HAL
open science

A further note on simulating word association behaviour in a second language

Clarissa Wilks, Paul Meara, Brent Wolter

To cite this version:

Clarissa Wilks, Paul Meara, Brent Wolter. A further note on simulating word association behaviour in a second language. Second Language Research, 2005, 21 (4), pp.359-372. 10.1191/0267658305sr251oa . hal-00572083

HAL Id: hal-00572083

https://hal.science/hal-00572083v1

Submitted on 1 Mar 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Second Language Research 21,4 (2005); pp. 359–372

A further note on simulating word association behaviour in a second language

Clarissa Wilks *Kingston University,* **Paul Meara** *University of Wales Swansea* and **Brent Wolter** *Idaho State University*

This article explores some critical methodological and theoretical issues that emerge from recent research into word association behaviour in second language (L2) learners. The studies that we discuss here all use computer simulations as a tool to investigate L2 lexical networks, and to compare these networks with those of first language (L1) speakers. This article broaches some previously unacknowledged complexities in this kind of research, and draws attention to the importance of which assumptions are built into simulation models. The article queries some of the assumptions of our 2002 article (Wilks and Meara, 2002), and provides a reinterpretation of some of the data that we presented there. The article argues that simulation modelling forces us to make critical analyses of assumptions in a way that is not always necessary in less exacting experimental environments.

I Introduction

In an earlier article published in this journal (Wilks and Meara, 2002), we reported data from an experiment in which we tested the ability of first language (L1) English speakers to recognize associated pairs in small sets of French words. These students (Ss) were first year university students, who had previously studied French for seven years in school. The material used consisted of a 40-item questionnaire. Each item in the questionnaire comprised a set of five words randomly chosen from the Français Fondamental list: approximately the first thousand most frequent words in French excluding grammatical items (Gougenheim *et al*., 1956). The participants were instructed to read

Address for correspondence: Paul Meara, Centre for Applied Language Studies, University of Wales Swansea, Singleton Park, Swansea, SA2 8PP, UK; email: p.m.meara@swansea.ac.uk

© 2005 Edward Arnold (Publishers) Ltd 10.1191/0267658305sr251oa

each set of words and circle any two words in the set that they considered to be associated. Ss were provided with written instructions which asked them to read each set of words and to circle any two words that they considered to be associated, for example:

blouse *cheminée* *coûter* *feu* *tort*

The instructions stipulated that where informants perceived no links between any of the words in the set, they should write nothing. In addition, the instructions stated that where they considered more than two of the words of the set to be associated they should circle the two with the strongest link. Ss were instructed to work instinctively and to spend no more than 20 minutes on the entire exercise. Instructions were given in English to the learners of French who participated in the study and in French to the French native speakers. These instructions were developed over a number of studies, and seem not to encourage frivolous responses. A typical item might look like example one below:

1) blouse cheminée coûter feu tort

If they saw more than one pair of associated words in the set, the participants were instructed to circle only the two words with the strongest link. In example (1), for instance, they might circle *cheminée* (chimney) and *feu* (fire). If they found no links between any of the words they were instructed to write nothing, and continue to the next item.

Alongside this group of L1 English speakers, we also ran a group of L1 French speakers, who carried out the same task. Our intention was to compare the data of the L1 English speakers with the native speakers of French, and we expected, of course, to find that our L1 English speakers were less adept at identifying associated pairs than the L1 French speakers were. Not surprisingly, this turned out to be the case ($t = 6.47$, $p < .001$). The data we reported are presented in Table 1.

These data clearly confirm that there is a difference between the two subject groups, and the most obvious explanation of this difference is

Table 1 Mean hit rate per group

	Non-native speakers	Native speakers
Mean hits	19.00	30.90
Standard deviation	7.65	5.74
Number of items	40	40
Number of Ss	30	30

that the association network of the L1 group is 'denser' than that of the second language (L2) group, in the sense that L1 words have more associative connections than L2 words do. This density metaphor is one that frequently occurs in the literature on L2 word associations, but its implications are rarely developed. Aitchison (1987), for example, talks about the lexicon as 'a gigantic multidimensional cobweb', and most researchers appear content to operate on this descriptive level. Wilks and Meara, however, attempted to show that it was possible to move beyond imprecise metaphorical descriptions, and to develop more specific quantitative models instead. We did this by comparing the experimental data with data generated by an association simulator.

The simulator was a computer program that modelled a small lexicon in which each word was linked with a number of other words in the lexicon. The number of links between each word and the rest of the lexicon – the NLinks parameter – could be varied, and Wilks and Meara showed that the probability of two associated words appearing in a small set of words varied with the value of this parameter. We then used this data to look again at the data generated by real subjects, and estimated what the real data implied about the density of interword connections in the mental lexicons of our test-takers. Our initial guess had been that the L1 English speakers would have relatively few connections between words in their L2 lexicons, perhaps as few as four or five. However, the results generated by the simulator forced us to revise that estimate. We concluded that the data implied a much denser set of connections, even for L2 speakers, perhaps as many as 30 or 40 links for each word. Wilks and Meara (2002) considered the implications of this for the way we normally interpret word association data generated by L2 speakers, and we concluded that the density of connections between words would have to be considerably higher than most researchers assumed it to be. This had significant implications for the way we thought about word association networks in an L2.

II Rethinking the simulator

A number of people, notably Brent Wolter, the third author of this article, pointed out to us that our simulator had in fact made some very severe assumptions about the way associations in a lexicon might work, and

that this might have resulted in unrealistically high estimates for the number of links between words. In order to clarify this objection, we need to explain the detailed workings of the simulator.

Basically the simulator consists of a large array of numbers. We set the size of this array at 1000, an arbitrary number, but one which is sufficiently large for us to argue that our results did not merely apply to very small lexicons. Each element in the array represents one 'word' in the lexicon. The items are not in fact real words, merely numbers that stand in for real words. Next we set the NLinks parameter, which determines the number of associations each word has. Let us suppose, for the purposes of illustration, that this parameter is set at three. The program next selects three random numbers between 1 and 1000 for the first 'word' in the lexicon. These random numbers point to three other 'words' in the lexicon, and represent the three associates of word number 1. The program then repeats this selection process for all the other words in the lexicon. The result is a structure that looks like Table 2.

For this particular simulation, word 1 is linked with words 123, 145 and 160; word 2 is connected to words 99, 182 and 279, and so on. Other runs, making different random choices, would produce different numbers, but the basic structure would be the same. Setting the NLinks parameter at a different value, say 5, would generate five associations for each word rather than three.

The critical next step is for the simulator program to select sets of five words and check whether or not they contain a pair of associated words (a "hit event"). In Wilks and Meara (2002) we determined that a hit event had occurred if the number that identified one of the five selected words also appeared in the association list of another word in a set. This is shown in example (2):

2)	wd29	15	123	135	138	742	881
	wd367	29	421	435	567	665	678
	wd456	71	139	156	489	543	820
	wd552	81	140	172	495	681	729
	wd699	10	259	273	682	695	891

In example (2), we have a set of five words – 29, 367, 456, 552 and 699 – each directly linked to six other words in the lexicon. Word 29 appears in the association list for word 367, and this is taken to mean that there is a direct association between word 29 and word 367. Given these data, the simulator program would record a hit. Compare this with

Table 2 Part of a simulated lexicon with the association parameter set at 3

word 1	123	145	160
word 2	99	182	279
word 3	129	182	761
...			
word 999	135	856	687
word 1000	72	65	321

what we find in example (3). Here, we have the same set of five target words – 29, 367, 456, 552, and 699 – but we have altered the association lists, and none of these words appears in the association list of the other words in the set. Given the data in example (3), the simulator program would record a NoHit event.

3)	wd29	15	123	135	138	742	881
	wd367	99	421	435	567	665	678
	wd456	71	139	156	489	543	820
	wd552	81	140	172	495	681	729
	wd699	10	259	273	682	695	891

Our thinking at the time was that a strong associative link between two words would normally result in the index of one word appearing in the association list of the other. For instance, if word 29 had been BLUE, and word 367 had been SEA, we might well expect one of these words to appear in the association list of the other: associations of this type are easily identified in word association norm lists (e.g., Postman and Keppel, 1970). In modelling terms, 29 might appear in the association list for word 367, and 367 might appear in the association list of word 29. Either event would have been sufficient for the program to register a hit. We assumed that most of the associations recognized by our Ss would be strong associations of this type, and that this interpretation of an associative link was a reasonable way of modelling association behaviour.

Wolter argued that this is actually a very narrow interpretation of association behaviour. It is certainly the case that some common associations have close connections of this kind: RED elicits BLUE, for example, or BIG elicits LITTLE in this way. In terms of our model, this would mean that BLUE would appear in the association list for RED, and LITTLE would appear in the association list of BIG, and the word association norms confirm that this is the case. However, not all associations are of this type. Test-takers are relatively consistent in the

associations that they **produce**, but they appear to be much less consistent in the associations that they **recognize**. In recognition experiments, the number of idiosyncratic associations is generally rather high, and test-takers will often find associations between words that are only loosely connected with each other. For example, test-takers will frequently identify pairs of words as associates, even when the pairs do not appear in the standard word association lists. Given a word set like RUN, SLEEP, BUS, BLUE, CLOUD, test-takers will sometimes identify RUN~BUS as an associated pair (because they sometimes run for the bus), or SLEEP~BUS (because they usually sleep on the bus), or BLUE~BUS (because the buses in their town happen to be painted blue). Loose associations of this type are particularly likely to be reported if there are no other stronger associations available. This behaviour clearly implies a much looser definition of an association than the one we were working with in Wilks and Meara (2002).

Wolter suggested that one obvious possibility for scoring looser associations of this sort would be to recognize that real associations between words often rely on a common link shared with a third word. For example, BIRD might be associated with AEROPLANE because they both FLY. This association feels like a strong one, even though BIRD is not commonly associated with AEROPLANE, and AEROPLANE is not commonly associated with BIRD. The Edinburgh Associative Thesaurus (Kiss *et al.*, 1973) reveals that no-one offered AEROPLANE as a response to BIRD or vice versa in that dataset. Wolter further pointed out that it would be relatively straightforward to reprogram our simulator so that its definition of a hit event reflected these looser associative connections. Specifically he suggested that the simulator might be reprogrammed to register a hit if any one of the words associated with word X also occurred in the list of associates for word Y. In example (4), for instance, word 138 occurs in the association set for word 29 and for word 456, and this might be interpreted as showing that word 29 and word 456 are linked together indirectly in some way.

4)	wd29	15	123	135	138	742	881
	wd367	99	421	435	567	665	678
	wd456	71	138	156	489	543	820
	wd552	81	140	172	495	681	729
	wd699	10	259	273	682	695	891

Suppose word 29 was BIRD, and word 138 was FLY and word 456 was AEROPLANE, then the simulator would register that both BIRD and AEROPLANE share a common associate FLY.

III Results

Wolter's hunch that it would be easy to adapt our simulator to operate in this way is correct, and Table 3 presents a set of data that shows what happens when we make this adjustment.

The data is the probability that the simulator program will register at least one hit in a set of five randomly selected words when the number of associated words for each entry in the lexicon varies from 1 to 20. Each figure in the table is based on 1000 simulated word sets.

It is immediately obvious that these probabilities are considerably higher than the figures we reported in Wilks and Meara (2002). The data suggests that the probability of a hit increases rapidly as the NLinks parameter varies between 2 and 14; for values of NLinks above 15, the probability of finding a hit in a set of 5 words is very close to 1.

We can now use these figures to re-interpret the data we reported in Wilks and Meara (2002). Wilks and Meara reported that their L1 French speakers, the Native Speaker group, registered a hit on approximately 30 of their 40 trials, i.e., about 75%, but with a substantial standard deviation from this mean. The closest match to these figures in Table 3 is when the NLinks parameter is set to 11 links per word. Wilks and Meara reported that their L1 English speakers, the Non-native Speaker group, returned a mean hit rate of 19/40 (=47.5%) but with a very substantial standard deviation from this mean score. The closest match to these figures in Table 3 is when the NLinks parameter is set to 7 links per word.

Table 3 Probability of registering an associated pair in a random set of five words

Number of links per word	1	2	3	4	5	6	7	8	9	10
p (hit)	.037	.077	.143	.195	.310	.382	.442	.563	.654	.714
Number of links per word	11	12	13	14	15	16	17	18	19	20
p (hit)	.772	.819	.852	.909	.926	.957	.962	.979	.980	.983

Table 4 shows Wilks and Meara's original data alongside some further data generated by the revised simulator program. In this simulation, we ran a set of 60 cases. For half these cases, the NLinks parameter was set at 7, and for the other half, the NLinks parameter was set at 11. Each case consisted of 40 five-word stimulus sets. In this way, the simulation is an exact parallel of the real data collected in Wilks and Meara's study. Table 4 shows that the mean scores in the simulated data and the real data are very close, but the simulator seems to generate smaller standard deviations, which are about half the size of the standard deviations in the real data.

These results suggest that our new model generates data that is a reasonably close fit to the data generated by real test-takers. However, the implications of this new dataset suggest rather different conclusions from the ones we drew from our earlier attempt at modelling associations in L2 speakers. In our earlier model, the chances of finding a hit in a random set of five words is fairly low, and the only way to account for the relatively high numbers of matches reported by our test-takers is to argue that the number of links between words must be correspondingly high. In our new model, relatively high hit rates can be achieved with relatively small numbers of connections between the words in the lexicon. Clearly, this changes the game substantially. In our new model, it looks as though 11 links per word is sufficient to account for the native speaker data, while 7 links per word is sufficient to account for the data generated by the L2 speakers. The difference between L1 speakers and L2 speakers is still real, but it now appears to be much smaller than the difference we were suggesting in our earlier account.

Table 4 Real and simulated data, 60 cases, each tested on 40 five-word stimulus sets

	Mean hits	SD
Real data		
Native-speaker group	30.90	5.74
Non-native-speaker group	19.00	7.65
Simulation		
NLinks = 11	29.90	3.17
NLinks = 7	18.56	3.54

IV Discussion

Two general points seem to emerge out of this re-evaluation of our Wilks and Meara (2002) data.

1 Simulations as a test bed for research tools

The first general point is that simulations clearly have a role in the development of testing instruments that has not been exploited hitherto in SLA research. Specifically, simulations can sometimes be useful because they allow us to identify what experimental conditions are necessary to test the hypotheses that we are working with.

As a simple example of this problem, consider the large amount of research that has looked at incidental vocabulary acquisition in L2. Typically, this work has measured take-up of vocabulary as a result of reading extended texts (see also Huckin and Coady, 1999), and it has done this by presenting Ss with a short list of target words that the Ss did not know when they started to read the text, and measured whether they did know them as a result of having read the text. On the face of it, this looks like a good design: if Ss acquire words by reading, then they should have higher scores on the post-test than they do on the pre-test. In practice, however, most studies of this type report only minimal increases in vocabulary knowledge. It is not difficult to work out why this result keeps recurring. Simply by quantifying what we mean by incidental vocabulary acquisition, we can easily see that some testing instruments will not be sensitive enough to pick up vocabulary growth even if it really exists. Let us suppose that there is a 5% chance of an S picking up a new word from a single encounter in a text (Nagy and Herman, 1985). Now suppose that we run an experiment that involves 20 target vocabulary items each repeated once in the text that the Ss read. How much improvement would we expect to find between the pre-test and post-test? The answer is about 5% of 20 words, that is, just one word would be expected to show an improvement. Clearly, a difference of this size is very likely to be swamped by variation within the subject group, and would be very unlikely to show up as a reliable improvement in a set of experimental scores. Even if we tested 100 words in this way, we would still expect to find an increase of only five words between pre-test and post-test. This simple quantitative

model, then, allows us to say unequivocally that a testing instrument that uses a small number of target words (e.g., the 12-item test used in Hulstijn, 1992: Experiment 1) is not going to be adequate as a tool for evaluating this particular hypothesis.

When it comes to evaluating more complex claims about lexical organization, it becomes more difficult to develop relevant armchair simulations like the one described in the previous paragraph. For example, most people agree that the lexical organization of L2 speakers is different from the lexical organization of L1 speakers, and that the association structure of advanced L2 speakers will be rather different from that of less advanced L2 speakers, but it has proved rather difficult to pin down these differences in practice. The data reported in this article suggests that word association tasks might be capable in principle of demonstrating a difference of this sort, but considerable care needs to be taken in designing experiments to test these claims.

Suppose, for instance, that we have a general hypothesis that lexical organization becomes more complex when language learners spend an extended period in a country where their target language is spoken, and let us suppose that we want to test this hypothesis using a word association recognition task. Now let us suppose that we have a group of students who, before spending a period abroad, score 38% on a test like the one we described in Section III, i.e., they behave as if their L2 lexicons have an average of six links per word. If, as the simulations suggest, the difference between native speakers and advanced learners is typically about four additional associative links, then we might expect that an extended period abroad would increase the average number of links slightly, so let us hypothesize that our learners increase their lexical links from six to seven words as a result of their stay abroad. With a test like the one Wilks and Meara (2002) used, this improvement ought to show up as an increase in the test scores from 38% to 44%. On a 50-item test, however, this difference is a mere three additional hits: an improvement that would be very difficult to detect in noisy data. Even with a 100-item test, the model suggests that we would expect to measure an improvement of only 6 additional hits. The point here is that relatively small differences in the overt behaviour of the test-takers could be pointers to quite significant changes in the underlying structure of their lexicons.

Arguments like this suggest that developing good test instruments for evaluating hypotheses about vocabulary development may be more difficult than we have typically supposed. Simply comparing the associations of L2 learners and native speakers, using *ad hoc* lists of words, as much of the research in this area has done, begins to look like a very unsatisfactory approach to assessing L2 lexical competence. Indeed, blunt research tools of this kind may be intrinsically incapable of evaluating the hypotheses we think we are researching. Careful simulation studies provide a way of testing out the capabilities of these instruments before they are widely used in real experiments.

2 The complexities of simulation research

The second general point to emerge from this study is that working with simulations is perhaps not as straightforward as it might seem to be. Clearly, the results we get from a simulation study are only as good as the assumptions that go into the simulator. If the assumptions are totally wrong, then the results of the simulation that uses them will be completely meaningless. Fortunately, cases of this sort are generally easy to identify. More difficult to work with are cases, like the one reported here, where the assumptions built into the simulation model are not exactly wrong, but limit the possible range of outcomes in some significant way. In the simulations in Wilks and Meara (2002), we chose to model association processes by looking for stimulus words that occurred in the association list of other stimulus words. This was one of a number of plausible implementations that we could have chosen as a way of modelling the association process and, at the time, it did not strike us as problematical or contentious to model these processes in this way.

With hindsight, it is obvious that the very large number of connections between words implied by the results of that simulation should have made us query this particular set of assumptions sooner. The fact that we did not do this perhaps suggests that researching with simulations is very different from the kind of research that we are used to in SLA. Typically, SLA research involves developing a loosely formulated but broadly testable theoretical position, and evaluating it by collecting data from groups of L2 learners. Because empirical data of this kind is generally hard to collect, it is highly valued, and we often accept data

that is partial and fragmentary (small groups of Ss, small numbers of stimulus items, and so on) as a compromise. At the same time, empirical data of this sort is usually treated uncritically: the profession as a whole tends to take empirical findings very much at face value, and it is rare to find data of this sort exposed to close critical scrutiny. Very often we end up accepting data that is really quite unsatisfactory as evidence in support of one theoretical position rather than another.

In simulation research, the relationship between data and theory is much more complex. It is actually very easy to generate vast quantities of data, and because of this, it suddenly becomes very easy to evaluate many different theoretical positions in a way that is logistically impossible using experiments with real L2 learners. However, this facility comes at a price, and the price is that simulations require us to be absolutely explicit about the processes that we think we are modelling, and the actual implementation of a model becomes critical to the research process. The corollary of this is that we need to do much more ground work at the level of theory than we need to do when we are dealing with data collected from human test-takers. Specifically, we can afford to pay much more attention to the details of our theoretical assumptions, simply because the implications of these assumptions can be evaluated much more easily. Potentially, this introduces a significant shift into the way we think about second language acquisition in general, and second language lexicons in particular. Simulations force us to explore the effects of building different assumptions into our theories: by modelling in different ways the processes we are interested in, we can explore in considerable depth exactly how these processes contribute to the performance of our model. This, in turn, can often force us to re-evaluate some of the fundamental assumptions that we are working with, and this process, when pushed hard enough, can significantly alter the way we look at things.

In these studies, we have been attempting to model the underlying structures that generate word associations. Most work in this field takes it for granted that word associations are the result of direct connections between one word and another in a mental lexicon. Our first simulation showed that this specific assumption only works if we allow very large numbers of connections between words. Our second simulation showed that a model with fewer connections will also work, but only if we

significantly redefine what we mean by an association. On balance, we think that the second model is a more plausible model than the one we developed in Wilks and Meara (2002): it is mathematically more tractable than the earlier model, and it suggests that the learning burden faced by L2 learners is perhaps not that big. However, it also suggests that previous work on L2 word associations may have been very naive in assuming that word association behaviour was a direct reflection of the immediate connections between words. If this assumption is not correct, then much of the early work on word associations – and not just the work on L2 word associations – may need to be re-evaluated. This is clearly a far-reaching outcome, and bearing in mind the relatively small scale of these simulations, it clearly illustrates the power of this kind of work.

V Conclusions

Ideally, simulation research is an iterative process, in which we formulate many different models, and test them against each other in a competitive environment. Unfortunately, we do not really have a tradition of this kind of work in SLA. Research in L2 vocabulary acquisition, at least, is dominated by a relatively informal attitude towards theory, and the literature predominantly consists of a huge number of one-off studies (see Meara, 1992), which tend to reinforce the general lack of theory. These two factors have prevented the development of the kind of on-going, critical dialogue that is required to allow the vocabulary field to move forward in theoretical terms. The position is made all the more difficult by the fact that few people have the computational skills necessary to work critically through the implications of a simulation model, or to suggest how the behaviour of a model might be affected by changing its fundamental assumptions. The obvious solution to this problem would be for simulation research to become a standard component in the training received by young SLA researchers. We hope that this article will stimulate some readers to think about providing training of this kind in future.

VI References

Aitchison, J. 1987: *Words in the mind*. Oxford: Basil Blackwell.

Gougenheim, G., Michéa, R., Rivenc, P. and **Sauvageot, A.** 1956: *L'élaboration du français élémentaire.* Paris: Didier.

Huckin, T. and Coady, J. 1999: Incidental vocabulary acquisition in a second language: a review. *Studies in Second Language Acquisition* 21, 181–193.

Hulstijn, J. 1992: Retention of inferred and given word meanings: experiments in incidental vocabulary learning. In: Arnaud, P. and Béjoint, H., editors, *Vocabulary and applied linguistics*. London: Macmillan.

Kiss, G., Armstrong, C., Milroy, R. and **Piper, J.,** editors 1973: *An associative thesaurus of English and its computer analysis.* Edinburgh: Edinburgh University Press.

Meara, P. 1992: Vocabulary in a second language, Volume 3. Special issue of *Reading in a Foreign Language* 9(1), 761–837.

Nagy, W. and **Herman, P.** 1985: Incidental vs. instructional approaches to increasing reading vocabulary. *Educational Perspectives* 23, 16–21.

Postman, L. and **Keppel, G.,** editors 1970: *Norms of word association.* New York: Academic Press.

Wilks, C. and **Meara, P.** 2002: Untangling word webs: graph theory and the notion of density in second language word association networks. *Second Language Research* 18, 303–324.