Bilingual representation and learning in a connectionist framework

CS784: Language Acquisition

Supervisors:

Prof. H Karnick
Prof. A M Raina

Arindam Biswas (Y1072)
Indradeep Saha (Y1162)

Introduction

Over half of the world's population is bilingual. Thus bilingualism being a world-wide phenomenon merits great attention. Bilingual memory research started about half a century ago but it was only in the last decade that sophisticated experiments and computational techniques including use of connectionist frameworks let researchers come up with interesting results and answers. The central and pertinent question which remained in the limelight of research and this is whether bilingual speakers have two separate lexicons, one for each language, or one large bilingual lexicon. In other words, whether the bilinguals have two mental dictionaries to recognize the words in a language or a single integrated mental dictionary. Another question which cropped up was if there are separate conceptual and lexical levels in the memory of a bilingual. Researchers more or less agree on the presence of a shared conceptual level but specific lexical representations for each language.

In our project we have simulated the Bilingual Simple Recurrent Network which provides a possible way of acquiring a bilingual organization within a single distributed lexicon in memory. A simple recurrent network is trained on two small languages of 12 words each and it is shown after performing a cluster analysis on the hidden-unit representations of the SRN that indeed there is a separation at the language level and even at the word level.

The bilingual simple recurrent network (BSRN) model of bilingual learning

Robert M. French in the year 1998 first showed that the connectionist framework of bilingual memory can be achieved by a simple recurrent network.

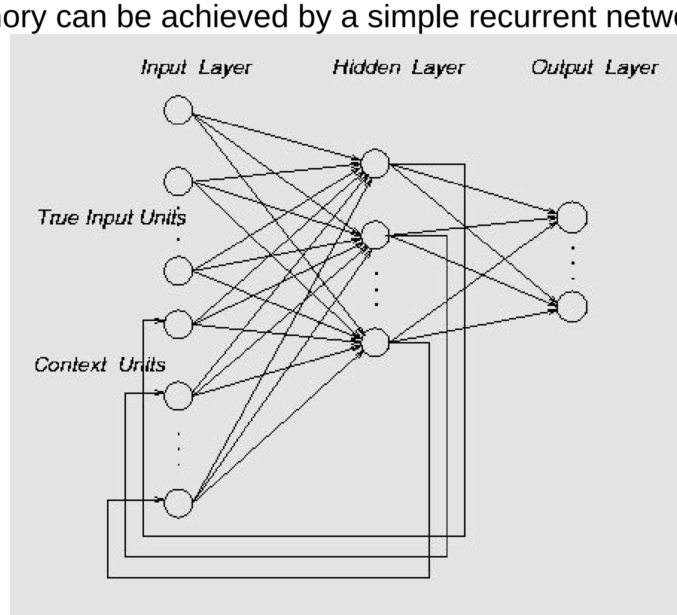

Fig. 1. Elman Network

BSRN is based on an Elman network. Elman network is nothing but a simple two layer recurrent network. As shown in the figure the output from the hidden layer is fed into the network along with true inputs. These feedback units are termed as context unit. Due to this recursive nature an Elman network is suitable for predicting or detecting time varying patterns. It is sensitive to statistical regularities in sequences. No semantics is required in this model.

Procedure:
In the simulation of BSRN we have chosen two micro languages alpha and beta each consisting 12 words each. These 12 words are divided into 3 categories subject-noun, verb and object-noun.
The structure of alpha and beta is as follows:

Alpha	**Beta**
Subject Nouns: BOY, GIRL, MAN, WOMAN	Subject Nouns: GARÇON, FILLE, HOMME, FEMME
Verbs: LIFTS, TOUCHES, SEES, PUSHES	Verbs: SOULEVE, TOUCHE, VOIT, POUSSE
Object Nouns: TOY, BALL, BOOK, PEN	Object Nouns: JOUET, BALLON, LIVRE, STYLO

Using dictionary words like BOY GIRL carry no semantic meaning. Any words even only letters can be used in this purpose. The only reason to choose such dictionary word is to make sure that language of the sentences generated from these words would be easily recognizable.

We have used a language generator to generate continues sequence of strings containing sentences randomly from the two languages. The language stream was generated using the following rules.

- The sentences generated are of the form [Subject-Noun] [Verb] [Object-Noun].
- The language switching probability is set to 0.001 to make sure the network get input from a particular language for reasonable amount of iteration.
- No language switching was allowed in the middle of the sentence.

The stream generated had no explicit markers either between sentences or between languages. So what the SRN sees is:

BOY LIFTS TOY MAN SEES PEN WOMAN TOUCHES BOOK GIRL PUSHES BALL FEMME SOULEVE STYLO FILLE PREND LIVRE GARCON VOIT BALLON WOMAN SEES BOOK BOY PUSHES PEN…

Training the network:

In our simulation we have created a 24-32-24 Elman network i.e. network with 24 input nodes, 32 hidden nodes and 24 output nodes. The learning rate was sate to

0.1 and momentum was set to 0.9. At each iteration one word from the stream is fed into the network. Before feeding the words into the network they are coded into 24 bit binary digits with i_{th} word coded as 1 at i_{th} position and rest 0. As for example:

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

The task is to predict the following word (e.g. given BOY on input, produce LIFTS on output; given LIFTS on input, produce TOY on output; etc.). In our simulation we simulated the model for different number of iteration and did hierarchical cluster analysis to see the pattern of the cluster of the hidden nodes. The stable cluster was formed after 90,000 iterations i.e. after encountering 30,000 sentences. The following diagram gives the cluster pattern of hidden nodes after 90,000 iterations.

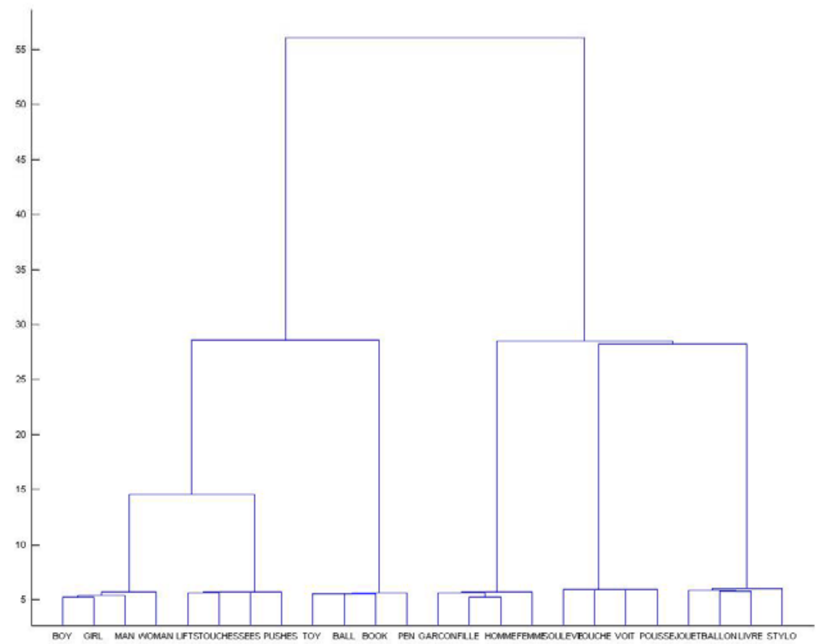

Fig. 2. Cluster of the hidden node activation patterns after 90,000 iterations

Figure 3 and Figure 4 gives closer view of the two bigger clusters formed.

The clusters have formed not only for the parts of speech in each language, but also for each language (Figure 2, 3 & 4). The network has separated the two languages into distinct clusters of hidden-unit representations.

BSRN can also explain the effect on bilingualism in case of brain trauma in bilinguals.

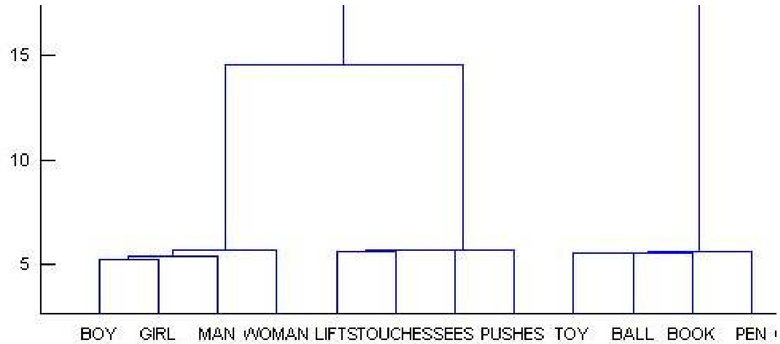

Fig. 3. Left cluster

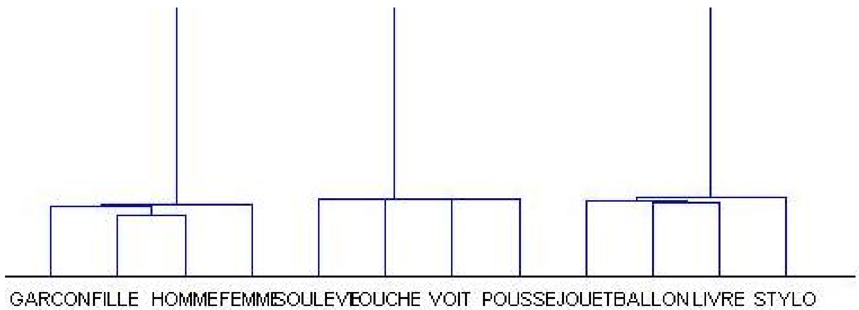

Fig. 4. Right cluster

In general, brain trauma in bilinguals does not result in the loss of one language or even extensive language-mixing. The organization of these internal representations is generally highly resistant to damage. Once it had learned Alpha and Beta, it was very hard to disrupt the organization of the clusters it had developed. After learning, nodes were removed from the hidden-layer and a cluster analysis was performed on the activation patterns of the remaining nodes. Even after some nodes were removed, the organization of the representational clusters remained essentially unchanged.

Simulation Environment

We used Matlab to perform the above simulations. The neural network Toolbox was used to create the Elman network. After training and simulation a dendrogram of the hidden point activation levels were drawn to come up with the above figures. While clustering an agglomerative hierarchical cluster analysis using a Euclidean distance metric and Ward's method to determine linkage was used.

Analysis

The results suggest that a bilingual representation can emerge within a single distributed lexicon from a simple recurrent network. The SRN model can explain some of the effects that have been observed in bilingual memory. 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. Even when we remove some nodes the internal representation does not change significantly and this is what is really observed in cases of brain trauma in bilinguals which does not result in the loss of one language or even extensive language-mixing.

Literature Survey

Initial Models

Initially, three major experimental paradigms dominated bilingual memory research–namely, word association and naming, recognition and recall, and language transfer and interference. The experiments based on these methods, however, produced highly contradictory results concerning knowledge organization in bilinguals.
It was Kolers (1963), who tried to explain bilingual memory from an information processing approach to human cognition. This approach is considered influential in changing the direction of research in this field. Based on Kolers' initial ideas about a bilingual's memory system, two competing hypotheses were advanced and two bilingual memory models were proposed:

- The **common storage model** under the interdependence hypothesis
- The **separate storage model** under the independent hypothesis

Controversial findings have been reported to support either of these two hypotheses throughout the debate on bilingual memory. If the independence hypothesis is correct, balanced bilinguals should react as monolinguals in both of their languages. For example, bilinguals either respond differently in their two languages or fail to transfer from one language to the other. Evidence for supporting the separate storage model has been found in studies using association tasks, free recall and language tagging.

On the other hand, if the interdependence hypothesis reflects a common memory store, it is supported by evidence indicating that intralingual behavior does not differ from interlingual behavior in bilinguals. Studies involving recall, recognition, pair associative learning, processing time seemed to support the interdependence hypothesis.

The **Dual coding theory (DCT)** was proposed by Paivio in 1971. According to the dual coding theory, cognitive activities in human memory are mediated by two symbolic systems; one specialized for processing verbal information and the other for nonverbal (imagery) information. These two representational systems are presumed to be interconnected, but capable of functioning independently. Interconnectedness means that representation in one system can activate the other systems. For example, pictures can be named and images can occur for words. Independence implies, among other things, that nonverbal and verbal memory codes, aroused directly by pictures and words or indirectly by imagery and verbal encoding tasks, should have additive effects on recall tasks. Experiments have also clearly showed that stimuli presented in pictures are recalled more than stimuli presented verbally. Dual coding theory entails assumptions of three levels of processing. First, representational processing involves the independent activation by a stimulus. Thus, words activate verbal representation, whereas objects or their pictures activate imagery representation. Second, referential processing refers to activation of representations in one system by the other through their interconnections. For example, naming an object and generating a mental image for a word are all referential responding. At a third level, associative processing refers to connections between linguistic units and between images in each system.

In 1980 Paivio & Desrochers presented their **bilingual dual coding theory (BDCT)**. It was an extension of the earlier dual coding theory and stated that bilinguals have two verbal representational systems, one for each language, in addition to a representation in the imagery system. These three systems are functionally independent and autonomous from each other but are interconnected at the referential level. In the bilingual dual coding model, the nonverbal imagery system is assumed to be functionally independent of both verbal systems. The assumption implies that bilinguals can perceive, remember, and think about nonverbal objects and events without the intervention of either language system and, conversely, that they can behave or think verbally without constant input from the nonverbal sys tem. On the other hand, the systems are at the same time functionally interconnected at the referential level, so that verbal activity in either language system can be influenced by the imagery system and vice versa. The authors of this theory claimed that their model solved the independence-interdependence controversy of bilingual memory as it encompassed flavors of each.

Experimental verification of BDCT

To validate BDCT, Paivio & Lambert carried out a series of experiments on a group of balanced French-English bilinguals. Balanced bilinguals are those who are considered equally fluent in both the languages they know. The subjects were shown pictures, French words and English words. After this some were asked to write down the name of the picture, some were asked to translate the shown word and then write it down while others to simply write down the word shown. The subjects were then unexpectedly asked to recall the list of words they had written. The results showed that subjects recalled 47% of the items they named, 31% of the items they translated, and 18% of the items they just copied. The findings of this study supported the independent aspect of the bilingual dual coding model, since the results demonstrate that verbal-nonverbal coding or bilingual coding has an additive effect on recall. To explain the higher recall for picture-named items than translated items, Paivio & Lambert relied on a depth of processing approach. According to Paivio and Lambert, verbal- nonverbal processing requires a deeper and a more elaborate level of processing than does bilingual coding alone so that subjects remembered more of the picture- named items than translated items.

To see whether BDCT could be extended to unbalanced bilinguals Sung & Padilla (1991) replicated the earlier Paivio & Lambert study with 18 unbalanced Korean-English bilinguals who were dominant in Korean. The subjects recalled 43% of the items in the Picture-Naming condition, 44% of the translated and 22.5% of the copied items. The approximate ratio of recalled items in the three coding conditions was 2: 2: 1. In Paivio & Lambert, in contrast, the ratio was 3: 2: 1.

The findings represent that the difference in the two studies could be due to the subject's being balanced or unbalanced in two languages or due to the linguistic relationships and orthographic differences in the two languages. Both English and French use almost the same alphabets while Korean has a very different orthography. In order to better understand memory representation in bilinguals, a more elaborated research study was designed and conducted by Sung and Padilla. Two different bilingual groups were compared: Korean-English and Spanish-English bilinguals. Spanish and English is more orthographically similar than Korean and English. Moreover the two groups consisted of bilinguals of three types of fluency – some where balanced, some dominant in one language and the others dominant in the other language. The same type of experiment was conducted and the results analyzed. The results showed that subjects in the present study recalled the items in the Picture-Naming and Translation conditions significantly more than those in the Copying condition. This pattern of recall was consistently found regardless of the subject's proficiency level in the two languages. This was also true regardless of the language combinations of bilinguals (Korean-English or Spanish- English bilinguals). Thus linguistic similarity or dissimilarity did not appear to alter the pattern of recall. The finding supported the bilingual dual coding hypothesis. To explain why unbalanced bilinguals performed better in the Translation tasks it was hypothesized that while translating they needed deeper processing power than balanced bilinguals and this led to better recall.

Some research methods used in bilingual memory research

One of the principal tools that researchers have used to investigate bilingualism is the lexical decision task, usually for visually presented words. Two types of methods are often used.

The first is priming, whereby researchers examine whether word recognition in one language affects later recognition in the other language. Priming experiments show that short term semantic priming occurs between as well as within languages. However long term lexical priming between the first and second presentations of a word is only found for repetitions within a language and not between translation equivalents in different languages.
The second sort of evidence relies on the fact that for many pairs of languages, there are word forms that exist in both languages. Here researchers examine whether such words (*homographs*) behave differently from matched words existing in only one of the languages (*Singles*). *Non-cognate homographs* are words that have the same form but a different meaning in each language (e.g. MAIN and FIN in English mean ‘hand’ and ‘end’ in French). Since they have a different meaning, these words often have a different frequency of occurrence in each language. Results have shown that the same word form is recognized quickly in the language context where it is high frequency, and slowly in the language context where it is low frequency. The fact that these words show the same frequency response as Singles suggests that their behavior is unaffected by the presence of the same word form in the other language, and in turn, that the lexical representations are therefore independent. In support of this view, presentation of a non-cognate homograph in one language context does not facilitate later recognition of the word form in the other language context.
On the basis of the above findings, researchers have tended to conclude that the bilingual has independent representations for a word and its translation equivalent at the lexical level, but a common representation at the semantic level. But as in most cognition hypotheses there are caveats. While the general picture is that lexical representations are independent, nevertheless under some circumstances, between language similarity effects are found. That is, words in one language show a differential behavior because of their status in the other language. For e.g., experiments have been found that non-cognate homographs are recognized more slowly than matched *cognate homographs* (words which have the same form and meaning in each language, such as TRAIN in English and French). It was found that cognate homographs in a bilingual’s weaker language were recognized more quickly than Singles of matched frequency, as if the stronger language were helping the weaker language on words they had in common.

Hierarchical Models

This conceptual/lexical separation is the basis for the broad class of three-node ‘hierarchical models’, consisting of the word-association, concept-mediation, mixed and revised-hierarchical models(See Figure below). All these models share a common architecture consisting of two separate lexical stores (one for each language) and one common conceptual store. The type of hierarchical model is determined by the location and weighting of the links between the L1 (first language) and L2 (second language) lexical nodes and the Conceptual node. In addition, data from bilingual patients with brain lesions have shed considerable light on bilingual memory organization. For example, certain bilingual aphasics also show translation disorders that would seem to support the revised hierarchical model. Various disorders can be described in terms of breakdowns of various links in this model.

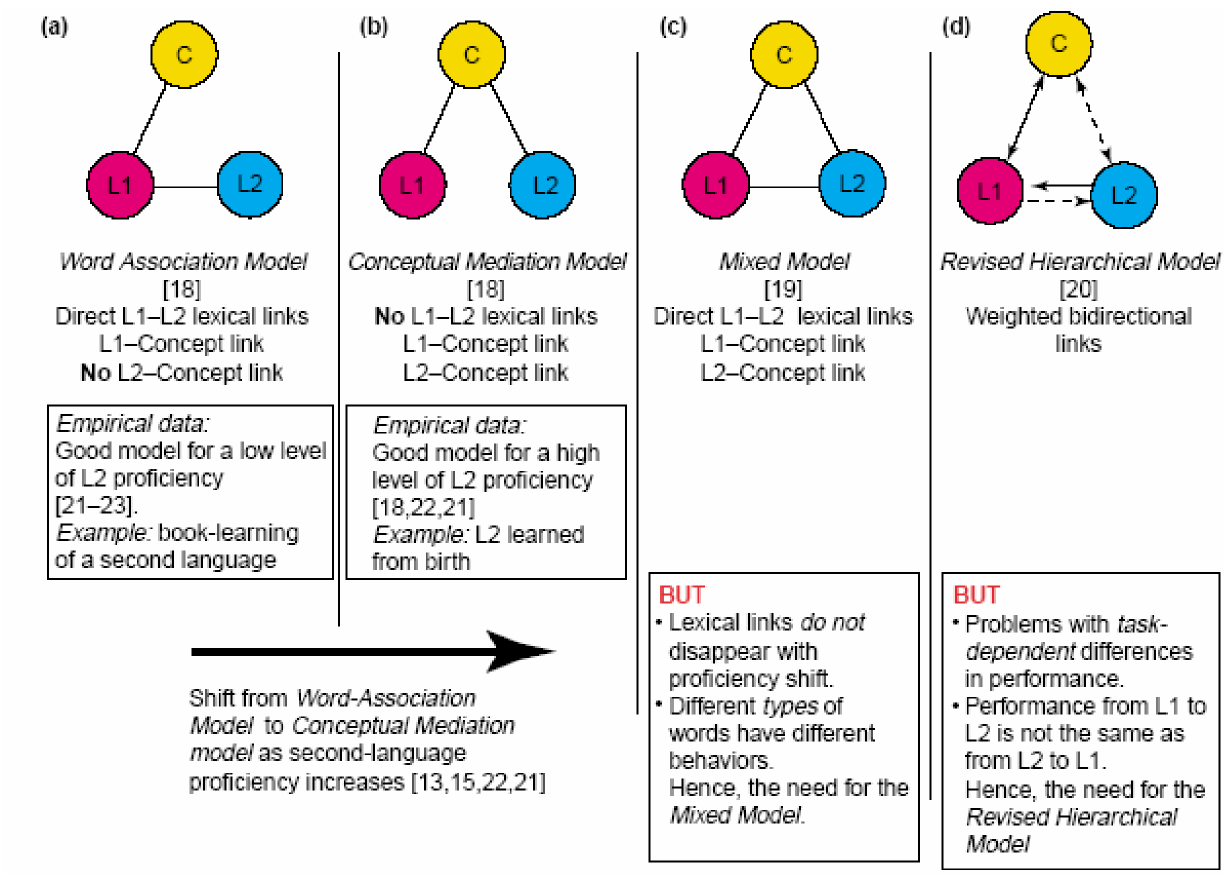

Image Source: "TRENDS in Cognitive Sciences"

Fig. 5 Hierarchical models. There are four types of hierarchical models (a–d), all of which have a single, shared semantic or concept level (C), thereby allowing cross language semantic priming, and two separate word-form lexicons for the two languages (L1 and L2), justified by the absence of cross-language repetition priming. The models each differ in the number, type and location of the links between these three nodes.

Critical Analysis of Hierarchical Model:

Hierarchical models have been criticized by several authors because the memory structure for an individual bilingual seemed to vary depending on numerous factors, including the concreteness or abstractness of a given word, its part of speech and, especially, whether its translation was a cognate or not (for a review, see J.G. Van Hell, PhD thesis, University of Amsterdam, 1998). In hierarchical models, although the two lexicons can interact to varying degrees, they are nonetheless separate. This proposed separation was due to the lack of long-term repetition priming: having seen the word "chien" in a list of French words does not produce faster subsequent word recognition (or related behaviors) of its orthographically dissimilar translation equivalent 'dog' in a list of English words. This argument for a separate-lexicon structure for bilingual memory is, nonetheless, open to a 'level-of-observation' problem. If we consider the phenomenon at the perceptual level, repetition priming involves similar perceptual components. When we fail to observe repetition priming between orthographically dissimilar synonyms, we do not conclude that each word is part of a different lexicon, so why should we arrive at that conclusion when we fail to observe equivalent priming effects between orthographically dissimilar translation-equivalents?

The two other powerful arguments for the separate lexicon view of bilingual memory organization come from:

1. *Release from proactive interference*. A release from proactive interference is observed by changing the language between two lists to be memorized.
2. *Language recall*. Language specific recall of previously presented words is performed well by bilinguals.

However, these arguments for a separate-lexicon structure for bilingual memory are once again; open to a 'level of observation' problem. Consider the release e from proactive interference. This is a well-known and widely investigated effect in monolingual studies, achieved by changing the semantic category of the two lists to be learned. Yet no one concludes from this that there are two 'lexicons', one for each category. Why should an identical result in the bilingual case cause us to propose two separate language lexicons? Furthermore, the role of context in recall performances is also well-established. Marian and Neisser have recently shown that language acts as a context cue in memory retrieval. Good language-recall performance might, therefore, be a product, not of separate language storage, but of the contextual effect of the specific language on recall. We discuss this in more detail below in the discussion of the role of the task. Support for separate lexical stores and separate language processing has also been weakened by numerous overlapping empirical studies on language at the neuroanatomical level.

Bilingual Interactive Activation (BIA) model:

BIA is an extension to McClelland and Rumelhart's well-known proto-connectionist Interactive Activation model. An integrated lexicon is the basic assumption of this model and it has been very successful in extending single-language effects to bilinguals.

When a string of letters is presented to the BIA model, this visual input affects particular features at each letter position, which subsequently excite letters that contain these features and at the same time inhibit letters for which the features are absent. The activated letters next excite words in both languages for which the activated letter occurs at the position in question, while all other words are inhibited. At the word level, all words inhibit each other, irrespective of the language to which they belong. Activated word nodes from the same language send activation on to the corresponding language node, while activated language nodes send inhibitory feedback to all word nodes in the other language. The main function of the language nodes is to collect activation from words in the language they represent and inhibit active words of the other language. The activation of the language nodes reflects the amount of activity in each lexicon [10].

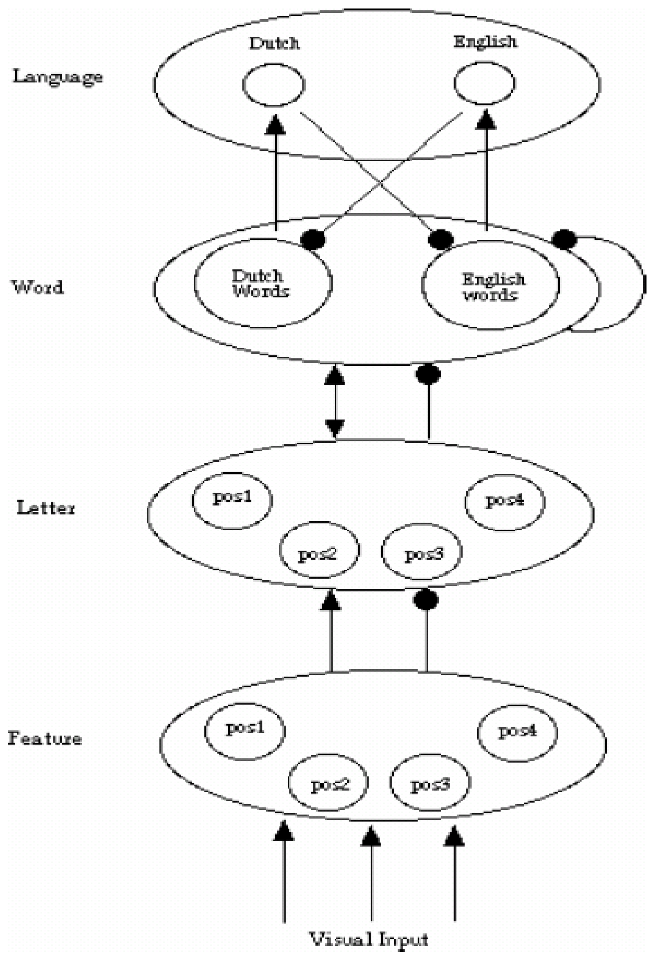

Fig. 6. BIA

An extended version of BIA known as BIA+ has been proposed which speaks of:

· An automatic ('bottom- up') process within the bilingual lexico-semantic system, essentially driven by stimulus input involving modification of the level of activation in the bilingual lexico-semantic system

· An intentional ('top-down') process that alters how the individual responds to signals coming from the bilingual lexico-semantic system, but does not modify activation levels within the system itself .

The main problem with the BIA model is that though it speaks of language nodes it does not speak how they came to form in the first place. Moreover though it speaks of an integrated lexicon, the division into two language nodes somehow blurs this approach. The other problem is that even though researchers in general have agreed upon the presence of a separate semantic or conceptual level in bilingual memory structure there is no such concept in BIA.

Conclusion

As shown by the simulations on BSRN it seems that Elman network might be able to serve as a useful model for bilingual memory. The model is quite successful to explain the usual stability of bilingual memory even after brain trauma. It also can show the fact that there may exist some situations as in cases having severe brain trauma which involve complete removal of one language memory (though we have not simulated these particular aspect of BSRN). The recurrent network used in this model is very simple, more studies and research in this path (SRN) may yield better networks explaining all the facts about bilinguals.

Bibliography

[1] Robert M. French. *A Simple Recurrent Network Model of Bilingual Memory*

[2] Jacquet, M. & French, R. M. (2002). *The BIA++: Extending the BIA+ to a dynamical distributed connectionist framework. Bilingualism, 5(3), 202-205.*

[3] Michael S.C. & Thomas Kim Plunkett. *Representing the bilingual's two lexicons.*

[4] Hyekyung Sung & Amado M. Padilla. *Bilingual Memory*: *Representation in Korean-English and Spanish-English Bilinguals*

[5] Robert M. French and Maud Jacquet. TRENDS in Cognitive Sciences Vol.8 No.2, February 2004. *Understanding bilingual memory: models and data*

[6] David W. Green. *Regulating the bilingual lexico-semantic system*

[7] Jing Liang. *Bilingual Language Learning in Connectionist Networks*

[8] Michael S. C. Thomas. *Bilingualism and the Single route / Dual route debate*

[9] Michael S. C. Thomas. *Distributed representations and the bilingual lexicon: One store or two?*
[10] Walter J. B. van Heuven and Ton Dijkstra, *Orthographic Neighborhood Effects in Bilingual Word Recognition*
[11] *http://www.semioticon.com/virtuals/talks/tomioka.htm*

Thank you