on how many and which entries to include, which meanings to give, how to order definitions, and how to illustrate usage.

Usage ordering is chosen primarily because the dictionary maker wants to make access of definitions correlate as closely as possible to the user's experiences in speech and writing. The lexicographer should use this more scientific method for deciding the order of grammatical forms in a general monolingual dictionary. Frequency counts have served as the basis of numerous studies about the English language: the Brown Corpus, for example, was prepared as a grammatical analysis of one million words of text in which all of the words were tagged, or given specific grammatical designators. The tagging procedure, which was semiautomatic, assigned each word token an unambiguous symbol based on a taxonomy of eighty-seven grammatical categories (cf. Kučera 1969).

Several computer scientists/linguists have been involved in furthering semantic analysis by computer. Dictionaries have been transcribed into machine-readable form for processing their information toward developing a formal semantic description of English. This work is usually based a a study of semantic primitives to discover the hierarchies of meaning. For example, a set of words occurring in sense descriptions is extracted and a set of words occurring in those words' sense descriptions is extracted until the members of the smallest set are found. Sets of definitional primitives may help determine the choice of semantic components -- which in turn may be regarded as sufficient to describe the senses of all the words in a language.

A group of word senses similar in meaning may be contrasted, yielding 'semantic fields', which are needed for working out a formal semantic description of a language. No person or processor can be expected to accurately select word senses to make a comprehensive list of semantic fields -- but a combination of human and machine resources are being used for studies in this area.

Manual disambiguation can be done of all the senses in a machine-readable dictionary and from that a frequency of the use of each sense can be determined. This process can show what semantic concepts are used to define other words in a dictionary. Dictionary text definitions are a specialized form of ordinary text.

Edward Kelly and Philip Stone (1975) made significant progress on automatic disambiguation of word senses. Lexical disambiguation was defined by Robert A. Amsler (1980) as the process of determining which of a fixed set of meanings is assignable to a given lexical item in a given context. In addition to being able to tell which senses are assignable to a lexical item in context, it would be desirable for all senses to be ranked in terms of their likelihood of assignment. All of these disambiguation tasks can be greatly facilitated by the use of dictionary definitions. It is therefore reasonable to expect that a fully disambiguated dictionary could form the basis for a disambiguation procedure on any text. Frequencies of definitions by parts of speech and numbers of senses can provide new information for further semantic hypotheses about the language.

Even historical dictionaries may be improved by computer assistance. Lexical data bases may reveal antedated citations or new evidence which can be used in correcting the order of definitions. The richness of archives and data bases may also aid the historical lexicographer in employing logical ordering when necessary.

It is only a lack of faith in the computer's capabilities that keeps the lexicographer from using frequency counts, concordances, and other machine-assisted products in preparing dictionaries. There are few arguments which could now place traditional slip-collecting and human judgment alone as criteria for ordering; the accuracy of computer-assisted products for guidance on sense arrangement must be heeded as the way of the future in lexicography.

Conclusion

There is no single system which would be both powerful and detailed enough to be used unequivocally and alone as the basis for ordering senses and which would command general authority and recognition. This is a result of the considerable number of possibilities of each word's application -- which causes different meanings to present themselves with different strength and clarity to different speakers.

This variation is also demonstrated by the reciprocal fact

that different dictionaries with different presentations can render equally good services, provided they are materially and factually correct. Respect for the two outstanding traditions of usage ordering and historical ordering should be maintained.

One problem is that all words cannot best be presented by any single method, that is it would probably be best if some words were presented in chronological order, others were presented in decreasing order of frequency, and still others presented by grouping basic meanings together into sub-categories. Thus, even though each dictionary editor chooses one of these orders, each seems to find it best in certain instances to break with the chosen method.

In order to maintain the goal of serving the user best, it could be proposed that consistent ordering in pocket and general college/desk dictionaries may be welcomed by the user. The consistency might be a catalyst in encouraging dictionary use: the reader will have less discouragement in alternating among work, school, library, and home dictionaries. Chances are very slim for cooperation or agreement on this premise, but it is an interesting proposition.

However, lexicographers should be encouraged by the prospects, that, at least for usage ordering, computer-assisted processes may remove the burden of making human judgment and data from traditional reading programs the only sources for determining the ordering of senses within entries.

References

Amsler, R.A. (1980) The Structure of the MERRIAM-WEBSTER POCKET DICTIONARY. Ph.D. thesis (Report TR-164 Department of Computer Science and Linguistic Research Center) Austin: University of Texas

Barnhart, C.L. (1967) "Problems in editing commercial monolingual dictionaries" in Problems in Lexicography ed. by F.W. Householder and S. Saporta. Bloomington, Indiana: Indiana U.P.

Hiorth, F. (1954) "Arrangement of meanings in lexicography" Lingua 4: 413-424

Hornby, A.S. (1965) "Some problems in lexicography" English Language Teaching 19: 104-110

Kelly, E. and Stone, P. (1975) Computer Recognition of English Word Senses. Amsterdam: North-Holland

Kučera, H. and Francis, W.N. (1967) Computational Analysis of Present-Day American English. Providence, Rhode Island: Brown U.P.

Kučera, H. (1969) "Computers in language analysis and lexicography" in AMERICAN HERITAGE DICTIONARY ed. by W. Morris

Lorge, I. and Thorndike, E.L. (1938) A Semantic Count of English Words. New York: The Institute of Educational Research/Columbia University Teachers College

Read, A.W. (1974) "Dictionary" Encyclopaedia Britannica 5: 713-722