

Philippe HUMBLE, Universidade Federal de Santa Catarina (Brasil)

The use of authentic, made-up and 'controlled' examples in foreign language dictionaries.

Abstract

The aim of this paper is to suggest a criterion for the choice of examples in foreign language dictionaries, monolingual or bilingual, from the viewpoint of *encoding*. It is claimed that made-up examples are better suited for beginning learners, authentic examples for advanced ones, and a third kind –found in 'controlled' corpora– for intermediate learners. Frequency, not only of words, but mainly of senses of words is suggested to be the main criterion for determining what kind of examples should be used in each entry.

Keywords: foreign language learning, examples

1. Introduction

Until recently, and because dictionaries were used much more as a *decoding* than as an *encoding* tool, the issue of examples seemed relatively simple. It became a topic of discussion only from 1987 on, when Cobuild started putting authentic language verbatim in its dictionary. Before that, it was generally assumed that examples had to be made up by the people compiling the dictionary. The main reason for this may have been the fact that there was no easy way of retrieving instances from large corpora, but it was also taken for granted that lexicographers were capable of illustrating the uses and senses of a word in a much clearer way than random language users. Apart from this, real language was deemed to be too difficult for learners to cope with. The Cobuild team, on the other hand, was convinced that the intuitions of lexicographers were untrustworthy and that made-up examples gave learners a false idea of the language (Fox, 1987, *passim*).

Although traditional dictionaries reacted against this viewpoint, most learner's dictionaries, as well as the major bilingual dictionaries, subsequently incorporated corpora as a standard lexicographic tool. They would not put, however, unmodified utterances into their entries. As a result, most dictionaries went on using made-up examples, whereas only Cobuild would use strictly authentic ones. Maybe because of the way the diverging standpoints were defended, the bone of contention came to be considered almost as an ideological problem and not as a practical one for which the solution may lie somewhere in the middle: each kind of example was considered intrinsically better or worse.

In my opinion, and for *encoding* purposes, the choice between made-up and authentic examples depends on the learner's level of knowledge, and the kind of example depends on each individual entry.

2. Learners' levels

Learners are traditionally subdivided into three groups: beginners, intermediate and advanced. Since little attempt has been made to define what characterises each of these groups, I drew up my own subdivision based on a few Hallidayan and Sinclairian concepts. In the following table I give a number of features of each group of learners.

	<i>beginners</i>	<i>intermediate</i>	<i>advanced</i>
<i>vocabulary</i>	- command of a small number of frequent <i>content</i> words - problems with the use of <i>function</i> words	- command of a great number of frequent and a few randomly acquired infrequent <i>content</i> words - few problems with <i>function</i> words (except prepositions)	- command of a great number of frequent and infrequent <i>content</i> words - no problems with <i>function</i> words
<i>syntax</i>	no subordination	subordination	subordination
<i>lexico-grammar</i>	- congruent mode of expression - no style variations	- congruent mode of expression - incipient style variations	- congruent and metaphoric mode of expression - style variations

Figure 1 Characterisation of foreign language learners according to their level.

In order to assist learners in producing a foreign language, dictionaries should take these characteristics into account. The frequently quoted input-theory (Krashen, 1982), and the experience of language teachers, tells us that learners profit most from an input which is slightly beyond their capacities, although still comprehensible and reproducible after analysis. According to this reasoning, it is pedagogically useless to feed learners instances of language which they are either unable to understand, or do not tell them anything new. This implies that each type of learner should be given a specific kind of information.

As a matter of course, publishing companies distinguish between groups of learners and make dictionaries for each of them separately. There is supposedly a way of defining and of giving examples appropriate to each group, but in practice the difference between each kind of dictionary lies much more in the number of entries, even if a current beginner's dictionary still has over 30,000 headwords.

In the cases in which lexicographers take the level of their audience into account as far as style and examples are concerned, frequent and infrequent words are dealt with in much the same manner. In dictionaries such as the *Oxford Wordpower*, aimed at beginning and intermediate learners, frequent and infrequent words are treated in fundamentally the same way, as the entry for *abscond* illustrates.¹

abscond

to run away from a place where you should stay, sometimes with something that you should not take

She absconded with all the company's money.

Figure 2 'Abscond' in *Oxford Wordpower*.

Here we have the case of an infrequent word, dealt with in an easy way, and not basically different from the way a frequent word like *sleep* is dealt with:

sleep to be in a state of sleep for a period of time
 Did you sleep well last night?
 I only slept for a couple of hours last night.

Figure 3 'Sleep' in Oxford Wordpower.

The made-up examples for *sleep* may not be perfect, but they show only congruent uses and no subordination, all of which makes them easier for beginning learners. In the *abscond* entry, the example was presumably made up and contains words that are much more frequent than *abscond* itself. This makes one wonder what its purpose is. The definition is sufficiently clear to help *decoding* learners. I assume therefore that it is intended to help with *encoding*. However, *encoding* learners using a word like *abscond* should be quite advanced and are looking for more complex information. Moreover, the simplicity of the example can induce them to think that *abscond* can be used in a straightforward, albeit formal, context. In this case, requirements valid for the general target audience prevailed over what the target word demands. Beginning and intermediate learners are given an advanced word in a beginner to intermediate style.

An example of the opposite are the Cobuild examples. These are often too difficult for entries or sections of entries that are likely to be looked up only by beginners. *Sleep*², in the sense of 'when you sleep you rest with your eyes closed and your mind and body inactive', is such a case.

During the car journey, the baby slept...
 I've not been able to sleep for the last few nights...
 a pool surrounded by sleeping sunbathers.

Figure 4 'Sleep' in Cobuild2.

What kind of learner is targeted in this entry? No advanced learner will look up *sleep* to know what the word means. If they look it up for *encoding* purposes, they will probably be after some idiomatic use, not one abiding by the open-choice principle. But the examples show only one of these idiomatic uses (*able to sleep*), while beginners will, on the other hand, wonder what *journey* means, *last few*, *pool*, *surrounded*, and *sunbathers*. Since one can *sleep* in any language, the only reason why this entry will be looked up is *productive* use. Beginning and intermediate students, however, will not find these examples in accordance with what they themselves would be able to produce with a little help from a dictionary. In other words, a frequent word was treated as if the target audience were advanced, and the policy followed was the policy of the dictionary in general, not of the word in particular. On the other hand, in the case of an infrequent word such as *abscond*, the superiority of authentic examples shows immediately:

He was remanded in care and ordered to appear the following day, but absconded...
 A dozen inmates have escaped or absconded from Forest jail in the past year.
 Unfortunately his partners were crooks and absconded with the funds.

Figure 5 Examples for 'abscond' in Cobuild II.

Authentic examples convey much more information than made-up ones. The first Cobuild example shows that one can *abscond* without mentioning whereto nor where from. *Remanded in care* and *ordered to appear* characterise the word as formal, probably juridical. Even advanced learners will presumably not be acquainted with the verb *remand* and this is a way of indicating the register. The first example shows that *abscond* does not necessarily need a complement; the two last examples show two typical prepositions; “have escaped or absconded” indicates that there is a difference between *escape* and *abscond*; and all of the examples show a great number of typical collocates (*inmates, jail* and *crooks*). This makes authentic examples fitted for advanced learners using infrequent words and who have the linguistic means to use this information.

3. Made-up or authentic?

Learners should be given examples resembling the sentences they aim to produce. Words can combine to make sentences according to the *open-choice principle* or to the *idiom principle*. Frequent words have a tendency to form idioms often, but can be used on their own as well, in an *open-choice* setting, even if this modality is not always the vernacular one and examples of these congruent uses are hard to find in a corpus. Made-up examples, however, if congruent, use the words in their core meaning, and are especially useful for beginning and intermediate learners. If they are hard to find in a corpus, the possibility of making them up should be considered and lexicographers with experience in foreign language teaching and learning should have no difficulty in doing this. Frequent idioms involving frequent words could be treated similarly, as single lexical items.

In the case of infrequent words or senses of words, simple examples are not sufficient and a lot more information is needed. Can one *abscond with books, with an umbrella, from home, from a supermarket, to the city, abroad*? It will be difficult for a lexicographer to think of all the possible cases and it will even be difficult to recognise them consciously if one comes across them. This is where authentic examples, in as large numbers as possible, do a much better job. Especially since they inevitably incorporate information which the lexicographer had not even thought of putting in.

4. ‘Controlled’ examples

Sleep and *abscond*, which I discussed above, are each at opposite extremes of the frequency continuum. A word like *mimic*, on the other hand, can be considered of intermediate frequency.³ Cobuild gives the following examples:

1. He could mimic anybody, and he often reduced Isabel to helpless laughter...
He mimicked her upper-class accent.
2. Don't try to mimic anybody. You have to be yourself if you are going to do your best...
The computer doesn't mimic human thought; it reaches the same ends by different means.

Figure 6 ‘Mimic’ in Cobuild II.

In the case of a mid-frequent word, these authentic examples can be considered adapted to an intermediate dictionary user. No word in these examples is blatantly less frequent than the

target word; the sentences are long enough for the learner to infer what the context is; it is clear after these examples that one can *mimic something* as well as *someone*; and that no preposition is needed, etc. Still, there are problems with this kind of example. First, there are too few of them. Second, it is difficult to determine what the context of the utterances is and one is consequently left in doubt as to their register. In this case, another kind of example, found in corpora such as encyclopaedias on CD-ROM, could prove to be extremely helpful. These examples are easily retrievable and generally numerous, intermediate learners can be sure that the register will be neutral enough to be appropriate in most situations and, additionally, in an encyclopædia learners can make the context just as long as they need it to be. The following are just a few examples for *mimic* as they feature in the *Grolier Encyclopaedia* on CD-ROM (*bold is red in the original*).

endocrine system

Its symptoms **mimic** those of primary aldosteronism, and it is related to high blood pressure and disorders characterized by edema, such as cardiac failure.

dolphins

Bottle-nose dolphins have become well known performers in many aquariums; they are capable of spectacular tricks and may **mimic** the sounds of a few human words.

diabetes

In order to **mimic** the closely linked changes in blood glucose and insulin concentrations that occur in healthy nondiabetics, several new techniques for insulin delivery have been developed.

crow

It can be tamed and can **mimic** human speech; research indicates that it is able to solve uncomplicated puzzles.

Figure 7 Occurrences of 'mimic' in the Grolier Encyclopaedia (4 out of 59)

These examples have the advantage of being clear, and written in a syntax which allows vocabulary problems to be solved easily. There are none of the incongruent uses which characterise everyday native adult language use and the occurrence of several infrequent words (*aldosteronism, edema, cardiac failure, glucose, insulin delivery*) will not deter a learner, since they are imbedded in a context which indicates that they belong to a scientific register and knowing exactly what they mean has no bearing on the use of *mimic*. Moreover, these hard words are all monosemous and have a meaning which does not depend on syntax or context. As for the style of these examples, encyclopaedias allow little or no variation. What may look like a disadvantage for advanced learners, is in fact an advantage for intermediate learners. Style variations, which include finesses such as situational, regional and context variations are the production aim of the advanced, not of the intermediate learner. The latter would benefit particularly from these encyclopaedia examples.

5. Conclusion

What I propose is a dictionary in which the examples are adapted to the audience. Since this audience is, among other things, characterised by the kind of words they use, the choice of examples can be made dependent on the word in question. Beginning learners will need examples of frequent words in their frequent senses. Advanced learners will need examples of infrequent words, and infrequent senses of frequent words. Made-up examples can therefore illustrate the former, and authentic ones the latter. Additionally, examples retrieved from corpora such as encyclopaedias can be put to good use, especially in the case of words of intermediate frequency.

How can the limit between frequent and infrequent be established? I did a few experiments myself based on the Cobuild frequency list and I expect that the 600 first ones on the list can be considered extremely frequent. If one takes them out of a text, one is left with only *content* words and these are of the kind that still gives the reader a good idea of the topic. Of these 600 initial frequent words, the *function* words should be singled out and given special treatment. The *Collins Shubun* is a magnificent example of how this can be done and how useful this is.⁴

A further 1400 words on the list can be considered relatively frequent and the traditional barrier of 2000 words can be taken as the border between beginning and intermediate learners. All of these 2000 words, with their most frequent senses, should be illustrated at least also by means of made-up examples, even if authentic examples should not be excluded. For the intermediate part of the lexis, learners could be given a choice between 'controlled' and authentic examples, although I personally prefer 'controlled' examples myself. For infrequent words, used by advanced learners, authentic examples offer the advantages I mentioned.

Maybe we do not need a complete dictionary for every kind of learner. What we need is a dictionary that caters to the needs of each level by giving the right example for each sense of a word.

6. Notes

- ¹ Frequency according to the Cobuild corpus: absconded (158), absconding (64), abscond (51) in approximately 250 million words.
- ² Frequency according to the Cobuild corpus: sleep (19146), sleeping (8190), sleeps (924), slept (4188).
- ³ Frequency according to the Cobuild corpus: mimic (741); mimicking (300), mimics (262), mimicked (262).
- ⁴ In the *Collins Shubun English-Japanese Dictionary* 300 'keywords' are given extended treatment.

7. References

- Chi, A. (1995). A critical user-oriented review of the *Collins COBUILD English Language Dictionary*: definitions and examples in Wong P. Y. Kitty and Green F. Christopher

- (1995). *Thinking language. Issues in the study of language and language curriculum renewal*, Language Centre, The Hong Kong University of Science and Technology, pp. 45-56.
- Chi, A. and Yeung, S. (forthcoming). A case study of the use of dictionaries to assist the teaching and learning of English in Hong Kong.
- Collins COBUILD English Language Dictionary*. Second edition. (1995). HarperCollins. London.
- Collins Shubun English-Japanese Dictionary*. (1993). HarperCollins/Shubun, London.
- Drysdale, P. D. (1987). The role of examples in a learner's dictionary in *The Dictionary and the Language Learner, Papers from the EURALEX Seminar at the University of Leeds*, 1-3 April 1985, Niemeyer, Tübingen pp. 213-223.
- Fox G. (1987). The case for examples, in Sinclair J. (ed.) *Looking Up*, London and Glasgow: Collins ELT.
- Halliday, M. A. K. (1994). (second edition) *An introduction to functional grammar*, Edward Arnold, London.
- Krashen, S. (1982). *Principles and practice in second language acquisition*, Pergamon.
- Laufer, B. (1992). Corpus-based versus lexicographer examples in comprehension and production of new words in *EURALEX '92 Proceedings I-II Papers submitted to the 5th EURALEX International Congress on Lexicography in Tampere, Finland, Part I*, *Studia Translatologica ser. A vol. 2*, Tampere pp. 71-76.
- Laufer, B. (1993). The effect of dictionary definitions and examples on the use and comprehension of new L2 words, *Cahiers de Lexicographie*, No 63, 131-142.
- Nesi, H. (1996). The role of illustrative examples in productive dictionary use *Dictionaries: The Journal of the Dictionary Society of North America* No 17 198-206
- Oxford Wordpower* (1993). Oxford University Press. Oxford.
- Sinclair, J. (1991). *Corpus, Concordance, Collocation*, Oxford University Press, Oxford.
- Summers D. (1988). The role of dictionaries in language learning. in: Carter R. and McCarthy M. *Vocabulary and Language Teaching*, Longman, London and New York.