Lexicon and Syntax in a Bilingual Connectionist Network

Valderes Aparecida Rinaldi (valderesrinaldi@yahoo.com.br)
Department of Linguistics, State University of Campinas
P.O.Box 6045, 13083-970 Campinas SP BRAZIL.

Ricardo Basso Garcia (rbgarcia@gmail.com)
Department of Linguistics, State University of Campinas
P.O.Box 6045, 13083-970 Campinas SP BRAZIL

Edson Françozo (edson@unicamp.br)
Department of Linguistics, State University of Campinas
P.O.Box 6045, 13083-970 Campinas SP BRAZIL.

Abstract

If bilingualism is a common phenomenon, research on bilingualism using recurrent networks is not. Few models using recurrent neural networks were developed in the past years in an attempt to address, among others, the question: is the lexical knowledge about two (or more) languages kept apart? French's (1998) results have shown that there are two separate regions in the lexicon, each one corresponding to one specific language. The present study extends French's model by including syntactic differences between the languages in addition to the exclusively lexical ones in the original simulation. Our network not only kept the lexicons apart, but also represented more fine grained structural differences between the languages in the hidden layer activation space.

Keywords: bilingualism; connectionism; language acquisition.

Introduction

Studies in bilingual knowledge and learning have grown into a full blown research field during the past twenty years (or so). In the past few years, experimental and sophisticated tools have allowed for considerable advances in the area, making room for the investigation of central questions about memory and language processing in bilinguals.

Computational modeling is also part of this landscape (for reviews, see French & Jacquet, 2004, and Thomas & van Heuven, 2005) and connectionist models are plenty: *BIA* (*Bilingual Interactive Activation*; Dijkstra & van Heuven, 1998), *BIMOLA* (*Bilingual Model of Lexical Access;* Grosjean, 1997), *SOMBIP* (*Self-Organizing Connectionist Model of Bilingual Processing;* Li & Farkas, 2002), *Hasta La Vista, Baby* (Scutt, 1997), *BSRN* (*Bilingual Simple Recurrent Network*; French, 1998) and *BLLCN* (*Bilingual Language Learning in Connectionist Networks*; Liang, 2004). Among these, Simple Recurrent Networks (SRNs) simulations allow for important aspects of the problem to be explored – on the one hand, the bilingual acquisition of language and, on the other, the dynamics of the mechanisms responsible for bilingual knowledge organization.

French (1998) put forward a simulation in which two small artificial languages are acquired by an SRN. Among the interesting results he obtained, it is of special interest to us the fact that in his simulation the very same network was able to separately represent the lexical knowledge of each language. Another result was that within each language cluster, word classes (i.e., subject nouns, verbs, and object nouns) were correctly identified and categorized. This was interpreted as an indication of the sensitivity of the system to the statistical regularities of the input, namely, the transitional dependencies present in word co-occurrence patterns in sentences.

Clearly, French's (1998) is a step towards a richer understanding of the dynamics underlying bilingual knowledge and organization. In special, it is a step towards understanding how lexical and structural bilingual distinctions are accomplished through time, much in the way posited by Elman (1990). In this connection, a question can be posed: what if syntactical complexity is added to the grammar of the model, so that languages are distinguished not only by their lexicons, but also by their different structures? It is hoped that by adding syntactical complexity light will be shed on the mechanisms underlying bilingual knowledge.

Computational simulations are of interest in connection with questions like this, because they offer the chance of manipulating variables believed to be active during bilingual language acquisition, but which cannot be controlled for in human data gathering.

The Simulation

An SRN was trained on two small artificial languages, just like in French (1998). For each of the languages, 4 adjectives were (optionally) added in the subject position. Thus, for each language, 16 words of were used (Table 1), resulting in 4 different syntactic structures (Table 2; notice that both objects and adjectives were optional). The intended syntactical difference between languages is thus brought about by differences in adjectives position: they always occurred before nouns in alpha and after them in beta (which, for mnemonic reasons, are referred to, respectively, as English and French). It must be noted that words and lexical categories do not carry any semantic content in the simulation.

Table 1: Lexicon for *alpha* and *beta*.

Alpha	
Adjectives	big, small, heavy, Chinese
Subject names	girl, woman, boy, man
Verbs	lifts, touches, sees, pushes
Object names	toy, ball, book, pen
Beta	
Adjectives	grand, petit, lourd, chinois
Subject names	fille, femme, garçon, homme
Verbs	souleve, touche, voit, pousse
Object names	juet, ballon, livre, stylo

Table 2: Sentence frames for *alpha* and *beta.*

Pattern	*Alpha*	*Beta*
1	S V	S V
2	S V O	S V O
3	A S V	S A V
4	A S V O	S A V O

Architecture and training

Using the categories and patterns from Tables 1 and 2, 5,000 sentences were generated, thus creating a bilingual environment similar to French's, except by the added grammatical complexity. Following Scutt (1997) and French (1998), no language difference was tagged nor was any semantic constraint furnished to the network. To simulate actual bilingual interaction, language switching was included (switch probability was set to 0.001). No language switching was allowed in the middle of a sentence.

A 33 node input (and output) network was created, with 44 nodes in the hidden and context layers. For both alpha and beta, a localist vector was used to represent the 32 word inputs to the network, with an extra vector representing the end of sentence. The learning rate was 0.1 and momentum 0.9. Training comprised 12 epochs (60,000 sentences).

Results

Once the network learned its task, testing was performed by presenting the same 5,000 sentences used in training. Cluster and Principal Components Analyses (PCA) were then performed on the hidden layer activations.

Learning Two Languages The inspection of Figure 1 makes it readily apparent that the network was able to separate the two languages as well as word categories within each language.

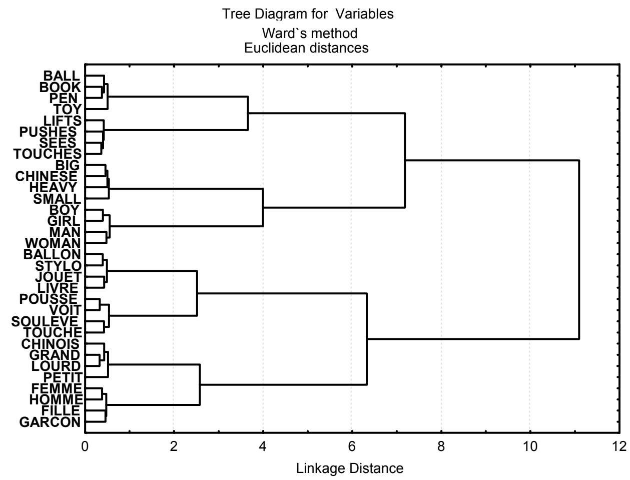

Figure 1: Cluster analysis showing the hierarchical knowledge after 12 epochs (60,000 sentences).

From the differences in linkage distance in Figure 1, it is also possible to argue in favor of the idea that language separation is prior to word categorization within each language. Thus, the present simulation achieved the same separation of languages as in French (1998).

Additional evidence concerning the representation of knowledge of the two languages can be gathered from the output of the network. At each cycle in a test run, we studied the average activation levels of words grouped according to their syntactic function, i.e., we added together the activation levels of all adjectives, subject names, verbs and object names, in both languages. In Figure 2, below, network predictions in instances in which English adjectives were fed to the input level are shown. The activation level of English subjects, the natural prediction when English sentences are started by adjectives, is by far the highest (approximately one third of the total activation).

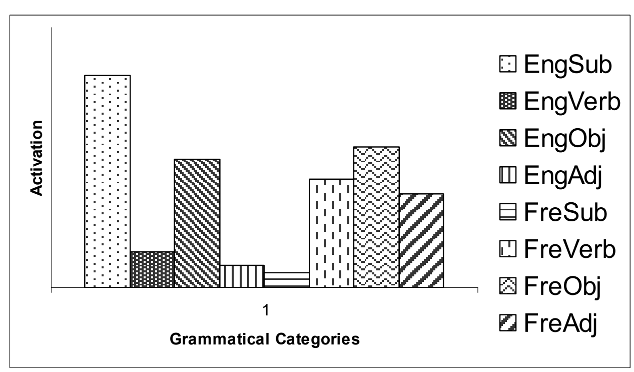

Figure 2: Average levels of activation of word categories when English adjectives are input to the network.

Overall, in Figure 3, the activation of French word categories is higher than English ones when French adjective words are input to the network. Moreover, French verbs are more activated than any of the other categories, which is an outcome of the fact that in the training set, French adjectives can only be followed by verbs.

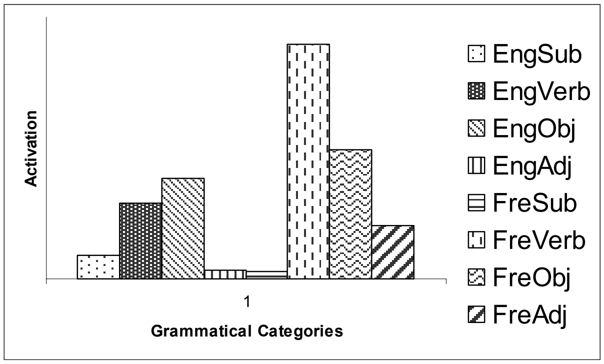

Figure 3: Average levels of activation of word categories when French adjectives are input to the network

One general trend that emerges from Figures 2 and 3 is that all word categories display some activation, even if not related to the language being processed. In Figure 2, for instance, an English adjective, while mostly predicting English subject nouns, also activates the other categories in English and French, as well. That is, both languages seem to be activated at the same time.

Syntax However, the task presented to the SRN was not as simple as French's, since categorizing adjectives, which could or could not occur in the different sentence positions in the two languages, posed an altogether different job to the network.

In order to account for this, one needs to examine the mechanism which, in time, performs sentence processing. By means of a PCA, we investigated the trajectories of pairs of sentences containing adjectives, one for each of the languages. In Figure 4, one such pair is depicted.

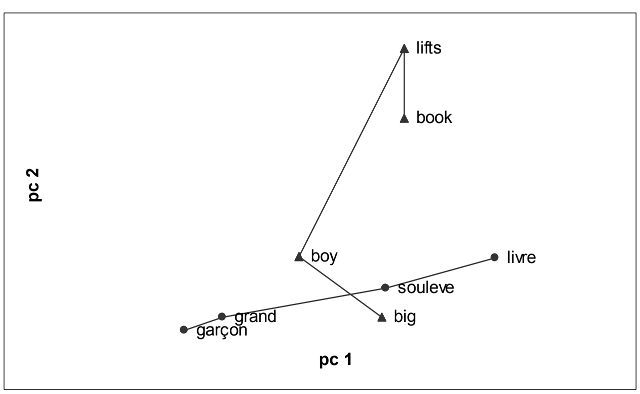

Figure 4: Trajectories in state space (factor plane 1x2) for *big boy lifts book* and *garçon grand souleve livre*.

In this figure, the trajectories for both languages are clearly distinguished, almost orthogonal, showing that language distinction is not only a matter of static differences regarding lexical representation, but also concerns different dynamics in sentence processing. One interesting feature of such dynamics is that subject nouns and adjectives, in both languages, are positioned closer to each other in comparison to verbs and object nouns. It seems clear, in this case, that the network abstracted from word order to some sort of structural knowledge, i.e., factor 2 seems to encode the fact that adjectives always come together with subject nouns, irrespectively of 'superficial' word order. In other words, the network has learned something about constituency: subject nouns and adjectives bound together to form a higher order constituent.

As far as trajectories between subject nouns and adjectives in factor plane 1x2 are concerned, they are opposed for the two languages – in English the system departs from the region of adjectives toward that of subject nouns; in French, it only reaches the very same region having departed from the subject nouns. That is, the opposing trajectories seem to reflect the differences in syntax or word order. The effect of word order is better viewed in yet another factor plane. Consider Figure 5, below.

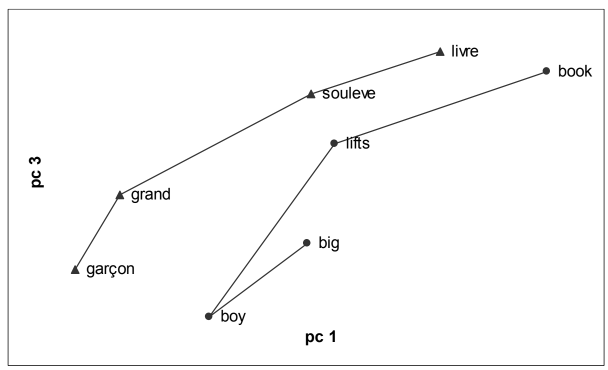

Figure 5: Trajectories in state space (factor plane 1x3) for *big boy lifts book* and *garçon grand souleve livre.*

Now, the trajectories are similar for both languages, even though adjective order is different. That is, in plane 1x3, the surface order of the sentence determines the path of the system. An additional feature of the system is also apparent in this figure: subject nouns, adjectives, verbs, and object nouns are placed in distinct "slices" of plane 1x3 (from left to right and from bottom to top), irrespective of language. This is an indication that different components, within the same system, simultaneously code for different features in the training set: two languages, four syntactical functions, two different word orders, and so on.

In order to further check the dynamics of the system, we decided to check what happens when a syntactically anomalous sentence is processed. In Figure 6, a comparison is made between *big boy lifts book* and *boy big lifts book**.

When canonical word order is changed, the trajectory in state space 1x2 is consequently adapted to reflect the new context. In this case, not only trajectories change, but also the representation of the adjective that is brought closer to the verb and farther from the subject noun (in comparison with the legal sentence). We believe this shows how the system copes with grammatical knowledge disruption – in this case, a disruption in constituency, and not simply in word order.

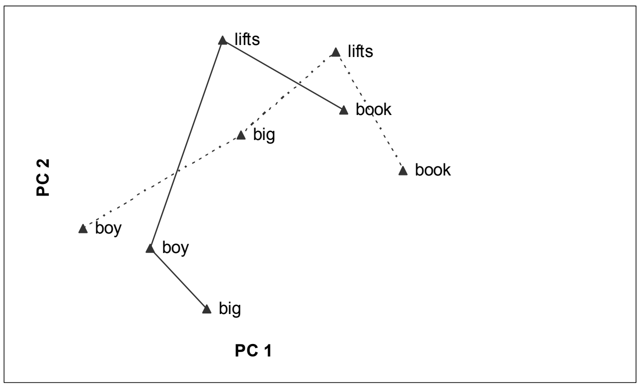

Figure 6: Trajectories in state space (factor plane 1x2) for *boy big lifts book** and *big boy lifts book.*

To sum up, while the hierarchical cluster analysis revealed how the network statically represented lexical knowledge with the state space, PCA let us see how, dynamically, the network processes sentences in both languages. In sentence processing, grammatical knowledge appears as trajectories in the state space of the system which are, *inter alia*, capable of coding for constituency. Insofar as constituency in both the languages in our study differs, different trajectories do portray language differences.

Syntax and Code Switching One interesting question, as far as bilingualism is concerned, is code switching. In Figure 7, below, we present the trajectories for *garçon grand souleve livre* and four instances of code switching, each of them in a different syntactic position.

Three of the code-switched sentences are gathered in the same region of the state space as the original sentence. But they display variations in trajectories that are due to code switching in different syntactic positions. As can be seen, code switching causes disturbances in the dynamics of the system.

In the case of *boy grand souleve livre** (Figure 7), an English word in the beginning of a French sentence led the system to behave differently in relation to the other three language mixed sentences. In the present case, it seems that the English word tricked the system into a different solution, that is, into a different state space region. If we compare the trajectory of this language mixed sentence with *[big] boy lifts book* in Figure 4, it seems that the English word *boy* has forced the system to solve the French sentence in the region normally employed to process English sentences. This can be thus seen as evidence of the importance of context in the bilingual processing of sentences.

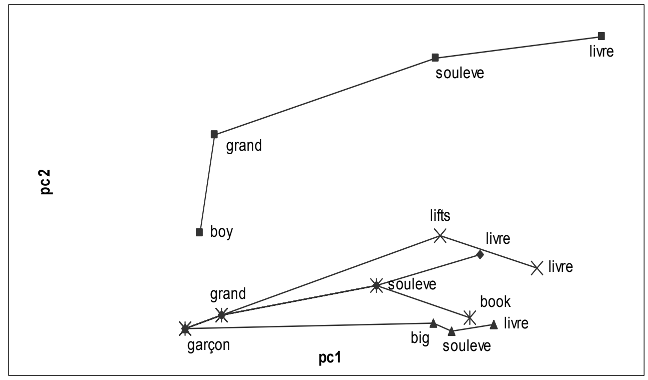

Figure 7: Trajectories in state space (factor plane 1x2) for sentences with language switching in French (*boy grand souleve livre**, *garçon big souleve livre**, *garçon grand lifts livre**, and garçon grand souleve book*)

General Discussion

We would like to highlight the role we believe language switching had in the experiment. In bilingual language acquisition, a child interacts with others who speak only one language at a time. For instance, it is reasonable to imagine that a bilingual learner uses one language at school and another at home, or that she employs different languages while interacting with each of her parents. As a consequence, simulation studies of bilingual language acquisition have to account for this in model design. Scutt's (1997) simulations of bilingualism, although quite realistic as far as the scope of its coverage of syntactic phenomena is concerned, have employed a training regimen which did not take language switching into consideration (language varied freely within the training set). His failure to obtain language separation, it seems to us, stem from overlooking the importance of this variable in bilingual language acquisition. In contrast, both our study and French's (1998) may have been able to obtain language separation by explicitly modeling language switching.

Another point of interest is the outcome of our experiment with code switching, which highlighted the role of language context in bilingual sentence processing. For instance, the literature on visual word recognition suggests that bilinguals use language context information to facilitate word recognition (e.g., Grainger, 1992; Scarborough et al, 1984). The same kind of facilitation may be operative in our coded switching experiment. As can be seen in Figure 7, whenever the code switching occurs in any but the initial word of a sentence, only small disturbances in trajectories arise. This could be credited to the system having contextual information about the language (in this case, French) being processed. On the contrary, in the case of code switching in the first position, the context establishes the expectation that an English sentence will follow; however, since the continuation is in French, remarkable shifts in trajectories result.

A further point of interest concerns the levels of activation of both languages in the system (cf. Figures 2 and 3). The relevant question is whether bilinguals process languages in parallel or not. Recent investigation on bilingual language processing using brain imaging and eyetracking (Marian et al., 2003) suggests that bilinguals can activate both languages in parallel even when in monolingual situations. Our simulation concurs with such a view.

One final point concerns the inclusion of grammatical complexity in our simulation. It is a fact that, in natural bilingual learning, languages vary not only in their lexicons, but in structure as well. Thus, including adjectives in the present simulation lends strength to the claim that connectionist simulation is suitable for the study of at least certain aspects of bilingual language acquisition. But the complexity which was added was quite modest, and warrants no evidence concerning, for instance, the role of grammatical knowledge in bilingual language acquisition. This was felt as a needed cautionary step leading toward a system which is veridical, enjoys data contact and input representativeness (Christensen & Chater, 2001).

Acknowledgments

Research supported by FAPESP (01/00136-2) and CNPq.

References

Christiansen, M. & Chater, N. (2001). Connectionist psycholinguistics: capturing the empirical data. *Trends in Cognitive Sciences*, *5*: 82-88.

Dijkstra, A., & Van Heuven, W. J. B. (1998). The BIA model and bilingual word recognition. In J. Grainger & A. M. Jacobs (Eds.), *Localist Connectionist Approaches to Human Cognition*. Mahwah, NJ, USA: Lawrence Erlbaum Associates.

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

French, R. M. (1998). A Simple Recurrent Network Model of Bilingual Memory. In Gernsbacher, M. A., & Derry, S. J. (Eds.), *Proceedings of the Twentieth Annual Cognitive Science Society Conference* (pp. 368-373). NJ: LEA.

French, R. & Jacquet, M. (2004). Understanding bilingual memory: models and data. *Trends in Cognitive Science*s. *8*(2), 87-93.

Grainger, J. (1992). Visual Word Recognition in Bilinguals. *In The Bilingual Lexicon (ed.)* R. Schreuder, B. Weltens. Amsterdam: John Benjamins Publishing Co.

Grosjean, F. (1997). Processing mixed language: Issues, findings and models. In de Groot, A. and Kroll, J. (Eds.). *Tutorials in Bilingualism: Psycholinguistic Perspectives*. Mahwah, NJ: LEA.

Li, P., & Farkas, I. (2002). A self-organizing connectionist model of bilingual processing. In Altarriba, R. H.J. (Ed.), *Bilingual Sentence Processing* (pp. 59-85). North Holland: Elsevier Science Publisher.

Liang, J. (2004). Bilingual language learning in connectionist networks. *Proceedings of the Sixth International Conference on Cognitive Modeling*. Mahwah, N.J.

Marian, V., Spivey, M. & Hirsch, J. (2003). Shared and separate systems in bilingual language processing: converging evidence from eyetracking and brain imaging. *Brain and Language, 86,* 70-82.

Scarborough, D.L., Gerard, L. & Cortese, C. (1984). Independence of lexical access in bilingual word recognition. *Journal of Verbal Learning and Verbal Behavior, 23*, 84-99.

Scutt, T. W. Hasta La Vista Baby: 'Bilingual' and 'Second-Language Learning' in a Recurrent neural network trained on English and Spanish, paper presented at the 1997 GALA Conference, Edinburgh.

Thomas, M. S. C. & Van Heuven, W. (2005). Computational models of bilingual comprehension. In J. F. Kroll & A. M. B. De Groot (Eds.) *Handbook of Bilingualism: Psycholinguistic Approaches*. Oxford University Press