



The Dígame Project

Paul Meara *Swansea University*

This paper is an informal account of part of a series of experiments on the acquisition of lexis in a second language. It falls into two parts: part one describes the development of a large volunteer subject panel with the help of the BBC; part two details some of the work that we have carried out with the help of his panel.

Part one

Over the last couple of years my students and I have completed the number of experimental studies on the acquisition of vocabulary in a second language, and on word handling in bilinguals. Some of these studies have already been described elsewhere (cf. Meara 1982, 1983, 1984a, 1984b). One unsatisfactory aspect of the studies is that they have all used very small numbers of subjects -- a shortcoming which is not uncommon in experimental psycholinguistics, but one which is often underplayed where L2 learners are concerned. Native speakers tend to behave in a relatively homogeneous fashion in the more common experimental paradigms, and this means that small numbers of subjects can produce viable and meaningful results. With non-native speakers, however, performance seems to be much more varied, and this makes it difficult to draw valid conclusions from experiments where only a handful of subjects are involved. In the UK it is difficult to get hold of large groups of subjects willing to take part in experimental research, and this difficulty is compounded if you work with a "minority" language, or if your work involves subjects visiting a specially equipped laboratory. A combination of both these factors is a recipe for disaster!

The biggest single source of language teaching in the UK is the British Broadcasting Corporation (BBC) whose department of continuing education is responsible for some six hours a week of radio and television courses in a wide range of languages. Audiences for these programmes are very large. The recent BBC Italian series *Buongiorno Italia*, for instance, was broadcast at peak times on Sundays, and reached an audience of over 250,000. However, this group of language learners has remained largely untapped as a source of research data.

In the spring of 1983, I made a series of approaches to the BBC which culminated in their agreeing to put out a number of short advertisements for me at the start of their beginners' course in Spanish *Dígame*. These advertisements explained that my team was carrying out a study of vocabulary acquisition, and asked for people following the course who also owned a microcomputer to get in touch with us. The reason for limiting replies to people with a microcomputer was principally because this allowed us to plan a set of studies which were more sophisticated than anything possible using a pencil and paper technology. We are also slightly apprehensive about the number of volunteers who might reply, and this was a convenient, if somewhat arbitrary, way of limiting the numbers to manageable proportions. We predicted that we might get 50 or 60 people for our panel, a significant improvement on the groups of 10 or 15 than we had been using heretofore. In the event, the advertisements

produced over 700 volunteers, and figure which far exceeded our expectations and indeed our ability to cope with them.

Our original plan had been to prepare a series of four or five experiments each using 50 subjects or so, and our budgets had been drawn up on the assumption that these estimates were about right. Each volunteer would receive a series of packages containing a cassette tape on which the experiment would be recorded, a datasheet to be returned to us at the end of the experiment, and prestamped envelopes to encourage the return of the data. The cost of each package worked out at about 75p, a figure which compared favourably with the going rate of payment (at 1984 rates) for volunteer subjects in the UK. The total budget for this project was a modest £500 which meant that if we sent a letter to all the people we were unable to use in the project we would have used up 25% of the total budget before we even got started. In the end we decided to go ahead with the original programme, and not attempt to cater for all the volunteers. One consequence of this decision was a large number of "unsatisfied customers", and a significant part of our efforts had to be diverted into coping with those who contact us to find out why they had not been chosen for inclusion.

We were eventually able to run a series of four experiments in a longitudinal study. This study is described in more detail below, but is also raised a number of more general issues which might be of interest to people using a panel of this sort. Our first mailing was sent out to 50 volunteers, and produced a return rate of 70 percent. This is reasonably good, though we think it is actually worse than it might have been, due to a number of unexpected technical problems. The most serious of these was that a significant proportion of the tapes we sent out turned out to be faulty and had to be replaced. This seems to be due to poor quality control by the manufacturer, but it would obviously be a serious problem if you were sending out hundreds of tapes rather than a few tens. We are currently exploring the possibility of having our tapes copied commercially in bulk. This would increase the cost of the operation, but economies to offset this increase could be made relatively easily. For example instead of posting several tapes to each volunteer, a single tape containing separate experiments could be used. It might be necessary to make it difficult for subjects to gain access to certain parts of the tape before time, but this is a relatively trivial problem. The second technical problem which caused some difficulty was that earlier versions of the microcomputer which we were using were not wholly compatible with the later versions on which the programmes had been tested. This meant that a number of volunteers got tapes which loaded but refused to run, and these tapes, too, had to be replaced. We suspect that most of the initial dropout rate was due to these technical causes. The return rate on subsequent mailing was between 80 and 90 percent, and most of those who failed to reply later explained that they had dropped out because they were unable to keep up with the course.

Towards the end of the course, i.e. six months after the volunteers originally got in touch with us, we attempted to contact another 50 people with a view to running a second set experiments. This attempt was less successful, however, and only a handful of people responded. Our guess is that this reflects the natural dropout rate on BBC's language courses, which tends to be high. Large numbers of viewers fail to get beyond the first ten lessons of a course of this sort. This high dropout rate suggests that a subject panel set up in the way outlined above may have a fairly limited lifespan, though once involved in a project like *Digame*, people seem more than willing to continue to be involved. It should also be

possible to use a panel intensively, and run a series of one off studies in the early weeks of a course. This strategy is one which will be actively exploring in 1984-85.

Part two

This section describes in more detail some of the work that has been carried out as part of the *Digame* project, and presents an interim report on the main longitudinal study.

The work that we initially wanted to carry out with the subjects concerned and their ability to recognise foreign language words. There are many reasons why this is an important line of research, but it is probably not appropriate to discuss them here. Interested readers will find a full discussion in Meara (1984a).

In a conventional laboratory, the obvious way of looking at this issue would be to run a series of experiments involving recognition thresholds. In these experiments, subjects would be seated in front of a display screen, and words would be flashed on a screen a very short durations. The subjects' task is to say what the word is. If they cannot read the word then it is flashed for a longer period, until eventually an exposure time is reached where they have no difficulty in identifying the word on screen. Typically L1 words can be recognised by normal adult native speakers in exposures of about 600-700 milliseconds.

Unfortunately, it is not possible to run standard experiments of this sort on the equipment that is likely to be available to home computer owners. Home computers linked to a domestic TV set are not sufficiently accurate for work on recognition thresholds. There are two reasons for this. Although the picture on your TV looks as though it is very stable, the picture is actually renewed once every 20 milliseconds as an electron beam sweeps across the screen. This cycle of 20 milliseconds means that it is practically impossible to present something on the screen a less than 20 milliseconds, or to increase an exposure length in steps smaller than 20 milliseconds. In laboratory work on word recognition, it is common to use increments of two to five milliseconds. A further technological obstacle is to be found in the way most home micros check their keyboards for input. This again works on a cycle principle: the better machines check their keyboards every 10 milliseconds, while others may check every 20 milliseconds. Under these circumstances, it is difficult to measure reaction times very accurately. Most of the standard laboratory paradigms for word recognition are looking for differences in the region of 50 milliseconds or so. A difference of this size is only slightly bigger than the margin of error on a home Micro configuration.

Our first task then, was to develop a word recognition paradigm that was reasonably robust, but produced response times in the region of one to two seconds. After some trial, and lots of error, we eventually developed the method described below.

The word recognition task we used consisted of presenting subjects with a string of letters in a line. This string is 20 letters long, and contains a single word embedded in it. The remaining letters are random, but reflect the likely occurrence of letters in the language been tested. (Technically, they are a first order approximation to the language of the stimulus.) The resulting stimuli looked like this:

```

*****
*
*   weolsulusimpletggiha   *
*
*****

```

Preliminary experiments with this type of display showed that native speakers have very little difficulty in identifying the hidden word. In the example above, most native speakers recognise SIMPLE immediately. For non-native speakers, however, the task seems to be considerably harder, with recognition speeds varying dramatically even for words that cause no difficulty to native speakers. The recognition times produced in a series of pilot studies suggested that native speakers typically take one or two seconds to recognise a word presented in this way. Learners take considerably longer than this; indeed many learners are unable to recognise even simple words in their L2 when the display is presented for as long as ten seconds. This pilot work, then, suggested that the reaction times we were getting were easily measurable by the equipment available.

A number of practical problems remain to be solved. If we were relying on people to carry out experiments at home, it was important to ensure that they were not cheating, and this meant that all the data had to be encrypted, and ways found of making sure that people did not do the test two or three times. In addition, the actual preparation of the programme tapes, and the distribution of the materials cause some headaches, and revealed some unexpected incompatibilities between machines. Most of these technical problems were eventually solved, however, and the volunteers selected for the first set experiments eventually received a series of packages which contained a cassette with a program on it, questionnaires and datasheets for them to fill in them and instructions on how to run the program.

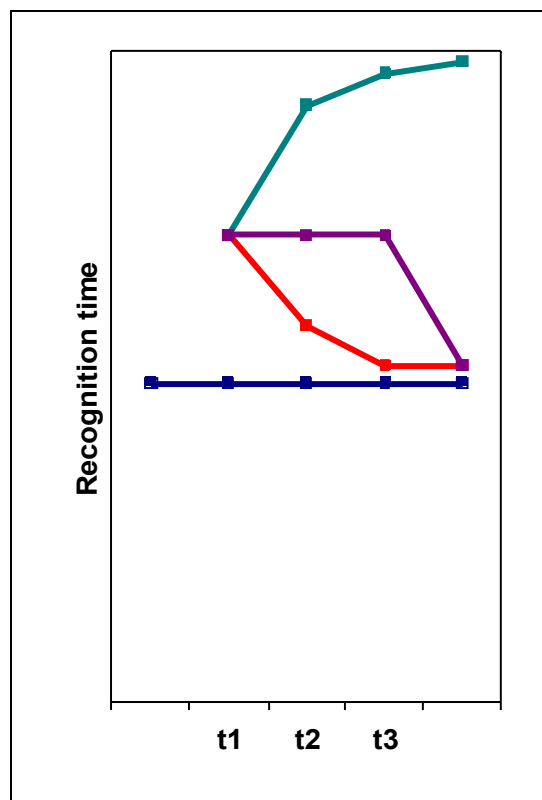
The main study carried out in 1983-84 was a longitudinal study in which the "progress" of a set of words was observed. Previous work by Lambert (1955) had suggested that the time speakers require to recognise a word was basically a function of how well they know the language. On the whole bilinguals are able to recognise words in their stronger language faster than they can handle words in their weaker, Lambert claims. This model seems to me to be intrinsically implausible and, it seems much more likely that many frequent words in the L2 can be handled as easily as frequent words in the L1, though obviously the relative amount exposure to each language might also be expected to have an effect.

Lambert's work seems to imply that it is possible to treat learners' vocabularies as two relatively separate wholes. It is not clear why he makes this assumption, however. We already know that there are significant differences to be found among the words of L1 speakers' lexicons. L1 speakers typically react differently to words of different lengths, different frequency, different degrees of spelling regularity, and so on, and it seems likely that similar differences will be found among L2 words. One important variable for an L2 seems to be how recently learned a particular word. Most L2 speakers of French seem to handle high frequency French words which they learned in school with a high level of facility. Relatively unfluent speakers of French (e.g. people did a school French course 15 years

previously and had travelled a bit since then) seem to recognise words like *femme* or *fenêtre* or *porte* as quickly as they can recognise the English equivalents. On the other hand, every learner can provide anecdotal evidence about the difficulty of recalling a word they learned only five minutes previously.

One way of approaching these phenomena is to elaborate models in which the acquisition of L2 vocabulary is seen not as an instantaneous occurrence, but something which progresses slowly over a fairly long period of time. You may be able to recognise a word you have just learned fairly easily, but it is not really an integral part of your personal word stock until you can handle it just as you would handle an L1 word, and this might take many weeks of constant use. Assuming that word recognition tests are a valid way of assessing how well the word is integrated into your personal word stock, models of this sort would predict that the time you needed to recognise an individual word in L2 would vary systematically as a function of time since learning. Such models fall naturally into three main classes, and these are illustrated in figure 1.

Figure 1:
Three models for word recognition in a foreign language



In this diagram, the average time necessary for a speaker to recognise an L1 word is shown by the dotted line. Since these words are reasonably well learned, and recognition time is reasonably stable, this line does not change height.

The first class of model for describing what happens to L2 words is represented by line A.

This model suggests that word recognition time decreases as a function of time, i.e. the longer ago you learned the word, and assuming the word is not forgotten since, you are more likely to handle it like an L1 word. This type of model raises two interesting questions. Firstly is it the case that L2 words asymptote out at the same level as L1 words, as is implied in the diagram? Lambert's work seems to imply that this is not the case, and that L2 words always remain at a higher level on the graph than L1 words, unless the two languages are properly balanced. The second question that this model raises is how long does it take for an L2 word to reach asymptote? At the moment we have no way of predicting whether the timescale for this change covers a few hours or several months.

The second class of model for describing what happens to L2 words is represented by line B. This model suggests that newly acquired words in the L2 are essentially stable, and that the time required to recognise them does not actually change a lot. At some stage however, a catastrophic change occurs, and the word suddenly crosses some threshold which causes it to be treated like an L1 word. Again, this model raises a number of subsidiary questions: what sort of experience might cause the catastrophe, and whether the catastrophe affects only individual words, or if it is part of a major restructuring of whole areas of the L2 word stock.

The third class of model for describing what happens to L2 words is represented by line C. This model suggests that words in the L2 may actually become harder to recognise as time passes. This may seem implausible at first glance, but it seems more likely if one considers the learners are typically exposed to large numbers of new words. It is relatively easy to recognise one member of a small set of items, but increasing the size of the set often makes recognition of individual members harder. In this case, it may become harder to recognise L2 words if the total L2 vocabulary increases too fast.

The obvious way to distinguish between the models is to run a longitudinal study in which a set of words learned early in a course is observed at intervals. Each model makes a different prediction. Model B corresponds with a null hypothesis, in which there is no change in recognition time, but recognition time should consistently exceed recognition time for L1 words. Model A and Model C both predict that recognition times will change over time. Model A predicts that recognition times will decrease, while Model C predicts that recognition times will increase over time.

A series of experimental studies designed to test the claims was carried out using 24 words taught in the first five lessons of *Digame*. All 24 words tested were six letters long. The frequency was not controlled in any way. There were two reasons for omitting to control for frequency. The principal reason was the standard frequency counts do not well reflect the use of individual items in a beginners' course. The second reason was that the only feasible alternative measure of frequency -- the actual occurrence of items on the course -- could not take account of individual variation in the use of the materials. Recognition times were obtained using the paradigm outlined above. Four experiments were carried out, but this interim report covers only the first three experiments of the series.

In Experiment One, each subject was tested on the 24 Spanish words and 24 English words of the same length and similar frequency. The experimental session was preceded by a lengthy practice session. Each subject's results were displayed at the end of the session, in

an easily comprehensible graphic form, and the actual results were then displayed on screen for the subject to copy onto the results sheet. These results were encrypted so as to avoid cheating.

The three subsequent experiments took basically the same form, except that Spanish words were used in place of the English ones. In these experiments, words learned later on the course were compared with the original 24 words learned in lessons one to five.

Results

The results of this set of experiments proved to be more difficult to analyse than had been expected. The initial pilot work suggested that the experimental paradigm produced results with a fairly small standard deviation and fairly small individual differences. This was certainly not the case with the results recorded here, however where very large individual differences were found, together with large differences between individual words. A close examination of the detailed results suggests that there was something like bi-modal distribution in the data. Almost all the responses to English words took less than four seconds, but responses Spanish words, while generally in the upper part of this range, showed a large proportion of very slow responses. In order to reduce the very large variances, we adopted a cut-off point of four seconds for the data reported in this analysis. This is not entirely satisfactory procedure, because it has a greater effect on the Spanish words than on the English words. It also has the effect of reducing the difference between the two types of words, and thus making the Spanish words seem more likely English words than they really were. Nevertheless, it seemed fairly clear that when subjects took a very long time to recognise word, whatever they were doing was something quite different from the apparently instantaneous recognition that seems to occur with L1 words. For the purposes of analysis then, scores of more than four seconds were treated as errors.

Table 1

Mean median recognition times for words in English and Spanish (in seconds)

English	Spanish(1)	Spanish(2)	Spanish(3)
1.47	2.15	1.74	1.55

Table 1 shows the mean median identification time for words in English Spanish. Figures for the Spanish words were collected in weeks 7, 12 and 17 of the twenty lesson course. The same data is also displayed in figure 2.

There is significant difference within the Spanish words [$F(2,29)=14.9$, $p<.001$]. This indicates that recognition times for Spanish words decrease with time. Closer analysis suggests that most of the variance on this dimension can be accounted for in terms of the difference between test 1 and the other two tests (2 and 3). Scores on Spanish(1) are significantly higher than those of Spanish(2) and Spanish(3), but the latter two scores do not differ significantly.

On the face of it, the results seemed to fit most closely the prediction made by Model A. That is it looks as though new words in a foreign language start off being recognised very slowly, but over a period time they are handled with greater facility. This facilitation period

seems to take about ten weeks, but by the end of this time, the L2 words are handled with the sort of facility one would normally expect of L1 words.

Two other considerations suggest that this interpretation is something of an oversimplification, however. In the first place, it should be borne in mind that the scores used in this analysis treat any score over four seconds as an error. Such errors were not included in the analysis, and as a result the figures presented in table 1 represented only those words which are responded to in a way which is broadly comparable with L1 words. In fact a large quantity of data is excluded on this criterion. The mean number of slow responses in each of the four tested is shown in table two. This table also shows errors of other sorts

Figure 2:

Mean median recognition times for 24 Spanish words over a period of 12 weeks.

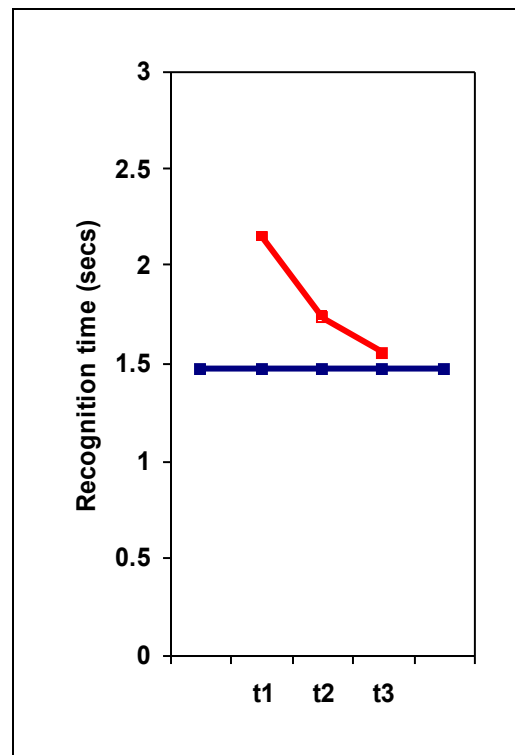


Table 2:

Mean number of responses over 4 seconds and errors

	English	Spanish(1)	Spanish(2)	Spanish(3)
Slow	2.0	4.5	3.5	1.3
Error	1.1	3.5	2.2	1.5

These figures for errors and slow responses show that a large proportion of the L2

responses were not included in the analysis reported here. The proportion of discounted responses declined with time until by the third test it is not very different from what one would expect of English words. To some extent, therefore, these figures confirm the analysis of the main results. There is a large difference between L2 and L1 words at the first test, but this difference gets smaller over testing period.

A more important problem is to be found in the detailed responses to individual words. The discussion so far has suggested that all L2 words start off being recognised very slowly, and gradually come to be handled in a more L1-like a way. Again a more detailed analysis of the results for individual words suggest that this conclusion is rather wide of the mark. From the data available so far, it appears as though almost all subjects produced faster recognition times for almost all words on the second test compared with the first. Recognition times on the second and third tests seem to vary in an unpredictable way, within certain loosely defined limits. The most likely interpretation of these figures is the words start off being recognised fairly slowly, but soon begin to behave like L1 words. On this interpretation, the relatively smooth curve which appears in figure two may actually be an artefact, in which series of discontinuous jumps has been smoothed out by averaging. Certainly, most of the action seems to have taken place between the first two tests, but the data available does not allowed to specify what the nature of this "action" is. In terms of the three models we discussed earlier, either Model A or Model B seems plausible. The process by which individual words move from relatively slow recognition times to reasonably faster ones could be a gradual process (Model A) or a sudden one (Model B). In either case, it looks as though the whole process is completed within five to ten weeks. The data currently available does not allow us to distinguish between these guesses. It should be possible, though, to establish which is the correct description by sampling recognition speeds at closer intervals and was possible here.

In addition to these reservations, which arise directly out of the work reported, another serious reservation arises out of the second series of experiments carried out the same subjects. These experiments monitored the way learners react to words that they met later on the course, and our main aim was to examine whether new words easier to handle as your vocabulary grows; whether all new words go through the same sort of processing as we found for the words tested in the first set of experiments, or whether learning new words actually gets harder as your L2 vocabulary increases. The most striking thing to come out of our preliminary analyses of this data is that there is a huge increase in the number of errors and failures to recognise words as the course progresses. Some words are handled relatively easily -- i.e. they are recognised in much the same sort of way as the new words in the early lessons of the course. Presumably, these words, like their earlier counterparts, are absorbed fairly quickly, though this claim is only a hunch at this stage, and needs to be properly tested. A very high proportion of words fail to be recognised in the four seconds available -- indeed, many of the test words fail to be recognised at all, even when the time available was extended to ten seconds. This finding suggests that the generalisations outlined above they apply only in the very early stages of learning language.

There are two possible explanations for this high failure rate. Firstly, it is possible the words introduced were peculiar in some way, and that this demotivated the students and inhibited their learning. For example, it is possible that the frequency characteristics of the words introduced in the later lessons showed a bias towards less frequent (and less useful) words.

This sort of explanation would suggest that how words are chosen for learners and how they are presented should be considered the main factors in determining the growth of vocabulary in L2. An alternative sort of explanation might be to suggest that the vocabulary presented in this course exceeds the number of words that learners of this sort can handle. In this case, we would be dealing with some sort of intrinsic limitation in L2 learners, rather than with an accidental effect of presentation or word choice. Maybe, for instance, learners can only handle 200 new words in a month, even under ideal conditions. Exceeding this limit would then produce an apparent failure rate of the sort we noted here. Again, given the data available at present, it is not possible to distinguish between these two types of explanation. It should be obvious however, that the methods reported in this paper could be used to sort out which type of explanation is best.

Conclusion

This paper has presented an interim report on a project where the acquisition of Spanish words by English learners was studied. The first part of the report outlined how the project used the new possibilities raised by widespread ownership of microcomputers, and discussed a new word recognition technique which was developed around this technology. The second part of the paper discussed interim results of one part of the study. The results suggest that in the early stages of learning, at least, word recognition times decrease gradually with experience of using the word. There was, however, some evidence to suggest that this may be in oversimplification. There are very large differences between individual learners, and individual words, both of which will require further investigation.

The main point to emerge from this report is that investigations of word recognition in L2 can provide interesting insights into the nature of vocabulary acquisition. A simple technique like the one discussed here is ideal for use with a highly motivated and willing group of subjects, such as adults following BBC courses, and raises the possibility of looking at vocabulary acquisition in a completely new light. We hope soon to be in a position to extend this project considerably and we intend to run a series of very large studies of vocabulary acquisition in a range of different languages in the near future.

References

Lambert, WE

Measurement of linguistic dominance in bilinguals. *Journal of Abnormal and Social Psychology* 50,2(1955), 197-200.

Meara, PM

Word associations in a foreign language *Nottingham Linguistics Circular*, 11,2(1982), 29-83.

Meara, PM

Vocabulary in a Second Language. London: Centre for Information on Language Teaching. 1983.

Meara, PM

Word recognition in foreign languages. In: **A Pugh and J Ulijn** (Eds.) *Reading for Special Purposes*. London: Heinemann Educational Books. 1984a.

Meara, PM

The study of lexis in interlanguage. In: **A Davies, C Criper and A Howatt** (Eds.) *Interlanguage*. Edinburgh: Edinburgh University Press. 1984b.

Notes:

This paper was first published in: **VJ Cook** (Ed.) *Experimental approaches to second language learning*. Oxford: Pergamon Press. 1986. 101-110.

The author would like to thank Maddalena Fagandini of the BBC for her help in making possible the project described in this paper.