OPEN ACCESS

Vocabulary Learning and Instruction
ISSN 2981-9954
Volume 14, Number 2 (2025)
https://doi.org/10.29140/vli.v14n2.102660

Castledown

"From Zero to Three Hundred" – Intensive Acquisition Techniques for the 300 Most Frequent Content Words in Welsh

Tess Fitzpatrick
Swansea University, Singleton Park, Swansea, UK SA2 8PP
t.fitzpatrick@swansea.ac.uk

Steve Morris
Swansea University, Singleton Park, Swansea, UK SA2 8PP

Abstract

Around twenty thousand adult learners enrol onto Welsh language classes each year, but many do not progress beyond *Mynediad level* (= Entry, CEFR A1). The study reported here explored techniques to maintain early learning momentum, maximise learners' returns for their learning efforts, and accelerate initial vocabulary acquisition. Novice learners of Welsh were challenged to learn 300 Welsh vocabulary items over a period of 50 days. They were provided with physical flash cards and information about learning techniques and were asked to spend ten minutes per day on their learning. Tests administered immediately after and two weeks after completion of the learning period showed that acquisition success varied considerably between participants, and an end-of-study questionnaire found that the more successful learners had used highly systematic approaches to word selection and revision and had frequently used the keyword technique of word learning. Test results also revealed that some target words were substantially less successfully acquired than others, and a list of words ranked by "learnability" was compiled. Outcomes from this study are informing curriculum revision by the National Centre for Learning Welsh, the main provider of Welsh language courses for adults.

An earlier, Welsh language paper reporting this study was published in the journal *Gwerddon* (Fitzpatrick & Morris, 2024).

Keywords: vocabulary learning, flashcards, beginner learners, learnability, learning burden, Welsh/Cymraeg, second language acquisition

Learning efficiency and learner motivation are fundamental challenges across the field of language education. Techniques that can give learners the best return for their learning effort and approaches that support learner momentum and motivation can be valuable in addressing these challenges. With this in mind the study reported here investigated the outcomes for beginner learners of an intensive period of studying Welsh language vocabulary. Adult participants who had no, or negligible, previous knowledge of Welsh were tasked to learn 300 words over a 50-day period. Immediate and delayed post-learning tasks revealed how effective their learning had been, and questionnaire data provided information about learning techniques, attitudes and motivation.

An overview of Welsh language and its use will help to contextualise this study. Wales is a devolved nation within the UK and has two official languages, English and Welsh (*Cymraeg*). In the 2021 census, approximately 538,000 of the 3.1m population of Wales identified as Welsh speakers (Office for National Statistics, 2022). This is a decrease from the 2011 figure of 562,000, and despite Welsh being a compulsory school subject for ages 3-16, the biggest decrease since 2011 was seen in the 5–15 age category, from 40% in 2011 to 34% in 2021[1]. The *Welsh for Adults* (*Dysgu Cymraeg*) learning sector is active, with around twenty thousand learners enrolling on classes each year, mostly through the courses offered by the National Centre for Learning Welsh (NCLW https://learnwelsh.cymru/). Others use apps such as Duolingo (https://www.duolingo.com/) and SaySomethingIn (https://www.saysomethingin.com/). Although most Welsh speakers reside in Wales, there is an established Welsh-speaking community in Patagonia, South America. More generally, migration from Wales has resulted in a sizable Welsh-speaking diaspora, and the increased offering of online courses by NCLW since 2020 has attracted a considerable number of new adult learners from outside Wales. As is the case with many minoritised languages, almost all learners of Welsh are expert users of the majority language – in this case, English. Both English and Welsh are members of the Indo-European family of languages, but whereas the roots of English are in the Germanic languages branch, Welsh is one of six languages in the separate Celtic branch (along with Irish, Scottish Gaelic and Manx, and sister Brythonic languages Breton and Cornish). Features of Welsh that are unfamiliar, and challenging, to English speaking learners include sounds, letters and diacritics that do not occur in English, grammatical gender, word order (Welsh is VSO), and the use of

[1] Note that 23% of children in Wales attend schools where at least half of the subjects are available through the medium of Welsh (Estyn, 2023).

mutated initial consonants according to specific syntactic environments. Those features relating to vocabulary acquisition will be returned to later in this paper.

The study was designed in the context of recent policy objectives and provision reviews in Welsh language education, along with newly released resources. These include:

- the CorCenCC corpus, released in 2020 (Knight, Morris, Fitzpatrick, et al., 2020, https://corcencc.org/). The corpus enables the creation of corpus-informed frequency-based lists of contemporary Welsh language vocabulary. The target words in this study are the most frequent 300 nouns, verbs or adjectives in the CorCenCC corpus (which contains over 11 million words).
- Cymraeg 2050. The stated objective of this policy document is for there to be one million speakers of Welsh by 2050, noting that "the production of more high-quality lexicographical, corpus and terminology resources" will "support learners..." (Welsh Government, 2017). This study exemplifies how new resources, in particular the CorCenCC Welsh language corpus (Knight, Morris, Fitzpatrick, et al., 2020) can be used to develop aids for learners.
- the recommendations of the Rapid Review of the National Centre for Learning Welsh (NCLW) (Morris, 2021) and the Welsh Government's response to those recommendations (2021). These include considerations such as increasing learner numbers, reducing learner drop-out rates, and broadening the remit of NCLW, for example by "sharing resources and good practice with the second language sector in schools" and extending "adult provision" to include 16–18-year-olds. NCLW statistics (Dysgu Cymraeg, 2024) suggest that retention of learners through *Mynediad* (= *Entry*; CEFR A1) and into *Sylfaen* (= *Foundation*; CEFR A2) levels is most problematic. This study focuses on the highest frequency words, thus mapping onto A1 course material.

Accelerated learning of the most frequent 300 words can provide a strong foundation for subsequent learning, and immediate communicative potential. The most frequent words of a language represent a disproportionately high coverage of running text; Zipf's (1936) law holds that there is inverse proportionality between the frequency rank of a word and the number of times it occurs in natural language use. This means that a relatively small number of high frequency words account for a high percentage of text. Kremmel (2016), for example, estimates that the 500 most frequent words of English represent 66% coverage of the entire content of the Corpus of Contemporary American English (COCA; Davies, 2008). The release of the National Corpus of Contemporary Welsh (CorCenCC, Knight, Morris, Fitzpatrick, et al., 2020) has made it possible to make similar calculations for Welsh (Knight, Morris, Tovey-Walsh, et al., 2020). We find that the 500 most frequent words of Welsh represent 71% coverage of the contents of that corpus. Further, and in line with Zipf's law, we find that the 300 most frequent words of Welsh represent 65% coverage of CorCenCC. Because our focus in this study was on learning individual words, it is important to note that many of the highest-ranking words in frequency lists are function words: Kremmel estimates that these represent around 40% of corpus coverage (2016, p. 982). Function words typically gain meaning from context, so as Dang and Webb note "it is more reasonable to incorporate teaching function words with other components of language lessons" (2016, p. 166).

Turning, then, to content words, Kremmel estimates that the 500 most frequent content lemmas in English give 26% coverage of the entire contents of COCA. An examination of CorCenCC reveals that in the Welsh data a similar percentage of coverage is obtained (25%) using the target words from the study reported here, i.e. the most frequent 300 Welsh nouns, verbs and adjectives (see below for more detail on target word selection). In terms of what can be done with knowledge of 300 words, Nation and Crabbe (1991) suggest that 120 words/phrases can constitute a "survival vocabulary" (in English), and Nation notes that with knowledge of 100-450 English word families, the easiest graded readers can be accessed (2022, p. 27).

The findings of Fitzpatrick et al. (2008) indicate that acquiring 300 words is manageable within a relatively short period of time. In that study, a participant with no previous knowledge of Arabic was required to learn 300 Arabic words over 20 days from flashcards; in an immediate post-test, 94% of target word forms were recalled, and for 95%, the meaning was recalled. Elgort and Piasecki (2014) also used flashcards for learning, though the target items in their study were pseudowords. They also report a 94% initial recall rate for their 48 target words, this time after one week of learning. Informed by these findings, the current study set a period of 50 days, with 10 minutes study per day, for the learning of 300 target words. This represents a significant initial boost to the usual estimated rate of lexical acquisition (see Milton & Meara, 1998).

In terms of word learning methods, Nation (2022) considers deliberate learning activities such as use of flashcards to be efficient and effective. Learners need to know several thousand words to become effective communicators, and there are rarely enough class hours to cover them all. It is therefore necessary to equip learners with strategies for autonomous vocabulary learning. Word learning is generally recognised to have multiple stages and aspects (see Nation, 2022; Richards, 1976); flashcards support the initial stage, connecting word meaning to word form. Teachers are encouraged to train learners in the use of strategies, and for flashcard learning this might include word retrieval practice, spaced repetition, and deliberate noticing (for further information see Nation, 2022, p. 407–429; Webb & Nation, 2017, p. 112–115). Nation (2022) and others promote the use of flashcard software, which often builds in features of strategic learning. The study reported here, though, aimed to investigate participant techniques for word learning, and to produce a list of words ranked by "learnability". Because software can regulate techniques, spacing, and repetitions of specific words, physical paper flashcards were used in this study.

As detailed below in the methodology section, study participants were given some learning strategy suggestions along with the task instruction. These focused on three techniques: *using flashcards, keyword technique*[2] and *word family technique*[3], and were informed by research overviews including Webb and Nation (2017), and

[2] The keyword technique integrates creative imagery: "learners attend to phonetic or orthographic features of the target item, link these to a familiar 'keyword', usually in the L1, and create a mental image of the link" (Fitzpatrick et al., 2018, p. 34). For example, an English speaker learning the Welsh word *ci* (= dog) might create a mental image of a dog opening a door with a *key*. See also Beaton et al. (1995); Dolean (2014).

[3] The word family technique entails focus on the constituent parts of a word, and deliberate learning of stems and affixes. See Webb and Nation (2017); Bauer and Nation (1993).

Fitzpatrick et al. (2018) which identified the keyword technique as particularly effective in word learning. This simple menu of effective learning techniques provided participants with a toolkit from which they could select approaches appropriate for specific target words and for their own learning preferences.

Words differ in terms of their 'learning burden' (Nation, 2022) and the ease with which they can be acquired. Having established the value of acquiring the 300 most frequent content words, it should be recognised that some will likely require more strategic effort, time, and classroom input than others. By testing learners' capacity to recall the form and/or the meaning of target words at the end of the learning period, any harder-to-acquire words can be identified, and flagged to learners and teachers. The creation of pedagogical wordlists ranked by learnability has received relatively little attention to date in second language research (though see, for example, Ellis & Beaton, 1993; Schmitt et al., 2021). Findings from this study will indicate whether this is likely to be a fruitful area for further investigation and application.

Study Aims

The study had four key aims:

i. To rank the 300 most frequent content words in Welsh according to "learnability"[4], in terms of both productive, or form-recall knowledge (English>Welsh) and receptive, or meaning-recall knowledge[5] (Welsh>English).
ii. To identify the high frequency words that are the most challenging to learn, and the features (e.g. orthographic, semantic, morphological) that might contribute to that learning burden.
iii. To determine whether knowledge of some words is more resilient to attrition (over a two-week period) than others.
iv. To identify effective acquisition techniques for the 300 most frequent content words.

Methodology

This mixed methods study entailed participants completing a vocabulary learning intervention stage followed by three post-learning tasks. The data generated by these tasks was subjected to both quantitative and qualitative analysis.

Selection of Target Words

The number of target items was set at 300. As explained above, this was small enough to be a feasible learning target, and large enough to represent a meaningful front-loading of vocabulary for learners.

[4] Here we use the term "learnability" to mean the ease, or difficulty, of learning a particular word – i.e. it is a feature of the individual word, not of the learner (this follows, for example, Carter and McCarthy (2014); Ellis and Beaton (1993); Laufer (1990)).

[5] We have adopted the terminology used by Laufer et al. (2004); form-recall measures productive word knowledge by requiring learners to supply the L2 target word, and meaning-recall requires them to demonstrate understanding of the meaning of the L2 target word.

The starting point for target item selection was a list of the most frequent 750 words from CorCenCC, as listed in Appendix 1 of the *Geirfan* report (Knight, Tovey-Walsh, et al., 2023). *Geirfan* is a pedagogical word list, adapted from the *Amliadur* lists (Knight, Morris, Tovey-Walsh, et al., 2020). Working through that list starting from the most frequent, target words for the study reported here were selected according to the following criteria:

- Only nouns, verbs and adjectives were selected as target words. Adverbs were systematically excluded because of the way many of them are formed in Welsh (requiring a separate particle and often a mutation). Function words were excluded as they tend not to have straightforward translation equivalents, and their connection with syntactic structures means that many are difficult to learn or use in isolation.
- Where the same letter string was used for different word classes or meanings, only the more frequent of these was included. Examples of this are the inclusion of the more frequent *de* (= *right*) but not of the less frequent *de* (= *south*).
- Where two candidate target words had the same English translation, and no other English word was a suitable translation, only the more frequent was included. Examples of this are the inclusion of the more frequent *dydd* (= *day* as complement to night), but not of the less frequent *diwrnod* (= *day* as twenty-four-hour period).
- Whereas a single word form is often used in English for both verb and noun (e.g. *work*), Welsh uses different word forms for each word class, for example *gwaith* = *work* (n) and *gweithio* = *work* (v). To enable participants to distinguish between nouns and verbs, English translations of verbs were preceded by "to".
- Following the above protocol, modal verbs that cannot be expressed with an infinitive "to", e.g. *dylai* (= *should*), were excluded.
- Words only used in formulaic expressions were excluded (e.g. *bodd (≈ liking)*, used in *wrth fy modd (= I'm delighted)*[6]).
- Words that do not have a straightforward equivalent English translation were excluded (e.g. *eisteddfod* = a festival specific to Wales, celebrating Welsh language poetry, music, literature and culture, with competitions and awards).

This selection process led to the exclusion of 178 words; the last (by frequency) of the 300 words to be selected was *llafur* (= *labour*), which ranks 478 on the frequency list from CorCenCC. The resulting set of target words, in order of frequency, can be seen in Appendix A (all appendices to this paper can be found via https://osf.io/xn7ac/).

Participants

Participants were recruited from the student population at a university in South Wales, and through personal contacts (students were encouraged to tell their friends about the study).

[6] Through the mutation system in Welsh, the initial letters of some words systematically change according to grammatical context, gender, etc. In the example given here, the pronoun *fy* (= *my*) has caused the initial letter of *bodd* to mutate to *m*.

A relatively high drop-out rate was anticipated, given the extended length and intensive demands of the task. Forty participants were recruited, on the basis that a dropout rate of 50% would still result in a workable sample of twenty. Recruitment was through information presented in lectures, and through an email to all students in the Faculty, inviting them to contact the research team if they were interested in participating in the study.

Those who made contact were directed to a Google Forms document with information about what participation entailed (see Appendix B), an eligibility check, and a consent form[7]. The first 40 participants who completed the consent form were recruited, and replies were sent to all other expressions of interest with an email that included information about Welsh language courses offered by the NCLW. As can be seen in Appendix B, the commitments required of participants were made clear. Participants were also told that as a reward for completion of the study, they would receive an Amazon voucher, be invited to attend an optional Welsh language taster session designed specifically for participants in this study and receive a certificate of participation with wording appropriate for use in a cv or job application.

The study was designed for participants with no prior knowledge of Welsh. However, Wales is a bilingual country, with official signs and communications in both Welsh and English. As all participants either lived/studied in Wales or had a connection with Wales (hence their interest in the study), they inevitably had had some exposure to Welsh through the linguistic landscape. The criteria for eligibility were therefore set as follows: "Participants should be first language or highly proficient users of English and should not be Welsh speakers. As a guide, if you know the Welsh words for two or more of these words: *go, give, big, new, people, children*, you are NOT eligible to take part in this project." Those test words were selected from the top 20 most frequent words in the target word list. The requirement for English proficiency was to ensure participants would know the English translations of the Welsh target words. The questionnaire administered at the end of the study revealed participants' language backgrounds as follows: 14 were first language users of English (one of these bilingual with Hindi), and the other first languages reported were Japanese, Turkish, Arabic, Hindi, Malayalam, and Czech. Three participants reported having no second languages. Others' second language knowledge varied from "basic French" (two participants) to knowledge of 4 other languages (four participants).

Twenty-three[8] participants completed 50 days of learning[9] and the immediate post-learning task. Of these, 22 also completed the delayed post-learning task, and of those, 20 also completed the end-of-study questionnaire.

---

[7] The project was approved by the Swansea University Faculty of Humanities and Social Sciences Ethics Committee (approval number 1-2023-7871-6856).

[8] A 24th participant completed the immediate post-learning test too late to be included in the results.

[9] The questionnaire included the question: "We realise that 10 mins per day of vocabulary learning is a big commitment. Please tell us whether there were any days that you missed (how many), and whether you used extra days to make up for this. Please note that we do not need to know why you missed days, and also that there was no obligation to add extra days for any you missed." Most reported that they had missed a few days; some (not all) had made up for this by tagging on extra days.

Seven participants contacted us during the study to say they had to withdraw because they lacked the time to undertake the learning. A further nine did not complete the post-learning task, and did not contact us further, despite regular reminder emails.

Materials

Participants were provided with a learning pack which contained:

- 300 flash cards (see Appendix C for sample). The size of each flashcard was approximately 140mm x 65mm and they were included in the packs in no particular order. In addition to the target word and English translation, the flashcard included a grid with 50 boxes, so that participants could mark each day they had looked at that card. Where words on more than one flashcard were from the same word family, the shared letter strings were in bold (see Appendix A).
- a 2-page instruction sheet (see Appendix D), with:
  - information about when to begin learning, and how to record the learning schedule
  - suggestions for planning their learning – e.g. how many words to attempt per day
  - suggestions about how to use flashcards, the keyword technique, and the word family technique
  - a note about parts of speech, and pronunciation
  - instructions about what to do on completing the 50 days of learning
- a "keep track of your 50 days of learning" guide. Participants started on different days (for their own convenience). This guide enabled them to note their own day 1, 2, etc. and to mark the end of their 50 days of learning.
- a guide to Welsh pronunciation.
- a copy, for their reference, of the information and consent form they had signed online.

Three post-learning tasks were prepared:

i. An immediate post-learning task, presented and completed on Google Forms. The instructions, which can be seen in Appendix E, were followed by a two-section task. The first section measured productive knowledge (form recall), presenting the English translations of the 300 target words, and requiring participants to provide the Welsh. The second section measured receptive knowledge (meaning recall) of the 300 words, requiring participants to supply the English translations of the list of Welsh target words.
ii. A delayed post-learning task. This was identical to the immediate post-learning task, and a link was sent to participants 12–14 days after they had completed the above. Participants were instructed not to look at their flashcards between completing the two tasks.
iii. A final questionnaire, asking for responses to questions about language background, strategies, and the experience of studying in this way (see Appendix F for list of questions).

The post-learning tasks were all presented on Google Forms.

Procedure

Participants started their learning during November 2023. Throughout the learning period, participants were sent a weekly email acknowledging the work they were contributing to the study and encouraging them to keep going. Participants were asked to contact the research team when they had completed their 50-day learning period. At that point they were sent an email with a link to the immediate post-learning task. 12-14 days after they had completed that task, they were sent a link to the second post-learning task. When they had completed that, they were sent a link to the final questionnaire.

Post-learning tasks were marked as follows, by two raters:

- 2 points were awarded for every correct answer
- 1 point was awarded for answers that contained minor spelling errors, or were spelled according to English orthography
- 0 points were awarded for incorrect answers or blanks

A full set of marking protocols can be seen in Appendix G.

Participants were repeatedly assured that the purpose of the post-learning tasks was to ascertain which words were easiest/most difficult to learn, rather than to assess their performance as learners. Nevertheless, it should be noted that it was not possible to determine whether participants had referred to external sources (e.g. Google Translate, dictionaries, Welsh-speaking friends/family) when completing the tasks.

Results

Word Learnability

The learnability of the 300 target words was assessed via the immediate post-learning task. The task was marked according to the protocol listed above and detailed in Appendix G, and the total number of points scored for each word, both English > Welsh (E > W; form recall) and Welsh > English (W > E; meaning recall) was calculated. Scores were interpreted as indications of learnability, i.e. the words with highest scores were considered to be most learnable. The maximum possible score per word, if all 23 participants produced correct answers (at 2 points each) was 46. That maximum score was awarded to just one item: in the W > E task, all participants provided the correct English translation for *Cymraeg* (= *Welsh*). A score of zero was also awarded to just one item: no participants provided the Welsh word for *to hear* (= *clywed*). Table 1 presents descriptive statistics for scores per word (median scores are reported as data was not normally distributed). The full list of target words ranked by learnability, with their respective scores out of 46, can be found in Appendices H (E > W) and I (W > E).

Table 1 indicates that meaning-recall scores are higher than form-recall scores. Following a significant Shapiro-Wilk test, a Wilcoxon's signed-rank test showed that scores on the W > E test were significantly higher than on the E > W test ($W = 2568$, $p < .001$). This concurs with the widely accepted view that receptive knowledge

Table 1 *Scores Per Word (learnability scores) – Immediate Post-learning Task*

Test Mode	Minimum	Maximum[a]	Median	SD	MAD
English > Welsh	0	42	17	9.03	6
Welsh > English	4	46	21	9.77	7

[a]maximum possible = 46 per mode.

precedes productive knowledge of a word (see e.g. Nation, 2022). The figures in the table indicate that items differed considerably in terms of whether their form or meaning was recalled by participants. This supports the premise of this study that some words are more challenging to learn than others.

The lists in Appendices H and I reveal both similarities and differences in the ranking of form-recalled (E > W) and meaning-recalled (W > E) words. For example:

- cognate words (with similar or identical orthographic form in both languages, typically due to borrowing) rank high in both lists;
- of the 43 words[10] ranking highest (i.e. most learnable) in each list, 31 are translation pairs – i.e. they were well learned in both form and meaning recall modes;
- of the other 12 words that ranked in the top 40 of the W > E task scores, most only marginally missed the 43 cut-off point in the E > W list. A clear exception to this was *newyddion* (= *news)*, which was the 39th most learnable word in meaning recall (W > E), but the 124th most learnable in form recall[11]. Scrutiny of the data shows that many participants lost points because they misspelled *newyddion*. Misspelling, especially of longer words, was a pattern throughout the data, contributing to the lower scores for form recall than for meaning recall.
- observable characteristics of the least learnable words (i.e. those awarded the lowest scores), included that they tended to be slightly longer; and they tended to include graphemes that are specific to Welsh (*ch, dd, ff, ng, ll, ph, rh, th* are distinct graphemes in Welsh), or that are used differently in Welsh than in English (*w* and *y* are vowels in Welsh, for example). Of the 29 target words beginning with the letters *cy* (anecdotally challenging for learners), 19 were among the 49 least learnable words (W > E list, scores of 12 and under).

Word Retention

Word retention was assessed via the delayed post-learning task. The period between the two tasks was intended to be two weeks, but in fact varied between 13 and 32 days because of participants' availability (M = 17.2, SD = 4.4). The task was marked,

[10] 43 was used as a cut-off point in the ranking to include words with the same scores. This included all E > W items with a score of 30 or above, and all W > E items with a score of 36 or above.

[11] Note that the ranking figures given here are indicative: the ranked lists include items with identical scores. See Appendices H–K for detail.

Table 2 *Scores Per Word (retention scores) – Delayed Post-learning Task*

Test Mode	Minimum	Maximum[a]	Median	SD	MAD
English > Welsh	0	43	11	8.48	5
Welsh > English	2	44	15	9.78	5.5

[a]maximum possible = 44 per mode.

and scores calculated, in the same way as the immediate post-learning task. This time, the words with highest scores were considered to be the most well retained (i.e. they were remembered after a period during which they had not been encountered). Because only 22 participants completed this task, the maximum possible score was 44. Again, the maximum score was awarded to just one item: in the W > E task, all participants provided the correct English translation for *Cymru* (= *Wales*). Table 2 presents descriptive statistics for retention scores by word. The full list of target words ranked by retention, with their respective scores out of 44, can be found in Appendices J (E > W) and K (W > E).

Of the words ranked in the top 43 for learnability (E > W), 33 also ranked in the top 41[12] for retention. Of the remaining ten, seven were within the top 60. There were three exceptions that fell considerably in the rankings between the immediate and delayed tasks: *family* > *teulu* (116th in the delayed test); *matter* > *mater* (78th) and *middle* > *canol* (97th).

As in the immediate post-learning task, following a significant Shapiro-Wilk test, a Wilcoxon's signed-rank test showed that scores on the W > E test were significantly higher than on the E > W test ($W = 1973.5$, $p < .001$).

Participant Learning Behaviour

Mean scores on the W > E task were higher than on the E > W task in both the immediate and delayed post-learning tasks (see Tables 3 and 4). A paired samples *t*-test showed the difference to be significant for the immediate post-learning task: (t (22) = 6.046, $p < .001$), with Cohen's d (1.261) suggesting that this is a large effect; and for the delayed post-learning task: (t (21) = 3.958, $p < .001$), with Cohen's d (0.844) again suggesting a large effect (see Plonsky & Oswald, 2014). Two of the participants scored higher on the E > W task. One possible reason for this is that the W > E was the second of the two tasks, and with the entire task taking 60–90 minutes, fatigue may have set in for those two participants. The mean difference between the E > W and the W > E scores on the immediate post-learning task was 56.43 points (*SD* 44.77), and on the delayed post-learning task was 47.09 (*SD* 55.8).

One of the study aims was to identify effective acquisition techniques for the 300 target words, and the questionnaire data provides an insight into the participants' learning strategies and approaches. Below we report relevant information from the six

[12] 41 was used as a cut-off point in the ranking to include words with the same scores. This included all E > W items with a score of 25 or above.

Table 3 *Scores Per Participant – Immediate Post-learning Task (N = 23)*

Test Mode	Minimum	Maximum[a]	Mean	SD
English > Welsh	37	588	235.13	152.62
Welsh > English	56	598	291.57	151.42
Combined score E>W + W>E	93	1186	526.7	300.73

[a]maximum possible = 600 per mode.

Table 4 *Scores Per Participant – Delayed Post-learning Task (N = 22)*

Test Mode	Minimum	Maximum[a]	Mean	SD
English > Welsh	15	587	187.77	151.26
Welsh > English	19	592	234.82	156.85
Combined score E > W + W > E	78	1179	422.55	303.1

[a]maximum possible = 600 per mode.

participants who scored highest on both the immediate and the delayed post-learning tasks (the full questionnaire data set is in Appendix L). These six participants' combined scores (E > W + W > E) on the two post-learning tasks are shown in Table 5, along with information about their language backgrounds. For ease of discussion the participants are coded p01, p02, p03 and so on.

As can be seen in Table 5, there is significant variation between the attrition rates of participants' vocabulary knowledge, with one participant in fact scoring higher on the delayed than the immediate post-learning task, and one scoring 354 points lower. All the top scorers knew at least one second language, and on average knew 2.5 second languages, in comparison to an average of 1.6 second languages known by the other 14 participants. While this might be interpreted as support for claims of a multilingual advantage in language learning (see e.g. de Bot & Jaensch, 2015), there is insufficient data to calculate whether this difference is significant, and indeed participants p04 and p05 claimed that their knowledge of French was basic. Only participant p02 knew a second language from the Celtic language family (Scottish Gaelic). However, it is notable that all six of these top scorers, but only one other participant, reported using their other L2s to help them learn and remember the Welsh vocabulary. Four of the six cited specific instances of this, especially in connection with the keyword technique (e.g. "For instance, the Welsh word 'arian' means 'money.' In my L1 - Malayalam, there is a word 'Ari' (Ah-REE) which means 'rice.' So, to remember 'Arian,' I associated it with needing money to buy rice." p03).

The questionnaire responses reveal detailed information about the approaches and techniques of these six and all the other participants. Salient responses of the top-scoring participants include:

Table 5 *Scores and Language Backgrounds of the Six Top-scoring Participants*

Variable	p01	p02	p03	p04	p05	p06
First language(s)	English	English	Malayalam	English	English	English
Other languages known	German, Spanish, Portuguese	German, Scottish Gaelic	English, Tamil, Hindi, Bengali	French	French	Italian, French, German, Spanish
Immediate post-learning task	1186	1127	913	861	796	773
Delayed post-learning task	1179	1056	559	742	808	679
Difference between immediate and delayed post-task scores	–7	–71	–354	–119	+12	–94

- They looked at 8-20 words per day, except for p01, who looked at around 30 for the first 10 days, and around 75 on subsequent days.
- In contrast to most other participants, they devised systematic approaches to deciding which words to look at each day. These included:
  - "I came up with a system on the first day. I split them roughly into verbs, adjectives and nouns then put them in groups of ten cards. I then did the maths of how many words I would need to learn a day to cover all 300 and came up with a 10-day system. Every 10 days looked like this: days 1 and 2 I learnt new sets of 10 words, day 3 I covered those 20 words, days 4 and 5 I learnt another two sets, day 6 I covered those, days 7 and 8 I learnt two more sets, day 9 I covered those last two and then on day 10 I revised all 60 words I had learnt. That way, over 5 sets of this routine, I managed to cover all 300 words." (p04)
  - "I tried to introduce some new words each day, and then would shuffle the deck of previous flashcards and draw some to practice." (p02)
  - "At the beginning of the 50 days I split my flashcards into words that looked 'easy' and 'hard', and started on the easier words, all the way through I would pick words that looked easier for me to learn. I feel in a way this may have set me up for failure from the beginning, however, as I did not feel motivation to attempt the 'hard' words, and therefore never ended up learning those." (p05)
  - "When doing 10 words a day to start off with, it was completely random. Once I'd studied all 300 words at least once, I'd go through them all again. The ones I remembered correctly I would put in the 'correct' pile. The ones I didn't remember correctly, I would put in an 'incorrect' pile. Once I had gone over all 300 words again, I would revisit the 'incorrect' pile and repeat the process, adding the words I could now correctly recite into the 'correct' pile and returning the ones I still couldn't recite correctly into the 'incorrect' pile. And so on, until all words were in the 'correct' pile. I would then repeat the process with the 300 words." (p01)

  - "I initially looked though them all and put to one side the ones I'd learned over the years (*English, England, Welsh, Wales,* service (from motorways services) *Christmas, thank, night, news*) and didn't look at them again till the end (green pile). Then I extracted the ones that were either the same or very similar to english (green pile). Then I made a pile of words where I'd successfully given myself a way of remembering them (orange). Finally I had a pile of words I didn't have anything for (red). I went through the red pile a few times and some of the words made it into the orange pile. As I gradually learned words from the orange pile, they made it into the green pile. A few times I went back to the red pile, but it was really just so that I could say I'd tried." (p06)
  - "I did not follow a specific criteria in choosing which words to learn each day. However, I must admit that at the beginning of Welsh vocabulary learning, I did prioritize nouns and learned them first over verbs." (p03)
- All but one of the six reported that they used the keyword technique, and were able to provide multiple examples, which included links made via their second language(s), and creative links. Several of the other learners also reported using the keyword method but provided few or no examples.
- Four of these six participants reported using the word family technique, with attention to the word parts that were in bold on the flashcards. Three noted that they found this helpful. (In scoring the post-learning tasks, it was noted that participants often confused different members of the same word family).
- In terms of other learning techniques used, general advice, and suggestions for improvements, responses from these six participants included:
  - saying the words out loud – using the pronunciation guide and/or Google Translate;
  - keep revisiting the same words, both E>W and W>E;
  - be organised in terms of sorting the cards into piles – decide on a technique for this and stick to it;
  - split longer words into chunks to remember;
  - more information or resources relating to pronunciation would be helpful, including information about the effect of the *to bach* (circumflex).

Participants were asked about what they found enjoyable, and less enjoyable, about the learning task. Responses from all 20 respondents were overwhelmingly positive (it should be noted that these were the participants who completed all tasks; it is likely that those who withdrew from the study felt less positive). The challenge of learning a new and very different language and/or being able to connect their learning with the linguistic landscape in Wales were aspects noted as particularly enjoyable. Sticking to the learning routine was noted by nine participants as a less enjoyable, or more challenging, aspect. Three participants (including p03 and p04) reported that they found it difficult not knowing how to pronounce the target words. The most positive response was from participant p01, who commented that "there was no part of [the task] that I didn't enjoy..... It would be good if the university could do this with other languages, I would happily do this again with another language. I am grateful for the opportunity to have learned some Welsh, and this was a good way to do so as it was only 10 minutes

a day, which meant it was a manageable commitment. Having all the information and cards in one clear zip-up folder was also very helpful."

It is perhaps testament to participants' positivity about the task (and the desire to know more about Welsh pronunciation) that 12 of them attended an online evening Welsh language session at the end of the study. This was one of the "rewards" offered to participants who completed all tasks. Feedback from participants indicated that they found the session useful and enjoyable.

Outcomes, Limitations, and Suggestions for Future Developments

The findings and outputs reported above and in the appendices have capacity to inform Welsh learning curricula, course materials, assessments, teacher education/methods, and self-study techniques, by providing information about which essential words require most attention, the incorporation of effective techniques within course material, addressing attrition of vocabulary knowledge, and so on.

The study generated a rich set of data for each participant, but this came at the cost of sample size; the heavy and sustained demands on participants resulted in a relatively high participant drop-out rate, and statistical findings reported in this paper should be interpreted with this in mind. Similarly, it should be noted that the immediate and two-week delayed post-learning tasks represent a somewhat narrow definition of learning and attrition, respectively. As with most studies of this kind, there is a reliance on accurate self-reporting by participants; we have assumed that participants followed the learning and post-learning task instructions, but cannot conclusively claim this to be the case. Systematic replications, including with extended/supervised post-learning test times, or different target languages, might address the above caveats. There is also potential to extend analysis of the data generated by the study reported here; further research might include:

- analysis of errors in the post-learning tasks to:
  - ascertain stages of incremental word knowledge, and to identify any common patterns of misspelling;
  - distinguish between items not recalled at all, and those partially recalled;
  - identify which items are commonly confused due to orthographic similarity (e.g. *bod/dod (= to be/to come)*, *newid/newydd (= to change/new)*, *agos/aros (= near/to wait*) or semantic similarity (e.g. *ifanc/bach (= young/small)*, *diweddaf/olaf (= last (most recent)/last (in a list));*
- examination of the connections between individuals' reported learning behaviours for specific words, and their post-learning scores on these words;
- looking at the relationship between participants' previous language knowledge and their performance on the post-learning tasks;
- examination of individuals' patterns of vocabulary knowledge attrition, and connections between this and individual learning behaviours (there is some evidence, for example, that keyword methods result in stronger word retention);
- investigation of the relationship between the number and spacing of encounters with each word, and acquisition success;
- devising acquisition techniques for those high frequency words that are the most challenging to learn.

The study here has demonstrated that for well-motivated learners it is possible to acquire most of the 300 most frequent nouns/verbs/adjectives in Welsh by learning for 10 minutes per day across 50 days. This makes it feasible to front-load a syllabus with useful vocabulary, and supplementing this with opportunities to use newly acquired words in context will deepen and embed word knowledge further. The ongoing motivation of participants who completed the tasks indicates that learners might continue their learning schedule beyond 50 days, applying acquisition strategies beyond the first 300 words.

Afterword

The opportunity to conduct research into lexical acquisition in Welsh has afforded the authors valuable perspectives, which in turn have highlighted aspects of vocabulary learning and instruction that are less evident in majority language contexts; (see also Morris, 2011; Fitzpatrick et al., 2018; Knight, Fitzpatrick, et al., 2023; Knight et al., 2021; Morris et al., 2019). Firstly, research, policy, and practice are closely integrated in the Welsh language context. Through funding, collaboration, and/or co-authorship, the Welsh Government and/or national education providers have been actively involved in the research reported in all the forementioned studies, enabling alignment with policy objectives and with practitioner needs. The findings of the study reported here have already informed curriculum changes; those harder-to-learn frequent items now receive more attention in the A1 course material.

Secondly, unlike beginner adult learners of English, many adult learners of Welsh will have had negligible exposure to their target language through film, TV, music or social media. Many are "incomers" to Wales so have had little exposure to the bilingual landscape (signage etc.). This opportunity to examine vocabulary acquisition in true beginner learners can shed light on threshold stages of acquisition, and the way newly acquired items are integrated into the nascent target language lexicon – or existing L1 or L2 lexicons. As is the case in minoritised languages, motivation to learn is typically intrinsic and integrative; there are no monolingual users of Welsh, and in conversation there is always the option to revert to English, so opportunities for learners to use the target language must be sought out.

Lastly, working with a lesser resourced language necessitates a critical and creative approach to materials and curriculum development. The first comprehensive Welsh corpus, CorCenCC (Knight, Morris, Fitzpatrick, et al., 2020) is still relatively new, and despite its principled sampling frame, is much smaller, at 11 million words, than corpora typically used to inform materials, frequency lists etc. in English. Novel methodologies for building pedagogic word lists that do not rely on corpora, and that harness learner experience, such as the ranked difficulty lists here, can act as blueprints for materials development in other under-resourced languages (see Knight et al., 2021).

We have found in our Welsh language research that differences in language learning contexts can be as informative as similarities; perhaps more so, in that they force us to question assumptions and to devise creative approaches to instruction. They also offer opportunities to highlight specific variables (e.g. motivation type) for contrastive

analysis. Finally, the close relationship between language policy, language research and language instruction in a small nation that holds its language dear, makes for an iterative cycle of scholarship that accelerates and enhances knowledge growth for all parties. We hope that VLI readers find the perspectives offered by this paper interesting and useful in connection with their own investigations and reflections.

Funding

The study reported in this paper was co-funded by the National Centre for Learning Welsh, and AHRC (Arts and Humanities Research Council) Impact Acceleration Award R874 (Swansea University). The authors are grateful to Gosia Rutecka for assistance with data processing, and to Helen Prosser for guidance and advice on study design. We also acknowledge the extended contribution and intensive engagement of study participants, without which the study would not have been possible. Finally, the authors are indebted to two anonymous reviewers for their careful and insightful feedback on this paper.

References

Bauer, L., & Nation, P. (1993). Word families. *International journal of Lexicography*, *6*(4), 253–279. https://doi.org/10.1093/ijl/6.4.253

Beaton, A., Gruneberg, M., & Ellis, N. (1995). Retention of foreign vocabulary learned using the keyword method: A ten-year follow-up. *Second Language Research*, *11*(2), 112–120. https://doi.org/10.1177/026765839501100203

Carter, R., & McCarthy, M. (2014). *Vocabulary and language teaching*. Routledge. https://doi.org/10.4324/9781315835860

Dang, T. N. Y. & Webb, S. (2016). Making an essential word list for beginners. In Nation, I. S. P., (Ed.), *Making and Using Word Lists for Language Learning and Testing* (pp. 153–167). John Benjamins. https://doi.org/10.1075/z.208.15ch15

Davies, M. (2008). *The corpus of contemporary American English: 520 million words, 1990-present*. http://corpus.byu.edu/coca/

De Bot, K., & Jaensch, C. (2015). What is special about L3 processing? *Bilingualism: Language and Cognition*, *18*(2), 130–144. https://doi.org/10.1017/S1366728913000448

Dolean, D. D. (2014). Using the keyword method in the classroom: Is the interacting imagery necessary? *System 45,* 17–26. https://doi.org/10.1016/j.system.2014.04.003

Dysgu Cymraeg. (2024, April 2). 2022-2023 statistics. https://learnwelsh.cymru/about-us/statistics/2022-2023-statistics/

Elgort, I., & Piasecki, A. E. (2014). The effect of a bilingual learning mode on the establishment of lexical semantic representations in the L2. *Bilingualism: Language and Cognition*, *17*(3), 572–588. https://doi.org/10.1017/S1366728913000588

Ellis, N. C., & Beaton, A. (1993). Psycholinguistic determinants of foreign language vocabulary learning. *Language learning*, *43*(4), 559–617. https://doi.org/10.1111/j.1467-1770.1993.tb00627.x

Estyn (2023). *The Welsh language in education and training.* https://annual-report.estyn.gov.wales/2023/the-welsh-language-in-education-and-training/

Fitzpatrick, T., Al-Qarni, I., & Meara, P. (2008). Intensive vocabulary learning: A case study. *Language Learning Journal, 36*(2), 239–248. https://doi.org/10.1080/09571730802390759

Fitzpatrick, T. & Morris, S. (2024). 'O sero i dri chant': technegau caffael dwys ar gyfer y 300 gair cynnwys a ddefnyddir amlaf yn y Gymraeg. *Gwerddon, 38*, 22–76. https://doi.org/10.61257/HEOY7440

Fitzpatrick, T., Morris, S., Clark, T., Mitchell, R., Needs, J., Tanguay, E., & Tovey, B. (2018) *Rapid evidence assessment: Effective second language teaching approaches and methods.* Welsh Government Report. https://gov.wales/effective-second-language-teaching-approaches-and-methods-rapid-evidence-assessment

Knight, D., Fitzpatrick, T., Morris, S., Tovey-Walsh, B., Prosser, H., & Davies, E. (2023). Corpus to curriculum: Developing word lists for adult learners of Welsh. *Applied Corpus Linguistics*, *3*(2), https://doi.org/10.1016/j.acorp.2023.100052

Knight, D., Morris, S. & Fitzpatrick, T. (2021) *Corpus design and construction in minoritised language contexts: The national corpus of contemporary Welsh*. Palgrave Macmillan.

Knight, D., Morris, S., Fitzpatrick, T., Rayson, P., Spasić, I., Thomas, E-M., Lovell, A., Morris, J., Evas, J., Stonelake, M., Arman, L., Davies, J., Ezeani, I., Neale, S., Needs, J., Piao, S., Rees, M., Watkins, G., Williams, L., Muralidaran, V., Tovey-Walsh, B., Anthony, L., Cobb, T., Deuchar, M., Donnelly, K., McCarthy, M., & Scannell, K. (2020). *CorCenCC: Corpws Cenedlaethol Cymraeg Cyfoes – the National Corpus of Contemporary Welsh*. Cardiff University. http://doi.org/10.17035/d.2020.0119878310. https://corcencc.cymru

Knight, D., Morris, S., Tovey-Walsh, B., Fitzpatrick, T., & Anthony, L. (2020). *Yr Amliadur: Frequency Lists for Contemporary Welsh.* Cardiff University. http://doi.org/10.17035/d.2020.0120164107

Knight, D., Tovey-Walsh, B., Davies, E., Morris S., & Prosser, H. (2023). *The Geirfan wordlist: A vocabulary list for adult learners of Welsh*. https://corcencc.org/wp-content/uploads/2023/03/Geirfan-v2.0.0-English.pdf

Kremmel, B. (2016). Word families and frequency bands in vocabulary tests: Challenging conventions. *TESOL Quarterly, 50*(4), 976–987. https://doi.org/10.1002/tesq.329

Laufer, B. (1990). Ease and difficulty in vocabulary learning: Some teaching implications. *Foreign Language Annals*, *23*(2), 147–155. https://doi.org/10.1111/j.1944-9720.1990.tb00355.x

Laufer, B., Elder, C., Hill, K., & Congdon, P. (2004). Size and strength: Do we need both to measure vocabulary knowledge? *Language testing*, *21*(2), 202–226. https://doi.org/10.1191/0265532204lt277oa

Milton, J., & Meara, P. (1998). Are the British really bad at learning foreign languages? *Language Learning Journal*, *18*(1), 68–76. https://doi.org/10.1080/09571739885200291

Morris, S. (2021). *A rapid review of the National Centre for Learning Welsh [Online]*. Cardiff: Welsh Government. https://www.gov.wales/rapid-review-national-centre-learning-welsh

Morris, S. (2011). Geirfa Graidd i'r Gymraeg: Creating an A1 and A2 core vocabulary for adult learners of Welsh – a Celtic template? *Journal of Celtic Language Learning*, *15*(16), 27–43.

Morris, S., Meara, P., Fitzpatrick, T. (2019) *Geirfa Graidd Lefel Canolradd [Online]*. https://questionbank.wjec.co.uk/qualifications/welsh-for-adults/welsh-for-adults-intermediate/Geirfa%20Canolradd%20Ionawr%202019.pdf?language_id=2

Nation, I. S. P. (2022). *Learning vocabulary in another language* (3rd ed.). Cambridge University Press. https://doi.org/10.1017/9781009093873

Nation, P., & Crabbe, D. (1991). A survival language learning syllabus for foreign travel. *System 19*(3), 191–201. https://doi.org/10.1016/0346-251X(91)90044-P

Office for National Statistics (2022). *Welsh language, Wales: Census 2021*. https://www.ons.gov.uk/peoplepopulationandcommunity/culturalidentity/language/bulletins/welshlanguagewales/census2021

Plonsky, L., & Oswald, F. L. (2014). How big is "big"? Interpreting effect sizes in L2 research. *Language learning*, *64*(4), 878-912. https://doi.org/10.1111/lang.12079

Richards, J. C. (1976). The Role of Vocabulary Teaching. *TESOL Quarterly 10*(1), 77–89. https://doi.org/10.2307/3585941

Schmitt, N., Dunn, K., O'Sullivan, B., Anthony, L., & Kremmel, B. (2021). Introducing Knowledge-based Vocabulary Lists (KVL). *Tesol Journal*, *12*(4). https://doi.org/10.1002/tesj.622

Webb, S., & Nation, P. (2017). *How vocabulary is learned*. Oxford University Press.

Welsh Government. (2017). *Cymraeg 2050: A Million Welsh Speakers [Online]*. Cardiff: Welsh Government. https://www.gov.wales/sites/default/files/publications/2018-12/cymraeg-2050-welsh-language-strategy.pdf

Welsh Government. (2021). *Response to the recommendations of the rapid review of the National Centre for Learning Welsh [Online]*. Cardiff: Welsh Government. https://www.gov.wales/response-recommendations-rapid-review-national-centre-learning-welsh-html

Zipf, G. K. (1936). *The psycho-biology of language*. Routledge.