What's Wrong with Welsh Adjectives?

Cornelia Tschichold*

Department of English Language and Literature, Swansea University, Swansea, Great Britain

Abstract. Why are some words harder to learn than others? In a long-term CASLR (computer-assisted second language research) study, a vocabulary flashcard program that employs spaced repetition for explicit vocabulary training was used in order to arrive at data on the difficulty of individual words. The vocabulary content of a beginner's Welsh course was periodically entered into the program as one learner progressed through the course and studied vocabulary with the help of the electronic flashcards. The Welsh words were trained both receptively and productively, and in a few cases also as part of a short phrase or sentence. The program automatically collects statistical information for each individual electronic card, including the number of times each card had been seen. Data was collected for an initial period of two years of non-intensive learning, and the resulting statistics for the individual flashcards allow an interesting insight into the very highly variable number of repetitions needed for each word.

Keywords: CASLR, vocabulary, electronic flashcard programs, spaced repetition.

1. Background

Vocabulary flashcard programs have the crucial advantage of providing immediate feedback to the learner (Nagata, 1993). This is more effective than delayed feedback, and is especially suitable for procedural and conceptual knowledge building, including verbal tasks. Immediate feedback is also better for low-achieving learners and for beginners (Shute, 2008). Learners are normally aware of the need to study vocabulary, so there is a natural market for software that promises to help them do just this. However, many commercially available applications are poorly designed and contain poor quality content. The multimedia advantage of CALL and MALL (mobile-assisted language learning) is rarely used well, as any critical look at some of these applications will show. Poor illustrations abound, and mistakes are easily found. Of course it is quite easy to add a good illustration for a noun such as *sheep*,

* Contact author: c.tschichold@swansea.ac.uk

In L. Bradley & S. Thouësny (Eds.), *CALL: Using, Learning, Knowing, EUROCALL Conference, Gothenburg, Sweden, 22-25 August 2012, Proceedings* (pp. 292-295).

but a preposition such as *by* is considerably harder to illustrate well. For this reason, the experiment described here does not contain any pictorial material at all, despite the fact that the software used allows the addition of both picture and sound files. The software used is VTrain (www.vtrain.net), a flexible system based on the Leitner learning principle of spaced repetition.

2. Learning Welsh vocabulary

A long-term, single-subject study on learning Welsh vocabulary was started in 2009. The learner was a complete beginner at the start of the course and had very little contact with Welsh outside of class, despite living in Wales. Welsh is natively spoken by a minority of the population in Wales, and while many road signs are bilingual, English clearly dominates most Welsh inhabitants' lives in almost all respects. The learner entered all the vocabulary contained in the Welsh language course into the VTrain database as the classroom-based course progressed. The taught element of the course was a non-intensive beginners' class of one hour a week (30 weeks per academic year). The course material ("Cymraeg i oedolion" – Welsh for adults) consisted of the course book and CD-ROMs containing audio files of much of the material in the course book. The course material takes a broadly communicative approach to teaching. One lesson in the book typically contains between 20 and 30 items of new vocabulary and was covered in two classroom lessons. In addition to this, the learner normally spent a few short (15-20 minutes) sessions every week working on the vocabulary that had been entered into the VTrain database, with longer breaks over the summer.

The software was set up with 10 'boxes' for the word cards, with a one-day interval from the first to the second box, and roughly doubling the interval length for every subsequent box. These intervals provided the guideline for revision. With the increase in the number of cards in the database, the number of word cards due for revision increased as well, and in the latter half of the training period, there were always cards due or overdue for revision. The boxes were set up to alternate between Welsh to English and English to Welsh questions.

Word cards progress through the system if the user types in the correct answer, or return to the first box if the learner's answer is not correct. Answers are evaluated by simple string matching, thus only recognized to be correct if they are spelled exactly right. If the question is "What is the Welsh for: *some time*?", the correct answer is "rhywbryd". The following forms are not recognized "rywbryd", "rhwybryd", "rhywbrud" and cause the word card to be returned to the first box. In the direction of L1, the difficulties for the learner are somewhat different. To the question "What is the English for: *dechrau*?", only the string "to start" is accepted as correct; the variants "start", "begin", or "to begin" are not recognized as correct and also cause the word to be returned to the first box.

3. Data collection and analysis

After two years, the learner had covered the first 30 lessons of the Welsh course and the VTrain database contained over 900 cards. At this point, the statistical information that is compiled by the system was retrieved. For every flashcard, VTrain records the total number of repetitions, the total numbers of correct and of incorrect answers, and the highest 'box' the flashcard reached. For the present study, all double entries, phrases, and items that were entered in order to practice grammatical aspects were deleted, so for example, the data for *go* was kept, but the card data for *I went* was deleted. This resulted in data from a total of 549 word cards. The averages of the statistics for all cards are given in Table 1.

Table 1. Average number of repetitions across all word classes

Average total number of repetitions	29.13
Average number of correct answers	17.50
Average number of incorrect answers	11.71
Average of highest 'box' reached	7.26

It should be kept in mind that as the system was set up with ten 'boxes', every word that ends up in the last box will have accumulated a minimum of ten repetitions. On the other hand, few cards had actually reached the last box at this point. The 549 single words were then sorted into word classes, with the distribution shown in Table 2.

Table 2. Number of words in each word class

verbs	72
nouns	278
adjectives	89
adverbs	33
numerals	20
other (prep, conj, dem, etc.)	67
TOTAL	549

The number of repetitions was then broken down by word class, resulting in the averages seen in Table 3.

Table 3. Average number of repetitions for individual word classes

	NOUN	VERB	ADJ	ADV	FUNCTION	ALL WORDS
Average total	12.37	39.64	40.30	31.70	39.60	29.13
Average correct	9.34	20.43	20.87	17.91	24.26	17.50
Average wrong	3.08	19.21	19.44	14.09	16.23	11.71
Average highest	6.99	7.66	6.98	6.94	7.39	7.26

This analysis showed that nouns were the easiest word class to learn, with adjectives the hardest. Adjectives are often taught in semantic groups, e.g., colours or in pairs of antonyms, a presentation mode that is not conducive to learning (Tinkham, 1997). The next step involved a closer look at those words that had particularly high numbers of repetitions, as these clearly presented more difficulties to the learner. Analysis of words with 50 or more repetitions showed that certain spelling patterns correlate with increased difficulty as measured by the number of repetitions needed by the learner. Because completely accurate spelling is critical for the program to recognize the learner's answer as correct, it could of course be argued that exact spelling is given far too much weight in this context, and that the learner would ideally be given partial credits for otherwise correct answers.

4. Concluding remarks

One interesting finding is that the spelling of Welsh words seems to present a major obstacle to the beginning learner despite the fact that Welsh is said to have a shallow orthography, which should therefore be relatively unproblematic to acquire. Another conclusion this long-term study suggests is that learners need considerably more repetitions than the figures of five to twelve typically found in the literature on vocabulary acquisition (cf. Nation, 2001). Despite the obvious shortcomings on the part of the software used, the analysis sheds some new light on the complexities of the long-term process of incremental vocabulary learning.

References

Nagata, N. (1993). Intelligent Computer Feedback for Second Language Instruction. *Modern Language Journal, 77*(3), 330-338. doi: 10.1111/j.1540-4781.1993.tb01980.x

Nation, I. S. P. (2001). *Learning Vocabulary in Another Language*. Cambridge: Cambridge University Press.

Shute, V. J. (2008). Focus on Formative Feedback. *Review of Educational Research, 78*(1), 153-189. doi: 10.3102/0034654307313795

Tinkham, T. (1997). The effects of semantic and thematic clustering on the learning of second language vocabulary. *Second Language Research, 13*(2), 138-163. doi: 10.1191/026765897672376469

Published by Research-publishing.net
Dublin, Ireland; Voillans, France
info@research-publishing.net

CALL: Using, Learning, Knowing
EUROCALL Conference, Gothenburg, Sweden
22-25 August 2012, Proceedings
Edited by Linda Bradley and Sylvie Thouësny

Typeset by Research-publishing.net
Cover design: © Raphaël Savina (raphael@savina.net)
Aquarelle reproduced with kind permission from the illustrator: © Sylvi Vigmo (sylvi.vigmo@ped.gu.se)
Fonts used are licensed under a SIL Open Font License

ISBN13: 978-1-908416-03-2 (paperback)
Print on demand (lulu.com)

British Library Cataloguing-in-Publication Data.
A cataloguing record for this book is available from the British Library.

Bibliothèque Nationale de France - Dépôt légal: décembre 2012.