Fritz Neubauer

THE LANGUAGE OF EXPLANATION IN MONOLINGUAL DICTIONARIES

Introduction

As some of us know from experience, foreign users of monolingual dictionaries are often not very satisfied with the results of their inquiries because they are either given circular definitions or they are provided with an explanation they cannot understand, unless they look up the words given in the first explanation with the hope that they will do better in the second operation. Such experiences make them turn to bilingual dictionaries where they, at least, do not face the problem of not understanding the words provided (cf. Bensoussan et al. 1981). Teachers and tutors, on the other hand, seem to believe that monolingual dictionaries are better for them, and sometimes they are the only ones allowed in school or during exams.

Some of the reasons for this kind of negative attitude towards monolingual dictionaries by users consist of low-level circularity and of using words for the explanations of a headword which are more difficult than the words to be explained. If, for instance, we look at a definition in entry E_1 from the OXFORD ADVANCED LEARNER'S DICTIONARY OF CURRENT ENGLISH (ALD):

E_1: instrument ... implement, apparatus, used in performing an action, esp. for delicate or scientific work

it seems to be obvious that the words implement and apparatus are more difficult to understand for a learner of English and much less frequent than the actual headword (instrument) that was supposed to be explained. Or, in other words, somebody who does not know what an instrument is, is even less likely to know what an implement or an apparatus is. If this user is not yet discouraged and continues his search to find out about implement and apparatus, he will get back to instrument very soon, via entries E_2 and E_3.

E_2: implement tool or instrument for working with

E_3: apparatus set of instruments or other mechanical appliances put together for a purpose

Controlled defining vocabulary

An established strategy for avoiding these particular pitfalls in dictionary production consists of controlling the vocabulary used in the explanations by adhering to an explicitly compiled defining vocabulary as was originally implemented by M. West in his NEW METHOD ENGLISH DICTIONARY in 1935. Through the application of such a defining vocabulary, the word instrument is thus explained in the LONGMAN DICTIONARY OF CONTEMPORARY ENGLISH (LDOCE) as follows:

E_4: instrument ... an object of help in work

where each of the words appear in the defining vocabulary listed at the end of the volume.

Another strategy occasionally also followed in the LDOCE consists of allowing words not listed in the defining vocabulary to occur in the explanations as long as they can be broken down into words from the defining vocabulary. Such 'mixed definitions', as Petöfi (1977b) calls them in his linguistic theory, allow for the explanation of technical terms with a precision that cannot be achieved with the words from the defining vocabulary alone and avoids the problems of oversimplification that was criticized by Jain (1981). So if we look at example E_5 from LDOCE,

E_5: hydrocarbon any of several chemical compounds of HYDROGEN and CARBON, such as gas or petrol

hydrogen and carbon are naturally not listed in the defining vocabulary, but can be found explained elsewhere in the dictionary.

In this context, it should be pointed out that monolingual dictionaries with explicitly listed defining vocabularies for languages other than English do not seem to have been published so far (cf. Neubauer 1980); in some German monolingual dictionaries words occurring in the explanations are not explained at all in the dictionary.

In Neubauer (1980) I have also shown how by a small increase in the number of words listed in the defining vocabulary, a much larger number of headwords in the dictionaries can be explained: while it took West about 1500 words to explain 24,000 words, the LDOCE needs only 2280 to explain more than twice that number. The rise in the number of words in the defining vocabulary is less steep than in the number of headwords. While the introduction of defining vocabularies thus constituted a step forward, especially for learner's dictionaries, the figures mentioned are not really as impressive as they may look at first glance. If we look again at the words in the explanation of instrument in the LDOCE, the conscientious user of the dictionary perhaps not sure of the meanings of the words in the explanation would run into difficulties if he/she would try to look up what these words from the defining vocabulary are really supposed to mean. In Table 1 we see that the four words listed in the defining vocabulary of the LDOCE are shown to have about 55 senses in the definitions of the same dictionary, some of which are not used for defining. Still, about 20 more basic ones can be found in West's General Service List of English Words (1936). The only differentiation undertaken in the LDOCE is to indicate word classes but only as far as fly, n,v or arm, n,v is concerned, the insect and the act of flying are still represented by the same word form. As Petöfi (1977a) has shown, West's 1480 word forms of the defining vocabulary correspond to 4607 senses if one only counts the senses indicated by West himself in his dictionary.

Table 1

instrument	an object	used	to help	to work
	↓	↓	↓	↓
number of senses indicated in definitions of LDOCE	object,n 6	use,n,v 15	help, n,v 7	work,n,v 30
number of senses indicated in West's General Service List	2	5	5	9

The problem of not distinguishing between words with different senses and word forms is also common to frequency counts where word forms are usually counted without recourse to the meanings involved. On the other hand, data from frequency counts have been successfully used to compile lists of words for basic vocabularies, for instance, in synonym deletion when one has to decide which of two synonyms should be retained in a corpus for teaching purposes, a look at frequency lists may be useful (cf. Gougenheim 1964).

Here it can easily be seen that there is a connection between compiling defining vocabularies for dictionaries and the compilations of corpora for language teaching purposes, text books or easy readers, since the purpose of explaining unknown words is common to the learner's dictionary and the teacher of a foreign language who incidentally usually does not have the time to think as long as a lexicographer before giving explanations in the foreign language. Foreign language teachers and their students would be helped in their task if they could be equipped with a working 'defining vocabulary' in the extended sense of not only serving for dictionaries, but a vocabulary in terms of which words could be explained better than they have been so far.

A further connection between defining vocabularies and the classroom language of the foreign-language teacher is easily observable if one compares corpora for language teaching purposes and defining vocabularies as shown in Tables 2 and 3.

There B stands for the basic vocabulary contained in the GRUND- UND AUFBAUWORTSCHATZ of the particular field, compiled by Weis; T indicates whether the word in question is included in the vocabulary of Threshold Level English (van Ek 1975); L indicates that the word in question is listed in the defining vocabulary of the LDOCE.

Table 2 Comparison of the vocabulary for the description of the human body from different sources

word	GRUND- UND AUFBAU-WORTSCHATZ	THRESHOLD LEVEL ENGLISH	LDOCE	LLOCE	frequency in Kučera & Francis (1967)	frequency per million in Thorndike & Lorge (1944)
arm	B	T	L	D		
back	B	T	L	D		
beard	B	-	L!	-	31	32
to bite	B	-	L	-		
blood	B	-	L	D	121	< 1oo
breath	B	-	L	-	53	< 5o
to breathe	B	-	L	-	16	< 5o
eye	B	-	L	D	122	< 1oo
finger	B	-	L!	-	4o	< 1oo
fist	B	-	-	-	26	23
foot	B	T	L	-	7o	< 1oo
hair	B	T	L	D	148	< 1oo
head	-	T	L	D	424	< 1oo
heel	B	-	L!	-	32	< 5o
leg	-	T	L	D	125	< 1oo
lung(s)	-	-	L	D	36	15
neck	-	T	L	D	83	< 1oo
organ	-!	-!	L	D	12	48
skin	-	-	L	D	47	< 1oo
sweat	B	-	-	-	23	19
tongue	B	-	L	-	35	< 5o

Improving defining vocabularies

Before we continue with the analysis of the information in Tables 2 and 3, I would like to point out that the purpose of this comparison constitutes a further step: after welcoming the existence of defining vocabularies for various applications and regretting some of their insufficiencies I would now like to move on to discuss in what ways we could improve them. As I have already mentioned, the lack of sense differentiation could be remedied by coding the words in the defining vocabulary with the sense indicated in the dictionary explanations with superscripts. Another method not really attempted yet on a large scale for this purpose (cf. Neubauer 1977) is to use the language of dictionary definitions not only as an object for improvement, but as a data-base to find out which words with what senses are really needed for word explanations, be it in a learner's dictionary or

in explanations in the classroom. This strategy would be supported by evidence from semantic theory and artificial intelligence where the search for semantic primitives has equally come to the conclusion that they are not things that are waiting to be discovered, but that they are the semantic units shown to be required for meaning construction in natural language as well as artificial language.

One difference between linguistic theory and lexicography is however, that in linguistic theories I can make up words and assign them meanings if I want to have unambiguously defined meaning-bearing units, while in natural-language dictionaries we have to take the words that exist in a particular language.

As to the strategy of using dictionary definitions for the purpose of finding out what words are really required for explanations, one such small-scale attempt is shown in Tables 2 and 3. D there indicates that the word in question occurs with a relatively high frequency in the explanations of the LONGMAN LEXICON OF CONTEMPORARY ENGLISH (LLOCE) relevant for the field.

Let us now look at these tables in detail. The words shown there are only meant as illustrations for the strategy of vocabulary selection chosen for that particular purpose, they do not represent a complete list of words for that particular field. Table 2 shows that arm and back (naturally with various senses) occur in all the four cases analyzed here. The same applies for hair the inclusion of which in the lists is also supported by frequency considerations (see the last two columns). Head, lungs, neck and skin do not occur in the basic vocabulary of this particular field. I would have thought that they were pretty basic parts of the human body, an opinion that is supported by the frequency data about the word forms in question. On the other hand, beard, finger and heel form part of the defining vocabulary in the LDOCE, although they do not occur frequently in the dictionary explications and I do not think that they are difficult to explain or needed for definition. Similarly, in Table 3 you can see that bicycle and taxi are included in LDOCE's defining vocabulary. For me personally the bicycle is rather basic because I cycle on it to work every day, but this still does not convince me that it should necessarily be part of the defining vocabulary of a dictionary, and my opinion is also supported by dictionary and frequency evidence.

A typical difference between basic vocabularies based on frequency counts and vocabularies for defining purposes is illustrated by the words organ in Table 2 and vehicle in Table 3. In both cases their relatively low frequency stopped them from making the basic vocabulary or Threshold English, but they show the highest frequency in the data from the dictionary explanations. They are therefore really needed and I would have thought them very useful words to have for a teacher if he has to explain words like stomach, kidney or fire engine in class. Here frequency data, even if they were more reliable, would have to be overridden.

This brief comparison is only meant to show on the basis of what information one could decide to modify the language used

Table 3 Comparison of the vocabulary for the description of transport from different sources

word	GRUND- UND AUFBAU-WORTSCHATZ	THRESHOLD LEVEL ENGLISH	LDOCE	LLOCE	frequency in Kučera & Francis (1967)	frequency per million in Thorndike & Lorge (1944)
(air)plane	-	T	L	D	135	< 50
to arrive	B	T	L	-	108	< 100
bicycle	B	T	L!	-	3	11
boat	B	T	L	-	72	< 100
bus	B	T	L	D	34	9
car	B	T	L	-	393	< 100
customs	B	T	-	-	18	-
departure	-	T	L	-	17	23
engine	B	T	L	D	52	< 50
flight	B	T	L	-	60	< 50
to fly	B	T	L	D	33	< 100
heaven	B	-	L	-	43	< 100
passenger	B	T	L	D	35	50
petrol(eum)	B	T	L	-	1!	8
to sail	B	-	L	-	12	< 100
speed	B	T	L	-	83	< 50
taxi(cab)	-	T	L!	-	20!	17
train	B	T	L	D	82	< 100
vehicle	-!	-!	L	D!	88	13
wing(ed)	B	-	L	-	18	< 100

in dictionary explanations and, consequently, to revise the defining vocabularies. Also, in the age of machine-readable dictionaries our computational colleagues should be able to assist us in this kind of work and they should do a quicker and better job than I could do with paper and pencil (cf. the paper by Michiels and Noël at this conference). But even a computer cannot help with word sense differentiation if the input has not been coded for it in advance. Another interesting point is whether defining vocabularies thus compiled in different languages could be harmonized in the long run.

It is always easier to criticize dictionaries than to do the work of compiling them, but this should not stop us, on the one hand, from using ideas from theoretical linguistics to improve

our dictionaries (or teaching materials) and, on the other hand, from using the language of existing dictionaries to provide valuable data for further lexicographical research and applied linguistics.

References

Bensoussan, M. et al. (1981) "The effect of dictionary usage on EFL text performance compared with student and teacher attitudes and expectations" in AILA 81, Proceedings I ed. by B. Sigurd and J. Svartvik. Lund University

Gougenheim, G. et al. (1964) L'élaboration du français fondamental (1er degré). Paris: Didier

Jain, M.P. (1981) "On meaning in the foreign learner's dictionary" Applied Linguistics 2: 274-286

Kučera, H. and Francis, W.N. (1976) Computational Analysis of Present-Day American English. Providence, Rhode Island: Brown U.P.

Neubauer, F. (1977) "Einige Aspekte umgangssprachlicher Lexika" in Das Lexikon in der Grammatik - die Grammatik im Lexikon ed. by J.S. Petöfi and J. Bredemeier. Hamburg: Buske

Neubauer, F. (1980) Die Struktur der Explikationen in deutschen einsprachigen Wörterbüchern (Papers in Text Linguistics 27). Hamburg: Buske

Petöfi, J.S. (1977a) "The defining vocabulary of Michael West" Mimeo

Petöfi, J.S. (1977b) "Lexikoneintragungen in der kanonischen Repräsentation von Sätzen und Texten" in Das Lexikon in der Grammatik - die Grammatik im Lexikon ed. by J.S. Petöfi and J. Bredemeier. Hamburg: Buske

Thorndike, E.L. and Lorge, I. (1944/72) The Teacher's Word Book of 30,000 Words. New York: Teachers College Press

Van Ek, J. (1975) Threshold Level English. Oxford: Pergamon

West, M. (1936/53) A General Service List of English Words. London: Longman