Renewal of the Neurophysiology of Language: Functional Neuroimaging

JEAN-FRANÇOIS DÉMONET, GUILLAUME THIERRY, AND DOMINIQUE CARDEBAT

Institut National de la Santé et de la Recherche Médicale U455, Hôpital Purpan, Toulouse, France; and School of Psychology, University of Wales, Bangor, United Kingdom

I. Introduction: Lessons From Aphasia

II. Functional Neuroimaging Techniques
   A. Overview of neuroimaging tools
   B. Tracking brain activations
   C. Outstanding questions

III. Language in the “Healthy” Adult Brain
   A. Single-word processing
   B. Sentence and discourse processing
   C. Beyond the “standard” language
   D. Outstanding questions

IV. Language and Brain Plasticity
   A. Developmental plasticity
   B. Postlesional plasticity
   C. Outstanding questions

V. Conclusions

Démonet, Jean-François, Guillaume Thierry, and Dominique Cardebat. Renewal of the Neurophysiology of Language: Functional Neuroimaging. Physiol Rev 85: 49–95, 2005; doi:10.1152/physrev.00049.2003.—Functional neuroimaging methods have reached maturity. It is now possible to start to build the foundations of a physiology of language. The remarkable number of neuroimaging studies performed so far illustrates the potential of this approach, which complements the classical knowledge accumulated on aphasia. Here we attempt to characterize the impact of the functional neuroimaging revolution on our understanding of language. Although today considered as neuroimaging techniques, we refer less to electroencephalography and magnetoencephalography studies than to positron emission tomography and functional magnetic resonance imaging studies, which deal more directly with the question of localization and functional neuroanatomy. This review is structured in three parts. 1) Because of their rapid evolution, we address technical and methodological issues to provide an overview of current procedures and sketch out future perspectives. 2) We review a set of significant results acquired in normal adults (the core of functional imaging studies) to provide an overview of language mechanisms in the “standard” brain. Single-word processing is considered in relation to input modalities (visual and auditory input), output modalities (speech and written output), and the involvement of “central” semantic processes before sentence processing and nonstandard language (illiteracy, multilingualism, and sensory deficits) are addressed. 3) We address the influence of plasticity on physiological functions in relation to its main contexts of appearance, i.e., development and brain lesions, to show how functional imaging can allow fine-grained approaches to adaptation, the fundamental property of the brain. In closing, we consider future developments for language research using functional imaging.

I. INTRODUCTION: LESSONS FROM APHASIA

Following seminal works by pioneers such as Paul Broca (46) and Karl Wernicke (420), the early conceptualization of mind-brain relationships was based on the clinical anatomical method. Brain lesions described post mortem were tentatively related to cognitive deficits evaluated in vivo, and vice versa, to establish structure-function relationships. Shortly after the first description of aphasia (46, 420), a primary functional model of language emerged in the form of a temporal frontal network (214).

In the century following the foundation of clinical neuropsychology, a crucial advance in cognitive neuroscience was the advent of radiological and isotopic meth-
methods for imaging brain structures and functions. X-ray computerized tomography (CT scanner) first appeared in 1967 (Hounsfield, Nobel Prize 1979) and allowed the seat and extension of brain lesions to be located with precision in vivo. Meanwhile, cognitive models of language and memory processes were developed and helped to characterize more precisely the nature of cognitive deficits observed in neurological patients. Twenty years ago, the further progress of science in the domains of nuclear physics, magnetic fields, informatics, and, more generally, electronics triggered a revolution in the history of neurophysiology. The advent of functional neuroimaging has made the dream of structure-function researchers come true: it is now possible to directly correlate mental operations with indices of brain activity. The great enthusiasm generated by these revolutionary techniques must, however, be modulated by the significance of the results obtained so far in the domain of language, which will be discussed in this review. Nevertheless, the profusion of neuroimaging studies, inquiring into various aspects of language processing and, in particular, those identified by psycholinguistic models, have accumulated enough data over two decades to complement, and even challenge, the classical aphasia model.

At this point it must be noted that the aphasia model is far from being adequate when it comes to deriving representations of brain functions. Although some of the first lesion-related findings have never been totally invalidated (e.g., the involvement of Broca’s area1 in speech production and the critical role of the left superior temporal gyrus in auditory verbal comprehension; for contemporary examples, see Refs. 138, 196), there are several reasons that make the interpretation of lesion studies complex and, in some cases, impossible. Four major limitations of the classical aphasia model should be considered in particular.

1) Language-related brain regions are embedded in complex and highly interconnected networks. It is therefore very unlikely that accidental lesions selectively affect specialized neural networks, such as those related to language, without damaging other, possibly less specialized, functional systems. In particular, lesions may impair language processes by affecting neural structures (such as the basal ganglia or the thalamus) that are connected to specialized brain areas and/or their mutual connections without damaging the specialized areas themselves. Consequently, anatomical lesions located a fair distance away from language-related regions can still affect language function (33).

2) Despite the massive development of diagnostic tools and tests (e.g., Boston Diagnostic Aphasia Examination, Goodglass and Kaplan, 1972, Psychological Assessment Resources, Lutz, FL 33549), the classical syndrome-based approach to aphasia has proven insufficiently specified (423). Not only do syndromes such as “Broca’s aphasia” correspond to poorly defined entities in terms of the cognitive components involved, but also aphasic syndromes as a whole (e.g., agrammatism) cannot be trivially related to reproducible and consistent lesion sites (13). Even when considering more specific disorders such as aphasic symptoms (e.g., word finding difficulties; Refs. 295, 423), necessary and sufficient correspondence between a specific lesion site and a symptom is rarely established. There are several reasons for the scarceness of such one-to-one relationships. It is generally accepted that lesion studies in aphasia can indicate which brain region is necessary for implementing a language process by observing language disorders following focal brain lesions. Notwithstanding the oversimplification of real conditions implied by this logic (see the first point above), the validity of this assumption is generally not assessed in formal, Bayesian terms. Indeed, “sufficiency” implies that the presence of a specific lesion can predict the symptom of interest in any patient. On the other hand, “necessity” implies that the lesion must be present for the symptom to be observed. Although necessity or sufficiency relationships are sometimes validated in clinical data (see Table 1), their coexistence is very rare. Effectively, since language, like other higher cognitive functions, is thought to rely on the interplay of many different brain areas (89, 240, 316), lesion-symptom relationships are likely to be influenced by a set of distributed regions rather than a single, circumscribed area (4).

3) The approach to language-brain relationships in aphasic patients is usually too static. Considerable changes in both language behavior and brain functions take place after a lesion such as focal ischemia has occurred. The neural rearrangement following damage depends on the nature and extent of the lesion, changes over time, and may also be influenced by various therapeutic interventions. These rearrangements give birth to a palliative “brain architecture” that may not be just the “normal” network minus one (damaged) component. In fact, well-described aphasic syndromes may often be the result of neural rearrangements that take place after the trauma which caused the disruption. Indeed, acute aphasia post-stroke is in most cases unclassifiable (138).

4) Numerous subject-dependent factors, such as gender, age, handedness, and literacy (see sect. 14C1 and Table 2), whose precise effects are still poorly understood, seem to substantially influence aphasic symptoms (35, 63). For instance, stroke more frequently affects elderly patients; therefore, the influence of age on aphasic...
TABLE 1. Lesion/symptom relationships in a sample of 18 right-handed aphasic patients observed at the chronic stage of left hemispheric, monofocal, corticocortical, ischemic strokes

<table>
<thead>
<tr>
<th>Left Posterior STG</th>
<th>Paraphasias Present</th>
<th>Paraphasias Absent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Damaged</td>
<td>6</td>
<td>0</td>
</tr>
<tr>
<td>Spared</td>
<td>6</td>
<td>6</td>
</tr>
</tbody>
</table>

A. Phonemic paraphasia in a word repetition task

<table>
<thead>
<tr>
<th>Left PIFG</th>
<th>Nonfluent Speech</th>
<th>Fluent Speech</th>
</tr>
</thead>
<tbody>
<tr>
<td>Damaged</td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td>Spared</td>
<td>1</td>
<td>9</td>
</tr>
</tbody>
</table>

Six of the 18 patients had lesions in the left posterior superior temporal gyrus (STG). In the same group of patients, 8 individuals had lesions in the left posterior inferior frontal gyrus (PIFG). A significant imbalance of contingency tables (Fisher exact test, \( P < 0.05 \)) was observed for two region/symptom relationships: 1) damage to the left posterior STG and presence of phonemic paraphasias and 2) damage to the left PIFG and reduction of speech fluency. A: relationship between phonemic paraphasias and lesion in the left posterior STG. Six patients with a lesion in the left STG produced phonemic paraphasias. Therefore, this symptom shows a sensitivity of 100% to the lesion site: damage to this region always induces the symptom. However, the specificity of this relationship is low (50%) as only one-half \((n = 6)\) of the patients with spared STG \((n = 12)\) were symptom free. Because phonemic paraphasias were observed in 6 other patients with spared STG, it must be inferred that damage to other brain region(s) can also induce this symptom. B: for the same patients, a different pattern of relationship between lesion in the left PIFG and reduction of speech fluency. Reduction in speech fluency can be described as a relatively specific symptom: all patients had a spared left PIFG, but one had normal speech fluency, making the symptom specificity 90%. All patients had reduced speech, but one had a lesion in the left PIFG and the lesion did not systematically induce the symptom as some patients \((n = 3)\) with a lesion in the left PIFG did not have speech fluency reduction. In sum, the symptom lacks “sensitivity” to the lesion. The lack of lesion/symptom relationships establishing both high sensitivity and high specificity suggests that 1) aphasic symptoms generally depend on the integrity of a distributed network of brain structures, 2) the lesion of distinct nodes in the network can induce the same symptom (e.g., phonemic paraphasia), and 3) when the symptom is not observed despite lesions in a given region (e.g., left PIFG region), it may be that spared nodes can alleviate the impact of the lesion by implementing palliative language processes (Démonet et al., unpublished observations). Symptoms cannot easily be dissociated from the etiology of the lesion.

Six of the 18 patients had lesions in the left posterior superior temporal gyrus (STG). In the same group of patients, 8 individuals had lesions in the left posterior inferior frontal gyrus (PIFG). A significant imbalance of contingency tables (Fisher exact test, \( P < 0.05 \)) was observed for two region/symptom relationships: 1) damage to the left posterior STG and presence of phonemic paraphasias and 2) damage to the left PIFG and reduction of speech fluency. A: relationship between phonemic paraphasias and lesion in the left posterior STG. Six patients with a lesion in the left STG produced phonemic paraphasias. Therefore, this symptom shows a sensitivity of 100% to the lesion site: damage to this region always induces the symptom. However, the specificity of this relationship is low (50%) as only one-half \((n = 6)\) of the patients with spared STG \((n = 12)\) were symptom free. Because phonemic paraphasias were observed in 6 other patients with spared STG, it must be inferred that damage to other brain region(s) can also induce this symptom. B: for the same patients, a different pattern of relationship between lesion in the left PIFG and reduction of speech fluency. Reduction in speech fluency can be described as a relatively specific symptom: all patients had a spared left PIFG, but one had normal speech fluency, making the symptom specificity 90%. All patients had reduced speech, but one had a lesion in the left PIFG and the lesion did not systematically induce the symptom as some patients \((n = 3)\) with a lesion in the left PIFG did not have speech fluency reduction. In sum, the symptom lacks “sensitivity” to the lesion. The lack of lesion/symptom relationships establishing both high sensitivity and high specificity suggests that 1) aphasic symptoms generally depend on the integrity of a distributed network of brain structures, 2) the lesion of distinct nodes in the network can induce the same symptom (e.g., phonemic paraphasia), and 3) when the symptom is not observed despite lesions in a given region (e.g., left PIFG region), it may be that spared nodes can alleviate the impact of the lesion by implementing palliative language processes (Démonet et al., unpublished observations). Symptoms cannot easily be dissociated from the etiology of the lesion.

Given the complexity and the limitations of the aphasia model, functional neuroimaging techniques have had a positive and immediately significant impact on the exploration of brain-language relationships because they constitute an independent source of evidence.

Based on different physical principles, positron emission tomography (PET), functional magnetic resonance imaging (fMRI), and multichannel electro- or magnetoencephalography (EEG, MEG) make it possible to measure various indices of ongoing neural activities arising from the brain “in action.” Depending on their relative spatial and temporal resolutions, they depict the functional anatomy of cognitive operations and/or index their time course. However, the above-mentioned limitations of the aphasia model should be kept in mind because most similarly apply to the study of language physiology in normal subjects.

We first address various technical and methodological issues on neuroimaging. Second, we review a series of significant contributions to the understanding of language based on the study of normal subjects. Third, we discuss the matter of language processing in the context of developmental or postlesional brain plasticity.

II. FUNCTIONAL NEUROIMAGING TECHNIQUES

Functional imaging sensu stricto refers to PET and fMRI. We will address various technical and methodological issues related to these techniques before addressing complementary issues in electrophysiological methods (EEG and MEG).

A. Overview of Neuroimaging Tools

In contrast to the static clinical anatomic paradigm, the “activation” method used in neuroimaging experiments is fundamentally dynamic. This method relies on recording changes in indices of cerebral activity (for reviews on technical issues, see Ref. 403). These variations are recorded in the form of tomograms, i.e., successive sets of slices through the brain which allow measurement of the regional cerebral blood flow (rCBF) in different areas. In most studies the main independent variable generating the statistical variance of interest is time. Experiments are based on repeated functional imaging measurements and alterations of controlled experimental parameters over a series of blocks (“block design”) or trials (“event-related design,” see sect. II A2 and Fig. 1). Time scales differ dramatically from one neuroimaging technique to another. In the original, rudimentary use of PET, the temporal resolution was on the order of several minutes (>10 min/measurement). In contrast, the maximal temporal resolution of fMRI is in the range of a few hundreds of milliseconds (122); the typical time needed to acquire a single functional brain slice is in the region of 50 ms. However, the repetition time used to acquire more than one functional slice is typically on the order of 2–4 s, and the time needed to sample the entire hemodynamic response is in the range of 12 s (11). Whatever the temporal resolution, the activation method is based on the comparison of signal level across different experimental conditions (e.g., one condition eliciting a particular cognitive process versus another in which the process is not likely to occur). A basic activation experiment features an “active” experimental task (involving stimulation, cognitive computation, and response) and a “rest” condition in which none of these stages is supposed to be involved.
The statistical difference between condition-specific patterns of activation is then likely to reflect neural activities associated with the process under study. In spite of many unresolved methodological issues (see sect. II), this approach has generated great enthusiasm and has produced attractive, colorful activation maps that were soon found to be congruent with physiological predictions.

In the past decade these techniques as applied to the exploration of language function have provided an enormous amount of data that have proven to be reproducible

<table>
<thead>
<tr>
<th>TABLE 2. Factors that may bias results of language-related neuroimaging experiments</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Subject Specific</strong></td>
</tr>
<tr>
<td>Gender (360, but see 129)</td>
</tr>
<tr>
<td>Age (220)</td>
</tr>
<tr>
<td>Handedness (405)</td>
</tr>
<tr>
<td>Literacy (70)</td>
</tr>
<tr>
<td>Motivation (182)</td>
</tr>
<tr>
<td>Emotion (172)</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

Reference numbers are given in parentheses.

The statistical difference between condition-specific patterns of activation is then likely to reflect neural activities associated with the process under study. In spite of many unresolved methodological issues (see sect. II), this approach has generated great enthusiasm and has produced attractive, colorful activation maps that were soon found to be congruent with physiological predictions.

In the past decade these techniques as applied to the exploration of language function have provided an enormous amount of data that have proven to be reproducible

**FIG. 1.** Block design and event-related design. Block-design experiments (left) are characterized by blocks in which stimuli pertain to one single experimental condition because the response of the brain is cumulated over the entire block. Activated voxels correspond to those parts of the image volume in which signal variation follow a pattern consistent with the theoretical signal generated from block alternation. Comparison of effect sizes (level of activation in each condition found in a sample of participants, bottom left) then allows drawing of statistical inferences. Event-related designs (right) allow the mixing of different experimental conditions, since the hemodynamic response is classically sampled entirely after each stimulus. In addition to statistical evaluation of differences across experimental conditions, this method gives access to the time course of event-related hemodynamic responses (bottom right).
and to contribute to a better understanding of the neural substrate involved. Yet such influential methods actually rely on the measure of indirect and rather rough metabolic indices whose physiological determinism is poorly understood and which have to be recorded from large functional networks distributed throughout the brain.

1. Basic principles of PET and fMRI

Neuroimaging techniques (such as PET and fMRI) have poor temporal resolution relative to typical neural firing rates of neurons, but have reasonably good spatial resolution throughout the entire brain volume. The typical precision of images (cubic millimeter) remains unsatisfactory vis-à-vis functional exploration at the neuron level, however. This limited anatomical resolution might be only an apparent drawback because high-order cognitive operations such as language processing are likely to involve large neural assemblies rather than microscopic circuits.

Signals exploited in PET (gamma emissions provoked by the positron emitting $^{15}$O isotope incorporated in water and injected into the systemic blood supply) or fMRI (magnetic susceptibility of hemoglobin indicating blood oxygenation status) relate to local changes in vascular parameters deduced from the local concentration of the tracer in microvessels. Considering a population of $\sim 10^4$ synapses in which the energetic demand varies suddenly as a consequence of sensory stimulation, motor response, or any kind of neural operation, such vascular changes have been estimated to happen in a brain volume a few hundred micrometers in diameter, using optical cortical activity imaging (404).

In fMRI experiments, the most studied signal is known as BOLD (blood oxygenation level dependent). It is based on the measure of changes in magnetic susceptibility of hemoglobin, depending on whether it conveys a dioxygen molecule or not (271). A sudden increase in synaptic metabolism is thought to be followed by a transient drop in oxyhemoglobin concentration, and consequently in the BOLD signal, in vessels neighboring activated neurons (429). A major increase in oxyhemoglobin concentration then occurs as a consequence of vessel dilatation, with a peak observed $\sim 5$–$6$ s after stimulus onset time (SOT). This massive local vascular response provides more metabolites than needed by neural activity. The physiological relation between blood flow in gray matter vessels and local variations of neural metabolism remains largely unknown, although recent advances have begun to unravel some important characteristics of this phenomenon (217), such as the role of astrocytes as "energy transducers" interposed between capillary walls and neurons (222).

2. Recent evolution in fMRI

Combining resolution in space and time is a general requirement for neuroimaging of cognitive functions; it is especially crucial when studying the neural correlates of language functions. Indeed, if one accepts that written language is an acquired artifact, human language depends primarily on auditory and vocal functions, which are linked to time in very essence, as they rely on a continuous stream of events. Because it is now possible to record fMRI signals from the whole brain volume in $\sim 1$ s, the temporal resolution of the BOLD effect is much higher than that of the $^{15}$O-PET response, which can only be computed after integration of gamma activity over a minimum of 30 s. Although some parameters, such as pulse or respiratory rates, have to be carefully controlled in fast fMRI acquisition, this method is emerging as the state-of-the-art approach to brain mapping, in paradigms such as single-trial acquisition.

Two current methods are used to acquire fMRI data: block design and single-trial or event-related design (see Fig. 1). In the first method, alternation of different conditions (activation/rest) is used as an entry function to convolve the hemodynamic response. The different blocks consist of different conditions, during which stimuli are presented and/or responses are required from subjects. A statistical analysis then identifies voxel signals that correlate with the alternation of experimental conditions (11). Such a correlational approach improves the amount of information that can be extracted from fMRI data because the signal-to-noise ratio is low when classical subtractive $t$-test analyses are employed.

In the second method, single-trial or event-related fMRI (55, 337), the hemodynamic response corresponding to each single stimulation is acquired individually. This is achieved using two different procedures: 1) stimuli are presented tens of seconds apart, allowing the complete sampling of the hemodynamic response between two stimuli (i.e., over at least 12 s, Ref. 10), and 2) alternatively, stimuli are presented at a faster rate (e.g., every second) and unitary hemodynamic responses are reconstructed by deconvolving the summed response acquired over the entire series (162, 241). Compared with block design, single-trial design makes it possible to present stimuli in a randomized order and therefore reduce habituation effects. Some studies have even described the recording of single events, even though signal-to-noise ratio in association cortices is in most cases insufficient (e.g., Ref. 270).

Some authors have pushed the temporal resolution of event-related fMRI to its actual limit (i.e., the time necessary to acquire one slice) to temporally discriminate activated clusters. For example, Menon et al. (238) used a single-trial design to show that the BOLD response elicited by a visuomotor task in the premotor cortex was
delayed compared with the BOLD response of the primary visual cortex. Moreover, they found a robust correlation between participants’ reaction times and the lag between onsets of hemodynamic response in the primary visual area and the supplementary motor area. Some preliminary results in the language domain also appear very encouraging (154, 391, 395, 396; see sect. nC2). Single-trial fMRI might well become the best procedure for collecting spatial and temporal information simultaneously because its temporal resolution, which is in the range of 100 ms/slice, can still be improved (for a review on the potential of time-resolved fMRI, see Ref. 122). Important questions currently being addressed are as follows: 1) What is the reproducibility of the hemodynamic response in different conditions, regions, subjects, or scans? 2) Does the slow time course of the BOLD effect impair its capacity to temporally discriminate brain activations? 3) Can we model the relationship between the amplitude of the signal and its time course in different brain regions?

3. Electrophysiology and imaging

Electrical and magnetic neuroimaging techniques are based on the noninvasive recording of electrical and magnetic field variations induced by neural activity. Neural electrical activity can be divided into two categories: action potential (AP) and postsynaptic potentials (PSPs). The AP corresponds to the propagation of ion flux bursts along the axon of a neuron and can be described as a quadrupole, the magnetic and electric fields of which decay rapidly (422). PSPs can either be excitatory or inhibitory and are larger bursts of ion exchanges on the surface of postsynaptic neurons. In contrast to the AP, PSPs can be described by current dipoles active for several tens of milliseconds. As a consequence, magnetic and electric fields recorded over the scalp derive from the summation of PSP dipoles rather than AP sources (218).

Pyramidal cells mainly found in layer V of the cortex are tall and parallel to one another. Their orientation is perpendicular to the surface of the cortex, and groups of a few hundred thousand pyramidal cells (i.e., a few mm² of cortex, see Ref. 331) activated simultaneously can produce electrical and magnetic activity deriving from PSPs and measurable on the surface of the scalp. Scalp electromagnetic activity is therefore mainly a consequence of synaptic discharge and global cellular polarization in colinear neurons. The recording of electrical activity using electrolyte gel and highly conductive electrodes is known as EEG, while MEG is the recording of correlative magnetic field variations using very sensitive sensors. Here, neural mechanisms underlying surface effects are better known than in the case of tomographic techniques. However, the precise localization of electrical and magnetic sources is complex. On the one hand, predicting surface pattern from the location, orientation, and intensity of brain sources (forward modeling) implies a comprehensive approach to the propagation of electromagnetic flow throughout the brain and the head tissues. On the other hand, modeling brain sources on the basis of surface recordings may be misleading because this backward modeling problem has an infinite number of solutions, especially if multiple and deep sources are likely to be involved. With that said, source analysis has recently benefited from the integration of whole brain structural anatomy provided by high-resolution three-dimensional MRI.

From EEG and MEG are derived event-related potentials (ERPs) and evoked magnetic fields (EMFs). ERPs and EMFs are based on the averaging of a large number of recordings time-locked to the occurrence of a stimulus to compensate for their low signal-to-noise ratio. Averaging over a large number of trials progressively cancels spurious brain electrical or magnetic activity that does not relate to the cognitive task performed by the participant. Conversely, electromagnetic activity relating directly to the processing of the event (stimulus) and subsequent cognitive operations are enhanced by averaging and emerge in the form of a series of positive and negative deflections (see Ref. 342; for a review on ERP components elicited by language processing, see Ref. 202). This dominant approach overlooks the value of studying changes in the spectral power of electromagnetic signals (quantitative EEG) and synchronization phenomena across different recording sites (coherence analysis; see, for instance, Refs. 332, 370, 382). Such new approaches could be applied to the investigation of language processing, however (86, 195).

4. New imaging techniques

Several new neuroimaging techniques, such as optical or near-infrared cortical imaging (383, 411), spectroscopic magnetic resonance imaging (e.g., Ref. 365), diffusion tensor imaging (e.g., Ref. 85), or physiological techniques that are not brain centered (e.g., pupillometry, cardiovascular measures, and electrodermal activity; for a review, see Ref. 200), are likely to complement current approaches to exploring the physiological substrates of cognitive processes and the timing of their involvement.

Computerized image processing software implementing elaborated mathematical unfolding procedures used in conjunction with high-resolution MRI techniques will provide significant improvements in our understanding of brain functions. For instance, cortical unfolding routines have proven especially useful in the exploration of the visual system (88, 197). Systematic statistical approaches to the morphometry of cortical regions may reveal significant differences in terms of brain functions in individual subjects (e.g., Ref. 223). With the use of diffusion tensor imaging, it is now possible to track the three-dimensional
geometry of white matter bundles and fascicles (85), allowing the modeling of neural networks in vivo. For instance, imaging diffusion and perfusion a few hours after ischemic stroke can reveal functionally impaired regions that are not at the core of the lesion. This advance will lead to a better understanding of the brain-symptom relationships in the acute phase of aphasia (160). Very brief and localized magnetic pulses produced by transcranial magnetic stimulation (TMS) can be used to transiently stimulate neural populations and induce either facilitation or inhibition of cognitive operations (277). This technique can demonstrate the intervention of particular cortical areas in a precise time window, from the onset of the stimulus in a confrontation naming task for instance, and can be used in combination with functional imaging (116). In the future, TMS should help to address complex language issues, such as the role of the right hemisphere in functional compensation of aphasia.

B. Tracking Brain Activations

1. The activation paradigm

   The first way to conceptualize the cognitive structure of “activation” experiments was an additive model, akin to the “pure insertion” hypothesis (128). In this model, an active condition involves several cognitive components that are thought to be independent from one another (e.g., input, intermediate, and output components). The “deletion” of one component in a second task is hypothesized to leave the other cognitive components unaltered. Consequently, the contrast (logically called “subtraction”) between the patterns of activity observed in the first condition and those obtained in the second condition are supposed to reveal the neural correlates of the removed component. The “additive-subtractive” model is usually used as the first step in data analysis because it has proven capable of leading to straightforward inferences. This is the basis of “hierarchical” designs that involve several tasks of increasing complexity. As cognitive components are progressively added to the higher order tasks, activation maps are thought to reflect the “add-up” effect in neurofunctional terms. Therefore, the additive-subtractive model appears to be the transposition of the transparency hypothesis formulated by Caramazza (65) to functional neuroimaging. It thus allows neuroimaging results to be compared with clinical anatomical findings. However, this model has important limitations, as stressed by Friston et al. (128). Its hierarchical structure implies that all components of lower order tasks are entirely embedded in any higher order tasks. In the case of complex functions such as language, a given task does not require a simple series of successive and independent processing stages that can be added or subtracted at will. Rather, such tasks involve different cognitive processes implemented in a parallel and interdependent fashion and need to be approached accordingly. Depending on the experiment, the manipulation of two or more experimental factors may be such that their combination would induce changes in neural activity that are not the simple, straightforward addition of activations elicited by each of them, but are, for instance, greater than this sum. Signal changes induced by activation tasks are thus frequently nonlinear and make it necessary to consider factorial interactions.

2. Statistical approaches to structure-function relationships

   A) Factorial design. Following this line of research, the Friston and Frackowiak group in London have set up a general method and dedicated software, Statistical Parametric Mapping (SPM, www.fil.ucl.ac.uk), in which a voxel-by-voxel analysis is performed to test experiment-induced signal changes according to the general linear model. This group has emphasized the interest in building up experiments in which cognitive components can be used as orthogonal factors so that their potential interdependency can be directly investigated.

   The statistical analysis is based on a factorial design in which each cognitive component corresponds to a main effect, and the interaction between these factors can be formally tested. Other independent variables, such as subject-specific characteristics (e.g., handedness, gender), can also be involved in such analyses (for an example, see Fig. 2).

   Paulesu et al. (279) provided one of the first clear examples of the advantages of such a factorial approach in a study of phonological verbal working memory using PET. The authors built an experimental design involving two binary factors yielding four conditions. Two tasks performed by English-speaking subjects were used: a memory task and a similarity judgment task on Latin letters and unknown Korean letters. The four conditions were thus verbal (Latin letters) versus nonverbal (Korean letters) and load (memorization) versus no load (similarity judgment) in working memory. In the factorial analysis of the PET data, a main effect for each factor was shown. Moreover, the authors found an interaction between the two factors, i.e., they found evidence for selective activation when verbal and memory load factors were combined. This analysis corresponds to the result of a “double” subtraction, namely (verbal load – verbal no load) – (nonverbal load – nonverbal no load). Disproving the null hypothesis for this subtraction means that, in some brain regions, the neural response is not reducible to the strictly additive effects of each factor, but that interactions exist between them. The authors showed that rCBF increase in the inferior part of the left supramarginal gyrus was significantly greater than what would result from the addition of the “verbal over nonverbal” and the “load over no
load" factors. In other words, this particular portion of cortex is likely to play a specific role in loading verbal items in working memory.

B) COGNITIVE CONJUNCTIONS. An interesting alternative to the additive-subtractive model is the conjunction approach. Instead of contrasting different conditions to relate a specific brain region to a specific cognitive process, one can investigate which parts of the brain are systematically active in various tasks sharing a defined cognitive component (304). This method is useful to demonstrate that different tasks may require common neural substrates in spite of their particularities. Combined with random effect analysis across groups of subjects, it is particularly efficient for identifying neurofunctional crossroads, i.e., cortical regions in which functionally distinct neural networks overlap with each other and show repeated activations although the cognitive tasks used are very different. Price and Friston (304) used this approach to characterize a common network involved in four different visual naming tasks, each of them associated with a specific reference task controlling for the effects of basic perceptual processes.

C) PARAMETRIC APPROACH, CORRELATIONAL APPROACH, AND STRUCTURAL EQUATION MODELING. Activation phenomena appear richer than what was originally observed in the framework of the additive model. Friston et al. (125) showed that principal component analysis may reveal a lot about the functional systems involved in cognitive experiments by unraveling functional connectivity in the brain (for an example of this method in a reading task, see Fig. 3). Without an a priori hypothesis about the differences to be expected from a particular contrast, activated networks may be related to the influence of continuous variables such as time and/or behavioral indices (e.g., Ref. 142). In the language domain, systematic research carried out by Price et al. (307, 311, 312) has focused on the influence of lower order stimulus-dependent factors (see sect. B4) and has shown that nonlinear correlations between such variables and changes in rCBF might be even more significant than linear ones.

As stressed earlier, activation experiments mainly rely on time-related signal changes. However, some functional neuroimaging studies (often designated as "correlational" studies) have specifically investigated changes in functional signals from one subject to another. Carried out with a group of subjects, these studies are based on the analysis of metabolic correlations across regions (164) or between metabolism in a given region and cognitive performance (e.g., Refs. 87, 108, 165). This line of research was further developed to explore the impact of independent factors on the functional relationships in a distributed network whose anatomical structure is known a priori. This approach is known as structural equation modeling, in which anatomical knowledge about connections across several key regions is used to constrain models of effective connectivity and explore its modulation by experimental conditions such as language requirements or memory load (52).

3. Convergence of neuroimaging and electrophysiology

As mentioned above, the temporal resolution of electrophysiology (EEG and MEG) is compatible with the speed of cognitive processes, while its spatial information remains poor. These properties are in striking opposition to those of tomography (PET and fMRI). Used independently, each type of technique requires highly specific activation paradigms focusing either on the temporal or the anatomical dimension of the signal, respectively. This specialization has rendered electrophysiological and tomographic data impossible to compare for at least 20 years. Nevertheless, different authors (9, 96, 374, 393) have reported language experiments performed with PET or fMRI on the one hand and ERPs on the other in an attempt to provide complementary results (see, for instance, Refs. 396 and 392). Given the evident complexity of correlating PET and ERPs, for instance, it is necessary to consider a single and unique paradigm that is compatible with both techniques. One possible procedure is to identify a set of activations via tomographic methods and then explore the relative intensity of corresponding brain...
generators (see Fig. 4) using source analysis software such as BESA (353) or Curry (Neuroscan). This procedure offers a way to bypass the inverse problem raised by brain source analysis (cf. sect. II B3). The functional significance of localized activations can then be revisited in terms of the kinetics of activation in neural assemblies. The opposite procedure can be proposed: first elaborate a tomography-compatible paradigm eliciting a well-defined evoked component such as the P300 or the N400, and then seek for its generator with tomographic techniques (e.g., Ref. 367). A promising procedure is the replication of ERP or MEG experiments using single trial fMRI (96, 395). However, such protocols cannot just be adapted from one technique to the other but must be specifically developed (see sect. II C3).

C. Outstanding Questions

1. Influence of subject-dependent and stimulus-dependent parameters

Whatever the procedure for functional neuroimaging analysis, subject-specific parameters and task generic parameters of language experiments have a profound, albeit frequently disregarded influence on the results. As the effects of these parameters were not anticipated in the pioneering experiments, they were not studied in a systematic way. In fact, their impact was established progressively, sometimes as a by-product of studies designed for other purposes. The most relevant of these variables are listed in Table 2.

For example, the earliest studies of brain activation using nontomographic isotopic blood flow measurements incidentally pointed out the overwhelming influence of motivation (higher activation and network modulation being observed in subjects showing high motivation; e.g., Ref. 204) and emotional state (a reduction of anxiety along a time series of brain recordings being associated with a global decrease in activation). The effects of these variables have been specifically investigated in recent and sophisticated studies (e.g., Refs. 172, 182).

Although the vast majority of functional neuroimaging studies have been conducted in young, well-educated subjects with a predominance of males (usually undergraduate or graduate students recruited in universities or laboratories), subject-dependent parameters such as gender, age, handedness, or literacy have been found to influence activation patterns drastically. Male subjects, for instance, were originally thought to display a stronger left-greater-than-right asymmetry for language (314, 360), although these results failed to be confirmed by further studies conducted in larger groups (129, 376, 410).

If one excludes the case of infant development (see sect. IV A), the influence of aging on language processing has been largely overlooked so far. In a study of visual recognition, however, Madden et al. (220) described an
increase of activity in the anterior part of the ventral visual system in healthy elderly subjects compared with younger controls. One cannot estimate the impact of this parameter on already published data.

Handedness is frequently viewed as a major factor influencing hemispheric dominance for language. Notwithstanding differences between right- and left-handers in terms of structural anatomy (e.g., Ref. 406), systematic neuroimaging studies including large subject samples have recently demonstrated that left-handers tend to present a left hemispheric dominance for language. Although the right hemisphere is less involved than the left in left-handers, the functional asymmetry is less marked than in right-handers. Activation has only rarely been observed in the right hemisphere in isolation (315, 376, 405).

Other studies have stressed the importance of more general, though less obvious, sources of signal modulation. A series of experiments performed by Price and co-workers (306, 312) focused on the influence of low-order stimulus-dependent factors, such as exposure duration and rate of stimulation, together with absence/presence of an overt utterance while reading. These authors demonstrated massive and unexpected effects in a variety of areas including “key” regions such as the left occipital temporal cortex and premotor areas. The response function relative to these stimulus-related parameters varied dramatically even between two neighboring areas, e.g., fusiform versus lingual gyrus.

While keeping these low-order parameters constant, other general factors might also bias brain mapping of language functions. For instance, subjects’ familiarity with the task may dramatically alter the pattern of activation, as demonstrated by Raichle et al. (320). These authors compared activations measured during a verb generation task with activations recorded in the same participants doing the same task after extensive training with the specific word list used. Much of the activation observed at the naive stage in the left inferior frontal areas disappeared at the trained stage and was seen again, though to a lesser extent, when subjects were presented with another word list.

An important source of modulation of activation patterns is the degree of task “difficulty” that can be manipulated via several experimental features, such as perceptual ambiguity between targets and distractors (e.g., phoneme targets among phonetically similar versus dissimilar distractors, Ref. 106) or the number of candidates among which an item has to be chosen in a word generation task (401).

2. New methodological challenges

A) TIME COURSE OF EVOKED HEMODYNAMIC RESPONSES. The issue of hemodynamic response variability between brain regions and between individuals has been extensively addressed in the last decade (2, 10, 53, 55, 110, 184, 206, 241, 352, 391). In two studies of evoked hemodynamic responses (EHRs) recorded during language tasks, Thierry...
and co-workers (391, 395) found a sequence of hemodynamic peak latencies that was compatible with physiological expectations (e.g., primary auditory cortex early, superior temporal regions intermediary, inferior prefrontal regions late).

However, it must be kept in mind that the hemodynamic response has proven too variable across regions in terms of timing, amplitude, and shape to enable direct comparison between different parts of the brain (10, 53, 55, 206, 352). Such regional differences might be due to variable influences of microscopic and macroscopic blood flow (76), to differential vascular sampling, or to real differences of neural activity (55, 241, 352). Although the hemodynamic response of one region is susceptible to being dysphased by several seconds across subjects (55, 184, 241), its grand-average latency and amplitude have proven reproducible for groups of subjects as small as \( n = 6 \), i.e., the central tendency of the EHR can be reproduced in different groups of subjects and, a fortiori, in the same group of subjects across experimental blocks, with a precision of tenths of seconds (55).

More recently, several authors have proposed a temporal analysis of averaged fMRI signals in cognitive tasks, called time-resolved fMRI (for a review, see Ref. 122). Formisano and Goebel (122) have proposed that the main processing stages of cognitive operations can be temporally differentiated using time-resolved fMRI, although fast neural exchanges between two interconnected regions are unlikely to be distinguished.

Thierry et al. (396, see Fig. 5) found that EHR peak latencies were significantly delayed by one experimental factor (maintenance of information in verbal working memory) in one region of the brain (inferior prefrontal cortex) while they remained identical in another region (superior temporal regions). According to Miezin et al. (241), the hemodynamic response in a given region is nearly identical from one data set to another (time to peak correlation \( r^2 = 0.95 \) across sets) so that the significant difference found by Thierry et al. (396) for the inferior prefrontal cortex can only relate to the difference introduced by condition or task variations and not to spurious hemodynamic effects. If cognitive tasks can significantly influence the time course of hemodynamic response, event-related fMRI provides a unique opportunity to merge spatial and temporal information in a single approach.

However, the temporal resolution of event-related fMRI remains poor to date (in the range of 50 ms for one slice), even though further technological developments have been announced. The most important limitation comes from the slow kinetics of the hemodynamic response itself (typically peaking at 6 s after SOT). The search is clearly on for a sound mathematical model of EHRs, taking into account regional specificities (175).
Another promising approach might be the investigation of the initial dip observed at the onset of the BOLD response, and thought to relate to a transient drop in oxygen concentration (238, 429). A spatial temporal approach to brain activations would allow powerful physiological models of brain function to be developed. If fMRI cannot provide the degree of precision needed in both space and time, it might be of interest to attempt the merger of functional anatomical data with temporal measures from another source, such as electrophysiology.

B) FUSION OF NEUROIMAGING AND ELECTROPHYSIOLOGY. Simultaneous recording of both BOLD and electrophysiological signals has already been reported in investigations of memory (343) and vision experiments (39, 377). Complex EEG artifacts generated by this procedure and relating to pulsatile blood flow seem to vary from one participant to another but can be corrected accordingly (pulse artifacts; see Ref. 5). Conversely, the magnetic susceptibility of EEG electrodes and electrolyte can induce fMRI image distortion (38). Nevertheless, these technical difficulties can be overcome, and studies reporting simultaneous EEG and fMRI recording are already being published (226).

The time course of EHRs cannot be directly compared with that of ERPs, and the procedure for statistical analysis is very different. The question is to what extent EHRs and ERPs can provide convergent sources of information about the same cognitive process. fMRI essentially provides anatomical differences between conditions while EEG contributes temporal windows of differences between these same conditions, but a direct correspondence between a region of the brain and a moment of involvement can be established only in the case of highly focal brain activations (e.g., activation of primary sensory regions or motor cortex). As soon as one considers distributed cognitive networks, such as those involved in language processing, the number of combinations [region of interest (ROI), equivalent brain generator, real time latency of activity] becomes overwhelming. No statistical method is available to date for such a four-dimensional mapping of brain activation. Therefore, one has to consider what the technique can offer to elaborate the methodology used. In other words, experiments using simultaneous fMRI and EEG recording need to rely on specific spatial and temporal hypotheses that can be tested independently by the two techniques. Thus the real advantage of using the two simultaneously is the guarantee that cognitive processes underlying the anatomical results are identical to those eliciting electrophysiological effects.

3. Limitations of neuroimaging

We have chosen to address the issue of constraints imposed by the physics of the scanner and the biophysics of brain metabolism in section iiC3a, and we address questions relating to experimental parameters such as stimuli, tasks, timing, and statistical analysis of the data in section iiC3b.

A) HARDWARE CONSTRAINTS. A fundamental drawback of current functional brain mapping methods is that they do not reflect neural metabolism per se but only indirect, vascular phenomena. Hence, it is not possible to distinguish between excitatory and inhibitory neural processes as they are both thought to induce energy consumption resulting from local synaptic activity, and to thus correspond to an increase in vascular signal. The intensity of signal changes might be much less for inhibitory synaptic populations, however, and such populations are thought to be less widely distributed in the cortex (413). New fMRI approaches using water diffusion tensor imaging (e.g., Refs. 85, 205), direct imaging of neural firing (37), or spectroscopic imaging (365) may provide fruitful alternatives to traditional BOLD monitoring as they permit more direct exploration of neural tissue metabolism. In addition, diffusion tensor imaging allows the tracing of neural pathways. Abnormalities in fiber bundles connecting cortical areas involved in language have been correlated with language impairments, despite the fact that the cortical areas themselves are spared (187).

Recent studies gathering data from C13 magnetic resonance spectroscopy, high-field fMRI and extracellular recordings in anesthetized rats explored the relationships between the BOLD signal, oxygen consumption rate, and cellular firing during a sustained stimulation of the somatosensory cortex (169, 372). Under such experimental conditions, a coupling between oxygen consumption and firing rate was found in a cortical layer mainly reflecting the activity of glutamatergic neurons. Smith et al. (372) stressed that the amount of energy consumed at "baseline" (via oxidative glycolysis) is massive compared with minute stimulation-induced changes. They also reported stimulation-induced decreases of electrophysiological signal, recorded in ~10% of electrodes.

Nevertheless, deactivation has been observed in various PET studies (e.g., Ref. 364) and fMRI studies (26, for a review see Ref. 147). The fact that similar deactivations can be seen using both techniques substantiates the fact that BOLD and PET signals are linked with rCBF variables. If rCBF variations reflect synaptic activity, inhibitory groups of neurons, like excitatory ones, should participate in brain "activation," although to a lesser extent (413). Thus deactivation loci should correspond to regions that are inhibited (rather than inhibiting).

Deactivations probably involve more than one physiological mechanism and include both local and large-scale effects (364). The latter probably have a major impact in language studies. When subjects are at rest, i.e., not focusing on any particular cognitive process, attentional resources are widely distributed over cortical territories. However, when subjects engage in higher cogni-
tive operations such as language processing, attentional resources are reallocated as a result of mutual competition between different processing pathways or subnetworks. The logical consequence of this reallocation is an increase in activity in the operative network and a decrease in activity in irrelevant functional systems (see Ref. 152 for discussion of cross-modality suppression effects). Such competitive mechanisms have been proposed to account for disorders of attention caused by thalamic lesions (417) and are congruent with neuropsychological models of attention (203). From a physiological point of view, Gusnard and Raichle (147) have suggested that deactivated areas, i.e., the posterior cingulate cortex, the posterior temporal/parietal cortex, and the medial frontal cortex, are involved in a “default” mode of brain functional status linked to nonspecific conscious experience. These authors proposed that the energy metabolism in these areas, especially the posterior cingulate cortex, is characterized by a constant and tight coupling between oxygen and glucose consumption, while phasic activation in other territories results in transient anaerobic episodes characterized by a very limited increase in oxygen consumption concurrent with a large increase in glucose consumption and local blood flow. Recent findings indicating a positive correlation between oxygen consumption and neural firing rate under stimulation in anesthetized rats (169) do not seem to support this hypothesis, but this apparent discrepancy might relate to methodological differences between experiments. Further basic physiological experimentation is needed to clarify this issue.

Because of its peculiar and reproducible anatomical distribution, deactivation was proposed to reflect implicit verbal elaboration in subjects supposed to “keep at rest” (26). Covert inner speech phenomena could account for apparent deactivation in some areas of the temporal parietal association cortex when rest is compared with cognitive tasks that are known to recruit other areas. Even if the neural basis for deactivation has not yet been clearly elucidated, the contribution of deactivation to the understanding of the physiology of language should be considered since 1) increase in activity in some regions of the brain during highly demanding, explicit language operations may be mirrored by dimmed metabolic signals in other regions, and 2) implicit, covert language processing may alter cross-condition comparisons by inducing an apparent decrease in activation in association cortices.

Aside from such conceptualization related to high-order phenomena, it remains the case, as pointed out earlier, that an immense gap exists between measurements of signal changes in language-specific large-scale networks and the recording of activities at the neuron level. The ultimate technical goal is to describe fundamental neuronal mechanisms that generate signal changes recorded by neuroimaging methods such as fMRI. Simultaneous recordings of intracortical neural activity and BOLD signals in monkeys’ striate cortex during visual stimulation suggest that the BOLD effect mainly reflects dendritic input rather than spiking output (217). This study also substantiates decreases in BOLD signal in the periphery of the activated cortical area as an index of corticocortical inhibitory mechanisms.

Even though simultaneous intra/extracortical recordings cannot be obtained under standard conditions when normal participants perform language tasks, they provide insight into the principles of “local” cortical physiology. The open challenge for the neuroimaging of cognitive functions such as language is to build up a general model integrating elementary and local cortical physiology into the global dynamics of large-scale neural networks.

B) METHODOLOGICAL CHOICES. Methodological choices directly determine how hardware constraints can be tackled to answer essential questions. For instance, is it possible to determine the invariant parameters of brain activation from one individual to another? Is it possible to characterize the determinants of individual variability in brain activation and describe their influence? These questions not only affect the theoretical significance of empirical studies, but are also critical to clinical applications, since only the most reproducible results can be reliably used in presurgical mapping of language in the brain.

From a physiological standpoint, constructing experiments and interpreting the results requires that one takes fundamental physiological facts into account. 1) Neural substrates implementing cognitive functions are distributed over the entire encephalon, with functional crossroads or “nodes” being closely interlaced in some “bottle-neck” regions (240), such as the left STG, the angular gyrus or the basal temporal language area (cf. infra). 2) Significant changes in patterns of activity are minute in the energy range (≈10⁻² of the measured signal or less) compared with the baseline level of activity measurable in the brain. 3) These changes occur over several tens of milliseconds or more, i.e., in a time scale that clearly overrides the time range of neural events. 4) Subject-specific variables and various experimental parameters (see sect. II C 1) can influence the functional state of neural networks more than cognitive tasks, making the signal-to-noise ratio insufficient vis-à-vis interindividual variance. In other words, despite the rush to collect neuroimaging data depicting the neural basis of human cognition, it is now obvious that one cannot freely manipulate all the variables that may alter patterns of activity in the brain. Activation experiments can only tackle transient and minor alterations of complex patterns under the influence of carefully selected experimental stimuli, instructions, and, most importantly, in the framework of clearly established hypotheses deriving from robust cognitive models.
In fact, neuroimaging language (or any cognitive process) is equivalent to dealing with more or less fuzzy pictures or echoes from a complex and moving landscape (a problem very similar to that of mapping subtle ocean streams from a satellite). Over and above the subtlety of language-related brain signals, a critical issue is signal variability from one subject to another, relating to background noise and region location within the three-dimensional structure of the cerebral cortex and subcortical nuclei. In the past two decades, neuroimaging studies of language have largely overlooked subject variability and have based their conclusions on averaged data obtained in small groups of subjects (typically <20). In spite of such limitations, this approach has proven empirically valid (see sect. III). However, having reached the point of validation, this new domain of physiology now faces the challenge of specifying the many sources of signal variability, or at least, of defining the conditions under which such variations can be optimally reduced by making pertinent methodological decisions.

Such issues have been recently addressed by several authors, especially by the group of the Wellcome Department of Imaging Neuroscience in London. Friston et al. (127) proposed a “random-effect” approach to statistical comparison between groups of subjects. Earlier neuroimaging studies typically used “fixed-effect” analyses in which the results of planned contrasts between conditions apply exclusively to the studied subject sample. In such an approach, significant effects can be induced by signal changes in one or two subjects, possibly yielding spurious results at the group level. The more recent random-effect method implemented in SPM uses estimation of between-subject rather than within-subject variance, and degrees of freedom relate to the number of subjects rather than to the number of scans. Consequently, group analyses are performed using contrast images involving one image volume per subject per contrast, and they allow generalization of the findings to the whole population.

Another recent advance in this field is the neuroimaging of single subjects (e.g., Ref. 308; see sect. IV.B). Although crucial to the renewal of single case studies in neuropsychology, especially for assessing the biological impact of therapeutic interventions (209), this analysis of brain activity changes in one patient requires further investigation to define the optimal statistical analysis (conjunction and contrast statistics).

With respect to the distinction between block-design and event-related design described in section II.A2, it is worth mentioning that methodological problems inherent in these two experimental modes remain unsolved. When engaged in a block-design experiment, participants are likely to show strong habituation effects; they may become tired or even drowsy (especially in the confined environment of the fMRI scanner), and, because they are exposed to the same experimental condition for a substantial period of time (in the range of 30 s), they may develop task-specific strategies (see, for instance, Ref. 294). Although this issue can be partly addressed by the counterbalancing of experimental blocks and by independently manipulating task difficulty (e.g., Ref. 394), it must be kept in mind that the spontaneous nature of brain processing is less likely to be observed in a series of very similar trials than in randomized series of trials. The use of an event-related design solves the problem of habituation and is meant to reduce strategic effects; however, it brings with it other methodological pitfalls. When a participant is exposed alternately to different experimental conditions in each trial, task switching and attentional mechanisms are likely to contribute substantially to the pattern of activity found. To overcome this, some authors have resorted to running their experiment using both block-design and event-related design (e.g., Ref. 82). Although rather laborious to implement, this approach makes it possible to check for the consistency of results in different cognitive contexts (75).

III. LANGUAGE IN THE “HEALTHY” ADULT BRAIN

In the last two decades, the neuroimaging of language has produced a profusion of data using a wide range of cognitive contexts, even though the tasks and paradigms used might a priori appear to be similar. In general, these studies have focused on averaged data (and thus blurred images) obtained in relatively small groups (typically <20 participants, often <10) involving young, well-educated normal subjects. The following review is not exhaustive. It attempts to summarize core findings regarding the classical levels of information processing hypothesized by the major cognitive models of language and essentially established for single-word processing (212, 213, 227, 228, 235, 209). Such levels range from word perception to word production, with the intermediate levels corresponding to semantic representations and processes. Figure 6 provides an overview of the main regions of cortex mentioned in the text hereafter.

A. Single-Word Processing

Most studies have addressed language physiology on the basis of its two main input routes: audition (spoken words and environmental sounds) and vision (written code, sign language, scene and picture viewing). However, some exceptions can be noted. For instance, the functional mapping of brain regions related to olfactory input and their links to verbal representations (318) have shown that familiar odors associated with verbal labels...
FIG. 6. Main brain regions involved in language processing. Top: lateral view of the left hemisphere of the human brain. Bottom: medial sagittal view of the right hemisphere, which has been tilted externally to show the basal surface of temporal lobe (shaded surface). Cut tissue is hatched. The main anatomical structures involved in language processing and mentioned in the present paper are colored and labeled. Numbers indicate Brodmann's areas (BA) separated by dotted lines as listed in Reference 113.
yield specific activation in the left cuneus, suggesting a particular involvement of mental imagery.

1. Auditory input

A classical issue concerning speech perception is the dominance of the left temporal cortex. Several neuroimaging studies (e.g., Refs. 24, 103, 105, 106, 259, 279) initially located the structures involved in the processing of language-specific sounds in the left superior temporal association cortex surrounding the primary auditory cortex (i.e., the medial part of Heschl’s gyrus, Ref. 421). More recently, numerous publications including meta-analyses (27, 30, 158) have highlighted the involvement of the anterior part of the left superior temporal gyrus and the superior temporal sulcus in both hemispheres as the main neural substrates involved in the auditory representation of speech components, including those specific to the human voice (21). Attempts to localize neural responses that are specific to the human voice or to speech components do not seem to point to a single, homogeneous, and clearly left-sided area, although a slight left-greater-than-right asymmetry in the temporal structures was noted by Binder et al. (27).

Several factors might influence the functional dominance of the left superior temporal gyrus in language processing.

1) The rate of change over time in speech signals. Fast changing temporal cues seem to elicit preponderant activations in sound spectral structure, a major feature of temporal cortex. Binder et al. (22) showed that the right superior temporal sulcus is sensitive to auditory object distinctiveness, for example, the ability to identify the characteristic sound of the human voice (21). Attempts to localize neural responses that are specific to the human voice or to speech components do not seem to point to a single, homogeneous, and clearly left-sided area, although a slight left-greater-than-right asymmetry in the temporal structures was noted by Binder et al. (27).

2) The nature of phonological entities. Démonet et al. (106) have shown that an increase in processing load during phonological tasks of graded difficulty results in sustained activations in left-sided temporal regions.

3) The verbal/nonverbal coding of the input. Thierry et al. (394) have recently obtained evidence for a left/right functional dissociation between the left and right superior temporal gyri depending on whether the information processed is verbal or nonverbal. Using highly controlled semantic tasks involving verbal and nonverbal auditory stimuli and stimulus-specific baselines, the authors found greater activations in the left and the right superior temporal regions for accessing semantic information on the basis of spoken words and environmental sounds, respectively.

Besides the issue of functional lateralization for speech perception, the neural correlates of auditory perception in the superior temporal cortex appear to be organized along a rostrocaudal gradient surrounding the primary auditory cortex (center of the system). This rostrocaudal gradient relates to a what/where functional distinction comparable to that observed in the visual modality (see, for instance, Refs. 3, 221). Echoing studies of the auditory cortex in nonhuman primates (177, 322–324), Wassinger et al. (421) used fMRI to describe a “core” region involved in the perception of pure tones surrounded by “belt” regions, which are selectively activated in humans by sounds with greater spectral complexity. This distinction between core and belt regions mirrors the organization of auditory cortical areas in nonhuman primates probed by recordings of neural activity in the superior temporal cortex. Considering the rostral part of this system, Rauschecker and Tian (324) identified small areas in the anterior lateral belt region which respond selectively to specific “meaningful” stimuli such as species-specific vocalizations. In humans, Zatorre et al. (433) recently showed that the right anterior superior temporal sulcus is sensitive to auditory object distinctiveness, for example, the ability to identify the characteristic sound of a trumpet independent of the produced melody.

In the language domain, Scott and Johnsrude (356) have proposed, on the basis of a detailed meta-analysis of
neuroimaging in humans, that a “rostral” pathway, running from the posterior middle part of the superior temporal cortex towards the temporal pole, would map low-level speech features onto lexical representations (see also Ref. 30). In line with this hypothesis, studies by Wise and colleagues (for a review, see Ref. 424) emphasized the importance of the portion of the left superior temporal gyrus located anterior to Heschl’s gyrus for speech comprehension. Using a speech deterioration paradigm, the authors implicated the anterior part of the left superior temporal gyrus (a component of the rostral pathway) in the interpretation of intelligible speech samples (355). Nevertheless, a specific involvement of the posterior STG in the same processes cannot yet be discarded (see below and Ref. 262).

Studies of posterior belt sites in animals have been linked to sound localization (for a review, see Ref. 324) rather than identification. Another, related, role of this caudal region is the coding of spectrottemporal cues relating to sound source motion (416). Results in monkeys (299) suggest the existence of a large overlap in the upper bank of the superior temporal sulcus between the motion-related visual stream and an auditory domain responding to various sounds including motion-related ones. In humans, recent imaging studies have shown that specific portions of the posterior belt are activated by detection of sound localization (221) and sound motion (14, 416). It should be noted that the neural network identified in these experiments not only involves the posterior end of the auditory cortex but also the inferior parietal cortex, notably in the right hemisphere.

Even so, the relationships between fundamental auditory processing and speech perception per se in the periphery of the primary auditory cortex remain difficult to characterize. One major challenge is to provide a theoretical account of why and how spatial acoustics and sound motion processing relate to speech-specific processing in the caudal part of the human belt region. Indeed, a number of neuroimaging studies have reported activations in the caudal part of auditory association cortex for processing various types of verbal and nonverbal auditory cues (29, 72, 143), although Griffiths and Warren (143) reported more speech-related than non-speech-related activities in the left planum temporale, an important section of the caudal auditory cortex. Whereas speech-selective activations appear to concentrate in the rostral component of the auditory system in both hemispheres (e.g., Ref. 30), lesions of the classical “Wernicke’s area,” i.e., the caudal part of the left superior temporal gyrus, have long been associated with phonological deficits, especially in the context of repetition tasks (12, 196). Moreover, recent results from cortical stimulation tend to show that phoneme perception can be disrupted by stimulation of the caudal component (36). A similar trend emerges from the relatively rare studies of aphasia in which the anatomy of lesions inducing phonological perception deficit was specifically addressed (61). The absence of lesion-based data corroborating the role of the rostral component of the auditory system in phoneme perception could however relate to a “vascular” bias, which would make lesions in the caudal part of the Sylvian territory following stroke more likely (161). It must also be kept in mind that posterior lesions can interrupt pathways conveying phonological information on their way from rostral sites to the inferior frontal cortex.

Overall, these results have shed new light on the classical aphasia-based model of auditory language comprehension. The posterior part of the left STG (formerly referred to as Wernicke’s area) seems to be functionally heterogeneous as it is activated by a variety of experimental conditions from phonological perception, access to lexical representations (166), monitoring the speaker’s own voice, to word retrieval from semantic memory (for a review, see Ref. 426). Wise and colleagues (426) distinguish two distinct subregions in the posterior left STG. The posterior part of the left superior temporal sulcus, which is equally activated by perception and generation of words, may represent a temporary buffer in which the sublexical components of a word could be transiently stored as the appropriate sequence of speech sound units (i.e., phonemes). These transient representations would be matched to the phonological form of words stored in lexical long-term memory. This left-lateralized component might be complemented by a homotopic region in the right hemisphere whose activity would depend on the frequency of the phonotactic structures in focus (225). Wise et al. (425) also described a small subregion, located dorsally to the superior temporal sulcus, at the junction between the superior temporal gyrus and the supramarginal gyrus, which is activated during articulatory speech movements. The authors suggested that this region could contribute to the interfacing of auditory speech representations with their motor counterparts. Therefore, speech information coded by the auditory system is likely to be further transcoded into motor speech acts, when corresponding signals are sent dorsally from posterior temporal regions to the inferior parietal cortex (especially the supramarginal gyrus) and, ultimately, to the inferior dorsal premotor cortex via the arcuate fasciculus (20, 25, 158). This transcoding process requires short-lived maintenance of speech representation via a system of phonological working memory.

Paulesu et al. (279) and Démonet et al. (105), in experiments addressing the neural substrates of phonological working memory, described specific activations located at the junction between the posterior superior temporal cortex and the inferior part of the supramarginal gyrus and proposed that this region harbors the neural substrates of the transient phonological store as defined in Baddeley’s model (see Ref. 8). This hypothesis does not
compete, but rather concurs, with the proposal of Wise et al. of a phonological buffer located more ventrally, i.e., in the left superior temporal sulcus. The specific requirements of phonological working memory tasks in which subjects are prone to activate the articulatory representations of phonological stimuli might well influence the localization of the activation peaks, tending to shift them towards dorsal subregions in the vicinity of the posterior temporal/inferior parietal junction.

2. Visual input

Reading is a skill acquired under cultural pressure, and it is recent in evolutionary terms. One may therefore wonder whether any brain area could be specifically dedicated to reading processes, from the integration of graphic features associated with phonemes or syllables to accessing word meaning. Even when restricting the matter to Western languages and adult subjects, the results of functional neuroimaging studies of reading (for review, see Refs. 119, 301) are characterized by their sensitivity to various experimental parameters, such as exposure duration, rate of stimulation, coverness of reading, and baseline tasks (17, 50, 155, 246, 301, 306, 312, 345). The impact of these factors on activation patterns makes the meta-analysis of neuroimaging studies of reading especially difficult.

After low-level perceptual analysis in the primary visual cortex, early processing of graphic stimuli elicits activation in the association visual cortex bilaterally, especially in its ventral and medial parts (81, 385). Petersen and co-workers (291, 292), who conducted seminal studies on the functional neuroimaging of reading, stressed the importance of the left medial extrastriate cortex (lingual gyrus), which is activated by words and pseudowords but not by consonant strings. Silent reading of words compared with false font viewing activates portions of the left posterior inferior aspect of the medial fusiform gyrus, the posterior part of the left superior temporal gyrus, and the cerebellum (301). Some authors (81, 94) claim that visual information resulting from low-level sensory analysis enters the “visual word form area” (VWFA) that they argue lies in the middle portion of the left fusiform gyrus, this being also one of the major lesion sites associated with severe alexia (31, 81; see also Ref. 234). Using visual hemi-field presentation, Binder and Mohr (31) demonstrated that this region responds to words and legal pseudo-words in contrast with consonant strings whatever the stimulated hemi-field, an effect not observed for striate cortex activations. Furthermore, this region shows repetition priming effects that are insensitive to changes in letter-case (95). In addition, the involvement of the left midfusiform region is proposed to start ~200 ms after stimulus onset time on the basis of ERP and MEG recordings (81, 385). Convergent results have been obtained from direct cortical recording of neural populations of the inferior temporal cortex in operated patients (266).

However, the specificity of this “VWFA” to words and even to visual input has been challenged and is now the subject of intense debate (303). Indeed, this region was not identified in one of the first studies which aimed at identifying the input orthographic lexicon, i.e., the PET study by Howard et al. (166), who contrasted reading words to saying “crime” when viewing false fonts. In this study, the authors instead found critical activations in the left posterior STG. Furthermore, Moore and Price (246) showed activation in the VWFA for both reading and naming objects versus baseline tasks and failed to find any differential activation when contrasting reading to object naming. More recently, Price and Devlin (303) have claimed that it is misleading to label this portion of the midfusiform gyrus as the VWFA because it is activated by a wide range of stimuli and tasks that do not require access to visual word form representations (see also Ref. 309). In particular, the hypothetical specificity of the VWFA seems to vanish whenever controlled cognitive tasks, especially semantic ones, are involved (see next section). The debate on the VWFA arose because the left midfusiform region is not only part of the visual ventral stream but also part, or a close neighbor, of the “basal language” area, which is activated by auditory or tactile stimulations (303). Price and Devlin (303) have therefore proposed three hypotheses regarding the functional role of the left midfusiform gyrus: 1) this region appears functionally multimodal because it harbors distinct subcomponents of different functional networks which are tightly intermingled and which challenge the resolution power of standard imaging techniques; 2) the function of this portion of cortex is indeed unique, but it is more complex, has yet to be determined, and cannot be limited to visual word form processing; and 3) the region embraces a multimodal neuronal population (convergence area), which participates in different cognitive functions depending on neural interactions with other distributed brain areas. One way to weight these hypotheses against one another is to examine the whole set of regions involved in reading experiments and to carry out large-scale statistical comparisons. Although difficult because of many confounding factors (see Table 2), meta-analysis has begun to emerge as a possible path toward functional modeling.

For instance, Jobard et al. (173) managed to analyze results from 35 PET and fMRI studies relevant to reading using a largely automatized procedure. While considering only left-sided foci of activation, the authors confirmed the absence of any reproducible site of brain activation that would respond more to words than to any nonlexical stimuli. Whatever the specificity of the so-called VWFA for processing written material, this region may still be
conceptualized as an “entry point,” providing access to lexical, semantic, and phonological associations of printed words. [“The role of this occipital temporal junction would be to segment, classify, and relay visual word information to other regions for further analysis” (173).]

Jobard et al. (173) also attempted to find out whether the two routes of reading described by the well-known dual-route models first described by Coltheart and colleagues (e.g., Ref. 84) could be substantiated as distinct patterns of brain activation. The main and important result of this work is the establishment of a link between the grapho-phonological route and the posterior part of the left STG and the left supramarginal gyrus. Moreover, distinct sites of activation were found in the left PIFG (“Broca’s area”), i.e., in the pars opercularis and the pars triangularis, respectively, the latter being interpreted by the authors as reflecting semantic access while reading and, hence, the direct, lexical semantic route.

Most neuroimaging studies of reading have also described selective activation in the superior/middle posterior temporal cortex (301). It is not clear at present whether this pattern of activation relates to lexicon access (166), phonological access, or both (119). A possible interpretation of the right superior temporal activation in reading aloud tasks relies on the fact that subjects can perceive their own voice while reading (301, 311).

Other perisylvian regions activate while reading, especially when grapheme-to-phoneme conversions and general phonological manipulations are encouraged. For instance, the supramarginal gyrus seems especially involved in studies dealing with pseudo-words or unfamiliar letter combinations (302). In reading aloud tasks, the left PIFG is systematically active, possibly in relation to a late processing stage before the articulatory stage is reached (see Refs. 302, 312). However, activation of the left PIFG has also been clearly established in silent and even implicit reading (301, 310). Therefore, reading conditions are likely to involve phonological retrieval even when there is no detectable speech output. Nevertheless, in the case of frequent, regular words, reading does not require precise phonological recoding, and phonological retrieval corresponds to direct and automatic access to their phonological form. In this condition, the neural network involved in reading (fusiform gyrus, the left STG, the supramarginal gyrus, and the left PIFG) might be modulated in a way that reduces activation in regions that are associated with phonological recoding (i.e., temporal and parietal areas).

Apart from the influence of word regularity within a given language, differences have been shown across languages. Paulesu et al. (282) showed that reading English requires preponderant access to a lexicon of orthographic patterns that activates the left fusiform gyrus and the left PIFG, whereas Italian reading, in which letter-to-sound conversion is the predominant process, specifically activates the left STG.

Japanese represents an important example of a language in which two different written codes coexist: Kanji (ideographic system) and Kana (syllabic system). Although previous behavioral studies suggested a specific involvement of the right hemisphere for processing Kanji, recent studies using neuroimaging techniques have found selective left-sided activation (e.g., Ref. 261). Sakurai et al. (348) recently stressed the specific participation of the left basal temporal language area for reading Kanji, whereas Kana processing recruited left temporal parietal regions.

3. Semantics

The exploration of semantic memory and semantic processes are at the heart of cognitive neuroscience dealing with language because the crucial function of language is to convey meaning between individuals.

A first line of research has been devoted to semantic comprehension tasks. The major result obtained demonstrated consistent involvement of a complex neural system involving the inferior temporal cortex, the middle and posterior temporal cortex (including the angular gyrus), and frontal association areas (e.g., Refs. 24, 105, 301, 409; see Figs. 6 and 7). This large system extends widely in the left hemisphere, and the contribution of right hemispheric cortical areas appears to be fairly limited (105). In experiments devoted to mathematical semantic knowledge, however, the right intraparietal sulcus seemed as intensely involved as its left homolog (97). This extra-sylvian system includes both the “entry points” described previously (i.e., the left STG for auditory processing and the posterior part of the left fusiform gyrus for visual processing) and “core semantic regions” whose interplay is likely to implement semantic processing directly (89). The core regions include the posterior superior temporal/inferior parietal cortex (BA 39), the “basal temporal language area” (mainly corresponding to BA 19, BA 37, and BA 20), and some parts of the lateral frontal cortex (BAs 44, 45, and 47). Whereas the role of the basal temporal area in semantic processes is clearly established, the involvement of the inferior frontal regions is less clear and has been disputed (e.g., Refs. 131, 307, 401).

The basal temporal language area is argued to implement the ability to organize the perceived world into distinct categories, as lesions in this part of the brain induce category-specific deficits (91). Patients with damage to this region are selectively impaired for retrieving semantic information relative to a particular category of objects, for instance, animals, whereas semantic knowledge of other categories is largely preserved. Damasio et al. (91) performed a large study combining analysis of lesion topography in patients and activation loci in normal volunteers with a focus on the temporal lobe. The authors claimed that both approaches pointed to an an-
terior-to-posterior organization of the inferior temporal lobe with respect to different semantic categories (people, animals, and tools). However, these results were modulated by other studies in which extended networks were found active for different semantic categories rather than localized monofocal activations (e.g., Refs. 230, 287).

Other studies addressing the topic of semantic categories suggest that the perceptual features of objects (e.g., function or color), which participants are encouraged to retrieve in task-specific contexts, are the key factors orienting activations towards different perceptual poles in the brain (e.g., Refs. 229, 251). Neural territories that activate for retrieval of corresponding lexical representations have been shown to stand a few millimeters around the areas known to be involved in perceptual processing. However, more recent exploration using both PET and fMRI (111) has shown that category-specific differences in functional anatomy are totally cancelled out when semantic categories are accessed on the basis of highly controlled lexical stimuli (that is, controlled for familiarity, concreteness, lexical frequency, neighborhood sizes, number of letters, and number of syllables).

The basal temporal language area in fact appears to be highly heterogeneous and may be composed of many different subparts, each of them modality specific but closely intermixed (137). This amalgam implies a multimodal or even amodal pattern of response (see Ref. 52 for an example of multimodal response and Ref. 56 for discussion). As far as language is concerned, this region could be viewed as an interface between at least two cognitive realms, dedicated to (visual) perception and language, respectively. Although activity in the different subparts is likely to be modulated in a complex manner, the weight of perceptual processes seems higher in the posterior part of the region (proximity of the striate cortex), whereas a reverse gradient would apply to language processes leading to lexical retrieval (246; see also Ref. 167).

Lexical semantic categorization cannot be restricted to the classification of real objects. Indeed, category-specific effects have been observed for higher order classification across words, namely, abstract versus concrete items or grammatical categories (see Refs. 17, 181). The few neuroimaging studies that have dealt with the neural basis of abstract word processing have commonly implicated the right temporal cortex (144, 181, 288) in addition to a large left-sided pattern of activation. This result still awaits clear interpretation and could merely reflect higher cognitive demand for processing abstract, i.e., less familiar words. Given behavioral data indicating that the right hemisphere is particularly unskilled at processing abstract words, supplementary activations in the right hemisphere are difficult to interpret.

Some parts of the left frontal lobe have been repeatedly activated in studies using semantic tasks (e.g., Ref. 131; for an exception, see Ref. 307). The question is to determine which components of these semantic tasks can be attributed to the frontal areas. Most of these studies involved production tasks (as opposed to comprehension tasks), such as stem completion, word fluency, or word generation. In this kind of task, the left PIFG has been associated with the selection of relevant features of semantic knowledge from a set of competing alternatives (401). However, it could be that only the inferior parts of the left PIFG (i.e., BA 45, and BA 47) are implicated in this function, whereas the superior part (BA 44) would be more involved in sublexical aspects of such tasks (131).

Word categorization in terms of grammatical class could also yield specific functional parcellation of language-related neural structures. Lesion-based studies have suggested that the left frontal cortex is associated with processing verbs (as lexical entities denoting mainly action), whereas temporal cortex lesions more specifically affect object names. Functional neuroimaging studies have attempted to verify this assumption, but again the results have not been as straightforward as expected. For instance, Perani et al. (288) showed a specific activation of the inferior part of the left PIFG and the left middle temporal gyrus when a lexical decision was performed on verbs compared with nouns. Their proposed interpretation relates to the dual processing of verbs for which semantics per se would be processed in the temporal regions, whereas the syntactic dimension would be implemented by the frontal cortex. The relationship between verb processing and neural activities in the frontal cortex has also been substantiated by rTMS studies showing specific effects of repetitive magnetic stimulation of this region on production of either verbs, pseudo-verbs (359), or action names (64).

Our understanding of the neural basis for action representation has benefited greatly from the discovery of “mirror” neurons in monkeys. Mirror neurons, found in the frontal cortex homotopic of the human PIFG in monkeys, respond to the programming of an action and the observation of similar actions performed by other individuals in a selective fashion (112). Links between linguistic representation of actions and neural activities in the prefrontal cortex have been reinforced, from a physiological point of view, by convincing evidence of the existence of similar neurons in the human premotor cortex (e.g., Ref. 340; for a discussion, see Ref. 328). Rizzolatti and Arbib (328) have speculated that gesture-based communication between individuals and its neural substrate might represent an archaic system from which the human linguistic code has developed. Congruent with this hypothesis, a recent fMRI study has shown an overlap of activated clusters in the left inferior frontal cortex during perception of action and verb generation tasks (150).
4. Speech output

The process by which meaning is turned into speech involves a phonological stage (i.e., the retrieval of the word form to be pronounced; for a review, see Ref. 171). Although it is difficult to conceive experiments that would permit the correlates of the various subcomponents to be separated out (see Ref. 296 for a critical review; for discussions, see Refs. 104, 297), Price and Friston (304) proposed a conjunction study of four different naming tasks that share speech output processes. Apart from structures involved in visual perception, they showed the involvement of two main structures: the left basal temporal area and a large cluster encompassing the left anterior insular cortex and the opercular frontal cortex. One may note the absence of both the left STG and the left PIFG in this pattern, as also shown by Etard et al. (117) who proposed a “direct” pathway for naming. These findings are partly confirmed by Wise et al. (425), who stressed the role of the anterior insular cortex for both listening and repeating words and found no activation in the left PIFG for a speech output task. The right cerebellum, although again located outside the classical language areas, also appears to be linked with articulatory levels of speech (1), even though cerebellar activation has typically been associated with more central cognitive components, such as the sustained search for lexical information (109, 131).

5. Written output

The neural correlates of actual written output have been explored in very few studies, unlike studies of speech output. Rinjties et al. (327) approached this topic in an avant garde study of hand versus foot signature. The authors showed a common response in secondary premotor and parietal cortices, whatever the limb involved.

Due to the interest of the coexistence of two writing codes, several studies of the Japanese language have been recently published showing, in general, that the particular complexity of the ideographic Kanji code is associated with extra processing demand, especially in the left inferior temporal cortex (231). These studies, including a recent one dealing with the writing of English (239), have also described a left functional network that particularly involves the superior parietal lobule and the inferior posterior frontal cortex (179, 232).

B. Sentence and Discourse Processing

Connected language processing (fluent speech and written text) has been explored in neuroimaging and electrophysiological studies using various comprehension tasks. Early attempts using ERPs focused on the physiological consequences of semantic expectation violation, i.e., the presentation of a spoken or written word that did not fit the global meaning of a sentence (201). Subsequently, authors have focused on the differences between the processing of meaning and the processing of syntactic information. Finally, several recent neuroimaging studies have addressed the delicate issue of discourse and global semantics in an attempt to characterize language processing in more “natural” conditions.

1. Sentence context and semantics

Ever since the fundamental electrophysiological experiment by Kutas and Hillyard (201), lexical semantic violations in a sentence context have been known to elicit a negative ERP component peaking ~400 ms after SOT on average, the N400. Although it is now accepted that this component is not specific to the processing of language (202), it clearly indexes the cognitive demand incurred by the integration of a meaningful stimulus (such as a word) into a more general semantic context (such as a sentence).

The N400 is sensitive to various properties of words, such as lexicality (300), concreteness (194), and typicality (379), but it shows a relative (and still debated) independence vis-à-vis lexical frequency (379). Interestingly, the N400 is not an on/off response. When incongruent words that relate in meaning to the best completion word are presented at the end of sentences, the N400 is significantly reduced compared with fully incongruent conditions. Federmeier and Kutas (118) have shown that this graduated response, which is observed for words presented to the right visual field (left hemisphere), is not seen when words are presented in the left visual field (right hemisphere). They interpreted this as a sign of a predictive processing trend in the left hemisphere (predictions are made on the basis of the semantic context) contrasting with an integrative processing trend in the right hemisphere (conclusions are made on the basis of context reanalysis).

2. Dissociating semantics and syntax

Processing linguistic information embedded in sentences brings in specific processes concerned with the order in which words are perceived and the rules that govern this order. Early electrophysiological studies demonstrated that syntactic processes are independent from semantic processing to some extent because syntactic violation elicits positive ERP components that are very different from the N400. The P600 or “syntactic positive shift” (SPS, e.g., Ref. 253) is elicited upon presentation of a word that is grammatically incorrect or when sentences are made abnormally complex. By varying the locus of syntactic anomalies, it has been possible to subdivide syntactic processing into different phases: 1) early morphosyntactic processes (such as syntactic parsing) reflected by a left anterior negativity (ELAN), and 2) highly
controlled reanalysis of the syntactic structure and repair processes reflected by the late positivity (P600/SPS) that is more centrally distributed over the scalp (123, 146).

Tomographic techniques have been extensively used to elucidate the anatomical basis of the cognitive mechanisms revealed by electrophysiology. Although the involvement of the inferior frontal cortex in syntactic processing has long been hypothesized (58), lesion studies of aphasic patients with agrammatism failed to delineate clearly the crucial brain regions involved in syntax processing. Caplan (62), in particular, stressed the variety of brain lesions associated with syntax comprehension impairments. Most functional neuroimaging studies have shown the involvement of the left inferior frontal region and/or adjacent areas as the neural substrates of syntax processing, either for comprehension or for production (e.g., Refs. 59, 60). However, whether such activations reflect highly specific grammatical/syntactic processing or orthogonal working memory involvement remains to be determined (396).

In an attempt to segregate activations relating to semantic processing from those induced by syntactic processing, Dapretto and Bookheimer (92) manipulated semantic and syntactic complexity independently in a $2 \times 2$ factorial design. They proposed a specific role for two different focal portions of the left inferior frontal cortex: 1) a ventral one (pars orbitalis) associated with semantic judgment and 2) a dorsal one (pars opercularis) associated with syntactic judgment. A similar study replicated the anatomical dissociation for syntax and semantics using a more implicit task (265). Ni et al. (265) presented their volunteers with sentences featuring semantic or grammatical errors with no instruction to attend to anomalies and found specific activation in the left PIFG for grammatical errors. Caplan et al. (60) addressed the question of the syntactic specificity of activations in the left PIFG, given its major role in verbal working memory. They demonstrated that syntactic activation of BA 45 was obtained for syntactic processing whether participants were involved in interfering repetitive utterance of a word or not, suggesting independence vis-à-vis verbal memory. An important aspect of syntax processing, the production of sentences, has not been studied as widely as comprehension. Indefrey et al. (170) described a graded activation localized in the left rolandic operculum, adjacent to the classical Broca’s area rather than the left PIFG itself, when subjects produced word sequences, noun phrases, and full sentences.

Musso et al. (256) recently involved monolingual subjects in the learning of the grammatical rules of two foreign languages in two different experiments. Increased activation in the pars triangularis of Broca’s area was significantly correlated to the accuracy of learning performance but only when learning was based on the principles of universal grammar (79) as opposed to arbitrary rules created by experimenters.

Overall, variation in syntax complexity can modulate activity in a much larger neural network than the classical Broca’s area, including the left PIFG, the posterior part of the left STG, and to a lesser extent, their right counterparts (176). Moro et al. (247), for instance, have replicated selective activation of the left PIFG and right prefrontal regions for syntactic (and morphosyntactic) processing versus semantic processing, but they also found selective activation of the left caudate nucleus and insula.

3. Discourse-level processing

Beyond the mechanisms of detailed sentence structure analysis, a major goal is to characterize the neural architecture involved in connected speech, which is the natural condition of language perception and production. Early studies involving story listening (e.g., Ref. 233) attempted to identify subparts of the language network that are specific to coherent, connected language samples. The authors reported the involvement of the anterior polar aspects of the superior temporal gyrus. Since then, various studies (168, 224, 355) have confirmed the crucial involvement of this portion of the brain for processing discourse. In addition to anterior temporal activation, Maguire et al. (224) found activation in the medial parietal posterior cingulate cortex and proposed that the latter was involved in linking the current understanding of a story with prior knowledge. As mentioned before, Scott et al. (355) have shown that a “rostral” region of the left superior temporal sulcus was selectively activated by “intelligible” stimuli, intelligibility being manipulated by degrading understandable speech (e.g., noise-vocoded voice samples). This study provides major evidence for the existence of a ventral or anterior “what” stream. In an experiment involving words and environmental sounds in series, Thierry et al. (394) also found activation in the superior medial frontal cortex (BA 8) when contrasting attempts to elucidate the overall meaning of the series (i.e., attempts to make up a story) with semantic categorization of single items (i.e., dealing with words and sounds individually; unpublished results). BA 8 has been shown to take part in planning strategies when subjects are required to make plans that are endogenous to the task (189) and might be crucially involved in discourse level comprehension.

In an elegant study St. George et al. (378) explored the neural basis for processing global discourse coherence. They presented a text sample word by word and compared a condition in which coherence was prompted by a title with a condition in which no title was given. They found a crucial involvement of the right middle temporal sulcus in the untitled condition, which was
meant to be more demanding in terms of coherence extraction.

Other studies have investigated the physiology of indirect sources of meaning in language. Prosody, for instance, conveys supralexical semantic information such as interrogative versus assertive mode or emotional connotation (happiness, anger, sadness, etc.). The few studies addressing this question (e.g., Refs. 51, 134) have reported activation in the right hemisphere, especially in the right inferior frontal gyrus, for the processing of emotional prosody versus neutral intonation. This finding is remarkably congruent with neuropsychological studies of aprosodias (338, 339).

Meaningful information is also modulated in daily life by the connotative dimension of discourse (e.g., “March came in like a lion”). Comparing metaphorical sentences to matched literal sentences, Bottini et al. (44) implicated the right hemisphere when processing metaphors, a finding which is again consistent with studies of patients with damage to the right hemisphere (49).

Finally, very few studies have focused on the production of connected speech. The first results obtained so far show that production of propositional speech activates a number of left hemisphere regions remote from the classical perisylvian “language area” (see, for instance, Refs. 33, 45).

C. Beyond the “Standard” Language

One drawback of the generalization power expected from neuroimaging lies in the fact that it implies focusing primarily on “normal” populations. However, many individuals do not fit the standard profile of right-handed young subjects with normal visual and auditory skills, mastering a single idiom and a single alphabetical written code. A number of neuroimaging studies have focused on the effects of nonstandard individual factors on the representation of language in the brain.

1. Illiteracy

Although it has had a major impact on the development of civilization, literacy is a new skill in evolutionary terms, and it is far from being universal. However, its impact on language organization in the brain appears to be far reaching. For instance, in a pioneering study, Castro-Caldas et al. (70) demonstrated that illiterate subjects activated similar brain areas when repeating words and nonwords, whereas literate matched controls did not. In a further analysis using structural equation modeling, Petersson et al. (293) showed that the weight of functional connectivity between subparts of the perisylvian network differed across illiterate individuals but not across controls. This result suggested variable attentional modulation in the network when illiterate subjects attempted to repeat pseudo-words.

2. Multilingualism

Mastering more than a single language is a common and socially important feature in human societies; functional neuroimaging has been used to explore the neural counterparts of this ability. Two main questions have been addressed: 1) Are two different languages implemented through a common neural network or several spatially segregated networks in the brain? 2) What are the determinants of cerebral organization of language in bilinguals?

Kim et al. (183) showed that first language (L1) and second (L2) language are represented in segregated portions of the left inferior frontal gyrus in late bilinguals (i.e., individuals who have acquired a second language in adolescent years) while a major overlap can be seen in subjects who have been exposed to both L1 and L2 in early childhood. However, Kim et al. (183) used a covert verbal recall of recent autobiographic episodes and further studies on language production using single-word tasks (e.g., word generation and fluency) failed to replicate these findings. While no differences in brain activation were found in very early bilinguals (73, 156), a common neural network was also identified in late bilinguals when participants were highly proficient in both languages (74, 186). Yetkin et al. (430) suggested that a less frequently used language yielded larger cerebral activations than a more regularly spoken language. These findings suggest that both attained proficiency and duration of exposure to a second language might be more important factors than age of acquisition to account for the cerebral representation of language networks in bilinguals. Nevertheless, Perani et al. (286) very recently proposed that the pattern of brain activation in high-proficient bilinguals might be modulated by both age of acquisition and language exposure.

Studies of language comprehension have given rise to less variable results. In highly proficient early bilinguals, a single and common neural network appears to be involved in the processing of the two languages (73, 290). For late bilinguals, a critical factor is the degree of language proficiency: highly proficient late bilinguals display a pattern of activation similar to that observed in early bilinguals, while a much more variable pattern is found in moderately proficient late learners (93, 289).

Translation, or simultaneous processing of both languages, represents an especially sophisticated skill that has been studied only recently. Price et al. (305) used reading aloud and translation tasks involving either homogeneous conditions (L1, German or L2, English) or alternating conditions (i.e., both languages in the same run). Translation elicited activation in the anterior cingu-
late cortex and subcortical structures bilaterally, thought to relate to greater demands on coordination and inhibition of parallel lexical processing. In addition, language switching was associated with activation in the left inferior frontal gyrus area and the supramarginal gyri, two regions which are well-known for their role in phonological processing. Phonological processing might also be enhanced in highly proficient bilinguals, witnessing indirect access to the lexicon of a target language. When required to ignore one of two languages, bilingual subjects have been shown to activate the left posterior inferior frontal cortex and the left planum temporale, a pattern suggesting that subjects resorted to a presensory phonological selection of the target language (333).

Dynamic effects on the functional anatomy of language in bilinguals also involve attrition effects observed when subjects cease to use one of their languages. Such modifications can be observed even in the case of individuals renouncing their mother tongue. For instance, Pallier et al. (275) showed that Korean children adopted by French families forgot their mother tongue and exhibited activation patterns no different from those observed in native French speakers when perceiving Korean.

3. Sensory deficits

One can speculate that sustained auditory or visual deprivation of linguistic input from birth should have an impact on language-related neural organization. Most studies involving deaf patients using American Sign Language (ASL) reveal that activations in the left hemisphere are very similar for ASL and oral language, although ASL recruits right perisylvian regions in addition to the classical left-sided areas (15, 264).

Patients benefiting from cochlear implants show modified patterns of activation. In such patients, regions involved in attention and sublexical processing are overactivated while association cortices relating to semantic processing tend to be less active (136). Moreover, Giraud et al. (135) recently highlighted the participation of early visual areas in perceptual processes involving fine-grained phonemic distinctions in implanted deaf patients.

In blind patients, functional neuroimaging of Braille reading emphasized cross-modal effects showing specific activation in the visual areas in blind participants performing tactile reading (347). These findings suggest that, in spite of visual deprivation, the visual cortex can be recruited for reading from tactile input and that this could even be indispensable for correct Braille reading (83). Furthermore, the inferior temporal region (BA 37), which has been hypothesized to implement visual word form processing, is not exclusively dedicated to higher visual processing. It may be recruited by other sensory systems as it is selectively activated in blind subjects during Braille reading (52). Plasticity effects for spatial auditory localization have also been found in blind subjects with activations in parts of the dorsal visual system during auditory perception (418).

D. Outstanding Questions

This section has attempted to provide key examples of the various (and ever multiplying) approaches to the neurophysiology of language. In spite of their invaluable contribution to a renewed neuroscience of language, numerous experiments conducted in the field have generated unexpected results that can be interpreted only with difficulty, if at all, while fundamental questions remain unanswered. A central problem in the neuroimaging approach to language is whether imaging results are specific to language processing, and consequently, whether part of this higher order function can be reduced to more basic and physiologically decipherable mechanisms. In this section, we attempt to sketch three major directions for future research: 1) the ever fascinating question of hemispheric predominance for language; 2) the hypothetical functional segregation between a ventral and a dorsal processing pathway; and 3) the opposition between controlled and automatic processes bridging language to global cognitive functions such as attention and memory.

1. Hemispheric predominance for language

Language activation limited to the left hemisphere is rarely observed. In fact, a substantial number of experiments performed in right-handed subjects have shown bilateral activation without asymmetry (e.g., the bitemporal neural responses reported for speech signals, at least in passive conditions; see sect. IV A1 and Table 3). Although this observation has been made repeatedly (105, 124), it is not overtly discussed, perhaps because it does not fit the classical left-greater-than-right asymmetry model. A nonexhaustive qualitative survey indicates that weak (or nonexistent) left-greater-than-right asymmetry tends to be associated with 1) auditory input rather than visual input, 2) figurative rather than symbolic types of stimuli, 3) single-word rather than sentence or syntax processing, and 4) comprehension rather than production tasks.

The visual confrontation naming task, for example, is one of the most widely used tests to detect and assess aphasia in clinical practice. Any damage caused to the left hemisphere, at least at the acute stage, impairs the ability to retrieve object names to some extent. Such disorders are only rarely observed following right hemispheric lesions in right-handed patients. In spite of its well-established clinical robustness, this task elicited weak functional asymmetry in most neuroimaging experiments (see Table 3). Corresponding activation was repeatedly found in the right homolog of the left PIFG and in the right...
Table 3. Synopsis of activations found in neuroimaging studies of confrontation naming tasks

<table>
<thead>
<tr>
<th>Author (Year)</th>
<th>Reference No.</th>
<th>BA/Region</th>
<th>Left Hemisphere</th>
<th>Right Hemisphere</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>x  y  z  Z score</td>
<td>x  y  z  Z score</td>
<td></td>
</tr>
<tr>
<td><strong>Occipital</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kosslyn et al. (1994)</td>
<td>193</td>
<td>17</td>
<td>−6 −79 0</td>
<td>3.36</td>
</tr>
<tr>
<td>Kosslyn et al. (1995)</td>
<td>192</td>
<td>17</td>
<td>−36 −68 −16</td>
<td>4.1</td>
</tr>
<tr>
<td>Bookheimer et al. (1995)</td>
<td>42</td>
<td>17</td>
<td>12 −83 −4</td>
<td>3.87</td>
</tr>
<tr>
<td>Menard et al. (1996)</td>
<td>230</td>
<td>17</td>
<td>12 −82 0</td>
<td>3.5</td>
</tr>
<tr>
<td>Martin et al. (1996)</td>
<td>230</td>
<td>17</td>
<td>2 −84 8</td>
<td>2.92</td>
</tr>
<tr>
<td>Kosslyn et al. (1994)</td>
<td>193</td>
<td>18</td>
<td>24 −87 4</td>
<td>3.51</td>
</tr>
<tr>
<td>Menard et al. (1996)</td>
<td>237</td>
<td>18</td>
<td>−26 −78 −4</td>
<td>3.81</td>
</tr>
<tr>
<td>Murtha et al. (1996)</td>
<td>254</td>
<td>S/MOG (18)</td>
<td>−16 −92 10</td>
<td>28 −87 16</td>
</tr>
<tr>
<td>Moore et al. (1999)</td>
<td>246</td>
<td>18</td>
<td>−8 −68 −8</td>
<td></td>
</tr>
<tr>
<td>Murtha et al. (1999)</td>
<td>255</td>
<td>18</td>
<td>−9 −95 5</td>
<td></td>
</tr>
<tr>
<td>Bookheimer et al. (1995)</td>
<td>42</td>
<td>18/19</td>
<td>−22 −78 0</td>
<td>3.37</td>
</tr>
<tr>
<td>Murtha et al. (1999)</td>
<td>255</td>
<td>18/19</td>
<td>22 −82 −16</td>
<td>3.51</td>
</tr>
<tr>
<td>Sergent et al. (1992)</td>
<td>357</td>
<td>MOG (19)</td>
<td>−40 −76 −6</td>
<td></td>
</tr>
<tr>
<td>Kosslyn et al. (1994)</td>
<td>193</td>
<td>19</td>
<td>−26 −78 20</td>
<td>22 −81 40</td>
</tr>
<tr>
<td>Kosslyn et al. (1995)</td>
<td>192</td>
<td>19</td>
<td>−36 −82 20</td>
<td>44 −68 24</td>
</tr>
<tr>
<td>Murtha et al. (1996)</td>
<td>255</td>
<td>19</td>
<td>−36 −87 −18</td>
<td></td>
</tr>
<tr>
<td>Murtha et al. (1996)</td>
<td>254</td>
<td>LG (30)</td>
<td>−13 −64 12</td>
<td>3.22</td>
</tr>
<tr>
<td><strong>Temporal</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sergent et al. (1992)</td>
<td>357</td>
<td>FFG (37)</td>
<td>−33 −22 −12</td>
<td>2.33</td>
</tr>
<tr>
<td>Kosslyn et al. (1994)</td>
<td>193</td>
<td>FFG (37)</td>
<td>−30 −48 −12</td>
<td>5.09</td>
</tr>
<tr>
<td>Kosslyn et al. (1995)</td>
<td>192</td>
<td>FFG (37)</td>
<td>−32 −46 −16</td>
<td>4.35</td>
</tr>
<tr>
<td>Bookheimer et al. (1995)</td>
<td>42</td>
<td>FFG (37)</td>
<td>−28 −58 −16</td>
<td>42 −44 −12</td>
</tr>
<tr>
<td>Martin et al. (1996)</td>
<td>230</td>
<td>FFG (37)</td>
<td>−35 −60 −10</td>
<td>43 −61 −16</td>
</tr>
<tr>
<td>Kanwisher et al. (1997)</td>
<td>178</td>
<td>FFG (37)</td>
<td>−40 −73 −12</td>
<td>44 −68 −13</td>
</tr>
<tr>
<td>Murtha et al. (1996)</td>
<td>254</td>
<td>FFG (37)</td>
<td>−44 −54 14</td>
<td></td>
</tr>
<tr>
<td>Moore et al. (1999)</td>
<td>246</td>
<td>FFG (37)</td>
<td>−40 −74 −11</td>
<td>44 −68 −14</td>
</tr>
<tr>
<td>Murtha et al. (1999)</td>
<td>255</td>
<td>37, 19</td>
<td>−40 −74 −11</td>
<td></td>
</tr>
<tr>
<td>Sergent et al. (1992)</td>
<td>357</td>
<td>ITG</td>
<td>−55 −39 −17</td>
<td></td>
</tr>
<tr>
<td>Kosslyn et al. (1995)</td>
<td>192</td>
<td>IT</td>
<td>−48 −32 −16</td>
<td>3.98</td>
</tr>
<tr>
<td>Damasio et al. (1996)</td>
<td>91</td>
<td>IT</td>
<td>−37 −35 −15</td>
<td></td>
</tr>
<tr>
<td>Menard et al. (1996)</td>
<td>237</td>
<td>IT (19)</td>
<td>−36 −65 −4</td>
<td>3.79</td>
</tr>
<tr>
<td>Sergent et al. (1992)</td>
<td>357</td>
<td>MTG (21)</td>
<td>−53 −9 −11</td>
<td></td>
</tr>
<tr>
<td>Bookheimer et al. (1995)</td>
<td>42</td>
<td>MTG</td>
<td>−56 −32 0</td>
<td>2.53</td>
</tr>
<tr>
<td>Menard et al. (1996)</td>
<td>237</td>
<td>MTG</td>
<td>−41 −46 8</td>
<td>3.23</td>
</tr>
<tr>
<td>Kosslyn et al. (1995)</td>
<td>192</td>
<td>MTG</td>
<td>−47 −27 −8</td>
<td>3.42</td>
</tr>
<tr>
<td>Bookheimer et al. (1995)</td>
<td>42</td>
<td>AST</td>
<td>−50 −12 4</td>
<td>2.39</td>
</tr>
<tr>
<td>Murtha et al. (1996)</td>
<td>254</td>
<td>20</td>
<td></td>
<td>36 −18 −34</td>
</tr>
<tr>
<td>Moore et al. (1999)</td>
<td>246</td>
<td>20</td>
<td>−38 −34 −14</td>
<td></td>
</tr>
<tr>
<td>Damasio et al. (1996)</td>
<td>91</td>
<td>TP</td>
<td>−30 0 −31</td>
<td></td>
</tr>
<tr>
<td>Murtha et al. (1999)</td>
<td>255</td>
<td>22, 42</td>
<td>−42 −40 3</td>
<td>55 −25 3</td>
</tr>
<tr>
<td><strong>Parietal</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sergent et al. (1992)</td>
<td>357</td>
<td>7</td>
<td>−32 −60 60</td>
<td></td>
</tr>
<tr>
<td>Kosslyn et al. (1994)</td>
<td>193</td>
<td>7</td>
<td>−24 −71 40</td>
<td>3.08</td>
</tr>
<tr>
<td>Kosslyn et al. (1995)</td>
<td>192</td>
<td>7</td>
<td>−20 −71 40</td>
<td>2.84</td>
</tr>
<tr>
<td>Sergent et al. (1992)</td>
<td>357</td>
<td>40</td>
<td>−53 −18 18</td>
<td>52 −23 32</td>
</tr>
<tr>
<td>Kosslyn et al. (1995)</td>
<td>192</td>
<td>IPG</td>
<td>−35 −50 48</td>
<td>3.57</td>
</tr>
<tr>
<td><strong>Frontal</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kosslyn et al. (1994)</td>
<td>193</td>
<td>FEF</td>
<td>−13 21 40</td>
<td>3.13</td>
</tr>
<tr>
<td>Murtha et al. (1996)</td>
<td>254</td>
<td>FEF</td>
<td>−4 15 45</td>
<td></td>
</tr>
<tr>
<td>Kosslyn et al. (1995)</td>
<td>192</td>
<td>FEF</td>
<td>−10 33 48</td>
<td>3.33</td>
</tr>
<tr>
<td>Kosslyn et al. (1994)</td>
<td>193</td>
<td>10</td>
<td></td>
<td>24 58 0</td>
</tr>
<tr>
<td>Bookheimer et al. (1995)</td>
<td>42</td>
<td>IFG (47)/I</td>
<td>−34 14 4</td>
<td>5.02</td>
</tr>
<tr>
<td>Moore et al. (1999)</td>
<td>246</td>
<td>47</td>
<td></td>
<td>42 26 12</td>
</tr>
<tr>
<td>Martin et al. (1996)</td>
<td>230</td>
<td>IFG</td>
<td>−28 16 8</td>
<td></td>
</tr>
<tr>
<td>Murtha et al. (1996)</td>
<td>254</td>
<td>IFG</td>
<td>−50 24 21</td>
<td></td>
</tr>
<tr>
<td>Kosslyn et al. (1995)</td>
<td>192</td>
<td>MFG</td>
<td>−33 44 4</td>
<td>4.27</td>
</tr>
<tr>
<td>Kosslyn et al. (1994)</td>
<td>193</td>
<td>AC</td>
<td></td>
<td>7 34 24</td>
</tr>
<tr>
<td>Bookheimer et al. (1995)</td>
<td>42</td>
<td>AC</td>
<td>−2 18 36</td>
<td>4.4</td>
</tr>
<tr>
<td>Murtha et al. (1999)</td>
<td>255</td>
<td>AC</td>
<td>−1 13 44</td>
<td></td>
</tr>
<tr>
<td>Sergent et al. (1992)</td>
<td>357</td>
<td>GR</td>
<td>−2 21 −19</td>
<td></td>
</tr>
<tr>
<td>Murtha et al. (1996)</td>
<td>254</td>
<td>GR (25)</td>
<td></td>
<td>3 8 −16</td>
</tr>
</tbody>
</table>

MOG, medial occipital gyrus; MTG, medial temporal gyrus; TP, temporal pole; FFG, fusiform gyrus; LG, lingual gyrus; FEF, frontal eye field; GR, gyrus rectus; IFG, inferior frontal gyrus; MFG, medial frontal gyrus; I, insula; AC, anterior cingulate; IPG, inferior parietal gyrus. Further activations (not reported in the table) were found in the motor cortex (42, 246), the thalamus (42, 230, 255), and the cerebellum (230, 254, 255). Note the frequent involvement of bilateral structures (178, 192, 237, 357).
temporal association cortex. In contrast, more demanding production tasks such as verb generation tend to elicit more activation in the left hemisphere than the right. This shift seems to relate to greater computational and attentional demands.

Considering the sensitivity and specificity of anomia in the case of left hemispheric lesions, one may speculate that right-sided activation during naming does not directly relate to naming performance in healthy right-handed subjects. As proposed by Price et al. (308; see sect. IV), functional neuroimaging of language in normal subjects may show a set of brain regions whose activation is “sufficient” but not “necessary” to perform the task of interest. In other words, the “lesion model” has the unique ability to reveal the necessary (left-sided) regions (see Table 1).

Activations observed in the right hemisphere during naming tasks must nevertheless have functional significance. They could reflect persistent, automatic activation in a formerly active system, whose function would have been competitively diminished by the preponderant activities of the left-sided system. Although studies of split-brain patients have shown that, in isolation, the right hemisphere has limited naming capacities, such a residual system could partially compensate for the effects of left-sided lesion and aphasia, at least at some stages of postlesional recovery (see sect. nB).

In sum, the predominance of the left hemisphere for language, an idea that has been the fundamental cornerstone of aphasiology for more than a century, still awaits contemporary validation. Future studies will hopefully elucidate the empirical conditions and levels of language processing which bias language activation towards the left hemisphere and those which trigger balanced activation across hemispheres. Kimura (185) proposed, for instance, that the right-sided bias for hand preference observed in the human population could relate to the particularly efficient spatial and temporal motor programming abilities of the left premotor area and that such efficiency would empower language capacities.

Kimura’s theory of hemispheric specialization can be linked with the so-called “temporal” hypothesis, according to which the left hemisphere is specifically able to process rapidly changing acoustic signals. Such ability is clearly crucial for speech comprehension. The temporal hypothesis has recently been explored using functional neuroimaging, although behavioral evidence was accumulated for more than 25 years by Tallal and co-workers (380, 381; see also Ref. 115). For instance, Zatorre and Belin (431) described activation in the left superior temporal gyrus for rapidly changing auditory stimuli (over a period shorter than 80 ms) and a contralateral bias for more steady acoustic events. Similar findings have been described by Poldrack et al. (298) but related to the inferior posterior frontal cortex close to the left PIFG. Correspondingly, in a recent study from our group (341), this region displayed an inverted U-shaped response curve showing activation decreases for speech samples stretched or compressed so as to make them unintelligible. Interestingly, non-speech auditory stimuli containing rapid transitions similar to those present in speech activate left-sided perisylvian structures as much as genuine speech stimuli do (432). However, even though such complex nonspeech sounds were not intelligible, it is impossible to rule out the possibility that subjects assimilated these stimuli to speech sounds, as both types of stimuli were presented to the same subjects during the same experiment (for an example of the influence of top-down modulation nonspeech auditory perception, see Ref. 358).

In other words, the temporal hypothesis stipulates a left-sided hemispheric specialization for perceiving and processing rapid acoustic transitions relying on stimulus driven, bottom up, on-line feeding of verbal working memory. Another hypothesis that is more complementary than contradictory to the temporal hypothesis is the category hypothesis. The latter proposes that left hemispheric structures store long-term representations of phonemes, which permit the categorization of those phonemes under the influence of top-down modulation mechanisms. Categorical perception thus reduces acoustic variability by sorting inputs into discrete classes defined by stored prototypes. According to Kosslyn’s theory of visual perception (191), this function mode is opposed to coordination (estimation of relative distances between each component of a stimulus). Several behavioral studies have suggested a left hemispheric predominance for categorical perception, at least in the context of visual spatial perception, but recent findings suggest that this leftward bias may also exist for the object-based, “ventral” visual stream (276). Even if categorical perception is a basic functional strategy applying to nonlanguage stimuli in various animal species, its efficacy is especially compelling for discriminating phonetic contrasts during speech perception. Recent neuroimaging findings (72, 341) showed activation in the left supramarginal gyrus and the left PIFG for categorical perception of spoken syllables.

2. Ventral and dorsal language functional pathways

Ungerleider et al. (408) described a large-scale organization in the visual system of monkeys, a structural model which has soon proved to be applicable to the human brain (151). The ventral, or “what,” pathway is devoted to object categorization and recognition on the basis of explicit long-term memory access. The dorsal, or “where,” pathway is concerned with the spatial localization of potential targets in relation to real-time action decisions in the surrounding space, which implies that this pathway is also a vision-for-action or “how” pathway.
The discovery of the global organization of the visual system was followed by a similar conceptualization about the auditory system in the macaque, even though anatomical evidence had accumulated well before (see Refs. 322–324). As mentioned in previous sections, the binary distinction between ventral and dorsal pathways could apply to the neural architecture implementing language function (e.g., the “rostral ventral” region of the superior temporal sulcus involved in speech intelligibility, Ref. 355). One can speculate that the fundamental linguistic distinction between whole word and sublexical aspects of language processing corresponds to a functional anatomical segregation at the neural level, i.e., a dual-route organization. However, the notion of dorsal/ventral dichotomy is only relative in physiological terms, which implies that the two pathways 1) operate in mutual interaction, 2) have a complex inner structure, and 3) may be closely intermixed at both their origin and their termination in “convergence areas” (140). The “lexical pathway” could be viewed as the “language what” system, dealing with meaningful language objects stored in long-term memory in relation to lexical and semantic categories. The “sublexical pathway” would correspond to a “language where/how” system, in which sequences of lexical subunits constitute spatial-temporal dynamic structures that are transiently stored in verbal working memory and that are subject to sensory/motor transcoding (maintenance or transformation in the articulatory loop).

However, how much empirical support has this attractive hypothesis received?

Looking at the visual stream dealing with language input, some results from neuroimaging studies of reading support the existence of a dorsal pathway associated with phonological assembling. According to Pugh et al. (313) and Jobard et al. (173), and congruent with results obtained by Price et al. (302) and Paulesu et al. (282), the dorsal phonological stream would start in the left inferior temporal occipital junction and progress through the caudal part of the left superior temporal region and the inferior supramarginal gyrus en route to the left inferior frontal gyrus. The alternative lexical semantic or direct route for reading would involve other portions of the left middle and inferior temporal cortex. This second hypothesis is less clearly substantiated, according to Jobard et al. (173). Evidence for involvement in the lexical semantic route appears limited to the most posterior part of the left middle temporal gyrus (166), and no clear-cut distinction between the two reading pathways can be found in the motor output system, i.e., the left PIFG.

Looking at the auditory stream, as mentioned earlier in section II A 1, the rostral ventral pathway involving the anterior superior temporal sulcus and the pole of the left STG has been clearly linked to “speech object” access and/or access to intelligible lexical entities (355, 394, see Fig. 7). It should be noted, however, that access to semantic representations from auditory language input have been shown to elicit distinct, though similarly ventral, activations in the midpart of the middle and inferior temporal gyri (BAs 21 and 37, see for instance Refs. 103 and 400). The situation is less clear for the caudal component of the auditory language system. Whereas a reliable dorsal pathway dealing with sound localization and source motion has been demonstrated in monkeys (322–324), the role of the human dorsal caudal pathway in language processing is less obvious. Hickok and Poeppel (158) suggested that this component may not be active in natural conditions of speech perception. Nevertheless, two sources of evidence substantiate the involvement of the dorsal caudal system in sublexical language processing.

1) Speech is made of a stream of rapidly evolving acoustic events. An analogy can be drawn between the particular sensitivity of the left planum temporale to fast evolving cues (e.g., voice onset time for speech perception, see Ref. 216) and the existence of neural responses coding for sound source motion in the same region (416), which can be viewed as an auditory equivalent of area V5 for processing motion in the visual system.

2) Despite the evidence for sound localization specificity of the posterior auditory cortex (325), no topographical maps have yet been identified, suggesting a more complex neural coding of spatial cues than that found in the visual system (324). One can speculate that, in the putative human dorsal pathway, spatial and motion features of speech can be transiently reordered as series of concatenated lexical subunits such as phonemes, or even syllables. This hypothesis is congruent with findings of Wise et al. (426), who proposed that a subregion of the posterior part of the left STG transiently stores phonological forms of perceived or internally generated words.

There are arguments to include the left temporal occipital junction and the left inferior parietal lobule (supramarginal gyrus) in the caudal sublexical pathway. Whereas the posterior part of the left STG would be involved in the extraction, assembly, and storage of sublexical components from auditory input, the inferior left temporal occipital junction and inferior parietal lobule would implement similar operations on information flowing from the visual input, e.g., would organize series of sublexical units in the course of grapheme-to-phoneme conversion and transiently maintain them in an amodal abstract space. Supporting this view, Price (302) stressed the functional convergence of the left inferior temporal region (found active in tasks requiring access to phonological representations from visual orthographic stimuli) and the supramarginal gyrus (activated when participants are dealing with ordered series of phonological units on the basis of auditory or visual stimuli). The supramarginal gyrus could be conceptualized as an abstract notepad in which higher order representations derived from sensory inputs or long-term memory are manipulated over short
periods of time before decisions or actions can be programmed. This conceptualization converges with the fine-grained juxtaposition of phonological and spatial functions shown by Dehaene and co-workers (366). Furthermore, the phonological function of the supramarginal gyrus in phonological processing makes it a critical component of the “articulatory loop” in Baddeley’s model of verbal working memory (8), namely, the “phonological store” (279). Further projections of this pathway would then orient towards the left PIFG and, more specifically, the upper part of BA 44, which is critically involved in verbal rehearsal and speech output programming (158, 279).

Some evidence for a ventral/dorsal functional subdivision reflecting the dichotomy observed in the auditory and visual systems has also been obtained in studies of the left PIFG. The left PIFG is connected to superior temporal and inferior parietal structures by two main anatomical projection streams (reviewed in Ref. 424). The rostral ventral part of the auditory association cortex is connected to the ventral part of the PIFG (pars triangularis and pars orbitalis, BAs 45 and 47, respectively) via the uncinate fasciculus. The dorsal caudal part of the human auditory system and the inferior parietal cortex are connected to the upper part of the left PIFG (pars opercularis, BA 44) via the arcuate fasciculus. Interestingly, Romanski and Goldman-Rakic (335), who have described a region of the prefrontal cortex responding to various complex auditory stimuli in monkeys, showed that this region was located just ventrally to a region responding to visual cues. Signs of functional distinction in the human PIFG oppose BA 44, which has been associated with phonological processing, and BAs 45 and 47, which have been linked to lexical semantic processes. For instance, Paulesu et al. (281) have shown that a fluency task involving phonological cues activates the posterior dorsal aspect of the left PIFG (i.e., BA 44), whereas a fluency task based on categorical semantic cue activates the rostral ventral aspect of the left PIFG (i.e., BA 45). Converging evidence from a variety of language-related neuroimaging studies using different input and output modality combinations, including accessory or compensatory modalities such as lip-reading (283), has been accumulating (for a review, see Refs. 40, 298).

The most likely interpretation of the involvement of the left PIFG, a part of the premotor cortex, in phonological or semantic tasks, relates to working memory (see, for instance, Ref. 396). The involvement of the left PIFG, in particular its upper part, in the articulatory loop of verbal working memory has long been supported by neuroimaging studies (for reviews, see Refs. 40, 173; for a discussion, see Ref. 104). However, the interpretation of the role of the PIFG has been less straightforward for semantic than for phonological processing. Some authors proposed that the left PIFG is not concerned by semantic processes per se but rather participates in selection among lexical competitors (400). Two major review papers (40, 298) have highlighted the role of the inferior part of the PIFG (especially BA 47) in working memory. 1) According to Poldrack et al. (298), the semantic function of the inferior part of the PIFG is “analog to the spatial and object working memory functions that have been suggested for the prefrontal cortex on the basis of neurophysiology (139, 334) and neuroimaging (373)”; and 2) Bookheimer (40) proposed that this region is important for executive aspects of semantic processing such as directing semantic search. Although these hypotheses await validation from dedicated experiments, the bulk of evidence suggests that the dorsal part of the left PIFG is involved in constructing production-oriented phonological structures, whereas its ventral part implements executive processes operating at the lexical semantic level.

Considering the established relationship between the PIFG and sentence-level processing, we already mentioned that this area provided one of the rare examples of highly replicable brain responses elicited by decision or production requiring syntactic processing. However, in the important domain of syntax, the functional subdivision of the left PIFG into ventral and dorsal components is more problematic. Dapretto and Bookheimer (92) associated the activation of a dorsal part of the PIFG with judgments on syntactic properties of sentences and a ventral region with acceptability judgments based on meaning. However, the association between syntactic processing and the upper part of the PIFG is not consistently observed (40). Should an anatomical functional landmark be specified, one might want to highlight the relationship between syntactic processing and activations in BA 45 (see, for instance, Ref. 256), with large interexperiment variability. The observed variability might relate to the influence of several important confounding factors, e.g., the nature of the task (grammaticality judgment, reading, sentence production), syntactic complexity, and working memory load (for the latter, see Ref. 396). Sentence production has been less explored than comprehension tasks but seems to recruit subregions located posterior to the main sites described above, i.e., in BAs 44 and 6 (163, 170). Nevertheless, the precise contribution of different subregions of the left posterior inferior frontal cortex to grammatical and syntactic processing remains to be clarified. In keeping with the dorsal/ventral dissociation as a global organization principle which may apply to language (335, 336), syntactic processing may emerge from the integration of several sources of information converging in the left frontal cortex. A subpart of the ventral pathway running from the anterior superior temporal cortex to the inferior part of the PIFG would process the semantic content of verbs, which could be mapped onto the corresponding representation of actions likely to be implemented by the lateral frontal cortex (90).
A dorsal pathway could deal with the stream of auditory and visual language information, transcoded in their motor (articulatory) counterparts and maintained in working memory. The contents of verbal working memory might be processed in a more elaborate format than the simple linear sequence of incoming information flow. Such representation would rather be constructed in an abstract space, manipulated, and reorganized according to a syntactic arborescence (78), involving hierarchical relationships between sentence components. For instance, in a given language, this hierarchy would map 1) the expected, canonical order of subject, verb, and object and 2) phrases of the main clause as higher order components relative to phrases of a subordinate clause. Although the neural correlates of syntactic reorganization for sentence comprehension or production are likely to integrate the ventral and dorsal streams of information, the process by which this integration is implemented remains to be understood. In effect, recent neuroimaging results (256) lend support to the provocative Chomskian theory on the innate origin of syntactic arborescence.

3. Controlled versus automatic language processes: attention and memory

Human’s unique ability to implement and make use of language relies on more general and transversal cognitive functions such as memory and attention. Investigations into the neural bases of language cannot ignore such fundamental cognitive foundations for information processing.

We already mentioned that serial processes underlying phonological awareness crucially involve working memory while lexical semantic operations rather depend on long-term memory. These relationships concern explicit memory systems in adult subjects performing language tasks in their mother-tongue and imply the manipulation of participants’ attentional resources to focus on specific aspects of language processing (28, 105). However, other memory systems are likely to be involved in other cases, such as children learning oral or written language or adults learning a second language.

The case of first language acquisition provides compelling evidence of implicit memory involvement. The impact of implicit memory has been directly demonstrated in neuroimaging experiments in the form of decreases in activation during repetition priming experiments. For example, Dehaene et al. (96) facilitated the processing of language targets by presenting covert primes and found significant decreases in activation in the visual association cortex.

A well-known example of interference between routine language production and a less automatic task is the Stroop color interference task (presenting color names printed in a discordant ink color). Neuroimaging investigations of the Stroop task (e.g., Refs. 23, 47, 211) have shown that activation in the anterior cingulate cortex is associated with the voluntary enhancement of the less automatic response (e.g., naming the color of the letters instead of reading the color name). Neural correlates of implicit, irrepresible reading attempts have been described by Price et al. (310), who showed similar activation patterns for implicit and explicit (reading aloud) tasks. Although some studies have addressed the routinization of verbal tasks such as verb generation in adults (320), little is known about the neural mechanisms underlying the process by which subjects become acquainted with verbal material and procedures (e.g., children learning to read). This line of research should be especially fruitful for neuralinguistics as a whole and would provide new insight into the mechanisms of language disorders, which can involve a selective loss of the ability to use automatized language procedures.

Notwithstanding the dual (declarative/procedural) route for language learning developed in the model of Ullman (407), the rare studies devoted to language learning in adults have mainly tapped into procedural learning, including phonological discrimination (57, 414), artificial syntactic rules (120, 274), and real syntactic rules (256). Such studies have highlighted the crucial role of the left PIFG for learning both natural and artificial rules. The declarative component of Ullman’s model has received less support in terms of empirical neuroimaging results (for an example of the impact of episodic and semantic memory system on lexical learning, see Ref. 319). Very few functional neuroimaging studies have explicitly addressed the role of episodic memory in language processing and learning, even though it is likely to have significant influence, for instance, on remediation in aphasic patients (209; see sect. ivB).

IV. LANGUAGE AND BRAIN PLASTICITY

Language evolves throughout our life span and recovers following disruptions such as aphasia. Such adaptation is the consequence of brain plasticity that is observed in both developmental processes (for a review, see Refs. 263, 268) and postlesional recovery (e.g., Ref. 77a). Brain plasticity is the process by which synaptic systems change via molecular and cellular transformations to increase communication efficiency in neural networks.

A. Developmental Plasticity

1. Normal development

For obvious ethical reasons, isotopic studies cannot be conducted in normal children. Studies conducted with these techniques involve children with focal brain lesions
for instance, can be found in the planum temporale found in the normal right-handed adult, whereas results from normal children essentially come from fMRI investigations (e.g., Ref. 100) and electrophysiological studies (e.g., Refs. 99, 242, 397).

Lesion studies constitute the main source of neurophysiological information available to date concerning language development. Studying the correlations between the age at which an infant suffers cerebral tissue destruction and the corresponding cognitive impairments has made it possible to establish critical periods in language development.

Almost 40 years ago Lenneberg (210) proposed the hypothesis of equipotentiality of the two hemispheres at birth for implementing language. Lenneberg (210) indicated a relatively wide range (2–12 years) for the critical period in which hemispheric dominance is established and argued that individuals less than 2 years old suffering a brain lesion could achieve normal language development whichever side of the brain is damaged. The hypothesis of equipotentiality has been challenged by numerous studies on language lateralization in the last three decades, however. The slight anatomical asymmetry of the planum temporale found in the normal right-handed adult, for instance, can be found in ~70% of newborns (412, 427). Some authors speculated that this leftward asymmetry might be linked to language lateralization in adults and, therefore, that the left hemisphere was more likely to implement language as soon as an individual is born. Partial support for this hypothesis comes from the observation that lesions incurred before the age of 2 are likely to shift hemispheric dominance for both manual control and language processing, while lesions incurred between 2 and 10 years of age are more likely to induce shifts in language lateralization only or, when the damage is limited, even local reorganization within the damaged hemisphere (349). More generally, child age seems to be the primary factor predicting recovery from aphasia, with prognosis far better in children under age 6 (e.g., Ref. 248).

In the course of maturation, the brain shows a great potential for recuperation of language function compared with low-level functions (e.g., primary visual processing or motor command), possibly because the time course of language maturation is more extended than that of elementary sensory or motor brain systems. Lesions in primary cortices or central sensory-motor relays (e.g., basal ganglia and thalamus) recover generally less well than lesions in association areas (7, 249). The size and the location of lesions are also factors of prime importance. Focal lesions are often followed by local reorganization, whereas massive brain lesions more often induce palliative recruitment of contralateral tissue (80).

Although indirect, results from neuroimaging studies in children suffering brain damage and/or pathology have shed light on the neurophysiology of language development. Studies conducted in both epileptic and aphasic children have suggested a metabolic and functional superiority of the right hemisphere in the first years of life (77). For instance, in children who recovered good language abilities after left hemispheric lesion, a stronger right-sided activation was found in the group of youngest children (age <6 years; Ref. 248). Recently, Hertz-Pannier et al. (157) described the case of a 10-year-old child who underwent hemispherectomy at the age of 9. A reasonable level of comprehension was recovered in the year after the operation, but expressive language was less well recovered. Comparison of fMRI scans before and after surgery showed an important takeover by the right hemisphere of processes initially recruiting the left hemisphere.

fMRI (as compared with PET) has renewed interest for neuroimaging in childhood because it is noninvasive and can be used with normal children. However, this remains a challenge because 1) the scanner environment is not particularly child friendly and is not suitable for testing very young infants on their own, 2) it is difficult to obtain absolute immobility in infants and children, and 3) the developing brain differs from the adult brain in many ways (e.g., myelination, chemical composition, metabolism, and vascularization).

Language activation studies have recently provided language lateralization normative data that could be compared with data recorded in brain-damaged and/or epileptic patients (132, 207). Although larger than the activation foci observed in adults, activated regions appear very similar in child groups (age >7) during the same tasks. However, infants’ BOLD signals might be influenced by several physiological factors that are specific to early stages of brain development, such as immature myelination or atypical vascular response. For instance, these factors could account for blood flow decreases instead of increases observed in the primary visual cortex (43). In spite of coarse spatial resolution, near-infrared spectroscopy seems an interesting, noninvasive, complementary procedure to study physiological cerebral vascular responses of infants or neonates to stimulation of sensory cortical areas (272).

Using fMRI in 3-mo-old babies, Dehaene-Lambertz et al. (100) have recently shown left-sided activations elicited by speech and backward speech. Although no specificity to language stimuli was found, the authors interpreted their results as a sign of early lateralization for processing fast formant transitions, which are characteristics of spoken language (see also Ref. 432). In the visual modality, Schlaggar et al. (354) found differences between adults and school-age children performing a single-word processing task in activation of the frontal cortex and the left visual association cortex. To validate these differences as task specific rather than age specific, the authors identified those regions showing no sensitivity to

Physiol Rev • VOL 85 • JANUARY 2005 • www.prv.org
task performance. Gaillard et al. (133) have shown that by 10 years, reliable signs of left-lateralization for language processing can be found. However, activations in 10 year olds are more diffuse than in adults, and the right frontal cortex is significantly more activated.

Even though neuroimaging studies in children are now being undertaken, one cannot overlook 30 years of electrophysiological research addressing the following questions: 1) What is the specificity of neural responses associated with the earliest stages of language development? 2) When can we observe the emergence of hemispheric dominance for language? 3) What are the main stages of development (for reviews, see Refs. 386, 387, 390)?

Mismatch negativity (MMN) is a modulation of the N2 ERP component which indexes the automatic detection of a variation in the auditory input (258). The MMN is classically elicited in the context of an oddball paradigm, in which series of standard events are randomly disrupted by infrequent (deviant) events. Various types of auditory stimuli have been used in MMN paradigms, such as harmonic tones (199), phonemes (259), and syllables (317; and the effect probably extends to word forms, see Ref. 258). The use of an MMN paradigm has permitted Dehaene-Lambertz and co-workers (see Refs. 98, 99), for instance, to obtain evidence of phoneme discrimination in 2-mo-old babies by presenting series of five identical syllables (/ba/) and series of four /ba/ ending with a deviant syllable (/ga/). The deviant syllable elicited a significantly different ERP waveform, which they interpreted as being a MMN-type response. However, Dehaene-Lambertz and Pena (101) observed that while neural responses elicited by auditory stimuli were clearly preponderant over the left temporal region in 4-mo-old infants, responses to speech were not significantly larger than those elicited by nonspeech stimuli. This result is not directly congruent with recent outcomes from fMRI studies (100), but anatomical hypotheses made on the basis of ERP topography are much less reliable than those coming from neuroimaging investigations.

Several attempts to characterize the first steps of word recognition have been made using groups of words that are thought to be known or unknown to individual infants aged 13–20 mo (242). Mills et al. (242) used 10 known words and 10 unknown words matched in length and number of syllables (but not in phonotactic structure). Word knowledge was established by asking parents to rate words on a vocabulary checklist and by testing children’s comprehension using a forced-choice word-picture matching task. Three main components were found: P100, N200, and N375, which can be compared with the main components identified by Kushnerenko et al. (199) for harmonic tone processing. The latter two peaks showed a sensitivity to word familiarity, particularly in the infants who had the highest comprehension scores. Mills et al. (242) further suggested that a lateral difference found in ERPs (e.g., greater amplitude of the P100 component in good comprehenders on one side of the scalp) was an indication of early language lateralization, which could account for good comprehension skills.

Thierry et al. (397) recently reported a MMN-like effect in 11-mo-old babies induced by the familiarity of English words. They presented 58 familiar words randomly intermixed with 58 rare words matched for length, number of syllables, and phonotactic structure. Despite the equal proportion of familiar and rare words, familiar words elicited a significant modulation of the N2 component peaking within 250 ms after stimulus onset. This modulation was compared with the MMN because 11-mo-old infants are likely to know only a very limited number of the words selected a priori as “familiar words.” Thierry et al. (397) interpreted this effect as a sign of automatic attentional involvement triggered by the phonological familiarity of the initial diphone or triphone of the stimuli.

2. Developmental disorders: the case of developmental dyslexia

Understanding of the physiopathology of developmental disorders affecting cognitive and language processes will greatly benefit from advances in neuroimaging methods. Little reliable evidence has accumulated so far for disorders such as autism (344), whereas developmental dyslexia, a specific impairment of visual and phonological processes affecting ~5% of children, has yielded a substantial set of results in the past few years (e.g., Refs. 165, 363, 389; for a review, see Refs. 148, 321).

Structural abnormalities affecting the corpus callosum (329), the temporal parietal regions (330), or perisylvian regions such as the inferior posterior frontal cortex and the cerebellum (114), have been described in dyslexic individuals. Although previous claims that dyslexics display abnormal asymmetry of the planum temporale have been challenged by recent MRI findings (48, 346), structural abnormalities in the perisylvian areas have been confirmed in a large controlled study of twins (285).

Diffusion tensor imaging has recently been used to demonstrate a disorganization of neural fascicles located beneath the supramarginal gyrus in dyslexics compared with controls, suggesting that abnormalities in neural structures in dyslexic individuals might not only involve gray matter but also subcortical connectivity (187).

Most functional imaging studies of developmental dyslexia have involved adult subjects displaying sequelae of dyslexia (for a review, see Refs. 107, 148, 313). The main outcome can be summarized in three main characteristics: 1) the disorganization of the language network, 2) the existence of a common neural basis for reading disorders, and 3) evidence for developmental plasticity.
Evidence for a disorganization of the normal network dealing with language processing in dyslexics has been provided by many functional imaging studies and has been characterized by 1) lack of activation of retro-rolandic areas (areas posterior to the central sulcus) such as the left STG and the angular gyrus (346), 2) higher-than-normal activation in the left PIFG (363), 3) disruption of the normal activation pattern involving perisylvian areas (280), and 4) absence of correlations between the activities recorded in the left fusiform gyrus and in the left angular gyrus (165). These regions have all been implicated in reading in normal participants (see sect. mA2).

Among these regions, the left fusiform gyrus has been singled out in a recent cross-linguistic study (278) because it was shown to be hypoactive in dyslexic subjects compared with normal subjects whatever their language (Italian, French, or English). In spite of major differences between these languages and their “standard” neurofunctional correlates (282), it seems that a particular region, which is also involved in object recognition and naming, fails to activate normally during reading tasks in adult dyslexics. These findings have been confirmed in a study of large groups of dyslexic and control children (361). However, dyslexics did perform well during the experiment, which entailed reading frequent, concrete words and legal pseudowords. Therefore, compensatory mechanisms, possibly reflected in the particularly intense activation of motor and premotor areas, are likely to be involved.

Studying 10-year-old dyslexic children with MEG, Simos et al. (368) have observed that impaired readers showed the same level of neural activity in the left fusiform gyrus as normal readers (which contrasts with findings in adults) and higher-than-normal activity in the right posterior temporal cortex. One may hypothesize a progressive disengagement of the left fusiform gyrus, the efficacy of which would progressively drop as dyslexia is being compensated in the course of adolescence, while other regions in the right hemisphere or in the premotor areas would progressively become overactive. Evidence for functional reorganization in compensated adult dyslexics has been provided recently by Shaywitz et al. (362).

Evidence for brain plasticity in dyslexic adults has been obtained by Temple et al. (389), who demonstrated the impact of exposure to slowed down speech in improving the processing of complex sounds. Recent neuroimaging studies have also involved dyslexic children and demonstrated therapy-induced changes in language performance and brain activity (198, 369). These authors have shown that a normal level of performance and normal activity could be restored in the left temporal region of dyslexic children after intensive phonological and orthographic training. Furthermore, a restoration of close to normal activity in the left perisylvian cortex after phonological training has been shown by both an MEG study (369) and an fMRI study (388).

Finally, dyslexia can be considered as an interesting and fruitful model for neuroimaging studies of language because it involves defective language organization in the absence of manifest brain lesions such as those observed in aphasic patients.

B. Postlesional Plasticity

1. Language recovery poststroke

Neuroplastic phenomena intervene almost systematically in the course of vascular aphasia, and spontaneous recovery is usually observed for some language deficits. The involvement of the right hemisphere in the compensation of aphasia has long been hypothesized (141). This hypothesis has even been used as the conceptual basis for several neuroimaging studies of lesion-related language disorders (e.g., Refs. 54, 67, 273, 402, 419). The contribution of specific right-sided regions in language recovery from aphasia has gained support from recent results. For instance, Leff et al. (208) presented aphasic patients who had recovered single-word auditory comprehension after left posterior temporal infarction with spoken words at different rates and found significant changes in the physiological responsiveness of the right posterior superior temporal sulcus. In chronic patients with predominant damage in the left frontal region, Blasi et al. (34) showed that learning to retrieve words can downregulate activities in a right frontal-occipital network, suggesting a compensatory role for this cortical network. Indeed, the response of this right-sided network to practice mimicked that observed in the left frontal cortex in control subjects. In the same vein, lesions localized to the pars opercularis of the third frontal gyrus on the left have been shown to affect activation of homologous right-sided regions during production of propositional speech compared with both healthy subjects and patients with frontal lesions sparing this critical lesion site (32). Converging evidence comes from studies using TMS which have inhibited right hemispheric areas in recovered patients. Flitman et al. (121), for instance, showed that magnetic stimulation may affect the compensatory functions of right-sided areas as it induces transient language disorders in these patients. However, contradicting this finding, other results have suggested that inhibition of right-sided premotor regions can improve naming abilities in nonfluent aphasic patients at a very late stage (260).

Overall, the role of the right hemisphere in compensatory mechanisms remains a matter of contention. An early activation study by Knopman et al. (188) suggested the implication of right-hemispheric structures may be restricted to early stages of poststroke evolution, whereas spared portions of the left hemisphere would become
crucially implicated in late recovery. Minura et al. (244), who studied correlations between language scores and rCBF measured at rest, showed the inverse pattern, however. The compensatory role of the right hemisphere was also the focus of several recent activation studies which in fact showed that in the end spared regions of the left hemisphere were the main substrate of recovery mechanisms. In the same vein, Belin et al. (19) and Warburton et al. (415) proposed that persistent activations in the right hemisphere could even compromise the restoration of language functions in the left hemisphere. The critical impact of the level of activity in the left superior temporal cortex for recovery was repeatedly stressed by Heiss and his group. Heiss et al. (153) have used a word repetition task as a common activation task in comparatively large groups of aphasics. They demonstrated that the relative sparing of this region corresponded to both recovery of language functions and restoration of left temporal activation at a later stage. Indeed, massive lesions of the left superior temporal region prevented this region from activating at later stages and was associated with partial recovery of comprehension as well as activation of the right temporal cortex only. Contributions of the left superior temporal region and the right homotopic cortex to the recovery of lexical production were also observed by Cardebat et al. (68), who showed correlations between signal change over a long period of time and behavioral performance on word generation tasks.

The heterogeneity of these findings comes not only from the many confounding factors due to subject-specific and lesion-specific variability (63), but also from task heterogeneity (in most cases, tasks are not specific to patients' deficits, and they are often compared with irrelevant control conditions, Ref. 102). An adequate control task should target undamaged language processes, whereas the active task should target a specific dysfunction. Moreover, discrepancies between (small) group studies emphasize the interest of single case studies in which striking dissociations in psycholinguistic performance can be observed (e.g., Ref. 67). In such studies, patterns of activation observed in a given patient should be compared with a group of control subjects or, even better, to each control subject (415).

Price et al. (308) proposed that, in a patient performing a semantic task correctly, the pattern of activation reveals a set of regions that are necessary to achieve the task. In contrast, the patterns of activation found in normal subjects on the same tasks may involve supplementary areas that are not indispensable but, rather, represent accessory mechanisms or strategies used by some subjects to optimize their performance. In this context, using parametric designs in activation experiments would appear to be a particularly fruitful strategy. Dissociations imply poor performance in one task while another task remains possible because of preservation or compensatory mechanisms. When a patient attempts to perform a task but fails, numerous “parasite activations” arise that do not reflect task-specific processes but rather the mental effort and effects linked to repetitive and unsuccessful attempts. Thus correlating a subject’s performance with activation signals in the whole brain would allow one to tease apart activations reflecting information processing from those merely reflecting parasite phenomena.

Exploring the influence of therapeutic intervention on brain functions in aphasic patients is a new topic. The few available reports concern studies that have mainly focused on behavioral revalidation, with the exception of one report on neuroimaging and pharmacological therapy in aphasia (180). This placebo-controlled activation study demonstrated a significant improvement in patients treated with piracetam, a GABA derivative. Other molecules are likely to be worth testing, such as cholinergic agonists, as suggested by Tanaka et al. (384). Serotonergic or amphetamine-like agents may also be of interest as they have been shown to improve motor recuperation. Indeed, neuroimaging experiments have demonstrated their influence on activation patterns during motor tasks using fMRI (219).

Belin et al. (19) were the first to address the relationships between specific language therapy in aphasia and neural reorganization explored with PET. These authors focused on speech production and observed better performance in patients who displayed left-sided perilesional activations after melodic intonation therapy. Small et al. (371) studied a single case of acquired dyslexia and performed fMRI before and after an intensive remediation program. They demonstrated that phonological training resulted in specific activation in the left association visual cortex (lingual gyrus). Similarly, Léger et al. (209) studied an aphasic patient who exhibited a massive speech output deficit and was involved in a language therapy program devoted to rehabilitating the output lexicon. Using fMRI, they found a specific activation in the left P10G and the superior part of the left supramarginal gyrus posttherapy, after taking into account the activation level before therapy. Conversely, a relationship between right-sided activation and improvement of auditory comprehension after training was observed by Weiller and co-workers (257), who demonstrated a correlation between activity in the right temporal cortex and comprehension scores. Similarly, training on sentence processing was associated with changes in the right hemisphere during a matching task between spoken sentences and pictures in a patient described by Thompson (398).

Overall, the following pattern seems to emerge: remediation and/or testing of comprehension elicits activation in the right hemisphere, whereas remediation based on phonological processing favors recruitment of the left hemisphere; it may be that some other factors interact with these effects, such as damage versus sparing of...
2. Language reorganization in neurodegenerative diseases

Degenerative diseases are characterized by a loss of cognitive aptitudes and especially semantic knowledge in the language domain. Grossmann et al. (145) described a lack of activation in the left temporal/parietal cortex in patients with Alzheimer’s disease compared with control subjects during a semantic task involving category judgment. In other neuroimaging studies using the resting state, significant correlations were found between focal hypometabolism and specific language deficits (e.g., relationships between phonological-to-lexical ratio and supramarginal-to-angular ratio observed as in Ref. 284). However, some activation studies carried out on patients with degenerative dementias have shown signal increases in patients compared with controls, especially in the frontal cortex, even when patients have poorer behavioral performance (e.g., Ref. 69). Therefore, regions that are hypofunctional when measured at rest can activate to a greater extent in patients than in control subjects when participants are involved in a cognitive task. In some studies, especially those related to memory retrieval, a compensatory role has been proposed for these increased activations (18, 351, 428). Sonty et al. (375), for instance, compared patients with primary progressive aphasia engaged in phonological and semantic tasks with control subjects using both voxel-based morphometry and fMRI. Patients showed more atrophy in temporal and parietal regions, but regions activated in the control group were similarly active in patients (despite decrease in patient performance). In addition, patients showed activations in areas not activated in control subjects. The authors proposed two interpretations: the involvement of compensatory mechanisms or defective inhibitory processes.

Another interesting clinical context is that of “semantic” dementia as a variant of frontotemporal dementia, a degenerative disease specifically affecting semantic knowledge. In these patients cortical atrophy is preponderantly observed in the anterior and lateral part of the left temporal lobe. Activation studies have either shown an increase in activity for adequate processing (66) or a decrease in activity related to semantic memory deficit (250) in spared portions of the temporal cortex. The relationships between activation and performance in patients with degenerative diseases appear complex and biased by local atrophy of brain structures. For instance, brain atrophy of patients with semantic dementia measured by voxel-based morphometry in the left anterior temporal cortex has been shown to correlate inversely with performance on semantic tasks (252). One possible explanation posits that remote effects of atrophy in some areas may account for lack of activation in other areas involved in the targeted cognitive process (66). Johnson et al. (174) studied both brain activation (using a semantic judgment task) and degree of atrophy in healthy senior subjects and patients with Alzheimer’s disease. While atrophy was more marked in patients than in controls matched for age in the superior temporal and inferior frontal regions, Johnson et al. found no difference for brain activation in the left inferior frontal region. Moreover, a positive correlation was found between atrophy and activation in the frontal region in the patient group only. This surprising finding may reflect adaptive, although inefficient, mechanisms in brain regions that normally implement the corresponding cognitive processes.

Further developments in this field of research could come from two sets of results: on the one hand, positive effects of some drugs such as cholinomimetic agents on memory have been demonstrated in patients with Alzheimer’s disease, and on the other hand, a modulation of prefrontal cortex activity has been shown in normal subjects performing a memory task while receiving physostigmine (130). Combining these two approaches might help to demonstrate drug-induced functional changes in brain-damaged subjects.

C. Outstanding Questions

Unlike studies of normal adults, functional neuroimaging studies of language networks in patients remain scarce. More studies are needed to determine the optimal method for comparing functional data in patients and control subjects. The major problem here is to deal with the heterogeneity of patients in terms of brain abnormalities and language deficits. This variability weakens the rationale of group average approaches, used in normal subjects.

One method for studying groups of patients is to exploit interindividual variability by seeking correlations between behavioral measures (e.g., performance on a language task) and signal changes in spared areas of the brain. This approach could permit the neural correlates of success and failure accompanying mental effort to be teased apart. In this respect, single-event procedures are the most promising methods because they allow the segregation of trials in relation to behavioral performance in single subjects.

Price et al. (308) demonstrated the value of the neuroimaging approach to single-case study. They showed that in some cases a cognitive task can be performed well above chance level by the patient. If a patient performs reasonably well in a given task and suffers from a lesion located in one of the regions activated in normal subjects
by the same task, the damaged region is unlikely to be necessary for the corresponding cognitive operations. This line of reasoning raises the question of whether other neural structures would take over damaged regions or compensate for defective mechanisms within the language-dedicated network. In the latter case, to what extent can the network cope with lesions and what are the factors that modulate this adaptive capacity?

Rapid recovery of language after major lesions in early childhood indicates that normal developmental changes in neural structures may help postlesional reorganization. Further longitudinal studies with functional neuroimaging should determine the time course and the variables that play a role in developmental recovery. In the same vein, systematic exploration of the neural correlates of recovery in adult patients, especially when targeted therapy is administered, will help to evaluate the efficiency of the treatment and to choose the appropriate strategy.

Speech therapy is the classical intervention for language deficits. However, its neural correlates are only now being established (209, 369). Although evidence has been obtained for neural changes induced by therapy, it remains unclear whether these changes are specific. Indeed, the possible relationships between the processes recruited during therapy and those that work during functional imaging experiments have generally not been assessed directly. A striking example comes from the study of developmental dyslexia. Two recent studies reported language improvement and palliative neural activities after intensive training using either language-specific procedures (phonological and orthographic exercises; Ref. 369) or nonverbal audiovisual “therapy” (198). Further studies are needed to disentangle the respective neural correlates of therapy-induced strategies and those related to language improvement per se.

Alternative and very recent interventional procedures consist of modulations of the neural network by TMS or by drugs. While TMS has the advantage of allowing focal interventions to be made, its physiological mechanisms need to be better understood in terms of inhibition and excitatory mechanisms as well as long-term effects. The use of drugs to improve acquired or developmental language deficits is still preliminary. Although encouraging, the first results have to be confirmed and better specified in terms of physiological impact. Functional neuroimaging seems a particularly appropriate method to investigate drug effects on cognitive functions in humans, and it could be used to guide biochemical development of specific drugs.

Finally, functional neuroimaging of language is likely to become useful for clinical research in the future. Indeed, the recent developments of perfusion and diffusion imaging make it possible to image very early stages of ischemia and to study the fluctuations of language deficits in real time as a function of perfusion in key cortical areas (159). This type of method will have an important role in establishing a prognosis for aphasia even at very early stages, making it possible to plan optimal therapy programs.

V. CONCLUSIONS

The massive scale of results already obtained using neuroimaging techniques in the domain of normal and pathological language physiology itself demonstrate the promise of this approach. Several meta-analyses recently published have underscored the coherence and reliability of the core data. On the other hand, it is difficult to keep track of the multiple and rapidly developing methodologies and, consequently, to integrate the various facets of language physiology into a single synthetic perspective.

Concerning clinical applications, functional neuroimaging has become a fundamental approach and could be even more important in the future, by improving the precision of diagnosis, prognosis, and therapeutic interventions in brain-damaged patients. For instance, the diagnosis of neurodegenerative diseases such as Alzheimer’s disease has clearly benefited from functional neuroimaging data. It has been suggested that its sensitivity and specificity could be further improved by combining results obtained at rest with those observed in activation studies (41) and biological markers such as genotype of apolipoprotein (for a review, see Ref. 71).

Studying the relationships between genotype and functional neuroimaging profiles could lead to great future developments, with the aim of elucidating the genetic determinants of cortical architecture (399), or of improving the prognosis in at-risk subjects before obvious symptoms appear (e.g., for Huntington’s disease, see Ref. 6). A first compelling example of this type of research in the domain of language is the description of a family (KE) in which several members suffer from a severe developmental language deficit (including speech and syntax impairments), corresponding to a mutation in the FOX P2 gene (215). An abnormally low level of activation (especially in the frontal cortex and the striatum) was observed during fMRI (using language tasks) in affected relative to nonaffected members of this family.

Finally, fMRI has now become a common preoperative procedure for mapping motor and cognitive functions before neurosurgical intervention. Results obtained in this field are congruent with intraoperative electrophysiological data, and many teams are currently addressing the same issues for language representation in the cortex (245).

Despite these desirable medical applications of functional neuroimaging as used today, electrophysiology and tomography remain technically and methodologically lim-
not only the formal aspects of language are under investigation. However, future functional neuroimaging studies will benefit from direct metabolic measurements (e.g., Refs. 190, 350), and perhaps give access to direct neural activity imaging. Such methodological advances would greatly improve our knowledge of brain-language relationships.

From a physiological standpoint, functional neuroimaging could prove a powerful tool for acquiring scientific insight into gene-brain-behavior links, which would provide an invaluable complement to more traditional research based on purely anatomical approaches to brain structures (326).

Although there is still scope for space-time integration of fMRI with electromagnetic data, functional neuroimaging clearly represents a fascinating tool for approaching on-going neural activity while perceiving, understanding, or producing language. However, from a cognitive point of view, the greatest challenge is to approach the real function of language, i.e., interindividual communication. Effectively, neuroimaging researchers are caught in a difficult dilemma: on the one hand, their mission is to conceive well-controlled cognitive tasks involving well-controlled stimuli to elucidate the relationship between brain activation and cognitive components; on the other hand, they are conscious that the context of neuroimaging experiments is very artificial. In the past two or three decades, studies clearly tended to ensure a good level of control over experimental parameters. Only now have studies begun to address the issue of the ecological validity of the findings, which may represent our first steps in the direction of the physiology of natural language.

In the future, the physiology of language might evolve from a subject-centered experimental perspective towards an interactive, dialog-based perspective, in which not only the formal aspects of language are under investigation but also, other human-to-human communication abilities, such as that of symbols and emotions.

We thank Annukka Lindell, Marilyn Vilman, and two anonymous reviewers for discussions and helpful comments.

Addresses for reprint requests and other correspondence: J.-F. Dénonet and D. Cardebat, INSERM U455, Federation de Neurologie, CHU Purpan, 31059 Toulouse Cedex, France (E-mail: jean-francois.denonet@toulouse.inserm.fr and cardebat@toulouse.inserm.fr); G. Thierry, School of Psychology, Univ. of Wales, LL57 2AS Bangor, UK (E-mail: g.thierry@bangor.ac.uk).

REFERENCES


Ross ED, Thompson RD, and Yenkosky J.

Rosa ED.

Rugg MD. Rugg MD.

Rugg MD.


Rumsey JM, Horwitz B, Donohue BC, Nace K, Maisog JM, and Ross ED, Thompson RD, and Yenkosky J.

Rumsey JM, Horwitz B, Donohue BC, Nace K, Maisog JM, and Andreason P.

Rumsey JM, Horwitz B, Donohue BC, Nace K, Maisog JM, and Andreason P.


