

# Simulating Frontotemporal Pathways Involved in Lexical Ambiguity Resolution

Jean-Philippe Thivierge (jpthiv@ego.psych.mcgill.ca)

Debra Titone (dtitone@hebb.psych.mcgill.ca)

Thomas R. Shultz (thomas.shultz@mcgill.ca)

Department of Psychology, 1205 Penfield Avenue  
Montreal, Québec, Canada H3A 1B1

## Abstract

Lexical ambiguity resolution is the process whereby one meaning of an ambiguous lexical item is activated over an alternative meaning. This paper introduces a preliminary model of lexical ambiguity resolution based on known cortical regions that are activated in this process, including areas of the prefrontal and temporal lobes. Through the interaction of these regions, the model is able to simulate findings associated with the time-course of meaning activation in context for ambiguous words whose alternative meanings differ in frequency.

## Introduction

Research in neurocomputing is interested in developing formal models that integrate findings across a number of fields, including cognitive sciences, neurophysiology, linguistics, mathematics, and computer science. A key issue in developing such models is to combine known anatomical structures with mathematical descriptions of the mechanisms postulated to be involved in various cognitive processes. The aim is to extract some meaningful principles, formalize existing theories in a concrete instantiation, and generate novel hypotheses about the interaction between neurophysiological structures, cognitive processes, and experimental manipulations.

The current study reviews neurophysiological evidence relating the prefrontal and temporal areas to the process of lexical ambiguity resolution. A computational model is proposed and employed to capture the influence of meaning frequency on lexical ambiguity resolution in context.

## Psychological Evidence

Lexical ambiguity resolution is the process whereby one meaning of an ambiguous lexical item is activated over an alternative meaning. For instance, the word BANK is a homonym with at least two alternative meanings. In the absence of a contextual bias, the following effects are reported (for a review see Titone, 1998):

- (1) If two lexical entries have equal frequency, both intended ( $i$ ) and unintended ( $u$ ) meanings gain some activation  $a$  in the following way:  $a_i > a_u > 0$ .
- (2) By increasing the frequency  $f_i$  of the intended meaning, activation of the unintended meaning approaches zero.
- (3) By increasing the frequency of the unintended meaning,  $a_i < a_u$  initially, but this relationship is eventually reversed given enough processing time.

- (4) In cases 1-3, activation of unrelated words remains close to zero.

## Neurological Evidence from ERP and fMRI

In the neuroimaging literature, the language system is described as a number of relatively small but tightly clustered and interconnected modules, each with unique contributions to language processing (Bookheimer, 2002).

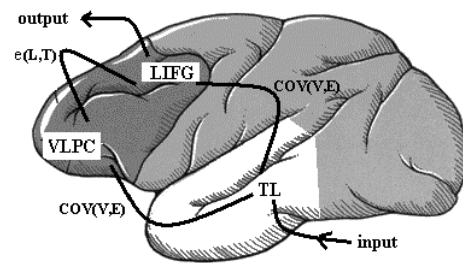


Figure 1: Frontotemporal regions involved in a task of lexical ambiguity resolution (left sagittal view). Two functions are postulated between these regions: covariance maximization ( $cov$ ) and error minimization ( $e$ ).

## Three modules involved in lexical disambiguation

A recent fMRI study has identified the following regions of interest in lexical disambiguation (Rodd, Davis, & Johnsrude, 2005):

- (1) the left anterior and posterior ventro-lateral areas of the prefrontal cortex (VLPC);
- (2) the left inferior frontal gyrus (LIFG);
- (3) the left occipito-temporal and left parieto-temporal areas of the temporal cortex (TL).

This suggests a tri-partite model involving frontotemporal pathways that connect the VLPC, LIFG, and TL regions (see Figure 1). Although these are not the only regions involved in meaning disambiguation, they arguably play a central role. All three regions are modality-independent. As such, they can receive input from either auditory or visual pathways, and thus spoken or written language.

EEG studies also provide evidence for the centrality of these brain regions. One study reports a rhythmic coupling in the theta band between the prefrontal cortex and the temporal lobe (Bastiaansen, van Berkum, & Hagoort, 2002). Other authors report a stream of activation going from the temporal lobe to the prefrontal cortex has been shown in ERP (Löw et al., 2003), with early negativity (N400) in the occipito-temporal lobe, and late positivity (P600) in the

frontal lobe. Evidence from magnetoencephalography (MEG) also confirms a posterior to anterior response sequence to word processing (Dhond et al., 2001).

These results suggest that the TL is the first structure to receive language input. This structure seems to play a role in encoding information and guiding the activation of other structures (Stowe et al., 2002). Downstream of the TL, another structure implicated in ambiguity resolution is the VLPC. This module acts as a working memory (WM) for semantic information. The VLPC retrieves pertinent information from long-term memory (LTM; Thompson-Schill, D'Esposito, & Kan, 1999) and re-formulates it in a useful form for the task at hand (Bunge et al., 2003). Once some representations are retrieved from LTM and activated in the VLPC, they compete through mutually inhibitory interactions (Miller & Cohen, 2001). After this competition process, activity is maintained in a small subset of the retrieved representations, and suppressed in others. This activity is maintained for the duration of the task, and decays afterwards if no longer solicited (Wagner, 2001).

Finally, both the VLPC and TL relay information to the LIFG. This structure is involved in the selection of semantic knowledge among competing alternatives, by integrating information between the TL and VLPC, and outputting a lexical decision. fMRI revealed that activation in the LIFG is correlated with word generation when word competition is high (Bookheimer, 2002). Studies using ERP argue for inhibitory processing of homonyms (Gunter, Wagner, & Friederici, 2003). With these results, we can conceive of the LIFG as a "gate" that integrates the representations activated in the VLPC with the input information encoded in the TL, and suppresses undesirable entries.

### Goals of the frontotemporal pathways

Three main functional goals have been associated with the frontotemporal pathways involved in lexical ambiguity resolution. One goal is to find correspondences between input stimuli (TL) and prior encounters of those stimuli stored in memory (VLPC). Mathematically, this goal can be described as an objective function ( $cov$ ) that maximizes covariance between the activities of these two regions. In the absence of relevant knowledge in memory, a similar function can be used between the TL and LIFG regions.

A second goal is to reduce error between information coming into the loop (TL), and information going out (LIFG), thus reproducing the signal as faithfully as possible. In mathematical terms, this can be described by an error-minimizing objective function ( $e$ ). Research on ERP error-related negativity has argued for such an error-driven process, by demonstrating that subjects' incorrect motor responses became phase-locked with prefrontal theta-band activity (Luu, Tucker, & Makeig, 2004).

A third goal is to be able to store new words. This should be accomplished simply by exposure, without requiring an error signal, or any type of feedback. For instance, new words can be extracted from a conversation or reading material. Mathematically, while the first two objective functions require a gradual, iterative process, the goal of storing new words should be addressed by a quick

automatic process. Computationally, this will be performed by a "knowledge injection" technique.

Although the three goals proposed are not presumed to fully capture the neurophysiological processes involved, our simulations will show that they are sufficient to capture the main results associated with frequency effects in meaning disambiguation.

### Integrating modules and goals

Integrating the previously described modules and functional goals serves as the basis of our preliminary account of disambiguation. Initially, the TL lexical representations are activated by an incoming word stimulus. Activation then spreads to the VLPC and LIFG regions, and an error signal is produced by comparing the signal from the TL and LIFG, through a set of direct connections between those two regions. At the same time, the VLPC retrieves a small subset of word representations from LTM that have some correspondence to the TL signal. Through an interaction between the TL and VLPC, these representations then compete in a winner-take-all fashion to provide an interpretation that best matches the TL signal. Finally, the LIFG acts as a gate that selects a winning representation, and outputs the final interpretation. Connections from the VLPC to the LIFG can be adjusted in order to reduce any residual discrepancy between the winning representation and the expected output signal.

### A Computational Model

A computational model of lexical ambiguity resolution involving the frontotemporal pathways requires both a distributed network between the brain regions involved and some clusters of neurons that are functionally specialized. The input to the model described will be the part of speech, meaning, pronunciation, and spelling of words. As in Kawamoto (1993), LTM memory consists of distributed lexical knowledge, and contextual constraint is simulated by "pre-activating" the meaning units of a given word's lexical representation. The network processes the input by matching its activation to representations retrieved from LTM. At the output, the network reproduces the input as faithfully as possible. Our model does not claim to capture the retrieval process from LTM but rather the subsequent competition that ensues among alternative representations, as well as the final interpretation of the input word.

The proposed model of ambiguity resolution is based on an autoencoder network. The goal of this network is to reproduce its input onto its output, as faithfully as possible, by adjusting the strengths of the connections linking the neurons together. The algorithm used to adjust weights is a modified version of the Rule-based Cascade-correlation network (RBCC; Thivierge, Dandurand, & Shultz, 2004). Word activation is typically measured by reaction time, and autoencoder networks have been successfully employed in biologically plausible models of reaction time studies (Sirois & Mareschal, 2004). In addition, autoencoding is known to exist in the neocortex (Rolls & Treves, 1998).

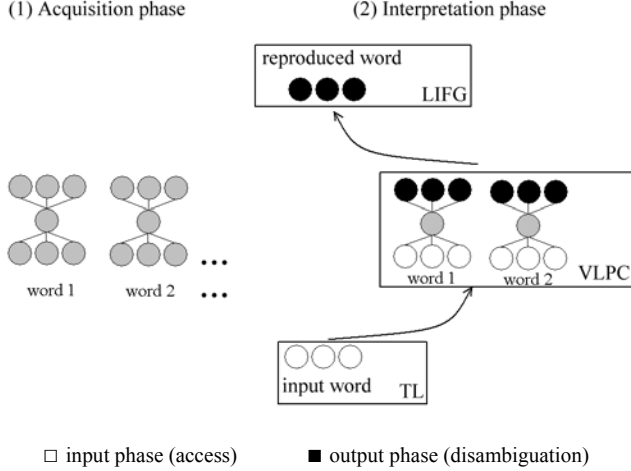


Figure 2: Acquisition and interpretation phases of the RBCC model. In the acquisition phase, various words are encoded in network form, and stored in memory. In the interpretation phase, the stored networks are retrieved and compete by adjustment of the weights between the nodes in white. Once a winner is determined, a disambiguation process takes place by adjustment of the weights between the nodes in black.

The functioning of the RBCC networks used here can be divided into two distinct phases, namely acquisition phase and interpretation phase (see Figure 2). In the acquisition phase, the goal of RBCC is to encode a number of words that will later be activated in WM. In order to do so, each word is first converted to a Horn clause, and then injected in the network. Each word is represented by a set of  $m$  random binary features that can only be in one of two states (on/off). In Horn clauses, the outcome is placed on the left hand side, and the antecedents on the right hand side. Absence of a feature  $f$  is denoted  $\text{not}(f)$ . Here is one instance of a Horn clause transforming a word into a vector:

$$\text{BANK\_FINANCIAL} := 1, \text{not}(2), \text{not}(3), \dots, m$$

These Horn clauses are directly encoded in a network with three layers: input, hidden, and output. Weights between the input and hidden layers are equal to  $W=4$ , except the bias weight  $q=(m+1-2n)W$ , where  $n$  is the number of positive features in a word vector of size  $m$ . Weights between the hidden and output layers are equal to  $x$  (the vector of input patterns). These networks encode words in an autoencoder fashion, hence there are  $m$  inputs and  $m$  outputs.

Once all words in a corpus are encoded in network form according to Horn clauses, the acquisition phase ends, and RBCC moves on to the interpretation phase. In this phase, words are presented at the input of the system, and a correct interpretation must be provided based on information retrieved from the acquisition phase. The process of interpreting a given word is initiated when the TL module receives some input activation, which is then propagated through the VLPC and LIFG modules before an output is

produced. At that time, activation is sent back to the TL module, and an error signal  $e$  is computed based on the discrepancy between the LIFG ( $L$ ) and TL ( $T$ ) activations:

$$e = \sum_o \sum_p (L_{o,p} - T_{o,p})^2, \quad (1)$$

summed over all neurons  $o$  and patterns  $p$ . Training connection weights according to Eq.1 is equivalent to the "output phase" of standard RBCC networks. As the error signal is generated, several word representations get activated in VLPC. In order to enable competition of these representations, the weights between the TL and VLPC are adjusted to maximize a score of covariance:

$$\text{cov}_c = h_c \frac{\sum_{o_c} \sum_o \left| \sum_p (V_{o_c,p} - \bar{V}_{o_c}) (E_{o,p} - \bar{E}_o) \right|}{\#O_c \cdot \#O \cdot \sum_o \sum_p E_{o,p}^2} \quad (2)$$

between the error signal  $E$  and the output  $V$  of each of the  $c$  competing representations in VLPC, where  $o_c$  indexes an output for each candidate. Training connection weights according to Eq. 2 is equivalent to the "input phase" of standard RBCC networks. At the end of a number of training epochs, the word representation with the highest covariance score wins, and forms the basis of the interpretation of the input word. Activation in the winning representation is maintained for the duration of the task at hand, while the activation of the other competing alternatives falls back to zero.

This process must be slightly modified if no relevant representations are available in VLPC. In this case, a novel interpretation is generated online by adding neurons between the TL and LIFG, and maximizing the covariance of these units with the error of the network.

Eq.2 introduces a scoring factor  $h_c$  for each competing word that allows the network to be sensitive to the frequency at which words occur in the vocabulary. This scoring factor is updated at the end of each interpretation phase:  $h_c = b_c G_c$ . Thus,  $h_c$  is determined by the final covariance value of each competing word at the end of the interpretation phase, weighted by a factor  $b_c$  corresponding to a learning rate that can be chosen independently for each competing word (here,  $b_c = 10$ ). As experiments below will demonstrate, the  $h_c$  parameter is essential to capture frequency effects.

Information processing in the brain seems consistent with the idea that, once a given representation is selected, its chances of being selected again are higher. This principle represents an extension of Hebbian learning: if a sending neuron  $A$  activates a target neuron  $B$ , their connection will be strengthened such that upon the next activation,  $A$  will more easily activate  $B$ .

RBCC networks start training in input phase, followed by output phase. In linguistic terms, the input phase of RBCC corresponds to an *access* process (i.e., retrieving some lexical item from the mental lexicon) and the output phase is analogous to a *disambiguation* process (i.e., correctly parsing ambiguous linguistic input).

## Simulations

In the following simulations, RBCC networks encoded three words in the acquisition phase, namely a homonym connected to one particular meaning ("BANK" - money), the same homonym connected to an alternative meaning ("BANK" - river), and an unambiguous control word. All three words are postulated to be available for processing by retrieval from LTM. Each word was composed of  $m=195$  features in total: 48 for spelling, 48 for pronunciation, 96 for meaning, and 3 for part of speech. These features were encoded as random binary vectors.

During the interpretation phase, the network was exposed to the target word. Each pattern consisted in a vector of length  $m$ . For each training iteration, the network received one component of the target word (meaning, spelling, pronunciation, or part of speech), while the features corresponding to the other components were set to zero. In this procedure, meaning was presented prior to any other component, thus allowing the network to pre-activate the influence of the context.

### Processing advantage of previously learned entries

One goal of the simulations was to assess the advantage of having previously acquired word entries during the acquisition phase that would become relevant for a given context. For this purpose, we ran twenty networks that had access to contextually relevant entries. The performance of these networks was compared to that of twenty networks that did not have access to those contextually relevant entries (i.e., they had access to other lexical entries as well as logistic units containing no word representation). In these networks, no representations are activated in VLPC, so the alternative TL-LIFG route is employed. In order to solve the disambiguation problem, networks must build a novel representation online, by using simple logistic units. In this process, networks were allowed to recruit up to 25 units.

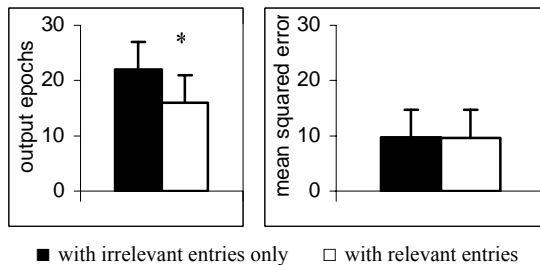


Figure 3: Comparison of networks with relevant or irrelevant lexical representations. \*  $p < 0.05$  (t-test).

Results revealed that having the correct lexical entry stored in memory decreased the average number of training epochs in output phase (50 input phase epochs were always required), while the final error remained unchanged (see Figure 3). Thus, the network recognized and made use of the lexical entries it had previously learned. Without these entries, networks prefer to recruit simple logistic units, which allow them to build a novel interpretation online, a

process that is more time-consuming than acting upon a representation already retrieved and available in WM.

### Time-course of word activation

A second goal of the simulation was to assess the impact of frequency on the time-course of word activation in context. Forty independent networks were exposed to one acquisition and two interpretation phases, and coding was randomized for every network. In order to settle into a stable interpretation, the network was allowed a maximum of 100 weight-updating epochs. To manipulate alternative meaning frequency, we created two conditions of twenty networks each that varied the intended meaning of the second interpretation phase. Note that in both conditions, the first interpretation phase simulated, by definition, the processing of equibaised homonyms (i.e., homonyms with two meanings of equal frequency) in context. This is because during acquisition, each meaning of a homonym had equal weighting. However, the two conditions manipulated meaning frequency in the second interpretation phase. This is because the immediate prior exposure to a particular meaning in the first interpretation phase rendered that meaning dominant over the other meaning for the subsequent second interpretation phase. Thus, we simulated a dominant meaning embedded in a supporting context when the contextually intended meanings in the first and second phases matched. In contrast, we simulated a subordinate meaning embedded in a supporting context when the contextually intended meanings in the first and second phases did not match. Here, we manipulated functional dominance as opposed to actual dominance, by letting the network create its own representation of which entry has more frequency (and is thus more dominant).

One limit of this simulation is that frequency is a gradual effect of long-term exposure to a lexicon, while in our model, frequency effects are obtained by a single exposure. However, the end result will be the same: selective exposure to certain entries over others will favour one interpretation over another. The details of how this effect is accumulated over time are left for further modeling.

Using this simulated manipulation of frequency and context, we are able to plot activation of contextually intended meanings for equibaised homonyms (based on the first interpretation phase results), activation for biased homonyms in a dominant context (based on the results of the second interpretation phase when it matched the first interpretation phase), and activation for biased homonyms in a subordinate context (based on the results of the second interpretation phase when it did not match the first interpretation phase).

Figure 4a shows the time-course of activation of an equibaised homonym in a context that biases one meaning. Here, the contextually intended meaning gains a large amount of covariance with the error signal (i.e., activation), and eventually wins in the process (i.e., is selected). The alternative contextually unintended meaning gains some slight amount of covariance, but flattens out after a few epochs (i.e., is inhibited). Finally, the unambiguous

unrelated control word gains virtually no covariance during the training process.

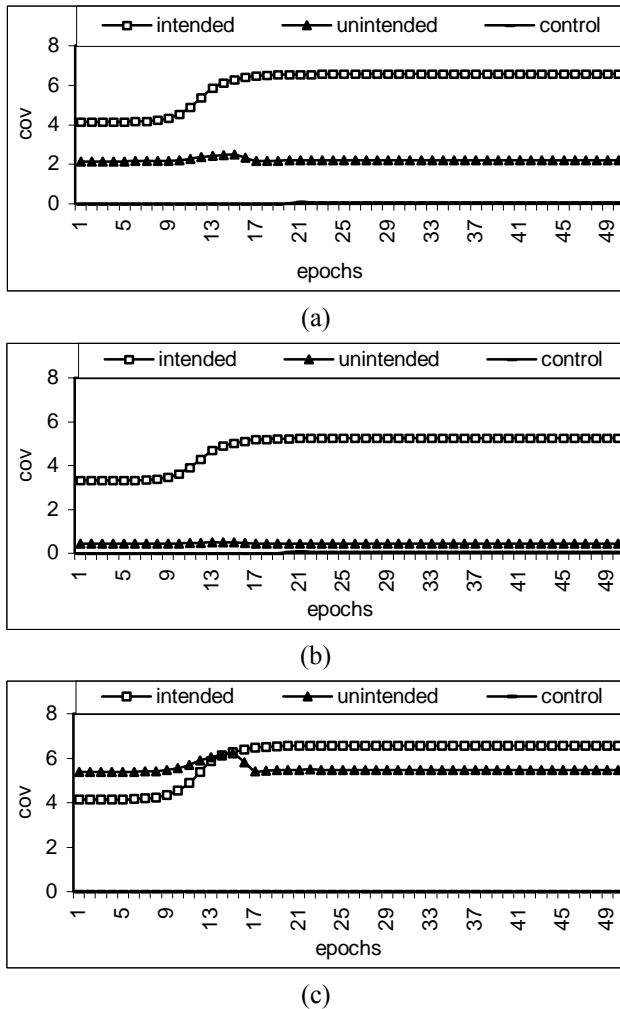


Figure 4: Average covariance of competing words that have (a) equal frequencies (1<sup>st</sup> interpretation phase); (b) higher frequency of the target (2<sup>nd</sup> interpretation phase); and (c) higher frequency of the homonym (2<sup>nd</sup> interpretation phase).

Figure 4b shows the time-course of activation of a biased homonym in a context that supports the dominant meaning. Here, the contextually intended meaning has increased in frequency as a result of the previous interpretation phase, and there is thus more discrepancy between the target and alternative interpretations.

Figure 4c shows the activation of a biased homonym in a context that supports the subordinate meaning. As a result of the previous interpretation phase, the unintended meaning has a higher frequency score than the intended meaning (6.56 vs. 2.17 respectively). Hence, its covariance is initially higher than the intended meaning. However, through weight adjustment, the system is able to eventually select the contextually intended meaning.

In summary, RBCC networks can make use of relevant entries to perform ambiguity resolution. In addition, they show an influence of previous exposure to word meanings.

### Scalability

Although the number of words tested is limited, it can be argued that the results obtained would generalize to larger corpora. Firstly, adding more lexical entries to the knowledge base of the network will likely not create conflicts because these entries would be equivalent to unrelated controls words. Thus their covariance would remain close to zero. Secondly, the mathematics of the model make it relatively easy to scale-up; the computational resources required to compute the covariance of additional words grow in quadratic time  $O(mn)$ , where  $m$  is the number of features of each lexical entry, and  $n$  is the total number of lexical entries (further analytical demonstrations will be included in a longer paper). Thirdly, we have shown that RBCC is able to compute over a large number of lexical features (195 in the experiments presented). Using such a large number of random features for each lexical entry increases our confidence that the results obtained are not due to fortuitous aspects of the encoding.

### Discussion

The current paper describes a preliminary neurocomputational model of ambiguity resolution based on frontotemporal pathways that have been shown by neuroimaging to be of prime importance. Other computational models have addressed the problem of ambiguity resolution, including Hopfield networks (Kawamoto, 1993; Borowsky & Masson, 1996) and models of prefrontal processing (Cohen & Servan-Schrieber, 1992; O'Reilly, 2002). However, these models do not emphasize to the same degree interactions between several brain regions that have been shown to be of major importance in lexical ambiguity resolution.

### Recirculation of activation

Although neuroimaging research has identified potential regions of interest in lexical disambiguation, the precise time-course of activation of these regions is still largely unknown. EEG research suggests a direction of activation going from the TL to prefrontal regions (Löw et al., 2003). Our model mostly agrees with these findings, and shows that a single direction of processing is sufficient to account for disambiguation. As seen from Fig.3a-b, a single epoch (i.e., a single pass forward) is sufficient to disambiguate the intended and unintended meanings. However, our simulations show an exception in the case where frequency biases against the intended interpretation (Fig.3c). In this case, a single pass forward will not allow disambiguation; on the first epoch of input phase, the unintended meaning is more highly activated than the intended meaning. Recirculation of information is necessary to raise activation in the intended meaning above that of the unintended meaning. This recirculation implies a longer time-course of activation than the other conditions.

## Future research

Future work will focus on lesioning various regions of the model, and simulating the impact of various neurological pathologies associated with language processing. In this way, we may capture several results. For instance, aphasia caused by focal LIFG lesion has been linked to activation of both meanings of an ambiguous word regardless of the context (Grindrod & Baum, 2002). In our model, this could be interpreted as an inability of LIFG to "gate" the entries.

The model constitutes a first attempt at integrating neuroimaging findings with a computational account of lexical ambiguity resolution. As highlighted throughout, several aspects of the model need to be modified to better account for the phenomena. In addition, our account is only one of several possibilities. Finally, our account of brain processes based on presumed goals of brain modules is at a very abstract level. Thus, our model is not an account of what neurons do, and future models should address how these goals can be represented in neurologically plausible rules (e.g., Hebb's "fire together, wire together" rule).

## Acknowledgments

This research was supported by a scholarship from FCAR (first author), grants from NSERC, CRC, and CFI (second author), and a grant from NSERC (third author).

## References

- Bastiaansen, M.C.M., van Berkum, J.J.A., & Hagoort, P. (2002). Syntactic processing modulates the  $\alpha$  rhythm of the human EEG. *NeuroImage*, *17*, 1479-1492.
- Borowsky, R. & Masson, M.E.J. (1996). Semantic ambiguity effects in word-identification. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*, 63-85.
- Bookheimer, S. (2002). Functional MRI of language: New approaches to understanding the cortical organization of semantic processing. *Annual Review of Neuroscience*, *25*, 151-188.
- Bunge, S.A., Kahn, I, Wallis, J., Miller, E.K., & Wagner, A.D. (2003). Neural circuits subserving the retrieval and maintenance of abstract rules. *Journal of Neurophysiology*, *90*, 3419-3428.
- Cohen, J.D., & Servan-Schreiber, D. (1992). Context, cortex, and dopamine: A connectionist approach to behavior and biology in Schizophrenia. *Psychological Review*, *99*, 45-77.
- Dhond, R.P., Buckner, R.L., Dale, A.M., Marinkovic, K, & Halgren, E. (2001). Spatiotemporal maps of brain activity underlying word generation and their modification during repetition priming. *The Journal of Neuroscience*, *21*, 3564-3571.
- Grindrod, C.M., & Baum, S.R. (2002). Sentence context effects and the timecourse of lexical ambiguity resolution in nonfluent aphasia, *Brain and Cognition*, *48*, 381-385.
- Gunter, T.C., Wagner, S., & Friederici, A.D. (2003). Working memory and lexical ambiguity resolution as revealed by ERPs: A difficult case for activation theories. *Journal of Cognitive Neuroscience*, *15*, 643-657.
- 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.
- Löw, A. et al. (2003). Semantic categorization in the human brain: Spatiotemporal dynamics revealed by magnetoencephalography. *Psychological Science*, *14*, 367-372.
- Luu, P., Tucker, D.M., & Makeig, S. (2004). Frontal midline theta and the error-related negativity: Neurophysiological mechanisms of action regulation. *Clinical Neurophysiology*, *115*, 1821-1835.
- Miller, E.K., & Cohen, J.D. (2001). An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience*, *24*, 167-202.
- O'Reilly, R. C., Noelle, D. C., Braver, T. S., & Cohen, J. D. (2002). Prefrontal cortex and dynamic categorization tasks: Representational organization and neuromodulatory control, *Cerebral Cortex*, *12*, 246-257.
- Pulvermüller, F. (2001). Brain reflections of words and their meaning. *Trends in Cognitive Sciences*, *5*, 517-524.
- Rolls, E.T., & Treves, A. (1998). *Neural networks and brain function*. Oxford: Oxford University Press.
- Rodd, J.M., Davis, M.H., & Johnsrude, I.S. (2005). The neural mechanisms of speech comprehension: fMRI studies of semantic ambiguity. *Cerebral Cortex*, advanced access.
- Sirois, S., & Mareschal, D. (2004). An interacting systems model of infant habituation, *Journal of Cognitive Neuroscience*, *16*, 1352-1362.
- Stowe, L.A., Withaar, R.G., Wijers, A.A., Broere, C.A.J., & Paans, A.M.J. (2002). Encoding and storage in working memory during sentence comprehension. In: Merlo, P., & Stevenson, S. *The lexical basis of sentence processing : Form, Computational and Experimental Issues*. (pp. 181-205). Philadelphia, PA: John Benjamins.
- Thivierge, J.P., Dandurand, F., & Shultz, T.R. (2004). Transferring domain rules in a constructive network: Introducing RBCC. *IEEE International Joint Conference on Neural Networks 2004*.
- Thompson-Schill, S.L., D'Esposito, M., & Kan, I.P. (1999). Effects of repetition and competition on activity in left prefrontal cortex during word generation. *Neuron*, *23*, 513-522.
- Titone, D. (1998). Hemispheric differences in context sensitivity during lexical ambiguity resolution. *Brain and Language*, *65*, 361-394.
- Wagner, A.D., Pare-Blagoev, E.J., Clark, J., Poldrack, R.A. (2001). Recovering meaning: left prefrontal cortex guides controlled semantic retrieval. *Neuron*, *31*, 329-338.