A Simple Recurrent Network Model of Bilingual Memory

Robert M. French (rfrench@ulg.ac.be)
Department of Psychology,
University of Liège, 4000 Liège, Belgium

Abstract

This paper draws on previous research that strongly suggests that bilingual memory is organized as a single distributed lexicon rather than as two separately accessible lexicons corresponding to each language. Interactive-activation models provide an effective means of modeling many of the cross-language priming and interference effects that have been observed. However, one difficulty with these models is that they do not provide a plausible way of actually *acquiring* such an organization. This paper shows that a simple recurrent connectionist network (SRN) (Elman, 1990) might provide an insight into this problem. An SRN is first trained on two micro-languages and the hidden-unit representations corresponding to those languages are studied. A cluster analysis of these highly distributed, overlapping representations shows that they accurately reflect the overall separation of the two languages, as well as the word categories in each language. In addition, random and extensive lesioning of the SRN hidden layer is shown, in general, to have little effect on this organization. This is in general agreement with the observation that most bilinguals who suffer brain damage do not lose their ability to distinguish their two languages. On the other hand, an example is given where the removal of *a single node* does radically disrupts this internal representational organization, similar to rare clinical cases of bilingual language mixing and bilingual aphasia following brain trauma. The issue of scaling-up is also discussed briefly.

Introduction

One of the central questions in the field of bilingual memory involves the lexical organization of bilinguals' two languages within the brain. Are these languages organized in a highly independent, selectively accessed manner or rather in a highly overlapping, distributed manner, much like monolingual memory with twice as many words? In short, are there two lexicons or one?

Research throughout the past decade has left little doubt that, at least functionally, there is a great deal of interaction between the two lexicons. Numerous experiments have consistently shown evidence of inter-language priming and interference, even when bilingual participants are carefully put into strictly monolingual contexts. Even a partial list of the work on cross-lingual priming and interference is extensive and would include: Kolers (1966), Meyer & Ruddy (1974); Schwanenflugel & Rey (1986); Beauvillain & Grainger (1987), Grainger & Beauvillain (1988); Chen & Leung (1989), De Groot & Nas (1991), Beauvillain (1992), Keatly & de Gelder (1992), Grainger & O'Regan (1992), Keatley, Spinks, & De Gelder (1994), French & Ohnesorge, (1996, 1997), Fox (1996), and so on. Further, interlingual Stroop tests have also shown clear evidence of considerable cross-language interference (Dyer, 1971; Preston & Lambert, 1969; Mägiste, 1984).

In this paper, therefore, I will take for granted that these inter-language priming and interference effects are real and that some form of interactive-activation model (McClelland & Rumelhart, 1981) can provide a reasonable explanation for them. However, interactive-activation models of memory are localist connectionist networks that, in general, are not designed to learn. This is the case, in particular, for Grainger's (1993) Bilingual Interactive Activation model This paper explores the question of what type of distributed connectionist architecture, capable of learning, might be able to produce some of the effects whose explanation currently requires an interactive-activation framework. I will suggest that a simple recurrent connectionist network (SRN) — frequently referred to as an Elman network (Elman, 1990) — is an appropriate non-localist connectionist framework in which to study bilingual memory. This SRN network exhibits:

- progressive development of hidden-unit representations that cluster according to grammatical forms (subject, verb, object) and languages, even though there are no explicit markers on input distinguishing the languages or their grammatical forms;
- inter-lingual interference effects;
- considerable resistance to lesioning;
- significant disruption of internal organization that can be produced, on rare occasions, by lesioning a very small number of nodes.

Related work has been done by Cleeremans (1993) in his connectionist simulations of implicit learning of Reber grammars (Reber, 1967). In addition, there is currently at least one other non-localist connectionist model of bilingual memory. This is a model developed by Thomas & Plunkett (1995) and Thomas (1997). However, the latter model is not recurrent and therefore cannot be used to study the sequential acquisition of language. Kawamoto (1993) proposed a recurrent connectionist network to study word disambiguation. This model could almost certainly also be adapted to the case of bilingual memory.

Simulation Environment

The two micro-languages that were used were called Alpha and Beta. Each language consisted of 12 items: subject nouns, verbs and object nouns. These were broken down as follows:

Alpha

Subject Nouns: BOY, GIRL, MAN, WOMAN
Verbs: LIFTS, TOUCHES, SEES, PUSHES
Object Nouns: TOY, BALL, BOOK, PEN

Beta

Subject Nouns: GARÇON, FILLE, HOMME, FEMME
Verbs: SOULEVE, TOUCHE, VOIT, POUSSE
Object Nouns: JOUET, BALLON, LIVRE, STYLO

It is important to remember that the words "BOY", "FILLE", "VOIT", "PUSHES", etc. carry no semantic information. For the purposes of this simulation, I could just as easily have chosen single letters, or any other arbitrary symbols. The reason I chose these particular words was so that they, and the sentences produced by them, would be immediately identifiable as belonging to one language or another. Thus, we know BOY TOUCHES BOOK is from Alpha, whereas FILLE SOULEVE STYLO is from Beta.

Sentences in each language have the following simple SVO grammatical structure: $NOUN_{SUBJECT}$ – VERB – $NOUN_{OBJECT}$. A "language generator" (a finite-state machine) generates sequences of legal sentences in both languages (Figure 1). It is designed to simulate an Alpha-Beta bilingual environment. It has a fixed probability of 0.001 of switching from one language to another. This probability does not change during the course of a single run. In other words, if the switching probability is 0.001 at the beginning of the run, it will be the same 10,000 sentences later. Language switching was only permitted at the end of a sentence. (This constraint was relaxed in other experiments and the clusters of hidden-layer representations for both languages remained essentially the same as in the case where language-switching was only permitted at the end of sentences.)

BOY LIFTS TOY MAN SEES PEN MAN TOUCHES BOOK GIRL PUSHES BALL WOMAN TOUCHES TOY BOY PUSHES BOOK FEMME SOULEVE STYLO FILLE PREND STYLO GARÇON TOUCHE LIVRE FEMME POUSSE BALLON FILLE SOULEVE JOUET WOMAN PUSHES PEN BOY LIFTS BALL WOMAN TAKES BOOK...

No explicit markers between languages (or between individual sentences).

Figure 1. A typical Alpha-Beta language stream generated by the language generator. This stream of input will be fed to the SRN. Notice that, as in real (spoken) language, there are no explicit markers either between sentences or between languages.

Methodology

Each word in the sequence was presented to a 24-32-24 Elman network with a bias node. The learning rate was set at 0.1, momentum at 0.9, with a sigmoid squashing function and using a Fahlman offset of 0.1 (Fahlman, 1989). Input to the network consisted of individual words from a long string of sentences (Figure 1) generated automatically by a finite-state machine. For each word in the sequence, the network's task was to predict the following word. For example, in the sequence shown in Figure 1, the network would first get BOY on input and try to predict LIFTS on output, then LIFTS on input, trying to predict TOY on output, and so on.

Unlike standard sequence learning, the network never returns to the beginning of the sequence. A single weight change is made per input word (i.e., it learns for one epoch per presentation). The network has no hope of actually memorizing the two-language sequence because the sequence, as in real language, is non-deterministic. The expedient of localist input coding was used for the twenty-four words comprising the combined Alpha and Beta vocabularies. This coding was done as follows:

```
BOY  = 100000000000000000000000,
GIRL = 010000000000000000000000,
MAN  = 001000000000000000000000, etc
```

The Emergence of Language-Specific Clusters of Internal Representations

In this simulation we allowed training to continue until 300,000 items (100,000 sentences) had been seen. At various points in the run, the hidden-layer activation patterns were collected for each of the 24 words in the Alpha and Beta and were subjected to agglomerative hierarchical cluster analysis using a Euclidean distance metric and Ward's method to determine linkage.

Figure 2 shows the SRN's hidden-layer representations of Alpha and Beta. It can be seen that these representations are highly distributed and overlapping, but they are also clustered (Figure 3). After exposure to 60,000 items (Figure 3), stable clusters have developed that correspond not only to grammatical structures (Subject nouns, Verbs, and Object nouns), but also to each of the two languages. Alpha words lie in a distinct cluster from Beta words. It is in this sense that we can talk of language separation. In addition, this clustering has occurred in the absence of any explicit language (or sentence) "marking."

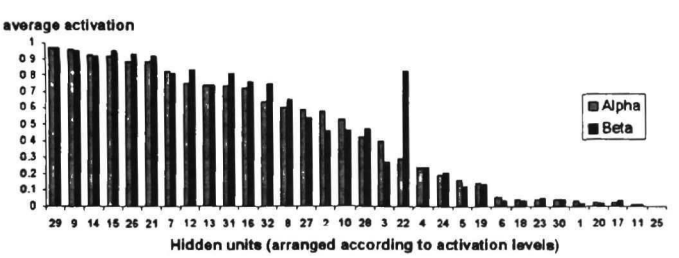

Figure 2: The highly distributed overlapping representations of Alpha and Beta at the hidden layer after exposure to 20,000 sentences (60,000 items).

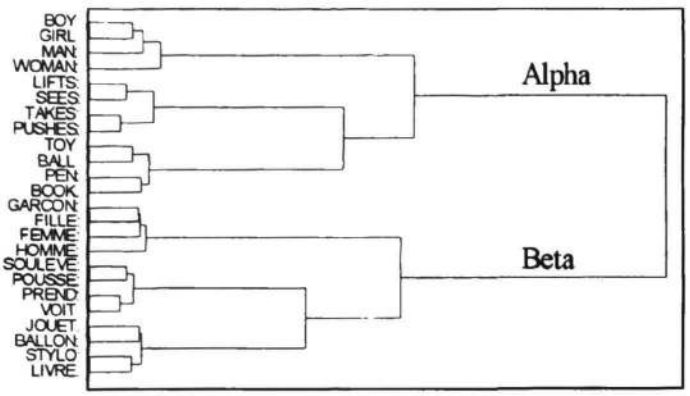

Figure 3. After 20,000 sentences (60,000 items) clusters have formed not only for the parts of speech in each language, but also *for each language*. The network has separated the two languages into distinct clusters of hidden-unit representations.

The Stability of the Language Clusters to Disruption by Lesioning of the Hidden Layer

In general, brain trauma in bilinguals does not result in the loss of one language or even extensive language-mixing. Connectionist networks modeling bilingual memory organization should also be able to display this ability of resistance to damage. In general, the ability to function when damaged is one of the most significant advantages of distributed connectionist systems. In the case of the present SRN model, once it had learned Alpha and Beta, it was very hard to disrupt the organization of the clusters it had developed. Following learning, nodes were removed from the hidden-layer and a cluster analysis performed on the activation patterns of the remaining nodes. In some cases, *up to 30 nodes* (out of 32) were removed and the organization of the representational clusters remained essentially unchanged.

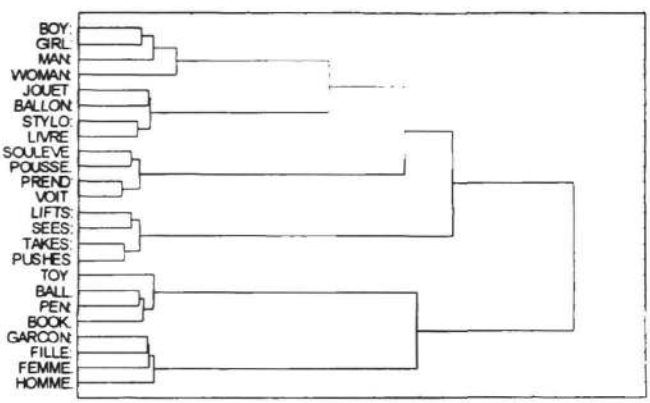

Figure 4. The separate Alpha-Beta language clusters are completely disrupted following the removal of Node 22.

On the other hand, in certain cases of brain trauma, fortunately quite rare, bilinguals can lose one of their languages or become incapable of distinguishing their two languages (Albert & Obler, 1977; Paradis, 1977; etc.). In the SRN model presented here, while this type of disruption is rare, it has been observed. The case presented in this paper (Figure 4) was provoked by *the removal of a single node* (Node 22) from the hidden layer after learning. If we refer to Figure 2, we notice a very large difference in the average activation of Node 22 for words in Beta compared to words in Alpha. It turns out that if this node (or any combination of nodes including this crucial node) is removed, the Alpha and Beta clusters disintegrate. Figure 4 shows the powerful effect on the language clusters of the removal of this node.

Discussion

It can be seen in Figure 2 that the representations for both languages are highly distributed and overlapping but what allows the network to distinguish Alpha from Beta is, ultimately, *differences* in the overall activation patterns for each language. In this case, it turns out that Node 22 accounts for 27% of this difference, compared to an average contribution of the other nodes of only about 3%. It is for this reason that its removal has such a significant effect on the overall organization of language clusters.

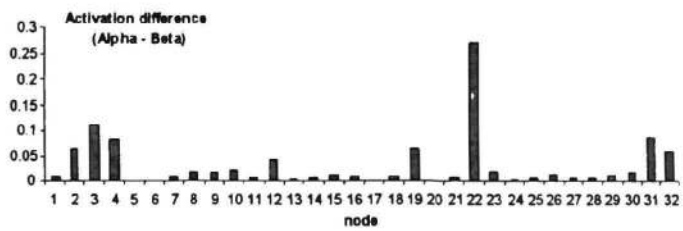

Figure 5. The activation difference between Alpha and Beta measured as the difference in the overall activation of all words in both languages.

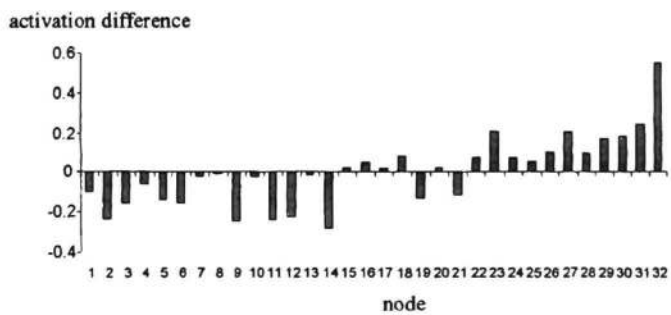

Figure 6. In general, the differences between the two sets of average representations for the words in each language are distributed over many nodes. Each node only accounts for a small fraction of the total difference and its loss would therefore be much less significant than the loss of Node 22 in Figure 5.

In most cases, the pattern of differences is much more evenly distributed over the entire hidden-layer. For

example, the distribution of differences shown in Figure 6 (from another run of the program), is far more typical. In the latter case the loss of a single node is not enough to seriously disrupt the overall difference between the two sets of representations.

The Decrease of Homographic Priming

French & Ohnesorge (1997) reported a disappearance of homographic priming for bilinguals in a mixed French-English condition compared to an All-English condition. They looked at a series of interlexical homographs —words that have distinct meanings in two separate languages; for example, words like, FIN (which means "end" in French), "PAIN" (= "bread" in French, etc.) — and paired them with words that they strongly primed in English. So, for example, the homograph FIN was paired with the target word SHARK, since in a word-nonword recognition task SHARK will be recognized to be a word much faster when it has been immediately preceded by FIN; we say that FIN *primes* SHARK. The homograph PAIN was paired with the target word HURT, and so on. The participants' task was to say whether the target word was a word *in English*. The

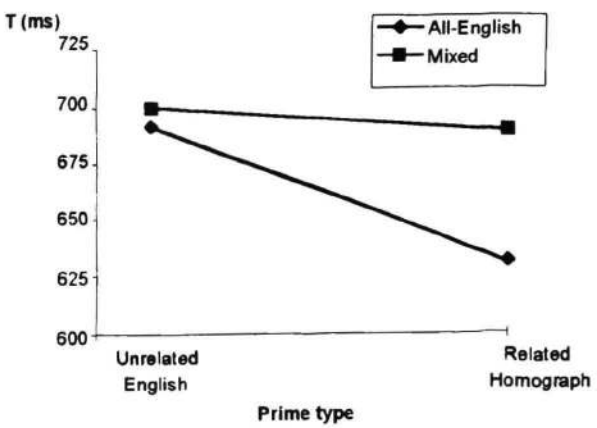

Figure 7. The substantial priming effect of SHARK by FIN in the All English condition is significantly reduced in the Mixed Condition.

amount of priming of the target words was compared in two conditions: an All-English condition and a Mixed condition (where the participants saw a mixture of half English and half French items). In the All-English condition, there was significant priming of the set of target words by the interlexical homographs. However, when French words were included in the Mixed condition (the task, identifying the item as a word in English, remained the same), the amount of priming was greatly reduced. The results in Figure 7 (French & Ohnesorge, 1997) show a priming effect of 62 ms in the All-English condition but very little effect (12 ms) in the Mixed condition.

"Spreading activation" in an SRN framework

Overlapping patterns of activation at the hidden layer can give rise to the perception of activation "spreading" throughout a language in an interactive-activation model. In other words, if the hidden-unit representation for BOY is active and the hidden-unit representation for GIRL is close by (i.e., the two representations have highly shared patterns of activation), this means that a very small change in the weights of the network can transform its internal representation of BOY into its representation for GIRL. On the other hand, the transformation from BOY to, say, SOULEVE, will be considerably more difficult since the respective representations of BOY and SOULEVE are considerably farther apart than BOY and GIRL, thereby requiring, on average, greater weight changes to transform one into the other. Clusters of representations are, by definition, activation patterns that are close together. It is therefore reasonable to suppose that if the hidden-layer activation clusters for each language are clearly separated (as in Figure 3), activation will "spread" within a language before it "spreads" to the other language,

Simulating the Decrease of Homographic Priming in the SRN Model

Based on this notion of spreading activation, we will say that in our SRN model, the amount of "priming" of a target word by a prime word will be determined by the how far the output of the network is from the target word after the prime word has been presented to it on input. So, for example, when a prime word W_p is presented to the network, an output vector X_p will be produced. This vector will be a certain Euclidean distance d_i from each word, W_i, known to the network. The word W_i that is best primed by W_p is the one for which d_i is the smallest. In the simulation described below, we trained the network on 10,000 sentences and then ran the priming tests.

For the following simulation, we first created an Alpha-Beta interlexical homograph similar to the French-English homograph, FIN. GARÇON and BOY were replaced by a single made-up word which, for no particular reason, I called TRAT. (Keep in mind that the "words" in Alpha and Beta have no semantic content. Their resemblance to French and English words serves only to identify the language they come from. Lexically speaking, TRAT could have come from both.) Recall that the French-English homograph FIN is a low-frequency word in English that strongly primes SHARK, and a high-frequency word in French. TRAT was designed to simulate this type of homograph. This was done as follows: 95% of the time that TRAT appeared in an Alpha sentence, it was followed by LIFTS. As a result, TRAT strongly primed LIFTS, according to the definition of priming given above. On the other hand, when the program was being given Beta sentences, TRAT was made to occur more frequently than other Beta subject nouns: When a Beta subject noun was needed, TRAT was selected 40% of the time instead of the usual 25%.

The network was first tested in an "All-Alpha condition." The All-Alpha context was created by giving

the network 10 Alpha words: five randomly chosen pairs made up of an Alpha input word and a legal Alpha successor to that word. The Mixed context was created by giving the network 10 Alpha or Beta words consisting of 5 randomly chosen legal pairs of either Alpha or Beta words. Learning remained on. Data was gathered over 100 independent runs of the program.

Using the definition of priming given above, it can be seen that in the SRN model priming of the word LIFTS by TRAT is significantly decreased in the Mixed condition (Figure 8). The Y-axis shows the Euclidean distance between the network's actual output and the target word.

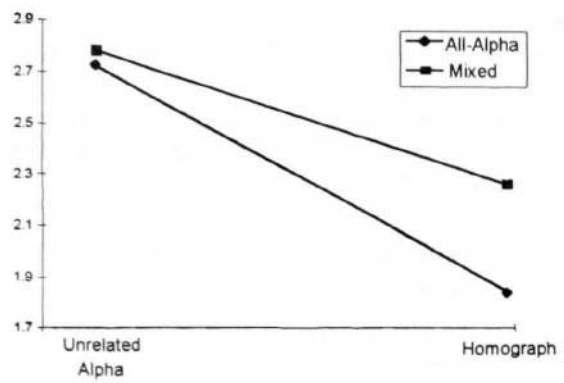

Figure 8. Priming of LIFTS by TRAT is significantly reduced in the Mixed Condition. The Y-axis represents the Euclidean distance between the output of the network after the prime word (either an unrelated Alpha word or TRAT) is presented and the word LIFTS. The X-axis shows the type of prime word. Compare with Figure 7.

Means for the English-Unrelated and Homograph-Related conditions were calculated for each run of the program and submitted to a mixed ANOVA. The interaction of Context(All-Alpha, Mixed) x Prime-Relatedness(Alpha-Unrelated, Homograph-Related) was significant, $F(1,183)= 4.4$, $p<0.05$. The priming effect in the All-Alpha condition was 0.85, compared to 0.49 in the Mixed condition.

In other words, using a canonical definition of priming as a measure of the proximity of the output of the network to the target, the SRN model exhibits a loss of priming comparable to that observed in real bilinguals.

Can the SRN Model Scale Up?

Real languages obviously contain more than 12 words and have a far more complex grammatical structure than either Alpha or Beta. Consequently, it is necessary to say a few words about the problem of scaling up.

The present model has been tested on as many as 1536 words (256 words per category) with the same simple SVO grammar. For the simulations involving 1536 words, the six categories of words were coded with 8 bits per word category, as follows:

```
Alpha_noun1=    0000000000.....0,
Alpha_noun2=    0000000100.....0;
Alpha_noun3=    0000001000.....0;
...               8-nodes
Alpha_noun256= 11111111000....0;
```

etc.

The network had 48 input and 48 output units and 100 hidden units. After exposure during learning to 30,000 sentences (90,000 items), the pattern of hidden-unit clusters shown in Figure 9 developed (calculated from a random sample of 100 words).

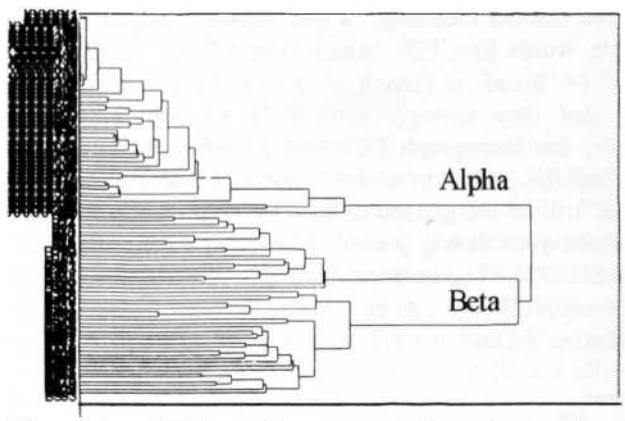

Figure 9. Hidden-unit representation clusters for 1536 Alpha and Beta words after 90,000 items (Alpha words are preceded by a series of X's to distinguish them from Beta words, which are preceded by a single "b")

The overall internal organization of the hidden-unit representations is basically the same as it was for the much smaller set of 24 words. Although a great deal more work needs to be done on this model, this result for over 1500 words would seem to indicate that there is at least some reasonable hope that the SRN model could scale up.

Conclusion

This paper suggests that bilingual memory phenomena that have been explained by an interactive-activation model of the BIA type (Grainger, 1993) may also be able to emerge from a simple recurrent connectionist network (SRN) model (Elman, 1990; Cleeremans & McClelland, 1991; Cleeremans, 1993). The SRN model presented here receives as input a long, undifferentiated sequence of sentences in two micro-languages and can reproduce a certain number of important effects that have been observed in studies of bilingual memory, in particular, those related to cross-lingual priming by interlexical homographs. In addition, the internal representations that the SRN develops reflect not only the divisions between the two languages, but also to the grammatical structure within each language. The organization of these internal representations is generally highly resistant to damage but, under certain exceptional circumstances, can also exhibit severe disruption following the removal of a as little as one node from the hidden layer.

Even though this initial work would seem to demonstrate that an Elman network might be able to serve as a useful

model for bilingual memory, it is also clear that a great deal of work still needs to be done. However, these initial positive results would suggest that further exploration of the SRN model of bilingual memory is justified.

Acknowledgments

This work was supported in part by Belgian science foundation grant PAI/4-19. The author would especially like to thank Michael Thomas and Axel Cleeremans for their numerous contributions to key ideas in this paper.

References

Albert, M. & Obler, K., (1978) "Neuropsychological studies of bilingualism" in *The Bilingual Brain*. New York: Academic Press.

Beauvillain, C. & Grainger, J. (1987) Accessing Interlexical Homographs: Some Limitations of a Language-selective access. *Journal of Memory and Language, 26*, 658-672.

Beauvillain, C. (1992) Orthographic and Lexical Constraints in Bilingual Word Recognition. In *Cognitive Processing in Bilinguals* by R. Harris (ed.) Amsterdam: Elsevier Science Publishers. 221-235.

Chen, H. & Leung, Y. (1989) Semantic facilitation and translation priming effects in Chinese-English bilinguals. *Memory and Cognition, 17*, 454-462.

Cleeremans, A. & McClelland, J. (1991). Learning the structure of event sequences. *Journal of Experimental Psychology: General, 120*, 235-253.

Cleeremans, A. (1993). *Mechanisms of Implicit Learning*. Cambridge, MA: The MIT Press.

De Groot, A. & Nas, G. (1991) Lexical representation of cognates and noncognates in compound bilinguals. *Journal of Memory and Language, 30*, 90-123.

Dyer, F. N., (1971) Color-naming interference in monolinguals and bilinguals. *Journal of Verbal Learning and Verbal Behavior, 10*, 297-302.

Elman, J. (1990) Finding structure in time. *Cognitive Science, 14*, 179-211.

Fahlman, S. (1989). Fast-learning variations on back-propagation: An empirical study. In *Proceedings of the 1988 Connectionist Models Summer School* (eds.) D. Touretzky, G. Hinton, & T. Sejnowski. 38-51. San Mateo: Morgan Kaufmann..

Fox, E. (1996). Cross-language priming from ignored words: Evidence for a common representational system in bilinguals. *Journal of Memory and Language, 35*, 353-370.

French, R. M. & Ohnesorge, C. (1996) Using interlexical nonwords to support an interactive-activation model of bilingual memory. In *Proceedings of the Eighteenth Annual Cognitive Science Society Conference*, New Jersey: LEA. 318-323.

French, R. M. and Ohnesorge, C. (1997). Homographic self-inhibition and the disappearance of priming: More evidence for an interactive-activation model of bilingual memory. *Proceedings of the 19th Annual Cognitive Science Society Conference*, New Jersey: LEA 241-246.

Grainger, J. & Beauvillain, C. (1988) Accessing interlexical homographs: Some limitations of a language-selective access. *Journal of Memory and Language, 26*, 658-672.

Grainger, J. & O'Regan, J. (1992). A psychophysical investigation of language priming effects in two English-French bilinguals. *European Journal of Cognitive Psychology, 4*(4) 323-339.

Grainger, J. (1993) Visual Word Recognition in Bilinguals. In *The Bilingual Lexicon* (ed.) R. Schreuder, B. Weltens. Amsterdam: John Benjamins Publishing Co.

Kawamoto, A. (1993) Nonlinear dynamics in the resolution of lexical ambiguity: A parallel distributed processing account. *Journal of Memory and Language, 32*, 474-516.

Keatley, C., Spinks, J., & De Gelder, B. (1994) Asymmetrical cross-language priming effects. *Memory and Cognition, 22*(1), 70-84.

Keatly, C. & de Gelder, B. (1992). The bilingual primed lexical decision task: Cross-language priming disappears with speeded responses. *European Journal of Cognitive Psychology, 4* (4) 273-292.

Kolers, P. (1966). Interlingual Facilitation of Short-term Memory. *Journal of Verbal Learning and Verbal Behavior, 5*, 314-319.

Mägiste, E. (1984) Stroop tasks and dichotic translation: The development of interference patterns in bilinguals. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10*, 304-315.

McClelland, J. & Rumelhart, D. (1981) An Interactive-Activation model of Context Effects in Letter Perception, Part 1: An Account of Basic Findings. *Psychological Review, 88*, 375-405.

Meyer, D. & Ruddy, M. (1974). Bilingual word-recognition: Organization and retrieval of alternate lexical codes. Paper presented at the meeting of the Eastern Psychological Association, Philadelphia.

Paradis, M., (1977). Bilingualism and Aphasia. In Whitaker, H. A. & Whitaker, H. (eds.) *Studies in neurolinguistics, 3*. New York: Academic Press.

Preston, M. & Lambert, W. (1969) Interlingual Interference in a bilingual version of the Stroop Color-Word Task" *Journal of Verbal Learning and Verbal Behavior, 8*, 295-301.

Reber, A. (1965). Implicit learning of artificial grammars. *Journal of Verbal Learning and Behavior, 6*, 855-863.

Schwanenflugel, P. & Rey, M. (1986) Interlingual semantic facilitation: Evidence for a common representational system in the bilingual lexicon. *Journal of Memory and Language, 25*, 605-618.

Thomas, M. & Plunkett, K. (1995). Representing the bilingual's two lexicons. *Proceedings of the 17th Annual Cognitive Science Society Conference*. NJ: LEA. 760-765.

Thomas, M. (1997). Knowledge representation and connectionist networks. Unpublished doctoral dissertation, University of Oxford, Cambridge.