

Should Psychology Ignore the Language of the Brain?

Peter Hagoort

F.C. Donders Centre for Cognitive Neuroimaging, Radboud University Nijmegen, and Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands

ABSTRACT—*Claims that neuroscientific data do not contribute to our understanding of psychological functions have been made recently. Here I argue that these criticisms are solely based on an analysis of functional magnetic resonance imaging (fMRI) studies. However, fMRI is only one of the methods in the toolkit of cognitive neuroscience. I provide examples from research on event-related brain potentials (ERPs) that have contributed to our understanding of the cognitive architecture of human language functions. In addition, I provide evidence of (possible) contributions from fMRI measurements to our understanding of the functional architecture of language processing. Finally, I argue that a neurobiology of human language that integrates information about the necessary genetic and neural infrastructures will allow us to answer certain questions that are not answerable if all we have is evidence from behavior.*

KEYWORDS—*language; brain; ERP; fMRI*

From a sociology of science perspective, cognitive neuroscience is a tremendous success. In part due to the enormous technical progress in noninvasive recordings of activity in the living human brain, a whole industry of research on brain and cognition has developed. The number of neuroimaging research centers has grown exponentially over the last decade, as has the number of publications reporting functional magnetic resonance imaging (fMRI) results. However, this success is not without its critics. The criticisms follow roughly the following logic: If you are interested in the genetic and neurobiological infrastructure underlying and implementing human cognition, that is just fine. Go ahead and do your research. However (according to the critics), if you are interested in characterizing human cognition itself, it does not help to know that “it happens somewhere north

of the neck” (Fodor, 1999). Despite the vast amount of neuroimaging studies, theories of human cognition have not (yet) profited from measuring the brain. The reason can be principled or practical. The principled reason is that our explanation of the mind “abstracts away from the biological realizations of cognitive structures” (Block, 1990, p. 261). The relation between mind and brain is not transparent enough to result in sufficiently strong constraints for cognitive theories by virtue of our understanding of the brain (Fodor, 1974; Page, 2006). The less severe variant of the critique is that one could potentially derive useful information from neuroimaging but, in practice, no actual examples that successfully distinguish between competing psychological theories have been provided.

The first thing to note about this debate is that the critique, at least implicitly, expresses doubts that cognitive neuroscience is a fruitful scientific endeavor. In this context, it is surprising that all of the criticisms focus on fMRI results only (Coltheart, 2006; Page, 2006; Uttal, 2001). The arguments are thus largely based on selective shopping in the toolkit of cognitive neuroscience. This toolkit has a lot more to offer than just fMRI; it also includes event-related brain potentials (ERPs), magnetoencephalography (MEG), transcranial magnetic stimulation (TMS), and measurements of brain structure. Moreover, theories in cognitive neuroscience are informed by animal models and studies using single- or multi-unit recordings of action potentials and/or local field potentials (i.e., measures of neuronal activity at the micro-scale of brain organization). The general issue is, therefore, not what fMRI contributes but, rather, how far cognitive neuroscience is a viable scientific enterprise using whatever research tools are at its disposal. This is the question that I will address here, for one particular, central cognitive skill: our capacity to communicate by means of natural language. But first, we need to specify the criteria for success.

To establish the criteria that an adequate neurobiology of language has to meet, we first need to clarify what we take our *explanandum*—that is, the thing we want to explain—to be. If one is interested not only in the cognitive architecture of language but also in the only machinery that so far has been able

Address correspondence to Peter Hagoort, F.C. Donders Centre for Cognitive Neuroimaging, P.O. Box 9101, NL-6500 HB Nijmegen, The Netherlands; e-mail: peter.hagoort@fcdonders.ru.nl.

to instantiate natural language (i.e., the human brain), it is obvious that the bridge between psycholinguistics and neurobiology has to be crossed. However, it is a valid position to restrict one's explanandum to the *cognitive* architecture of language functions. In that case, the brain facts will only be relevant in so far as they can be used to develop, select, or constrain a cognitive-architecture model for the language function of interest. The cognitive architecture then specifies the levels of representation needed and the processing steps required for accessing representational structures and for performing the necessary computational operations on them. Even in this case, brain facts might be relevant.

One example of a brain fact relevant for cognitive models relates to the nature of the flow of information. Strictly feedforward models of language comprehension (e.g., Cutler & Clifton, 1999) predict a fixed spatiotemporal pattern of brain activity that is not seriously modulated by attention or output-related factors. Such models are compatible with a serial model of perception and action, in which a perceptual stage is followed by central cognition (e.g., executive function), which is then followed by appropriate action (cf. Fodor, 1983). One of the arguments for strictly feedforward systems is that only these will guarantee the high speed that characterizes many aspects of human cognition. For instance, we easily recognize and understand three or four words per second. How would this be possible if all information that we have in memory about the entities to which these words refer could, in principle, have an impact on language perception?

Despite the design argument in favor of feedforward models, recent findings in cognitive neuroscience raise serious doubts about the general tenability of strictly feedforward serial models. Certainly, the neural architecture of the human brain supports feedback options. From all we know now, recurrent loops are a fundamental characteristic of the mammalian neocortex. All layers of the cortex are heavily back-connected to earlier regions in the neural processing chain. Information flow in the brain is, therefore, not one-way. The cerebral cortex is basically a feedback system, and the lack of top-down influences on perception (informational encapsulation) can thus not be based on the architecture of the cortex. This does not necessarily mean that information flow, in terms of a processing model, cannot be feedforward. If, under conditions of normal operation, the input system provides its input to the next level in the processing hierarchy before feedback can have its effect, informational encapsulation might still be achieved. However, in this case, it is speed that buys the system informational encapsulation, rather than the reverse. Moreover, it is only a soft form of encapsulation (i.e., not hardwired in the cortex), since with additional time, feedback will start to have its effect. Importantly, much recent evidence in cognitive neuroscience suggests that perception is influenced by the observer's attentional state, the task, and the observer's strategies. These seem to be relevant considerations for functional-architecture models of human cognition.

An adequate neurobiology of language might thus provide data that are relevant for cognitive models of language functions.

At the same time, the relevant brain facts can only be obtained in neuroscience research that is guided by state-of-the-art psycholinguistics in terms of theoretical models and experimental materials. The criteria for an integrated neurobiology of language are thus (a) specifications of the neural principles of language functions that are adequate in relation to behavioral data and the cognitive architectures derived from these data (*upward adequacy*) and (b) specifications of the cognitive architectures that are adequate in the light of our understanding of the principles of brain function (*downward adequacy*). The underlying assumption is that there is a systematic relation between cognitive states and brain states. Despite claims made in the past that these two levels of description and explanation might not be related in a lawful or transparent way (e.g., Fodor, 1974; Mehler, Morton, & Jusczyk, 1984), many believe that cognitive neuroscience has made sufficient progress to warrant this assumption a certain face validity.

Here I give a few examples to indicate where the contribution to psycholinguistics could be seen for two different methods used in cognitive neuroscience—namely, ERPs and fMRI. However, I want to stress that, at the end of the day, it is not a single method that is going to make the difference.

A FEW EXAMPLES

The recording of ERPs is the oldest and cheapest method in the toolkit of cognitive neuroscience. ERPs reflect the sum of simultaneous postsynaptic activity of a large population of mostly pyramidal neurons recorded at the scalp as small voltage fluctuations in the electroencephalogram (EEG), time locked to sensory, motor, or cognitive processes.

Study of the electrophysiology of language started with the discovery by Kutas & Hillyard (1980) of an ERP component that seemed especially sensitive to semantic manipulations. Kutas and Hillyard observed a negative-going potential (a brain wave with a negative amplitude) with an onset at about 250 milliseconds (ms) after a word stimulus appeared on the screen, and a peak around 400 ms (hence the name N400), whose amplitude was increased when the semantics of the eliciting word (e.g., *socks*) mismatched the semantics of the sentence context—as in “He spread his warm bread with socks.” Since 1980, much has been learned about the processing nature of the N400. We know now that the N400 effect does not depend on a semantic violation. Subtle differences in semantic expectancy can modulate the N400 amplitude as well (see Fig. 1; Hagoort & Brown, 1994). Modulations of the N400 amplitude are generally viewed as directly or indirectly related to the processing costs of integrating the meaning of a word into the overall meaning representation that is built up on the basis of the preceding language input (Brown & Hagoort, 1993).

A different set of ERP effects has been observed in connection to the processing of syntactic information. The two most salient syntax-related effects are an anterior ERP with a negative am-

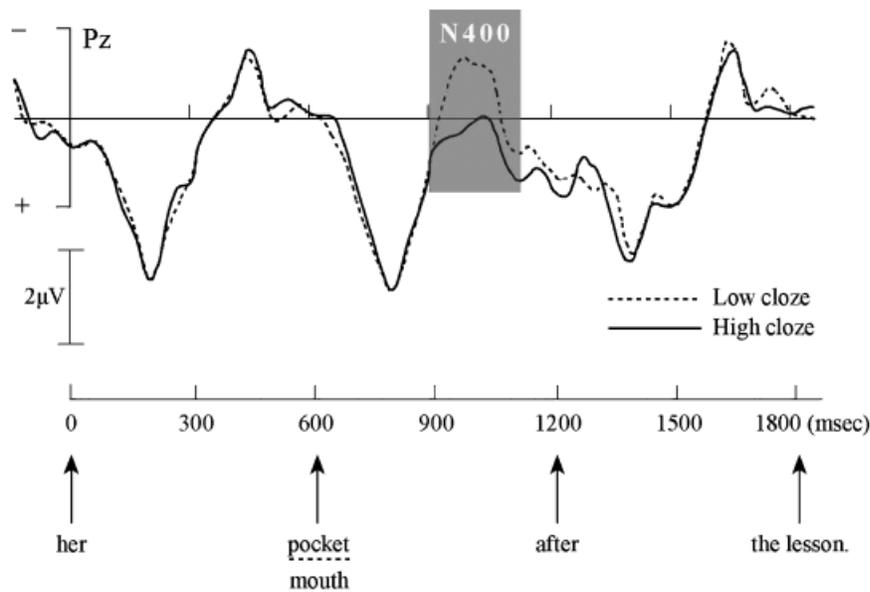


Fig. 1. Modulation of the N400-amplitude event-related brain potential as a result of a manipulation of the semantic fit between a lexical item (either the word *pocket* or the word *mouth*) and its sentence context (“Jenny put the sweet in her ____ after the lesson”). The grand-average waveform is shown for electrode site Pz (parietal midline) for the best-fitting word (high cloze; solid line) and the word that is less expected in the given sentence context (low cloze; dashed line). The sentences were visually presented word by word, every 600 milliseconds (msec). The critical words are shown at 600 msec on the time axis. (Negative is up.) Adapted from *The Neurocognition of Language*, C.M. Brown & P. Hagoort, eds., 1999, Oxford University Press, p. 281. Copyright 1999, Oxford University Press. Adapted with permission.

plitude (left anterior negativity, or LAN), and a more posterior ERP effect with a positive-going amplitude, usually referred to as P600. I will here focus on the P600 (Hagoort, Brown, & Groothusen, 1993). This effect is triggered by a violation of a syntactic constraint or a difference in syntactic complexity. If, for instance, the syntactic requirement of number agreement between the grammatical subject of a sentence and its finite verb is violated—for example, “The boys kisses the girls” (see also sentence 1b below)—a positive shift in the ERP waveform is observed that starts at about 500 ms after the onset of the violation and usually lasts for at least 500 ms. An argument for the independence of this effect from possibly confounding semantic factors is that it also occurs in sentences where usual semantic/pragmatic constraints are not present (Hagoort & Brown 1994). Removing such constraints results in sentences like the following, where one (1a) is semantically odd but grammatically correct, whereas the other (1b) contains an agreement violation (marked by the asterisk):

1a. The boiled watering-can *smokes* the telephone in the cat.

1b. The boiled watering-can *smoke** the telephone in the cat.

As can be seen in Figure 2, these ERP effects in response to syntactic processing are qualitatively different from the N400. Even though the generators of these effects are not yet well determined and not necessarily specific to language, the existence

of qualitatively distinct ERP effects to semantic and syntactic processing indicates that the brain honors the distinction between semantic and syntactic processing operations. Thus, the finding of qualitatively distinct ERP effects for semantic and syntactic operations supports the claim that these two levels of language processing are *domain specific*. That is, the ERP evidence indicates that syntactic computations cannot be collapsed into a general-purpose language processor, whose internal machinery does not maintain the distinction between different types of information (lexical, syntactic, semantic), as suggested in connectionist models (Tabor & Tanenhaus, 1999).

However, domain specificity should not be confused with modularity (Fodor 1983). The modularity thesis makes the much stronger claim that domain-specific levels of processing operate autonomously without interaction (i.e., informational encapsulation). Although domain specificity is widely assumed in models of language processing, there is much less agreement about the organization of the interaction between the different levels of processing. Recently, new light has been shed on this issue by a series of ERP studies reporting a P600 associated with thematic role assignment. In this case, a P600 is elicited when constraints for grammatical role assignment are in conflict with thematic role biases. For instance, Kim and Osterhout (2005) report a P600 to the verb *devouring* in the sentence “The hearty meal was devouring . . .”, where the first noun phrase is not a good agent but would be fine as a theme. The fascinating possibility suggested by these results is that a strong thematic bias

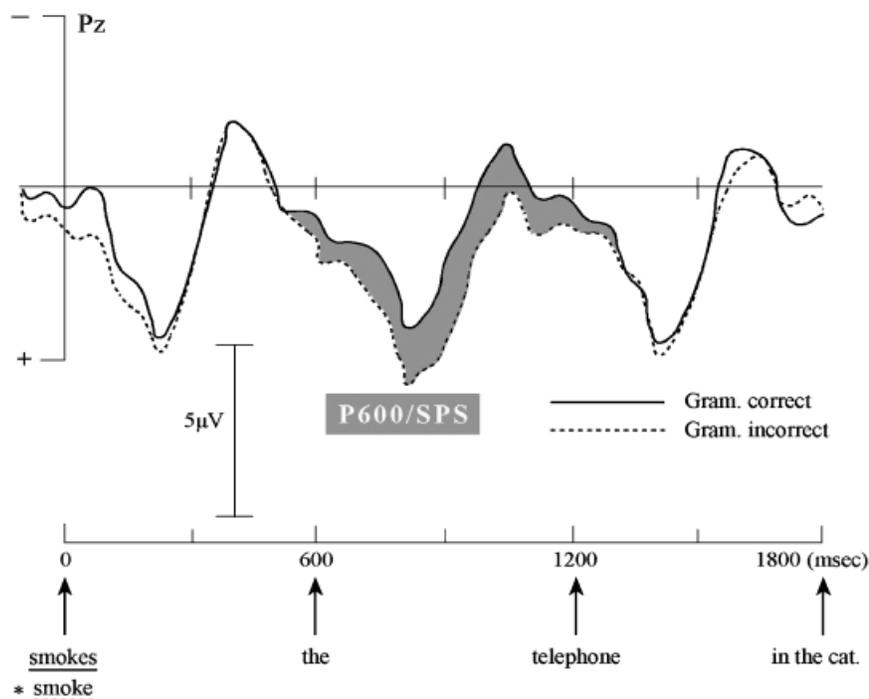


Fig. 2. Event-related brain potentials (ERPs) to visually presented sentences without a coherent semantic interpretation. A positive-going brain wave (P600/SPS) is elicited by a violation of the required number agreement between the subject-noun phrase (“The boiled watering-can”) and the finite verb of the sentence (*smoke*). The averaged waveforms for the grammatically correct word (solid line, “smokes”) and the grammatically incorrect word (dashed line, “smoke”) are shown for electrode site Pz (parietal midline). The word that renders the sentence ungrammatical is presented at 0 milliseconds (msec) on the time axis. Words were presented word by word, with an interval of 600 msec. (Negative is up.) Adapted from *The Neurocognition of Language*, C.M. Brown & P. Hagoort, eds., 1999, Oxford University Press, p. 287. Copyright 1999, Oxford University Press. Adapted with permission.

could induce a tendency to detect a grammatical error where there is none (e.g., “-ing should be -ed”) or to assign the grammatical role of object to the first noun phrase, whereas the syntactic cues indicate that it is the subject of the sentence. As a result of this conflict between thematic role biases and syntactic cues, a P600 results. In this case, semantic factors are so strong that they seem to impose a syntactic structure onto the input that is not provided by the syntactic cues themselves. As these results show, the fact that ERPs provide potentially qualitatively different effects for qualitatively different processes with a temporal resolution in the milliseconds range is a distinct advantage over unidimensional measures such as reaction time and, to some degree, measuring eye movements during reading.

These and many other results (for further arguments, see Hagoort & Van Berkum, 2007) are in line with the *immediacy assumption*, which states that all available information types are brought to bear on language interpretation as soon as they become available, without giving priority to the syntax-constrained combination of lexical-semantic information.

ERP research on language these days is not present only in cognitive neuroscience journals but also in journals of experimental psychology and psycholinguistics. Being a method based on neurophysiological activity in the cortex, ERPs nevertheless

play an important role in guiding and testing purely functional models of language processing.

Skepticism about the contribution to our understanding of the cognitive architecture is more substantial for fMRI than for ERPs. With respect to the language system, this criticism is partly justified. I had the privilege to review the language abstracts for the annual meeting of the Organization for Human Brain Mapping for a number of years. Overall, the psycholinguistic quality of the majority of these submissions is disappointing. The sophistication in psycholinguistics in carefully controlling for numerous potential confounds in the materials (frequency, familiarity, morphological structure, phonological structure, etc.) and in addressing issues based on explicit models of speaking, listening, reading, or writing is often not present in neuroimaging studies on language. This situation is, however, improving, as it should be. The lack of psycholinguistic sophistication of many fMRI studies on language does not mean that there is any principled reason why fMRI data would be useless for our understanding of the cognitive architecture of language. Let me give one hypothetical example that explains the logic of inference. This example provides an argument against the claim that on the basis of fMRI data nothing would change regarding our appreciation of the functional architecture of language processing (Page, 2006).

A few years ago, Kempen (2000) proposed an explicit computational model of syntactic processing that deals with both syntactic encoding and grammatical decoding (parsing). For a number of reasons (such as speaker–hearer alignment during dialog; Garrod & Pickering, 2004), a common mechanism (in terms of cognitive resources) for grammatical encoding and decoding is attractive. Nevertheless, the common-mechanism view goes against the standard view that assumes separate mechanisms for syntactic encoding and parsing. To decide empirically between the one- and two-mechanisms architectures, brain facts might be relevant. For instance, a common-mechanism view would be hard to reconcile with neuroimaging data that show a clear segregation of areas activated by encoding and areas activated by decoding. Under the reasonable assumption that a common-mechanism view and a separate-mechanism view have consequences for the hypothesized neural organization of grammatical encoding/decoding, brain facts do contribute to the body of empirical data that might guide the choice for one cognitive-architecture option over the other.

Apart from contributions of fMRI to our understanding of the cognitive architecture of language, I expect that we will soon see more evidence on the consequences for other cognitive functions from having a symbolic system (language). A striking example is a recent study from the group of Edmund Rolls in Oxford (de Araujo, Rolls, Velazco, Margot, & Cayeux, 2005), showing how the olfactory system can be influenced by language. In an event-related fMRI study, a test odor was delivered to the subjects. In half of the trials, the test odor was paired with the verbal label “cheddar cheese,” in the other half, with the verbal label “body odor.” Subjects rated the same test odor as unpleasant when labeled “body odor” and as neutral when labeled “cheddar cheese.” The authors found that activation in the medial orbitofrontal cortex was modulated by the verbal label that accompanied the test odor. The region where word labels modulate olfactory processing was found to be within the primary olfactory cortex, where main effects of odor were observed. Thus the activation in the olfactory input system produced by a test odor could be modulated by a cognitive marker provided by simultaneously presented words. This example illustrates something that would not so easily be found out with a behavioral method: that language information acts directly in the olfactory input system. In my opinion, such information is highly relevant for how we construe the architecture of human cognition.

EPILOGUE

I have mainly focused on the contribution of cognitive neuroscience to functional models of cognition, particularly language. Personally, I am interested in more than this. I would like to see a neurobiology of language that integrates the genetic and neural infrastructures necessary for human language with psycholinguistic models of language processing. There are many questions that can only be answered by measuring aspects of brain and genome. For instance, which aspects of the neuronal machinery are shared by language and other cognitive systems?

To what extent are learned language skills such as reading and writing built on the evolutionary hard-core system of spoken language? In what way is the neural infrastructure of people gifted in language (e.g., simultaneous interpreters; people who speak multiple tongues fluently) different from that of the average language user? At the same time, it is clear that answering these questions requires a detailed understanding of the cognitive architecture of the different language skills. Linguistic sophistication and psychologically motivated fractionations of complex language functions into their elementary components are necessary ingredients for asking the right questions about the underlying neurological and genetic infrastructures. This is clear from the history of neurolinguistics. The classical Wernicke-Lichtheim theory (1885; see Beaumont, Kenealy, & Rogers, 1996) about the neural organization of language was based on the implicit assumption that language consisted merely of words. Higher levels of organization such as sentence prosody, syntax, and semantic interpretation beyond the single-word level were not part of the Wernicke-Lichtheim framework and therefore were not examined in terms of the neural organization of language. Today, based on fine-grained cognitive models of language processing, the neural organization of language beyond the single-word level is an active area of research. This proves how crucial psychological theories are for our understanding of the brain. The major challenge for the coming decades is to connect the different levels of analysis and to determine how their mutual constraint relations are to be understood.

Recommended reading

- Brown, C.M., & Hagoort, P. (Eds.). (1999). *The neurocognition of language*. Oxford, UK: Oxford University Press. Provides functional blueprints of speaking, listening, and reading, as well as specifications of the neurobiology of language.
- Cutler, A. (Ed.). (2005). *Twenty-first century psycholinguistics: Four cornerstones*. Mahwah NJ: Erlbaum. An overview of recent developments in the field of psycholinguistics.
- Jackendoff, R. (2002). *Foundations of language: Brain, meaning, grammar, evolution*. Oxford UK: Oxford University Press. An interesting and influential linguistic account of the components of the human language faculty.
- Kutas, M., & Federmeier, K.D. (in press). Event-related brain potential (ERP) studies of sentence processing. In G. Gaskell (Ed.), *Oxford handbook of psycholinguistics*. Oxford, UK: Oxford University Press. A review of ERP studies on sentence processing.
- Osterhout, L., & Holcomb, P.J. (1995). Event-related potentials and language comprehension. In M.D. Rugg & M.G.H. Coles (Eds.), *Electrophysiology of mind* (pp. 171–215). Oxford, UK: Oxford University Press. A review of ERP studies on language comprehension.
-

REFERENCES

- Beaumont, J.G., Kenealy, P.M., & Rogers, M.J.C. (1996). *The Blackwell dictionary of neuropsychology*. Cambridge, MA: Blackwell.

- Block, N. (1990). The computer model of the mind. In D.N. Osherson & E.E. Smith (Eds.), *An invitation to cognitive science, Vol. 3: Thinking* (pp. 247–289). Cambridge, MA: MIT Press.
- Brown, C., & Hagoort, P. (1993). The processing nature of the N400: Evidence from masked priming. *Journal of Cognitive Neuroscience, 5*, 34–44.
- Brown, C.M., & Hagoort, P. (Eds.). (1999). *The neurocognition of language*. Oxford, UK: Oxford University Press.
- Coltheart, M. (2006). What has functional neuroimaging told us about the mind (so far)? *Cortex, 42*, 323–331.
- Cutler, A., & Clifton, C.E. (1999). Comprehending spoken language: A blueprint of the listener. In C.M. Brown & P. Hagoort (Eds.), *The neurocognition of language* (pp. 123–166). Oxford, UK: Oxford University Press.
- de Araujo, I.E., Rolls, E.T., Velazco, M.I., Margot, C., & Cayeux, I. (2005). Cognitive modulation of olfactory processing. *Neuron, 46*, 671–679.
- Fodor, J.D. (1974). Special sciences (or: The disunity of science as a working hypothesis). *Synthese, 28*, 97–115.
- Fodor, J.D. (1983). *The modularity of mind*. Cambridge, MA: MIT Press.
- Fodor, J.D. (1999). Let your brain alone. *London Review of Books, 21*. Retrieved February 21, 2008 from http://www.lrb.co.uk/v21/n19/fodo01_.html
- Garrod, S., & Pickering, M.J. (2004). Why is conversation so easy? *Trends in Cognitive Sciences, 8*, 8–11.
- Hagoort, P., & Brown, C. (1994). Brain responses to lexical ambiguity resolution and parsing. In C. Clifton Jr., L. Frazier, & K. Rayner (Eds.), *Perspectives on sentence processing* (pp. 45–81). Hillsdale, NJ: Erlbaum.
- Hagoort, P., Brown, C.M., & Groothusen, J. (1993). The Syntactic Positive Shift (SPS) as an ERP measure of syntactic processing. *Language and Cognitive Processes, 8*, 439–483.
- Hagoort, P., & van Berkum, J. (2007). Beyond the sentence given. *Philosophical Transactions of the Royal Society of London, B: Biological Sciences, 362*, 801–811.
- Kempen, Gerard. (2000). Could grammatical encoding and grammatical decoding be subserved by the same processing module? *Behavioral and Brain Sciences, 23*, 38–39.
- Kim, A., & Osterhout, L. (2005). The independence of combinatory semantic processing: Evidence from event-related potentials. *Journal of Memory and Language, 52*, 205–225.
- Kutas, M., & Hillyard, S.A. (1980). Reading senseless sentences: Brain potentials reflect semantic anomaly. *Science, 207*, 203–205.
- Mehler, J., Morton, J., & Jusczyk, P.W. (1984). On reducing language to biology. *Cognitive Neuropsychology, 1*, 83–116.
- Page, M.P. (2006). What can't functional neuroimaging tell the cognitive psychologist? *Cortex, 42*, 428–443.
- Tabor, W., & Tanenhaus, M.K. (1999). Dynamical models of sentence processing. *Cognitive Science, 23*, 491–515.
- Uttal, W.R. (2001). *The new phrenology: The limits of localizing cognitive processes in the brain*. Cambridge, MA: MIT Press.