

Corpus-Based versus Lexicographer Examples in Comprehension and Production of New Words

ABSTRACT: The paper examines the differences in effectiveness of two types of examples in comprehension and production of new words: a. examples taken from authentic language corpora and b. examples written by lexicographers for the particular entry. Twenty low-frequency words were given to a group of EFL learners. Ten were illustrated by authentic examples, ten by lexicographers' examples. The subjects were tested both on comprehension of the target words and on their ability to use these words in sentences of their own. The results of the experiment suggest that, for users of learner dictionaries, naturalness of information is not necessarily more effective than its clarity.

1. Introduction

1.1. Word knowledge

In most linguistic analyses a word is described as a set of properties, or features (Lado 1972, Chomsky 1985, Gibson and Levin 1975, Richards 1976). Thus knowing a word would ideally imply familiarity with all its properties as is often the case of an educated native speaker. When a person 'knows' a word, he/she knows the following: the word's pronunciation, its spelling, its morphological components, if any, the words that are morphologically related to it, the word's syntactic behaviour in a sentence, the full range of the word's meaning, the appropriate situations for using the word, its collocational restrictions, its distribution and the relation between the word and other words within a lexical set. (For a discussion of word knowledge see, for example, Nation 1990 and Laufer 1991).

Unlike an educated native speaker of a language, the foreign language learner knows a much smaller number of words. In some European countries and Israel lexical knowledge is about 2000-3000 word families, in some Asian countries - 750-1500 (Laufer 1989). In many cases, word knowledge is only partial, i.e. the learner may have mastered some of the word's properties but not the others. In fact, the plurality of word features to be learnt increases the probability of a word being only partially learnt.

One possible source of information about a totally new word, or a partially learnt word is the learner's dictionary. The phonetic script shows the right pronunciation, the grammatical specifications provide information about the syntactic behaviour of the

word, the derivatives in the entry show which words are morphologically related to the looked-up word. The definition explains the meaning, or meanings of the word, register specification adds information about the affective and pragmatic aspects of meaning. While each of the above mentioned parts of the entry provides one piece of information about the word, the example illustrating the word combines in it most of the information the learner needs to have about the word. This is so because a correct and natural use of a word in a sentence, or several sentences, will necessarily bring out the grammatical, semantic, pragmatic and collocational characteristics of the word.

1.2. Corpus-based examples and lexicographer-written examples

Since the function of an example is to illustrate the most typical properties and contexts of the new word, it is often claimed that the lexicographer's best source on information about dictionary examples is a large corpus of language that comprises real language as used by native speakers. Such language corpora are indeed available in the major dictionary companies and are used by them in the compilation of lexical entries. An issue of contention is whether dictionary examples should be authentic, with minor modifications for low-frequency words, or whether they should be only corpus-oriented. In the latter case, the corpus would serve as a guideline for the lexicographer, but the lexicographer would be free to make up his/her own examples. Fox (1987) argues for the superiority of authentic examples over lexicographer-made examples. She points out that the latter are often isolated, self-contained sentences, because lexicographers tend to produce sentences with too much information in them. Communication, on the other hand, is context dependent and so are the authentic examples. Authentic examples are not only grammatically correct, but also situationally appropriate as they were actually used in real life, unlike the made-up examples which are sometimes odd and not very likely to occur in a communicative act. Artificial examples, Fox argues, may not reveal the most typical usage of a word. This usage is shown by concordances from which the authentic examples are taken.

However, one can question the argument that lexicographer-made examples are not natural enough. Lexicographers who are educated native speakers of the language are bound to have correct intuitions about their mother tongue, about the grammaticality of the word, its typical use and its typical environment. These intuitions are not necessarily less correct than the intuitions of those language users who are represented in the corpus and are therefore not less reliable. Second, even if we accept the argument that lexicographer-made examples are less natural than the authentic ones, we might still prefer to see them in learner's dictionaries if their pedagogic value proved to be greater than that of the authentic ones. Both teachers and learners expect the learners' dictionaries to clarify the meaning of new words in the best possible way and help the learners use these words in speech and writing of their own. If 'artificial' examples are not different from authentic, or even better in attaining these objectives, one cannot sensibly argue against their inclusion in the learner's dictionary.

1.3. Previous research

The question of naturalness and pedagogic value as perceived by teachers of English was investigated by Rundell and Maingay (1990,1991). A sample of 25 words was given to a

group of EFL teachers. Each word was illustrated either by an authentic example or by a lexicographer-made example. The teachers were asked to state the source of each example and to evaluate its pedagogic value. The results showed that teachers were in general unable to distinguish between the two types of examples. There was also no correlation between the source of an example and its perceived pedagogic value. The first part of the experiment (distinguishing between authentic and made-up examples) was replicated with 'naive' native speakers i.e. who were not language teachers. They too were in general unable to spot the source of the examples.

If native speakers consider 'artificial' examples to be just as natural as the 'natural' ones, then the naturalness argument put forward in favour of the authentic examples is not very convincing. The claim for the pedagogic superiority of the authentic examples is doubtful too, since they were not rated as superior to the made-up examples on the perceived value test. One should not forget, however, that the pedagogic value measured by Rundell and Maingay was the PERCEIVED value as judged by native speakers who were familiar with the words that were illustrated. In order to investigate the REAL pedagogic value of a teaching tool we must test it by testing the performance of the learners who use it. If one teaching tool is better than another, then the use of the better tool should result in better performance by comparison to the use of a less successful tool. Therefore, if the authentic examples are superior to the lexicographer's examples, then illustration of new words by these examples should lead to better comprehension and use of these words than illustration of the same words by lexicographer's examples.

2. The Study

2.1. Aims

The aim of the study was to examine the difference in the effectiveness of two types of examples in the dictionary entry: a. examples taken from an authentic language corpus and b. examples written by lexicographers for the particular entries. The dictionary we are interested in is the learner's dictionary and the user of the dictionary is an advanced foreign language learner whose proficiency is roughly equivalent to the Cambridge First Certificate of English.

2.2. The subjects

The subjects were two classes of altogether 57 adult EFL learners in the University of Haifa from different departments. At the time of the experiment they were taking a course in English for Academic Purposes which emphasized reading skills.

2.3. Test items

Twenty low frequency words were chosen as the target words to be tested. The words were taken from the list of 25 words in the experiment by Rundell and Maingay (1990) and so were the two types of examples for each word. We used 20 words which were unfamiliar to the subjects. They were not a part of the high school syllabus and were not taught in the EFL university course prior to the experiment.

2.4. Procedure

The tests were taken during class time and each test was considered a regular language exercise similar to other exercises required in the course. The differences between the two types of examples were tested in two conditions: 1. the 'example only' condition and 2. the 'definition + example' condition. In condition 1 the subjects received a list of 20 target words and their illustrations. In condition 2 the subjects received the same 20 target words; only this time they were defined by a dictionary definition and also illustrated by an example. In condition 1, only comprehension of the new words was tested. In condition 2, we tested both comprehension and production of the new words. Comprehension was checked by students' translation of the target words; production by writing a sentence with each of the words. To compare the two types of examples, each test sheet included 10 items which were illustrated by authentic examples and 10 which were illustrated by lexicographer examples. To avoid a situation where all the students would have the same words illustrated by the same type of example, each half of the tests had different ten words illustrated by the same type of example. This way, each word was tested by two types of examples and each student was exposed to two types of examples. The scoring procedure was as follows: for each correct translation or a correct use of word in a sentence the subject got 2 points, for an approximate translation or word use he got 1 point; for an incorrect - 0 points. Several days later one class (27 students were present) was also given the Vocabulary Levels Test (Nation 1983) which measures the size of the learner's passive vocabulary.

3. Results

3.1. Condition 1 (examples only)

As stated earlier, only comprehension of the new words was tested. It was felt that the task of using a new word on the basis of an example only was too difficult. 26 subjects participated in experimental condition 1. The task of understanding the meaning of new words from the context of examples only was difficult, irrespective of whether the example was authentic or made up by a lexicographer. In the case of the authentic examples, out of the maximum score of 20 (10 words x 2 points), the mean score of the group was 2.15, range 0 - 8, sd 2.17; in the case of lexicographer examples, it was 4.15, range 0 -11, sd. 3.27. The difference between the means of the two types of examples was significant ($z = 3.39$, $p = .007$, Wilcoxon Matched-pairs Signed-ranks test). This shows that the lexicographer's examples provided better clues to the understanding of the new words than the authentic ones. One should not forget, however, that no matter what examples were used, if the example was the only source of information, comprehension of the words was unsatisfactory.

3.2. Condition 2 (Definition and example)

31 subjects took part in this experimental condition. The results for comprehension were as follows: understanding new words from definitions together with examples was much easier than understanding from examples only. In the case of words illustrated by authentic examples, the mean score was 9.32 (out of 20), sd 4.61, range 2-20. With words illustrated by lexicographer examples, the mean was 10.45, sd 4.28, range 0-18. The

difference between the two means was significant ($z=2.06, p=.038$) This shows that even if the new word is defined, the additional information which is provided by the lexicographer's example will contribute to the understanding of the word significantly more than the information provided in the authentic example.

As mentioned before, in condition 2 (definition + example), subjects were required to write sentences with the target words. The 'production' results were as follows: with the authentic examples, the mean score was 7.36 (out of the maximum of 20), range 0 - 20, sd 5.03. With the lexicographer examples, the mean was 8.36, sd 4.26, range 0 - 16. The difference between the two means was not significant ($z = -1.55, p = .12$). Thus, in production, unlike in comprehension, the subjects' performance was not significantly affected by the different type of the examples. A comparison was also made between comprehension scores in condition 1 (example only) and those in 2 (definition + example). With authentic examples, the comprehension scores were much higher in condition 2 ($z = -5.63, p = .0000$); the same was the case with lexicographer's examples ($z = 4.85, p = .0000$). This shows that an example alone cannot be expected to provide as much information as a definition with an example. (Another study which compared learners' comprehension of new words from definitions only with comprehension from examples only, (Laufer in preparation) suggests that examples alone provide less information than definitions alone).

3.3. Lexical level and the example-type.

Since some learners took the vocabulary size test we could calculate correlations between a. lexical level and comprehension scores for each type of example and b. lexical level and production scores for each type of example. A slightly higher correlation was found between lexical level and scores for words demonstrated by authentic examples than between lexical level and scores for words demonstrated by lexicographer's examples: between comprehension and authentic examples .45; between comprehension and lexicographer's examples .39; between production and authentic examples .55; between production and lexicographer's examples .49. This may suggest that the possible benefit from lexicographer's examples is less dependent on vocabulary level of the dictionary user than the benefit from authentic examples (even though the latter are modified for low frequency words).

4. Conclusion

The basic concern of this study was the pedagogic value of two types of examples: authentic and lexicographer-made. It was assumed that if one type was superior to another, this superiority would be reflected in better learner performance on new words. Below is the summary of the findings. 1. Comprehension of new words is better when the words are both defined and illustrated than when they are only illustrated by an example. 2. Lexicographer's examples are more helpful in comprehension of new words than the authentic ones. In production of the new word, lexicographer's examples are also more helpful, but not significantly so. 3. The usefulness of made up examples seems to be less dependent on the learner's general lexical knowledge than the usefulness of the authentic examples.

The findings suggest that lexicographer-made examples are pedagogically more beneficial than the authentic ones (further studies would be useful to substantiate this claim). If this is so, can they be considered as unacceptable on the grounds of lack of naturalness? One would assume that a group of lexicographers working together and consulting a language corpus should be able to suggest examples which are typical, natural, thought-provoking and surrounded by typical context. These are the main requirements stated in Fox (1987). As for the argument that one should only use examples that occurred in real life, we may ask ourselves how much of real life communication is an exact duplicate of what has previously been said or written. Certainly, some of it is. But if our ultimate goal of language teaching is achieved, new words will be used by learners in new contexts, in sentences never heard before or never produced before. The use of words may be modelled on the patterns the learner studied, but it will not necessarily be identical to them. A lexicographer, one hopes, can provide learners with appropriate and effective models of vocabulary use.

ACKNOWLEDGEMENT: I would like to thank my student Hedva Brochstein for collecting the data and Hagai Kuppermintz for the statistical guidance in the analysis of the data.

Bibliography

- CHOMSKY, N. (1975): *Aspects of the Theory of Syntax*. Cambridge, Mass. The MIT Press
- FOX, G. (1987): "The case for examples". In *Looking Up. An account of the COBUILD project in lexical computing*. Ed. by J.M. Sinclair. London. Collins
- GIBSON, E. J. and H. LEVIN. (1975): "On the perception of words: and application of some basic concepts". In *The Psychology of Reading*. Ed. by E. J. Gibson and H. Levin. Cambridge, Mass. The MIT Press.
- LADO, R. (1972): "Patterns of difficulty in vocabulary". In *Teaching English as a Second Language*. Ed. by H. B. Allen and R. N. Campbell. New York. McGraw Hill
- LAUFER, B. (1987): "A case for vocabulary in EAP reading comprehension materials". In *Beads or Bracelet? How Do We Approach LSP?* Ed. by A-m. Cornu, J. Vanparijs, M. Delahaye, L. Baten. OUP.
- LAUFER, B. (1991): "Knowing a word: what is so difficult about it?" *The English Teachers' Journal* 42: 82-88
- LAUFER, B. In preparation. "The effect of dictionary definitions and examples on the use and comprehension of new L2 words".
- MAINGAY, S. and M. RUNDELL. (1990): "What makes a good dictionary example?" Paper presented at the 24th IATEFL Conference, Dublin.
- MAINGAY, S. and M. RUNDELL. 1991: "What we have learnt about writing dictionaries from students and teachers". Paper presented at the 25th IATEFL Conference, Exeter.
- NATION, I. S. P. (1983): "Testing and teaching vocabulary". *RELC Journal (Guidelines)*: 12-24.
- NATION, I. S. P. (1991): *Teaching and Learning Vocabulary*. New York: Newbury House
- RICHARDS, J.C. (1976): "The role of vocabulary teaching". *TESOL Quarterly* 10: 77-89.

KEYWORDS: dictionary use, authentic examples, lexicographer-made examples, foreign learners' perspective, comprehension/production