# The use of corpus examples for language comprehension and production

**Ana Frankenberg-Garcia**

Centre for Translation Studies, University of Surrey

a.frankenberg-garcia@surrey.ac.uk

## Abstract

One of the many new features of English language learners' dictionaries derived from the technological developments that have taken place over recent decades is the presence of corpus-based examples to illustrate the use of words in context. However, empirical studies have generally not been able to produce conclusive evidence about their actual worth. In Frankenberg-Garcia (2012a), I argued that these studies – and indeed learners' dictionaries themselves – do not distinguish sufficiently between examples meant to aid language comprehension and examples that focus on enhancing language production. The present study reports on an experiment with secondary school students carried out to test the usefulness of separate corpus examples for comprehension and production. The results support the need for different types of examples for comprehension and production, and provide evidence in support of data-driven learning, particularly if learners have access to more than one example.

**Keywords**: corpora, corpus examples, data-driven learning, dictionaries

## 1. Introduction

One of the main contributions of technology towards the remarkable developments of pedagogical lexicography over recent decades is the emergence of

learners' dictionaries supported by example sentences or phrases that have been copied or adapted from corpora. However, a series of empirical studies to assess the effects of corpus-based examples in dictionaries has presented limited or inconclusive evidence about their benefits. Using as subjects an unspecified number of students attending a post-First-Certificate (upper-intermediate) level class at the Bell School in Cambridge, Summers (1988) compared the effects of definitions only, examples only and definitions plus examples on language comprehension and production. She found that the differences between the three experimental conditions were not statistically significant. Laufer (1993) carried out an analogous experiment with a group of 43 Israeli undergraduates. 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 test, the combination of examples plus definitions produced the best results, but the differences between presenting learners with just examples or just definitions were not significant. Nesi (1996) asked 40 learners of English in Britain to produce sentences with 18 difficult words with the help of the *Longman dictionary of contemporary English* (LDOCE) (Summers 1996/2003) after the examples of nine of the entries had been removed. Going against her expectations, she was not able to find evidence that examples had helped.

In Frankenberg-Garcia (2012a), I argued that one of the reasons why these early studies on the value of examples presented inconclusive evidence could be that they did not distinguish sufficiently between language comprehension and language production. The idea that receptive tasks need to be clearly distinct from productive tasks is not new (for example, see Van Patten 1994). In the language production tests conducted by Summers (1988), Laufer (1993) and Nesi (1996), however, the subjects were required to write sentences with words that they had to look up for comprehension first. Apart from

the intuitive fact that it is not natural for writers to employ words that they are not familiar with[1], the comprehension factor may have constituted an important intervening variable in the production tests.

Another problem is that, as Humblé (2001) points out, dictionaries that make use of corpus examples do not seem to make a satisfactory distinction between examples intended to aid comprehension and examples meant to benefit production. It is important to remember that corpus-derived input can serve two distinct purposes: (a) enriching meaning representations to foster comprehension, and (b) making the learner aware of morpho-syntactic and distributional properties in support of production. While example sentences meant to facilitate the comprehension of a previously unknown word should contain sufficient context to enable a learner to infer what that word means, example sentences for language production should focus on which other words frequently go together with the target word (collocation) and on the grammatical preferences and constraints of the target word (colligation).

In dictionary example (1) taken from the entry for *showcase* supplied in the *Oxford advanced learner's dictionary* (OALD) online[2], *found a film role* and *all his talents* can help us understand what *showcase* means – it is primarily a decoding example. On the other hand, the example from the entry for *daytime* (2) does not provide enough context to help us guess what *daytime* might mean, but provides information on colligation, showing that the preposition *in* is used before *daytime* and that the definite article that precedes *daytime* is optional. Example (3) from the entry for *failure*, in turn, illustrates that *doomed* is a frequent collocate of *failure*, but because *doomed* is a rather 'difficult' and infrequent word[3], it may not serve to clarify the

[1] Summers (1988) and Laufer (1993) themselves recognize this limitation.
[2] OALD online available at http://oald8.oxfordlearnersdictionaries.com/
[3] With only 613 hits in the British National corpus, against 8,759 hits for *failure*.

meaning of *failure.* Both examples (2) and (3) are therefore primarily encoding examples.

(1) Jack found a film role that showcased all his talents.

(2) You don't often see this bird in (the) daytime.

(3) The attempt was doomed to failure.

Examples that help with encoding and decoding simultaneously are not very easy to find. Moreover, while simple decoding examples do not always exhibit the target collocation or colligation required to satisfy a learner's encoding needs, examples that focus on encoding like (3) can be overly complex for a learner looking up a word for decoding purposes. This could explain in part why, in a more recent experiment, Al-Ajmi (2008) found that examples from the OALD failed to help a group of Kuwaiti learners of English in a comprehension test.[4]

Bearing this problem in mind, Frankenberg-Garcia (2012a) carried out an experiment with 48 Portuguese undergraduates to test the usefulness of separate corpus-based examples for encoding and for decoding. Care was taken to make sure the examples provided for the comprehension test contained contextual clues to help decoding, and the examples supplied in the production part of the test were hand-picked so as to ensure that they illustrated the target collocation or colligation. Unlike the previous studies discussed, the results obtained in Frankenberg-Garcia (2012a) provide clear evidence in support of data-driven learning (Johns 1991): examples containing contextual clues to aid comprehension helped the learners understand new words in

[4] Another reason why Al-Ajmi's (2008) experiment failed to provide evidence in favour of examples was that the student translations given in the comprehension test were only considered correct if they matched the translations supplied in a bilingual Arabic-English dictionary, when in actual fact other translations might have been acceptable.

English, and examples illustrating collocation and colligation helped these learners correct typical L2 mistakes.

At this juncture it is important to clarify that the term data-driven learning, as proposed by Johns (1991), simply refers to the ability to use corpus data to figure out – as opposed to being told – what words mean or how they are used. It is strictly in this sense that I refer to data-driven learning. This is must not be confused with the use of the term 'learning' by second language acquisition (SLA) scholars, where learning can only be said to have taken place when subjects are able to recall and retain the new language they have been exposed to. In other words, data-driven learning is about learners being able to reach their own conclusions about language simply by being exposed to data. Whether data-driven learning will lead to actual learning in the SLA sense is a different question. According to the noticing hypothesis (Schmidt 1990), language input does not become intake unless it is consciously registered. While there have been several empirical studies supportive of this hypothesis (e.g. Leow 2000; Izumi 2002; Mackey 2006), it is important to note that there are two separate processes involved here: first, noticing, and second, converting the input that has been noticed into intake. Frankenberg-Garcia (2012a) and the present study are simply about whether corpus examples will help learners notice what words mean and how they are used. No claims about subsequent learning are being made.

It is also important to understand that a single corpus example only provides learners with a one-shot exposure to new vocabulary and grammar, and this can be insufficient for noticing to take place. As any experienced corpus user will know, it often takes more than one concordance line to help figure out what a word means or how it is used. Indeed, Frankenberg-Garcia (2012a) found that multiple examples were

more effective than a single one, both in helping learners understand new words and in helping them correct typical L2 mistakes.

Despite the promising results obtained in Frankenberg-Garcia (2012a), it is important to acknowledge that it was a one-off study conducted in a specific higher education setting – as indeed are most of the data-driven-learning experiments described in the literature (Boulton 2010). The need for replication of isolated pieces of research in applied linguistics is amply discussed in Porte (2012). Without it, it is not possible to separate the knowledge obtained from one particular study from the context in which it was originally found. According to Gast (2010), replication allows one to evaluate whether previous findings are reliable (i.e. internal validity), whether they are generalizable (i.e. external validity), and to learn more about conditions under which the previous findings do not apply. One of the aims of the present study is to find out whether the general findings of Frankenberg-Garcia (2012a – henceforth referred to as the 2012 study) are reproducible in the context of secondary education.[5] The kind of replication attempted here is what Porte (2012) and Abbuhl (2012) identify as "conceptual replication", where in addition to using a different population sample, two important changes in the research design were also made. First, the data collection procedure to test the value of encoding examples is different: while the 2012 study asked participants to correct sentences containing typical L2 errors in the use of the target lexical items tested, the present study asked them to write their own sentences using the target lexical items. Second, in order to compare the performance of subjects using corpus examples with that of subjects using dictionary definitions, one of the experimental groups in the 2012 study did the test with the help of definitions from the

[5] As Boulton (2010) observes, most data-driven-learning studies draw on experimental work with university students, and there is a general lack of data-driven-learning studies with younger, school-age learners. See also Boulton's 2012 update to Boulton (2010), available at http://corpuscall.eu/file.php/5/0_DDL_empirical_survey_2012_July.pdf

fourth edition of the *Longman dictionary of contemporary English* (Summers 1996/2003). However, not all definitions provided exhibited the target collocation/colligation required for language encoding. In the present study, care was taken to make sure that all the definitions used in the production test showed the target collocation/colligation.

The aim in the present study is therefore twofold. First, to test whether corpus examples will help a different population of language learners, namely, secondary school students instead of undergraduates. Second, after introducing several modifications to the research design of the 2012 study, to obtain a clearer and more detailed picture of the value of separate encoding and decoding examples. More specifically, the research questions are:

1. Will decoding examples help learners understand words that are not familiar to them?
2. Will definitions help learners understand words that are not familiar to them more than decoding examples?
3. Will encoding examples and definitions that exhibit target collocation/colligation help learners in language production?
4. Will a definition that exhibits target collocation/colligation help learners in language production as much as a single encoding example?
5. Will multiple encoding examples help learners in language production more than a single example and more than a definition that exhibits target collocation/colligation?

The expected answers to these questions are summarized in Table 1. For comparative purposes, the questions posed and the results obtained in the 2012 study are summarized in Table 2 .

**Table 1.** Expected results of the present study

| | **Helps secondary school students to understand new words?** | | **Helps secondary school students to use error-prone words?** |
|---|---|---|---|
| **single decoding example** | yes | **single encoding example** | yes |
| **multiple decoding examples** | yes, and more than a single example | **multiple encoding examples** | yes, and more than both a single example and definitions |
| **definition exhibiting target collocation / colligation** | yes, and more than a single example, but not more than multiple examples | **definition exhibiting target collocation / colligation** | yes |

**Table 2.** Summary of the findings reported in Frankenberg-Garcia (2012a)

| | **Helps undergraduates to understand new words?** | | **Helps undergraduates to identify typical L2 errors?** |
|---|---|---|---|
| **single decoding example** | yes | **single encoding example** | yes |
| **multiple decoding examples** | yes, and more than a single example | **multiple encoding examples** | yes, and more than both a single example and definitions |
| **LDOCE definition** | yes, and more than a single example, but not more than multiple examples | **LDOCE definition** | no |

## 2. Method

### *2.1 Subjects*

The participants were 50 students attending the eleventh (i.e. second last) grade of a secondary school in Portugal, with seven years of school English and generally at

B1 level. The female/male ratio was balanced (24:26) and the participants were on average 16.6 years old.

The subjects were randomly assigned to four different groups – three experimental and one control – in order to take the comprehension and production tests described in sections 2.2 and 2.3. The first experimental group (12 students) took the tests with the help of dictionary definitions (Definitions group), the second group (14 students) accessed a single corpus example per test item (1Ex group), and the third group (13 students) referred to three corpus examples (3Ex group) per test item. The fourth, control group (11 students) was not allowed to consult any references. Details about the definitions and corpus examples used are provided 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 in a multiple-choice translation test. The words selected for this test were chosen randomly, but had to satisfy the following two criteria: (1) they should not have cognates in Portuguese, and (2) they had to lie outside the Oxford 3000[6] list so as to ensure that they were not among the words that learners of English are most frequently exposed to. As shown in Figure 1, the target words were presented in bold within sentences taken from a general English corpus (the BNC, COCA or UKWaC)[7], some of which were slightly shortened or adapted in order to remove contextual clues that could have allowed the students to infer meaning. The distractors in the multiple-choice options were then selected with the help of the DeepDict tool[8] (Bick 2009) and Word Sketches from the Portuguese TenTen Corpus (Kilgarriff,

[6] See http://oald8.oxfordlearnersdictionaries.com/oxford3000/
[7] 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, Husák, McAdam, Rundell and Rychly (2008).
[8] Available at http://gramtrans.com/deepdict/

Jakubiček, Pomikalek and Whitelock 2012)[9] so that all three options fitted the context of the sentences supplied. For example, in test item 4, the three verbs in the multiple-choice options collocate frequently with *indiferença* (*indifference*), but only the last one is the translation of *feigning*.

**Figure 1.** Comprehension test (target English words in bold, multiple-choice options in italics, correct Portuguese equivalents in bold)

1. Use a soft brush to loosen the **grime**.
   [*pelo/**sujidade**/superfície*]
2. Stella crossed the street and **loitered** outside the store window.
   [***ficou a vaguear**/ficou a esperar/ficou a sonhar*]
3. Two hundred tons **silage** went up in flames.
   [*combustível/**feno**/lenha*]
4. I lie back on the grass, **feigning** indifference.
   [*olhando com/reagindo com/**simulando***]
5. Monsieur Pallon was a **portly** man.
   [*elegante/**gordo**/poderoso*]
6. She believed they had the ability to **mend** things.
   [*inovar/produzir/**reparar***]
7. For a second he felt like **retching**.
   [*rir muito/**vomitar**/festejar*]
8. Joe and Sarah **slackened** their pace.
   [*apressaram/acertaram/**abrandaram***]
9. He had dark hair, pale skin, **brooding** dark eyes, and broad shoulders.
   [***meditativos**/arregalados/penetrantes*]
10. He became a philanthropist, and a **denizen** of the city's night life.
    [***frequentador**/promotor/defensor*]

*2.3 Production test*

[9] Available at http://www.sketchengine.co.uk/

Another ten words were tested for language production. The words selected had to be words likely to be familiar to learners of English at B1 level (they were taken from the above-mentioned Oxford 3000 list), but at the same time they had to generate typical errors by Portuguese learners of English. For example, most secondary school students have no difficulty in understanding the verb *call* meaning 'telephone', but because of the influence of Portuguese, they often produce the error *call someone to* (rather than *at* or *on*) *a particular number*.

Whereas in the 2012 study the data had been elicited via an error-correction test, in the present study the students were given ten sentences in Portuguese and were asked to write down an English equivalent for each. The production test items are listed in Figure 2 along with the typical errors associated with them. In test item 1, for example, the students had to write down how they would say *Vota no teu DJ favorito* (*Vote for your favourite DJ*) in English, using the words supplied in brackets in their sentences: *vote, favourite* and *DJ*. The target word in this test item was *vote*, and the specific area of difficulty was the preposition used after it, for Portuguese speakers often say **vote in someone* rather than *vote for someone.*

To dissuade the students from translating word for word, they were specifically instructed not to give literal translations, but rather to give natural equivalents. The English vocabulary given in brackets also meant the students were not confronted with the added difficulty of having to think how to say Portuguese words in English, although they were, of course, forced to select appropriate colligations/collocations for the words supplied.

**Figure 2.** Production test (target words in bold) and typical errors (in italics)

1. Vota no teu DJ favorito. (**vote**, favourite DJ)
   **Vote in your favourite DJ*

2. Acho que depende da situação política. (**depends**, political situation)
   **I think it depends of the political situation*
3. Ligue-me amanhã para o 212 745 975. (**call** me, tomorrow)
   **Call me tomorrow to 212 745 975*
4. Escolha sapatos que combinem com a cor do vestido. (choose, shoes, **match**, dress)
   **Choose shoes that match with the colour of the dress*
5. Napoleão usava sempre uma camisa vermelha. (Napoleon, always, a red **shirt**)
   **Napoleon always used a red shirt*
6. Já paguei a viagem. (**paid**, trip)
   **I already paid the trip*
7. Telefone para o banco amanhã de manhã. (**telephone**, bank, tomorrow morning)
   **Telephone to the bank tomorrow morning*
8. Não podemos aprovar esse tipo de comportamento. (**approve**, type of behaviour)
   **We can't approve this type of behaviour*
9. A piscina tem vinte metros de comprimento. (swimming pool, 20 **metres**)
   **The swimming pool has 20 metres*
10. Nunca fui bom a matemática. (never been, **good**, maths)
    **I have never been good in maths*

*2.4 Definitions*

The Definitions group was given a reference sheet stating which test item each definition pertained to, and only the senses that were relevant to the test items in the study were supplied. 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 sense[10]. Because the focus of the study was on the effect of definitions alone, all other dictionary information such as part of speech, transitivity, examples and usage notes were left out.

[10] Bogaards and van der Kloot (2002) distinguish between the "findability" and "usability" of the information in dictionaries; the present study aimed to test only usability, without the intervention of findability.

Unlike the LDOCE definitions supplied in the 2012 study, all the definitions supplied for the production test of the present study exhibited the target collocation/colligation. Many of them were copied from COBUILD online[11], for it is a policy of the COBUILD dictionary to provide learners with definitions that not only explain meaning, but also help with the ways in which words are used (Hanks 1987). However, the COBUILD definitions did not help with the target collocation/colligation of six production test items. For example, COBUILD's definition of the sense of *call* that is relevant to the test is "If you call someone, you telephone them". This does not help the students with production test item 3 (see Figure 2). As it was not possible to find ready-made definitions that met the experimental criteria in other dictionaries either, customized definitions were conceived for the six test items in question. The reference sheet with the definitions used is supplied in Appendix 1.

*2.5 Corpus examples*

The two experimental groups that consulted corpus examples were supplied with reference sheets containing concordances selected from general English corpora (again, the BNC, COCA and UKWaC)[12]. Care was taken to ensure that all the examples supplied for the comprehension test contained contextual clues to facilitate decoding. For example, the following concordance lines were used to help the students understand the meaning of *feigning*:

(a) Corbett rubbed his eyes, **feigning** to be more exhausted than he really was.

(b) Patrick thought about **feigning** surprise, but knew the older man would see through him.

[11] Available at http://www.mycobuild.com/free-search.aspx

[12] A few samples were shortened to make them fit on a single line.

(c) Attention-seeking patients who **feign** illness or deliberately fabricate symptoms.

As can be seen, *to be more exhausted than he really was* in (a), *the older man would see through him* in (b), and *deliberately fabricate symptoms* in (c) all signal that *to feign* means 'to pretend'.

The examples supplied for the production test did not necessarily provide contextual clues about meaning, but purposefully focused on the target collocation/colligation addressed in the test. For example, to show that the preposition *for* was required after *paid* in production test item (6), the following concordances were selected:

(d) Peggy helped pay for the honeymoon.
(e) And they actually paid for the guns?
(f) Nicole sold photos to pay for the funeral.

The reference sheet with the full set of concordances supplied to the 3Ex group is available in Appendix 3. The 1Ex group was given just the first line of each three. Just as in the definitions reference sheet, the corpus examples reference materials also indicated which test sentences each example or set of examples pertained to.

*2.6 Procedure*

The experiment was carried out in test conditions after the students had been randomly assigned to three experimental and one control group. The students 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 materials and one group (the control) would take the test without any reference materials. The students in the experimental groups were instructed to read the 'help' provided for each test question before they attempted to answer it. The examiner made sure the students understood the instructions written on the test papers by explaining them again in Portuguese. The students were also told that they could ask for clarification if there was any word in the production test that they did not understand, but no questions actually arose.

When marking the tests, one point was given to each correct answer in the multiple-choice comprehension test, such that the maximum score was ten. Two students left one question in their tests blank, so the maximum score in their tests was nine, and a proportional mark out of ten was then calculated. In the production test, one point was given to each correct English sentence that matched the Portuguese prompts. The elicitation procedure used in the production test was successful in preventing the students from writing sentences that avoided the use of the target area of difficulty, for only two instances of avoidance were detected. The test items in question were therefore excluded from the analysis using the same procedure as for the comprehension questions that had been left blank. Mistakes that were unrelated to the target area of difficulty in the test (such as spelling mistakes) were ignored.

## 3. Results

Table 3 summarizes the results obtained in the comprehension test. It shows that the Control group obtained the lowest mean and median comprehension scores. Of the three experimental groups, the Definitions group obtained the highest mean and median, but the 3Ex group came very close, and the1Ex group did better than the Control group. The production test results are summarized in Table 4. As can be seen, the 3Ex group

was the one that performed best; next came the 1Ex group and finally the Definitions group.

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

| Group | Mean Score (out of 10) | Standard Deviation | Range | Median |
|---|---|---|---|---|
| **Control** | 4.40 | 1.43 | 2-7 | 4.4 |
| **Definitions** | 8.17 | 2.04 | 4-10 | 9 |
| **1Ex** | 6.41 | 1.73 | 4-9 | 6.5 |
| **3Ex** | 7.46 | 1.05 | 6-9 | 8 |

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

| Group | Mean Score (out of 10) | Standard Deviation | Range | Median |
|---|---|---|---|---|
| **Control** | 2.45 | 2.02 | 0-5 | 2 |
| **Definitions** | 3.78 | 2.74 | 0-8 | 4 |
| **1Ex** | 4.35 | 2.10 | 1-7 | 4.5 |
| **3Ex** | 6.59 | 2.43 | 2-9 | 7 |

To examine the comprehension results from the perspective of the research questions posed, a one-way ANOVA was applied comparing the comprehension means of the four groups, which was then followed by a Gabriel[13] test in order to find specific differences between groups (Tables 5a and 5b).

**Table 5a.** Comparison of comprehension means (ANOVA)

| | Sum of Squares | df | Mean Square | F | Sig. |
|---|---|---|---|---|---|
| Between Groups | 92.621 | 3 | 30.874 | 12.022 | .000 |
| Within Groups | 118.135 | 46 | 2.568 | | |
| Total | 210.755 | 49 | | | |

**Table 5b.** Comparison of comprehension means (Gabriel)

| Groups | compared | Mean Difference | Std. Error | Sig. | 95% Confidence Interval | |
|---|---|---|---|---|---|---|
| | | | | | Lower Bound | Upper Bound |

[13] Levene's test showed that for both reading and writing, there was no significant difference across the variances. The Gabriel test was thus considered to be the best post-hoc test option, especially due to the fact that the groups were of slightly different sizes.

| | | | | | | |
|---|---|---|---|---|---|---|
| Cont | 1Ex | -2.01429* | .64568 | .018 | -3.7820 | -.2466 |
| | Def | -3.76667* | .66894 | .000 | -5.6009 | -1.9324 |
| | 3Ex | -3.06154* | .65652 | .000 | -4.8606 | -1.2625 |
| Def | 1Ex | 1.75238* | .63044 | .045 | .0246 | 3.4802 |
| | Cont | 3.76667* | .66894 | .000 | 1.9324 | 5.6009 |
| | 3Ex | .70513 | .64153 | .847 | -1.0541 | 2.4643 |
| 1Ex | Cont | 2.01429* | .64568 | .018 | .2466 | 3.7820 |
| | Def | -1.75238* | .63044 | .045 | -3.4802 | -.0246 |
| | 3Ex | -1.04725 | .61724 | .444 | -2.7399 | .6454 |
| 3Ex | 1Ex | 1.04725 | .61724 | .444 | -.6454 | 2.7399 |
| | Cont | 3.06154* | .65652 | .000 | 1.2625 | 4.8606 |
| | Def | -.70513 | .64153 | .847 | -2.4643 | 1.0541 |

*Mean difference is significant at the $p<.05$ level.

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, it can be seen that all three experimental conditions had a significantly positive effect on comprehension when compared with the Control group. It can also be seen that the difference between the effect of definitions and the effect of single corpus examples was significant, but the difference between definitions and multiple corpus examples was not. Therefore, it can be concluded that whereas dictionary definitions are likely to help learners understand words that are not familiar to them more than a single corpus example, multiple corpus examples seem to help as much as definitions, exactly as was predicted in Table 1.

To answer the language production research questions, an ANOVA followed by a Gabriel test was applied to the language production data (Tables 6a and 6b).

**Table 6a.** Comparison of production means (ANOVA)

| | **Sum of Squares** | **df** | **Mean Square** | **F** | **Sig.** |
|---|---|---|---|---|---|
| Between Groups | 108.807 | 3 | 36.269 | 6.478 | .001 |
| Within Groups | 257.553 | 46 | 5.559 | | |
| Total | 366.360 | 49 | | | |

**Table 6b.** Comparison of production means (Gabriel)

| Groups | compared | Mean Difference | Std. Error | Sig. | 95% Confidence Interval | |
|---|---|---|---|---|---|---|
| | | | | | Lower Bound | Upper Bound |
| Cont | 1Ex | -1.90260 | .95338 | .265 | -4.5127 | .7075 |
| | Def | -1.32045 | .98771 | .699 | -4.0288 | 1.3879 |
| | 3Ex | -4.13776* | .96938 | .001 | -6.7942 | -1.4814 |
| Def | 1Ex | -.58214 | .93086 | .989 | -3.1333 | 1.9691 |
| | Cont | 1.32045 | .98771 | .699 | -1.3879 | 4.0288 |
| | 3Ex | -2.81731* | .94725 | .027 | -5.4148 | -.2198 |
| 1Ex | Cont | 1.90260 | .95338 | .265 | -.7075 | 4.5127 |
| | Def | .58214 | .93086 | .989 | -1.9691 | 3.1333 |
| | 3Ex | -2.23516 | .91138 | .101 | -4.7344 | .2641 |
| 3Ex | 1Ex | 2.23516 | .91138 | .101 | -.2641 | 4,7344 |
| | Cont | 4.13776* | .96938 | .001 | 1.4814 | 6.7942 |
| | Def | 2.81731* | .94725 | .027 | .2198 | 5.4148 |

*Mean difference is significant at the $p<.05$ level.

The ANOVA for the production scores in Table 6a shows that it is unlikely that the differences between the four groups were obtained merely by chance. The results in Table 6b indicate that the only experimental condition that had a significantly positive effect in language production was presenting learners with multiple corpus examples: the difference between multiple corpus examples and no reference materials was highly significant, at the 99.9% level.

The language production scores by the Definitions and the 1Ex groups were in neither case significantly different from those by the Control group. Thus contrary to what was predicted in Table 1, there is no evidence in support of the hypotheses that a single corpus example or that definitions exhibiting target collocation/colligation will help learners in language production.

The results in Table 5b also show that the differences between single and multiple corpus examples in language production were not significant. However, multiple corpus examples helped significantly more than definitions exhibiting target syntax and collocation.

## 4. Discussion

*4.1 Comprehension findings*

The results presented in the previous section indicate that all three experimental conditions – definitions, a single corpus example and multiple corpus examples – helped language comprehension, just as in the 2012 study. The comprehension results of the two experiments were remarkably similar, even though the subjects in the present study were younger students attending secondary school with fewer years of English instruction than the undergraduates in the 2012 study.

The earlier studies discussed in the introduction (Summers 1988; Laufer 1993; Al-Ajmi 2008) did not make a clear distinction between (a) examples containing contextual clues to facilitate comprehension, and (b) examples whose main function was to illustrate collocation/colligation, and were not able to show any evidence that examples assisted comprehension. However, when examples provide contextual clues that allow learners to infer meaning, both the 2012 and the present study indicate that one example can help a little, and three examples can help as much as definitions.

Having said this, examples and definitions should not be regarded as mutually exclusive options. Indeed, learners' dictionaries often provide a combination of the two. However, the examples they give do not always reinforce comprehension. The combination that should work best for language comprehension is a definition plus examples that specifically contain contextual clues to facilitate understanding, rather than examples whose main function is to illustrate collocation and colligation. I would therefore like to emphasize the need for learners' dictionaries to distinguish between examples for language production and examples for language comprehension and, in the case of less frequent words, to grant enough room to the latter, since learners are more likely to look them up for comprehension.

*4.2 Production findings*

In terms of language production, the results obtained in the 2012 and the present study are not directly comparable because there were major changes in the research design. As explained in section 2.3, one major difference was that only two of the LDOCE definitions used in the 2012 study exhibited the target collocation/colligation required for the error-correction test, while all definitions in the present study exposed the students the target collocation/colligation. However, contrary to what was expected (see Table 1), it seems that the learners were generally not able to derive any significant benefit from collocation/colligation patterns in definitions.

Also contrary to what was predicted in Table 1, the Ex1 group did not do significantly better than the Control group in the language production test. One reason why the Definition and the Ex1 groups did not do as well as expected in terms of language production may have been that that the students in the 2012 study had been told the sentences given to them contained mistakes and were therefore actively looking for errors and ways to correct them, whereas in the present study the students were not aware that the sentences they were required to produce were error-prone. The students in the present study may have felt they already knew the correct answers to the production test items – which looked quite easy but were actually rather tricky – and may therefore not have consulted the reference sheets provided to them as carefully as the students in the 2012 study, who had been specifically instructed to look out for mistakes. This is an important point to make. As argued in Frankenberg-Garcia (2011) and Ranalli (2013), one of the main reasons why learners are underusers of dictionaries and other language resources is that they are often not aware of their own language limitations and reference needs. There is no reason why learners should consult

dictionaries or other references if they think they already know how to use the language, even if they are wrong about this.

The only experimental condition which had an unequivocal positive effect on language production both in the 2012 study and in the present one was supplying learners with multiple corpus examples. Multiple corpus examples helped both undergraduates and younger, secondary school students. Moreover, in the production test, multiple corpus examples helped both the 2012 students who had been told the sentences contained mistakes and the present students who were not warned that the sentences were prone to errors. Thus multiple corpus examples seem to help not only learners who are consciously looking up linguistic information, but also learners who are not necessarily aware that their assumptions about language might be wrong.

Obviously, it is not just any set of multiple corpus examples that will help. All the examples in both the 2012 and the present study exhibited collocation and colligation patterns that were comparable to those required in the production tests. In contrast, the examples used in previous experiments (cf. Summers 1988; Laufer 1993; Nesi 1996) and the examples available in existing learners' dictionaries may not address a number of specific language production difficulties.

Another limitation of learners' dictionaries is that exemplification is not usually repeated. The importance of reiterated exposure to ways in which words are used cannot be overemphasized. This is what data-driven learning, as proposed by Johns (1991), is about. It is only when we have access to a number of examples of the same kind that we are able to detect conventional patterns of use.

The question that remains to be answered is how to provide learners with easy access to multiple corpus examples that focus on the patterns of language that they need to look up for language production. Both the 2012 and the present study were carried

out in controlled experimental conditions, where the learners were essentially spoon-fed with the right type of examples. In the real world, however, learners would need to discover those examples by themselves. One solution would be for learners to use corpora directly in order to look up whatever patterns of use they felt the need to explore. However, as pointed out in Frankenberg-Garcia (2012b), concordancing programs are not particularly user-friendly and raw corpus data is far more difficult to understand than the edited materials language learners are accustomed to using. With the technological advances that we see today, however, there is no reason why we should not attempt to fill in the existing gap that lies between the polished, albeit limited, linguistic information neatly systematized in dictionaries and the countless other linguistic facts that can be gleaned from corpora, but which only experienced corpus users are able to access.

We know from previous studies such as Bogaards and van der Kloot (2002), Dziemianko (2006) and Chan (2012) that examples in dictionaries are a favourite source of information about collocation and colligation. We therefore need to capitalize on that and invest in dictionaries that provide learners with better and more examples. Although the CD-ROM versions of many modern English learners' dictionaries already provide learners with extra corpus examples, these additional examples are either not sorted at all, and can thus be as hard to interpret as raw corpus data, or they are only sorted for meaning, i.e. according to the different senses of polysemous words. To help with language production, however, we need to ensure that examples are also sorted according to different lexico-grammatical patterns of use, and that there are enough examples of each type so as to provide learners with repeated exposure to such patterns. This is especially important in the case of high-frequency words, since these are the words that learners are required to use actively in everyday communication.

## 5. Conclusion

The idea of encouraging learners to arrive at their own conclusions about language as a result of exposure to corpus data has been under discussion for more than twenty years. However, there are not many empirical studies available that support Johns' (1991) model of data-driven learning, and the few data-driven-learning experiments described in the literature tend to be isolated pieces of research conducted in the restricted environment of higher education. In the present study, I aimed to test the robustness and generalizability of my previous findings (Frankenberg-Garcia 2012a) – where the participants were undergraduates – by conducting an analogous experiment with secondary school students and a few non-trivial adjustments in the research design.

The two studies showed remarkably similar results with regard to the effects of corpus examples upon language comprehension: one example was significantly helpful, and multiple corpus examples helped as much as dictionary definitions. The results obtained for language production were slightly different, which, as discussed earlier, may be have been because of the different ways in which language production was tested in the two studies. While in the 2012 study a single corpus example was significantly helpful, in the present study a single corpus example was not enough to reach accepted levels of significance. Ensuring that the definitions presented to learners exhibited target collocation/colligation was even less helpful for language production. Only multiple corpus examples were significantly beneficial in both studies, irrespective of the data collection method used and the subjects involved.

Of course, more replication studies are still necessary. In particular, it would be important to test how learners from different cultures, particularly non-western students from non Indo-European language backgrounds, react to corpus examples. It would also

be worth exploring the long-term effects of presenting learners with multiple corpus examples. Although learners are generally familiar with dictionaries, they are not usually accustomed to having to figure out how language works by observing corpus data, so it would be important to test whether the ability to profit from corpus examples can improve over time. Finally, there is the question of how many corpus examples are necessary. In both the 2012 study and the present one, the students using corpus examples were supplied with either just one or exactly three examples. Further research is needed to explore whether there is an optimum number of examples that will help learners.

**References**


Abbuhl, R. (2012) Why, when, and how to replicate research. In Mackey, A. and Gass, S. M. (eds.), *Research methods in second language acquisition: A practical guide.* Oxford: Blackwell, 296-312.

Al-Ajmi, H. (2008) The effectiveness of dictionary examples in decoding: The case of Kuwaiti learners of English. *Lexikos*, **18**: 15-26.

Bick, E. (2009) DeepDict: A graphical corpus-based dictionary of word relations. *Proceedings of NODALIDA 2009*. Tartu: Tartu University Library, 268-271.

Bogaards, P. and van der Kloot, W. (2002) Verb constructions in learners' dictionaries. In Braasch, A. and Povlsen, C. (eds.), *Proceedings of the tenth Euralex international congress*. Copenhagen: CST, 747-757.

Boulton, A. (2010) Learning outcomes from corpus consultation. In Moreno Jaén, M., Serrano Valverde, F. & Calzada Pérez, M. (eds.), *Exploring new paths in language pedagogy: Lexis and corpus-based language teaching*. London: Equinox, 129-144.

Chan, A. (2012) Cantonese ESL learners' use of grammatical information in a monolingual dictionary for determining the correct use of a target word. *International Journal of Lexicography,* **25**(1): 68-94.

Davies, M. (2008) *The Corpus of Contemporary American English (COCA): 410+ million words, 1990-present*. http://www.americancorpus.org

Dziemianko, A. (2006) *User-friendliness of verb syntax in pedagogical dictionaries of English*. Tübingen: Max Niemeyer Verlag.

Frankenberg-Garcia, A. (2011) Beyond L1-L2 equivalents: Where do users of English as a foreign language turn for help? *International Journal of Lexicography,* **24**(1): 97-123.

Frankenberg-Garcia, A. (2012a) Learners' use of corpus examples. *International Journal of Lexicography,* **25**(3): 273-296.

Frankenberg-Garcia, A. (2012b) Raising teachers' awareness of corpora. *Language Teaching,* **45**(4): 475-489.

Gast, D. (2010) Replication. In Gast, D. (ed.), *Single subject research methodology in behavioral sciences*. New York: Routledge, 110-128.

Hanks, P. (1987) Definitions and explanations. In Sinclair, J. (ed.), *Looking up: An account of the COBUILD project in lexical computing*. London/Glasgow: Collins ELT, 116-136.

Humblé, P. (2001) *Dictionaries and language learners*. Frankfurt: Haag & Herchen.

Izumi, S. (2002) Output, input enhancement, and the noticing hypothesis: An experimental study on ESL relativization. *Studies in Second Language Acquisition,* **24**: 541-577.

Johns, T. (1991) Should you be persuaded: Two examples of data-driven learning. In Johns, T. and King, P. (eds.), *Classroom concordancing*. *English Language Research Journal*, **4**: 1-16.

Kilgarriff, A., Husák, M., McAdam, K., Rundell, M. and Rychly, P. (2008) GDEX: Automatically finding good dictionary examples in a corpus. *Proceedings of the 13th Euralex international congress*. Barcelona: Universitat Pompeu Fabra, 425-432.

Kilgarriff, A., Jakubíček, M., Pomikalek, J. and Whitelock, P. (2012) Setting up for corpus lexicography. Paper presented at the *15th EURALEX international congress*. University of Oslo, 7-11 August.

Laufer, B. (1993) The effects of dictionary definitions and examples on the comprehension of new L2 words. *Cahiers de Lexicologie*, **63**: 131-142.

Leow, R. (2000) A study of the role of awareness in foreign language behavior. *Studies in Second Language Acquisition*, **22**: 557-584.

Mackey, A. (2006) Feedback, noticing and instructed second language learning. *Applied Linguistics*, **27**: 405-530.

Nesi, H. (1996) The role of illustrative examples in productive dictionary use. *Dictionaries*, **17**: 198-206.

Porte, G. (2012) Introduction. In Porte, G. (ed.), *Replication research in applied linguistics.* Cambridge: Cambridge University Press, 1-19.

Ranalli, J. (2013) The online strategy instruction of integrated dictionary skills and language awareness. *Language Learning & Technology*, **17**(2): 75-99.

Schmidt, R. (1990) The role of consciousness in second language learning. *Applied Linguistics*, **11**: 129-58.

Summers, D. (1988) The role of dictionaries in language learning. In Carter, R. and McCarthy, M. (eds.), *Vocabulary and language teaching*. London: Longman, 111-125.

Summers, D. (ed.) (1996/2003) *Longman dictionary of contemporary English* (3rd and 4th editions). Harlow: Longman.

Van Patten, B. (1994) Explicit instruction and input processing. *Studies in Second Language Acquisition*, **15**: 225-241.

## Appendix I. Definitions

**Question 1.** When you **vote** for someone, you indicate you choose that person in an election or competition.*

**Question 2.** If you say that one thing **depends** on another, you mean that the first thing will be affected or determined by the second.**

**Question 3.** If you **call** a person at a particular number, you dial that number to telephone them.*

**Question 4.** If one thing **matches** another, or if two things match, they look attractive together because they are a similar colour, pattern etc.***

**Question 5.** If you wear a **shirt**, you have it on your body.*

**Question 6.** When you **pay** for something, you give money to someone because you are buying that thing.*

**Question 7.** If you **telephone** someone or somewhere, you dial their telephone number**.***

**Question 8.** If you **approve** of an action, event, or suggestion, you like it or are pleased about it.**

**Question 9.** If something is a certain number of **metres** long, it measures that number of metres from one end to the other.*

**Question 10.** If you are **good** at something, you are skilful and successful at doing it.**

**Question 11.** **Grime** is dirt which has collected on the surface of something.**

**Question 12.** If you **loiter** somewhere, you remain there or walk up and down without any real purpose.**

**Question 13.** **Silage** is food for cattle that is made by cutting a crop such as grass or corn when it is green and then keeping it covered. **

**Question 14.** If someone **feigns** a particular feeling, attitude, or physical condition, they try to make other people think that they have it or are experiencing it, although this is not true.**

**Question 15.** A **portly** person, especially a man, is rather fat.**

**Question 16.** If you **mend** something that is broken or not working, you repair it, so that it works properly or can be used. **

**Question 17.** If you **retch**, your stomach moves as if you are vomiting. **

**Question 18.** If something **slackens** or if you **slacken** it, it becomes slower, less active, or less intense.**

**Question 19.** If someone's expression or appearance is **brooding**, they look as if they are thinking deeply and seriously about something, especially something that is making them unhappy**. ****

**Question 20.** A **denizen** of a particular place is a person, animal, or plant that lives or grows in this place.**

* customized definition
** COBUILD online
*** LDOCE, fourth edition

## Appendix II. Corpus examples

**Question 1**
I tried to **vote** for you today, but they told us we couldn't vote any more.
You **voted** for the president's healthcare plan.
Are they **voting** for the other guy?

**Question 2**
So it **depends** on where you come from.
Everything's going to **depend** on the next fighting season.
He **depended** on her so much.

**Question 3**
**Call** him directly at 214-555-049 to give him your response.
Give us a **call** at 800-989-8255.
For information or directions, **call** the church at 281-297-5900.

**Question 4**
His inflamed nose **matched** the color of his brick-red hair.
The carpets were also green and **matched** the walls and curtains.
Should a cake **match** the theme of a wedding?

**Question 5**
I wore a blue long-sleeved **shirt** with overalls.
She was wearing pink **shirt** and blue jeans.
You can't wear that **shirt**, Tommy.

**Question 6**
Peggy helped **pay** for the honeymoon.
And they actually **paid** for the guns?
Nicole sold photos to **pay** for the funeral.

**Question 7**
I must **telephone** the police.
She **telephoned** the office on a Sunday night using the after-hours number.
Bea **telephoned** the hospital and asked the nurse if Corren Jones had gone home.

**Question 8**
Her great-grandmother would not only **approve** of Julia's plan but applaud it.
He **approves** of her every action and would do anything for her.
She smiled at him, as if she **approved** of what he said, and then she let herself out the door.

**Question 9**
The tank was about ten metres wide and fifteen **metres** long.
What we have here is two **metres** long and about three centimetres wide.
The Panama Canal lock chambers are 305 **metres** long and 33.5 meters wide.

**Question 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.

**Question 11**
Years of accumulated **grime** had to be removed.
Your first step should be to clean off all the road **grime**, using a good-quality car shampoo.
The windows are covered with **grime**; the place looks as if it has not been occupied for several weeks.

**Question 12**
He had the habit of being seen **loitering** around places where beautiful women had just been strangled.
All day and all night he would be there **loitering** in the streets.

Five or six teenagers **loiter** in front of a newsagent, drinking and smoking.

**Question 13**
The sheep graze the grass in the fields and are fed grass **silage** during the winter.
As we descended a farmer was busy cutting **silage**.
Long winters meant considerable amounts of **silage** and hay had to be conserved.

**Question 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.
Attention-seeking patients who **feign** illness or deliberately fabricate symptoms.

**Question 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.

**Question 16**
We've got glue, we can **mend** the teapot.
I loved this new sweater so much, though, that I decided to **mend** it then and there.
How can you **mend** a broken heart?

**Question 17**
The sour taste in her mouth made her **retch** a couple of times.
At last I stopped **retching** and washed my hands and face with a cloth.
The sour taste in her mouth made her **retch** a couple of times.

**Question 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**.

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

**Question 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.