# Connectionist Networks and Knowledge Representation: The Case of Bilingual Lexical Processing.

Michael S. C. Thomas.
Linacre College.

A thesis submitted for the degree of Doctor of Philosophy
at the University of Oxford.

Trinity Term, 1997.

This thesis contains approximately 110,000 words.

# Connectionist Networks and Knowledge Representation: The Case of Bilingual Lexical Processing.

Michael S. C. Thomas.
Linacre College.

Thesis submitted for the degree of Doctor of Philosophy.
Trinity Term, 1997.

## Abstract.

This thesis is concerned with the implications of distributed representation for models of bilingual lexical processing. A review of the empirical literature shows evidence that the bilingual has an independent 'mental dictionary' for each language. The evidence comes predominantly from repetition priming data and frequency effects in bilingual lexical decision tasks. However, there are some indications of between language similarity effects, whereby, for instance words behave differently if they exist in both languages. Two hypotheses are considered as an explanation for these effects: (1) they arise from the nature of the underlying representations. A connectionist model of bilingual lexical word recognition, based on Seidenberg and McClelland's (1989) reading framework, is introduced. This model stores both languages over a *single* set of distributed representations and can demonstrate both behaviour suggesting separate dictionaries as well as the relevant between language similarity effects; (2) the similarity effects arise from the nature of the control processes co-ordinating the operation of independent representations (e.g. separate dictionaries compete or co-operate in recognising words). Experiments are presented using English-French bilinguals, which explore the role of between language similarity in the bilingual's attempts to co-ordinate responses according to each of their mental dictionaries. It is concluded that both of the two hypotheses have some merit, but that the representational account is more satisfactory in its explicit specification and in its parsimony. However, some difficulties remain for the distributed account with regard to second language acquisition. It is not obvious how a second language may be introduced into a network already representing a first language without damaging the pre-existing knowledge. Some ideas are presented as to how this problem may be overcome. Finally, some more general conclusions are drawn regarding the relation of distributed representations to single route and dual route models of cognitive processes. It is speculated that this distinction may dissolve using certain sorts of learning algorithm constructed to avoid catastrophic interference.

# Acknowledgements.

I have many people to thank in the researching and writing of this thesis. Firstly, I owe everything to Sharon McHale, without whose support, encouragement, and patience I could never have got this far. I would like to thank my supervisors, Kim Plunkett and Alan Allport for their help and encouragement. Thank you to Neil Forrester and Denis Mareschal, the other two musketeers, for making my time at Oxford so enjoyable, and for many helpful discussions. Thanks to Derek Besner, Max Coltheart, Ken Forster, Glyn Humphreys, and David Plaut, for their advice and discussions concerning the arguments in this thesis. Thanks to Ramin Nakisa for his help in the complex land of corpus counts, to Roland Baddeley for discussions on neural representation, and to Sam Perry for discussions on task switching. Thanks to Ann Baker, Becky Dalton, Sue King, Karen Nobes, and Peter Ward in the Department of Experimental Psychology, for seamlessly maintaining the resources necessary for research. Thanks to all those at King Alfred's College for all their support, encouragement, and the time they gave me, to Tony, Alison, Mike, Tony, and Sandie. Thank you to Chris and Sarah Brunsdon for their friendship and generous hospitality. Thanks to Gillian Sebestyen and Kia Nobre for their friendship and support. Finally, thank you to my parents and family.

# Long Abstract.


Distributed representations have been employed in a range of models of human cognitive processes. In a distributed system, many computations are carried out using the same representational resource. This project is interested with finding the edges of distributed representations; that is, when should we see sets of computations as falling within the same distributed representational resource, and when should we see them as falling within separate resources. This question is examined with regard to a specific case study, that of bilingual lexical representation. Here the aim is to extend the existing monolingual distributed model of word recognition (Seidenberg and McClelland, 1989; Plaut, Seidenberg, McClelland, and Patterson, 1996) to the bilingual case. When we use distributed representations, does it look like the bilingual has two mental dictionaries (one for each language) or a single distributed dictionary containing both languages?

We begin the thesis by introducing monolingual theories of lexical representation: the core empirical evidence which constrains them, and the principal models. These are the serial search model, the interactive activation model, and the distributed model. We will later see that the serial search and interactive activation models have been extended to the bilingual case, but that this has yet to be attempted with the distributed model. It is noted that only the distributed model offers the potential to generate a parsimonious account of how language representations might be acquired.

In Chapter 3, we review the evidence regarding bilingual lexical representation. By and large this research has sought to discover whether the bilingual has one combined ‘store’ for their word knowledge, or separate stores for each language. We review three types of research: neuropsychological, psycholinguistic, and developmental.

The neuropsychological evidence shows some evidence of differential impairment of languages in bilinguals after brain damage, but none of the evidence is sufficient to demonstrate anatomically separate language systems (Paradis, 1995).

The psycholinguistic approach to the one or two stores question is to find out whether operations in one language affect later operations in the other language: for example, if I recognise a word in English, does that help me recognise its translation equivalent in French ten minutes later? (The answer is no). This is an example of a priming effect, and these tend to be employed with experimental tools such as the lexical decision task. The reasoning behind the psycholinguistic approach is as follows: if recognition in one language operates independently of the recognition in the other, then the stores must be separate; if there is between language priming, then the languages must be stored in a single system which can mediate these priming effects. When the empirical evidence is brought to bear, the conclusion is that bilinguals have independent representations of lexical knowledge for each language, but a common set of semantic representations (Smith, 1991). Operations accessing word form information do not transfer between languages. Operations accessing semantic information do transfer between languages. One or two complications to this picture are also explored.

The developmental evidence is of two types: the simultaneous acquisition of two languages, and the later acquisition of a second language. Infant studies regarding simultaneous acquisition do not turn out to be useful for resolving questions of representation (Genesee, 1989). Second language acquisition appears to produce a set of lexical representations similar to those acquired by simultaneous acquisition (Potter, So, Von Eckardt, and Feldman, 1984).

In the rest of Chapter 3, we consider existing models of bilingual lexical representation. There are a number of views: that monolingual models can cope unchanged with the bilingual case, merely relying on the difference between words in each language to distinguish them (Kirsner, Lalor, and Hird, 1993); that the serial access model may be extended by postulating separate word lists for each language; that the interactive activation model can be extended by connecting the word units of each language to a separate ‘language node’ so differentiating their behaviour (Grainger and Dijkstra, 1992). The crucial evidence put forward to distinguish the models relates firstly to the fact that bilinguals take time to switch between recognising words in each language, and secondly to the fact that words in one language will be recognised more slowly if they resemble words of the other language more closely than they resemble words in their own (known as between language neighbourhood effects). On the basis of each model’s adequacy in accounting for these effects, Grainger and Dijkstra conclude that an extension of the interactive activation model, with added language nodes, is most appropriate. Once again, however, these models are static, final state accounts. They do not consider how their representations might be developed.

In Chapter 4 we consider possible ways to extend the distributed framework to the bilingual case. We consider three hypotheses: The No Change (NC) model, The Bilingual Single Network (BSN) model, and the Bilingual Independent Networks (BIN) model. The NC model uses the monolingual system to learn the words in both languages. Unfortunately, the model cannot learn word forms which have a different meaning in each language (non-cognate homographs, such as PAIN and FIN in French and English), since networks are unable to learn two different mappings from the same input. Nor can the model account for the fact that between language neighbourhoods are inhibitory while within language neighbourhoods are facilitatory, since it does not support the within/between distinction. On these grounds, the NC model is discarded. The BSN model employs a similar architecture to the monolingual model, but tags each word by its language membership: both languages are stored in the same set of distributed representations. In the BIN model, information about each language is stored in a physically separate network.

Our aim is to evaluate the BSN and BIN models (although in the final analysis, we will question whether they must necessarily be distinct). To explore the implications of the BSN model, we then run a 'toy' simulation. A small connectionist network is trained on two word sets, and its internal representations examined. The results show that language information stored in the same network will interfere if it is similar. The literature is re-examined with this result in mind, and a large number of studies (approximately 30) are found demonstrating between language similarity effects in bilingual lexical processing.

We then examine the implications of developmental evidence for the BSN and BIN models. The Single Network model is appropriate for simultaneous acquisition (the network is trained on both languages at once) but has problems with explaining second language acquisition: How can a second language be introduced into a single network without overwriting the first language already stored there? (This is the so-called problem of Catastrophic Interference.) The Independent Networks model deals straightforwardly with second language acquisition (use a different network) but has difficulty in justifying how an infant would know to employ separate representational resources when exposed to a world with two languages.

In sum, any model of bilingual lexical representation must explain the mixture of evidence pointing to the independence of lexical representations, and evidence of between language similarity effects. The rest of the thesis takes between language similarity effects to be the key data to distinguish between the BSN and BIN models, and examines how the respective models might account for them. First the BSN model is evaluated by constructing and testing a computer model (Chapters 5-8). Then the assumptions of the BIN model are evaluated by empirical experimentation (Chapters 9-10). In Chapter 11, the respective claims of the two models are evaluated.

The BSN account starts by filling in a missing step in the argument. The BSN model will be evaluated by how well it simulates empirical findings on, among other things, between language priming effects in the lexical decision task. However, it is not clear that the *monolingual* framework has any consistent account of the range of priming effects found in this task. Since priming affects form half of the evidence used in bilingual lexical processing, this is a serious shortcoming. To rectify this situation, in Chapters 5 and 6, initial simulations are carried out to show how the monolingual distributed framework can explain priming effects. This requires pulling together the strands of many approaches into a unified account.

Next the BSN model is constructed. Natural languages are too complex to employ in an initial model. Thus a connectionist network is trained to map between the word forms and meanings for two artificially created mini-languages. Words are tagged by language membership, but the network stores both languages across the same set of distributed representations. The model's performance is examined on how accurately it generates meanings for words existing in one or both languages and for words with different orthographic characteristics. Next, patterns of priming within and between languages are examined. The results show that the model demonstrates both evidence of independence and also the key evidence of between language similarity effects. These results suggest that the BSN model may account for much of the data on bilingual lexical processing, without the need to postulate the structural modifications to the monolingual framework proposed by the BIN model.

In Chapter 8, we pursue a prediction of the BSN model that is at odds with the empirical data. This prediction is that non-cognate homographs will not show a between language priming effect. However, Gerard and Scarborough (1989) have reported just such an effect. We attempt to replicate this result in a priming study using English-French bilinguals. The subjects perform a 'language exclusive' lexical decision task. In this task, they are presented with a string of letters and must respond 'Yes' only if the stimulus is a word in the currently active language. The currently active language changes every 50 trials. Items are repeated within or between languages, and the patterns of priming examined. The results of this study support the BSN model: there is within language priming for non-cognate homographs, but no between language priming. We offer reasons why Gerard and Scarborough may have found the results that they did.

The BIN model can straightforwardly explain independence effects - it does so by postulating independent networks. But how does it account for similarity effects? One explanation is that they result from the way activity from each lexicon is co-ordinated. Bilinguals can generate responses from just one lexicon (as in language exclusive lexical decision task used above). Similarity effects must therefore arise because they cannot silence activity coming from the context-irrelevant lexicon. In the BIN model, similarity effects are thus explained by the way bilinguals control their lexical representations. In Chapters 9 and 10, we evaluate this idea. In Chapter 9, we review what is known about control mechanisms acting over mental representations, and then specifically about the control of bilinguals' lexical representations. Much of the evidence comes from language switching experiments. Subjects are required to switch between recognising or naming words in each of their languages. These studies typically show that subjects incur a time cost to switch between responding in each language. Yet in Stroop experiments, subjects appear unable to ignore irrelevant language information, suggesting that there *is no input switch* to toggle recognition processes between one language and the other. We formulate a hypothesis about what the switch cost represents, and then in Chapter 10, carry out two experiments exploring factors affecting the switch cost. In these experiments, English-French bilinguals switch between performing lexical decisions according to their English and French lexicons every other trial. In the first experiment, we explore the effect of lexical status on the switch cost - is switching slower if the word that appears exists in both languages? - answer, yes (though this is a non-significant trend). Is the cost sensitive to the word's meaning? - answer, no. In the second experiment, we vary the orthographic characteristics of the stimulus, and find that this has a marked effect on switch costs, particularly when they are nonwords. Finally, we find that the switch cost depends on subjects' relative skills in each language.

In the light of the results, it is concluded that between language similarity does influence control processes, at least as they are revealed by switching. This is taken as support for the BIN model. It is suggested that the switch cost does not reflect the operation of an input switch as such, but a cost of reconfiguring responses. This is an important finding, since evidence of language switch costs has influenced a number of bilingual models. Some ideas are offered about the nature of the reconfiguration process. Lastly, because it is unclear how control processes would operate in the BSN, it is concluded that, while the results support the BIN model, they cannot rule out the BSN. It too could experience reconfiguration costs as language context changes.

In Chapter 11, we evaluate the respective models. The simulations supported the BSN account of independence and similarity effects. The experiments supported the BIN view, but could not rule out the BSN model. We explore how this tension between single route and dual route accounts runs through other domains in psycholinguistics where distributed models have been used (e.g. in the past tense and naming models). Using lessons from those debates, we try to separate the BSN and BIN. On grounds of parsimony, we support the BSN model. But the model would seem to run up against the problem of catastrophic interference when explaining second language acquisition. We explore the nature of this problem, and offer possible solutions that would support the BSN. In doing so, we suggest that the BSN and BIN approaches may well converge if the appropriate learning algorithm and network architecture are used. That architecture would initially use homogeneous representations, but would then self-organise according to the demands of the task domain, in this case, into separate representations where languages were different, overlapping representations where they were similar. This is to speculate that the distinction between one and two route models is not a meaningful one, although much work remains to be done to ground this speculation.

With regard to our broader question, the edges of distributed representation are to be found where there is an absence of between task similarity effects. Such similarity effects are the hallmark of a single set of distributed representations performing two tasks. They can be found in consistency effects in word naming, consistency effects and overgeneralisation errors in past tense formation, and in the similarity effects that are found when bilinguals recognise or name words in each of their two languages.