VOCABULARY DRILLING: A GHOST OF THE PAST – OR A RENEWED OPPORTUNITY?

Brigitte Plomteux
Haute Ecole Libre Mosane (BELGIUM)

Abstract

Vocabulary acquisition is key to progress in foreign language learning. If this statement probably needs no further argumentation, there are multiple questions as to how vocabulary should be taught and (hence hopefully) learned, and to the extent to which the efforts teachers and learners put in vocabulary teaching, learning, practice and testing prove successful in subsequent language production by the student. The gap between learning vocabulary and using it productively is not easy to bridge and assessing vocabulary in open or semi-open student production is a daunting task.

Interest in vocabulary learning in foreign teaching is not new, and a lot of work already has been done on among others description, acquisition and pedagogy [1], assessing vocabulary [2] and maybe more recently on the mental lexicon [3], all of which, and many more, paved the way to better understanding of what vocabulary is and how it could be taught, learned and tested.

On the other hand, the development of online learning possibilities, in this case the MOODLE learning platform, offers new opportunities in exercising vocabulary, with a wide variety of question types, immediate correction, and feedback possibilities.

The aim of this paper is to report on an experiment, covering four years, with students in the last year of a bachelor in foreign trade, in which traditional in-class tests were replaced by online testing of vocabulary, and to measure the extent to which this change in testing method affected the use of vocabulary in students' productions, as well as the students' reactions to the change in method and the impact this new method had on their feeling of proficiency. Attention will be given to detailed analysis of the use of various types of vocabulary, lexical density in students' production, and other relevant aspects.

Keywords: vocabulary, online testing, foreign language, distance learning, neurosciences in education.

1 VOCABULARY, IN A FEW WORDS

1.1 Words and more

It certainly is an impossible task to describe what vocabulary is in a few words. The research on this topic has evolved in such a way over the past few years that even attempting to summarise it would most certainly take dozens of times the space allotted for this paper. I will only mention here the importance, for vocabulary in general, but also more specifically for this study, of moving from individual words to words in collocations, in multi-word units, in idiomatic expressions, in formulas [4]. Only when moving from isolated words towards words in combination with others, in other meanings, in other contexts, can a learner improve comprehension on the receptive side of language knowledge, and fluency on the productive side.

1.2 What do we need to know about a word?

When learning a foreign language, knowing a "word" (and there can be discussion about the meaning of the word "word" itself) is certainly not limited to knowing the / a translation in the mother tongue. If a learner is to be able to use a word correctly, s/he also needs to know the different forms, the valence (what words, adjuncts or other complements go with it), the collocates, idiomatic expressions and more. Testing vocabulary can therefore not be limited to testing translation, and should include knowledge and correct use of multi-word units, contextualisation, register, idiomatic expressions, to name but a few.

It is also important to note that "knowing a word" is, and always will be, for probably every single word in a language, a work in progress: on a first encounter, one will get a rough impression of the meaning, and subsequent encounters will refine this knowledge, adding information on word grammar, collocations, range, register, and idiomatic expressions.
1.3 Neurosciences

From a focus on the structure of vocabulary (word families, derivations, collocations, prefixes, suffixes, for example), the last research in cognitive science seems to have shifted to the study of the brain. The newest technology of medical imaging, eye-tracking and other technical evolutions allow us to penetrate the brain, thus unveiling its mysteries and telling us more about how vocabulary is stored and retrieved, and will certainly offer new possibilities for teaching and testing. But to move forth in this direction, still more needs to be known about this field of research [5] in order to make use of the newest discoveries in vocabulary teaching and testing materials. Moreover, learning theories are also evolving fast, building on those discoveries, thus enticing teachers to explore new paths to teach and test [6]. In times when so many devices make it possible to externalize our memory, it is interesting to note that there is a current in neurosciences pleading for more memory practice, which especially for vocabulary seems to make much sense [7].

2 WHAT MOODLE MAKES POSSIBLE

Vocabulary testing is probably very old, and is still part of many language exams, be they official exams or the ones organized by schools that offer language classes. It can also be easy for a teacher to test and mark, if the simple option is taken of giving students a few words to translate from their mother tongue to the foreign language, or the other way around.

But we've come a long way from assessing vocabulary knowledge by asking students to translate isolated words, with at best the context of the current lesson to rely on. Not only do we have a better theoretical view of what vocabulary is and how it is, or supposedly is, stored in the brain, but the development of learning platforms with their testing tools have also dramatically changed the way a teacher can test vocabulary. There are two aspects that are relevant for this study: the ease of use, as well for teacher as for student, and the wide choice of possible question types that can be chosen from to test vocabulary knowledge.

Taking precious classroom time for vocabulary testing should really be a thing from the past: in this experiment, students were asked to perform the tests outside the classroom, with an unlimited number of attempts, and only the best result of each of the different tests being taken into account for their final grading, and for a reduction of the number of texts they had to read for the oral exam – students need these incentives if we want them to do their best. I am fully aware that authorizing students to do the tests at home makes it possible for them to use all resources available, from the course book to online resources to help from friends or people who are fluent in the language, but with an unlimited number of attempts possible, they should understand that it's in their advantage to take this opportunity to monitor their progress. They also perfectly know that there is an upcoming exam, and then they will have to rely only on their own personal knowledge to get passing marks.

2.1 Organisational ease with MOODLE

Enumerating the practical advantages of using an online Learning Management System (LMS), in this case MOODLE, should not lead us to believe it's an easy matter to put a testing system in place. The teacher has to create a huge database of questions if the system of unlimited number of attempts is to work: if students keep having the same questions in their second, third, or tenth try, there won't be enough variation for learning to take place. This means that the teacher may be relieved of correction work or other practical details such as providing copies of the paper tests, or overseeing students taking their exam to avoid cheating, but the workload has not diminished – it has only shifted to a more creative part, i.e. creating the questions and then the tests.

This paper doesn't aim to describe in detail how to use MOODLE for vocabulary testing. Suffice it to say that creating a test, with all its settings (opening and closing time, password, duration, marking, feedback) is quite straightforward, and adding questions to the question database, although time consuming, is very easy to do. The more important part is creating the questions: doing that requires mastery of the question tool of the LMS, but more importantly a clear vision of what one wants to test, and how to do this. Using ICT (Information and Communication Technology) certainly doesn't make the teacher's work redundant, on the opposite: the teacher still, or even more than before, needs to have a thorough knowledge of vocabulary theory to create relevant questions.
2.2 Interesting question types

Of course, when using online testing, one is limited to the question types provided by the platform, and that may imply a change in the way we formulate our questions. But many question types are suited for vocabulary testing, provided one keeps in mind the fact that correction by the computer may relieve the teacher from this often tedious work, but possible frequent mistakes should be anticipated, and providing feedback is part of the learning side of this testing scheme.

2.2.1 The classical question types

Fill in the blanks, match beginning and end of sentences, or adjectives with nouns, cloze questions, and multiple-choice questions can probably be considered as obvious question types for vocabulary testing. But even the obvious can be stretched further than what seems possible at first sight. Multiple choice questions, for example, are not limited to "what is the missing word?" but can also be used to have a student choose which is the best ending for a sentence (a way of testing collocations, for example), or to ask for the best translation of a sentence, or which word should be associated to a prompt, be it for collocations, associations (in meaning, "synonyms", hypernyms and more), register, and still other possibilities.

2.2.2 Less obvious question types

Some question types may at first sight seem less suited for testing vocabulary, but should be given second thoughts, as they can provide for variation and another entry in the knowledge of the word. "True-false" questions can be used to check understanding of a sentence from a vocabulary point of view. If the sentence includes a form of definition of the word to be tested, the correct answer can be an indication of some form of knowledge of the word in question. My favorite example is "someone who has been beheaded feels better afterwards": without understanding the verb "to behead", it's impossible to answer the question (unless one would want to answer it on a philosophical level, in which case it's not vocabulary understanding that is being tested and this goes far beyond the scope of this study).

Cutting a short text in parts, not at random but between elements of a collocation for example, and requiring students to put the parts in the correct order, can also give an indication of an existing connection in the student's brain between the elements that have been separated for the purposes of the exercise. By separating elements that normally go together, we create a want for bringing them back together, thus ensuring they are stored together in the brain and the link is reinforced – or so we hope.

3 THE STUDY: VOCABULARY TESTING FOR THIRD YEAR BACHELOR STUDENTS IN FOREIGN TRADE

3.1 The research question

Students learning a foreign language don't do so in order to get good marks in vocabulary exercises: depending on their study field and other factors, they want to be able to use the language, in what is commonly called the four skills: reading, listening, writing and speaking.

The aim of the online vocabulary testing thus is not to create better test-takers, but to help students use the vocabulary in their own production. It could, and can also, of course, be used receptively (and many students reported a better comprehension of texts after having taken the vocabulary tests), but in this case, the analysed data were written dialogues in the students' final exam. And the central question is the following: would the students, after having done the online tests, use more and better vocabulary in their production?

3.2 Context

The study on which this paper is based spans four years, and the essential information necessary to understand its evolution can be summarized in the following table:
The third-year Dutch course is based on a ten-episodes video story. In the year N-1, students would be tested in class, with tests after episodes 2, 4, 6 and 8 (episodes 9 and 10 were too close to the exam time to allow for more tests). From the year N, I shifted to online testing: limited classroom time, the urge to put the available time at use for more oral practice, for which teacher presence is or can be vital, ease of organization, and above all the possibility to differentiate according to students needs: as they all have different levels of proficiency and learning speeds, online testing offers them the possibility, to a certain degree, to pace their learning according to their needs and abilities.

Until the year N – 1, the final exam would include an oral exam (not relevant for this study, except maybe for the incentive), and a written exam. The written exam consisted of vocabulary questions, grammar questions, translation (from French – the language of education, and most students' mother tongue - to Dutch, based on the situations and dialogues of the video course), and essay questions. The essay questions included writing a dialogue using the characters of the video in a situation similar to those of the video, and discussion questions.

From the year N onwards, the vocabulary part of the exam was done online, using the same question database as for the compulsory tests. The database consists of more than 2500 different questions, so even students who had tried the tests several times would probably not have encountered all the questions available. Moreover, MOODLE offering the possibility to shuffle the elements of questions (drag and drop, multiple choice) and the questions in a test, the same question appearing for the second or third time in a student's subsequent attempts will possibly not be recognised as having been encountered already.

The students do the online part of the exam, and then move on to the grammar, translation and dialogue and essay writing part of the exam. That means that from the year N onwards, they didn’t any longer have the vocabulary exercises visible on the question paper, meaning they couldn’t just use the vocabulary that was on one of the previous pages. Unfortunately, it was impossible to account for this difference between year N – 1 on the one hand and years N, N + 1 and N + 2 on the other hand, for practical reasons.

The data analysed for this study were the dialogues written by the students in their final exam. It is important to remember that what we are trying to analyse is the difference in vocabulary use in the dialogue question. The students have not had other opportunities to practice writing a dialogue, so if progress is made from year N – 1 to subsequent years, one logical explanation would be that the online testing improved this skill without even practicing it, but by exposing students to the vocabulary they need or might need, and having them interact with it by replying to the questions.
3.3 Method

The vocabulary list available to students in their coursebook was put in an Excel file, with various information such as episode, scene, type of vocabulary (collocation, idiomatic expression, binomial, among others), and a tab for each student. For the exam, students were instructed to write a dialogue of about 200 words, with no specific emphasis on using the vocabulary that had been practiced in class and tested online. The dialogues produced by the students in their written exam were then re-read by the teacher, and the used vocabulary was put in the Excel file, using exactly the form written by the students, with possible mistakes, be it in word form, order, context or other aspects. The choice was made not to change a single thing in their writings, not even correcting spelling mistakes that could have been considered to have been caused by distraction or a lack of accuracy. If the choice had been made to correct mistakes, it would have been hard to draw the line between correcting distraction mistakes (how can I know the student was actually distracted and that caused the mistake?) and correcting other mistakes, which would have biased the research results.

Each used item was then given a logarithmic scaled note. The scaled note is explained in the table below:

<table>
<thead>
<tr>
<th>Note</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 000 000</td>
<td>Perfect use of the vocabulary item: context, form</td>
</tr>
<tr>
<td>100 000</td>
<td>Context still correct, but small mistakes (spelling, congruency of adjective, …)</td>
</tr>
<tr>
<td>10 000</td>
<td>Context correct, but bigger mistakes (subject – verb agreement, verb form, …)</td>
</tr>
<tr>
<td>1000</td>
<td>Item grammatically correct, but inappropriate context</td>
</tr>
<tr>
<td>100</td>
<td>Inappropriate context with small mistakes</td>
</tr>
<tr>
<td>10</td>
<td>Inappropriate context with big mistakes</td>
</tr>
</tbody>
</table>

As any teacher who has had to grade essays knows, it is practically impossible to classify all mistakes according to as simple a table as the above. This means that choices had to be made as to which note to choose, and some of these choices could probably be questioned.

This method has other drawbacks. It was very time-consuming, as a total of over 200 essays had to be re-read and the used vocabulary put in the excel file. This work could not be done in a short time span, which means that some slight differences might have occurred between beginning and end of this part of the work. Furthermore, as we were not working with a "one word – one entry" vocabulary list, it was sometimes difficult to decide where to write the form. Decisions had to be taken one item at a time. As the number of students involved in the experiment is rather important, we can hope that the influence of these decisions doesn't have too big an influence on final results.

4 RESULTS

4.1 Bell curve for number of items

Once we had all the data in the excel file, we could finally analyse them statistically. Fig. 1 below shows the bell curve for the four years, according to the number of items used by each student. The X axis gives the number of items used, ranging from 0 to 40, and the Y axis indicates density of probability: what is the chance of getting values near the corresponding points on the X axis?
On observing the graph, there is a noticeable difference in the mean for year N – 1 on the one side, and the other three years on the other one. A first rough estimation thus could lead to the conclusion that, indeed, online testing gave better results than the classical tests used before. But the means are only one part of the story.

It becomes even more interesting when we look at the standard deviation. Table 3 gives standard deviation for the four years.

<table>
<thead>
<tr>
<th>Year</th>
<th>Standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>N – 1</td>
<td>7.21</td>
</tr>
<tr>
<td>N</td>
<td>7.33</td>
</tr>
<tr>
<td>N + 1</td>
<td>7.49</td>
</tr>
<tr>
<td>N + 2</td>
<td>6.58</td>
</tr>
</tbody>
</table>

Values in standard deviation are very similar for the first three years, but there is a noticeable decrease in year N + 2. One explanation is that the students in year N + 2 had had, in their second year, 22 compulsory online tests, and had thus had, in comparison to the three previous years, more opportunity to practice (and also, maybe, a better mastery of the system of online testing).

Variance also evolved in the same direction: from 50.8 in year N – 1 to 52.79 in year N, then 55.36 in year N + 1 to drop to 42.46 in year N +2. Again, the same explanation can be given: the effect of more online testing.

The linear for year N + 2 also approaches the horizontal, which is the result to aim for if we want a majority, if not all, students to get the best score possible. A too large deviation from the mean can be interpreted as good students doing even better, when the weaker ones hopelessly lag behind. This is not what I as a teacher aim for.

### 4.2 Bell curve for logarithmic scale

As said above, data were put in the excel file, and a logarithmic score was given to each item used. Fig. 2 below shows the bell curve for these logarithmic scores.
The curves are similar to the ones for the number of items, with the same conclusion that can be drawn from them.

5 TO BE CONTINUED

Other data has not been analysed yet, due to lack of time, and could for the same reason not be treated in this paper. But it would probably be worth it to have a closer look at other aspects, such as more detailed analysis of the structure of vocabulary used. Especially for the year N + 2, which was added to this study because those students had not only had, as those of the years N and N + 1, eight compulsory tests in their third year, but also 22 tests in their second year, not on the same vocabulary, but on other topics covered in the second-year course. A look on the graph shows variance to be reduced in the results of that year, with average still at a level significantly above the pre-online-testing group. The hypothesis is that those students, having been drilled more than their predecessors, might have had more opportunity to raise their general level, which would have enabled them to obtain better results. But for the time being, this is still in the realm of the unverified hypotheses.

5.1 Lexical density

Lexical density can be calculated by dividing the number of different words in a text by the total number of words in the text. The higher the figure, the higher the lexical density. In a short text (200 words for a dialogue), this indication is maybe not very relevant, especially since in a dialogue one would encounter more repeated words, and that has a negative influence on lexical density. It still would be interesting to investigate this aspect further, in a next step in the study.

Although this may, for a dialogue, be less relevant than for an essay, other measures can be taken as well, such as sentence length. The longer the sentence, the more complex the language use.

5.2 Phrasal verbs

Dutch has in its verb categories an equivalent to English phrasal verbs, be it that they don't behave in the same way. These verbs are important in vocabulary, as the particle plays a vital role in meaning. Students of the year N + 2 have been tested on some of these verbs in their second year. It would according to me be worth looking into their use of those verbs in their dialogue: did they on average use more of these verbs in comparison to the students of the three other studied years?

6 ONE LAST WORD

This study, spanning four years and more or less 200 students and the dialogues they wrote for their exams, has yielded more data than could be analysed in this paper.
Some of the results are quite encouraging, and go beyond the simple popular wisdom of "the more you practice something, the better you get at it". Indeed, as we analysed not the results at the tests, but the production of a dialogue, a transfer from the exercises to the writing of the dialogue is visible.

6.1 For the teacher

The teacher's work in this system shifts towards the creation of online questions and tests. In order to be able to do this, the teacher needs to have a clear view of what vocabulary knowledge means, and how it can best be attained. Creating online exercises is only (but it is of course important) an easier, more up-to-date way of attaining the same goal, i.e. vocabulary acquisition, and more importantly transfer to use in own language production.

The work shift, from paper work and time spent overseeing students taking their tests, is an interesting one, as a more important proportion of time spent on this part of the teacher's job is used for productive purposes. But only an experienced teacher, with a thorough knowledge of the subject of vocabulary, can according to me create exercises that will help students master vocabulary in a way that enables them to use it correctly – whatever that may mean. Associating "words" with their translation is far from being sufficient, student needs other approaches, such as associations with hypernyms, collocates, expressions, that will give them a broader vision of how they can use the vocabulary they learn.

6.2 For the student

According to a survey taken after year N, students globally appreciated the online testing system, giving ease of use, feedback, and possibility of trying as often as they could as positive elements of this system. An overwhelming majority of them (98 %) voted, anonymously, for the continuation of this system.

As for the results on the long run, these cannot be measured objectively, but if we go back to fig. 1 and fig. 2 we can hope the effect of those tests will be lasting.

6.3 Learning

How can we know if learning actually took place? Do the statistics developed for this study show learning was improved in some way? Is the gain in vocabulary use worth the time and energy put in this system of online tests?

Again, the statistics show an improve between year N – 1 and subsequent years, which can be seen as a positive outcome. But measuring learning remains a difficult if not impossible task.

One fact though deserves to be mentioned. In their third year, students should be able to write the dialogue based on the course they have had, but that was not always the case.

In year N-1, 7 out of the 41 students didn't even try to write the dialogue. This number has fallen to respectively 0, 3 (in the year with the highest total number of students – 65) and 1 in the three following years. Even if we can question the gain between writing nothing and writing a dialogue that may have been full of mistakes, it probably was a wide gap to bridge for those students: their feeling of competence seems to have grown from doing the online exercises.

6.4 A ghost from the past – or a renewed opportunity?

As a final conclusion, let's go back to the question asked in the title of this paper. The narrow view of what drilling means – usually very repetitive, maybe boring rehearsal of items, lists, or other, has given way to a more open, up-to-date through the use of a learning platform, and varied way of somehow coercing students (with incentives for good results) into doing those tests and hence to learn their vocabulary. This is, according to me, clearly a renewed opportunity, and we should go along this path in order to help students enrich their vocabulary - thereby alluding to the Dutch word for vocabulary, "woordenschat", which can be literally translated as "treasure of words".

ACKNOWLEDGEMENTS

I would like to thank Antoine Derobertmasure from Université de Mons, Belgium, for his help along the different steps of this study. His comments and advice have certainly given the study the necessary rigor and structure.
Julien Perrez and Laurent Rasier from Liège University, Belgium, helped me with the organisation of the Excel table to enable easy and efficient statistical treatment. My thanks also go to my colleagues Julien Delhaxhe and Mikaël Degeer for their unvaluable help in the statistical treatment of the collected data. Sandrine Biemar and Izida Khamidoullina also followed the project and supported me through the different stadia. I am also indebted to Philippe Lehette, Anne Dister and Bénédicte Schoonbroodt from HELMo for their support and suggestions. With so many brilliant people to help and guide me, any shortcomings, inaccuracies or even mistakes are mine and I would be the one to be blamed for them.

REFERENCES