



On the neurobiological investigation of language understanding in context

Steven L. Small^{a,*} and Howard C. Nusbaum^b

^a *Departments of Neurology, Radiology and Psychology and Committee on Computational Neuroscience, Brain Research Imaging Center, The University of Chicago, 5841 South Maryland Avenue, MC-2030, Chicago, IL 60637, USA*

^b *Department of Psychology and Committee on Computational Neuroscience, The University of Chicago, 5848 South University Avenue, Chicago, IL 60637, USA*

Accepted 12 August 2003

Abstract

There are two significant problems in using functional neuroimaging methods to study language. Improving the state of functional brain imaging will depend on understanding how the dependent measure of brain imaging differs from behavioral dependent measures (the “dependent measure problem”) and how the activation of the motor system may be confounded with non-motor aspects of processing in certain experimental designs (the “motor output problem”). To address these problems, it may be necessary to shift the focus of language research from the study of linguistic competence to the understanding of language use. This will require investigations of language processing in full multi-modal and environmental context, monitoring of natural behaviors, novel experimental design, and network-based analysis. Such a combined naturalistic approach could lead to tremendous new insights into language and the brain.

© 2003 Published by Elsevier Inc.

1. Introduction

How do we understand stories? How do we engage in conversation? How do we give or receive commands? These are all fundamental questions about language use, and the disciplines that investigate language, such as linguistics, psychology, anthropology, or neuroscience, would agree on their importance. However, these different disciplines would probably not agree how best to address these questions. Traditionally, investigators from different disciplines have approached the study of language processing with different hypotheses and research methods, motivated by equally disparate theories and models, and starting with very different assumptions about what constitutes the fundamental phenomena of interest.

The advent of noninvasive brain imaging has led to increasing attention to the neurobiological mechanisms underlying language processing, providing yet another set of theories and models to explain language process-

ing. Of course, an interest in neurobiological mechanisms does not in itself dictate agreement on how to investigate them. At the simplest level of consideration, we can view neurophysiology as providing a new dependent measure of language processing that can address extant theories from psychology and linguistics. However, the fundamental differences between neuroimaging and behavioral measures offer an opportunity to examine language processing in terms of its interaction with other kinds of psychological processes in tasks that start to more closely mirror the natural uses of language.

The landmark 19th century work of Broca (1861) and Wernicke (1874), has shaped much of our understanding of the way language and the brain are related. The association between anatomical locations of brain injury and disruption of particular language behaviors (e.g., production and comprehension) has provided an important functional definition of language processing (Benson, 1979; Geschwind, 1971). Similarly, the psycholinguistic study of linguistic behavior affords another way to provide a functional definition of language processing using the patterns of error rates and reaction times in carefully designed tasks. Instead of starting

* Corresponding author. Fax: 1-773-834-7610.

E-mail address: small@uchicago.edu (S.L. Small).

from the assumption that lesion-deficit pairings define the functional characteristics of language processing, psycholinguistics typically starts with the assumption that behavioral sensitivity to variation in some linguistic property (e.g., verb regularity) defines processing. For example, the theoretical division between expressive and receptive language processing derives in part from gross deficits seen in patients with damage located in more anterior or posterior cortical regions, and the research questions emerging from this division focus on characterizing the processing of those regions (e.g., agrammatism vs. working memory deficits for Broca's area). On the other hand, the example of a theoretical division between rule-based processing and statistical regularity emerged from differences in performance on specific lexical processing tasks (Pinker & Prince, 1988; Seidenberg & McClelland, 1989). Thus, in part, research methods provide the rose-colored glasses that can shape our view of language processing phenomena.

With the increasing use of neuroimaging measures, the methods of lesion analysis and psycholinguistic experimentation seem to have formed the conceptual foundation for the methodological toolbox of functional brain imaging. An assumption underlying both of these approaches is the componential reduction of language processing, with a focus on language competence—basic linguistic knowledge—rather than language performance (Chomsky, 1965; de Saussure, 1959). The original motivation for this theoretical distinction is that linguistic performance—what is really said and what is really understood—constitutes an actual behavior, and is therefore intertwined with the operation of cognitive and motor systems. Constraints that appear in these behaviors may reflect a number of cognitive and motor system limitations that collectively distort measurements of purely linguistic ability. Over the past 50 years, we have learned a great deal about many levels of language processing, from phonology to discourse, by using this approach. However, this approach may be limited when it comes to neuroimaging studies, imposing a different set of distortions on the kind of results we obtain.

Studying linguistic competence by definition abstracts language processing away from its grounding in behavior. However, by shifting to studying language use rather than linguistic competence, we may gain, rather than lose, in our ability to understand language processing (see (Clark, 1996) for a discussion) when using neuroimaging measures.

There can be no doubt that language evolved for communication between people, or that language evolved for multi-modal, face-to-face communication, and that language use occurs in a rich environmental context that can ground communication for cognitive purposes. Rather than start from the position of looking for evidence of specific types of language processing “in” the brain or looking for evidence of language

processing by “the brain”, we suggest that it may be useful to examine cortical activity during language behavior that most closely matches conditions of evolution: language use by people at a time and place, aiming to understand and to be understood, fulfilling a purpose. The utility of this approach is that it considers how language processing, in service of specific goals and uses, interacts with a broad set of neural circuits that are involved in more general cognitive, affective, and social processing.

By examining the distribution of such network activity during language use, we can begin to investigate the richness of the neural interactions that occur in real time integrating linguistic knowledge with putatively non-linguistic processes such as motor activity, working memory, or attention. There has been a tendency in neuroimaging research to try to isolate language processing from these other kinds of processes using a variety of analytic and design methods. However, it is important to remember that language use in the real world interacts fundamentally with motor behavior—all language expression *is* motor behavior—and the systems for language use and motor behavior are functionally intertwined, affecting our ability to investigate and ultimately to understand the neurobiology of language. Furthermore, real language use entails cognitive, sensory/motor, and affective operations in addition to linguistic ones. In order to study the biology of language use, understanding the relationships among these inter-related neural processes will be a central aspect of the basic scientific problem.

2. Componential processing models

A common feature of both lesion analysis and psycholinguistic research is the emphasis on functional decomposition, which views the brain as organized into anatomically segregated parts (Gall, 1825) and complex behavior as being mediated by a collection of functionally independent units (Fodor, 1983). Recent work in dynamical systems theory (Freeman & Barrie, 1994) suggests an alternative approach: rather than viewing different patterns of behavior as the result of the operation of different and independent subsystems each responsible for a different pattern, such patterns of behavior can arise from a single complex system operating in different modes at different parameter values. This has produced significant scientific breakthroughs, including in psychology (e.g., see Smith & Thelen, 1993).

Our argument against strict functional decomposition is not an argument in favor of the older holographic view of the brain as a mass of equipotential tissue (Lashley, 1950). We do not assume that all parts of the brain participate equally in all behaviors. Nor do we assume that each part of the brain provides an identifi-

ably unique and functionally separate process. Rather, we postulate that the neural circuits that operate within and across different anatomical regions, are both interdigitated and interactive, and operate differently depending on their dynamic patterns of activity. This intrinsic neural context (McIntosh, 2000) complements the extrinsic environmental context, producing different modes of processing in different circumstances, leading to unique patterns of behavior. The apparent specializations of different anatomical regions may not have clear psychological interpretations, which has been an underlying assumption of much neuroimaging work.

The scientific tension between decompositional reduction and more global behavioral analysis in psychology is certainly not new. For example, the reflex arc concept decomposed behavior into a system of three processes of sensation, classification, and response, that could, in principle, be separately investigated. However, Dewey (1896) argued that the separation of a behavior into these descriptive components was really for the convenience of the scientist and should not be taken as reflecting the underlying causal properties of the brain or mind. He pointed out that what constituted real sensation for an organism often depended on the response to be performed and thus the units often function interactively.

The information processing era of the cognitive revolution led to a plethora of serial componential “boxological” models of behavior (Neisser, 1976). For example, language comprehension has been studied as a series of processing stages that match the propositional encoding of a sentence against a propositional encoding of a picture (Clark, Carpenter, & Just, 1973). This decomposition provided the basis for important experimental manipulations to investigate subprocesses of sentence comprehension. These information processing models assumed, however, that each processing stage was independent of the others and was necessarily completed before starting the next (Sternberg, 1969).

This approach to cognitive research has continued through recent times. Just as Fodor (1983) viewed the mind as composed of modules, the neurosciences have viewed the brain as modular, consisting of functionally specialized and independent locations (e.g., Shallice, 1988). In the study of language, the frontal operculum (Broca, 1861) and the posterior superior temporal region (Wernicke, 1874) have played special roles in this localizational view, representing the sites for language production (early view) or syntax (later view) and language comprehension (early view) or semantics (later view), respectively. In part, these componential views are rooted in other studies of biological specialization. Just as the heart and the lungs are anatomically and mechanically specialized for specific distinct physiological functions (but operate together as integrated systems), anterior and posterior cortices have been viewed

as specialized for motor and sensory functions, replicating the notion of structure–function relationships found elsewhere in biology.

However, many systems are not decomposable into independent functional parts (Runeson, 1977), even though the standard operating assumption in psychology is to reduce systems to putative functional components. In psychological research, this componential view is critical to the interpretation of response-time experiments: in broad terms, these experiments generally: (a) assume that the duration of any particular cognitive process is composed of the sum of a set of constituent subprocesses (Donders, 1868/1969) and (b) these putative subprocesses provide the basis for the manipulation of experimental variables from which to infer the processing characteristics of component subsystems (Sternberg, 1969).

Neurology has also taken a componential (anatomical decomposition) approach to understanding the neural mechanisms that mediate complex behaviors. The inferential logic of “double dissociation” (Shallice, 1988) depends on the notion that there are component mechanisms that have independent functions. Damage to one component should produce patterns of behavior change that are different and complementary to the change produced by damage to a different component.

Ultimately, this conceptual framework is the basis for many studies in functional brain imaging with PET and fMRI. In research on language and the brain, some studies have focused on validating certain models derived from information-processing psychology, which themselves have often been derived from the analytic considerations of theoretical linguistics. Consider the example of lexical access, in which the process of recognizing a spoken word is viewed as isolable from the rest of the language processing system by comparing neural activity produced by: (1) repeating words with (2) hearing reverse speech and uttering a standard word (Howard et al., 1992). This elucidates brain regions for lexical access, based on the assumption that the two tasks contain all the same components except one (i.e., the access component), in the same order and with the same feedback (Sergent, Zuck, Lévesque, & MacDonald, 1992).

Neuroimaging studies often assume a one-to-one correspondence between neural (brain locations) components and psychological (behaviorally isolable) components. Typical tasks used to study language in the brain include, at different levels of language processing: rhyme judgment and phoneme discrimination (phonological level), lexical decision (lexical level), or grammaticality judgment (sentence level). To carry out any of these tasks, responses depend on the use of a specific kind of linguistic competence. For example, to judge that two words rhyme, the listener must compare the phonological patterns of the words, thereby exercising

phonological processing. (Of course this assumes that the nature of the phonological processing used in a metalinguistic rhyme judgment task depends on the same phonological competence used in fluent language use.) By designing tasks based on well-defined (in theoretic terms) specific areas of linguistic competence, it is assumed that the operation of a component mechanism that mediates that competence will be selectively illuminated. The success of this approach depends on the assumption that the explicit judgment of a linguistic property of an utterance exercises the same kind of processing (i.e., same mechanism used the same way) as the implicit routine use of this processing in daily language use.

A study conducted in our laboratory illustrates this concern and the nature of the problem. This study compared phoneme discrimination with nonspeech tone discrimination in a context in which the former required phonological segmentation and another where it did not (Burton, Small, & Blumstein, 2000). By contrasting two discrimination tasks (one phonological, one auditory)—both calling for stimulus comparison and planned motor behavior—we intended to isolate those neural processing components that mediate phonological segmentation. We concluded that “it is the process of segmentation of the initial consonant from the following vowel, probably requiring articulatory recoding, that appears to involve left ... inferior and middle frontal [gyri]” (Burton et al., 2000).

Of course, the contrasts that are carried out in these kinds of studies assume that we understand a priori the componential structure of the tasks we use. Do listeners *actually* segment the speech stream into phonemes before recognizing the phonemes or do listeners just recognize linguistic units without segmentation? Are phonemes truly the basic unit of speech perceptual analysis or are syllables or diphones or onset-rime structures the basic unit of perception? Although these are *standard* assumptions in much speech research, and may reflect consistency in information conveyed in speech (Studdert-Kennedy, 1981) this does not necessarily license a neural reality for these assumptions. If tone discrimination and phoneme discrimination are carried out by complex neural networks that are simply modulated differently across conditions, the isolable anatomical components may have little or no relationship to the behavioral components, if there really are any (cf. Runeson, 1977). It is important to remember at this point Dewey’s (1896) cautionary note that the division of behavior into stages is for the analytic convenience of the scientist but may not reflect the psychological (or neuroanatomical) reality.

Indeed, it turns out that the conclusions of our first study depended critically on the specific nature of the task comparison, as we later learned: a follow-up study, using a different nonspeech tone discrimination control

task (requiring pattern segmentation, similar to the meta-phonological judgment made with syllables) found no frontal activation (Burton & Small, 2001) because this component was “subtracted off” when the more comparable speech–nonspeech comparisons were carried out.

Holding aside for the moment that listeners never need to make explicit phonological discriminations during real conversations (thus making discrimination a very unnatural task), the presence or absence of apparent frontal activity in this study depends on the comparison task that is used for subtraction, as should be the case. However this leaves us with a very real question: Which result is more indicative of real phonological perception, the involvement or non-involvement of the frontal lobe? If one nonspeech control task emphasizes working memory and the motor system more than another, this will moderate the appearance of neural activity in the frontal region during the phonological discrimination task. Since we can modulate this involvement easily with the control task, how can we ascertain the “correct” degree of match between control and target experimental tasks? The only possible way to make this decision is by an a priori theoretic assumption, which may be of questionable validity.

3. Inadvertent study of language/motor integration

Studies such as the phonological segmentation experiment are intended to investigate the independent components of a complex behavior as if the parts can be inserted or removed without changing *ceteris paribus* the functioning of the other components (Donders, 1868/1969). Since most experiments are designed with explicit decision-making components and overt motor responses, and these aspects of processing are not the focus of the scientific investigation, the contribution of these components to the dependent measures of brain activity must be eliminated. This requires that decision-making and button-pressing must be treated as (or at least assumed to be) independent and isolable from the cognitive and linguistic processes of interest in both behavioral terms and in the brain. In general, this has been a productive strategy for understanding some of the basic aspects of linguistic competence and cognitive functioning. However, to understand language use, rather than competence, it is important to understand the interactions that occur between language processes and cognitive, affective, and motor systems. With this research goal, it is likely that the assumptions regarding component isolability may be problematic, and that matched-task subtractions could mask or eliminate activity from brain regions of interest. Thus, applying the common experimental method for functional brain imaging to the study of language use may involve the in-

advertent study of language/motor integration in task-dependent (as suggested with the example of phonological segmentation) rather than language-use-dependent ways.

Consider the commonly studied rhyme judgment task as another example: in this task, a participant sees or hears two words, decides if they rhyme, and then makes a forced-choice button press response. Although this does involve reading or hearing words, the goal of processing the words is to carry out a rhyme decision, not to understand the words. While it seems likely that some aspects of understanding may be inadvertently involved, the processing focus is on the pattern properties of the words. This kind of focus has been demonstrated to skew the nature of processing compared to other kinds of more semantic decisions (McDermott, Petersen, Watson, & Ojemann, 2003). However, we do not mean to advocate one form of skewing processing over another—semantic and phonological metalinguistic decisions are still artificial compared to the psychological acts involved in language use. Neither task directs the participant in an experiment towards the goals of comprehension, production, or most other acts of human language use. A rhyme judgment task is intended to evaluate the processing characteristics of a particular language subcomponent—phonology—and this task may be useful in psycholinguistic experiments for understanding the way phonological information is accessed during word perception. However, this kind of task may have unintended effects when used in brain imaging studies. To understand the difference between the psycholinguistic experiment and its transplanted form in a brain imaging experiment, there are two things to be considered: first, how do dependent measures differ in brain imaging and psycholinguistic experiments, and second, what is the role of decision-making and motor output in producing the imaging result?

The dependent measure in fMRI brain imaging—hemodynamic response—is fundamentally and critically different from the dependent measures—response time or accuracy—in psycholinguistic investigations. Behavioral measures, such as response time or accuracy, typically give us a relatively univariate view of language processing, only providing a measure at the outcome of the overall process. In essence, this compresses a complicated network of neural computation into a single behavioral output. By contrast, neuroimaging gives us a multivariate data set reflecting all of the activity in this network over time. Every subprocess can manifest itself relatively simultaneously (depending on temporal sensitivity) and in parallel across the brain. In the behavioral measure, experimental manipulations of specific variables can modulate the mean difference across conditions such that the contribution of some subprocesses is swamped by the variance due to the “manipulated” subprocesses of interest. However, in a neuroimaging

study, the manipulated target subprocesses and the ancillary subprocesses are all manifest distributed across the dependent measure. We call this difference between behavioral and neurophysiological measurements the “dependent measure problem”.

In the laboratory, putative cognitive components are not really isolable, but given their overall characterization by a single univariate measure (e.g., reaction time), simple assumptions about the components’ respective contributions to overall processing and a limited set of conclusions can simplify interpretation. In these studies, experimental tasks are specifically engineered to produce patterns of results that emphasize processing variation within a single subcomponent of the overall system. The measured variation due to the independent variable has to exceed the random variation in all the other subcomponents (e.g., see Sternberg, 1969).

By contrast, brain imaging offers the opportunity to observe all the components operating in parallel, overlapping and distributed in time. However, unlike response time or error rate, the dependent measure reflects aggregate system behavior in a very different way. It is important to note that in neuroimaging the dependent measures are themselves directly linked to the system components of interest—*anatomy*. Variation in one dependent measure is no longer a reflection of the entire chain of processing in a task; rather the dependent measure can reflect the contribution of any one anatomical component to the task, as well as the modulation of that component by linked components. However, the relatively slow (in relation to mental time) changes of some neuroimaging measures could compress successive moments of processing into a single anatomical location. On one hand, the association between anatomically defined dependent measures and functionally defined processing components provides one of the incredible strengths of neuroimaging research. On the other hand, the lack of strong neurophysiological theories of psychological states, processes, and behaviors makes it difficult to separate out the contributions to any particular measure that result directly from any one event, from associations across events, or from multiple events occurring over the (low) time resolution of the method. As a result, the incredible strength of neuroimaging comes at a certain cost: it is not straightforward to use multiple control conditions to compare behaviors of interest along a single dimension.

A corollary issue then is that the decompositional or subtractive approach to imaging can lead to the inadvertent study of language/motor integration. We call this the “motor output problem”. It is obvious that virtually all measurable behavior involves the motor system. A central tenet of most neuroimaging studies has been to use measurable behavioral outputs (e.g., rhyme decision button presses) to establish that the brain activity being measured corresponds to the in-

tended (by the experimenter) processing that is being investigated. In other words, if listeners are making accurate rhyme decisions, they must be using phonological processing. The rhyme-based button pressing behavior itself is not the processing of interest in these studies. However, due to the dependent measure problem, without appropriate treatment, the cortical activity underlying the button-pressing behavior will show up in the dependent measures of putative phonological processing.

This has meant that for the results of imaging studies to be interpretable, it is necessary to assume that motor planning and control are independent of the cognitive process under investigation. This assumption would allow the motor activity to be subtracted off using appropriately matched control conditions. Yet this assumption seems questionable—we know that complex motor circuits interact with many other networks throughout the brain. In fact, the areas of the brain that have been associated with language (Broca, 1861; Burton et al., 2000; Zatorre, Meyer, Gjedde, & Evans, 1996), emotional experience (Lane, Reiman, Ahern, Schwartz, & Davidson, 1997), attentional control (e.g., Banich et al., 2000), and working memory (Cohen et al., 1997; Smith, Jonides, Marshuetz, & Koeppe, 1998) are also closely identified with motor processing. For example, it is known that much of the anterior cingulate gyrus, an area frequently implicated in attention mechanisms (Smith & Jonides, 1999), plays an integral role in motor processes (Grafton, Hazeltine, & Ivry, 1998; Morecraft & van Hoesen, 1998; Picard & Strick, 1996). If a motor task is imposed on a neuroimaging experiment to guarantee that the brain activity reflects the intended psychological processing, a significant degree of the dependent measure will reflect the motor system activity produced by aspects of the task that may be irrelevant to the psychological process under investigation. This activity may not be easily (if at all) dissociable from the cognitive process under investigation. Perhaps it should not be, but perhaps instead of being skewed to reflect highly artificial processing goals (e.g., rhyme judgment), it should be focused on more ecologically relevant goals such as motor behavior that is consistent with the psychological process under investigation.

Since all measurable behavior inherently depends on the motor system, and since the study of brain/behavior relationships requires careful assessment of both brain function and behavioral performance, it seems impossible to avoid the study of the motor system in every investigation of language and the brain. To interpret functional imaging data, the nature of the processing carried out during image acquisition must be carefully determined. The most common way to do this currently without a concurrently imposed task is to ask participants a series of questions after the experiment to assess compliance with the tasks. This approach has been suc-

cessfully used in several language comprehension experiments (Mazoyer et al., 1993; Schlosser, Aoyagi, Fulbright, Gore, & McCarthy, 1998; Tettamanti et al., in press). Of course, since these questions are answered after the processing has taken place, the answers may be contaminated by introspection and retrospective processes.

Clearly it would be important to monitor psychological processing during image acquisition rather than to try to assess it after the fact. Since it is not possible to inspect mental behavior directly, and since all observable behavior is motor, the only viable solution to the real time monitoring of psychological processing is to measure behaviors that do not interact with the language task under investigation or at least are consistent with more ecologically valid language use. One way to do this is to observe naturally occurring language behavior such as vocal responses to utterances, as in conversation, or eye movements that result from imperatives or requests regarding a visual display (e.g., Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995). The difference between this kind of motor activity and less ecologically valid activity (e.g., metalinguistic button pressing) is that the interactions that occur with ecologically valid motor activity may reflect typical processing interactions. For example, when eye movements are tracked in a real-time language understanding task, there is a very different pattern of processing—integration of diverse sources of knowledge—compared to a linguistic judgment task (Eberhard, Spivey-Knowlton, Sedivy, & Tanenhaus, 1995). Another way to approach this problem of measuring ongoing psychological processing is to record other naturally occurring physiological responses, such as sweat, pupillary diameter, and electromyographic responses, from which some aspects of processing (e.g., arousal or attention) can be inferred.

4. Advertent study of language/motor integration

In studying language use rather than component linguistic competencies, it may be possible to avoid or at least moderate both the dependent measure problem and the motor output problem. Rather than impose artificial metalinguistic probe tasks on participants, it is possible to use more ecologically plausible language tasks, such as conversation, comprehension, or instruction following. Brain activation patterns during such tasks might be particularly revealing, since these tasks are likely to have played a role in the ontogeny and phylogeny of brain development. These kinds of ecologically valid language processing tasks, in contrast with meta-linguistic judgment tasks, may be more closely suited to the nature of the dependent measure of brain imaging.

Brain imaging studies of ecological language processing in multi-modal naturalistic context might be a

valuable way to avoid the problems associated with componential modeling assumptions and decision-making tasks. This is not new to psychology or neurology. In fact, the “Chicago School” of psychology emphasized the study of cognitive processing in context, the interactivity of the component parts, and the investigation of naturalistic phenomena (Dewey, 1896; James, 1904). Further, Brunswik (1947) argued that psychological research should contrast conditions that display the full range of natural variation observed in behavior. While true ecologically valid language behavior is difficult under the conditions of neuroimaging, particularly with fMRI, it is possible to move studies more in that direction, both by changing the nature of the tasks used and by changing the kind of information provided to our participants.

Language evolved in the context of face-to-face communication, not in the context of telephone conversation. So perhaps it should not be surprising that visual information showing movements of the mouth and lips during talking enhances speech comprehension, even though we often think of speech perception as being infallible based on the acoustic signal alone (Sumbly & Pollack, 1954; Summerfield, 1992). Furthermore, other visual information about motor movements produced by an interlocutor while speaking are important to communication, such as information about the manual gestures that accompany speech, which clearly affect our understanding of that speech (McNeill, 1992). In addition, we have recently shown that manual gesturing while speaking improves cognitive efficiency as measured by memory capacity (Goldin-Meadow, Nusbaum, Kelly, & Wagner, 2001), suggesting an interaction between the language system and the motor system for cognitive functions.

In trying to understand why face-to-face language comprehension is easier than audio-only (e.g., telephone) language comprehension, brain imaging reveals a possible explanation, rooted in these interactions with the motor system. We tested the prediction that perception of the visual information from the oral–facial gestures that accompany speech during face-to-face conversation affects perceptual processing through associated motor system activity. Subjects were imaged with fMRI while listening to interesting stories (audio only), listening to stories while seeing the storyteller (audiovisual), or just seeing the storyteller (visual). We found far more activation in the inferior frontal cortex (BA 44/45) in the audiovisual condition than in either other condition (Skipper, Nusbaum, & Small, 2002; Skipper, Nusbaum, & Small, submitted for publication). Moreover, the presence of the visuo-motor information changed the laterality of the activity in superior temporal cortex, demonstrating the interaction in processing between face information and acoustic speech in more traditional speech perception areas.

It is important to note that listeners were required only to understand the spoken stories in this study, and not to perform any adjunctive metalinguistic task. If we had designed a specific judgment task to measure comprehension, the motor behavior in responding and the working memory used during judgment could have masked the Broca’s area activity observed during comprehension. However, the limitation of this approach is that without specific behavioral measures of comprehension processing, we cannot directly relate the patterns of cortical activity to the details of behaviors. While post-task questioning can establish gross aspects of processing, such as whether listeners understood the stories and some of what they remember, these measures are not sufficiently sensitive to diagnose more specific hypotheses. One challenge then is to develop new methods that allow us to assess more directly the relationships between brain activity and behavior without changing either. We can think of this as a kind of Heisenberg Uncertainty Principle in cognitive neuroimaging research.

5. Ecological brain imaging

Performing ecological functional brain imaging of language processing will require several advances in experimental design and/or analysis methods. As we have suggested, experimental design should be tailored to focus on real-world functions of language, in (relatively) natural contexts of presentation or behavior. This represents part of the challenge of this approach given the decidedly unnatural setting of an MRI scanner. Ideally, research designs should avoid imposing decision-making processes, such as meta-linguistic judgments, as well as motor planning and execution that are not part of the natural language behavior under investigation. All tasks that result in measurable behavior will necessitate motor system activity, attentional processing, and probably working memory loads. Tasks should not impose additional extrinsic cognitive demands on the participants that could mask language-use-relevant motor and cognitive cortical activity. It would be preferable to have the kind of motor and cognitive activity ecologically consistent with the kind of language use being investigated (e.g., vocal responses in a conversational setting, eye movements in response to questions or imperatives).

Furthermore, experimental design and data analysis should permit the interpretation of linguistic processing at different levels of representation simultaneously, e.g., phonological or lexical processing, within the full context of language use. From this perspective, it may be better to examine the phonological activity within the context of discourse comprehension than to attempt to artificially isolate phonological activity and in doing so,

distort the kind of processing that is taking place. Experimental design for language imaging in communicative and naturalistic contexts in principle should be the presentation of full discourse, rather than isolated phonemes, syllables, words, or even sentences. Ideally, this presentation involves audiovisual stimuli, rather than auditory-alone stimuli, and the goals of the listener should be defined in some meaningful social context. The lips, mouth, and hands of the speaker should be visible, and the prosody should be natural. Several stimulus design properties are less plausible than others, and of course, the environment of brain imaging (e.g., loud noise, constrained physical space, lack of dialogue) does constrain some of the ways in which language use may be studied. Yet some of the idealized goals are achievable and are highly desirable. For example, to study lexical processing, it would be important to focus the experimental design on the type of lexical processing that might actually occur during discourse comprehension or conversation and embed this usage in a task defined with more naturalistic communicative goals for the participant. It is then beholden upon researchers to develop strategies for data analysis capable of testing the specific questions of interest given the increased situational and stimulus variability. Clearly there are many ways in which an experiment can move closer to or farther from the idealized form of ecological language use. Depending on the specific research questions, a realistic design will likely reflect the kind of compromises as reflected in the Uncertainty Principle.

Given that the experimental methods are predicated on the idea that diverse neural networks will be interacting across different tasks or conditions, the nature of the data analyses must be sensitive to measuring these interactions. Rather than emphasize analyses that localize activity to specific cortical regions, data analysis for these imaging studies should examine the distribution of cortical activity across the complex neural networks involved in processing. It is almost certainly the case that localized regions of the brain perform different kinds of functions depending on their "neural context" (McIntosh, 2000), and can thus be best understood in the framework of regional connectivity and correlation of activity (with or without anatomical constraints and directionality) (Friston, Phillips, Chawla, & Buchel, 2000; McIntosh, 1999). Data analysis should be designed to illuminate the interconnectivity of different cortical areas and the modulation of activity across these areas in different conditions.

In this respect, analyses need to be sensitive to the effect of context on cortical activity. For example, in one recent fMRI study, we contrasted comprehension of sentences in a coherent discourse context with similar, matched sentences presented as an unstructured list (clearly an unnatural stimulus). To analyze these data, we used a hybrid of a block and event-related design

(Small, Uftring, & Nusbaum, 2003; Small, Uftring, & Nusbaum, 2002). Each story was analyzed both as a block (of sentences) and as an event (a single story). For this study, we used a standardized discourse structure (Trabasso & Suh, 1993), in which story protagonists set particular goals and subgoals, perform actions, and ultimately achieve or fail to achieve the goals (i.e., outcomes) (Table 1). As with any experiment, it was necessary to compromise some aspects of ecological validity (e.g., the presentation of the sentences was separated by short but unnatural intervals), but the choice of these compromises was made in consideration of emphasizing the mechanisms under investigation (e.g., some aspects of discourse coherence are achieved using working memory over such durations even in natural discourse).

Data were analyzed at both levels of interest: at the block level, we compared comprehension of stories with comprehension of unordered matched sentences. This addresses questions concerning the difference in brain activity for understanding stories vs. the sentences that compose those stories without narrative coherence. At the event level, we compared the goal-setting sentences that follow goal successes or failures (that is, sentences with a specific discourse role in the structure of narrative events) with temporally and structurally matched sentences from unordered lists of sentences. This examines how the contextually defined role of the specific sentences changes the processing of these constituent elements.

Considering both levels of analysis provides an interesting view of the process of discourse comprehension. The block-level analysis (Bandettini, Jesmanowicz, Wong, & Hyde, 1993; Levin & Uftring, 2001) showed the overall differences between listening to well-formed discourse and to (incoherent) sets of sentences. This analysis demonstrated activation in stories to be greater than for non-story sentences in the precuneus, left posterior superior temporal gyrus and angular gyrus (AG),

Table 1
Story structure

Setting
Event 1
Goal 1
Action 1
Outcome 1: Goal 1 Success or Failure
Reaction 1
Event 2
Goal 2
Action 2a
Action 2b
Outcome 2: Goal 2 Success or Failure
Action 3a
Action 3b
Outcome 3: Goal 3 Success

and the right premotor regions, right temporal pole, and the hippocampal formation bilaterally. Thus, there is something about the information that transcends individual sentences that results in this pattern of activity in comprehending stories. The event-level analysis (Ward, 2001) starts to illuminate some of the key aspects of discourse comprehension that turn on specific sentence roles in the structure of a story. This analysis showed the differences between the “goal response” sentences in a story and comparable sentences in the non-story blocks. Activation following failed goals was greater in both superior temporal gyri, left AG, left cerebellum, and limbic areas. Activation following successful goals was greater in both angular gyri, left superior temporal sulcus, and the right medial frontal region (Small et al., 2003). We cannot understand discourse processing simply by looking at how stories differ overall from sentences. However combining analyses across levels of processing can provide a clearer picture of this processing.

6. Network analysis methods

Although it would simplify matters tremendously if brain regions and behavioral functions mapped onto each other in a one-to-one fashion, this is unfortunately not the case. In fact, this relationship appears not only to be a many-to-many mapping, but to have dynamic properties as well, i.e., the mapping changes depending on a wide variety of environmental and intrinsic factors (Freeman & Barrie, 1994). These correspond to what Claude Bernard referred to as “milieu exterieur” and “milieu interieur” (Bernard, 1865) or what has been referred to here as real-world context (Small, 1987) and neuronal context (McIntosh, 2000). Therefore, although there may be some value in associating specific brain areas as important or even critical for particular functions, understanding how the processing within brain areas changes over different contexts may provide a deeper understanding of brain/behavior relationships. It is therefore important to be able to characterize the brain networks that participate in any particular psychological process and to examine how these networks change with different goals, expectations, and context.

The easiest type of “network” analysis is simply to examine correlations among activations in different regions and to examine how these correlations change across different tasks, either directly, or following an eigenvector transformation (Bullmore et al., 1996). These correlations indicate the degree to which processing changes in a similar way across cortical areas, independent of anatomical evidence of connections. However, a more advanced method takes into account what is known about the underlying anatomy of the system, such that relationships are only inferred between

regions that are actually known to have some physiological relationship. Structural equation modeling (Buchschafer & Friston, 1997; McIntosh et al., 1994) has been used successfully to delineate effective connectivity changes across different tasks in a variety of domains, including language (Peterson, Reis, Askelof, Castro-Caldas, & Ingvar, 2000).

In an elaboration on our study of audiovisual and audio-only language comprehension described above, we performed an analysis of the relationships among several of the participating regions to examine differences in the organization of the functional networks in

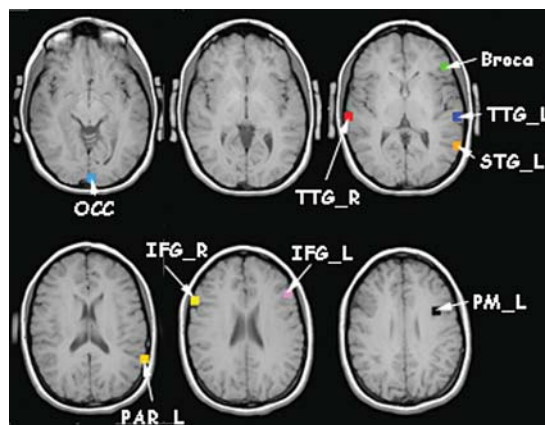


Fig. 1. Location of voxels used for time series correlations in principal components analysis and structural equation modeling.

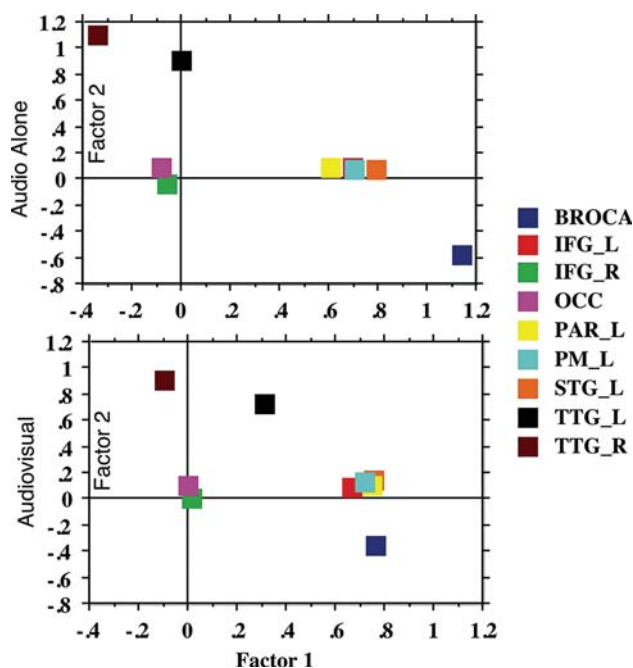


Fig. 2. Principal components analysis of activation time series from nine voxel locations for two conditions. First two principal components are shown for audiovisual and audio conditions.

these two conditions of speech understanding. We based the analysis on waveforms (vectors) from single voxels in a small number of relevant regions (see Fig. 1). A principal-components analysis showed several interesting features of the two-dimensional activation space defined by the first two eigenvectors (Fig. 2). For both conditions, there seemed to be four general clusters of regions in this space, including the left hemispheric language areas, a visual axis, an auditory axis, and the left frontal operculum. Of particular interest in this analysis are that in the audiovisual condition compared to the audio condition, both the left transverse temporal region and the left frontal opercular region are closer to the language areas. These pilot data suggest that the left auditory region and the left frontal operculum change the nature of the processing they carry out during language comprehension when the face and lips of the speaker can be perceived than when they cannot.

An extension of correlational approaches such as principal components analysis uses known anatomy to augment the functional information with structural connectivity information. Such structural equation modeling can be used to create models of both static and dynamic relationships (Horwitz, Tagamets, & McIntosh, 1999; McIntosh et al., 1994). We are currently working to use such network-based analyses in the study of language processing. The critical aspects of this work are to determine the (functionally relevant) anatomical pathways in the human brain and their relative strengths. Until in vivo human studies are possible, the data for such models necessarily comes from primates, who do not use language, and require inferences about analogous human pathways, their directionality, and their quantitative strengths. This is an important undertaking, but one requiring significant future work.

7. Summary and conclusions

Functional brain imaging provides a fundamentally new and different approach to studying language processing. Understanding the nature of this method and how it differs from previous approaches are critical to taking advantage of the strengths that neuroimaging provides. In part, this depends on understanding both the dependent measure problem and the motor output problem. In particular, in some cases, brain imaging experiments designed to isolate and examine specific subcomponents of language competence may be confounded by inadvertent language/motor interactions, since these experiments depend on complex metalinguistic decision-making tasks that require explicit motor responses. These experiments use metalinguistic tasks such as rhyme judgment, phoneme discrimination, lexical decision, or grammaticality judgment in order to focus on specific aspects of linguistic competence. The

interaction of decision-making processes and response-generation processes with linguistic processing may mask, distort, or insert (depending on the design) significant motor preparation and execution associated with language processing. This poses a problem for investigating the brain networks active during language use wherein we expect activity in motor and cognitive systems outside of linguistic processes. If the goal is to understand the richness of interaction among brain circuits, imposing specific metalinguistic judgments may distort the image of the brain processing during natural communication.

However, by shifting the focus of research questions to understanding language use, brain imaging allows us to investigate neural mechanisms that are responsive to a multi-modal and environmental contextual information to understand the richness of interactive neural activity during real language behavior. This approach will depend on the analysis of activation across network structures rather than in specific localized regions. This presents substantial new challenges for experimental design and image processing methods, but we believe that a hierarchical event-related design might provide the needed tools. This combination of context-dependent naturalistic imaging with monitoring of natural behaviors, novel experimental design, and network-based analysis could lead to tremendous new insights into language and the brain.

Acknowledgments

The support of the National Institutes of Health under grant DC-3378 is gratefully acknowledged. Additional support from the Brain Research Foundation and the McCormick Tribune Foundation is also acknowledged. We would like to thank Ana Solodkin and Jeremy Skipper for helpful discussions about these topics. Finally, we would like to thank Elizabeth Bates for many conversations over the past 10 years about the strengths and weaknesses of brain imaging for the study of human language.

References

- Bandettini, P. A., Jesmanowicz, A., Wong, E. C., & Hyde, J. S. (1993). Processing strategies for time-course data sets in functional MRI of the human brain. *Magnetic Resonance in Medicine*, 30, 161–173.
- Banich, M. T., Milham, M. P., Atchley, R., Cohen, N. J., Webb, A., Wszalek, T., Kramer, A. F., Liang, Z. P., Wright, A., Shenker, J., & Magin, R. (2000). fMRI studies of stroop tasks reveal unique roles of anterior and posterior brain systems in attentional selection. *Journal of Cognitive Neuroscience*, 12(6), 988–1000.
- Benson, D. F. (1979). *Aphasia, alexia, and agraphia*. New York: Churchill Livingstone.
- Bernard, C. (1865). *Introduction à l'étude de la médecine expérimentale*. Paris.

- Broca, P. P. (1861). Nouvelle observation d'aphémie produite par une lésion de la partie postérieure des deuxième et troisième circonvolutions frontales. *Bull Soc. Anat. Paris*, 6, 398–407.
- Brunswick, E. (1947). *Systematic and representative design of psychological experiments. With results in physical and social perception*. Berkeley: University of California Press.
- Buchel, C., & Friston, K. J. (1997). Modulation of connectivity in visual pathways by attention: cortical interactions evaluated with structural equation modelling and fMRI. *Cerebral Cortex*, 7(8), 768–778.
- Bullmore, E. T., Rabe-Hesketh, S., Morris, R. G., Williams, S. C., Gregory, L., Gray, J. A., & Brammer, M. J. (1996). Functional magnetic resonance image analysis of a large-scale neurocognitive network. *Neuroimage*, 4(1), 16–33.
- Burton, M. W., & Small, S. L. (2001). Functional neuroanatomy of segmentation of speech and nonspeech. *Neuroimage*, 13(6), S511.
- Burton, M. W., Small, S. L., & Blumstein, S. E. (2000). The role of segmentation in phonological processing: an fMRI investigation. *Journal of Cognitive Neuroscience*, 12(4), 679–690.
- Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, Massachusetts: The MIT Press.
- Clark, H. H. (1996). *Using language*. Cambridge: Cambridge University Press.
- Clark, H. H., Carpenter, P. A., & Just, M. A. (1973). On the meeting of semantics and perception. In W. G. Chase (Ed.), *Visual information processing*. New York: Academic Press.
- Cohen, J. D., Perlstein, W. M., Braver, T. S., Nystrom, L. E., Noll, D. C., Jonides, J., & Smith, E. E. (1997). Temporal dynamics of brain activation during a working memory task. *Nature*, 386(6625), 604–608.
- de Saussure, F. (1959). *Course in General Linguistics* (W. Baskin, Trans). New York: The Philosophical Library.
- Dewey, J. (1896). The reflex arc concept in psychology. *Psychological Review*, 3, 357–370.
- Donders, F. C. (1868/1969). On the speed of mental processes. *Acta Psychologica*, 30, 412–431.
- Eberhard, K. M., Spivey-Knowlton, M. J., Sedivy, J. C., & Tanenhaus, M. K. (1995). Eye-movements as a window into real-time spoken language comprehension in natural contexts. *Journal of Psycholinguistic Research*, 24(6), 409–436.
- Fodor, J. A. (1983). *Modularity of mind: an essay in faculty psychology*. Cambridge, Massachusetts: MIT Press.
- Freeman, W., & Barrie, J. (1994). Chaotic oscillations and the genesis of meaning in cerebral cortex. In G. Buzsaki, R. Llinas, W. Singer, A. Berthoz, & T. Christen (Eds.), *Temporal coding in the brain* (pp. 13–38). Berlin: Springer.
- Friston, K., Phillips, J., Chawla, D., & Buchel, C. (2000). Nonlinear PCA: characterizing interactions between modes of brain activity. *Philosophical Transactions of the Royal Society of London Series B—Biological Sciences*, 355(1393), 135–146.
- Gall, F. J. (1825). *Sur les fonctions du cerveau et sur celles de chacune de ses parties*. Paris: Ballière.
- Geschwind, N. (1971). Current concepts: aphasia. *New England Journal of Medicine*, 284(12), 654–656.
- Goldin-Meadow, S., Nusbaum, H., Kelly, S. D., & Wagner, S. (2001). Explaining math: gesturing lightens the load. *Psychological Science*, 12(6), 516–522.
- Grafton, S. T., Hazeltine, E., & Ivry, R. B. (1998). Abstract and effector-specific representations of motor sequences identified with PET. *Journal of Neuroscience*, 18(22), 9420–9428.
- Horwitz, B., Tagamets, M. A., & McIntosh, A. R. (1999). Neural modeling, functional brain imaging, and cognition. *Trends in Cognitive Sciences*, 3(3), 91–98.
- Howard, D., Patterson, K., Wise, R., Brown, W. D., Friston, K., Weiller, C., & Frackowiak, R. (1992). The cortical localization of the lexicons. Positron emission tomography evidence. *Brain*, 115(Pt 6), 1769–1782.
- James, W. (1904). The Chicago School. *Psychological Bulletin*, 1(1), 1–5.
- Lane, R. D., Reiman, E. M., Ahern, G. L., Schwartz, G. E., & Davidson, R. J. (1997). Neuroanatomical correlates of happiness, sadness, and disgust. *American Journal of Psychiatry*, 154(7), 926–933.
- Lashley, K. S. (1950). In search of the engram. In *Symposia for the society for experimental biology*. Number 4, Cambridge, England: Cambridge University Press.
- Levin, D. N., & Uffring, S. J. (2001). Detecting brain activation in fMRI data without prior knowledge of mental event timing. *Neuroimage*, 13(1), 153–160.
- Mazoyer, B. M., Tzourio, N., Frak, V., Syrota, A., Murayama, N., Levrier, O., Salamon, G., Dehaene, S., Cohen, L., & Mehler, J. (1993). The cortical representation of speech. *Journal of Cognitive Neuroscience*, 5(4), 467–479.
- McDermott, K. B., Petersen, S. E., Watson, J. M., & Ojemann, J. G. (2003). A procedure for identifying regions preferentially activated by attention to semantic and phonological relations using functional magnetic resonance imaging. *Neuropsychologia*, 41(3), 293–303.
- McIntosh, A. R. (1999). Mapping cognition to the brain through neural interactions. *Memory*, 7(5–6), 523–548.
- McIntosh, A. R. (2000). Towards a network theory of cognition. *Neural Networks*, 13(8–9), 861–870.
- McIntosh, A. R., Grady, C. L., Ungerleider, L. G., Haxby, J. V., Rapoport, S. I., & Horwitz, B. (1994). Network analysis of cortical visual pathways mapped with PET. *Journal of Neuroscience*, 14(2), 655–666.
- McNeill, D. (1992). *Hand and mind: what gestures reveal about thought*. Chicago: University of Chicago Press.
- Morecraft, R. J., & van Hoesen, G. W. (1998). Convergence of limbic input to the cingulate motor cortex in the rhesus monkey. *Brain Research Bulletin*, 45, 209–232.
- Neisser, U. (1976). *Cognition and reality*. San Francisco, California: W.H. Freeman and Company.
- Petersson, K. M., Reis, A., Askelof, S., Castro-Caldas, A., & Ingvar, M. (2000). Language processing modulated by literacy: a network analysis of verbal repetition in literate and illiterate subjects. *Journal of Cognitive Neuroscience*, 12(3), 364–382.
- Picard, N., & Strick, P. L. (1996). Motor areas of the medial wall: a review of their location and functional activation. *Cerebral Cortex*, 6(3), 342–353.
- Pinker, S., & Prince, A. (1988). On language and connectionism: analysis of a parallel distributed processing model of language production. In S. Pinker & J. Mehler (Eds.), *Connections and symbols* (p. 255). Cambridge, Massachusetts: The MIT Press.
- Runeson, S. (1977). On the possibility of smart perceptual mechanisms. *Scandinavian Journal of Psychology*, 18(3), 172–179.
- Schlosser, M. J., Aoyagi, N., Fulbright, R. K., Gore, J. C., & McCarthy, G. (1998). Functional MRI studies of auditory comprehension. *Human Brain Mapping*, 6(1), 1–13.
- Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review*, 96(4), 523–568.
- Sergent, J., Zuck, E., Lévesque, M., & MacDonald, B. (1992). Positron emission tomography study of letter and object processing: empirical findings and methodological considerations. *Cerebral Cortex*, 2(1), 68–80.
- Shallice, T. (1988). *From neuropsychology to mental structure*. Cambridge: Cambridge University Press.
- Skipper, J. I., Nusbaum, H. C., & Small, S. L. (2002). Speech perception and the inferior frontal neural system for motor imitation. *Journal of Cognitive Neuroscience*, 0, F103.
- Skipper, J. I., Nusbaum, H. C., & Small, S. L. (submitted). Listening to talking faces: motor cortical activation during speech perception.

- Small, S. L. (1987). Parsing, word expert. In S. Shapiro (Ed.), *Encyclopedia of artificial intelligence*. New York: John Wiley and Sons.
- Small, S. L., Uftring, S. J., & Nusbaum, H. (2003). A hierarchical design for contextual imaging of language comprehension [Abstract]. *Neuroimage*, 15(6), CD-ROM.
- Small, S. L., Uftring, S. J., & Nusbaum, H. C. (2002). Naturalistic language imaging: hierarchical event analysis [Abstract]. *Journal of Cognitive Neuroscience*, F94.
- Smith, E. E., & Jonides, J. (1999). Storage and executive processes in the frontal lobes. *Science*, 283(5408), 1657–1661.
- Smith, E. E., Jonides, J., Marshuetz, C., & Koeppel, R. A. (1998). Components of verbal working memory: evidence from neuroimaging. *Proceedings of the National Academy of Sciences of the USA*, 95(3), 876–882.
- Smith, L. B. & Thelen, E. (Eds.). (1993). *A dynamic systems approach to development*. Cambridge, MA: MIT Press.
- Sternberg, S. (1969). The discovery of processing stages: extensions of the donders' method. *Acta Psychologica*, 30, 276–315.
- Studdert-Kennedy, M. (1981). The emergence of phonetic structure. *Cognition*, 10(1-sup-3), 301–306.
- Sumby, W. H., & Pollack, I. (1954). Visual contribution of speech intelligibility in noise. *The Journal of the Acoustical Society of America*, 26(2), 212–215.
- Summerfield, Q. (1992). Lipreading and audio–visual speech perception. *Philosophical Transactions of the Royal Society of London Series B—Biological Sciences*, 335(1273), 71–78.
- Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, 268(5217), 1632–1634.
- Tettamanti, M., Buccino, G., Saccuman, M. C., Gallese, V., Danna, M., Perani, D., Cappa, S., Fazio, F., & Rizzolatti, G. Sentences describing actions activate visomotor execution and observation systems. *Journal of Cognitive Neuroscience*, in press.
- Trabasso, T., & Suh, S. (1993). Understanding text: achieving explanatory coherence through on-line inferences and mental operations in working memory. *Discourse Processes*, 16(1-2), 3–34.
- Ward, B. D. (2001). *Deconvolution analysis of FMRI time series data Milwaukee*. Wisconsin: Biophysics Research Institute, Medical College of Wisconsin.
- Wernicke, C. (1874). *Der aphasische symptomenkomplex*. Breslau: Cohn, Weigert.
- Zatorre, R. J., Meyer, E., Gjedde, A., & Evans, A. C. (1996). PET studies of phonetic processing of speech: review, replication, and reanalysis. *Cerebral Cortex*, 6(1), 21–30.