DISAMBIGUATION DURING SILENT READING: A SIMULATION STUDY OF THE INFLUENCE OF INTER AND INTRA TOPOLOGICAL CONNECTIVITY IN HEMISPHERES WITH LEARNING ON THE CORPUS COLLOSUM

Zohar Eviatar[1], Hananel Hazan[2], Larry Manevitz[2], Orna Peleg[1], Rom Timor[2]
[1]*Institute of Information Processing and Decision Making University of Haifa, Haifa, Israel*
[2]*Department of Computer Science University of Haifa, Haifa, Israel*
[hhazan01, manevitz]@cs.haifa.ac.il, [zohare, opeleg]@research.haifa.ac.il, romtimor@013.net

Keywords: Simulation, Neural network, Corpus collosum.

Abstract: This work shows some computer simulations of the interactions of two hemispheres which are distinguished in their intra-regional (orthography, phonology and semantical) connections and their inter-hemispherical connections (i.e. the corpus collosum). The intra-hemispheric connectivity follows the theory of Eviatar and Peleg; while the inter-hemispheric connectivity is chosen to have region to region connections only (i.e. orthography to orthography, phonology to phonology, and semantics to semantics) with different directional biases, and different biases on the potential weights on each connection during learning. This results of these simulations compare well with psychophysical experiments performed by Eviatar and Peleg; thereby helping to confirm their theory. In addition, simulating late "clues" resolving ambiguity revealed a very nice symmetric picture showing that for homophones the RH is slower and thus allows the entire system to resolve ambiguities correctly that the LH alone cannot; while for heterophones, the LH is slower and helps the entire system resolve the ambiguities in this case.

1 INTRODUCTION

Neuropsychological studies have shown that both cerebral hemispheres process written words, but they do it in somewhat different ways (e.g., Iacoboni & Zaidel, 1996, Grindrod & Baum, 2003).

Previous simulation has examined the activation of meanings of homophone ambiguous words with polarized meanings (where one meaning is much more frequent (dominant) in the language) and has shown that transfer of information from a 'right hemisphere (RH)' network to a 'left hemisphere (LH)' network, when context biasing to the nondominant meaning is presented after the initial presentation of the word, is the most efficient mechanism for "recovery" from erroneous activation of the dominant meaning. That is, there are systematic cases where the LH purported architecture could not recover by itself; nor could the RH perform at high levels of performance (Peleg et al., 2007). Other simulation (Weems & Regggia, 2004) suggests that different connections can produce different results.

This paper examines different possible connections between networks representing the two hemispheres and how these differences affect the results of processing homographs. (Monaghan & Pollmann, 2003) show that when stimuli have to be matched in a complex task (such as whether two letters have the same name), performance is better when stimuli are presented across the hemispheres of the brain. Furthermore, they argue that for simpler tasks (such as whether two letters have the same shape), better performance is achieved when stimuli are presented unilaterally. They show that this bilateral distribution advantage effect emerged spontaneously in a neural network model learning to solve simple and complex tasks with separate input layers and separate, but interconnected, resources in a hidden layer. They show that relating computational models to behavioral and imaging data helps to understand hemispheric processing and generating testable hypotheses.

This paper presents the computational advantage of having two networks that can exchange information: LH fully connected (Orthography, Phonology and Semantics) and RH lack of connection between Orthography and Phonology.

2 BACKGROUND

Behavioral studies have shown that the LH is more influenced by the phonological aspect of written words whereas lexical processing in the RH is more sensitive to visual form. A large amount of psycholinguistic literature indicates that readers utilize both frequency and context to resolve lexical ambiguity (e.g., Titone, 1998 , Peleg et al., 2004). Although hemispheric specialization for LH in language processing is assumed, it is also assumed that the RH plays a significant role in language function, especially when ambiguous words are presented in context (e.g. Burgess & Simpson, 1988).

Behavioral studies examining the disambiguation of homophones (e.g., “bank”) suggest that all meanings of an ambiguous word are initially activated in both hemispheres, but at different speeds. While the LH quickly activates both meanings and then selects one alternative (the contextually compatible meaning when prior contextual information is biased, or the salient, more frequent meaning when embedded in non-constraining contexts), the RH activates the nondominant meaning more slowly, and maintains both alternate meanings (including less salient, subordinate and contextually inappropriate meanings).

Previous studies also suggests that exchange of information between the LH and the RH networks will produce better performance and can help the LH recover the subordinate meaning, when it is appropriate to the context (This task the LH could not perform in isolation.)

2.1 Research Goals

The main goal is to investigate how different types of information (phonological, lexical and contextual) are utilized during silent reading in the two connected networks simulating the left and right hemispheres. Specifically the results are crucial for answers regarding inter-hemispheric relation during the disambiguation process of homophones.

We achieved this goal by building a neural network that can process word information (phonological, lexical and contextual) and resolve the meaning of ambiguous words in Hebrew. The network is based on (Peleg et al., 2007) LH and RH networks architecture while adding connections between regions (Orthography, Phonology and Semantics) in various ways thereby simulating the corpus collosum.

The connected networks after training, demonstrate the effects of context and frequency on the resolution of homophones. The computer model consists of "weakly coupled" neural networks that can deal with ambiguity of a written or a spoken word in Hebrew. The main idea is to investigate some questions regarding the "weakly coupled" connection properties such as when, how and where information is transferred and determines the degree of transferred information while shedding light on the division of labor between hemispheres. The "weakly coupled" networks should support the same properties when disconnected and additional or improved properties when connected.

Investigating the "weakly coupled" properties included searching the connection direction in the different tasks (homophones and hetrophones). This is performed by experimenting with fixed weights on the corpus collosum. After determine the connection direction property we experimented with various learning procedure (bounded and free) to check if the assumptions about the connection properties can be learned.

Furthermore, we measure the time it take for the connected networks to resolve the meaning and the paths the networks use to do so. Then we compare the results to existing psycholinguistics theories of how humans process the language. One of the reasons to build computational models is the ability to change parameters, aspects and connection properties of the models in ways that are not possible with human subjects. This provides us with an insight into the mechanisms of reading and understanding the meaning of words.

3 PREVIOUS WORK

Lexical ambiguity resolution is appropriate for the connectionist approach since models of cognitive process are computational, hence producing a response to a stimulus. Results of simulations in a computational model (such as errors and reaction time) can be compared to results of behavioral studies. By doing computational simulations we can replicate human experiments findings and analyze the data to answer the question why do these findings occur. Computational simulation has the advantage of modifying parameters in the simulation while repeating tests that cannot easily be performed on humans. To achieve this goal this work aims to seek the simplest model to emphasis the tested phenomenon. This paper model is based on or inspired from the following previous models.

3.1 Harm's & Seidenberg's network

This model (Harm & Seidenberg, 2004) is designed from the behavioural studies on how semantics is computed from the orthography and phonology where the phonology is mediating between the orthography and the semantics together with a direct link between the orthography and semantics. The model is focused on the homophones and not focused on the differences between homophones and hetrophones.

3.2 Kawamoto's network

Kawamoto (Kawamoto, 1988) designed his neural network model in such a way that the entire word, including its orthographic, phonological and semantic features occurs as an "attractors" in the recurrent network (Hopfield, 1982).

The connectionist network represents the spelling, pronunciation and meaning of a word. There are 216 units in this network all connected to each other.

According to his model, the more frequent a certain meaning of the word in a certain context is, the stronger attractor it will be, and the completion of other features (semantic and phonological) would usually fall into this attractor.

Another factor examined was the time lapse between accessing the dominant meaning and the time lapse of accessing the secondary (subordinate) meaning (Kawamoto, 1988; Kawamoto, 1993).

3.3 Hazan's network

Hazan (Peleg et al., 2007) designed a two-hemisphere model based on Kawamoto's model. The model includes two separate networks. One network incorporates Kawamoto's version, and successfully simulates the time course of lexical disambiguation in the LH. In the other network based on the behavior of the disconnected RH of split brain patients (Zaidel & Peters, 1981), a change was made in Kawamoto's architecture, removing the direct connections between orthographic and phonological units. (Peleg et al., 2007).

Hazan (Peleg et al 2007) focused on the different architecture between hemispheres while testing the effect of data transfer (mainly in homophones) between the hemispheres by copying the data at a specific point in time (during the network convergence). This procedure is external to the modelling and is corrected in this paper.

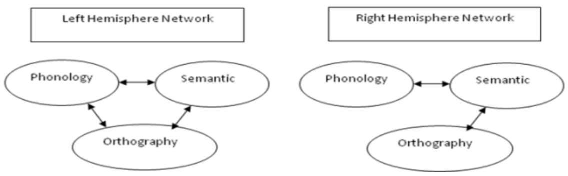


Figure 1: Hazan's network architecture

3.4 Weems & Reggia network

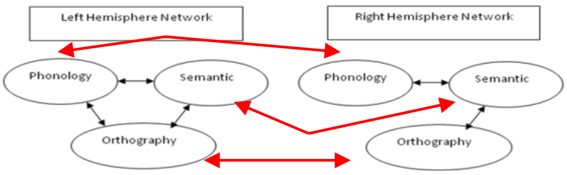
Weems & Reggia tested hemispheric specialization and independence for word recognition while comparing three computational models: Callosal Relay (strong right to left, minimal left to right connectivity, output from LH), Direct Access (minimal connectivity between hemispheres, separate outputs) and Cooperative (strong connection, single output) and showed advantage for the Cooperative model together with a slight performance dropdown (Weems & Reggia, 2004).

4 COMPUTATIONAL MODEL

The simulation is based on (Peleg et al., 2007) LH and RH network which includes the implementation as described by (Kawamoto, 1993) with some changes in the encoding. The simulation includes the "Corpus Callosum" (CC) that was implemented as a connection from LH units to RH units in a various ways:

1. "One to One" - Each neuron from LH/RH is connected to the corresponding neuron in the other hemisphere.

2. "One to Many" - Each neuron from LH/RH is connected to a group of neurons in the corresponding area of the other hemisphere.
3. Within regions and between regions -



Orthography, Phonology and Semantics.

Figure 2: Illustration of network that include CC connections between Corresponding regions of LH & RH

This work focuses on the different corpus collosum connections in order to realize in a more natural manner the assumption verified in Hazan's work (Peleg et al., 2007) that transfer of data between hemispheres is beneficial to this lexical decision task.

4.1 The learning stage

The network was trained with a simple error correction algorithm (Kawamoto, 1988) taking into consideration a learning constant and the magnitude of the error determining a bipolar activity of a single unit. This activity is determined by the input from the environment, the units connected to it (within the hemisphere and from the CC) and a decay in its current level of activity. The learning process was achieved by altering the weights between the units of the network to minimize the error between the activation level and the network input.

$$\Delta W_{ij} = \eta(t_i - i_i) * t_j \text{ , where } i_i = \sum_j W_{ij} t_j$$

n – Learning constant.
ti , tj – target activation levels of unit i and j.
ii – net value of unit i.

In a learning trial an entry was selected randomly from the lexicon. Dominant and subordinate meanings were selected with a ratio of 5 to 3 roughly based on linguistic considerations. The learning phase was divided to the following steps:

A. Initialization of units with random values.
B. Random order of sets of words.
C. The network was trained with 48 words.
D. The network was tested if more training is needed. If so another 48 words were chosen to continue the training. The testing had to fulfill these conditions:

- Presenting the orthographic part of word leads the network to select the dominant meaning.
- Presenting the orthographic part of word with a clue to the subordinate meaning leads the network to select the subordinate meaning.

The learning was stopped when the conditions were fulfilled for each group of words (homophones, heterophones and normal words) separately or when the training set ended.

4.2 Testing the model

After the networks were trained they were tested with the same lexicon used for training by presenting just the orthographic part of the entry as the input (to simulate neutral context) or by presenting part of the semantic (subordinate meaning) sub-vector after presenting the orthography (to simulate contextual bias). The clues were given at the point in time when the LH reached 80-100 percent convergence to the dominant meaning. In each simulation the input sets the initial activation of the units. Unit i was influenced from the following sources:

A. External stimuli (orthographic part of word or clues).
B. Previous values from the last iteration multiplied by the decay rate.
C. Sum of the intra connected units output.
D. Sum of the inter connected units output (Simulates the CC).

The activity of unit a at time t+1 is:

$$a(t+1) =$$
$$LIMIT\left[\delta a(t) + \left[\sum_j Wij(t)aj(t)\right] + Si(t)\right]$$

Where:

$$LIMIT\,[X] = \begin{cases} 1 & X>1 \\ -1 & X<-1 \\ X & \text{otherwise} \end{cases}$$
$$\delta = the\ decay\ variable$$

The decay variable was set dynamically starting from 0.6, increasing while the network progresses and ending at the value of 1 when the run is completed. This was done to strengthen the unit activity and weaken the environment as the network progresses and convergences to its destination. The name of the variable can be misleading since it indicates the decay of the current value of the neuron but its value increases during the network process. It should be noted that there are some suggested implementations to the decay variable that should perform the effect but in a more natural way. One suggestion is to incorporate the magnitude of activation changes in the network architecture and increase the value of a(t) while the activation changes are sufficiently small (indicating that the network is converging). This research did not pursue this direction.

In order to assess lexical access, the number of iterations through the network for all the units in the spelling, pronunciation or meaning fields of one of the LH or RH networks to become saturated, was measured. A response was considered an error if the pattern of activity did not correspond with the input but corresponded to a different meaning; non-convergent if all the units did not saturate after 50 iterations. (This means that after 50 iterations the units did not represent a word from the lexicon.)

Testing was done after training the connected LH and RH or after setting fixed weights on the CC. In the latter case in different experiments the weights were fixed uniformly at values that varied between 0.05 to 0.50 or one value was chosen for the weights from LH to RH and a different one for RH to LH.

4.3 Results analysis

In each simulation, 12 architecturally identical networks were used to perform 12 runs in order to simulate 12 subjects in an experiment by varying their initial weights sets randomly and by training them randomly. (12 networks were chosen to correspond to the protocol in a corresponding human experiment (Peleg et al., 2008)) During the testing phase the network received various inputs. First the orthography of a work and then the inputs included clues from the word meaning. The level was set to +0.25 if the corresponding input feature was positive, -0.25 if it was negative and 0 otherwise.

Result of each trail was recorded including the number of iterations needed for coverage and sum number of errors. The data was separated for Group of words (homophones , hetrophones and normal words), Type of clues (to subordinate or to dominate), Number of clues and CC weights or weight limitation. Mean and standard deviations were calculated.

In this paper the focus is on the different type of connections in the different ambiguity task (hetrophonic vs. homophonic). Therefore all results presented in the following chapters relates to giving the orthographic part of word with clues to the subordinate meaning. Performing the opposite tests (subordinate to dominant) produced the same results as presented in (Peleg et al., 2007) and did not contributed to the research regarding the different types of connections in the different ambiguity linguistic tasks.

5 SIMULATION RESULTS

Hazan's previous results (Peleg et al., 2007; Peleg et al., 2010) indicated that without transfer of data between LH and RH the LH cannot recover to the subordinate meaning after receiving semantic clues and thus selects the dominant meaning. The RH was able to perform this recovery and select the subordinate meaning. This phenomenon was called the "Change of heart".

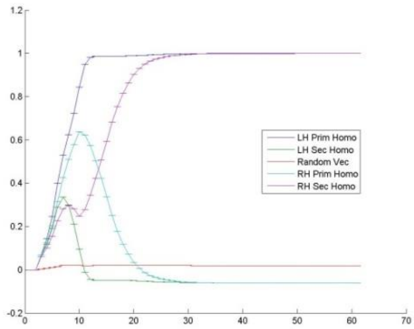


Figure 3: Network performance without CC. Only RH can perform the "Change of heart"

Hazan (Peleg et al., 2007) also showed that if we copy the RH data to the LH during the time course of convergence we can help the LH recover to the subordinate meaning. This is also e beneficial for the LH convergence time.

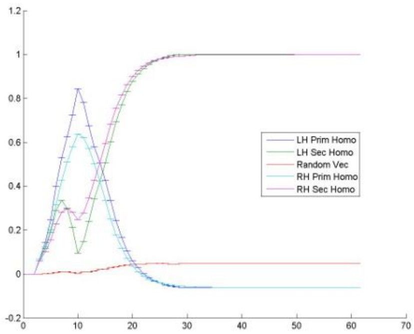


Figure 4: Network performance when copy data from RH to LH is performed. Both LH and RH can perform the "change of heart"

Accordingly, there is a reason to believe that connecting the LH and RH with CC can yield better results. The next results are aimed to find the specific properties of this connection.

5.1 Homophones

Table 1 shows the results of average convergence time (in iterations until the vector obtains 100% accuracy including standard deviation in parentheses) for LH and RH when presenting a homophonic word without clues, the sum of errors and non-convergences when presenting a word with clues to the subordinate meaning.

Table 1: Network performance in resolving homophone (various architectures). * Errors and Non-conv are out of 96 in each hemisphere.

#	Network architecture	LH	RH	Errors* LH/RH		Non-conv* LH/RH	
1	Without CC	40.32 (3.42)	41.54 (4.19)	29	0	37	0
2	With CC: Weights fixed at 0.25 (RH to LH)	39.18 (3.24)	41.06 (2.93)	11	0	14	0
3	With CC: Weights fixed at 0.25 (LH phonology to RH phonology, RH semantics to LH semantics).	39.77 (4.21)	40.69 (4.48)	23	12	19	7
4	With CC: Weights fixed at 0.25 RH to LH and 0.10 LH to RH.	39.84 (5.33)	40.14 (4.77)	21	9	17	11
5	With CC: Weights fixed at 0.25 (All, Both ways)	40.36 (6.03)	40.79 (5.64)	31	24	33	19

Architecture 1 is used as a baseline for network performance compared to previous studies (Peleg et al., 2007; Peleg et al., 2010) and for the subsequent architectures.

Architecture 2 and 4 are designed to focus on the assumption presented in the previous studies (Peleg et al., 2008) indicating that in homophones the RH can help the LH recover and perform the change of heart. According to this architecture 2 and 4 are focused on the transfer of data from RH to LH. In addition Architecture 4 includes a weaker transfer of data from LH to RH to examine the effect of such direction of information flow. The test in architecture 4 uses a weaker CC from LH to RH since a balanced CC was tested in architecture 5 and proven wrong.

Architecture 3 and 5 are designed to compare results with a balanced transfer of data. While architecture 5 is a full balanced CC, architecture 3 tests different weights in different areas. In architecture 3 the phonology is set from LH to RH and the semantics is from RH to LH. The phonology from LH to RH was chosen due to the LH inner architecture assuming that it will be beneficial for the LH to help the RH with data from the phonology because it should be more evolved during the network process than it is in the RH. Semantics from RH to LH was chosen because of previous studies [8] indicating that in homophones the RH can help the LH recover and perform the change of heart.

Figure 5 (Without STD) and 9 (With STD) shows the time course of convergence corresponding to case 2 in the table above (STD bars shows that the behavior is similar in all subjects in general) . Trials were performed for various CC weights between 0.1 to 0.3 (with 0.05 intervals), one way or both ways, same regions or between regions. Case 2 in the table above seems to be the optimum for resolving homophone ambiguity.

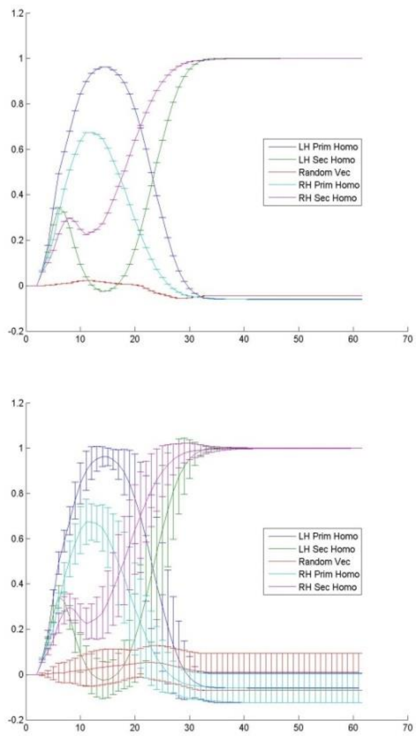


Figure 5: Network performance (with and without STD) when weights on CC are fixed at 0.25 (RH to LH). Both LH and RH can perform the "change of heart".

5.1 Hetrophones

Table 2 shows the results of average convergence time (in iterations until the vector obtains 100% accuracy including standard deviation in parentheses) for LH and RH when presenting a hetrophonic word without clues, the sum of errors and non-convergences when presenting a word with clues to the subordinate meaning.

Table 2: Network performance in resolving hetrophone (various architectures). * Errors and Non-conv are out of 96 in each hemisphere.

#	Network architecture	LH	RH	Errors* LH/RH		Non-conv* LH/RH	
1	Without CC	30.39 (4.88)	28.07 (5.14)	0	11	0	5
2	With CC: Weights fixed at 0.25 (LH to RH)	30.14 (5.11)	27.51 (5.37)	0	7	0	3
3	With CC: Weights fixed at 0.25 (RH phonology to LH phonology, LH semantics to RH semantics).	29.77 (6.52)	29.23 (5.93)	7	13	5	9
4	With CC: Weights fixed at 0.25 LH to RH and 0.10 RH to LH.	29.63 (7.13)	28.36 (7.20)	9	16	4	2
5	With CC: Weights fixed at 0.25 (All, Both ways)	29.32 (9.31)	29.95 (8.67)	19	18	12	15

Architecture 1 is used as a baseline for network performance compared to previous studies (Peleg et al., 2007) and the subsequent architectures.

Architecture 2 and 4 are designed to focus on the assumption that in hetrophones the LH can help the RH recover and perform the change of heart. This is assumed because of results in the previous section of this research that showed that in homophones the RH can help the LH since it has a slower time course. Therefore in hetrophones due to the slower time course of the LH it ought to be able to help the RH. According to this architecture 2 and 4 are focused on the transfer of data from LH to RH. Architecture 4 includes in addition a weaker transfer of data from RH to LH in order to examine the effect of such a direction of information flow. The test in architecture 4 uses a weaker CC from RH to LH since a balanced CC is tested in architecture 5 and proven wrong.

All architectures were chosen to be symmetrical in the opposite way to the selected architecture in homophones and to test specific hypothesis.

In addition architectures 3 and 5 are designed to compare results with a balanced transfer of data. While architecture 5 is a full balanced CC, architecture 3 tests different weights in different areas. In architecture 3 the phonology is set from RH to LH and the semantics is from LH to RH. The phonology from RH to LH was chosen to help the LH decide on the specific phonology since it was assumed that during the network process it is struggling with two different phonologies. Semantics from LH to RH was chosen due assumption that due to the slower time course of the LH it can help the RH in hetrophones to recover and perform the change of heart.

Figure 6 (Without STD) and 11 (With STD) shows the time course of convergence corresponding to case 2 in the table above (STD bars shows that the behavior is similar in all subjects in general). Trails where performed for various CC weights between 0.1 to 0.3 (with 0.05 intervals), one way or both way, same regions or between regions. Case 2 in the table above seems to be the optimum for resolving hetrophone ambiguity.

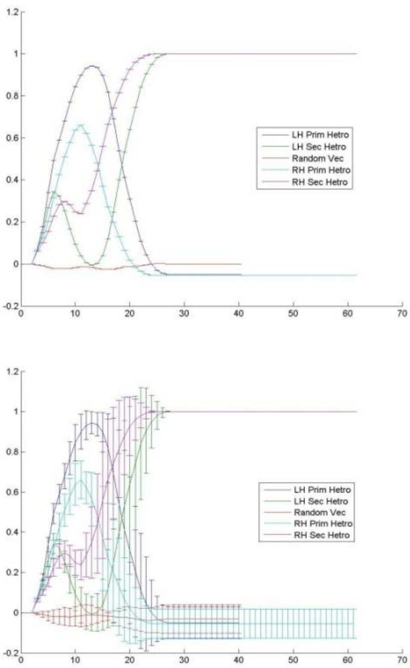


Figure 6: Network performance (with and without STD) when weights on CC are fixed at 0.25 (RH to LH). Both LH and RH can perform the "change of heart".

5.1 Connected Learning

Table 3 shows the results of average convergence time (in iterations until the vector obtains 80% accuracy) for LH and RH when presenting a hetrophonic word without clues, the sum of errors and non-convergences when presenting a word with clues to the subordinate meaning.

Table 3: Network performance in connected learning (various architectures). * Errors and Non-conv are out of 96 in each hemisphere.

#	Network architecture	LH	RH	Errors* LH/RH		Non-conv* LH/RH	
1	Free learning – homophones	16.03 (6.12)	16.66 (8.02)	28	25	44	47
2	Weights bounded to 0.25 (RH to LH) – homophones	35.94 (5.14)	33.17 (4.63)	13	0	16	0
3	Weights bounded to 0.25 (LH to RH) – hetrophones	28.66 (4.97)	25.21 (5.35)	0	3	0	5
4	Weights bounded to 0.25 (All, Both ways) – homophones	35.15 (6.37)	34.93 (7.11)	18	9	15	13
5	Weights bounded to 0.25 (All, Both ways) – hetrophones	27.45 (10.31)	27.02 (9.52)	16	13	11	6

Architecture 1 was selected as a baseline for network performance compared to architecture 1 in the previous sections (homophone and hetrophones results without connected learning) and to the next architectures. The rest of the architectures where selected to compare results with previous sections while selecting the best architectures in the previous sections. Architectures 2 is compared to architecture 2 in the homophones section using bounded weights on the CC in the direction from RH to LH under the assumption that in homophones the RH can help the LH improve network performance. Architecture 3 is compared to architecture 2 in the hetrophones section using bounded weights on the CC in the direction from LH to RH under the assumption that the LH can help the RH improve network performance. Architecture 4 and 5 are to compare results with a balanced transfer of data in the different type of words (homophone or hetrophone) and were chosen to test the immerged structure of the CC during a balanced bounded weight.

Free learning of CC weights caused the LH and RH to lose their special properties. LH became slower while selecting the dominant meaning and the RH lost its ability to perform the "Change of heart" when presented with clues to the subordinate meaning.

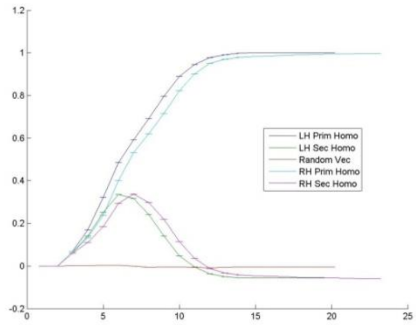


Figure 7: Network performance with CC (Free learning). RH & LH cannot perform the "Change of heart".

Restricted learning under specific restrictions enabled the LH and RH not to lose their special properties. Both RH and LH performed the "Change of heart" but LH recovery was partial.

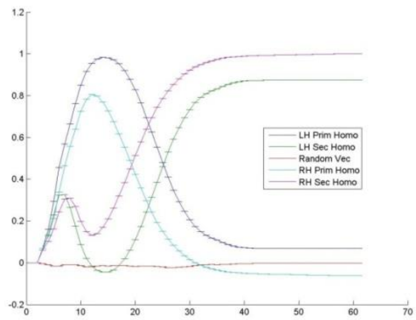


Figure 8: Network performance with CC (CC weights are bounded to 0.25). RH & LH can perform the "Change of heart". Note LH recovery is partial.

5 CONCLUSIONS

5.1 Homophones v Hetrophones

Hazan's previous work (Peleg et al., 2007; Peleg et al., 2010) showed in the case of homophones that running the LH without data transfer from RH has substantially worse performance, both in number of iterations to convergence and in the ability to perform the "Change of heart" when presented with clues to the subordinate meaning at a later point in time during the network processing.

Hazan demonstrated the above by transferring the data between the hemispheres artificially: After some iteration the data from the RH was copied to the LH and was clamped for some iterations. Copying the data returns, in some sense, the LH back in time and allowed it to react to the clues to the subordinate meaning. Transfer of data from RH to LH in homophones yield better performance for the LH even in cases when the RH had failed to perform the recovery.

Hazan results were reconstructed in this work while connecting the LH and RH in a more natural way (See Table 1 - Row 2 and Figure 5). The network demonstrated that in homophones running the LH without data transfer from RH has substantially worse performance in the number of iterations to convergence and the ability to perform the "Change of heart“.

This work also shows that:

1. Data transfer in homophones is more beneficial when done from RH to LH (See Table 1 - Row 2 compared to Table 1 Row 3-5).
2. Data transfer in hetrophones can be more beneficial when done from LH to RH (See Table 2 - Row 2 compared to Table 1 Row 3-5). Note that results are less dramatic than the above homophone case, since there were fewer errors and non-convergence to begin with. This is assumed due to the networks size and the specific lexicon in this research.
3. Connection between the hemispheres via the CC is "weakly coupled" as compared to the inner hemisphere connections (See Table 1 and Table 2 Rows 2-5)[1].

Word processing is different in LH and RH when comparing different tasks such as homophone and hetrophone disambiguate resolution. In homophones the RH has less error and non-convergence cases than LH but in the cost of a slower convergence time. Whereas in hetrophone the LH has less error and non-convergence cases then RH but again in the cost of a slower convergence time[2]. The convergence time drawback in performance is an advantage when trying to perform the "change of heart" from dominate to subordinate meaning. The ability to perform the "change of heart" is more efficient when transferring data between hemispheres. Transferring data from the "slower" network to the "faster" network allows the "faster" network to absorb the clues to the different (subordinate) meaning.

5.1 Connected learning v Separate learning

Results of connected learning also point out some interesting facts. In general connected learning has better performance in convergence time then in separate learning.

Also, it is shown that free learning of the CC weights causes the network to lose the "weakly coupled" proportions and therefore the LH and RH lose their special properties (convergence time and "Change of heart"). Furthermore, learning with bounded weights on the CC produces the desired properties only if the CC bounded weights are less in proportion to the interior hemispheric natural boundary of weights (1 to -1), thus forming a "weakly coupling" between the hemispheric networks.

Results of LH and RH after connected learning are slightly different then in separate learning. In performance variables such as convergence time there is a slight advantage to connected learning but

[1] Weights on the CC must be more than 0.05 in order to make a difference and less than 0.30 to prevent non convergence. Note that, in contrast, inner hemispheric weights vary from -1 to 1, and forms a relative strong intra-hemispheric connection between the hemispheric regions orthography , phonology and semantics.

[2] Note that in hetrophones the different time course of the LH is not so significant than in homophones and therefore the results are not as conclusive as in homophones.

in errors measurements connected learning has shown worse results but in proportion to the results demonstrated in separate learning.

It is mentioned above that the LH and RH have a different time course and that each hemisphere has a different time course in homophones and heterophones. In separate learning it is shown that the difference between homophone and hetrophones in the RH are not significant but are significant in the LH. Also, separate learning shows that the RH has a longer time course both in homophones and in hetrophones. The different time course is maintained in connected learning but it is noted that the significant difference between homophones and hetrophones is more prominent and that in the connected learning the time course of RH is longer only in homophones while in hetrophones the LH has a longer time course.

In connected learning we can see that there is a clear advantage to transfer data from RH to LH in homophones and help the LH recover whereas in hetrophone the transfer of data from LH to RH has a less significant effect. Note that in hetrophones the transfer of data from RH to LH has a negative effect on the LH ability to recover.

It should be noted that although connected learning with balanced bounded weights (architecture 4 and 5 in the connected learning section) resulted in worse performance than the unbalanced bounded CC weights (architecture 2 and 3) testing the network in the different tasks (homophones and hetrophones) while suppressing the negative direction (e.g. LH to RH in homophones) produced results similar to (architecture 2 and 3). It is therefore assumed that the learning should be performed with balanced weights on the CC.

5 HUMAN EXPERIMENTS

Recently, behavioral studies have been performed by Peleg and Eviatar (Peleg et al., 2008; Peleg et al., 2009) designed to test certain intra-hemispheric connectivity assumptions that they put forward. These studies combined divided visual field (DVF) techniques with a semantic priming paradigm.

The behavioral studies were conducted in Hebrew and combined a divided visual field (DVF) technique with a semantic priming paradigm. Subjects were asked to focus on the center of the screen and to silently read sentences that were presented centrally in two stages. First, the sentential context was presented for 1500 ms and then the final ambiguous prime was presented for 150 ms. After the prime disappeared from the screen a target word was presented to the left visual field (LVF) or the right visual field (RVF) for the subject to make a lexical decision. Targets were either related to the dominant or the subordinate meaning or unrelated. Magnitude of priming was calculated by subtracting reaction time (RT) for related targets from RT to unrelated targets. The most interesting results were observed in the subordinate-biasing context condition (“The fisherman sat on the bank”): At 250 SOA both meanings (money and river) were still activated in both hemispheres (Peleg & Eviatar, 2009). However, 750 ms later (1000 SOA), a different pattern of results was seen in the two visual fields.

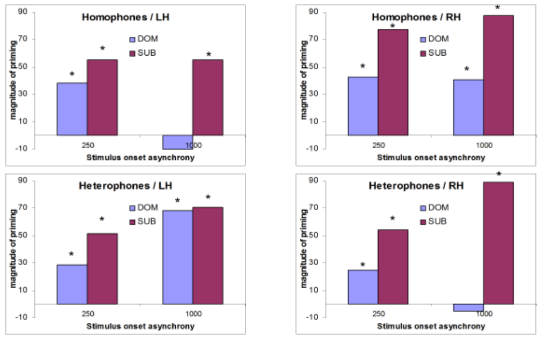


Figure 9: Study: For homophones, LH selects a meaning while both meanings are activated in the RH. For heterophones, it is more difficult for the LH to suppress the dominant contextually inappropriate meaning.

For homophones (e.g., “bank”), previous results (Peleg et al., 2007) were replicated: the LH selected the contextually appropriate meaning, whereas both meanings were still activated in the RH (See Figure 9). For hetrophones (e.g., "tear") we get an opposite pattern: the LH is unable to suppress the dominant meaning, while the RH is able to do so (Peleg et al., 2009). This observation fits results of previous work Peleg et al., 2010).

These studies, although limited to reaction time did succeed in implying different patterns of activation of both meanings in the two hemispheres. Our simulations, built to correspond to their intra-hemispheric connectivity assumptions produced results that fit well with those human experiments and thereby further support the theoretical underpinnings of Peleg and Eviatar. Here the interpretation of the similarity of activation to dominant and subordinate meanings at iterations is taken as parallel to maintenance of the corresponding meanings in the hemispheres.

5 SUMMARY

This work implemented a model of both the RH and LH, with architectural differences between the hemispheres as proposed by the theories of Peleg and Eviatar. The hemispheres are linked together in a natural fashion, both during learning and functioning. The results of the simulations show that the connections between the hemispheres allow additional functionality for the LH as observed in humans ("change of heart"); and the hemispheres also perform at comparative speeds that also qualitatively match human DVF experiments.

Results demonstrated in the behavioral studies (Peleg et al., 2008; Peleg et al., 2009) correspond to results reported by the simulations. This work suggests a refinement of these experiments to check as well the connectivity strength between hemispheres. One possible method to do this would be to use Dynamic Causal Modeling (Friston et al., 2003) to test the effective connectivity between hemispheres during fMRI studies.

Further, this work predicts connectivity strength between the two hemispheres in architectural regions; and thus suggests new human experiments.

Resent human experiments performed by Peleg and Eviatar tested the connectivity influence while presenting different stimuli to different visual field (one with clues to the subordinate meaning and one with "noise" to disable the one hemisphere influence) indicates some interesting findings about the interaction of the hemispheres. This experiment is reconstructed in the simulation described in this paper to support the theory and discover some properties about it.

The suggested architectures in this work implies different connection in CC when resolving homophones then when resolving hetrophones.

However, it should be possible to use a single architecture with trained weights in the CC for both cases by either i) changing the signal to noise effect by increasing the size of the network or ii) changing the architecture to include an asymmetrical inhibitory affect that depends on the difference in convergence between RH and LH phonology that indicates the specific task (homophone or hetrophone).

REFERENCES

Burgess, C., & Simpson, G.B. (1988): Cerebral hemispheric mechanisms in the retrieval of ambiguous word meanings. *Brain and Language, 33* (pp. 86-103).

Friston K.J., Harrison L. and Penny W.: Dynamic causal modelling. *Neuroimage 2003 Aug;19(4)* (pp. 1273-302).

Grindrod, C., & Baum, S. (2003). Sensitivity to local sentence context information in lexical ambiguity resolution: Evidence from left-right-hemisphere-damaged individuals, *Brain & Language*, 85, 502-523.

Harm M. W. & M. S. Seidenberg (2004): Computing the meanings of words in reading: cooperative division of labor between visual and phonological process. *Psychological Review, 111*, 662-720.

Hopfield J. J 1982: Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Science, USA, 79*,2554-2558.

Iacoboni M, Zaidel E (1996) Hemispheric independence in word recognition: evidence from unilateral and bilateral presentations. *Brain Lang* 53:121–140 14.

Kawamoto, A. H. (1988). Distributed representations of ambiguous words and their resolution in a connectionist network. In S.L. Small, G.W. Cottrell, & M. K. Tanenhaus (Eds.) *lexical ambiguity resolution: Perspectives from psycholinguistics, and artificial intelligence*, San Mateo, CA: Morgan Kaufmann.

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

Monaghan, P. & Pollmann, S. (2003). Division of labour between the hemispheres for complex but not simple tasks: An implemented connectionist model. *Journal of Experimental Psychology: General*, 132, 379-399.

Peleg, O., Giora, R., & Fein, O. 2004: Contextual strength: The Whens and hows of context effects. In I. Noveck & D. Sperber (Eds.), *Experimental Pragmatics* (pp.172-186)*.* Basingstoke: Pagrave.

Peleg,O., Eviatar, Z., Hazan, H,& Manevitz, L. (2007). Differences and Interactions between Cerebral Hemispheres When Processing Ambiguous Homographs. *Attention in Cognitive Systems, L. Paletta and E. Rome, Eds. Springer-Verlag Berlin Heidelberg 2007- Lecture Notes in Computer Science (LNAI) publication no.4840, 'Attention in Cognitive Systems', L. Paletta and E. Rome, Eds.*, pp. 367–380.

Peleg O. and Eviatar Z. : Hemispheric sensitivities to lexical and contextual information: Evidence from lexical ambiguity resolution. *Brain and Language 105 (2008)* (pp. 71–82).

Peleg O. and Eviatar Z.: Semantic asymmetries are modulated by phonological asymmetries: Evidence from the disambiguation of homophonic versus heterophonic homographs. *Brain and Cognition 70 (2009)* (pp. 154–162).

Peleg,O., Manevitz, L., Hazan, H., Eviatar, Z., (2010). Two Hemispheres - Two Networks A Computational Model Explaining Hemispheric Asymmetries While Reading Ambiguous Words. *Annals of Mathematics* *and Artificial Intelligence (AMAI)* Volume 59, Number 1, 125-147.

Titone, D. A. 1998: Hemispheric differences in context sensitivity during lexical ambiguity resolution. *Brain and Language*, 65, 361-394.

Weems, S. & Reggia, J., Hemispheric specialization and independence for word recognition: Acomparison of three computational models, *Brain and Language* 89 (2004) 554-568.

Zaidel, E., & Peters, A. M. 1981: phonological encoding and ideographic reading by the disconnected right hemisphere: Two case Studies. *Brain & Language, 14*, 205- 234.