Learners’ use of corpus examples

One of the distinguishing characteristics of corpus-based dictionaries is that most entries contain example sentences or phrases that have been copied or adapted from corpora. Although examples are generally regarded as positive and have high face validity among learners, the body of evidence about their actual benefits is limited and inconclusive. My aim in the present study is to revisit the idea of testing the usefulness of corpus examples. However, unlike previous research, in the present study different words are used to test language comprehension and the ability to correct typical language production errors, and a distinction is also drawn with regard to examples intended to facilitate decoding and examples meant to benefit encoding. In addition to this, because a single example might not be enough to help people understand what a word means or how it is used, in this study I also test the value of presenting learners with multiple corpus examples.

1. Introduction

One of the most distinguishing characteristics of corpus-based dictionaries, especially learners’ dictionaries, is that, in addition to definitions, most entries contain example sentences or phrases that have been copied or adapted from corpora. Although the presence of examples in dictionaries is generally regarded as positive (for example, see Béjoint 1981, Fox 1987, Humblé 2001 and Kilgarriff et al. 2008), the body of evidence about their actual benefits is limited and inconclusive.

In one of the first empirical studies in this field, Summers (1988) gave a group of students a reading passage and tested their comprehension of eight difficult words in a multiple-choice test. The students were then asked to produce sentences with these same words in order to test language production. To assess the value of dictionary definitions and examples, the students doing the test were given index cards with definitions only, with examples only, with definitions plus examples and empty cards. In the comprehension test, the differences between the three experimental conditions were not statistically significant. In the production test, the results were equally inconclusive, with the additional problem that some of the sentences produced by the students were too similar to the ones in the examples or not informative enough to indicate that the students had understood the words being tested.

A few years later, Lauffer (1993) tested the comprehension of 18 infrequent English words by asking 43 Hebrew undergraduates to translate them into their native language, and then asked the students to produce sentences with those words in order to test them for production. She found that, in the comprehension test, the effects of the definitions plus examples were slightly better than those of just definitions, which in turn were better than those of just examples. In the production tests, the combination of examples plus definitions also produced the best results, but the differences between presenting learners with just examples or just definitions were not significant.

In another study on the use of examples, Nesi (1996) asked 40 learners of English in the United Kingdom to produce sentences with 18 difficult words. The students were allowed to consult the Longman Dictionary of Contemporary English (LDOCE), but the examples present in nine of the entries were removed. Much against her expectations, she found that the dictionary examples did not appear to have a significant influence on language production. Although Nesi ingeniously prevented the students from writing sentences that were similar to the ones in the examples by requesting them to pair up each word being tested with a specified frequently-occurring word.
word, one of the problems of her experiment, as Humblé (2001) observed, was that all the 18 target words in the study were inadvertently of Latin origin. This means the words tested may not have been particularly difficult for the students if their mother tongue was a modern language like Portuguese, Spanish, French or Italian.

However, one of the main reasons why all three of the above studies present inconclusive evidence about the value of examples could be that they do not make a clear-cut distinction between language decoding and language encoding. As Summers (1988: 122) herself points out, in her experimental work, the “students were asked to do production tests for words which they also needed to look up for comprehension”. This is also true for the related studies carried out by Laufer (1993) and Nesi (1996), where learners were required to produce sentences containing words that were not familiar to them in the first place, which is something that does not usually happen in real life. As Laufer (1993:138) observes, “In the case of writing, people rarely use words which are entirely new to them”.

Another important problem might have to do with the very nature of the examples provided by dictionaries. For Humblé (2001:61-62), there is a difference between the kind of example learners need for encoding and the ones they need for decoding purposes: “A decoding learner is interested in the meaning of a lexical item, whereas an encoding learner is interested in a word’s syntactic features and collocates.” According to his analysis of five monolingual English learners’ dictionaries, the examples given tend not to deal with encoding and decoding needs in a satisfactory way. In fact, examples aimed to help learners with encoding by illustrating collocates and syntactic features can introduce an extra layer of complexity that might even end up compromising decoding.

This is what seems to have happened in the study by Al-Ajmi (2008) on the effects of dictionary examples on comprehension. 54 students of English in Kuwait were asked to translate into Arabic ten difficult words with the help of the *Oxford Advanced Learners’ Dictionary* (OALD). Half the students consulted definitions plus examples, while half consulted just definitions. The translations provided by the students were then checked for correctness against an English-Arabic bilingual dictionary, and Al-Ajmi noted that the dictionary examples had a negative influence on comprehension. These results come as no surprise. Because the examples given in dictionaries such as the OALD often provide added value to the core meaning expressed by definitions, the translations of those examples might not correspond to the more general translation equivalents favoured by bilingual dictionaries.

Thus despite lexicographers’ consensual recognition of the value of examples and their high level of face validity among dictionary users, research in the field presents unconvincing, contradictory and inconclusive results. I would therefore like to revisit the idea of testing the usefulness of corpus-based examples, but this time keeping words tested for comprehension and words tested for production apart, and differentiating between examples meant to help comprehension and examples intended to help production.

My starting point is the assumption that, one the one hand, a well-formulated dictionary definition summarizing a wealth of information extracted from corpora by trained lexicographers will aid comprehension more efficiently than corpus examples. On the other hand, when learners already know the meaning of a word and simply need to check how to use it in context, I believe judiciously selected corpus examples can be more helpful. However, as any experienced corpus user will know, a single example is often not enough to help people understand what a word means or how it is used.
Therefore, I would also like to test the value of presenting learners with multiple corpus examples. The present study thus set out to test the following four hypotheses:

H1 – Definitions will help learners understand words that are not familiar to them more effectively than corpus examples

H2 – Corpus examples will help learners use words that are already familiar to them more effectively than definitions

H3 - Multiple corpus examples will help learners understand words that are not familiar to them more effectively than a single corpus example

H4 - Multiple corpus examples will help learners use words that are already familiar to them more effectively than a single corpus example

2. Method

2.1 Subjects
The subjects participating in the present experiment were 48 second year undergraduates studying for a degree in Tourism at a Portuguese university. The native language of 46 participants was Portuguese. The two students who were not native speakers of Portuguese were nevertheless advanced users of the language. 14 participants were male and 34 were female, and they were on average 20.8 years old. They had studied English in school for an average of seven years, and at the time of the experiment they were on their third semester at university, each of which included three hours of English lessons per week, with a focus on professional English for Tourism. Common European Framework of Reference levels ranged from B1 to B2. Unlike language, linguistics, literature or translation students, these students were not in the habit of using dictionaries routinely.

Despite the fact that the participants were all studying for the same degree at the same university, some were naturally much better at English than others. They were therefore assigned to three different experimental groups and one control group in a completely random way. Each group consisting of 12 students each was required to do the comprehension and the production error tests that shall be described in the next couple of sections. The first experimental group consulted dictionary definitions, the second group accessed a single corpus example per test item, and the third group referred to three corpus examples per test item. The fourth, control group took the tests without any references at all. Details about the definitions and corpus examples used are provided further on, in sections 2.4 and 2.5.

2.2 Comprehension test
Ten different words that the participants were unlikely to be familiar with were tested for comprehension. These were words that satisfied the following two criteria after having been selected from a list generated at random: (1) they did not have a cognate in Portuguese (since all participants were either native speakers of Portuguese or had a good command of the language), and (2) they were unknown to a similar group of students (11 second year students studying for a degree in Hotel Management at the same university). The words selected are listed in table 1, along with their classification in terms of the five-diamond frequency-band system adopted in the COBUILD dictionary, where the higher the band the more common the word. As can be seen, of
the ten words selected for the comprehension test, one belongs to band 2, five to band 1, three have no band assigned and one was not frequent or important enough to deserve being included as a headword in COBUILD.\footnote{According to COBUILD’s front matter (p. xliii), words in bands 1 and 2 fall beyond the 3300 more frequent words in English, and words without frequency-band diamonds are words learners might read or hear but are unlikely to use very often.}

**Table 1.** Words selected for the comprehension test and their frequency bands in the COBUILD dictionary

<table>
<thead>
<tr>
<th>Word tested for comprehension</th>
<th>Frequency band in COBUILD</th>
</tr>
</thead>
<tbody>
<tr>
<td>grim</td>
<td>2</td>
</tr>
<tr>
<td>dwindle</td>
<td>1</td>
</tr>
<tr>
<td>listed (adj)</td>
<td>-</td>
</tr>
<tr>
<td>feigning</td>
<td>1</td>
</tr>
<tr>
<td>portly</td>
<td>-</td>
</tr>
<tr>
<td>mulish</td>
<td>not included as headword</td>
</tr>
<tr>
<td>plunder</td>
<td>1</td>
</tr>
<tr>
<td>slackened</td>
<td>1</td>
</tr>
<tr>
<td>brooding</td>
<td>1</td>
</tr>
<tr>
<td>denizen</td>
<td>-</td>
</tr>
</tbody>
</table>

The words tested for comprehension were presented in bold in the naturally-occurring context of a sentence taken from a general English corpus (the BNC, COCA or UKWaC)\footnote{Available at http://corpus.byu.edu/bnc/ (BNC), http://corpus.byu.edu/coca/ (COCA) and http://forbetterenglish.com/ (UKWaC via GDEX) – see also Davies (2008) and Kilgarriff et al (2008).}, some of which were slightly shortened or adapted in order to remove contextual clues that would allow the students to infer what the target words meant. The students were required to select the best Portuguese translation for those words in a multiple-choice test. To exemplify this, below is the test question for *feigning* (the full test is supplied in appendix 1).

*I lie back on the grass, feigning indifference.*

( ) olhando com ( ) reagindo com ( ) simulando

As can be seen, while the sentence context allowed the participants to infer which part-of-speech category the words being tested belonged to (in the above example, a verb), care was taken to make sure that the sentence itself was vague enough to prevent the students from deducing meaning from context.

The distractors in the multiple-choice options were then carefully chosen with the help of the DeepDict tool\footnote{Available at http://gramtrans.com/deepdict/} (Bick 2009) so that all three options fit the context of the sentences supplied. For example, in the above sentence, the three verbs in the multiple-choice options collocate frequently with “*indiferença*” (*indifference*), but only the last one is the translation of *feigning*.
2.3 Production errors test

Another ten words were then selected to test the students’ ability to correct typical language production errors. As already mentioned in the introduction, special care was taken to make sure that the words selected for this test did not pose any problems of comprehension: they were straightforward enough for the participants to understand, but at the same time tricky enough for them to have problems using them correctly in certain specific contexts. For example, the students had no difficulty in understanding the verb *call* meaning *to telephone*, but because of the influence of Portuguese, they often produced the error *to call someone to* (rather than *at* or *on*) *a particular number.* The words tested are listed in table 2, along with their frequency bands in the COBUILD dictionary and the type of production error associated with the word that had been repeatedly observed in class. All words in the production errors test fall within the 3300 most frequent words of English (bands 3-5), except for *disapprove*. However, as the latter is a cognate of the Portuguese “*desaprovar*”, it was highly unlikely that the participants would not understand it.\(^4\)

**Table 2.** Words selected for the production errors test, their frequency bands in the COBUILD dictionary, and typical errors involving the word

<table>
<thead>
<tr>
<th>Word used in the production errors test</th>
<th>Frequency band in COBUILD</th>
<th>Typical production error observed</th>
</tr>
</thead>
<tbody>
<tr>
<td>appreciate</td>
<td>3</td>
<td>*appreciate if you could/would do something</td>
</tr>
<tr>
<td>order (noun)</td>
<td>5</td>
<td>*to make an order for something</td>
</tr>
<tr>
<td>call (verb)</td>
<td>5</td>
<td>*to call someone to a particular number</td>
</tr>
<tr>
<td>foot (unit of measure)</td>
<td>5</td>
<td>*to have 6 feet</td>
</tr>
<tr>
<td>contract (noun)</td>
<td>4</td>
<td>*to finish a contract with someone</td>
</tr>
<tr>
<td>match (verb)</td>
<td>5</td>
<td>*to match something with something else(^5)</td>
</tr>
<tr>
<td>opposite</td>
<td>3</td>
<td>*to be opposite to something</td>
</tr>
<tr>
<td>recommend</td>
<td>4</td>
<td>*to recommend someone something</td>
</tr>
<tr>
<td>disapprove</td>
<td>1</td>
<td>*to disapprove something</td>
</tr>
<tr>
<td>good</td>
<td>5</td>
<td>*to be good in something</td>
</tr>
</tbody>
</table>

The ten words in table two were presented to the students in the context of sentences taken from general English corpora (again, the BNC, COCA or UKWaC) and then modified so as to exhibit the frequent errors involving the use of those words that had been previously observed in class. To exemplify this, below is the test question for *call* (the full test is supplied in appendix 1).

*Please call me to 919934487 if you have any questions.*

The participants were at first required to read the ten sentences and underline any word they did not understand, just to make sure that the word being tested for usage or indeed the sentences in which they were inserted did not pose any problems of comprehension.\(^6\) Next, the participants were informed that each sentence contained one

---

\(^4\) Note also that only the words in band five do not have cognates in Portuguese.

\(^5\) Note that *to match something with something else* is not considered to be grammatically wrong if the verb has a human agent. For example, *Emma is trying to match Harriet with Mr Walton.*

\(^6\) If the students underlined any of the specific words being tested, then the test item in question would be invalidated. If the students underlined any other word, the examiner would simply explain what it meant. See section 2.6.
mistake each and were asked to correct them either by crossing out words that were excessive, or by inserting words that were missing, or by replacing words that did not sound right with better ones. In the test question for call above, this would involve replacing the preposition to with at or on. The idea was not so much to give the students an abstract, decontextualized test, but rather, to emulate the real-life situation in which the students were requested to try and correct their own language production errors.\footnote{Chan (2011) used a similar elicitation procedure to test the use of grammatical information in a learners' dictionary. Frankenberg-Garcia (2011) too used sentences with common mistakes to find out what kind of reference learners would use when confronted with the need to look up information other than L1-L2 equivalents, such as collocations and syntactic features.} The students were told there was no need to rewrite the whole sentence and were asked to mark their corrections directly on the sentences supplied. The rationale behind this was to dissuade the students from rewriting entirely different sentences, avoiding the use of the target word being tested. If the students were unable to spot the mistake in a sentence, they were told to leave the sentence unchanged and move on to the next one (see full test in appendix 1).

2.4 Definitions
The experimental group that was allowed to consult dictionary definitions during the two tests was given a reference sheet containing definitions copied from the electronic version of the Longman Dictionary of Contemporary English (LDOCE), fourth edition. Bogaards and van der Kloot (2002) draw a very important distinction between the findability and usability of the information in dictionaries, and in this study care was taken to test only usability by making sure there were no problems of findability. The definitions reference sheet therefore made explicit reference to the test sentence each definition pertained to. Also, instead of presenting the learners with all possible senses of the headwords being tested, only the senses that were relevant to the test items in this study were given to the students. This was especially important in the case of polysemous words with long and complex entries, as the students might otherwise fail to find the relevant dictionary information by themselves. For example, the noun order has 25 main sense distinctions in the LDOCE, but only the seventh one (a request by a customer for a company to supply goods) was included in the reference sheet.

Because the focus of the study was on the effect of definitions alone, all other dictionary information such as part of speech, transitivity, countability, collocation, examples and usage notes were deleted. Note, however, that in some cases the information provided by the dictionary but hidden away from the students could have been of considerable help. For example, in the entry for the verb match, there is an explicit reference derived from the Longman Learners’ Corpus alerting users not to say match to or match with. This could have obviously helped the students to correct the mistake in test item 6 (A blue blouse to match with the colour of your eyes). However, as already mentioned, this information was not given to the students as it could interfere with the evaluation of the effects of definitions alone. The reference sheet with the complete set of definitions provided to the experimental group using definitions is supplied in appendix 2.

2.5 Examples
The two experimental groups that were allowed to consult corpus examples were supplied with reference sheets containing full-sentence concordances selected from
general English corpora (again, the BNC, COCA and UKWaC). There are several reasons for not using the examples supplied by the LDOCE itself (or by any other monolingual learners’ dictionary). To begin with, one of the experimental groups required three different examples per test item, but the LDOCE provided less than three or no examples at all for the specific senses of eight of the twenty words in the test: *denizen* and *listed* (0 examples); *brooding, feign, mulish, plunder, portly* (1 example each); and *slacken* (2 examples). Secondly, while some examples in the LDOCE are full sentences like *her family strongly disapproved of her behaviour*, others are just fragments such as *brooding eyes*. As the effects of providing the students with full sentences and fragments could differ, it was important to keep this variable constant so that it would not interfere with the results.

The third reason why it was not possible to use the examples already present in the dictionary was that the aim of the experiment was to test the use of suitable corpus examples rather than just any corpus example. In other words, for the ten words tested for comprehension, it was necessary to provide the students with examples that could facilitate decoding, and for the ten words in the production errors test, it was necessary to give the students examples that could help encoding. The examples provided by the LDOCE did not always meet these criteria. For instance, with regard to one of the words being tested for comprehension, it is not possible to infer what *slacken* means from the LDOCE example *Guy slackened his pace as he approached the gate*. Guy could very well be anxious to arrive and wish to hurry instead. Likewise, the LDOCE example *What time did Tony call?* does not contain any syntactic clue that, by analogy, could help a learner correct the test sentence *Please call me to 919934487 if you have any questions.*

So instead of using actual dictionary examples, the examples selected for the comprehension test were full-sentence concordances that provided contextual clues about the meanings of the words tested. For example, the following concordances were used to help the students understand the meaning of *dwindle*:

*The number of readers began to **dwindle** well before the economic crisis.*  
*The chances of getting back any of the money had **dwindled** to just about nothing.*  
*There were fewer boats docked every year, **dwindling** from fifteen to seven in as many years.*

In the first concordance, *economic crisis* signals that the number of readers is decreasing. In the second concordance, *to just about nothing* indicates that there is less rather than more money. In the third concordance, *fewer boats and from fifteen to seven* provide the contextual clues the reader needs to understand that the number of boats is diminishing.

The examples selected for the production errors test did not necessarily provide contextual clues about meaning, but focused instead on the target syntactic features and/or collocates of the words in the test. For example, to illustrate that you call someone *at* (or *on*) rather than *to* a particular number, the following concordances were used:

*Call him directly at 214-555-049 to give him your response.*

---

8 A few concordances were shortened to make the reference sheet fit into a single page.

9 Corpus evidence from COCA and the BNC suggests that *to call someone at particular number* is the preferred form for American speakers, while British speakers tend to favour *to call someone on a particular number*. The concordances selected for this particular test item were purposefully just examples of *call at* so as to avoid raising awareness to an added complication that was not relevant to the present test.
Humblé (2001:55) claimed that “Choosing an example to illustrate the meaning of an item for decoding is probably not as difficult as choosing one to show how an item can be used productively”. However, from my experience in selecting appropriate examples for the present study, I found quite the opposite was true. It was much easier to find concordances with appropriate collocations and syntactic features suitable for encoding purposes than ones with enough contextual clues for decoding purposes. In fact, for some comprehension test items, the concordances found for decoding purposes were far less satisfactory than the ones given for dwindle above. For example, as can be seen below, the concordances given to help learners understand the meaning of grim were a lot more subtle than the ones for dwindle. It is only indirectly that the collocates future, reality and news help the reader realize that grim is not something positive.

Whatever the answer, the future is grim.
It’s easy to forget the grim reality.
The news gets more grim for the Democrats every day.

The reference sheet with the full set of concordances supplied to the experimental group using multiple examples is available in appendix 3. The experimental group referring to a single corpus example was given just the first concordance of each set of three in that list. In the same way as in the definitions reference sheet, both the single and the multiple corpus examples reference sheets made explicit reference to the test sentences each example pertained to.

2.6 Procedure
The experiment was carried out in test conditions, during a 40-minute session. All the participants managed to finish the test well before the 40 minutes had expired. As already mentioned, the students were randomly assigned to four different groups. They were told that the test was part of an experiment involving English words and dictionaries and that, for this reason, three groups would be doing the test with the help of different types of reference sheets and one group (the control) would take the test without any references. The examiner made sure the students understood the instructions written on the test papers by explaining them again in Portuguese. The words being focused on in the production errors test did not pose any comprehension problems to the participants. However, some students asked questions about other words in the context. For example, some students were not sure what the word blouse meant in test item 6, some students wanted to know what lounge and flat-screen meant in test item 7, and one student wanted to know how much six feet in test item 4 was. All questions were clarified on the spot to make sure there were no comprehension problems in this section of the test.

When marking the tests, one point was given to each corrected sentence on the production errors test and to each correct answer in the multiple-choice comprehension test.

3. Results
The results for the comprehension test are summarized in table 3, where it can be seen how the four different groups performed.

**Table 3. Results for the comprehension test**

<table>
<thead>
<tr>
<th>Group</th>
<th>Mean Score (out of 10)</th>
<th>Standard Deviation</th>
<th>Range</th>
<th>Median</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control</td>
<td>3.17</td>
<td>0.94</td>
<td>2-5</td>
<td>3</td>
</tr>
<tr>
<td>Definitions</td>
<td>8.25</td>
<td>1.06</td>
<td>6-10</td>
<td>8</td>
</tr>
<tr>
<td>Single corpus example</td>
<td>4.58</td>
<td>2.39</td>
<td>1-9</td>
<td>4</td>
</tr>
<tr>
<td>Multiple corpus examples</td>
<td>8.08</td>
<td>1.93</td>
<td>3-10</td>
<td>8.5</td>
</tr>
</tbody>
</table>

As can be seen, the students who did not consult any references (control group) obtained the lowest average comprehension score. Of the three experimental groups, the students consulting definitions obtained the highest average, but the students referring to multiple corpus examples came very close. In fact, the median score was actually higher for the students using multiple corpus examples. The students with access to a single corpus example did not do much better than the control group. From the standard deviations and ranges, it can also be seen that the scores belonging to both groups using corpus examples varied more than the scores obtained by the students in the definitions and the control groups.

The results for the four different groups in the production errors test are summarized in table 4. It can be seen that the students referring to multiple corpus examples got by far the highest scores in this test. The students consulting a single corpus example came second. Both the students consulting definitions and the control group not using any references at all performed very badly when it came to correcting typical production errors, with averages below 1 point out of 10. The standard deviations and ranges show that, once again, substantial variation occurred among the two experimental groups using corpus examples, whereas the definitions and the control groups were more homogeneous.

**Table 4. Results for the production errors test**

<table>
<thead>
<tr>
<th>Group</th>
<th>Mean Score (out of 10)</th>
<th>Standard Deviation</th>
<th>Range</th>
<th>Median</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control</td>
<td>0.25</td>
<td>0.62</td>
<td>0-2</td>
<td>0</td>
</tr>
<tr>
<td>Definitions</td>
<td>0.33</td>
<td>0.49</td>
<td>0-1</td>
<td>0</td>
</tr>
<tr>
<td>Single corpus example</td>
<td>4.08</td>
<td>3.06</td>
<td>1-10</td>
<td>3.5</td>
</tr>
<tr>
<td>Multiple corpus examples</td>
<td>7.00</td>
<td>1.81</td>
<td>3-10</td>
<td>7</td>
</tr>
</tbody>
</table>

To test the first hypothesis, i.e., that definitions would help learners understand words that were not familiar to them more effectively than corpus examples, a one-way ANOVA was applied comparing the means of the four groups, which was then followed by a Games-Howell test in order to find out about the specific differences between the definitions and the three other groups. The results are summarized in tables 5a and 5b.

---

10 Note that, unlike the multiple-choice comprehension test, where the students had 1/3 of a chance of guessing the correct answer, the production errors test was not multiple-choice and the students could only score a point if they really knew the answer.

11 The Games-Howell test is suitable for data of unequal variance, as was the case in this study.
Table 5a. Comparison of comprehension scores to test H1

<table>
<thead>
<tr>
<th>ANOVA</th>
<th>Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between Groups</td>
<td>233.229</td>
<td>3</td>
<td>77.743</td>
<td>27.202</td>
<td>.000</td>
</tr>
<tr>
<td>Within Groups</td>
<td>125.750</td>
<td>44</td>
<td>2.858</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>358.979</td>
<td>47</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The F value in table 5a shows that it is highly unlikely that the differences between the four groups were obtained merely by chance. From the results in table 5b, in turn, it can be concluded that the dictionary definitions helped the students understand words that were not familiar to them more than no references at all and more than a single corpus example. However, one cannot conclude that the definitions helped comprehension more than multiple corpus examples, for the differences between the two groups were not significant.

Table 5b. Comparison of comprehension scores to test H1

<table>
<thead>
<tr>
<th>GAMES-HOWELL</th>
<th>Groups compared</th>
<th>Mean Diff</th>
<th>Std. Error</th>
<th>Sig.</th>
<th>95% Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>definitions</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>single corpus</td>
<td>3.667</td>
<td>.755</td>
<td>.001</td>
<td>1.49, 5.84</td>
</tr>
<tr>
<td></td>
<td>control</td>
<td>5.083</td>
<td>.407</td>
<td>.000</td>
<td>3.95, 6.22</td>
</tr>
<tr>
<td></td>
<td>multiple corpus</td>
<td>.167</td>
<td>.635</td>
<td>.993</td>
<td>-1.64, 1.97</td>
</tr>
</tbody>
</table>

To test the second hypothesis, that corpus examples would help learners use words that were already familiar to them more effectively than definitions, first a one-way ANOVA was applied comparing the means of the four groups. As the differences were significant (see table 6a), a Games-Howell test was applied next in order to find out about the specific differences between the corpus examples and the other groups. The results are summarized in table 6b, and they indicate that the null hypothesis can be rejected. Both a single corpus example and multiple corpus examples helped the students perform better in the production errors test than definitions or no references at all.

To test the third and fourth hypotheses, that multiple corpus examples would help learners understand words that were not familiar to them and use words that were already familiar to them more effectively than a single corpus example, two further comparisons rendered by the Games-Howell test were relevant. The results are shown in tables 7 and 8. As can be seen, in both tests, the differences between single and multiple corpus examples were significant, and the null hypotheses can be rejected.
Table 6a. Comparison of production errors scores to test H2

<table>
<thead>
<tr>
<th>ANOVA</th>
<th>Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between Groups</td>
<td>381.833</td>
<td>3</td>
<td>127.278</td>
<td>38.402</td>
<td>.000</td>
</tr>
<tr>
<td>Within Groups</td>
<td>145.833</td>
<td>44</td>
<td>3.314</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>527.667</td>
<td>47</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 6b. Comparison of production errors scores to test H2

<table>
<thead>
<tr>
<th>GAMES-HOWELL</th>
<th>Groups compared</th>
<th>Mean Diff.</th>
<th>Std. Error</th>
<th>Sig.</th>
<th>95% Confidence Interval</th>
<th>Lower Bound</th>
<th>Upper Bound</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>control</td>
<td>3.833</td>
<td>.901</td>
<td>.005</td>
<td>1.16</td>
<td>6.51</td>
<td></td>
</tr>
<tr>
<td></td>
<td>definitions</td>
<td>3.750</td>
<td>.894</td>
<td>.006</td>
<td>1.08</td>
<td>6.42</td>
<td></td>
</tr>
<tr>
<td></td>
<td>multiple corpus</td>
<td>-2.917</td>
<td>1.026</td>
<td>.049</td>
<td>-5.82</td>
<td>.02</td>
<td>5.82</td>
</tr>
<tr>
<td>single corpus</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>control</td>
<td>2.917</td>
<td>1.026</td>
<td>.049</td>
<td>.02</td>
<td>5.82</td>
<td></td>
</tr>
<tr>
<td></td>
<td>definitions</td>
<td>6.750</td>
<td>.552</td>
<td>.000</td>
<td>5.14</td>
<td>8.36</td>
<td></td>
</tr>
<tr>
<td>multiple corpus</td>
<td></td>
<td>6.667</td>
<td>.541</td>
<td>.000</td>
<td>5.07</td>
<td>8.26</td>
<td></td>
</tr>
</tbody>
</table>

To test the third and fourth hypotheses, that multiple corpus examples would help learners understand words that were not familiar to them and use words that were already familiar to them more effectively than a single corpus example, two further comparisons rendered by the Games-Howell test were relevant. The results are shown in tables 7 and 8. As can be seen, in both tests, the differences between single and multiple corpus examples were significant, and the null hypotheses can be rejected.

Table 7. Comparison of comprehension scores to test H3

<table>
<thead>
<tr>
<th>GAMES-HOWELL</th>
<th>Groups compared</th>
<th>Mean Diff.</th>
<th>Std. Error</th>
<th>Sig.</th>
<th>95% Confidence Interval</th>
<th>Lower Bound</th>
<th>Upper Bound</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>multiple corpus</td>
<td>3.500</td>
<td>.887</td>
<td>.004</td>
<td>1.03</td>
<td>5.97</td>
<td></td>
</tr>
<tr>
<td></td>
<td>control</td>
<td>4.917</td>
<td>.619</td>
<td>.000</td>
<td>3.14</td>
<td>6.69</td>
<td></td>
</tr>
<tr>
<td></td>
<td>definitions</td>
<td>-.167</td>
<td>.635</td>
<td>.993</td>
<td>-1.97</td>
<td>1.64</td>
<td></td>
</tr>
</tbody>
</table>

Table 8. Comparison of production errors scores to test H4

<table>
<thead>
<tr>
<th>GAMES-HOWELL</th>
<th>Groups compared</th>
<th>Mean Diff.</th>
<th>Std. Error</th>
<th>Sig.</th>
<th>95% Confidence Interval</th>
<th>Lower Bound</th>
<th>Upper Bound</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>multiple corpus</td>
<td>2.917</td>
<td>1.026</td>
<td>.049</td>
<td>.02</td>
<td>5.82</td>
<td></td>
</tr>
<tr>
<td></td>
<td>control</td>
<td>6.750</td>
<td>.552</td>
<td>.000</td>
<td>5.14</td>
<td>8.36</td>
<td></td>
</tr>
<tr>
<td></td>
<td>definitions</td>
<td>6.667</td>
<td>.541</td>
<td>.000</td>
<td>5.07</td>
<td>8.26</td>
<td></td>
</tr>
</tbody>
</table>

4. Discussion
The results presented in the previous section indicate that dictionary definitions helped the students with language comprehension more than a single corpus example, which is in accordance Laufer’s (1993) findings. However, going against my initial expectations, despite the fact that dictionary definitions constitute what is probably the most efficient
way of summarizing word meaning, in the present study they cannot be said to have aided comprehension more than multiple corpus examples. While one example was not enough to aid comprehension in any significant way, three corpus examples seemed to do the trick.

Of course, much can depend on the kind of example that is given. When examples provide added value to the core meaning expressed by definitions, they might indeed confuse more than help, which is what Al-Ajmi’s (2008) findings seem to suggest. However, when examples provide contextual clues that allow learners to infer basic meaning, as was the case in the present study, it appears that a series of just three corpus examples can help comprehension as effectively as carefully formulated definitions.

It should also be noted that examples for decoding are more likely to help when the target word is a low-frequency, difficult word like the ones in the present comprehension test, in which case the context will probably contain words that are easier than the target word itself, facilitating comprehension. For example, in the concordance The number of readers began to dwindle well before the economic crisis, dwindle is the least frequent and most difficult word in the sentence. The other content words in the context, i.e., number, readers, began and economic crisis are more common, easier words that are likely to help the reader infer the meaning of dwindle. If the target word is a very frequent word, however, it will be hard to find it occurring in the context of words that assist comprehension. Humblé (2001:64) refers to the following example for road in the COBUILD dictionary: There was very little traffic on the roads, where traffic is less frequent than road. Thus learners who do not understand the meaning of road are also unlikely to know what traffic means.

In the production errors test, the results obtained in the present study were very different from those of studies where students were required to create original sentences with previously unknown words. Neither Summers (1988), nor Laufer (1993), nor Nesi (1996) were able to find any substantial evidence in support of the use of examples in language production. In the very different circumstances under which language production problems were approached in this study, where the words tested were not unfamiliar to the students, but were words the students had trouble using in certain contexts, and where instead of asking the students to produce original sentences with those words (which would probably not elicit the target area of difficulty), they were simply asked to correct commonly observed production errors, resorting to corpus examples made all the difference. A single corpus example helped the students use those words appropriately more than definitions, and multiple corpus examples helped even more.

Another factor that may have counted in favour of these results is that instead of actual dictionary examples like the LDOCE ones used by Nesi (1996), the examples used in the present study were all full-sentence concordances exhibiting syntactic and collocation patterns that, by analogy, could help the students in the production errors test. In other words, the concordances chosen were purposefully biased towards the errors addressed in the test so as to stimulate what has come to be known as data-driven learning (Johns 1991), where learners are prompted to arrive at their own conclusions regarding language use as a result of exposure to corpus data. In contrast, the examples present in existing dictionaries, though often corpus-based, are not necessarily geared to language production errors and certainly do not provide repeated exposure to specific target structures that can be problematic to learners.

It should be noted, however, that information on the syntactic features and collocates of words is not a prerogative of corpus examples. Following the model
introduced in the first edition the COBUILD dictionary (Sinclair and Hanks 1987),
where full-sentence or contextual definitions not only explain meaning but also “help
people grasp the ways in which words are used” (Hanks 1987:135), the LDOCE
definitions of match and opposite shown in in appendix 2 were full sentences exhibiting
the target syntactic features which could have actually helped the students spot the
mistakes in match with and opposite to. The definitions of the other eight words in the
production errors test explained meaning, but did very little to help the students grasp
how the words tested were used in context. Instead of full sentences, they were what
Lew and Dziemianko (2006) refer to as single-clause and analytical definitions. A
closer look was therefore taken at the responses given by the definitions group to see if
they performed better in the test items involving the use of opposite and match than in
the other eight words selected for the production errors test. However, only one of the
twelve participants in that group managed to correct opposite to, and none of them was
able to correct match with. Their performance was not any different from that of the
control group, where one participant was able to correct opposite to without any
references at all. Thus there was not enough evidence to suggest that the definitions
group would have performed better in the production errors test if all ten definitions
given to them had been full-sentence definitions like the ones given for match and
opposite. 12

Another important point that needs further investigation is the question of
variance. In both the comprehension and the production errors tests in the present study,
the control and the definitions groups showed less variance in the scores obtained than
the two corpus-examples groups. In other words, in the comprehension test, the students
in the control group got consistently lower scores and the students in the definitions
group got consistently higher scores, but in the two corpus examples groups, some
students did very well and others did very badly. Likewise, in the production errors test,
the students in the control and the definitions groups got consistently bad marks, but in
the two corpus-examples groups, again, some students did very well and others very
badly. This could simply mean that in the two corpus groups there were both
exceptionally good and unusually bad students, while in the definitions and control
groups the students were more evenly distributed. However, because the students were
assigned to completely random groups, this is more likely to indicate that getting
information from corpus examples works better for some students than for others.
Interestingly, the scores by the students with access to multiple corpus examples
showed less variance than those by the students referring to a single example, so
multiple corpus examples seem to help more consistently than single corpus examples.
Of course, when discussing the question of variance, it should be noted that people are
generally familiar with dictionary definitions, but learning from corpus examples is
something the present group of students had not been requested to do before. It is
therefore possible that even the students for whom corpus examples did not work so
well in this test should learn how to profit from them with a little practice. In the future,
it would therefore be interesting to further explore this question of variance
longitudinally, and also with larger groups of students.

A final point that would need further investigation is whether similar results
would apply to groups of students from different cultural backgrounds. Western, Indo-
European students like the ones in this study may respond to corpus examples in a way
that is different from students from other parts of the world.

12 But see Dziemianko (2006) for an in-depth study on this use of definitions.
5. Conclusion and implications

Almost 25 years ago, Summers’s original experiment on the use of examples was motivated by the wish of the Longman team “to test the proposition that an ELT dictionary consisting only of examples would be more helpful than the usual combination of abstract definition plus examples” (Summers 1988:118). On the basis of that experiment, they “did not feel that the examples-only approach would be an improvement in ELT dictionaries” (Summers 1988:122).

Although the present study made no attempt to test the value of combining examples with other dictionary information, the results obtained suggest that it is quite possible for learners to understand new words and correct their own frequent mistakes by looking at examples alone, in what has come to be known among the teaching and language corpora community as data-driven learning (Johns 1991). Although a single concordance may not be enough, multiple concordances helped the students participating in the present experiment understand new words as much as dictionary definitions did, and also made a big difference when it came to helping them to correct the use of words that they understood but frequently misused.

Of course, none of this is of any practical value to the real world if learners are not able to find the examples that they need in the first place. It is only in controlled experimental conditions such as the ones described in the present study that it is possible to separate findability from usability. If finding the right definition is not always easy, especially in the case of polysemous words, finding examples that help decoding and examples that help encoding and differentiating between the two can be quite a daunting task.

Because learners’ dictionaries do not distinguish enough between examples for encoding and examples for decoding purposes (as noted by Humblé 2001), and because it is often hard to find in them enough or the right type of examples, the alternative would be to use corpora directly. However, resorting to the direct use of corpora can be just as challenging. The present researcher, who happens to be an experienced corpus user, had to consult three different corpora in order to find appropriate examples for this experiment. Language learners who need immediate answers to language comprehension and production questions clearly cannot be expected to do the same.

However, now that electronic dictionaries are becoming more and more mainstream, it is perfectly feasible to develop separate hyperlinks to: (1) a selection of examples with enough contextual clues to help decoding, and (2) another selection of examples with repeated syntactic and collocation patterns to support different possibilities of encoding. As space is not too much of a problem in electronic dictionaries, ideally it should be possible to make available not just one example per structure, but groups of three or four of each, so as to provide learners with concentrated patterns of language that would facilitate suitable generalizations.

On the one hand, for words with low frequency bands, emphasis should be placed on decoding examples, since most learners are likely to look them up for comprehension alone. On the other hand, for words with high frequency bands, special attention should be given to encoding examples, since these are the words that learners are required to use actively in everyday communication.

Acknowledgements

I would like to thank Robert Lew for his valuable feedback on the original idea for this study, Chris Butler for his helpful advice with statistics, and the anonymous reviewers.
and Paul Bogaards for their much appreciated comments and suggestions. The usual disclaimers apply.

References

a. Dictionaries


b. Other References


APPENDIX I: Production and comprehension test

Age: [ ] F [ ] M

Test type:

a. Read the sentences below and underline any words that you don’t understand.

1. I would appreciate if you could call me as soon as possible.

2. This morning he made an order for thirty-seven pairs of shoes in various styles.

3. Please call me to 919934487 if you have any questions.

4. The man has 6 feet and weighs 220 pounds.

5. After Manchester United finished his contract, he returned to Italy.

6. A blue blouse to match with the colour of your eyes.

7. In the lounge, a large flat-screen TV is opposite to the sofa.

8. I don't recommend you the experience, but I understand it.

9. He knew that I strongly disapproved the invasion.

10. He was really good in geography and design.

b. The sentences you have just read contain one mistake each. Can you correct them? There is no need to rewrite the sentences again. Simply mark your corrections directly on the sentences above either by crossing out words that are excessive, or by inserting words that are missing, or by replacing words that don’t sound right with better ones. If you can’t spot any mistake, do not mark anything.
c. Choose the best translation for the words in bold:

11. I took a **grim** satisfaction in seeing him suffer.
   (    )gratificante  (    )sombria  (    )enorme

12. The numbers seemed to **dwindle** each year.
   (    )duplicar  (    )aumentar  (    )diminuir

13. Surprisingly, the property is not a **listed** building.
   (    )classificado  (    )público  (    )administrativo

14. I lie back on the grass, **feigning** indifference.
   (    )olhando com  (    )reagindo com  (    )simulando

15. Monsieur Pallon was a **portly** man.
   (    )elegante  (    )gordo  (    )poderoso

16. She frowned, looking **mulish**, and clenched her hands in her lap.
   (    )teimosa  (    )nervosa  (    )preocupada

17. These are people who are paid to **plunder** and to rape.
   (    )destruir  (    )saquear  (    )atacar

18. Joe and Sarah **slackened** their pace.
   (    )apressaram  (    )acertaram  (    )abrandaram

19. He had dark hair, pale skin, **brooding** dark eyes, and broad shoulders.
   (    )perturbadores  (    )arregalados  (    )amendoados

20. He became a philanthropist, and a **denizen** of the city's night life.
   (    )frequentador  (    )promotor  (    )defensor
**APPENDIX II: LDOCE definitions**

**Sentence 1 – appreciate:** used to thank someone in a polite way or to say that you are grateful for something they have done

**Sentence 2 – order:** a request by a customer for a company to supply goods

**Sentence 3 – call:** to telephone someone

**Sentence 4 – feet:** a unit for measuring length, equal to 12 inches or about 30 centimetres

**Sentence 5 – contract:** an official agreement between two or more people, stating what each will do

**Sentence 6 – match:** if one thing matches another, or if two things match, they look attractive together because they are a similar colour, pattern etc.

**Sentence 7 – opposite:** if one thing or person is opposite another, they are facing each other

**Sentence 8 – recommend:** to say that something or someone is good, or suggest them for a particular purpose or job

**Sentence 9 – disapprove:** to think that someone or their behaviour, ideas etc are bad or wrong [≠ approve]

**Sentence 10 – good:** able to do something well

**Sentence 11 – grim:** looking or sounding very serious

**Sentence 12 – dwindle:** to gradually become less and less or smaller and smaller

**Sentence 13 – listed:** a listed building is one of historical interest in Britain, and is protected by a government order

**Sentence 14 – feign:** to pretend to have a particular feeling or to be ill, asleep etc

**Sentence 15 – portly:** someone who is portly, especially an old man, is fat and round

**Sentence 16 – mulish:** refusing to do something or agree to something, in an unreasonable way [= stubborn]

**Sentence 17 – plunder:** to steal large amounts of money or property from somewhere, especially while fighting in a war

**Sentence 18 – slacken:** to gradually become slower, weaker, less active etc, or to make something do this

**Sentence 19 – brooding:** looking thoughtful and sad

**Sentence 20 – denizen:** an animal, plant, or person that lives or is found in a particular place
APPENDIX III: Corpus examples

Sentence 1
I’d appreciate it if you could go down as soon as possible.
I’d appreciate it if you didn’t let the insurance company know about this.
I wouldn’t appreciate it if anyone told me to be more discreet.

Sentence 2
I placed an order for a 1995 Ford Mustang GT.
If you’re moved by something you see, you’re always welcome to place an order.
When someone calls before 10:30 a.m. to place an order, the order is shipped the same day.

Sentence 3
Call him directly at 214-555-049 to give him your response.
If you or someone you know is wrestling with these issues, give us a call at 800-989-8255.
For information or directions, call the church at 281-297-5900

Sentence 4
The average American woman is 5 feet 4 inches tall and weighs 140 pounds.
She was five feet eleven inches -- about average for a woman.
Since Zoe was six feet tall and skinny, the look was distinctive.

Sentence 5
I paid a lawyer five dollars to terminate the contract.
We’re terminating the contract, and we’re not disseminating any of this information.
If it was discovered that my conduct might damage the reputation of the company, my contract would be terminated.

Sentence 6
His inflamed nose matched the color of his brick-red hair.
Handmade shoes to match the dress.
Should a cake match the theme of a wedding?

Sentence 7
The Jewel Tower is opposite the Houses of Parliament.
The servants’ quarters are behind the kitchen opposite the stable.
I didn’t move until she was opposite the house.

Sentence 8
I have recommended the book to several friends.
Can you recommend a doctor?
I do not recommend this technique.

Sentence 9
Freud disapproved of the typical kind of dream book.
68 percent of the public disapproves of the way the governments is handling this.
Americans disapprove of spending more money rescuing automakers and banks.

Sentence 10
I am good at figures, I have an analytical mind, I think logically.
Why are so many men just not good at giving gifts to their ladies?
Women tend to be better at self-control than men
Sentence 11
Whatever the answer, the future is grim.
It’s easy to forget the grim reality.
The news gets more grim for the Democrats every day.

Sentence 12
The number of readers began to dwindle well before the economic crisis.
The chances of getting back any of the money had dwindled to just about nothing.
There were fewer boats docked every year, dwindling from fifteen to seven in as many years.

Sentence 13
It is an offence to demolish or to alter a listed building.
St James's Church was not listed as a building of architectural or historic interest.
The fact that the building is listed limited the opportunities for external alterations.

Sentence 14
Corbett rubbed his eyes, feigning to be more exhausted than he really was.
Patrick thought about feigning surprise, but knew the older man would see through him.
They are attention-seeking patients who feign illness or deliberately fabricate symptoms.

Sentence 15
By noon I was ready for lunch; portly people like me prefer our meals on time.
Portly old men should use belt and braces, in case their trousers burst open.
Portly Bavarians waddled out, having consumed vast quantities of beer with their enormous meal.

Sentence 16
They are silly, mulish, ego-deprived men.
Patti's mulish determination had its advantages.
It’s just plain mulish stubbornness.

Sentence 17
They were paid with privileges: food, tobacco, wine and medicines plundered from the dead.
Looters moved into the disaster area last night to plunder shops and abandoned flats.
Western allies continue to control and plunder the oil resources in the Arab region.

Sentence 18
Once at the door, we slackened our pace to a walk.
As it approached its speed slackened.
The rain was finally starting to slacken.

Sentence 19
In that one moment of brooding silence, Mom tried hard to find something to say.
He wore a sombre, brooding expression.
She’s known for her brooding, introspective portraits.

Sentence 20
 Specialists on every imaginable denizen of the ocean, from seaweed and starfish to octopuses.
Biologists began to study another desert denizen, the fat sand rat of the northern Sahara.
They too suffer horribly but, unlike the denizens of Hell, they have hope.