Fitzpatrick, Al-Qarni and Meara 2008

Intensive Vocabulary Learning: a case study.
Tess Fitzpatrick, Ibrahim Al-Qarni and Paul Meara

This paper is one of a series of studies in which we have investigated in some detail the behaviour of individual subjects in vocabulary learning tasks. Despite the massive growth in the literature on vocabulary that has occurred in the last twenty years, single subject case studies remain somewhat rare (Meara 1995). A small collection of studies of this sort appeared as a special issue of *Second Language Research* in 1995, and there has been a trickle of case studies since then (Newton 1995, Ridley and Singleton 1995, Altmann 1997, Horst and Meara 1999, Meara 2005, Miura 2005). Single subject case studies are, of course, a standard methodology in the study of vocabulary acquisition in an L1, and it is surprising to us that the methodology has not been more widely exploited in L2 studies. Because large-scale group studies are difficult to organise, they tend to be methodologically conservative, and are often able to look only at gross effects in vocabulary acquisition. Single case studies, on the other hand, allow us to be methodologically innovative, and to ask questions which are exploratory and risky. Good examples of this approach are Segalowitz, Watson and Segalowitz (1995), and Horst and Meara (1999), both of which explored innovative methodologies, which would have been very difficult to implement in the context of a large experimental study.

This paper, then, is a further contribution to this small but methodologically interesting literature that uses single subjects in vocabulary acquisition research. The question we are interested in here is whether a single subject can learn a relatively large vocabulary in a short space of time, and whether performance drops off if large quantities of input are maintained over an extended time period. We also ask whether there is a marked difference in the number of words that the subject learns "productively" and "receptively".

METHODOLOGY

The subject

The subject in this study is an L1 English speaker, who we will refer to as Sue. Sue is female, aged 41, teaches linguistics in a UK university, and at the time of the study had no knowledge of Arabic other than a couple of basic greetings.

The task

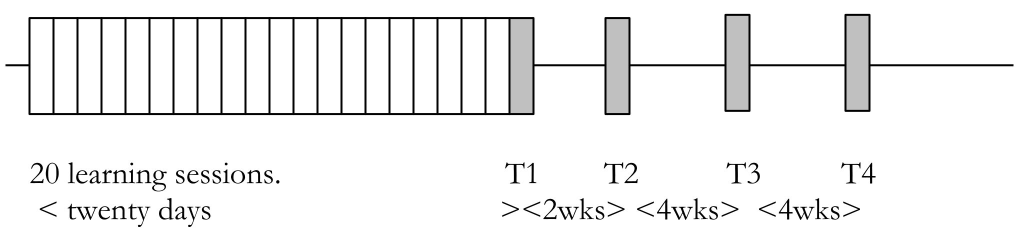
Sue was asked to learn a vocabulary of 300 Arabic words over a period of 20 days. A list of 300 relatively high frequency Arabic words was selected, and each of these words was coded on a file card. The cards contained the English transcription of the Arabic form of the word, and its English translation. The cards also contained 20 numbered boxes, one for each day of the learning period, so that the learner could indicate on which days she had encountered or revised each word. Sue was instructed to spend a maximum of 30 minutes each day on the learning task. During this time she was expected to learn fifteen new words which were assigned as that day's learning task, and to revise any words which she had learned previously. As well as recording every encounter with each of the target words, Sue kept a diary of her learning and the strategies that she employed.

Data collection

At the end of the learning period, four test sessions were administered. Each test session comprised two parts, a test of productive knowledge of the 300 words, and a test of receptive knowledge. These were straightforward translation tasks. The productive knowledge test (recall test) was always administered first, with the learner being asked to provide the Arabic translation (transcribed in Roman script) of English cues. In the receptive test (recognition test), English transcriptions of Arabic words were given and the learner provided the English translation. The tests were administered immediately after the last learning session, and two weeks, six weeks and ten weeks after the first test.

The procedure is summarised in Figure 1 below.

Figure 1: Summary of the learning procedure



RESULTS

Two analyses are reported in this section. The first analysis reports the overall levels of performance. The second analysis reports a more detailed account of the way each of the words was handled.

A: basic analysis

The basic results of this study are reported in Table 1 below. The table shows the number of words that were correctly recalled in each of the two test modes at each of the four testing times.

Table 1 : Number of correct responses at four test times:

	T1	T2	T3	T4
recognition test	286	262	221	219
recall test	283	191	135	149

The data clearly shows that, contrary to expectation, Sue had no difficulty in acquiring almost all of the 300 words she was required to learn, and she was able to retain these words as long as she was allowed to rehearse them. Our initial expectation had been that the cumulative effect of learning a new set of words, day after day, would eventually cause performance to drop off dramatically, but there is no evidence here that this is the case.

However, there IS some evidence that this acquisition is at best temporary. Of the 286 words that Sue recognised at T1, only 219 were still recognised at T4. There is some evidence that the number of words recognised at this point is beginning to plateau out. For words which are part of Sue's productive vocabulary, the position is considerably worse, with just over half of the 283 words correctly produced at T1 still capable of being recalled at T4.

B: Detailed analysis

The overall analysis reported in the previous section implies that the learning load imposed on Sue was relatively light, and that learning new words at this rate ought to be possible over an extended period. A more detailed analysis of the data suggests that this interpretation is something of an over-simplification.

Figure 1 below shows that the pattern of correct recalls is not as straightforward as this account implies. This graph shows how Sue's performance deteriorates as more words are learned, and her Arabic vocabulary increases in size. The graph shows the number of words which were correctly generated in the productive test at T4. The graph clearly shows that there is a strong downward trend in the data. Words learned very early in the experiment are more likely to be correctly remembered than words acquired in a later learning session, and only a third of the words introduced in sessions 18, 19 and 20 are actually being retained. Similar effects are found in the data for all the testing sessions.

The data for receptive vocabulary is rather more difficult to interpret. See figure 2. Here too there is evidence of a downward trend in the data, but the trend is much more gradual than the data reported in the previous paragraph. For most of the learning sessions, the number of words retained at Test 4 is about 66%, and only for words acquired in the final few learning sessions did retention fall below this level. Only three of the 15 words learned on day 20 were still recognised at Test4, but again, it is unclear whether this is an effect of vocabulary overload, or whether it is merely a reflection of the fact that words learned later have fewer opportunities to be rehearsed.

Figure 1: The effect of increased load on retention: words known productively at T4.

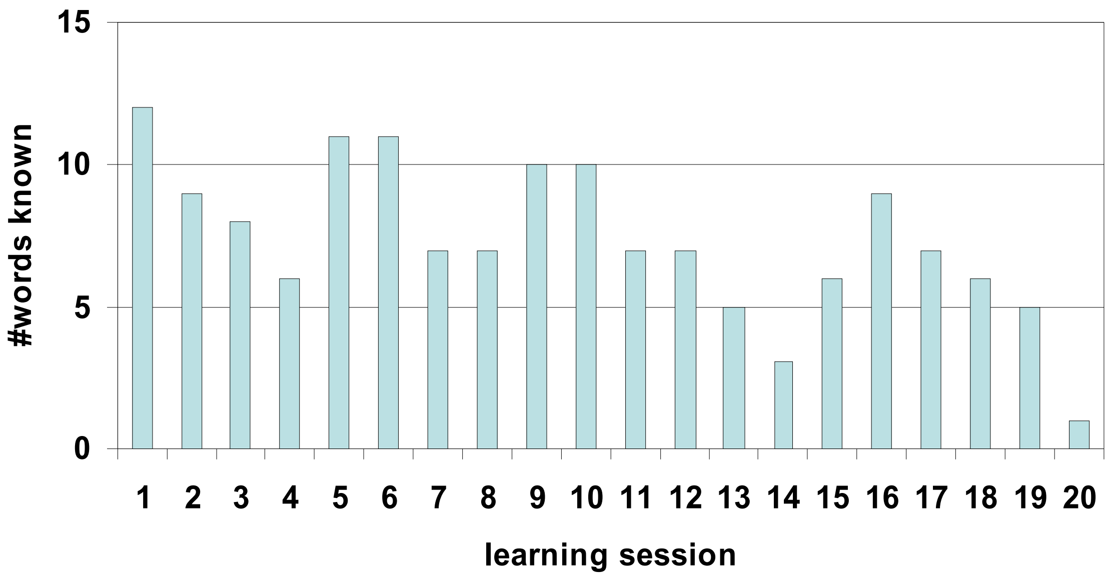
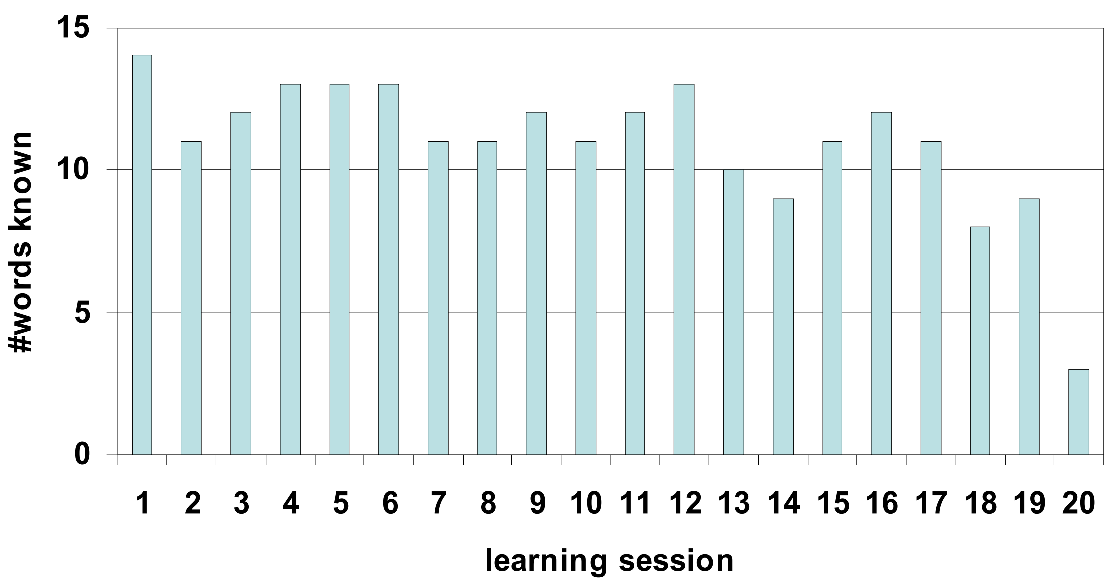


Figure 2: The effect of increased load on retention: words known receptively at T4.



Sue kept extensive notes about which words she reviewed, and these notes also allow us to analyse other aspects of the data set. The notes suggest that words which were encountered more often, as part of the review process, were more likely to be recalled in the subsequent tests ($t=2.79$, $p<.01$), that words for which Sue employed a conscious mnemonic were much more likely to be recalled on the subsequent tests ($X^2=15.50$ $p<.001$), and in general shorter words were easier to learn than longer words. This last factor echoes a finding reported by Laufer (1997).

DISCUSSION

This section will cover two discussion points: a) the methodological implications of this study, and b) a matrix analysis of the data which assesses the long-term stability levels of Sue's learning.

A: Methodological implications of this study.

In the introduction to this paper, we made the point that single subject studies allow us to investigate vocabulary acquisition in ways which would be very difficult to implement with large subject groups. The task in this paper was a longitudinal study which required our subject to learn a relatively large vocabulary over an extended period of time – a very clear example of a task which would be difficult to administer to 'normal' groups of subjects.

In some respects, however, the task was much less difficult than we had originally expected it to be. Our original expectation had been that fifteen new words each day would present a learning burden that was close to practical limits, and we fully expected that performance would deteriorate substantially as more and more words were added to the system. Surprisingly, this does not appear to have been the case. Sue had no difficulty dealing with this level of input. The limitations of the method only really become apparent after she has stopped rehearsing the new words, and attrition has begun to set in. There is a hint that performance is beginning to drop off towards the end of the learning phase, but this is nothing like the catastrophic decline in take-up that we would have predicted.

From a methodological point of view, then, two conclusions seem to emerge from this data. The first is that, given a 20-30 minute learning period per day, a learning burden of 15 words per day is not a significant load, at least for this subject. It remains an open question whether significantly different behaviour would emerge if the size of the learning burden were to be increased. Our original intention in this study had been to ask Sue to learn 25 words a day, and we revised our target after discussions with colleagues, all of whom thought that this figure was much too high a load for an average learner. The data here suggests that this assumption was incorrect, and that average learners could reasonably be asked to acquire vocabulary at a repeated rate of around 15 words a day.

The second methodological point is that we had expected learning to be less successful as the study progressed and the number of words that Sue was juggling simultaneously increased. There is some evidence that this was the case (see figure 2), but again, the effect was much smaller than we had anticipated. Sue's performance does seem to get worse as the study progresses, but again, there is no evidence of a catastrophic collapse in learning. Indeed, there is a case to be argued that after the first five or six learning sessions, her vocabulary uptake seems to be plateauing out, and the proportion of words that she knows does not decline as we might have expected. Conversely, there is no evidence here that learning words gets easier for her as the number of words she knows goes up.

This suggests to us that a study over 20 days may not be long enough for the really important features of vocabulary learning to emerge. Future studies of this sort should consider collecting data over a longer timespan.

These two points taken together suggest that the design of our study has somewhat underestimated the ability of our learner to acquire new words over an extended period. We recommend that future studies of this type should be carried out over a longer period, say, fifty days, rather than 20, and that more words should be introduced as learning targets for each day. A daily target of 25 words over a 50 day learning period would involve subjects learning 1250 words, a significantly higher number of target words than the 300 used in the present experiment. This figure ought to be sufficient for any effects of vocabulary overload to become apparent. However, the size of these targets once again highlights the importance of designs which use co-operative single subjects to carry out research tasks which simply could not be administered to larger groups in any practical way.

B: Matrix analysis

Meara and Rodríguez Sánchez (2001) have suggested that it might be misleading for researchers to report a single post-test as the definitive outcome of a study of this sort. They argue that post-tests are necessarily a snap-shot of the current state of the test-taker's vocabulary, and might not reliably reflect the long-term outcomes of vocabulary learning experiments. They propose that more reliable outcomes could be generated by taking a series of post-tests, and using these data to make a long-term forecast of how test-takers' vocabulary will develop. The methodology for making these long-term forecasts is explained in detail in Meara (1989). Basically, it involves computing a transitional probability matrix which describes how words move between a number of defined states. Meara showed that the long-term distribution of words in a test-taker's vocabulary should be defined by this matrix, rather than by the actual distribution of words at any particular time. Empirical data supports this view - the matrix forecasts reported in Horst and Meara (1999) for a long-term single subject study of reading for instance, are astonishingly accurate.

The data generated by this project are the product of Sue having learned a large number of words, and of word knowledge being measured in a series of post-tests. These two factors make the data suitable for a matrix analysis. Two analyses are reported in this section, one for words which our subject was able to produce in response to an L1 stimulus (her **productive** vocabulary), and a second analysis for words whose meaning she was able to recognise, whether or not she was able to produce the L2 word to order (her **receptive** vocabulary).

Table 2 reports the basic data for the productive vocabulary. The table shows the number of words which Sue was able to produce in the first and second post-tests, and the number of words which she was not able to produce in the same tests. Clearly there is some movement between the two test events.

Table 2: Data for the productive matrix analysis.

	Test 1	**Test 2**
# Known words	283	192
# Unknown words	17	108

A detailed analysis of the behaviour of individual words reveals that of the 283 words known at Test 1, 189 were also known at Test 2. Three words not known at Test 1 spontaneously regenerated and were known at Test 2. Similarly, detailed analysis of the unknown words at Test 2 showed that 14 of them were unknown at Test 1, but 94 of these unknown words had been successfully produced in Test 1.

We can summarise these data in a transitional probability matrix, in which the actual numbers are converted to probabilities, as in Table 3. The matrix - the shaded columns of Table 3 - shows that the probability of a word correctly produced in Test 1 also being correctly produced in Test 2 is .668, whereas the probability of a word failing to be correctly produced in either test is .824. The probability of a known word being forgotten is .332, while the probability of an unknown word spontaneously coming to mind is a meagre .176.

Table 3: Matrix analysis of Sue's productive vocabulary.

	T1			**T2**	**F1**	**F2**	**F3**	**F4**	**F5**	**F6**
Known words	283	.668	.332	192	147	125	114	109	106	105
Unknown words	17	.176	.824	108	153	175	186	191	194	195

Column F1 in Table 3 is a forecast of how Sue's vocabulary would develop if the transitional probabilities we have identified remain stable over time. The figures in column F1 are generated by applying the matrix to the data in column T2. Thus, at T2 we have 192 words which are known productively, and the probability of these words still being known at a subsequent test time is .668. We therefore forecast that 192*.668=128 of those words will be known at this time. We also have 108 words which are not known at T2, but some of these words will spontaneously regenerate with a probability of .176. This gives us 108*.176= 19 words to add to our total of known words, bringing us to a new total of 147 known words. The number of unknown words will therefore be 300-147=153 in this forecast.

The same process can now be applied to the F1 forecast. Our forecast for F2 is that Sue will know 147*.668=98 + 153*.176=27, a total of 125 known words. The remaining columns in Table 3 are generated in an identical way. Each set of figures is a forecast based on the preceding column. The figures clearly suggest that Sue's vocabulary is settling into a long-term equilibrium, where she knows about a third of the target vocabulary.

Figure 3 shows how well these figures compare with the actual data collected from Sue at the four test times. In this figure, the continuous line details the forecasts generated by the matrix, while the bars show the actual data collected at Test 3 and Test 4. The figure suggests that the matrix forecast fairly accurately matches the actual data collected at Test 3, but the forecast underestimates the number of words known at T4.

Table 4 and Figure 4 show a similar analysis for Sue's receptive vocabulary data – i.e. words she was able to recognise and supply the L1 meaning of, but could not generate in the L1>L2 translation task. In the recognition test at T2, 5 words which had been unknown at T1 were reactivated and correctly recognised.

Table 4: Matrix analysis of Sue's receptive vocabulary.

	T1			T2	F1	F2	F3	F4	F5	F6
Known words	**286**	**.899**	**.101**	**262**	**249**	**242**	**238**	**236**	**235**	**234**
Unknown words	**14**	**.357**	**.643**	**38**	**51**	**58**	**62**	**64**	**65**	**66**

Detailed analysis of the data suggest that words that are known receptively are fairly resistant to loss – the probability of a known word remaining in that state is .899. At the same time, the probability of an unknown word being correctly recognised in a subsequent test is also relatively high, at .357. These two factors combine to keep the number of known words very high. The long-term forecast here is that Sue ought to retain about 234 of the 286 words she knew at T1. Figure 4 shows a comparison between the forecasts and the data collected at T3 and T4.

Figure 3: Number of words known and projections: productive vocabulary

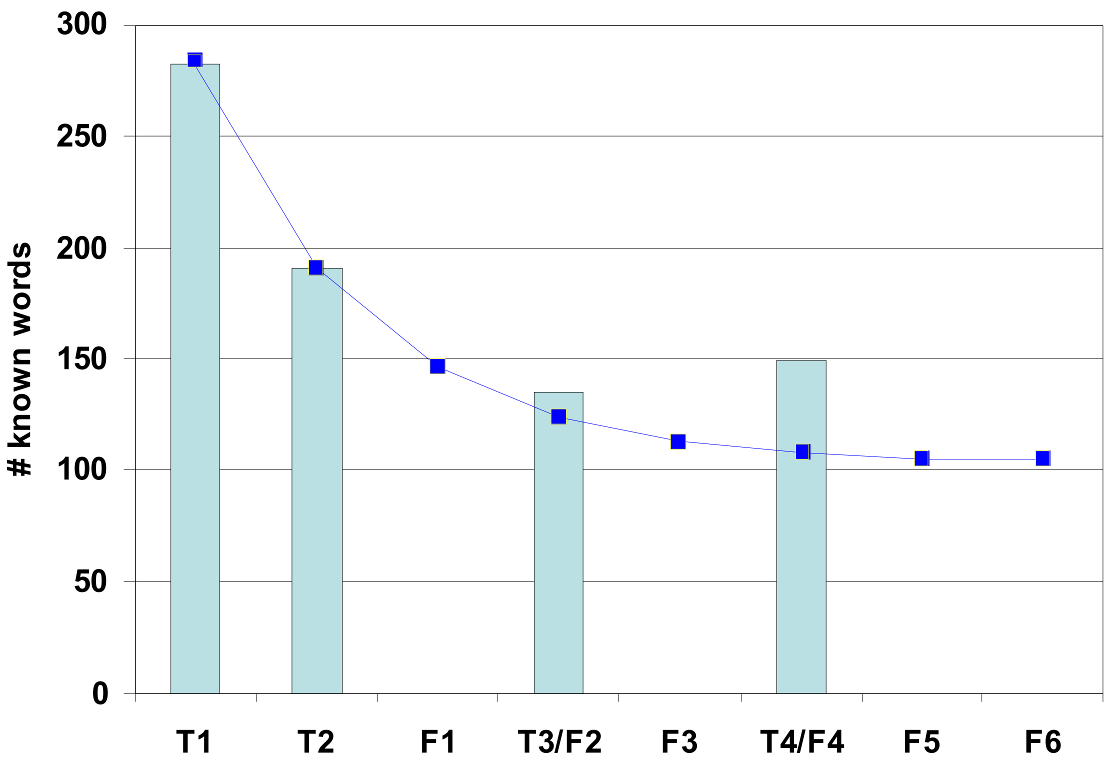
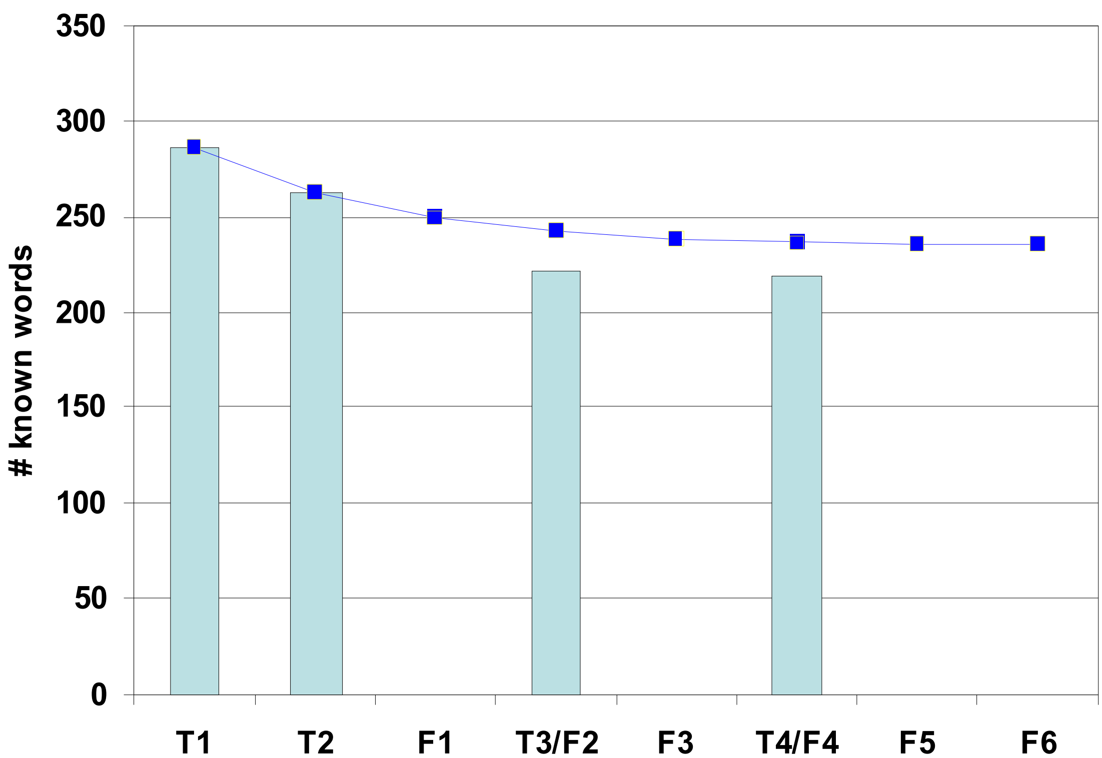


Figure 4 suggests that the matrix analysis slightly overestimates the number of words Sue knows receptively at Test 3 and Test 4. The forecasts are out by a mere 8% at T3 and 7% at T4; the difference seems to be getting smaller in successive tests. This suggests that the matrix model works well in forecasting test-takers' long-term knowledge of words learned receptively in experimental studies.

Comparison of the two sets of forecasts suggests that whereas Sue's long-term receptive knowledge of the 300 target items is about 78%, her long-term productive knowledge of these words would be about 35% . This figure is considerably lower than the equivalent figures we get from the T3 and T4 data. The matrix forecasts suggest that productive vocabulary declines at a much faster rate than the receptive vocabulary does, and that it takes some time for a stable ratio to be reached. This reinforces Meara's point about the dangers of relying on a small number of test events, and assuming that they provide a definitive picture of what will happen in the longer term.

The matrix analysis is not as close to the actual data as we had hoped, although the discrepancies, especially in the receptive vocabulary analysis, are impressively small. Only a small number of comparisons could be made with the actual data, since we only have four test events, covering a relatively short period of time. The obvious inference is that testing subjects' retention over a ten week period may not be sufficient to get at the long-term stable retention rates that we would expect to find in experimental studies of this sort. This is an interesting example of the way mathematical modelling of experiments can influence the way future experiments are designed.

Figure 4: Number of known words and projections: receptive vocabulary



CONCLUSIONS

This paper has presented some results for a single subject study of vocabulary acquisition. The subject, Sue, an L1 English speaker, was required to learn a large vocabulary in Arabic at a rate of fifteen new words a day over a period of 20 days. Sue's performance on this task was better than we expected, and the methodological implications of this were discussed. We argued that further studies of this sort need to be carried out over longer time periods, and should require the subjects to learn more words. We also looked at the factors which affected the uptake of specific words, and considered the long-term projections for vocabulary acquisition which are implicit in this data.

The data reinforce our belief that single-subject case studies are a neglected resource in work on vocabulary acquisition, and encourage us to believe that further work of this type might make a useful contribution to our understanding of the factors that affect vocabulary acquisition in adult language learners.

BIBLIOGRAPHY

Altmann, R
Oral production of vocabulary: a case study. In: J Coady and T Huckin (eds) *Second Language Vocabulary: a rationale for pedagogy*. Cambridge: Cambridge University Press. 1997.

Horst, M and P Meara
Test of a model for predicting second language lexical growth through reading. *Canadian Modern Language Review* 56,2(1999), 309-330.

Laufer, B
What's in a word that makes it hard or easy? Intralexical factors affecting the difficulty of vocabulary acquisition. In: **N Schmitt and M McCarthy** (eds.) *Vocabulary: description, acquisition and pedagogy*. Cambridge: Cambridge University Press. 1997.

Meara PM
Matrix Models of vocabulary acquisition. *AILA Review* 6(1989), 66-74.

Meara, PM
Single-subject studies of lexical acquisition. *Second Language Research*, 11,2(1995), i-iii.

Meara, PM
Reactivating a dormant vocabulary. *EuroSLA Year Book* 5(2005), 269-280.

Meara, PM and I Rodríguez Sánchez
A methodology for evaluating the effectiveness of vocabulary treatments. In: **M Bax, and J-W Zwart** (Eds.) *Reflections on Language and Language Learning*. Amsterdam: Benjamins. 2001.

Miura, T
A case study of the lexical knowledge of an advanced proficiency EFL learner. *The Language Teacher* 29,7(2005), 29-33.

Newton, J
Task-based interaction and incidental vocabulary learning: a case study. *Second Language Research*, 11,2(1995), 159-177.

Fitzpatrick, Al-Qarni and Meara 2008

Ridley, J and D Singleton
Strategic L2 lexical innovation: case study of a university level *ab initio* learner of German. *Second Language Research*, 11,2(1995), 137-148.
Segalowitz, N, V Watson and S Segalowitz
Vocabulary skill: single-case assessment of automaticity of word recognition in a timed lexical decision task. *Second Language Research*, 11,2(1995), 121-136.