Vocabulary acquisition: a neglected aspect of language learning.

Paul Meara

Introduction
Vocabulary acquisition is part of the psychology of second language learning that has received short shrift from applied linguistics, and has been very largely neglected by recent developments in research. This neglect is all the more striking in that learners themselves readily admit that they experience considerable difficulty with vocabulary, and once they have got over the initial stages of acquiring their second language, most learners identify the acquisition of vocabulary as their greatest single source of problems.

This article is an attempt to redress this neglect. It summarises the current work being done on vocabulary acquisition, and draws attention to a number of studies carried out by experimental psychologists which may have implications for the development of vocabulary in the second language. The article ends with a number of questions which have not been investigated in any depth, but which seem to me to be worth looking at more closely.

A: Bibliographies and general works
Despite the comments above about the general level of neglect in the study of vocabulary acquisition, there do exist a number of bibliographies of relevance to anybody working in this field. The most important of these is Dale and Razik 1963, a very extensive work, not primarily concerned with foreign language acquisition, but containing three relevant subsections with some 150 references. The 1963 edition of this work is actually a reworking of an earlier edition, and this may account for the fact that most of the references relate to second language work carried out in the 1930s. The most recent bibliography dealing specifically with L2 vocabulary acquisition is Twomey 1979. This work is patchy in its coverage, however, and a fuller bibliography, using a larger database, is in preparation (Meara, in prep.).

The principal impression that emerges from these bibliographies is that research in vocabulary acquisition has been largely atheoretical and unsystematic. There are no
clear theories of vocabulary acquisition, and the level of research activity is in general fairly low. Twomey, for instance, contains a large number of references which are short articles aimed at providing practical tips for teachers concerned with particular items of vocabulary for a particular target audience (e.g. Brown 1974; Martin 1976; Salt 1976; Ridout 1977), and on the whole, research in the field has avoided the serious theoretical questions that arise once one moves away from this very basic level.

A number of generalised approaches to vocabulary acquisition do exist. Galisson 1970 discusses vocabulary teaching at length, and considers the effectiveness of different types of vocabulary learning exercises. These ideas are expanded in Galisson 1979. An approach to vocabulary teaching based on contrasting lexical structure is to be found in Holec 1974 (cf. also Dagut, 1977, for a shorter argument on the same lines). Two other large-scale works, Dale, O'Rourke and Bamman 1971 and O'Rourke 1974 are also worth mentioning at this stage. They are both concerned exclusively with first language vocabulary development, but the ideas discussed are clearly of some relevance to second language learners.

The also exists a number of shorter articles which draw attention to the need for further work on vocabulary acquisition. Marton 1977 discusses the problems of idioms, which he sees as the biggest obstacle to fluent comprehension in advance learners. Lord 1974 draws attention to the importance of Leopold's 1948 study of semantic development in a bilingual child (cf. also Yoshida 1978 for an empirical study along these same lines). A brief but excellent critique of vocabulary research is Levenston 1979, who reiterates a number of points made in this paper, criticising applied linguistics for its general neglect of vocabulary learning in favour of the study of syntactic development. Levenston discusses a number of research projects which have only recently begun, and which are therefore difficult to assess satisfactorily. These include Levenston's own work, (Blum and Levenston 1978 and Levenston and Blum 1977), where it is argued that the lexical simplification strategies used by learners may follow universal rules; and Linnarud 1979, who suggests that foreign language learners may have characteristically low levels of vocabulary richness compared to normal speakers. (On this use of type token ratios as a measure of learners' vocabulary cf. also Webber 1977, Philpot 1977 and Meara 1978.) All these pieces of work are small-scale, however, and none of them adds up to a coherent and impressive body of knowledge at this stage.

Apart from these general works, and a few small pockets of isolated research work mentioned above, there also exists a number of fields which have been or are being investigated in a fairly systematic fashion, and are thus important because they comprise the bulk of the work on vocabulary acquisition.
B: Vocabulary control
The best developed and most systematic work in the field is to be found in attempts to justify the selection of vocabulary items for inclusion in courses and examinations on the basis of frequency counts and similar objective measures. This work is too well-known to need discussion here. Good summaries can be found in Bongers 1947; Mackey 1965; and Syracuse University Research Corporation 1973.

Despite their obvious merits, frequency counts are now very much out of favour with applied linguists (cf. Wilkins, 1972, for a fairly typical criticism of this sort approach). In contrast to the very carefully chosen vocabulary of earlier course books, many modern books appeared to rely almost entirely on subjective assessments of the usefulness of words. Van Ek (1977), for instance, contains no discussion of what criteria were used for the inclusion of words in the Threshold Level vocabulary (cf. also Hoffman 1976). In general, language teachers seem to be unaware of more recent developments in word counts. West’s General Service List 1953 is often cited by publishers and examining boards as guiding their choice of words in readers and examinations for learners in English, for example, despite the fact that both this list and the related Thorndike-Lorge list have now been superseded by the more recent Kucera-Francis count (Kucera and Francis 1967). Recent counts for a number of other languages commonly taught to learners also exist, and ought to be more widely known. These include Juillard and Chang-Rodriguez 1964 for Spanish; Juillard, Brodin and Davidowitz 1970 for French; and Juillard and Traversa 1973 for Italian. An important feature that differentiates most of these modern counts from their earlier counterparts is the use of computers in their preparation. All the early counts were carried out by hand (the Thorndike-Lorge count was in fact begun during the Depression as a way of providing work for the unemployed), and were accordingly both expensive and slow to reach completion. The rapid processing facilitated by computers makes it possible to produce word counts at minimal cost, and to keep them regularly updated.

This use of computers to carry out simple statistical analyses of texts is a development which is likely to be of some importance to language teachers, as it has considerable implications for the preparation teaching materials. Computer programs which will do word counts and similar basic statistics on continuous text have existed for some time, and they are now beginning to appear in the form of easy-to-use packages, designed for amateurs with no real experience of computer programming, and they require only minimal instruction before they can be used. An excellent example of an easy to use package of this sort is the Oxford Concorance Project (Hockey and Marriott 1979-80; Burnard, Hockey and Marriott 1979). This package produces basic word counts for texts of any reasonable length, alphabetical listings, frequency listings, concordances with
contexts of specified lengths, and so forth. Though the package is primarily aimed at literary scholars, its uses are of course not limited to literary texts. It could also be used, for example, to provide accurate frequency counts of the vocabulary used in 'special-purpose' situations or to prepare a glossary to accompany set text, and so forth. Use of tools of this kind seems likely to become increasingly important in the preparation of teaching materials (cf. Lyne and 1975; Culhane 1977; and Johnson 1972. An interesting general introduction to this sort of work is Morton 1979).

This outline of research into vocabulary control and selection has been deliberately brief, and is not intended to be a comprehensive one. On the whole this work is well-known and reasonably familiar. My main reason for raising it here is because work of this type has played such a preponderant part in the study of vocabulary acquisition that no review would be complete without at least a cursory mention. However, the work is also important because it illustrates two aspects of research into vocabulary acquisition which are characteristic of the field is whole, and therefore deserve further comment. In the first place, this work is characteristic in that it concentrates on what is basically a problem to do with the management of learning, rather than with the learning process itself - i.e. the object of this type of research is to decide what words are to be taught, not to find out how words are actually learned. This is an important point, and will be returned to later at the end of the next section. In the second place, this work also illustrates how easy it is to fall into the trap of accepting uncritically a whole set of assumptions, and to design a large research programme around these assumptions, without ever calling their validity into question. In this case, the central assumption is that it really is necessary to place a severely restricted upper limit on the number of words that the learners can reasonably be expected to acquire in a foreign language. Some simple arithmetic indicates that a vocabulary of 2000 words could be learned in 11 months if new words were acquired at a rate six per day. This figure does not appear to be wildly excessive, given what we know about the capacity of the brain to acquire new information. Nevertheless, most teachers would undoubtedly consider 2000 words to be well beyond the capacity of many learners, even over a five-year course instruction (cf. for instance, Wicklow 1974, and Barnard 1971, where these assumptions are made quite explicit). No doubt there is some practical justification, based on experience, for this general belief that learners cannot easily acquire a large foreign language vocabulary in a short space of time, but the theoretical basis for this agreement is by no means clear. This is obviously an area in which further research would be most useful.

C. Mnemonics
The second major area of research to be discussed is one which has, in a way, addressed itself to this problem, challenging the assumption that massive vocabularies cannot be
acquired by introducing mnemonic techniques into the teaching of vocabulary.

The best studied of these methods is the 'keyword' method which has been the object of considerable attention in the last few years particularly in the United States (Atkinson and Raugh 1975; Atkinson 1975; Raugh and Atkinson and 1974; Pressley 1977; Pressley and Levin 1978; Pressley (in press); Singer 1977). In this method, the target language words are associated with phonetically similar L1 words (called keywords) in the first stage of learning, and then, in the second stage, these keywords are associated with the L1 translation of the original target language word by means of a striking visual image. Thus for example Raugh and Atkinson suggests that the Spanish word CABALLO (pronounced cob-eye-oh [sic]) might be linked to English EYE and EYE to HORSE via the image of horse with a great side cyclopean eye in the middle of its forehead. Or more prosaically, CABALLO might be linked to the English word CAB, which in turn would be linked to HORSE via the image of horse drawing a cab. The papers listed above report a number of experiments which compare more traditional ways of learning vocabulary (e.g. list repetition) with this keyword method, and despite its initial implausibility, present some impressive results in support of this sort of practice. Raugh and Atkinson (1974), for example, report that learners using the keyword method can cope with very long list of words (60 items) and still get 80 percent correct on a subsequent test, a figure that is considerably better than that produced by learners using repetition and rehearsal methods. More importantly, the keyword groups preserve their advantage over time, and show less evidence of forgetting than is found with control groups.

This evidence is very impressive at first sight, but work of this kind is actually rather problematical at a deeper level, and needs to be treated with some caution. The most obvious problem is that experiments of this sort treat vocabulary items as discreet pairs of translation equivalents, and completely ignore the complex patterns of meaning relationships that characterise a proper, fully formed lexicon, as opposed to a mere word list. Learning vocabulary is not just a matter of acquiring translation equivalents: it is well-known that languages rarely map their lexical items onto each other in a one-to-one fashion. Some lexical structuring must go on even when the shortest word list is learned, and any view of vocabulary acquisition which treats the problem as a simple matter of pairing words with their translation equivalents is an oversimplified one, which cannot adequately account for how these semantic relationships are built up in a foreign language vocabulary. All the experimental studies of vocabulary acquisition which make use of mnemonic devices are basically subscribing to model of 'paired associate learning' which does not seem to me to be sufficiently rich to account for what is involved in the acquisition of a second language vocabulary (cf. for example Crothers
A second problem that characterises the keyword studies is that they are generally one-off experiments that do not study real language learners in the course of learning a language, but only subjects prepared to take part in a small number of experimental sessions in a laboratory setting. This means that the subjects tested often have initial vocabulary of zero, and it is unclear how far and results in such experiments might be generalisable to more advanced learners. More importantly, it means that the comparison between the keyword method and other methods is usually limited in practice to methods that can also be contained within a single experimental session, such as rote learning by repetition. Whether the keyword method is in the long-term more effective than other methods which are not readily comparable with it, such as the Silent Way, or Total Immersion, or even methods which place special emphasis on vocabulary acquisition such as Gouin’s Series Technique (Gouin 1880) or Barter’s Comparative Method (Barter 1970; Becker 1977), is typically a question which is not asked. Indeed, even within its own terms of reference, the comparisons made are rather spurious. It is difficult to imagine that even the most ardent believers in rote learning methods might require their students to use this method with lists containing 60 items. It seems important, then, that these laboratory tests should be complemented by properly controlled longitudinal classroom tests, before their findings are widely accepted.

A third problem is that the keyword method is used principally in situations where the target language word is required to evoke the native language equivalent. The phonetic link idea seems to work reasonably well in this respect, and the method does seem to have some value as far as recognition vocabulary is concerned. The value of the method for developing active vocabulary is much less clear, however, and it seems likely that the practice of stressing crude phonetic similarities between L1 words and target language words would in the long-term have a serious detrimental effect on the pronunciation of target language words.

A final point to be made about this work is that Raugh and Atkinson, at least, have a highly directive approach to the choice of keywords. One might have expected that this choice could easily be left to the individual learner, but Raugh and Atkinson actually go to some lengths to stress that not all keywords are equally effective, and only keywords shown by extensive research to be effective should be used.

A more recent mnemonic technique is the Hook Word technique, described by Paivio 1978 and Paivio and Desrochers 1979. This research is still in its infancy, but seems to
suffer some of the same problems as the keyword method. The same criticism also applies to Ott, Blake and Butler 1976, who report a number of studies using 'elaborative techniques' (Lado, Baldwin and Lobo 1967; Butler, Ott and Blake 1973; Knop 1971; Groberg 1972; Holley 1971; cf also Setzler and Clark 1976).

D. General comments
So far we have covered two areas of research in the broad field of vocabulary acquisition. These two areas may at first sight appear to be relatively unconnected. In fact, however, they are both linked in that they share common defect: their concern with the peripheral aspects of vocabulary acquisition rather than central ones. In the case of research into the uses of frequency counts, we have already seen how the management of learning rather than an understanding how learning takes place and what it involves, is the principal driving force behind the investigations reported. Basically, the learners' load is lightened for them by working out beforehand which words are likely to be of use to them, and which ones are not worth the trouble of learning. What happens to the words that are learned is not in question. With mnemonics, too, the emphasis is placed squarely on the management of the learning process, the main theoretical questions asked being ones concerned with the effectiveness of different forms of presentation.

Both these approaches, then, are concerned with what is basically the periphery of acquiring new vocabulary. Yet it must be obvious that these peripheral aspects, important though they may be, leave unanswered a large number of questions which are of considerable relevance to our understanding of how vocabulary is acquired. Learning new words is not an instantaneous process - if it were, and if presentation were the only critical variable involved, then words would not be forgotten and need to be re-learned. As it is, however, it seems that words are absorbed slowly over time, and that only gradually do they become fully integrated into the learners' personal stock of words, when they can use them with the same sort of fluency and that characterises the words they use in their native language. Some work that is relevant to this rather more difficult question is discussed in the sections that follow.

There does exist a reasonably large body of experimental work which has attempted to investigate how bilingual speakers store words in their mental dictionaries. These studies may not appear to be of direct relevance to language teaching and language learning, but in fact their relevance is often greater then appears at first sight. Their importance lies in the fact that they provide us with some clues about what the end product of learning a foreign language might consist of, and what sort of behaviour can be expected of a fluent bilingual. Information of this kind should, in theory at least, enable us to compare the behaviour of non-fluent bilinguals - i.e. language learners -
with that of fluent bilinguals, and so to make inferences about the way a developing vocabulary in a second language grows. Basically, this work can provide us with a model, albeit a sketchy one, against which to assess the more limited abilities of less fluent learners.

Most of the work to be described in the next sections is not ostensibly concerned with foreign language learners; the subjects used in these experiments are usually 'balanced bilinguals' - i.e. speakers who are judged to be equally fluent in both their languages. The main purpose of this research has been to compare two types of bilinguals - compound and coordinate - a distinction first drawn by Ervin and Osgood 1954. Ervin and Osgood argued that the way in which a language was acquired might be expected to produce different types of structuring in a bilingual's mental dictionary. Compound bilinguals, those who have acquired the two languages together in a single environment, usually in infancy, might be expected to have a single set of meanings tagged by two sets of labels, one for each language. In contrast, Ervin and Osgood argued that coordinate bilinguals, who learned their two languages in largely separate environments, might develop what are effectively two separate lexicons, one for each language, which function independently of each other.

Again, this work may not seem to be directly relevant to second language acquisition, but the relevance is there, nonetheless. Lambert has argued that certain types of teaching method may be more likely to produce behaviour characteristic of one type of bilingual, rather than the other. Direct Method teaching, for instance, which tries to eliminate reference to a learner's mother tongue, might be more likely to produce the learners with the characteristics of coordinate bilinguals. Experimental evidence to support this claim is not available, however, and a rather different claim has been put forward by Riegel 1968. Riegel argues that this dichotomy is a false one, and that there is actually a natural development from a sort of compound system to one which is closer in kind to the classical coordinate model, irrespective of the language acquisition background.

The reason why this work is of some relevance to second language acquisition is that many of the 'balanced bilinguals' the used as subjects in these experiments are actually far from equally fluent in their two languages. In fact, the criteria used to decide on an acceptable level of ability in the second language are often ill defined and crude, and in practice the term 'bilingual' can mean anything from fully and equally fluent in both languages, to someone who has only barely begun to acquire their second language and could not be considered equally fluent by any stretch of the imagination. The unsatisfactory nature of the tests used to measure second language fluency (often only
self-ratings, or impressionistic judgements made by the experimenter) is a major flaw in this literature. It does mean, however, that a number of the studies, despite their pretensions, are actually concerned with ordinary language learners.

The work to be considered in the next sections fall into two parts: (a) experiments concerned with the general question of whether the bilingual’s lexicons are separate or independent, and (b) experiments concerned with the semantic relationships that exist between words in the bilingual’s lexicons. A third area - the ability of bilinguals to handle words in each of their two languages in very demanding situations such as tachistoscopic recognition tasks, or dichotic listening tasks, and so forth - will not be considered here, due to lack of space. Interested readers are referred to Albert and Obler 1979, an excellent book with a superb bibliography (though cf. Green and Newman 1980 for critical review). This work suggests that there may be major differences between stronger and weaker languages, and provides some evidence for the claim that second language words may be processed less effectively by certain parts of the brain. In particular, a number of studies suggest that there are hemispheric asymmetries for different languages (cf. Walters and Zatorre 1978 for Spanish; Hamers and Lambert 1977 for French; Kershner and Jeng 1972 for Chinese; and Orbach 1953 for Hebrew). This type of research is becoming increasingly sophisticated and influential and looks like becoming one of the major growth areas in psycholinguistics in the immediate future.

E. Memory experiments
The work to be reviewed in this section consists of a number of experimental studies which have all attempted to test the claim that bilinguals have two separate, independently functioning lexicons, rather than a single fully integrated one.

For reasons which are not wholly clear, this claim seems to have been most often investigated by the use of memory tasks, and in particular by the use of tasks where interference from one language to another is observed. This is obviously a fairly crude tool, and the results found in these experiments are correspondingly limited. The general line of argument is that if bilinguals’ two sets of words were stored totally independently, then very little interference would be expected in tasks that require subjects to use both of their languages. Where interference is found, this is usually interpreted as supporting the claim that the two languages function interdependently, and are not wholly separated.

The more important studies of this kind are summarised in table 1. Useful discussions of this work will be found in Albert and Obler 1979 and McCormack 1977. The experi-
### Table 1: List learning and recall tasks

<table>
<thead>
<tr>
<th>Author</th>
<th>Langs</th>
<th>Task</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lambert</td>
<td>French</td>
<td>Learn English word list;</td>
<td>For compound bilinguals learning a French list facilitates recall of English words. For co-ordinates, both interference tasks are equally disruptive.</td>
</tr>
<tr>
<td>Havelka</td>
<td>English</td>
<td>Learn French word list or learn nonsense word list</td>
<td></td>
</tr>
<tr>
<td>Crossby 1958</td>
<td></td>
<td>Recall English word list</td>
<td></td>
</tr>
<tr>
<td>Ervin 1961</td>
<td>Italian</td>
<td>Name pictures in English</td>
<td>For compound bilinguals, dominant in Italian were easy to recall. Objects named fastest in English were recalled equally well in either language.</td>
</tr>
<tr>
<td>Langs or Italian</td>
<td>Practice with Italian</td>
<td>Recall original lists</td>
<td></td>
</tr>
<tr>
<td>Ervin</td>
<td>English</td>
<td>Name pictures in English</td>
<td></td>
</tr>
<tr>
<td>Kolers 1965</td>
<td>French</td>
<td>Recall of mixed language word lists with some repeated items</td>
<td>Interlingual and intralingual repetitions both increased the probability of recall.</td>
</tr>
<tr>
<td>Lambert</td>
<td>English</td>
<td>Recall of monolingual or mixed language lists, semantically categorised or not</td>
<td>Greater recall in the Ss' stronger language.</td>
</tr>
<tr>
<td>Ignatow 1968</td>
<td>Russian</td>
<td>Recall of mixed language lists</td>
<td>No effect for mixed language lists, but large numbers of translation errors are found when categories and languages conflict.</td>
</tr>
<tr>
<td>Nott 1968</td>
<td>French</td>
<td>Recall of mixed language lists containing semantically related items or not</td>
<td>Categories crossing languages produced worse recall.</td>
</tr>
<tr>
<td>Lambert 1968</td>
<td>English</td>
<td>Recall of mixed language lists containing semantically related items or not</td>
<td></td>
</tr>
<tr>
<td>Tulving 1970</td>
<td>French</td>
<td>Recall of mixed language lists</td>
<td>No difference in primary memory; large differences in secondary memory with the most proficient language being most impaired on multilingual lists.</td>
</tr>
<tr>
<td>Colotla 1970</td>
<td>English</td>
<td>Recall of mixed language lists</td>
<td></td>
</tr>
<tr>
<td>Kintsch 1970</td>
<td>German</td>
<td>Words presented in lists containing repetitions. Ss are asked to judge whether each word was new or repeated.</td>
<td>More accurate performance in the dominant language. Ss found it hard to treat translations as new words.</td>
</tr>
<tr>
<td>Wickens 1971</td>
<td>English</td>
<td>by a final list where one of the list characteristics is changed</td>
<td>Switching languages improves performance for balanced bilinguals, but greater dominance produces a smaller effect.</td>
</tr>
<tr>
<td>Saegert 1973</td>
<td>Spanish</td>
<td>Learn a short list then a longer list that contains items from the first list in a different order.</td>
<td>Learning the part-list in L1 and the whole list in L2 produces negative transfer. Positive transfer is found in the opposite condition.</td>
</tr>
<tr>
<td>Kazararian 1973</td>
<td>English</td>
<td>Recall of mixed language lists categorised semantically</td>
<td></td>
</tr>
<tr>
<td>Champagnol 1973</td>
<td>French</td>
<td>Recall of mixed language lists</td>
<td>Better recall in French, worse recall in mixed language lists, plus the usual category effects.</td>
</tr>
<tr>
<td>Lopez 1974</td>
<td>Spanish</td>
<td>Learn list A</td>
<td>When list B contained items translated from list A recall was better, though more so if list B was in English.</td>
</tr>
<tr>
<td>Young 1974</td>
<td>English</td>
<td>Learn list B</td>
<td>Translation errors are frequent where a language change was made.</td>
</tr>
<tr>
<td>Lopez</td>
<td>Spanish</td>
<td>Recall List B</td>
<td></td>
</tr>
<tr>
<td>Hicks 1974</td>
<td>English</td>
<td>Learn list of English word pairs, then a new list where items or language are changed</td>
<td></td>
</tr>
<tr>
<td>Young 1974</td>
<td>Arabic</td>
<td>Repeat learning of lists drawn from either a monolingual or a bilingual pool.</td>
<td>Performance deteriorates more in the bilingual condition.</td>
</tr>
<tr>
<td>Liepmann 1974</td>
<td>English</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
mental methods may seem rather obscure, but in general they are techniques which have been widely used in experiments with monolingual subjects, and which produce robust results which are reasonably well understood.

The bulk of the work reported in table 1 is generally taken as supporting the interdependence position, rather than the independence position: i.e. it seems to support the claim that words in a second language are indeed integrated in some way with words in the first language to form a complex whole lexicon. The logic of this argument is not always as clear as it might be, however and some of the data argues strongly against this position. Tulving and Colotla 1970 for example showed that in mixed language lists of words, recall was much worse than in single language lists, and that the greater part of this deficit seems to be due to the fact that L1 words are handled less satisfactorily in secondary memory than words from weaker languages. This finding is at odds with some of the other findings reported, and is not what would have been expected, but there are no obvious faults with this experiment, and it therefore needs to be taken seriously. Some of the other experiments, particularly the ones that use Spanish as one of the languages tested, are rather less satisfactory and need to be treated with some caution. The subjects in these studies are usually schoolchildren brought up speaking Spanish as their mother tongue, but being taught English as the medium of instruction in American schools. Most of the subjects rated themselves as more fluent in English than in Spanish, and were thus classified as English dominant, but this seems rather implausible. The use of self-ratings with subjects such as these seems to be an unreliable method of assessment in view of the obvious pressure that subjects must be under to overrate their competence in English and to underplay their abilities in Spanish.

Generally speaking, the results of these studies show that subjects are aware of the language in which words are presented, and that they use language as a classifying label more successfully than they use other more arbitrary coding features such as colour, but this last finding is hardly surprising in view of the artificiality and unfamiliarity of the colour coding task (cf. McCormack 1976). More interesting is the finding that it is much harder to remember the language of presentation when languages and semantic categories are confused, and the finding that subjects have some difficulty in recognising words in a list as new ones when they have previously been presented in translated form. These results show clearly than when some kind of cognitive operation other than simple recall of the phonetic form is called for, it does become extremely difficult to keep two languages apart. In this sense, forms in one language clearly evoke the corresponding related forms in the other language, a finding which would be very difficult to explain if the independent lexicons claim were true.
F. Semantic tests
The work reported in the previous section is rather unsatisfactory in that it all treats the bilingual’s lexicon as though it consisted of one or two undifferentiated wholes about which it was possible to make sweeping generalisations. There is, of course, no reason to assume that all the words in the lexicon should behave in the same way as each other, and just as words are differentiated in the lexicons of monolingual speakers in terms of frequency, length, and other similar objective characteristics, it is quite likely that similar features may produce subsets of words in the bilingual’s mental lexicon which also have quite different behavioural properties. Perhaps, then, it is inappropriate to ask whether the bilingual’s two words stores are integrated or independent as wholes, and it might make much more sense to assume that some words will be integrated while others will not, and thus to shift the focus of attention to individual words and relatively small semantic fields.

Some work in this kind has been carried out, and a good example of what can be done is to be found in a number of studies using the Stroop test with bilingual subjects (Stroop 1935). In this technique, subjects are given lists of words to read which have been printed in different colours. Typically three sets of words are used: (a) a set of neutral words or colour patches; (b) a set of colour words congruent with the printed colours e.g. red printed in red, green printed in green, etc.; (c) a set of colour words that are incongruent with the printed colours, e.g. red printed in blue, green printed in red, and so forth. The subjects’ task is to ignore the actual printed words and state the colour that the word is printed in. Thus, given red printed in green ink, the correct response is ‘green’ and not ‘red’. Task (c) proves to be extremely difficult in practice, since the colour words interfere with the naming of the printed colours. It is relatively easy to introduce a bilingual variation on the scheme, by producing further sets of words that contain the names of colours in the subject’s second language. Again the stimuli are either congruent or incongruent with respect to the colours of the print. Thus (d) rouge written in red, or bleu written in blue are congruent in French, while (e) vert written in blue, or rouge written in green would be incongruent. Subjects can then be asked to call out the names of colours either in their first language or in a second language, and this introduces an additional layer of interference.

A number of languages have been studied using this technique: Dalrymple-Alford and Budayr 1967 (English and Arabic); Dalrymple-Alford 1968 (English and Arabic); Preston and Lambert 1969 (English, French, Hungarian and German); Dyer 1971 (English, Greek, Italian, French, German and Spanish); Albert and Obler 1979 (English and Hebrew); Evans and Townsendson 1979 (English and Welsh). Hamers and Lambert 1972 who use an auditory version of the Stroop test in English and French, are also worth mentioning.
here. What this work shows, without exception, is that interlingual conditions do show strong evidence of interference, though less than would be expected in the case of monolingual subjects. Dyer showed that the level of interference was in part related to the phonetic shape of the words. Thus, for English speaker, for example when asked to name the word printed in red ink, the Spanish word \textit{azul} would be much less disruptive than the French \textit{bleu} because of the similarity between \textit{bleu} and the English word \textit{blue}. Other things being equal, however, the amount of interference caused by the foreign language words appears to depend on the level of proficiency of the subjects, but this correspondence has been measured only in very gross terms. Again, the logic of these data is that they support the idea that bilinguals’ two lexicons are integrated into a single whole. However, if bilinguals’ colour words in a second language were totally integrated with their L1 colour words, then one would expect to find Stroop interference equally great in either language. In as much as the weaker language produces less interference, this suggests that words in second language are not fully integrated at the semantic level, or at least than they are not treated with the fluency that is usually accorded to first language words. A similar argument based on data using semantic differentials in two languages is to be found in Jakobovits and Lambert 1961 and Lambert and Jakobovits 1960.

One of the most accessible and most easily understood methods of studying the structure of semantic relationships in bilingual lexicons is the use of word associations. In its simplest form, this technique involves the presentation of a number of single words to the subjects participating, and they are then instructed to reply with the first word that each of these stimulus words makes them think of. The bulk of the responses produced in tasks of this sort are noteworthy principally for their banality, at least as long as unemotive, common words are used. Far from being original, most people’s responses are characteristically shared with a large proportion of the rest of the population of normal adult native speakers (cf. Pollio 1966; Deese 1965; Postman and Keppel 1970). In English, for example, only 70 percent of people produce \textit{WHITE} in response to \textit{BLACK}, \textit{WOMAN} in response to \textit{MAN}, \textit{BUTTER} in response to \textit{BREAD}, and so on. This phenomenon, known as associational stereotypy, is found in all languages that have been investigated, though the absolute levels of stereotypy vary from one culture to another. English has particularly high levels of stereotypy compared to other languages. French, German, Italian and Polish are all significantly less stereotyped than English is (cf. Rosenzweig 1961; Szalay and Deese 1978; Vikis-Freibergs and Freibergs 1976; Kurcz 1966).

In addition to being very similar to the responses produced by other members of a similar population of subjects, responses to common words generally fall into one of two
major classes. *Paradigmatic* responses are responses that belong to the same major form class as the stimulus word. Thus, CAT, BONE, TAIL, COLLIE would all be classed as paradigmatic responses to the stimulus DOG. *Syntagmatic* responses are responses that typically combine with the stimulus words to form a phrase or syntactic whole. Thus, BITES, BARKS, FIERCE, STUPID and HOT would all be syntagmatic responses if produced in response to DOG. Normal adult native speakers tend to produce many more paradigmatic responses than syntagmatic ones, at least as long as the stimulus words are fairly common words. Infrequent words, such as ABDICATE or INITIATIVE are more likely to produce syntagmatic response forms such as KING or TAKE (Stoltz and Tiffany 1972). A good account of these typical responses is to be found in Clarke 1970, who also provides a good basic bibliography.

These normal response patterns are not preserved in abnormal native speakers (de Wolfe 1971) nor are they found in children. This latter group tends to produce response patterns that differ markedly from those of adults in than they contain a large number of syntagmatic associations in place of the more normal paradigmatic ones. Children also tend to produce lots of *clang associations*, i.e. responses that are phonetically related to the stimulus word, but fail to have any clear semantic connection with it. Typical examples of clang associations would be FIGHT, TIGHT and WHILE or WIDE produced in response to WHITE (Ervin 1961; Entwhistle 1966; McNeill 1963; Entwhistle, Forsyth and Muss 1964).

Table 2 below summarises the main studies of word association behaviour in which foreign language learners and bilinguals were used as subjects. These studies are very disparate in nature, and it is therefore rather difficult to summarise the findings briefly, or even to compare one study directly with another, since often the principal questions which the data were collected in order to elucidate do not have much in common. This variation can be seen in a large number of different entries in the column headed *subjects, methods and stimuli*. These differences are explained below.

*Explanatory notes on table 2*  
(a) *materials used*. There is almost no agreement over what sort of stimulus words to use in studies of word associations. Most of this work has used unstandardised and apparently unmotivated lists of words, some so extremely small in number that one is forced to wonder whether results based on these stimuli can be considered at all generalisable. The standard list of stimuli is the Kent-Rosanoff list, first used by Kent and Rosanoff 1910 in their study of word associations produced by mental patients. This list has the advantage that it has been vary widely used in a large variety of studies.
Table 2. Principal studies of word associations in a second language.

<table>
<thead>
<tr>
<th>author</th>
<th>language</th>
<th>subjects</th>
<th>main question</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lambert 1956</td>
<td>English, French</td>
<td>N=42 native speakers and students sixteen words in each language continuous associations 45 seconds</td>
<td>compared responses produced by groups at different levels of proficiency in terms of stereotypy and quantity</td>
</tr>
<tr>
<td>Kolers 1963</td>
<td>German, Spanish, Thai, English</td>
<td>N=38 10 Germans 14 Spanish 14 Thais 55 nouns in five different categories from the Palermo norms single responses</td>
<td>compares interlingual responses with intralingual ones, esp. frequency of translations</td>
</tr>
<tr>
<td>Lambert-Moore 1966</td>
<td>English, French</td>
<td>136 English speakers, 206 French speakers, 88 bilinguals Kent-Rosanoff list single responses</td>
<td>stereotypy and equivalence of responses</td>
</tr>
<tr>
<td>Macnamara 1967</td>
<td>Irish, English, Latin</td>
<td>72 bilinguals, 30 seminarians 3 words continuous associations with language switching. Three minutes</td>
<td>linguistic independence - compares switching conditions and how this effect the number of responses</td>
</tr>
<tr>
<td>Davis Wertheimer 1967</td>
<td>English, French</td>
<td>N=59 university students, postgrads and native speakers 8 English words, 8 French words, 8 ambiguous words continuous associations 15 seconds</td>
<td>relationship between number of responses and level of competence</td>
</tr>
<tr>
<td>Riegel, Ramsey 1967</td>
<td>English, Spanish</td>
<td>N=48 24 Americans, 24 Spanish 35 common nouns from the Kent-Rosanoff list restricted associations</td>
<td>stereotypy levels and overlap of responses in two languages</td>
</tr>
<tr>
<td>Lambert Rawlings 1969</td>
<td>English, French</td>
<td>20 bilingual students 60 French and 60 English words core concepts</td>
<td>and tea and different types of bilingual to recover stimulus word given a list associates</td>
</tr>
<tr>
<td>Dalrymple-Alford Aamiry 1970</td>
<td>English, Arabic</td>
<td>English/Arabic bilinguals 12 words from the Jenkins and Palermo norms and their Arabic equivalents single responses</td>
<td>stability of responses</td>
</tr>
<tr>
<td>Ruke-Dravina 1971</td>
<td>Swedish, Latvian</td>
<td>N=40 13 Swedish, 16 Latvian, 11 young bilingual 4 words in Swedish and Latvian continuous associations 5 minutes</td>
<td>total number of different responses and quality differences between the groups</td>
</tr>
<tr>
<td>Taylor 1971</td>
<td>English, French</td>
<td>30 undergraduates learning French 18 English and 18 French words continuous associations with</td>
<td>compares the effect of different switching rates</td>
</tr>
</tbody>
</table>
and there thus exist well-documented sets of response norms for these 100 words, covering different dates of collection, geographical locations, and types of subjects providing the responses. This word list has also been translated into a number of languages other than English (Rosenzweig 1961 for French, Italian and German; Haworth 1979 for Spanish) and this makes it possible to compare the responses of native English speaking learners of these languages with responses that would be expected of a normal native speaker population. The principal disadvantage of using the Kent-Rosanoff list is that a large proportion of the words that make it up consist of high frequency words which produce highly stereotyped response patterns which do not vary greatly from one language to another. This means that only a small subset of the list is of any real interest where the main purpose of the study is to make cross language comparisons.

(b) method. The standard word association method has already been described. A number of variations on this basic theme will be found in table 2, however. The chief of these is the continuous association method in which a single word is presented as the stimulus but instead of providing only a single word in response, subjects are required to produce a continuous stream of responses for a given length of time (usually in the region of one minute). These responses are then pooled and counted as in the standard
measure. This technique produces patterns of responses that closely resemble those produced with the standard method using single responses. The continuous association method has the practical advantage that it is viable with a relatively small number of subjects, however. Deese 1965 claims that a minimum of 50 subjects is necessary to produce stable norms for English using the standard method, and there is some evidence that other languages may require even larger numbers of subjects than this. With the continuous association method, however, stable response patterns can be found with as few as 15 subjects.

(c) *interlingual associations*. These are associations made in a language that is not the same as the one in which the stimulus word was presented.

(d) *restricted associations*. This term describes a technique principally used by Riegel and his associates. It consists of a basic elicitation technique for single responses, but with limitations imposed on the type of response allowed. Thus, for example, subjects might be instructed to produce a response that could describe the stimulus, or one that was a superordinate of it.

(e) *language switching*. This variation is used only in conjunction with the continuous association method. Subjects are required at specified intervals to stop producing associations in one language, and to change to their other one.

Despite these important differences in method, it is possible to draw some general conclusions from the studies of word association behaviour in bilinguals and foreign language learners. Firstly, on measures of fluency, such as number of responses, speed of responses and so forth, bilingual speakers are less adept in their weaker language than in their stronger one. Secondly, responses in a weaker language tend to be strikingly less stereotyped than responses in a stronger language. This finding is odd in view of the fact that learners must have smaller vocabularies than native speakers, and this would lead one to expect that the range of possible responses would be correspondingly more restricted and less variable. The general inference from these two points is that words in a second language are less well-organised and less easily accessible than those in the mental lexicon of native speaker. However there is some evidence that these differences diminish with increasing proficiency in the second language, and this suggests that given the right sort of coaxing, words from the second language do end up by becoming fully integrated into learners' personal lexicons.

Thirdly, clang associates (i.e. responses which are principally phonetically motivated, rather than semantically motivated) account for a large proportion of the responses
produced by less advanced students, suggesting that, at certain stages of learning at least, the formal phonetic properties of foreign language words may be more salient than their semantic forms (cf. also a similar claim made by Henning 1973 on the strength of a series of recall experiments). Finally, despite claims to the contrary (e.g. Randall 1980), there is no clear evidence to support the view that learners are like children in that they produce a higher proportion of syntagmatic responses than would be expected in a comparable group of native speakers. Politzer’s 1978 paper is the only one to offer clear evidence to support this claim, but this data is of limited value in that his subjects were absolute beginners and the stimulus list contains some unsatisfactory items. My own experience with the syntagmatic/paradigmatic distinction leads me to believe that it is largely unworkable in practice, as there are no clear criteria for deciding which category any individual response belongs to. Even if unambiguous decisions can be made in a large number of cases, there always seems to be a significant number of responses that cannot be classified with any degree of certainty, and this suggests that claims about syntagmatic/paradigmatic difference in foreign language learners need to be treated with caution.

In general, these conclusions are suggestive and interesting as far as they go, but in the same time they are rather unsatisfying. What seems to missing from this research is any overall strategy which would enable us to follow through these differences between native speakers and learners to the point where we would be a position to make important and relevant claims about the storing a second language vocabulary. In addition, it is a pity that this work is concentrated on the study of group responses, rather than the individual subjects who make up these groups. This sort of approach inevitably ignores information of a personal kind and fails to comment on what might be important individual differences.

G. General conclusions
Three principal points seemed to emerge from this review. In the first place, a very large proportion of the work on vocabulary acquisition has been concerned with vocabulary teaching rather than with vocabulary learning, and though this work is not without interest, it does not throw much light on how words are learned. In the second place, the more psychologically oriented work is also rather limited in scope, in that it has used a rather narrow range of investigative techniques, and looked principally at questions concerning the learner’s entire second language vocabulary, which it is treated as a single undifferentiated whole. I have argued here that this view is probably oversimplified, and that it is quite likely that major differences could be found for words of different types within an individual learner. Thirdly, the most comprehensive work in this field is the relatively large number of studies that have looked at word associations
in foreign language learners. However, even this work is unsystematic, and fails to show any signs of a coherent and coordinated research strategy. This work also suffers from the drawback that it is concerned principally with groups of learners in an area where large individual differences might be expected.

Clearly, then, the study of vocabulary acquisition is an area where the sort of research work out that has been carried out is far from satisfactory, and where a large number of questions still remain to be answered. Levenston 1979 concluded his discussion of some problems in vocabulary research with a list of questions and suggestions for further work. All these questions are worth pondering, but they might usefully be supplemented by some additional questions which have a slightly different emphasis. (a) are there any systematic differences between well-known and recently acquired words in a second language? (b) do newly acquired words in a second language pass through any identifiable stages of acquisition? (c) is it the case that L2 words ever produce behaviour that is indistinguishable from what would be expected with L1 words? (d) are there any clear thresholds which it is necessary for an L2 word to cross before it can be considered to be properly acquired? If so, what types of activity lead to these thresholds being crossed? (e) how is it that L2 words which are often learned as paired associates of their L1 translations eventually come to operate in a way that is relatively independent of their translation? (f) is the acquisition of new words affected by such considerations as the morphological structure of L2 words, or their phonetic structure? (g) are the lexical errors of learners (e.g. malapropisms) systematically different from those of native speakers?

Our current understanding of vocabulary acquisition has almost nothing to say on any of these points, and there is no doubt that work along the lines suggested here could be the beginnings of a very useful research programme.

References

Albert, M and L K Obler  

Arkwright, T and A Vian  

Atkinson, RC  
Atkinson, RC and M Raugh
An application of the mnemonic keyword method to the acquisition of Russian vocabulary. Journal of Experimental Psychology: Human Learning and Memory 1, 2 (1975), 126-133.

Barnard, H

Barter, AR

Becker, D
The etymological dictionary as a teaching device. Die Unterrichtspraxis 10,1(1977), 70-77.

Birmingham University Russian Language and Literature Department

Blum, S and E A Levenston

Bol, E and JAM Carpay
Der Semantisierungprozess im Fremdsprachenunterricht: Lernpsychologie, Experimente und methodische Folgerungen. [The process of semanticization in foreign language teaching: the psychology of learning, experiments and methodological conclusions.] Praxis des Neusprachlichen Unterrichts 19,2(1972), 119-133.

Bongers, H

Brown, DF

Burnard, L S Hockey and I Marriott

Butler, D, C Ott and R Blake

Champagnol, R

Clarke, HH

Crothers, E and P Suppes

Culhane, PT

Dagut, MB
Incongruities in lexical gridding - an application of contrastive semantic analysis to language teaching. *IRAL* 15,3(1977), 221-229.

Dale, E, J O'Rourke and H Bamman

Dale, E and T Razik

Dalrymple-Alford, E

Dalrymple-Alford, E and B Budayr
Examination of some aspects of the Stroop colour-word test. *Perceptual and Motor Skills* 23(1966), 1211-1214.

Davis, BJ and M Wertheimer

Deese, J

De Wolfe, AS

Desrochers, A

Dyer, FN

Entwhistle, D

Ervin, S

Ervin, S

Ervin, S and C Osgood

Evans, I and J Townsendson

**Galisson, R.**

**Galisson, R.**

**Genesee, F, J Hamers, WE Lambert, I Mononen, M Seitz and R Starck**

**Goggin, JG and DD Wickens**

**Gouin, F**

**Green, D and S Newman**

**Groberg, D**

**Hamers, J and WE Lambert**

**Hamers, J and WE Lambert**

**Hammerly, H**

**Haworth, S**
Spanish word association norms for the 100 words of the Kent Rosanoff list. MA thesis Birkbeck College London. 1979.

**Henning, G**

**Hockey, S and I Marriott**

**Holec, H**

**Holley, FM**
The mental lexicon: vocabulary acquisition as a problem of linguistics and human

**Jakobovits, L and WE Lambert**

**Johnson, DB**

**Judd, EL**
Vocabulary teaching and TESOL: the need for re-evaluation of existing assumptions. *TESOL Quarterly* 12,1(1978), 71-76.

**Juilland, A and E Chang-Rodríguez**

**Juilland, A, D Brodin and C Davidiwitch**

**Kent, GH and JA Rosanoff**

**Kershner, J and A Jeng**

**Klein, GS**

**Kintsch, W**

**Knop, C**

**Kolers, PA**

**Kolers, PA**

**Kolers, PA**

**Kucera, H and WN Francis**

**Kurcz, I**

**Lado, R, B Baldwin and F Lobo**

**Lambert, WE**

**Lambert, WE**

**Lambert, WE, J Havelka and C Crossby**
The influence of language acquisition contexts on bilingualism. *Journal of Abnormal and Social Psychology* 56(1958), 239-244.

**Lambert, WE, M Ignatow and M Krauthamer**

**Lambert, WE and L Jakobovits**
Verbal satiation and changes in the intensity of meaning. *Journal of Experimental Psychology* 60(1960), 376-383.

**Lambert, WE and N Moore**

**Lambert, WE and CR Nott**

**Lambert, WE and C Rawlings**

**Leopold, W**

**Levenston, EA**

**Levenston, EA, and S Blum**

**Liepmann, D and J Saegert**
Meara 1980


**Linnarud, M**

**Lopez, M, R Hicks and R Young**

**Lopez, M and R Young**

**Lord, R**

**Lyne, A**

**Mackey, W**

**Macnamara, J**

**Martin, A**
Teaching academic vocabulary to foreign graduate students. *TESOL Quarterly* 10,1 (1976), 91-98.

**Marton, W**

**McCormack, PD**
Bilingual linguistic memory: independence or interdependence: two stores or one?

**McCormack, PD**
Bilingual linguistic memory: the independence/interdependence issue revisited.

**McNeill, D**
The origin association within the same grammatical class. *Journal of Verbal Learning and Verbal Behaviour* 2 (1963), 202-262.

**Meara, PM**
Schizophrenic symptoms in foreign language learners. *UEA Papers in Linguistics* 7(1978),
Meara, PM
Learners' word associations in French. *Interlanguage Studies Bulletin* 3,2(1978) 192-211.

Meara, PM (in prep)

Miron, M and CC Pratt

Morton, AQ

Neufeld, G

O'Rourke, JP

Orbach, J

Ott, CE, RS Blake and DC Butler

Ott, CE, DC Butler, RS Blake and JP Ball

Paivio, A

Paivio, A and A Desrochers

Paivio, A and A Desrochers

Palermo, D and JJ Jenkins

Philpot, M

Pollio, H
Politser, RB

Postman, L and G Keppel

Pressley, M
Children's use of the keyword method to learn simple Spanish vocabulary words. Journal of Educational Psychology 69,5(1977), 465-472.

Pressley, M, JR Levin, JW Hall, GE Miller, and JV Berry

Pressley, M and JR Levin

Preston, M and WE Lambert

Randall, M

Raugh, MR
Teaching a large Russian vocabulary by the mnemonic keyword method. Psychology and Education Series, Technical Reports, 256, 1975.

Raugh, MR and RC Atkinson

Raugh, MR and RC Atkinson

Richards, JC
The role of vocabulary teaching. TESOL Quarterly 10(1976), 77-89.

Ridout, R

Riegel, KF
Some theoretical considerations of bilingual development. Psychological Bulletin 70,6 (1968), 647-670.

Riegel, KF, R Ramsey and R Riegel
Meara 1980


**Riegel, KF and IWM Zivian**

**Rosenzweig, MR**

**Ruke-Dravina, V**

**Saegert, J, S Kazarian and R K Young**

**Salt, MJ**

**Segalowitz, N and WE Lambert**

**Setzler, HH and RE Clark**
Research briefing: recent research on mnemotechnics for learning foreign language vocabulary. *Educational Technology* 16,8(1976), 43-44.

**Singer, JG**

**Stoltz, W and J Tiffany**
The production of childlike word associations by adults to unfamiliar adjectives. *Journal of Verbal Learning and Verbal Behavior* 11.

**Stroop, JR**

**Syracuse University Research Corporation**

**Szalay, LB and J Deese**
*Subjective meaning and culture*. Lawrence Erlbaum Asociates.1978.

**Taylor, M**

**Thorndike, EL and I Lorge**
The teachers wordbook of 30,000 words. Columbia University: Teachers’ College. 1944.

Tulving, E and VA Colotla

Tweedell, F
Vocabulary expansion in the ESOL classroom. TESOL Quarterly 7(1973), 61-78.

Twomey, E

van Ek, JA

Vikis-Freibergs, V and I Freibergs

Walters, J and R Zatorre

Webber, J

West, M

Wilkins, D

Wicklow, CR

Yoshida, M

This paper first appeared in Language Teaching and Linguistics: Abstracts 13,4(1980), 221-246.