Transfer and translation in L2 word associations: Comparing learner data across languages.

Alex Boulton

Alex Boulton. 2003.
Transfer and translation in L2 word associations: Comparing learner data across languages.

- Paper given at the 24th GERAS conference: *Transfert(s)*. Le Havre, France: Université du Havre. 20-22 March 2003. (draft version; please note that this is an expanded reconstruction, the original having been lost in the cyber-void)

Abstract

L1 transfer is often considered in a negative light as "interference" — something to be eradicated at all costs. In vocabulary studies, for example, interference characterises Meara's seminal 1978 article, the base reference for most work on L2 word associations over the last 25 years. His main conclusion is that L2 learners and native speakers provide fundamentally different responses, and this is automatically attributed to defective underlying mental representations resulting largely from negative L1 transfer.

We begin with a critical review of Meara's article and how it has influenced the field to date, and discuss some alternative interpretations for such data and what they can tell us about the structure of the L2 mental lexicon. We then provide new context for this by describing part of an ongoing series of experiments of our own amongst French learners of English. Rather than comparing the learners with natives of the L2, we compare the results of subjects performing the same word association task in English and in French. If their L1 responses are taken as the baseline, then any differences in the L2 data should be indicative of influence from the L2 irrespective of whether such responses approach L2 norms. In other words, we are looking for evidence of approximating towards the L2 rather than deficiency away from it.

Résumé

Le transfert L1>L2 est souvent perçu de façon négative comme une "interférence", et donc à limiter. Pour ne prendre qu'un exemple, elle caractérise l'article de Meara (1978), la référence de base pour les recherches en associations lexicales L2 depuis 25 ans. Sa conclusion principale est que les apprenants et les natifs fournissent des réponses fondamentalement différentes, ce qu'il attribue d'office à des représentations défectueuses résultant du transfert négatif de la L1.

Suivant une revue critique de cet article et de son influence dans le domaine à ce jour, nous examinons d'autres interprétations possibles de telles données et de leur importance pour comprendre la structure du lexique mental L2. Ensuite, nous décrivons une nouvelle série d'expériences auprès d'un public d'apprenants français de l'anglais L2. Le point de comparaison n'est pas, comme d'habitude, entre apprenants et natifs de la langue cible, mais entre les réponses L1 et L2 des mêmes sujets. En acceptant leurs réponses L1 comme point de départ, alors toute différence dans les résultats L2 pourrait indiquer l'influence L2, quelles que soient les normes de la L2. Autrement dit, nous raisonnons en termes d'une approximation vers la L2 et non pas d'une défaillance L2.

Since my PhD in 1998 I've been Maître de Conférences at the Université Nancy 2. I teach on distance degrees in English at the **CTU** (Centre de Télé-enseignement Universitaire), with courses in *didactique*, linguistics, lexicology, corpus linguistics, history of English, translation, etc. http://www.univ-nancy2.fr/CTU/

I'm a member of the **CRAPEL** (Centre de Recherches et Applications Pédagogiques en Langues; affiliated to the ATILF/CNRS), which is where I do my research. My first love was lexis, but current strong interests in data-driven learning and teaching/learning applications of corpus linguistics. http://www.univ-nancy2.fr/CRAPEL/

Contact:
alex.boulton@univ-nancy2.fr

CRAPEL-ATILF/CNRS
Université Nancy 2
3 place Godefroi de Bouillon
BP 3397; 54015 Nancy-cedex
France

Tel: 03.83.96.71.30
Fax: 03.83.96.71.32

Research in second or foreign language (L2) learning is often based on earlier research conducted with native speakers of that language. This is probably inevitable and often even desirable, as it means that pre-existing experimental designs can be used without having to reinvent the wheel or go through the same process of trial and error in the L2. It also means that the learner populations can be compared against data previously collected from natives of the L2. However, there are clearly two basic potential sources of difficulty here. Firstly, experiments which are designed with native speakers in mind may not transfer particularly well to an L2 situation. Secondly, comparing learners with native L2 speakers may be misleading. To highlight these issues, we look at the traditional word association experimental design as applied to L2 studies. In particular, we take a critical look at Meara's seminal 1978 paper and discuss some of the difficulties encountered here in transferring a test for native speakers to a group of learners. We then move on to discuss data from a similar study of our own which compares learners not with native L2 speakers, but with data they produce themselves in their mother tongue (L1).

Traditional word associations

Everybody knows what word associations are: if I say the word *black*, there is a fairly high likelihood that the first word you think of will be *white*. The use of word associations in linguistics stems from the work carried out in the medical diagnosis of psychological conditions; indeed, much work is based on norms collected from a list of 100 stimuli originally compiled by Kent and Rosanoff in 1910 in their 2-part *Study of Association in Insanity*. This initial transfer (from psychology and medicine to language) was not without its problems, as we shall see later. Postman and Keppel's 1970 collection, *Norms of Word Associations*, is perhaps the most widely cited in the field over the last half century, for two major reasons. Firstly, it clearly lays out a standard experimental protocol that can be easily reproduced — 100 written stimuli; subjects are to work through the list, writing a single response next to each one for "the first word it makes [you] think of" (Jenkins 1970: 2) — along with a number of minor variations, for example in the languages used. Secondly, it made lists of norms readily available: each paper contains a complete list of its response data (most importantly Jenkins'). Such language-oriented studies have allowed insights into the workings of the mental lexicon. Without going into detail here:

- native speaker populations tend to be remarkably homogeneous
- responses can be classified into broad categories: the majority are semantic — mainly paradigmatic (the same grammatical class as the stimulus, eg *table>chair*) but also syntagmatic (such as collocations and other "completions", eg *table>lay*) — while a very small number of responses have a purely formal link ("clang" responses, eg *table>able*)
- norms change predictably as a function of: protocol (time allowed, written or oral, single or chained responses etc), stimuli (familiarity, part of speech, priming etc), population profile (age, education, socio-economic category, sex etc), language (or language variety, eg GB/US English), and date of collection

Given the large body of work in L1, it is not surprising that word associations started to be used as an investigative tool in L2 research from the 1960s, with well-reported work by Kolers (1963), Lambert & Moore (1966), Ruke-Dravina (1971), Riegel & Zivian (1972), Arkwright & Viau (1974) and Champagnol (1974), among others. What is perhaps more surprising is that none simply transferred the L1 protocol to L2 learners with a view to collecting norms; rather, they tended to focus on bilinguals rather than learners, use ingenious but highly complex protocols, and pursue a variety of different and highly-specific aims.

Paul Meara was the first to simply transfer the standard procedure as laid out in the Postman & Keppel collection (1970) to a group of learners; indeed, he does not cite any other L2 studies in this article. He uses Rosenzweig's (1970) French translation of the Kent Rosanoff set of 100 stimuli and exactly the same procedure (written stimuli and single responses). His subjects consisted of a group of 76 English girls at London comprehensives starting their French O-level year (the equivalent of today's GCSEs), so presumably aged around 15 with at least 2-3 years of study behind them. This approach had the advantage of allowing direct comparison between the resulting data and substantial native speaker norms. His 1978 paper (along with complementary discussion of the same experiments in 1980 and 1982 in particular) was a landmark in the field, inspiring a number of L2 word association studies over the years. While many of these are very different in design and objectives, they frequently compare results back with Meara even today. Indeed, his paper is perhaps considered the foundation upon which all else rests.

However, as this was the very first L2 experiment to use the traditional design, there are not surprisingly a number of defects. Meara himself is the first to point these out, so our purpose here is to put these into perspective. It must be said that such re-appraisal does nothing to diminish the importance of the paper or indeed the researcher, who has published dozens of papers on L2 lexis including several more on word associations, notably in recent years his work on using associations as a measure of lexical competence (eg Meara & Fitzpatrick 2000), as well as lexical networks and complexity in graph theory (eg Meara & Schur 2001, Wilks & Meara 2002). The problems from the original paper can be divided into two broad categories.

Problems with design:

- transferring procedure: a protocol designed for native speakers to L2 learners
- transferring stimuli: their perceived characteristics may be different in the two languages; this is especially apparent where they have to be translated
- transferring data: comparing L2 learners with native "control" groups

Problems with results:

- data presentation
- primary responses, errors and selective reporting
- other conclusions, including heterogeneity and paradigmatic/syntagmatic response types
- further interpretation

More generally, the fact that earlier L2 studies did not attempt simply to apply the standard protocol to L2 situations may suggest that researchers were aware of some of these problems; the multitude of alternative procedures developed since may support this. For example, learners are generally found to take considerably longer, which may imply that they are less spontaneous in their associations, which could lead to radically different patterns. It might then be more prudent to remember that we are ultimately comparing sets of behaviour (indeed, most work on word associations was firmly rooted in the behaviourist traditions), and that any extrapolations to underlying mental structures is at best tentative. However, if we it is comparisons we are interested in, then it is clear that we have to use the same procedure.

Transfer

A first theoretical problem is that the protocol and stimulus sets devised for monolinguals may not be entirely appropriate when transferred wholesale to L2 situations. For example, it is well known that learners take longer to complete the traditional test than native speakers

unless some way is found to standardise time taken; this might imply they are not responding with the first word that occurs to them, but are perhaps reasoning their response, which may lead to quite different types of behaviour. Furthermore, the Kent-Rosanoff stimulus list from 1910 is still probably the most widely-used; otherwise researchers use their own equally idiosyncratic lists which are often not comparable across studies; experimental design also varies widely. Secondly, given the L1 research available, direct comparison between natives and learners seems inevitable:

> Such a test asks the question: How do L2 learners compare with native speakers in the typical associations they have with given lexical items? The motive for using word association tests with L2 learners has always been… to provide a convenient measure of 'nativeness'… The extent to which the L1 and L2 mental lexicon, respectively, are actually organised in the same way by virtue of their shared structural properties will clearly promote similar paths of association (Kruse et al 1987: 141-142).

Inevitable perhaps, but nonetheless a major source of theoretical difficulty, as the L2 learner is not and cannot be identical to a monolingual native speaker of the target language: as Grosjean (1989) puts it, "a bilingual is not a two monolinguals in one person" — s/he already has a mature cognition backed up by knowledge of the world and knowledge of a natural language. Given this, it seems highly unlikely that any bilingual mental lexicon can or indeed should be a perfect replica of a monolingual lexicon. Meara admits as much in a follow-up paper: "teaching a language aims to produce people who are bilingual, not mere replicas of monolingual speakers" (1982: 34). The consequences of this are important: while the ultimate goal may be native-like behaviour, it is far from certain that a native-like mental lexicon is necessary to achieve this, or even whether it is possible — pure monolinguals cannot translate, for example (Boulton 1999a).

Familiarity of stimuli

If the protocol may not transfer perfectly, it is clear that there may be even more severe difficulties transferring stimuli: meaningful semantic norms cannot be expected when the stimuli are unfamiliar to the majority of subjects, a particular problem with a small group. Meara seems to be aware of this, as he goes out of his way to describe the stimuli as "high frequency words which students at this level would be expected to know" (p194). While he defends them on the grounds that they feature in Gougenheim et al's (1956) *Français Fondamental*, words such as *mou*, *tige*, and *rugueux* are unlikely to be highly familiar to many learners, so the basis of comparison is unevenly weighted. Meara recognises this problem later (Schmitt & Meara 1997: 23), concluding that "it seems that word frequency by itself is not a reliable index of the likelihood of a word being known," and "association results can best be interpreted according to how well the prompt words were known to the individual subject." The results provide perhaps the best evidence for the unfamiliarity of the stimuli — in one case (*aiguille*) the primary response is provided by only 2 of the 76 subjects. Indeed, the lack of homogeneity in general, coupled with the relatively large number of formal associations, both of which Meara takes to be characteristic of L2 learners, may in fact merely indicate lack of familiarity with the stimuli. When Söderman (1989) used the same stimuli in English, he found he could only analyse data for 85% of the stimuli, as others were clearly unfamiliar to his subjects.

The choice of stimuli can be criticised on several more general counts, not least because "tried and trusted tools which work for L1 situations are rarely wholly appropriate for L2 situations" (Meara 1982: 34). Specifically, they are not particularly frequent (nor consistently rare); many appear rather dated now (the Kent Rosanoff list was first published in 1910); the list was compiled with clinical psychology in mind and so does not represent a logical set for linguistic purposes, and indeed are in many ways insufficient for this purpose. Meara's 1982

paper provides a detailed discussion of such defects, though he had already concluded 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" (1980: 115). Still, these are perhaps minor problems compared to Meara's more pressing need of having a large bank of native speaker norms against which to compare the learner data: the Kent-Rosanoff list fulfils this function.

Translation

More problematical is the translation of the list into French: in much the same way as total synonymy is probably non-existent for any items in a given language, so total synonymy of translation equivalents is arguably impossible too. As Rozenzweig remarks (1970: 101), comparisons between languages can only be made "to the extent that confidence can be placed in the translation". He is in fact quite candid in pointing out the difficulties in translating such lists, and mentions that Lambert and Moore's translation (1972) differs for 8 cases. The criteria involved included choosing a word with as similar meaning, connotation, frequency and coverage as possible. This gave a relatively clear choice for 60 item, but a number of problems remained. These mainly centre on the fact that one form rarely has one meaning. For example, one might hesitate before translating *light* as *léger* or *clair*, *sleep* as a verb (*dormir*) or a noun (*sommeil*), or *sweet* as *doux* (which might back-translate as soft). Another problem concerns cognates: these may have been preferred in many cases, but not for *whisky* which was considered a "foreign" drink. Similarly, *bible* was translated as *évangile*, as the French *bible*, according to Rosenzweig, was rarer and more commonly referred only to the Old Testament. This seems less true today for both words. The implications for comparing data in different languages are clear.

Furthermore, there is confusion as to some of the stimuli, as Rosenzweig uses different sets in different studies. The upshot is that he and Meara use different stimuli for at four of the items (*agneau*, *rivière* and *beau* rather than *mouton*, *fleuve* and *belle*; Rosenzweig in fact seems to use *mouton* twice in his data, though this is unclear).

Comparison groups

Meara compares his results extensively against Rozenzweig's data for French women psychology students, with passing comparison with Miller's study of British students (Figure 1).

study	collection date	language	n°	sex	age
Meara 1978	1978?	French	76	f	15?
Rosenzweig 1970	1955-56	French	184	f	18?
Miller 1970	1961-62	English	400	f + m	18+

Figure 1: Comparison sets

As Schmitt (2000: 41) remarks, "we cannot use *any* group of native speakers for baseline data; rather, it is necessary to organise a group as similar as possible to the non-native subjects in terms of education, age, and so on." However, as can be seen in the table above, the comparison groups do not correspond closely. The potential problems here are manifold.

Firstly, it is well known that norms change over time, so using 20-year-old norms is debatable at best.[1] Another difficulty concerns the population sample: it may be wondered

[1] The corollary of this is of course that Meara's findings should not be taken as a basis of comparison today, 25 years on. More recent norms are available, such as Moss & Older's (1996) Birkbeck word association norms; these present data from British subjects aged 17-45 for over 2000 words. More accessible is the "Edinburgh associative thesaurus" (EAT), available free on Internet (http://www.itd.clrc.ac.uk/activity/psych+267). This includes 100 responses for each of 8400 different stimulus words including all of the Kent-Rosanoff list, along with other frequent words; the responses were also used as stimuli to gather further data. A further advantage of

how much similarity could be expected between adult native speakers at university on the one hand, and 15-year-olds with only 2 or 3 years of the language behind them on the other. And while Rosenzweig's group comprises 184 women, Miller's data is from 200 men and 200 women. While it is generally acknowledged that sex makes relatively few major differences between primary responses, a number of interesting differences can still be seen comparing Rosenzweig's data for women against those for men; indeed, had Meara compared his learners against all of Rosenzweig's population samples, they may have appeared considerably more native-like.

This brings us on to the question of sample size. There are a number of dangers inherent in comparing groups of different sizes, as can be seen if we push the argument to its extreme: for a sample population of 1, every response will be idiosyncratic; for an extremely large group, the number of different responses will increase as a whole but decrease as a proportion of all responses. *Reductio ad absurdum* aside, a larger group may not significantly change what we know about the primary responses, but significant changes may occur with less frequent items. The only way to avoid such difficulties is to have comparable groups.

While Meara elsewhere argues that 50 subjects is usually sufficient (1986/1996a: 36), the standard for native speaker studies seems to be around 100 (Jenkins 1970 tops 1000). This is usually enough for a clear picture of the primary responses, but leaves a large grey area surrounding the more idiosyncratic responses. Meara's group of 76 subjects is therefore rather on the small side, especially as learners tend to be less homogeneous than natives. (Though in all fairness, many other L2 word association studies feature considerably smaller groups.)

An alternative approach to criticising the stimuli would be to use more advanced learners. While Meara (p194) deems them "moderately proficient", this is perhaps charitable to say the least: it might be wondered how learners with so little experience can possibly be expected to provide adult native-speaker-like norms to relatively ill-known stimuli (Randall 1980: C4). Either a larger group is needed, or better-known words, or more proficient subjects.

Data presentation

Unfortunately, Meara does not discuss exactly how he analyses his data, and provides complete data for only 3 stimuli (*pain*, *long* and *mémoire*); for the others, we only have the 3 primary responses, along with the number of occurrences and a figure for the total number of different responses for each stimulus. This means that it is not possible to repeat some of his analysis independently, so we shall be following his own arguments closely. Furthermore, he deals only in raw figures, providing no real statistical analysis whatsoever.

From our own research, we know that even in the best of circumstances, word association researchers have to make decisions about misspellings, non-words, multiple-word responses, foreign words, nonsense words, singular and plural forms, regular and irregular inflexions, derivations, etc. Meara is not forthcoming on this, but for the 3 well-known stimuli where the entire results are given, responses were elicited from 91%, 92% and 95%, of subjects respectively. For the same words in our data (below), this compares against 67%, 83% and 82% respectively — for students with on average over 7 years of the language behind them. While this might be an indication that less proficient learners are more likely to take risks, it might rather suggest that the subjects in Meara's experiments were encouraged to produce some kind of response no matter what; this would clearly have a significant result on the number of formal or nonsense associations.

---

this set of norms is the interactive nature of the presentation, though again they are starting to age somewhat. Recent French norms include Ferrand & Alario (1998) for 366 concrete nouns. That said, older norms are not totally without interest. Malrieu & Gronoff (1999) prefer to use Kent & Rosanoff's original (1910) norms rather even than Jenkins' (1970).

It is worth noting too that the data as given sometimes contain 2 or more primaries; but instead of choosing the primary alphabetically, as is the custom, he tends to choose the one that suits his argument. For example, for *fourneau* he gives *tourneau*, *couteau* and *chaud* as primaries in that order (there may be more, but we only have the 3 primaries for each); *chaud* features in the native norms, but *tourneau* is a clang, a feature he wishes to highlight. In 3 of these 7 cases, this leads to a change of category.

Primary responses, errors and selective reporting

Meara begins his data analysis with a consideration of the primary responses of his subjects, dividing them into 3 categories. We shall examine these in some detail.

a) learner primary responses are identical with native primary responses
b) learner primary responses occur in the native norms, but not as primary responses
c) learner primary responses do not occur in the native norms

First though, his classifications seem to include a number of errors:

- In category A where the L2 primary is not a direct equivalent in Meara's data and in Miller's: Meara finds only 4 cases; we find 7. While 6 of these rank in Miller's top 5, they represent on average less than 10% of his English norms, which may suggest learner behaviour which is not simply translation.
- In category B where the L2 primary is a translation of one of Miller's 5 strongest responses: Meara finds 25; we find 24.
- In category B where the L2 primary has a frequency of only one of Rosenzweig's list: Meara claims 6; we find 2.
- In category B, *sombre>soleil* does not occur among the female sample but only among Rosenzweig's male student and worker samples, while *agneau>mouton* does not occur in the norms for *mouton* at all; *aile* is Rosenzweig's primary response for *papillon*, not *fleur* as Meara gives; similarly, for *bleu, ciel* is the primary, not *mer*.
- In category C for L2 primaries do not feature at all in Rosenzweig's sample: Meara finds 37; we find 28. Admittedly, 8 of difference occur with a frequency of only one, so may have been given by one of the male students.
- In category C for clang primary responses: Meara finds 18; we find only 14.

These are of course mainly only minor details, but it is also worthwhile examining Meara's interpretation of the data. In total, he finds 63 L2 primary responses occur somewhere in Rosenzweig's norms for female psychology students. He emphasises that 6 are not typical French responses, occurring only once in the French list: this certainly constitutes a case of selective reporting, as it might be more revealing to stress that this leaves 57 which are more frequent, including 23 primaries.

Meara points out that many of these are identical in Miller's English norms, and so may reflect translation processes rather than the structure of the lexicon per se. This may be true at least to an extent, and it would therefore be interesting to devise a test which might be able to distinguish them, for example by reducing the time available. But even if such responses are indicative of translation and transfer from the L1, it is still interesting to see how and when they occur and to what extent they help or hinder L2 acquisition. It therefore seems unfortunate to dismiss such large numbers of responses — the most similar to the target at that — as "basically uninteresting" (p200), especially when ostensibly looking for evidence of similarity. In total, of the 63 L2 primary responses which occur in the French norms, Meara manages to find only 6 which are "genuinely French and un-English". This is perhaps

indicative of his tendency in this first article to see the proverbial glass as being half empty rather than half full. But there are other reasons for expecting overlap anyway — it is, to be sure, the same world they are dealing with in both languages, and English and French are in many ways culturally and lexically very close, as can be seen from the considerable number of primary translation equivalents by natives of the two languages.

The remaining 37 responses do not occur anywhere in Rosenzweig's sample. Meara qualifies these as "totally unFrench" (p200), though several of the responses can be found in Rosenzweig's companion study of workers' norms: *mémoire>tête*, for example, ranks 6= there. It is also perhaps unfair to include the learners' *beau>belle* here, as Rosenzweig had *belle* as the stimulus for his female students, and has *beau* as the fifth ranking response.

Meara also qualifies derivatives as "unFrench". He gives the examples of *confort>confortable, beau>belle* and *maladie>malade*. While such forms are not frequently to be found as native primary responses (though eg *ordre* has *désordre* as primary), such forms do feature in the French norms (*confortable* and *malade* are both present, if only at *f*=1). *Beau*, as we have seen, is a more complex case but seems to rank fifth, and furthermore contains two derivatives (*beauté* and *belles de nuit*) each occurring once. This highlights that the problem seems to be one of quantity rather than quality: such associations can and do occur in native norms, if considerably less frequently.

Clang responses also feature in the native norms in small numbers, including two of those Meara gives as L2 primaries (*profound>plafond* and *rugueux>rouge*). There are also a number of "misunderstandings" which may be considered indirect clangs. Occasionally the stimulus is apparently related to another French word (eg *jaune [>jeune?] >vieux*), but more commonly the French stimulus seems to be identified with an English word which is then associated back into French (eg *santé [>Santa?] >noël; carré [>car] >voiture*); sometimes the intermediary stage may be French or English (eg *colère [>couleur/colour?] >bleu*).

This only leaves two primaries that are uncertain: *tapis>eau* may be the result of *tapioca*, while *aiguille>train* remains mysterious, though this primary response was given by only 2 subjects. Nonetheless, for non-primary responses Meara finds "a strong tendency for totally extraneous words, related to the stimulus neither in form nor meaning" (p208). Whatever the cause of these, even seemingly inexplicable data should not be cast aside, as Fromkin (1971) has clearly shown: even the seemingly most anomalous language behaviour generally has some cause: it is merely that the cause is unknown.

As we have seen, formal or nonsense associations are perhaps more indicative of unfamiliar stimuli than of a structural tendency in the L2 mental lexicon — faced with an unknown word, semantic associations are simply not possible. Furthermore, as there is generally less exposure to the L2 than the L1, all L2 words are by definition relatively rare, as Meara himself points out (1982: 37). This might be an argument for comparing the best known L2 words with relatively infrequent L1 words for a more even balance.

Other results

It is perhaps not worth going into such detail for Meara's discussion of the secondary and tertiary responses, as similar comments can be made. More interestingly, he mentions several other findings briefly, in particular the relative heterogeneity of the learners' responses, and the relative dearth of syntagmatic responses.

Heterogeneity

It is clear that the learners are less homogeneous as a group than native speakers, both in terms of weaker primary responses and in terms of greater variety of different responses for each stimulus. While this may at first sight appear surprising, the most intuitively appealing explanation is that "words in a second language are less well organised and less easily

accessible than those in the mental lexicon of a native speaker" (p118). Other explanations should not however be ignored, such as a biasing effect of insisting on responses for unknown items. It should also be pointed out that French norms are less homogeneous than English norms anyway — French primaries have been found to account for 20.2% of responses, British 31.2%, and American 37.5% (Rosenzweig 1970, Miller 1970, and Jenkins 1970 respectively). Meara does not make much of this in his original article, probably because, on this criterion alone, at 22.5% his learners actually appear more homogeneous than Rosenzweig's French female students.

Furthermore, as with native speakers, the learners exhibit greater homogeneity with more common stimuli, insofar as there are fewer different responses and the primary responses account for a greater percentage of all subjects. And even in English, as Schmitt (2000) also points out, a concentration on primary responses should not ignore the high levels of infrequent answers:

> Even a cursory glance at… the published [L1] norms… shows that there are far more unstereotyped responses than there are stereotyped responses. Any attempt at a theory of lexical organisation must come to grips with this large amount of data as well as with the most frequent responses (Randall 1980: C11).

Paradigmatic and syntagmatic responses

One of Meara's original hypotheses was that learners would resemble children in their association patterns, in particular providing relatively more syntagmatic than paradigmatic responses. The opposite apparently turns out to be true in his experiment, although the figures are not given and cannot be fully derived from the incomplete norms provided. The discussion in Meara's 1978 paper again is rather brief on this point, though he concludes that "there is no evidence in the data as a whole that the learners produce syntagmatic responses in any systematic way" (p207). By giving higher than expected levels of paradigmatic responses, the learners would therefore seem to be exhibiting more mature, educated behaviour than might be expected given their profile. However, other findings conflict on this point: Söderman (1989) finds large numbers of syntagmatic responses at lower levels, rapidly giving way to more paradigmatic responses as proficiency increases.

Meara has returned to this point several times, first claiming that the paradigmatic/syntagmatic distinction "is largely unworkable in practice, as there are no clear criteria for deciding which category any individual response belongs to" (1980: 119). Singleton (1999: 233ff) agrees, lumping paradigmatic and syntagmatic responses together under the heading "semantico-pragmatic". This view will no doubt elicit much sympathy from many researchers actually trying to allot items to these categories: the distinction is so open to subjective interpretation that comparisons between studies are extremely uncertain (see also Schmitt 1999). Meara later (1982: 30) claims the distinction is "important", and refers to it in 2002 (Wilks & Meara) as a possible future direction for ongoing research.

Interpretation

Meara finds his data consistent with two possible interpretations. The first is that there may be "serious inadequacies in the learner's grasp of French" (p208). In particular here, he returns to the idea that the native speaker lexicon is semantically organised, while for the learners "this semantic organisation seems to be much less well established"; while the learners "do show some evidence of semantic organisation", this is "mainly dependent on translation between French and English" (p208). Two initial points here: firstly, as we have seen, semantic organisation depends on familiar stimuli, otherwise only formal associations are possible; secondly, the vast majority of his learners' primary responses are semantically motivated. The claim about translation is perhaps more significant, as Meara considers this as

interference to studying the structure of the L2 mental lexicon. It is however possible to think of translation as an entirely natural and common practice (especially so for less proficient learners, such as here), and therefore something worthy of study rather than a barrier to research. One may be left with the feeling that the learners cannot win in most cases: if they provide a common response they are merely translating from the L1, if they do not they are not native-like.

Meara's second major possible interpretation is that the major differences between learners and natives are "not really of any importance" if "all learners go through a phase when their foreign language lexicon is organised on non-semantic criteria" (p209). Again, the words that are organised along non-semantic lines tend to be those that are unfamiliar, suggesting that they may not be stored in the mental lexicon as such; if this is the case, then what is revealed in the experiment is not a non-semantic organisation of the lexicon itself, but non-semantic associational processes.

Summary

As we have seen here, applying the standard test to a group of learners was a hugely important step in the field of L2 WAs. Unfortunately, the study is potentially flawed in using a small group of learners at a low level of proficiency for stimuli of relatively unfamiliar frequency. There are also a number of inconsistencies or errors in reporting the results and comparing against the native speaker norms; the limited data provided means that other elements cannot be checked.

These are perhaps minor quibbles in most cases, but it is clear too that facts do not speak for themselves — they need interpreting, and not everyone has the opportunity to go through the data in such detail. It is in the manipulation of data and selective interpretation of results that more serious misgivings arise, given that the findings are still regularly cited in the field. In particular, Meara tends to highlight differences between learners and natives while neglecting similarities: as he himself pointed out in a later article critical of other studies (Meara 1982), the quantities of data resulting from word association experiments mean that it is possible to confirm virtually any hypothesis, and explain away the rest.

Ultimately what we are left with are the following cases; as it can be seen, most of the time the learner just cannot win (figure 2).

learner responses	Meara's explanation
similar to native speakers	probably the result of translation, so uninteresting
different from native speakers	different or developing lexicon, so not native-like behaviour
"clang" or formal in nature	reveals formal structure to lexicon, so not native-like behaviour

Figure 2: Response possibilities

Meara later admits that much of this original work is "rather coarse grained", and "suggestive rather than conclusive" (Meara 1984: 232). He has also on several occasions criticised his own study, along with much other contemporary research, for covering "old ground" and being "content merely to describe the sorts of responses that learners produce, together with a minimal statistical analysis" (1982: 32); this is ironical for someone who later uses graph theory (eg Meara 1992; Meara & Schur 2002) and who advocates greater mathematical rigour in applied linguistics (Meara 2001). Despite his own misgivings and the problems outlined here, the study is still commonly cited as proof that learners are less

heterogeneous than native speakers, that response types have different distribution, and that the L2 lexicon is more formally based.

An alternative study

Armchair criticism is of course a very easy game to play, but it is perhaps more worthwhile to try to produce something constructive. In this section, we discuss part of an on-going L2 word association study. Over several years, we are using the Kent-Rosanoff list in standard procedure, despite its faults, with non-specialist students in France. Given that the stimuli, protocol and norms exist and have been used extensively, we are for some purposes stuck with them for better or for worse, just as we may be stuck with a peculiar keyboard or driving on an original side of the road.

In this series of experiments, subjects are banded into three groups of approximately equal size according to level, giving an average of 124 sets of responses per year so far. This will allow us to track how results may change over time for the population, compare results between groups of different levels of L2 proficiency, and provide a large bank for other comparative uses. This basic design is frequently supplemented by variations or other word association tests, but as these are conducted by the same researchers with students of the same profile and within the framework of the same project, comparisons should be relatively unproblematic — if some elements of design and analysis are open to subjective interpretation, at least the same subjectivity will apply to the entire body of data collected. The rationale, design and results will be the subject of projected future work, but the basic aim is to provide a coherent framework to a set of large-scale traditional experiments along with a fairly exhaustive battery of variations on a theme.

In the present case, we wanted to design a follow-up experiment allowing us an alternative perspective to Meara's original study. When two native speaker samples are compared, differences are considered interesting, but not indicative of a deficient lexicon. With learners, however, any differences between the learner and the native speaker data tend to be taken to represent the learners falling short of their target (situation A in Figure 3). Meara's original paper attributed all significant differences to deficient structure of the learners' mental lexicon and implied a view of translation as interference that has to be overcome for successful learning to take place.

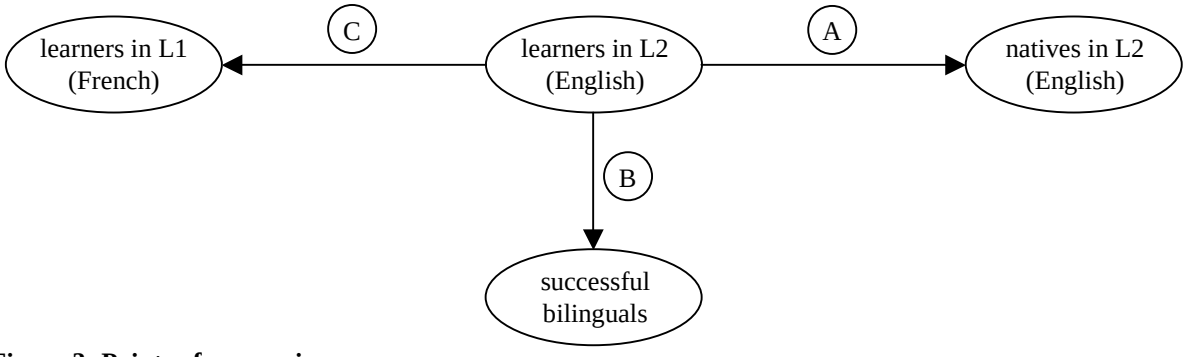

Figure 3: Points of comparison

We have already seen that successful bilinguals are unlikely to be indistinguishable from a native speaker of each language. Thus if word association tests are in any way indicative of the structure of the mental lexicon, then some differences between learner and native L2 responses may be necessary and desirable rather than to be avoided at all costs. Comparisons with natives are not without their value, but they are not necessarily as revealing as once thought. Thus Meara, on second thoughts, decided that "the apparent bonus of being able to

compare learners' responses with the published norms for native speakers turns out on closer inspection to be of doubtful value" (1982: 33), though some of his most recent work (eg Wilks & Meara 2002) still involves useful comparison. An alternative point of comparison would be between learners and successful bilinguals (situation B in Figure 3). As Kruse et al (1987: 142) point out, researchers working with learners are interested in the "interlexicon" rather than some idealised finished product of native speaker-like bilingualism. In many ways, such comparisons would be ideal, though it is hard to define and harder still to find a suitable comparison group of "successful bilinguals".

A third type of comparison compares the learners in their L2 and their L1 (situation C in Figure 3). This is the approach adopted in this study. Thus rather than considering an ultimate L2 goal to aim for, we consider the starting point of the learners' own L1; rather than considering any non-native-like (L2) behaviour as falling short of the target, we may consider any non-native-like L1 behaviour as an attempt at language acquisition.

In this experiment then, the same group of learners completed the test in the L2 (English) and in their L1 (French) to allow a direct comparison of the results. This design produces a number of quite practical benefits. In particular, using the same group of subjects twice provides an almost perfect comparison, as the language is virtually the only thing that changes; furthermore, group size is identical, and with close collection dates there are unlikely to be any significant differences in age, sex, personality and so on.

These benefits are balanced by a number of difficulties, such as that the first test may serve as a prime for the second: this is the reason why the L2 experiment was conducted first, as it is more likely to be influenced by the L1 than vice versa. The biggest problem is that we are dealing with two languages, which has two obvious but important consequences: firstly, translated stimuli are as we have seen not identical and will influence results; secondly, responses can also only be compared in translation, so any judgements depend on a degree of (partially subjective) interpretation. For these reasons, any results must be interpreted with caution and should be taken as indicative rather than absolutely conclusive. However, we feel this approach is useful in complementing results so far obtained in the field.

Method

The Kent-Rosanoff (1910) list of 100 stimuli in English was presented using the procedure in French given in Postman and Keppel (1970) and previously replicated by Meara (1978): briefly, to respond to each written stimulus with a single written response, the first word that came to them. Further explanation was given as necessary from earlier pilot trials: in particular, responses should be in English; subjects were asked to provide some response even if they did not know the stimulus, but to move on if really nothing came to them; time was limited to 13 minutes. The same procedure was employed two weeks later using the Rosenzweig (1970) French translation of the same word list as it features in Meara 1978; responses were here to be given in French.

Subjects were first year students at ESSTIN, an engineering college in Nancy, in the north east of France. Subjects are entirely male, 18-year-old native French speakers enrolled for the first time here. Results were not counted from students with other profiles, who did not understand the test instructions (eg by responding in the wrong language), or who attended only one of the two test sessions, leaving a total of 138 subjects. With between 5 and 9 years (median 7) of English behind them, they have English as a compulsory part of their engineering course (two 2-hour classes per week), and they are streamed according to an in-house placement test which allowed us to divide them into 3 approximately evenly-weighted groups for the purposes of this test: 44 upper intermediate; 50 intermediate; 44 lower intermediate; clearly though such groupings do not represent discrete classes. The experiment

itself was conducted 2 months into the academic year at the end of normal class time by the researcher.

Blanks, illegible responses and repetitions of the stimulus were counted as zero responses. Items were lemmatised for plural, verbal, adjectival and genitive inflections, though derivatives were counted separately, as were 2-word responses (except in the case of some grammatical particles). Items that include the stimulus as part of the response were counted together (eg *man* and *spiderman* were counted as the same response to the stimulus *spider*). Spelling was normalised, though at times this entailed subjective decisions due to the high number of cognates between French and English. The 3 primary responses for each stimulus in both English and French versions are given in the appendix, along with the number of zero responses and the total number of different responses for each stimulus.

Results and discussion

As has commonly been noted, word association experiments give rise to enormous amounts of data that can be analysed in countless different ways. As the purpose of this paper is to provide comparison with Meara's 1978 findings, we shall limit our discussion to similar analysis and relevant evidence. We thus focus in particular on the following:

- homogeneity: from primary response strength and total different responses
- translation equivalents: as a group as well as for each individual subject
- formal response types: in particular clang responses and "erroneous" associations
- typically "un-French" responses

But first, the large number of zero responses warrant a brief initial analysis: barely 10% of subjects provide a response for one stimulus (*stem*), and for the test as a whole, over a quarter of all possible answers in English are either left blank, illegible, or merely repeat the stimulus, with subjects averaging 72 responses for the 100 stimuli; this compares to over 99 for the French version. Such results echo Singleton's (1999) finding that L1 stimuli produce more chained associate responses than L2 ones on average. It is clear that zero responses were not a random feature, as the correlation between groups is extremely high at +0.94 (see Figure 4), which shows that it is consistently the same stimuli that pose problem.

	r (Pearson product-moment correlation coefficient)	r^2 (coefficient of determination)	*p*
zero responses	+0.9373	0.8790	<.0001

Figure 4: Average between-group correlations for zero responses / stimulus

Further, an ANOVA test followed by a Tukey HSD test show that all three groups are significantly different at $p<0.01$ level (Figure 5). This further shows that the number of zero responses declines significantly with proficiency.

	Gp A	Gp B	Gp C	Total
n	100	100	100	300
ΣX	971	1347	1526	3844
mean	9.71	13.47	15.26	12.8133
ΣX^2	19647	32757	38296	90700
variance	103.2181	147.6052	151.6085	138.6139
std. dev.	10.1596	12.1493	12.3129	11.7734
std. err.	1.016	1.2149	1.2313	0.6797

	SS	df	MS	*f*	p
treatment [between groups]	1604.81	2	802.4	83.41	<.0001
error	1904.53	198	9.62		
subjects	37936.21	99			
total	41445.55	299			

Tukey HSD: A/B, A/C, B/C: $p<0.01$

Figure 5: ANOVA and Tukey test summary for zero responses by group

One obvious assumption is that the zero responses are given to unfamiliar items. While frequency is not an infallible guide to familiarity, as has already been noted (especially with two languages with such a high degree of cognateness as English and French), some considerable overlap may be expected. Compared against the British National Corpus of over 100 million words (90% written text, 10% spoken), we found that 23 of the 100 stimuli did not feature within the commonest 4844 words which have a frequency of at least 20 per million. The average ranking for the others was 1712, with an average frequency of 213 per million (Leech et al 2001). This compares with Meara's comment that all but 7 of the French stimuli appeared in the first or second levels of the *Français fondamental* (Gougenheim et al 1956).

Homogeneity

There are two basic ways we can gauge the homogeneity of a group:

a) from the relative strength of the primary responses as a portion of all responses
b) from the total number of different responses for each stimulus

a) Primary responses

For the English experiment, primary responses account for 22.3% of all responses on average; the three primary responses for 36.5%. For the French version, these figures are rather higher at 32.2% and 51.5% respectively, suggesting that the subjects are indeed more homogeneous in their L1 than the L2. It is interesting to compare these scores against Rosenzweig's 1970 study of 104 male students (Figure 6): a mere 22.2% provide the primary response, considerably lower than our roughly comparable group of subjects in French, and similar to the L2 results in our study and in Meara's. This may be revealing of norms changing over time and becoming more homogeneous, though such discrepancies serve perhaps more to underline how dangerous it can be to compare dissimilar groups: even if all are male students, Rosenzweig's data were from a sample population finishing high school or enrolled in psychology at the Sorbonne in Paris in 1955-56 rather than first year engineering students in Nancy in 2002.

	L1 French	L2 English	Rosenzweig (male students) L1 French	Meara L2 French
Primary responses	32.2%	22.3%	22.2%	22.5%

Figure 6: Primary response strength

A more rigorous statistical analysis is necessary here. ANOVA and Tukey HSD tests on the primary responses show that there is no significant difference between our subjects in the L2 test and Meara's when taken as a percentage of number of participants (Figure 7); but there is a significant difference between L1 and L2 in both cases (p<0.01). Insofar as the proportion of subjects providing primary responses can be taken as an indication of homogeneity, this provides strong support for Meara's conclusion that natives are more homogeneous than learners, though we can add another dimension: this homogeneity applies whether the learners are compared against native L1 or native L2 norms. In other words, not only are learners not behaving like L2 natives, they are not behaving like L1 natives either.

	L2 (English)	L1 (French)	L2 (Meara)	Total
n	100	100	100	
ΣX	2226.79	3215.95	2191	7633.74
mean	22.2679	32.1595	21.91	25.4458
ΣX^2	77034.4567	132353.1481	79067.697	288455.3018
variance	277.2578	292.2202	313.7665	315.0792
std. dev.	16.6511	17.0945	17.7135	17.7505
std. err.	1.6651	1.7094	1.7713	1.0248

	SS	df	MS	*f*	p
treatment [between groups]	6767.47	2	3383.73	24.71	<.0001
error	27110.97	198	136.92		
subjects	60330.24	99			
total	94208.68	299			

Tukey HSD: A/B, B/C: p<0.01; A/C: no sig diff

Figure 7: ANOVA and Tukey test summary for percentage of primary responses between languages

If we apply the same reasoning and analysis to group differences, we find that the higher groups are indeed slightly but significantly more homogeneous than group C (p<0.01), though there is no significant difference between groups A and B (Figure 8). This can be seen too in the way the standard deviation decreases with proficiency. In fact, groups A and B are also more homogeneous than Meara's group of learners — it is group C which brings the average down. So it seems that learners do become more homogeneous in their primary responses as they progress in the L2, though it remains debatable whether such homogeneity would ultimately level off, for a group of successful learners, at levels comparable to their L1 or to the L2.

	Gp A	Gp B	Gp C	Total
n	100	100	100	300
ΣX	2440.9	2402	2113.66	6956.56
mean	24.409	24.02	21.1366	23.1885
ΣX^2	84823.2116	85788	74596.9388	245208.1504
variance	254.9827	283.7572	302.2359	280.5877
std. dev.	15.9682	16.8451	17.3849	16.7508
std. err.	1.5968	1.6845	1.7385	0.9671

	SS	df	MS	*f*	p
treatment [between groups]	639.13	2	319.57	9.3	0.000138
error	6803.28	198	34.36		
subjects	76453.32	99			
total	83895.73	299			

Tukey HSD: A/C, B/C: p<0.01; A/B: no sig diff

Figure 8: ANOVA and Tukey test summary for percentage of primary responses between groups

It may be interesting to take into account zero responses and to redo the calculations for the primary responses accordingly — not as a proportion of the number of subjects, but as a percentage of responses elicited; we use a *t*-test here, as Meara does not provide the number of zero responses for his data. While there remains a difference, this is markedly less

significant (p<0.05) and the gap narrows considerably to only 3.8%, with an average of 28.7% in the L2 and 32.5% in the L1.

This shows how easy it can be to manipulate the statistics to show whatever one wants, and at the same time how difficult it can be really to prove anything. On the one hand, if learners are instructed to respond no matter what, the statistics will suggest a high degree of heterogeneity compared to native speakers; on the other, if they are allowed not to respond and this is allowed for, then the differences disappear. Another perspective: over the 100 stimuli, the subjects were actually more homogeneous in the L2 than their L1 for 21 words (see appendix); 12 of these were translation equivalents of the L1 primary (*black, white, citizen, soldier, stomach, dream, yellow, bread, bible, hungry, butter, bed*), but the others were not (*music, sickness, house, foot, spider, sleep, working, religion, city*).

b) Total number of different responses

Another way to gauge homogeneity is to look at the total number of different responses provided by each group. Statistical analysis reveals no significant difference between groups A and B, but A/C and B/C show slight differences at the p<0.01 levels (Figure 9).

	Gp A	Gp B	Gp C	Total
n	100	100	100	300
ΣX	1604	1647	1383	4634
mean	16.04	16.47	13.83	15.4467
ΣX^2	28414	29709	21925	80048
variance	27.1297	26.09	28.2637	28.3216
std. dev.	5.2086	5.1078	5.3164	5.3218
std. err.	0.5209	0.5108	0.5316	0.3073

	SS	df	MS	*f*	p
treatment [between groups]	401.29	2	200.64	33.5	<.0001
error	1186.05	198	5.99		
subjects	6880.81	99			
total	8468.15	299			

Tukey HSD: A/C, B/C: p<0.01; A/B: no sig diff

Figure 9: ANOVA and Tukey test summary for total number of different responses by group

More important of course is to compare the L1 and L2 versions of the experiment; we cannot include Meara's learners here, due to very different sample sizes. A *t*-test shows a significant difference — but with the learners being *more* homogeneous in the L2 at a significance level of p<0.0001 (average 33.4 different responses per stimulus as opposed to 40.0 in the L1). The obvious explanation is simply that the learners provided enormous numbers of zero responses in the L2. Once this is taken into account (by dividing the number of different responses by the total number of responses for each stimulus), the averages change to 38.3 different responses per stimulus in the L2 and 29.3 in the L1, still at the huge level of p<0.0001. Viewed this way, the subjects are indeed more homogeneous in the L1 than the L2, just as we saw above. This suggests that our earlier allowance for zero responses is not just statistical manipulation. On the other hand, it might be argued that this in fact skews the results here: remember that our reason for not comparing against Meara's results was that the group sizes were very different. By discounting zero responses here, we are effectively creating different group sizes; there seems to be no obvious solution to this problem.

	unweighted		
	L2 (English)	**L1 (French)**	**Total**
n	100	100	200
ΣX	3342	4000	7342
ΣX^2	121744	172688	294432
subjects	10054.36	12688	24907.18
mean	33.42	40	36.71

	weighted for zero responses		
	L2 (English)	**L1 (French)**	**Total**
n	100	100	200
ΣX	3828.00	2928.00	6756.00
ΣX^2	1744.82	925.680	2670.5
subjects	279.46	068.36	388.32
mean	38.28	29.28	33.78

Figure 10: t-test data summary for total number of different responses by language

Translation equivalents

We now move on to the question of translation which formed the basis of much of Meara's discussion. As deciding what constitutes translation is necessarily partly a subjective judgement, we have generally assumed translation where there seemed to be room for doubt. For the primary responses in English, 47 have close translation equivalents that are also primaries on the French list; 11 others are in the top three. This figure is considerably higher than Meara's 23 primary overlaps. We therefore compared our L2 data against Miller's English norms and found 37 primaries in common; this is still considerably more than Meara. We also compared our French primaries against Rosenzweig's French norms and found only 52 primaries in common. Taken together, these figures suggest that we are unlikely to be dealing with a simple anomaly. The high L2 figures obtained here suggest two possibilities. Firstly, if WAs are a gauge of the structure of the mental lexicon, it may be that the learners, for the most part, see no need to structure their L2 lexicon any differently from the L1, and transfer all kinds of lexical elements from the L1 directly on to the L2. On the other hand, it may be that the subjects are resorting to the *process* of translation, especially for the less familiar stimuli. For example, none of the 3 primary responses for 12 stimuli feature anywhere in the L1 list (*smooth, sweet, whistle, needle, carpet, sour, cabbage, stem, swift, square, heavy, blossom*); in most cases, these words are clearly unknown to the vast majority of subjects, so they are much more revealing of can only lexical processes than lexical structure.

In order to compare translation equivalents between groups, we looked at individual students to see how many provided translation equivalent responses in the two tests. As expected, there is a linear difference between the 3 groups, though this difference is only significant between groups A and C ($p<0.05$). This increases to $p<0.01$ once zero responses are allowed for (ie as a percentage of those responses given), though is still not significant between the other groups (Figure 11).

	Gp A	Gp B	Gp C	Total
n	44	50	44	138
ΣX	924.56	963.21	647.81	2535.58
mean	21.0127	19.2642	14.723	18.3738
ΣX^2	24641.5846	23239.3787	11305.1315	59186.0948
variance	121.2572	95.59	41.1036	91.9557
std. dev.	11.0117	9.777	6.4112	9.5894
std. err.	1.6601	1.3827	0.9665	0.8163

	SS	df	MS	*f*	p
treatment [between groups]	932.52	2	466.26	5.4	0.005545
error	11665.42	135	86.41		
subjects					
total	12597.94	137			

Tukey HSD: A/C: $p<0.01$; A/B, B/C: no sig diff

Figure 11: ANOVA and Tukey test summary for average translation equivalents per subject as a percentage of responses

All this might sound like unnecessary statistics, but what is interesting is that group A (the most proficient group) on average gives significantly *more* translation equivalent responses than group C (the least proficient). Also of note is that their standard deviation is almost double, indicating greater variability among the more advanced learners. Taken at face value, this suggests that translation (if indeed that is what is going on) may be desirable or even necessary for proficiency, but that this is depends enormously on the individual learner. This clearly goes against the widespread idea that translation is indicative of an immature L2 lexicon and is not something to be encouraged — beginning learners use translation as they have few other tools at their disposal, but advanced learners should relate L2 words to each other rather than back to the L1.

As this finding goes against the grain of received wisdom, caution should of course be exercised here. One explanation seems obvious: translation can only be used when the learner knows the equivalent for both (L2) stimulus and (L1) response; clearly the more proficient learners are more likely to fit into this category. Additionally, less proficient learners may be apt to produce erroneous responses through misunderstanding of either stimulus or response. Nonetheless, this explanation suggests that translation may be conducive to learning in at least some cases, may be one tool out of many that contribute to a mature L2 lexicon, and as a result should not be discouraged out of hand. Indeed, given the wealth of linguistic and extra-linguistic knowledge that the learner has in his L1 or using his L1, it might be considered unnecessarily handicapping to try to deny him access to this in attempting to eradicate all translation and transfer. Conversely, it seems reasonable to draw on this in at least some cases, to explore how translation may be used most effectively by allowing reference to a vast pre-existing store of linguistic and extra-linguistic knowledge rather than trying to force the learners to relearn everything anew. This appears particularly pertinent given that it seems impossible to stop learners translating all of the time anyway.

A qualitative analysis finds that significant overlap (in translation equivalents) tends to occur between L1 and L2 primary response patterns in certain circumstances:

- stimulus and response are not language-specific in their meaning
- stimulus and response are close equivalents and are not polysemous (or are polysemous in similar ways)
- learners know both stimulus and response (there is a fairly strong inverse correlation between translation equivalents and zero responses: -0.63; $p<0.0001$)

If translation equivalents are thus dependent on the stimulus, this must go at least some way towards explaining why general statistical patterns are difficult to find and indeed

misleading if they are applied to all the items in a relatively random list such as the Kent-Rosanoff set. Where the criteria above apply, we should expect the items for the two languages to be embedded in similar network structures in association experiments; failure to show this indicates unfamiliarity with either the stimulus or the response or both, or some deviant network structure for other reasons. Consequently, translation equivalents are "a good thing" here. On the other hand, we might expect other items to be embedded in language-specific network structures; failure to show this in association experiments indicates unfamiliarity with either the stimulus or the response or both. Alternatively, there may be some other reason for a deviant network structure for other reasons; translation equivalents are "a bad thing" in such cases.

Formal response types

Meara makes much of clang responses, finding 18 among the primary responses alone. In our data we find only 5 clang responses among the primaries: *whistle>castle, needle>need, cabbage>garbage, stove>love, blossom>blood*. As these are some of the weakest primaries, together they account for only 26 responses out of a possible total of 13800, ie 0.19%; even for the three primaries the rate is still only 0.47%. The reduced importance of clang respones is found in other studies (eg Randall 1980, Singleton 1999b). It may well be that the subjects here are more familiar with the stimuli than are Meara's population sample, but this seems unlikely to provide a complete explanation for the disparity; it seems unlikely too that French is more conducive to clang associations than English. This leaves two major explanations: it may be inferred that the subjects either have radically different lexicons (probably as they are relatively more advanced than Meara's subjects), or are employing radically different procedures (if they are more likely to produce a zero response to unfamiliar items where otherwise only clang responses are possible), or a mixture of the two.

A second formal category explored by Meara includes "erroneous" responses resulting from confusions and misunderstandings; these may be accidental or, apparently in some cases, deliberate. In our data these represent only 2 primary responses (*sour>ear*, *swift>clean*) representing 0.04% of total responses possible; there are only 12 altogether in the 3 primaries, accounting for 0.3% of possible responses. While such items are clearly of interest and worthy of further study, such low figures show that it is as well not to overstate their importance.

Another area we might include here concerns lexical transparency between the two languages. In particular, English and French have many cognate items: Walter (2001: 121), for example, lists 3222 words which are spelt "absolutely identically" in English and in French and "cover exactly the same meaning or meanings" in the two languages; vast quantities more are similar in form and meaning. It is not surprising then that 41% of the English primary responses are likely to be known to the subjects in roughly similar form and with significant semantic overlap (eg *foot* having primarily different meanings in the two languages), and figure in the *Petit Robert* 2001; this compares to 48% of the stimuli. Similarly, some English words, while not full cognate borrowings and not in most dictionaries, are still known to many monolingual French speakers for ephemeral cultural reasons, mainly sport, cinema and television, music, computers and video games. Thus we have *war>hammer, heavy>metal, smooth>criminal, spider>man*, etc. In all 16% of stimuli and 14% of primary responses seem to feature in this category; combined with cognates, this brings the total to 64% and 55% respectively (see Boulton 1999b for more on this.)

Typically "un-French" responses

This leaves us with other response types that seem to be typically "un-French" in some way, and which may therefore be indicative of some kind of L2 influence, real or perceived.

But firstly, some more prosaic explanations need to be considered. In some cases French-like primary associations would be virtually impossible due to the translation of the stimulus list itself: *fleuve* gives *rivière*, while English has no equivalent distinction and has *river* as the stimulus; the obvious antonyms for *short* and *petit* are *long* and *grand* respectively, the primary for each. This is still interesting however, as it does mean that learners are focusing on the most relevant sense in each language rather than just transferring willy-nilly. In addition to the unfamiliarity of some stimuli (cf our discussion of zero responses above), it might be that a familiar word as stimulus triggers a concept for which the learner does not have a lexical item in the L2, or at least which is relatively unfamiliar to large numbers of subjects. This may explain why, for example, *papillon* produces 24 *chenilles*, but *butterfly* only 4 *caterpillars* — subjects may wish to make more direct use of L1 but be unable to for lack of competence in the L2.

That said, some words do seem to have quite English connotations. The primary response for *house*, for example, was *home*, which has no real equivalent at all in French, the closest perhaps being *chez soi* at rank 10. It also seems that *gardens* (*f*15) are more closely associated with *houses* than *jardin* (*f*5) with *maison*. *Sweet* also gives *home*, as the expression *home sweet home* is well known. Similarly, English *cottage* is associated with *country* and *field*, neither of which feature in the data for French *villa*, which may have a rather different meaning. A final example is *cheese*, common on fast food menus, thus more closely associated with *hamburgers* than dairy products as in French.

For several stimuli there seems no good reason for not choosing the primary:

- *want* is highly familiar, but only features in 3rd place in response to *wish*, while its translation equivalent *vouloir* is the primary;
- there seems no particular reason for *eyes* to be more closely associated with *head* than *oeil/yeux* are with *tête*, or for *water* to be closer to *ocean* than *sea* (as *mer* to *océan* in French);
- the primary for *beautiful* was *nice*, which has no obvious equivalent in the French, while *woman* is only the secondary response;
- *carpet* has the primary *magic*, and though French has the same common concept (as can be seen through *volant* in second position), *magie* itself has only *f*1;
- *working* is primarily associated with *sleep* (*f*27), but *travail* is mostly *dur* (*f*13).

A number of words have quite unusual patterns: 12 subjects, for example, responded *deep>forest*, though no equivalent featured anywhere in the L1 norms. Some associations seem to be language specific: English *music* elicits *rock*, but *musique* has more classical or traditional associations (*chanson*, *son*, *note*). These provide further evidence of the subjects exhibiting rather different behaviour in the two languages.

Conclusion

In this paper we have examined some of the problems in transferring word association experiments intended for native monolinguals to an L2 situation. In particular, these include protocol (flexible time for each response may encourage different behaviour), stimuli (the choice as well as the translation) and control group (size and recency, as well as subject profiles — age, sex, education etc). We critically examined an experiment carried out by Paul Meara (1978), the first to transfer the standard word association test directly to L2 learners. We analysed a number of general difficulties as well as a several more specific problems: these include some minor factual errors, but also certain problems of reasoning and interpretation.

We then carried out our own experiments with a similar design, the main innovation being to compare learners' word association responses in L1 and L2. Using the same subjects for both experiments provides the "perfect" control in terms of size and profile. While supporting Meara's major conclusions of different response patterns, our data suggest that the differences may not be as high as previously thought, and underline that they are quantitative rather than qualitative in nature (cf Boulton 1999a). Kruse et al (1987) found that some learners may even produce more "native-like" responses than natives; clearly, there is enormous overlap and differences are mainly to be found in statistical patterns.

This alternative paradigm also entails a shift in approach, as we are not merely looking for evidence that learners behave inappropriately in the L2. This is partly inherent when comparing learners with a native L2 control group, as any differences are taken as evidence of deficiency on the part of the learners.

In our design, comparing learner behaviour with their own L1 norms allows us to begin with the hypothesis that any differences are evidence of not just relying on the L1. The response patterns here reveal considerable overlap between the two sets of data. While Meara underlines that this may be evidence of translation and therefore not worthy of discussion, we have argued that such overlap may be a necessary feature of a successful bilingual lexicon rather than just a hindrance. Other researchers have concluded that the overlap is suggestive of L1 patterns being used in the L2 and not just of translation processes. Van Ginkel & Van der Linden (1996: 32), for example, find that L1 influence "remains obvious even at an advanced level of proficiency… building up an L2 lexicon does not mean that access to [the] L1 lexicon is inhibited in any important way." Nor should we wish it to be. Lewis (1993: 16) argues that "it is by no means obvious that naturally occurring data should provide either the model or the target for language learning"; the same could be said for experimental data:

> Knowing the native 'norm' would in no way facilitate the teaching or learning of foreign words — on the contrary: it would create additional (semantic) problems, as cultural differences would be involved. (Erdmenger 1985: 161)

Bibliography

Arkwright, T. & Viau, A. 1974. Les processus d'association chez les bilingues. *Working Papers in Bilingualism*, 2, p. 57-68.

Boulton, A. 1999a. *Le Lexique Mental L2: Connexions et Associations.* Lille: Presses Universitaires du Septentrion.

Boulton, A. 1999b. Associations lexicales interculturelles. *Actes de l'UPLEGESS.*

Champagnol, R. 1974. Association verbale, structuration et rappel libre bilingues. *Psychologie Française*, 19/1-2, p. 83-100.

Dell, G.S. 1989. The retrieval of phonological forms in production: Tests of predictions from a connectionist model. In Marslen-Wilson, W.D. (ed) *Lexical Representation and Process.* Cambridge MA: MIT Press, p. 136-165.

Erdmenger, M. 1985. Word acquisition and vocabulary structure in third-year EFL-learners. *IRAL*, 23/2, p. 159-164.

Ferrand, L. & Alario, F-X. 1998. Normes d'associations verbales pour 366 noms d'objets concrets. *L'Année Psychologique*, 98, p. 659-709.

Fromkin, V.A. 1971. The non-anomalous nature of anomalous utterances. *Language*, 47, p. 27-52.

Galton, F. 1879. Psychometric experiments. *Brain*, 2, p. 149-162.

Gougenheim, G., Michea, R., Rivenc, P. & Sauvageot A. 1956. *L'Elaboration du Français Elémentaire*. Paris: Didier.

Grosjean, F. 1989. Neurolinguists, beware! The bilingual is not two monolinguals in one person. *Brain and Language,* 36, p. 3-15.

Jenkins, J.J. 1970. The 1952 Minnesota word association norms. In Postman, L. & Keppel, G. (eds) *Norms of Word Association.* New York: Academic Press, p. 1-38.

Kent, G.H. & Rosanoff, A.J. 1910. A study of association in insanity. Part I: Association in normal subjects. Part II: Association in insane subjects. *American Journal of Insanity* 67, p 37-96; p. 317-390.

Kiss, G.R., Armstrong, C., Milroy, R. & Piper, J. 1973. An associative thesaurus of English and its computer analysis. In Aitken, A.J., Bailey, R.W. & Hamilton-Smith, N. (eds) *The Computer and Literary Studies.* Edinburgh: Edinburgh University Press, p. 153-165.

Kolers, P.A. 1963. Interlingual word associations. *Journal of Verbal Learning and Verbal Behaviour,* 2, p. 291-300.

Kruse, H., Pankhurst, J. & Sharwood-Smith, M. 1987. A multiple word association probe in second language acquisition research. *Studies in Second Language Acquisition,* 9/2, p. 141-154.

Lambert, W.E. & Moore, R. 1972. Word association responses: Comparisons of American and French monolinguals with Canadian monolinguals and bilinguals. In Lambert, W.E. (ed) *Language, Psychology and Culture.* Stanford University Press, p. 197-211.

Leech, G., Rayson, P. & Wilson, A. 2001. *Word Frequencies in Written and Spoken English: Based on the British National Corpus.* London: Longman. Companion website accessed 4/3/03: http://www.comp.lancs.ac.uk/ucrel/bncfreq/.

Lewis, M. 1993. *The Lexical Approach*. Hove: Language Teaching Publications.

Malrieu, D. & Gronoff, J-D. 1999. Associations libres et mémoire sémantique: Que nous apprend l'approche morphologique? Un retraitement du corpus de Kent et Rosanoff. *Cahiers de Lexicologie,* 75, p. 55-80.

Meara, P. 1978. Learners' word associations in French. *The Interlanguage Studies Bulletin,* 3/2, p. 192-211.

Meara, P. 1980. Vocabulary acquisition: A neglected aspect of language learning. In Kinsella, V. (ed) *Language Teaching and Linguistics Surveys. Volume I.* Cambridge: Cambridge University Press, p. 100-126.

Meara, P. 1982. Word associations in a foreign language: A report on the Birkbeck vocabulary project. *Nottingham Linguistic Circular,* 11/2, p. 29-37.

Meara, P. 1984. The study of lexis in interlanguage. In Davies, A., Howart, A. & Criper, C. (eds) *Interlanguage.* Edinburgh: Edinburgh University Press, p. 225-235.

Meara, P. 1992. Network structures and vocabulary acquisition in a foreign language. In Arnaud, P.J.L. & Béjoint, H. (eds) *Vocabulary and Applied Linguistics.* London: Macmillan, p. 62-70.

Meara, P. 2001. The mathematics of vocabulary. In Gill, M., Johnson, A.W., Koski, L.M., Sell, R.D. & Wårvik, B. (eds) *Language, Learning and Literature: Studies Presented to Håkan Ringbom.* Åbo: Åbo Akademi, p. 151-165.

Meara, P. & Fitzpatrick, T. 2000. Lex30: An improved method of assessing productive vocabulary in an L2. *System*, 28/1, p. 19-30.

Meara, P. & Schur, E. 2002. Random association networks: a baseline measure of lexical complexity. In Spelman Miller, K. & Thompson, P. (eds) *Unity and Diversity in Language Use*. *British Studies in Applied Linguistics*, 17. London: Continuum, p. 169-182.

Miller, K.M. 1970. Free-association responses of English and Australian students to 100 words from the Kent-Rosanoff word association test. In Postman, L. & Keppel, G. (eds) *Norms of Word Association.* New York: Academic Press, p. 39-52.

Moss, H.E. & Older, L.J.E. 1996. *Birkbeck Word Association Norms.* London: Psychology Press.

Postman, L. & Keppel, G. (eds) 1970. *Norms of Word Association.* New York: Academic Press.

Randall, M. 1980. Word association behaviour in learners of English as a second language. *Polyglot,* 2/2, p. B4-D1.

Riegel, K. & Zivian, I.W.M. 1972. A study of inter- and intralingual associations in English and German. *Language Learning,* 22/1, p. 51-63.

Rosenzweig, M.R. 1970. International Kent-Rosanoff word association norms, emphasizing those of French male and female students and French workmen. In Postman, L. & Keppel, G. (eds) *Norms of Word Association.* New York: Academic Press, p. 95-176.

Ruke-Dravina, V. 1971. Word associations in monolingual and multilingual individuals. *Linguistics,* 74, p. 66-84.

Schmitt, N. 1999. The relationship between TOEFL vocabulary items and meaning, association, collocation and word-class knowledge. *Language Testing,* 16/2, p. 189-216.

Schmitt, N. 2000. *Vocabulary in Language Teaching.* Cambridge: Cambridge University Press.

Schmitt, N. & Meara, P. 1997. Researching vocabulary through a word knowledge framework: Word associations and verbal suffixes. *Studies in Second Language Acquisition,* 19, p. 17-36.

Singleton, D. 1999. *Exploring the Second Language Mental Lexicon.* Cambridge: Cambridge University Press.

Söderman, T. 1989. Word associations of foreign language learners and native speakers: A shift in response type and its relevance for a theory of lexical development. *Scandinavian Working Papers on Bilingualism,* 8, p. 114-121.

Van Ginkel, C.I. & Van der Linden, F.H. 1996. Word associations in foreign language learning and foreign language loss. *Jyvaskyla Cross Language Studies,* 17, p. 25-33.

Walter, H. 2001. *Honni Soit Qui Mal Y Pense: L'Incroyable Histoire d'Amour entre le Français et l'Anglais.* Paris: Robet Laffont.

Wilks, C. & Meara, P. 2002. Untangling word webs: Graph theory and the notion of density in second language word association networks. *Second Language Research,* 18/4, p. 303.324.

Appendix

English primary responses (all groups combined)

N°	stimulus		Primary		Secondary		Tertiary	zero resp[1]	difft resp[2]
1	table	72	chair	18	desk	8	eat	5	26
2	dark	26	black	23	night	22	light	6	30
3	music	16	rock	11	song	11	sound	4	57
4	sickness	16	ill	7	illness	4	health	61	43
5	man	112	woman	7	girl	3	boy	1	16
6	deep	12	forest	8	hole	8	purple	47	34
7	soft	63	hard	16	(soft) ware	4	cool	17	31
8	eating	43	drink	17	food	7	meat	9	42
9	mountain	27	snow	10	sea	9	bike	3	52
10	house	29	home	15	garden	11	family	4	39
11	black	92	white	12	dark	8	colour	1	22
12	mutton	16	sheep	12	animal	8	cow	69	25
13	comfort	17	bed	15	sofa	7	armchair	30	39
14	hand	53	foot	30	finger	13	(hand) ball	8	17
15	short	63	long	8	little	5	small	7	39
16	fruit	41	apple	17	orange	15	vegetable	4	30
17	butterfly	10	flower	10	fly	9	bird	44	44
18	smooth	18	criminal	4	mouth	3	lip	84	27
19	command	20	(and) conquer	10	order	8	army	38	45
20	chair	64	table	13	sit	9	seat	17	20
21	sweet	15	home	13	shirt	7	good	39	38
22	whistle	5	castle	4	whisky	3	wind	90	33
23	woman	74	man	17	girl	5	beauty	2	28
24	cold	49	hot	18	warm	14	ice	11	27
25	slow	57	fast	21	speed	10	quick	10	25
26	wish	14	dream	11	hope	11	want	45	43
27	river	47	water	13	sea	10	ocean	14	29
28	white	76	black	10	colour	9	snow	4	30
29	beautiful	19	nice	17	woman	11	girl	17	46
30	window	28	door	17	computer	11	glass	11	44
31	rough	4	red	3	smooth	2	golf	107	24
32	citizen	19	city	16	person	7	town	32	50
33	foot	46	(foot) ball	24	hand	12	sport	3	31
34	spider	64	(spider) man	6	web	5	animal	13	33
35	needle	7	need	2	drug	1	animal	105	26
36	red	35	blood	21	blue	12	colour	5	36
37	sleep	58	bed	10	wake up	8	dream	11	31
38	anger	9	furious	4	cross	3	angle	69	43
39	carpet	13	magic	6	fish	4	dog	60	41
40	girl	65	boy	11	beautiful	9	man	4	31
41	high	27	low	15	mountain	7	building	22	41
42	working	27	sleep	21	hard	16	school	8	44
43	sour	2	ear	2	stone	1	blind	115	21
44	earth	14	world	10	(earth) quake	10	blue	19	46
45	trouble	27	problem	3	ill	3	peace	49	48
46	soldier	51	war	12	army	8	fortune	12	43
47	cabbage	6	garbage	5	bag	4	baggage	102	16
48	hard	57	easy	18	soft	12	difficult	10	30
49	eagle	35	bird	11	fly	10	eye	24	34
50	stomach	27	eat	8	food	6	(stomach) ache	50	38

1: Column 9: total number of zero responses, including blanks, illegible, and repeating stimulus

2: Column 10: total number of different responses for each stimulus

English primary responses (continued)

N°	stimulus		Primary		Secondary		Tertiary	0 resp	difft resp
51	stem	2	howard	2	stamp	2	step	123	11
52	lamp	69	light	3	petrol	2	electricity	44	22
53	dream	43	sleep	12	night	8	nightmare	13	38
54	yellow	41	sun	18	submarine	12	colour	7	32
55	bread	26	eat	12	butter	9	breakfast	46	35
56	justice	22	law	12	judge	7	balance	26	49
57	boy	87	girl	8	man	7	young	9	28
58	light	23	dark	23	sun	19	lamp	19	40
59	health	21	life	8	ill	8	sport	43	39
60	bible	42	religion	19	god	17	book	3	32
61	memory	18	head	13	brain	6	remember	25	54
62	sheep	13	cow	8	animal	7	white	60	35
63	bath	31	(bath) room	29	water	17	shower	27	25
64	cottage	14	house	3	country	3	field	82	34
65	swift	4	clean	3	swiffer	2	life	113	19
66	blue	34	sky	18	sea	14	colour	6	26
67	hungry	45	eat	12	food	7	angry	29	37
68	priest	19	religion	8	bible	6	church	82	19
69	ocean	25	water	20	sea	18	blue	5	32
70	head	15	eye	11	hair	9	brain	35	46
71	stove	6	love	2	rolling	1	fire	119	13
72	long	69	short	5	time	4	court	23	31
73	religion	25	bible	15	god	10	church	17	39
74	whiskey	53	alcohol	22	drink	9	good	26	26
75	child	14	parent	11	children	10	baby	15	46
76	bitter	4	beer	4	butter	3	dog	99	26
77	hammer	7	war (hammer)	4	car	4	hummer	73	42
78	thirsty	11	drink	5	hungry	5	twenty	82	28
79	city	43	town	8	village	7	citizen	10	38
80	square	18	garden	9	tree	7	park	39	42
81	butter	22	bread	13	(butter) fly	6	breakfast	46	30
82	doctor	16	health	12	nurse	10	ill	27	44
83	loud	7	music	6	noise	4	quiet	90	27
84	thief	4	bad	4	robber	4	steal	86	38
85	lion	28	king	26	jungle	15	animal	15	33
86	joy	16	happy	11	enjoy	4	(joy)stick	74	20
87	bed	79	sleep	17	(bed) room	4	night	19	19
88	heavy	27	(heavy) metal	6	hard	5	strong	68	26
89	tobacco	22	smoke	15	cigarette	10	bad	39	37
90	baby	21	child	11	mother	8	cry	27	51
91	moon	24	sun	20	night	13	earth	24	32
92	scissors	24	cut	4	knife	4	paper	93	15
93	quiet	12	silence	6	calm	5	noisy	49	49
94	green	22	grass	13	blue	10	colour	27	29
95	salt	16	pepper	15	sugar	7	sea	66	21
96	street	24	city	16	car	7	road	31	43
97	king	40	queen	13	lion	8	kingdom	31	32
98	cheese	14	hamburger	14	mouse	9	burger	42	35
99	blossom	2	blood	2	coat	2	heart	122	13
100	afraid	22	scare	12	fear	10	frighten	46	36

French primary responses (all groups)

N°	stimulus		Primary		Secondary		Tertiary	0 resp	difft resp
1	table	96	chaise	7	manger	5	bureau	0	21
2	sombre	54	clair	27	noir	16	nuit	0	26
3	musique	12	chanson	12	son	11	note	0	62
4	maladie	15	SIDA	12	cancer	10	médecin	0	45
5	homme	116	femme	3	fort	3	viril	1	18
6	profond	32	trou	15	mer	10	puits	0	52
7	mou	87	dur	4	molle	3	caramel	0	38
8	manger	64	boire	16	faim	5	dormir	2	37
9	montagne	33	neige	14	mer	11	plaine	0	47
10	maison	16	toit	10	famille	10	porte	1	55
11	noir	90	blanc	12	sombre	8	nuit	1	25
12	agneau	39	mouton	15	loup	14	brebis	0	36
13	confort	32	canapé	18	luxe	15	fauteuil	1	13
14	main	56	pied	39	doigt	18	bras	2	24
15	petit	98	grand	17	nain	2	gros	1	23
16	fruit	42	pomme	29	légume	12	orange	0	27
17	papillon	24	chenille	14	insecte	12	voler	0	45
18	lisse	66	rugueux	11	doux	11	plat	3	35
19	ordre	61	désordre	10	ranger	6	classer	0	39
20	chaise	68	table	16	asseoir	11	pied	1	27
21	doux	20	dur	10	lisse	7	laine	1	62
22	sifflet	32	arbitre	27	bruit	6	gendarme	0	44
23	femme	93	homme	8	belle	5	enfant	1	28
24	froid	66	chaud	19	hiver	15	glace	0	28
25	lent	85	rapide	16	tortue	12	escargot	0	22
26	désirer	54	vouloir	27	envier	11	aimer	0	29
27	fleuve	49	rivière	28	eau	8	lac	0	37
28	blanc	73	noir	23	neige	5	bleu	0	29
29	belle	35	moche	21	femme	15	bête	2	32
30	fenêtre	37	porte	12	rideau	11	vitre	1	43
31	rugueux	65	lisse	11	doux	10	dur	3	42
32	citoyen	14	ville	13	voter	8	citoyenne	2	46
33	pied	43	main	18	chaussure	13	jambe	0	31
34	araignée	42	toile	19	insecte	10	animal	0	32
35	aiguille	25	fil	21	piquer	14	couture	0	37
36	rouge	57	sang	17	noir	14	bleu	0	38
37	sommeil	44	dormir	32	lit	13	rêver	0	29
38	colère	18	énerver	14	rouge	10	calme	1	60
39	tapis	34	sol	9	volant	8	salon	2	56
40	fille	73	garçon	4	belle	4	enfant	2	39
41	haut	97	bas	8	immeuble	6	montagne	1	22
42	travail	13	dur	9	repos	8	santé	2	67
43	aigre	43	doux	20	vinaigre	12	acide	7	32
44	terre	15	mer	9	ciel	9	eau	0	55
45	difficulté	45	facilité	18	facile	6	dur	3	49
46	soldat	42	guerre	23	armée	12	plomb	0	41
47	chou	54	(chou) fleur	16	légume	10	bruxelles	3	38
48	dur	65	mou	10	doux	5	difficile	1	47
49	aigle	43	oiseau	14	rapace	12	royal	1	41
50	estomac	26	manger	17	ventre	16	faim	1	38

French primary responses (continued)

N°	stimulus		Primary		Secondary		Tertiary	0 resp	difft resp
51	tige	53	fleur	10	bâton	6	fer	5	47
52	lampe	74	lumière	15	chevet	12	ampoule	1	30
53	rêve	34	dormir	21	nuit	19	cauchemar	1	42
54	jaune	34	soleil	17	vert	14	œuf	2	35
55	pain	18	manger	13	mie	11	beurre	0	45
56	justice	27	loi	19	balance	9	tribunal	0	50
57	garçon	99	fille	4	jeune	3	homme	3	26
58	clair	61	sombre	12	foncé	12	obscur	2	33
59	santé	24	maladie	11	malade	9	fragile	2	47
60	évangile	39	religion	37	bible	8	dieu	4	33
61	mémoire	32	cerveau	22	souvenir	6	trou	0	60
62	mouton	42	laine	13	agneau	11	brebis	1	43
63	bain	32	douche	22	eau	11	mousse	1	38
64	villa	41	maison	10	riche	10	vacance	0	40
65	rapide	66	lent	6	vite	5	voiture	1	42
66	bleu	42	ciel	22	mer	17	océan	0	27
67	faim	38	manger	26	soif	10	nourriture	2	42
68	prêtre	35	église	24	religion	19	curé	1	42
69	océan	39	mer	23	bleu	11	eau	0	39
70	tête	24	cerveau	21	cheveu	12	pied	1	43
71	fourneau	35	pain	14	cuisine	11	chaud	1	31
72	long	74	court	21	large	5	grand	2	29
73	religion	14	dieu	12	croyance	12	église	2	57
74	cognac	82	alcool	7	whisky	5	boisson	3	35
75	enfant	20	parent	16	bébé	15	adulte	1	54
76	amer	24	acide	14	aigre	11	doux	3	46
77	marteau	33	clou	14	enclume	11	piqueur	2	38
78	soif	36	boire	35	faim	23	eau	1	29
79	ville	19	village	15	cité	8	campagne	1	59
80	carré	36	rond	24	rural	16	triangle	2	37
81	beurre	21	pain	15	tartine	13	mou	2	48
82	docteur	25	médecin	24	malade	11	maladie	1	48
83	bruyant	27	calme	17	silencieux	11	sourd	2	50
84	voleur	10	cambrioleur	8	brigand	8	police	3	66
85	lion	30	roi	19	tigre	16	jungle	1	40
86	joie	22	tristesse	18	heureux	16	bonheur	2	45
87	lit	50	dormir	13	sommeil	11	oreiller	1	36
88	lourd	73	léger	7	gros	5	plomb	4	43
89	tabac	32	fumer	31	cigarette	11	mauvais	2	31
90	bébé	30	enfant	9	naissance	7	biberon	2	54
91	lune	27	soleil	23	nuit	10	pleine (lune)	0	41
92	ciseaux	64	couper	15	couteau	10	papier	2	30
93	tranquille	45	calme	8	paisible	5	énerver	4	56
94	vert	27	herbe	15	jaune	12	bleu	1	35
95	sel	58	poivre	37	mer	11	sucre	4	27
96	rue	25	avenue	10	ville	10	voiture	3	51
97	roi	53	reine	16	lion	11	mage	2	36
98	fromage	16	lait	13	pain	12	odeur	1	50
99	fleur	23	rose	9	pétale	8	lys	1	59
100	effrayé	50	peur	29	apeuré	3	peureux	4	46