Paul BOGAARDS, Leiden University

Scanning long entries in learner's dictionaries

Abstract

Compilers of learner's dictionaries have devised different systems which are thought to be helpful in guiding users who are looking for some specific piece of information in long entries. In order to examine the relative efficiency of these different guiding systems, an experimental study was undertaken in which advanced learners of English were presented with long entries in which they had to look up one particular piece of information. Semantics-based access structures turn out to be more adequate than other systems.

Keywords: Dictionary use, Learner's dictionaries, Search strategies

1. Introduction

All languages have several hundreds of very frequent words which have many different senses and uses and are part of a lot of fixed phrases and expressions. In English this is the case with verbs like *to get* or *to put*, adjectives like *old* or *poor*, and nouns like *party* or *hour*, or else with words like *round*, which may belong to a number of different grammatical classes. All these words have long, sometimes even extremely long, descriptions in one or more entries, in monolingual dictionaries for native speakers as well as in so-called learner's dictionaries. It is this very type of entry that is most demanding on the dictionary compilers, who spend much of their time devising ever better lay-outs in order to present these very rich materials in as well-organized a manner as they can.

At the same time it is well known that dictionary users do not like long entries. Although as yet it is not entirely clear how important their aversion to long entries is, part of their search strategies have, no doubt, to be interpreted in terms of avoidance. Another part of these strategies affects longer entries as well. As was found by Bogaards (forthcoming), words which have familiar forms but which are used in unknown senses tend to be looked up far less often than words whose form and meaning are unknown. *Faux amis* as well as new senses of familiar word forms ('allosemic words') seem to be underestimated as sources of errors by most of the students participating in Bogaards' experiment.

The most frequent words belong to the core vocabulary of the language and, for that very reason, constitute the defining vocabularies that are used in the learner's dictionaries. That means that the meanings of these words must be known to the users. If they were not, foreign learners would not be able to use these dictionaries at all. Defining such simple words in the dictionary creates considerable problems, all the more so when the definition is to be given within the limits of the defining vocabulary. Indeed, how is someone who does not know the word *give* to understand its definition if this contains far less frequent words like *supply* or *provide*?

This makes for a somewhat paradoxical situation: the entries which require most inventiveness on the part of the dictionary makers tend to be ignored by the dictionary users. More

specifically, for learner's dictionaries one can add that these entries contain much information that cannot be really useful to this group of users because it constitutes at the same time an absolute precondition for the use of these dictionaries.

It rapidly becomes evident that there is no simple way out of this situation. A dictionary which does not include this type of word seems to be inconceivable, all the more so because many less well known senses and uses are related to the more frequent ones. Especially in learner's dictionaries one would like to find these relationships explicitly given, as is actually the case in most of them, because they may facilitate the enormous task L2 learners have to perform in learning not just word forms but all the particular uses these forms can be put to.

So, there are and there have to be long entries, even in learner's dictionaries. As a consequence, L2 learners will have to find their way in these entries whenever they are confronted with unknown senses or uses of forms they are already familiar with. The problem I would like to address in this paper concerns the ways in which L2 learners of English scan long entries in order to find the particular piece of information they need.

I will limit the discussion to the context of L2 reading, to the exclusion of listening as well as of productive use of the dictionary. Readers may profit from three types of information about unknown elements they come across in a text. The first source of information is on the level of discourse: readers may be aware of the type of text they are reading, whether it is on medicine or mechanics, whether it was written in Britain or in the U.S., whether it is a literary text or a text taken from a newspaper, etc. Secondly, on the level of semantics, they may have some idea about the global meaning of the form they are confronted with. Finally, at the level of grammar, they may know to what part of speech the unknown word belongs and in what type of structure it is used.

As was shown in Bogaards (1996: 287, see also Bogaards, to appear), the four learner's dictionaries that exist for English as a foreign language do not contain many indications about discoursal features, except for styles and registers. The semantic and the grammatical information seems to be the important aspects of the access structures used by the four learner's dictionaries. LDOCE and CIDE exploit the semantic information: the *sign posts* and *menus* in LDOCE and the *guide words* in CIDE try to lead the users as fast as possible to the part of the entry which may serve them best. In COBUILD the most eye-catching element is the extra column, which mainly contains grammatical information. Except for some very long entries, OALD does not offer information of a special kind in order to guide the user.

This variety in guiding principles in the longer entries in learner's dictionaries offers an interesting point of departure for the study of search strategies among advanced non-native learners of English. In the second part of this paper I will report on the research I have conducted on this topic. The research was inspired by the following considerations. As a rule, readers are interested in the content of their text. It may thus be expected that access structures which disturb the comprehension procedure as little as possible will be most appreciated and will give the best results. Therefore, it may be expected that dictionaries with a semantics-based access structure, like LDOCE and CIDE, will be more adequate than those which use grammar as the main guiding principle, as in the case of Cobuild, or where there is no clear policy at all, as in OALD. In the case of the latter two dictionaries, it is difficult to see why a grammar-based access structure should be more adequate than the absence of any clear access structure. The use of grammar forces the reader to take some distance from his

text in order to realize what part of speech or what special structure is at hand, whereas the absence of a clear access structure permits the reader to stay on the semantic level.

2. Experiment

In order to study the ways in which advanced learners of English as an L2 look for the particular piece of information they need in the body of longer entries in learner's dictionaries, I carried out an experiment where Dutch pre-university students were asked to indicate the information they had used for translating words which were marked in a number of sentences.

2.1. Materials

English words were selected which belong to the 2,000 most frequent ones and may be assumed to be, therefore, part of the vocabulary of advanced non-native students, at least as far as one of their core meanings is concerned. For each of these words one less well known sense was selected. This sense had to be clearly represented in each of the four learner's dictionaries used in the experiment (LDOCE, OALD, COBUILD and CIDE). Each word was inserted in a context which was conducive to the one particular sense of the word that had been selected, but in such a way that it was practically impossible to guess the exact meaning of the word from that context. The sentences were translated into Dutch, except for the word which formed the object of the research. An example is:

> Is she still wearing those awful camp hats?
> Draagt ze nog steeds van die vreselijke ..... hoeden?

The materials selected were screened in several ways. First teachers were asked whether the senses that had to be looked up would be known by their students. Whenever there was any doubt, items were removed so as to keep only unknown senses in the selection. Then contexts were constructed with the help of a native speaker. The words in context were given to a group of university students with a higher level of proficiency in English than the groups of pre-university students that were to be used as subjects in the experiment itself. The university students were asked to try and give a translation of the target words. In those cases where the contexts turned out to be more or less transparent, these were changed. The Dutch translations (without the translation of the target words) were then given to another group of university students who had to fill in whatever they thought would fit in the context. This again led to some modifications. All in all 20 words were used in the experiment, all of them in senses that were unknown to the students of the target group and in sentences which did not contain any clear indication as to these particular senses.

Each sentence was copied at the top of a separate page together with the truncated translation. On the same page was a copy of a dictionary entry of the non-translated word. At the bottom of the sheet there was room to write the Dutch translation of the target word.

There were four versions of the test, each containing the same words in the same contexts, but in different orders. Moreover, in each version, a given word was presented with a copy of the corresponding entry taken from a different dictionary. Each subject was presented with four groups of five words in sentences; each word of a given group was accompanied by an entry taken from one dictionary. Thus all subjects worked with five entries from each of the four

dictionaries. The four versions of the test have been distributed randomly. For an example of a test sheet see Appendix 1.

2.2. Subjects and procedure

The subjects were 54 pre-university students, 31 male and 23 female, who had been taught English for about seven years. They were in the pre-final year of the highest level of secondary education. They all had Dutch as (one of) their native language(s). Their mean age was 16 to 17 years old.

The subjects were asked to read the English sentence and its Dutch translation on each page and to look up the meaning of the target word in the dictionary entry that was copied underneath it. When they had found the relevant information, they had to underline it.

In order to make an estimation of the time needed for the look-up procedures they were asked to write down the time two times on each page, first before starting to read the English sentence, secondly immediately after having underlined the relevant information. The time was given by a digital clock in front of them.

Summarizing, the routine that had to be followed for each item was the following:

1. write down the time
2. read the English sentence and its Dutch translation
3. look up the target word in the dictionary entry and underline the relevant information
4. write down the time
5. give the Dutch translation of the target word
6. go to the next page.

Before starting the real test there was an example page. After the last item there were questions about which type of dictionary entry they liked best, which dictionary they were already familiar with as well as questions about age, sex and mother tongue.

2.3. Results

For each word a check was first carried out to ascertain whether each subject had been able to locate the relevant information in the lemma. Only the cases where the right information had been underlined were taken into account in the analyses. For these cases the times noted at the top of the page and at the bottom were taken to give an indication of the time that had been needed to find the relevant information. The difference between these two moments was noted in number of seconds. For these same cases the Dutch translation that was given was classified as 'correct', 'nearly correct' or 'incorrect'.

Table 1 gives the means and the standard deviations of the numbers of seconds needed to look up the relevant information in the four learner's dictionaries. As can be seen in the table there is a mean difference of about 10 seconds between LDOCE and OALD. With CIDE the highest number of subjects find the correct information (82%), with OALD this number is much lower (67%). The overall percentage of correct look-up is 75.

Dictionary	Mean number of seconds	Standard deviation	Number of valid cases
LDOCE	42.30	25.14	208
CIDE	44.79	27.77	221
Cobuild	50.50	30.28	204
OALD	52.42	29.00	181
Total	47.32	28.32	814

Table 1. Mean numbers of seconds needed to look up the relevant information per dictionary.

In order to see whether the place of the relevant information in the entries or the total length of the entries could play a role in the look-up behaviour of the subjects, two values were entered: Place and Length. Place is defined as the number of lines between the beginning of a dictionary entry up to the line where the beginning of the relevant information could be found. Length is defined as the total number of lines in the entry. Table 2 gives the total numbers of lines for Place and Length as well as the means of these values for each of the four dictionaries.

As can be seen in Table 2, the number of lines to be read from the beginning of the entries up to the point where the relevant information was to be found varied from one dictionary to the other, OALD having the lowest number of lines and CIDE the highest. The total length of the entries varied less. Here again OALD was the shortest, but this time Cobuild is the longest; this latter point may be due to the particular lay-out of the Cobuild entries.

Dictionary	Place		Length	
	N lines	Means	N lines	Means
LDOCE	671	33.6	1384	69.2
CIDE	1059	53.0	1398	69.9
Cobuild	726	36.3	1469	73.5
OALD	456	22.8	1240	62.0
Total	2912	36.4	5491	68.6

Table 2. Data about the factors Place and Length per dictionary.

In a test of between-subjects effects, it turned out that dictionary and item both had highly significant effects on the dependent variable Time (i.e. the number of seconds needed to find the correct information; $F = 7{,}773$, $p < .000$ and $F = 2{,}547$, $p < .000$ respectively). The interaction between dictionary and item was not significant. This means that the dictionaries as such as well as the particular items had a non-random impact on the search strategies of the subjects, whereas the subjects were not influenced by the combination of items and diction-

aries. In an analysis of variance (ANOVA) it turned out that Place and Length had rather low, non-significant correlations with the dependent variable Time.

As to the translations given by those who had reached the right information, the numbers of 'correct', 'incorrect' and 'nearly correct' answers are given in Table 3. Whereas the percentages of incorrect translations vary very little, there is somewhat more difference between dictionaries when it comes to (nearly) correct answers. Cobuild seems to be a less good guide to clear understanding than the three other dictionaries: it has the lowest percentage of fully correct translations, the highest score for nearly correct translations as well as the highest percentage of missing cases. In the context of the present study it is only possible to speculate about the reasons why Cobuild seems to be less adequate as a receptive dictionary, but this result does not come as a surprise. As was argued by Bogaards (1996: 310), Cobuild seems to be better for productive than for receptive use.

Dictionary	Number of valid cases	Incorrect	Nearly correct	Correct	Missing cases
LDOCE	208	17 (8%)	35 (17%)	145 (70%)	11 (5%)
CIDE	221	16 (7%)	32 (14%)	156 (71%)	17 (8%)
Cobuild	204	16 (8%)	42 (21%)	125 (61%)	21 (10%)
OALD	181	14 (8%)	33 (18%)	125 (69%)	9 (5%)

Table 3. Numbers and percentages of 'incorrect', 'nearly correct' and 'correct' per dictionary.

Although Cobuild was regularly used in class, only about half of the subjects said they were familiar with that dictionary; CIDE was said to be known by only 2% of the students. As to the question which of the dictionaries they liked most, 54.3% of the students who gave an answer preferred CIDE. LDOCE was liked best by 23.9%, Cobuild by 13.0% and OALD by only 2.2% (1 student), whereas 6.5% had no preference.

Taking these results together, a rather strong image seems to emerge. LDOCE and CIDE, which both have access structures which are based on semantic principles, give the best results: students find the information they are looking for more often and they need less time to find it. The one dictionary without any clear guiding principles, OALD, has the least success: students more often do not find what they are looking for and they need more time (24% more than for LDOCE, 17% more than for CIDE), even though there is less text to scan and even though the number of lines to be read before the relevant information is reached is much lower than in the other dictionaries. Cobuild, which has a system which is mainly based on grammatical features, is in between these two extremes: it has nearly the same success rate as LDOCE as far as location of the right information is concerned but the use of Cobuild takes about 19% more time than LDOCE and about 13% more than CIDE. In addition, Cobuild seems to give less clear descriptions of the meanings involved than each of the other dictionaries.

3. Conclusion

Several conclusions can be drawn from this research. First, semantic guiding principles seem superior to access structures without clear guiding principles, and also better, but to a lesser degree, than access structures which are based on grammar. This conclusion confirms what was expected.

Second, the results indicate that dictionary entries are not read like other texts: people do not necessarily begin on the first line and go on until they have found what they need. Nor do they seem to take into consideration the whole entry before they start looking for a particular piece of information. The figures presented above strongly contradict such interpretations. Dictionary users like to take shortcuts and make use of search strategies which take them as fast as possible to the information they need.

Third, the preferences expressed by the subjects seem to indicate that they like a good mix of speed and success. This explains why they prefer CIDE, a dictionary they were not familiar with but which has the highest success rate and takes only slightly more time even though the information needed is to be found almost at the end of rather long entries. On the other hand, almost nobody liked OALD, presumably because this dictionary does not present clear structures in longer entries, which may lead to a low success rate and relatively long search procedures.

These conclusions should not be generalized too far, however. In the present research only main senses of words were used. One could do the same type of research with other types of word senses and uses, like idioms or fixed expressions, which are often entered under these main senses and are, therefore, less directly accessible than the meanings that were studied in this research. The present research has clearly shown, however, that these aspects can be studied in an experimental way and that it is useful to do so.

Acknowledgements

I would like to thank Dr W. den Brinker, Dr M. Hannay, Dr V. van Heuven as well as the teachers and the students of the Da Vinci College in Leiden for their assistance at different stages of this research.

4. References

CIDE = *Cambridge International Dictionary of English,* Cambridge: Cambridge University Press 1995.

Cobuild = *Collins Cobuild English Dictionary*, London: HarperCollins 1995[2].

LDOCE = *Longman Dictionary of Contemporary English*, Harlow: Longman 1995[3].

OALD = *Oxford Advanced Learner's Dictionary of Current English*, Oxford: Oxford University Press 1995[5].

Bogaards, P. (1996) Dictionaries for Learners of English. *International Journal of Lexicography 9*, 277 - 320.

Bogaards, P. (forthcoming) What Type of Words do Language Learners Look up? In: B.T.S. Atkins, K. Varantola (eds.) *Studies of Dictionary Use by Language Learners and Translators*, Tübingen: Niemeyer.

Bogaards, P. (to appear) Access Structures of Learner's Dictionaries. Paper read at "The Perfect Learners' Dictionary (?)", Symposium held at the University of Erlangen, 27 - 30 April 1997.

APPENDIX 1:

NOTEER EERST DE TIJD: ..... min. ..... sec.

3. We have seen some of the animals, but we have not found their <u>earth</u>.
We hebben een paar dieren gezien maar we hebben hun ..... niet gevonden.
a31

S2 W2 **earth**[1] /ɜ:θ|ɜ:r θ/ *n*
1 ▶ **WORLD** ◀ [singular] also **the Earth** the world that we live in: *the planet Earth|The earth revolves around the sun.|The space shuttle is returning to earth.* —see LAND[1] (USAGE) —see picture at SOLAR SYSTEM
2 ▶ **SOIL** ◀ [U] substance that plants, trees etc grow in: *footprints in the wet earth|a lump of earth*
3 ▶ **LAND** ◀ [singular] the hard surface of the world, as opposed to the sea: *After six months at sea, it was good to feel the earth beneath my feet again.* —see LAND[1] (USAGE)
4 what/why/how etc on earth...? *spoken* used when you are asking a question about something that you are very surprised or annoyed about: *What on earth did you do that for?*
5 cost/pay/charge the earth *informal* to cost etc a very large amount of money: *What a beautiful necklace! It must have cost the earth!*
6 the biggest/tallest/most expensive etc on earth the biggest etc example of something that exists
7 come back/down to earth (with a bump) to stop behaving or living in a way that is not practical: *When he realized he'd spent all the money he really came back to earth with a bump.*
8 ▶ **ELECTRICITY** ◀ [C usually singular] *BrE* a wire that makes a piece of electrical equipment safe by connecting it with the ground; GROUND[1] (30) *AmE*
9 ▶ **ANIMAL'S HOME** ◀ [C] the hole where a wild animal such as a FOX lives
10 go to earth *BrE* to hide in order to escape from someone who is chasing you
11 nothing on earth a strong way of saying 'nothing': *Nothing on earth would persuade me to repeat the experience of marriage.*
12 look/feel etc like nothing on earth *BrE* to look or feel very strange: *It looks like smoked salmon, but tastes like nothing on earth.*
13 run sb/sth to earth *BrE* to find someone, especially by looking in many places: *I finally ran him to earth in the stockroom.* —see also DOWN-TO-EARTH, **move heaven and earth** (HEAVEN (10)), **hell on earth** (HELL[1] (1)), **promise sb the moon/the earth** (PROMISE[1] (3)), **the salt of the earth** (SALT[1] (2))

earth[2] *v* [T] *BrE* to make electrical equipment safe by connecting it to the ground with a wire; GROUND[2] (5) *AmE*: *The amplifier wasn't properly earthed.*

Betekenis onderstreept? **Noteer eerst weer de tijd: ..... min. ..... sec.**

Vertaling van het onderstreepte woord: .....

Ga door naar de volgende bladzij.