INTERDISCIPLINARY CROSS-FERTILIZATION OR CROSS-STERILIZATION? CHALLENGES AT THE INTERFACE OF RESEARCH ON BRAIN AND LANGUAGE

DAVID POEPPEL

In the last several years, the popular media have reported enthusiastically on progress in brain research, in general, and human brain research, in particular. Magazine titles such as "Demystifying the brain" suggest to the public readership that substantive progress has been made on topics ranging from the brain basis of learning to the brain basis of philosophical topics (neurophilosophy), religious topics (neurotheology), and economic topics (neuroeconomics). The techniques that have elicited the most discussion and the most widespread popular recognition are the non-invasive brain imaging techniques (PET, fMRI). The canonical result that we can read about in media ranging from *The New York Times* to *Der Spiegel* is an image of a brain in which a specific location is activated ("on fire"), suggesting that specific localizations constitute the basis for the psychological phenomena in question. Interestingly, the images created by art departments for magazine covers (localized flames inside of a glass head, say) are not as different from technical scientific visualization as one might imagine. For example, one cover of the popular neuroscience and psychology magazine *Gehirn & Geist*, an issue discussing how neuroscience embodies an attack on our image of humanity, is strikingly similar to images illustrating results from psycholinguistic brain imaging studies in a technical publication (*Journal of Neuroscience*).

The foundational concept that informs experimental design, the results, and their interpretation is the concept of functional localization. The idea that different places in the brain are associated with different processes and representations has a long history, dating back at least as far as Franz Josef Gall (1758–1828). Gall's phrenology and organology made explicit a view of psychology in which the human mind is made up of a list of mental facul-

---

This paper is a shortened and modified version of an evening lecture given in March, 2004, at the Wissenschaftskolleg. The lecture, making extensive use of visual and auditory examples not incorporated into this version, was given in German and was entitled "Vergebene Liebesmüh? Zur interdisziplinären Erforschung von Sprache und Gehirn".

ties, each of which is represented in a certain location. We now view this approach as hopelessly underspecified and psychologically misguided – yet it is at the basis of most of the results reported in brain imaging and, either implicitly or explicitly, constitutes a central part of the explanations provided for the brain basis of cognitive function.

In this brief overview I discuss some of the prospects and problems for the concept of localization, using examples from neurolinguistic research. I argue that the concept of functional localization is, in fact, profoundly useful – but only when coupled with a psychological or cognitive theory of a very different granularity than typically assumed. Specifically, I develop the argument that a linguistic domain, such as phonology or syntax, is not the right level of granularity to permit localization in any useful sense. The challenge for psycholinguistic as well as neurobiological research is to identify the correct level of abstraction to allow for *linking hypotheses* between physiological mechanisms and basic computations underlying different aspects of language.

Why is localization of function, particularly the idea of a spatial (topographic) map, so compelling? There is an easy answer: in the study of sensory and motor domains, the idea of a localized spatial map is actually correct. Research on the organization of the visual system has shown that the brain representation of the visual world is accomplished by the use of "retinotopic maps": adjacent objects in the visual world are adjacently represented in visual brain areas. Most of the 30–50 cortical areas identified in the primate visual system share this property of retinotopic organization.

Much like in the visual system, the auditory system uses a form of spatial mapping, namely mapping by frequency. It is often argued that brain areas sensitive to auditory information reflect a so-called "tonotopic organization". In such a map, low-frequency sounds are represented at one end of the map and high-frequency sounds at the other end. A given sound (speech, music, traffic, etc.) is broken down into its frequency components, which in turn are represented in their respective locations in these maps. Therefore, it is a pattern of locations and their activation that form the brain's representation of that sound.

Perhaps the most striking example of the principle of spatial representation is the representation of body surface in the brain: the somatosensory homunculus, or "somatotopy". The body surface is represented in an orderly way in the parietal cortex. Moreover, some areas of the body are dramatically over-represented (magnified) in the somatosensory cortex. Specifically, the brain space allocated to the mouth, face, and hands far exceeds the space dedicated to, say, the lower back. The mind/brain's view of the body is consequently a human shaped form with wildly exaggerated hands and head. The body is represented as a spatial map.

The spatial topographic principles of retinotopy, tonotopy, and somatotopy have been replicated across techniques and across species and therefore form one of our most strongly held beliefs about brain organization. Given the success of these concepts, it is not surprising that one might want to extend the notion of spatial topographic representation to domains *beyond* sensory and motor function. An example from the neural basis of processing words illustrates the extent to which we wish to broaden spatial mapping into even very complex cognitive domains.

How is word meaning represented in the brain? A satisfying answer to this question would not only reflect substantive progress in cognitive neuroscience but also begin to provide a real bridge to the humanities. Engaging a major issue such as "meaning" signals that the humanities and the natural sciences are focusing their joint efforts on common questions – raising the hope of interdisciplinary cross-fertilization that neutralizes the two-cultures divide. So, to illustrate this program of research, what does it mean to understand the word *kick* while reading it, and what distinguishes it from the word *pick*? Recent brain research, for example a study using fMRI, has addressed this question, and the results have received widespread attention. One interesting study (O. Hauk, I. Johnsrude, F. Pulvermueller. 2004. "Somatotopic Representation of Action Words in Human Motor and Premotor Cortex." *Neuron* 41, 301–307.) is based on the following intuition: verbs denoting actions, such as *kick* or *jump*, have strong sensorimotor components. According to certain theories of word comprehension, processing such a word has as its core operation the activation of those parts of the body associated with that word. The recent experiment mentioned takes a very strong view, predicting that even reading the words "lick", "pick", and "kick" will activate the corresponding areas in the somatosensory homunculus. In other words, reading the word "lick" will be reflected in the activation of the tongue area of the somatosensory homunculus, reading "pick" will be reflected in the finger area, and reading "kick" in the foot area. While this is, prima facie, a defensible hypothesis (some version of a "motor theory of word meaning"), a serious question is how far would one want to take such a theory. To test the prediction, these investigators measured the regional brain activation when subjects, while lying in a scanner, read verbs such as "kick", "pick", and "lick". Furthermore, in a control condition, the same participants wiggled their toes, fingers, and tongues. The results showed that there was some overlap between the brain activation elicited by moving the relevant body parts and the activation elicited by reading the corresponding verbs. This is certainly a spectacular result. It was interpreted as showing that understanding verb meaning is accomplished by activating sensory or motor representations. They argue that their experiment "rules out a unified meaning center in the human brain and supports a dynamic view . . . with cortical topographies that reflect word semantics". Can this be a plausible conclusion following from such data? What aspects of such a research program should make us nervous? Consider a few points that render such an interpretation highly problematic for all but the smallest set of action verbs.

(1) Brain imaging techniques such as fMRI (functional magnetic resonance imaging) reflect hemodynamic (blood-flow) activity over many seconds. While it is plausible to hypothesize that for verbs with a strong mental imagery component, such as "kick", motor representations can be co-activated, it is not immediately clear why one would interpret widespread hemodynamic brain activation integrated over seconds as the *primary representation* of a word's meaning. Language comprehension – including the retrieval of word meaning – is remarkably fast (current research estimates that word meaning is robustly available to speakers/listeners no later than 300–400 milliseconds after a word is presented). Therefore the assumption that brain activity recorded over many seconds reflects primary word meaning is extremely sketchy.

(2) If the meaning of a verb is given by the activation of the verb-associated motor area of the sensory and motor homunculus, why stop in the cerebral cortex? Brain areas and peripheral areas are massively interconnected, so why not include the spinal cord, which obviously also plays a crucial role in motor control? Why not the muscles in the fingers? Is there, in this view, still a usable and defensible notion of word meaning?

(3) What does such a view imply about patients with movement disorders or patients with injuries? Is meaning processing of verbs impaired for them because their somato-motor homunculus is compromised? That prediction is, in my opinion, much too strong. It would be tantamount to arguing that a congenitally quadriplegic patient cannot read a book about soccer because not having the necessary motor representations renders understanding of the associated words impossible.

(4) Only a small fraction of vocabulary can be covered by such a "direct matching" theory. What to do with verbs with no sensorimotor features (e. g. *seem* or *justify*) or, more generally, any abstraction (e. g. *Wednesday* or *political intrigue*)? We will need separate theories of word meaning for such items. Is it economical to stipulate separate word meaning representation theories for different types of vocabulary?

(5) What about so-called closed-class or function words (e. g. *or*, *some*, *not*, *through*)? These words have very precise meanings but do not map onto any topography we currently know of.

(6) Finally, what about all of *compositional semantics*, that branch of meaning associated with interpreting phrases in sentences, not just single words? The "direct matching" theory between sensorimotor brain representation and word meaning has nothing to say about how we construct sentential interpretations, not to speak of even higher-order, discourse-related interpretations. Distinguishing the sequence of words *dog bites man* from *man bites dog* requires the construction of meaning on a level more abstract than looking up the meaning of individual words.

On balance, the position articulated in a direct matching "motor theory of word meaning" does have merits for a small portion of vocabulary and implies that there is a tight connection between the many aspects of (action) verb meaning and the actual execution of the action. But notwithstanding this positive result, such a view drastically underestimates the complexity of word meaning, in general, and verb meaning, in particular. The cognitive neuroscience results that are brought to bear on these issues are of an extremely coarse granularity in comparison with the linguistic complexity of the items (see the extensive literature on word meaning in philosophy, linguistics, and psycholinguistics).

We briefly reviewed one piece of neurobiological (imaging) evidence. Presumably, a goal of such an interdisciplinary research program must be to relate the results from psycholinguistics to the results from neurolinguistics. Does the work illustrate scientific progress driven by interdisciplinary cross-fertilization? Or does it, rather, exemplify interdisciplinary cross-sterilization? The implicit presupposition in this type of research program is that by performing such brain and language research we will learn something about *how the brain works* and/or *something about how language works*. But is this presupposition correct? Is such "unification" of disciplines possible given our current state of knowledge? Is there a future for interdisciplinary cross-fertilization, or are we doomed to interdisciplinary cross-sterilization?

The problems that the research program is faced with are straightforward to articulate. Specifically, the "alphabets" (the *primitives* or *elementary constituents* or *ontologies*) of linguistics and neuroscience are profoundly distinct. Suppose we generate an inventory of the fundamental elements and operations postulated by the language sciences, on the one hand, and the brain sciences, on the other. Are there plausible links between the domains? The fundamental elements for language research include concepts such as distinctive feature, syllable, morpheme, noun-phrase or clause; fundamental operations on primitives include operations such as concatenation, linearization, and semantic composition. Basic neurobiological elements include structures such as dendrites, neurons, ensembles, cortical columns, and operations such as long-term potentiation, oscillations, or synchronization. Are there any links between the alphabet of one domain and the alphabet of the other? Can we formulate a connection between any of these listed concepts? Clearly not. In fact, there is a total absence of linking hypotheses by which we can explore how well established brain mechanisms form the basis for linguistic computation.

Why are there no linking hypotheses? There are, in my view, two major causes, the *granularity mismatch problem* (neurolinguistics in practice) and the *ontological incommensurability problem* (neurolinguistics in principle). The granularity mismatch problem (GMP) refers to the fact that linguistic and neurobiological studies *of language* operate with objects of different granularity. In linguistics, extremely fine-grained distinctions and generalizations are made, referring to very detailed aspects of linguistic representation and computation. In contrast, neurobiological studies work with much broader conceptual distinctions. For example, numerous cognitive neuroscience studies seek to investigate the neural basis of "syntax" or "phonology". However, a rich tradition of linguistic research shows that these domains are not monolithic but have many subroutines and parts. Therefore, linguistic research refers to subroutines such as "phrase structure building" or "linearization", each in turn one sub-computation in syntax and phonology. In short, the GMP simply illustrates that the grain size of the objects of study is not congruent in a manner that permits mapping and matching between domains of inquiry such as neurobiology and linguistics.

The ontological incommensurability problem (OIP) refers to the fact that the units of linguistic computation and units of neurobiological computation are incommensurable. Therefore, an attempt at reduction, in its most canonical sense, cannot be made in any convincing way. If reduction is, at the moment, not even approachable in principle, the next step might be to investigate "unification" of fields (cf. the unification between chemistry and physics at the beginning of the 20th century). But unification will, presumably, involve *conceptual change*: a rethinking of how the fundamental concepts of the fields can be rearranged so as to permit the expression of relations across the disciplines. In order to permit such mapping, or *Abbildung*, of the areas onto one another, conceptual change is mandated.

The GMP and the OIP are, of course, not unique to research on brain and *language*. When we consider questions in visual perception, memory, attention, reasoning and decision-making, or even higher-order experiential issues (free will, consciousness), analogous problems emerge. The psychological domains in question have articulated theories dealing with subtle aspects of perception, cognition, and so on; that is, the "granularity" of the questions addressed in these domains is very high. In contrast, the manner in which neurobiological research talks about these domains again reflects a much more coarsely grained research program. The general principle that we can derive as a challenge to this kind of interdisciplinary research is: the level of description and analysis in one domain tends to be dramatically mismatched to the level in the other domain. Therefore *true* interdisciplinary progress is not possible. If true progress means *substantive* new insight into either of the domains in question, the present state of affairs suggests that the research enterprise has been a spectacular failure.

Again, to illustrate this crisis from the perspective of brain-language research: while neurobiology operates with distinctions such as syntax, semantics, or phonology, linguistic research operates with carefully distinguished sub-categories, the collection of which makes up a domain such as syntax or phonology. On this view, what can be the focus of localization studies? In neurobiological approaches, the focus has tended to be on the levels of syntax vs. semantics vs. phonology – but linguistic research shows us that these domains are not elemental. And common sense (as well as biological knowledge) dictates that if anything is localized in nervous tissue, at least on the level of cellular ensembles or columns, it will be *elementary* computational functions.

Given this rather pessimistic assessment of the situation – an attempt at reduction is incoherent and unification requires extensive reconceptualization – what do we need to look toward to generate substantive interdisciplinary progress, or, more pointedly, to even develop a viable interdisciplinary program of research? Is there a positive proposal available, a way out of this rather negative assessment of interdisciplinary research in cognitive neuroscience?

My suggestion is to look to *computation*, broadly construed, as a mediating level of analysis that permits contact between the conceptual architecture of the humanities, the cognitive sciences, and the neurosciences. More concretely, I argue for the following research strategy. Linguists and psycholinguists owe a decomposition (or fractionation) of the particular linguistic domain in question (e. g. syntax) into *formal operations that are, ideally, elemental and generic*. The types of computations one might entertain, for example, include concatenation, comparison, or recursion. Generic formal operations on this level of abstraction can form the basis for more complex linguistic representation and computation. Cognitive scientists, and linguists in particular, should develop the set of elementary representations and computations (independent of particular theoretical predilections). The challenge to the neurobiological colleagues is then precise and clear. Specifically, what type of neuronal circuit is the basis for the hypothesized computations in question? If the stipulated computation is a foundational constituent for processing language, then it stands to reason that it must be implemented by neuronal tissue *somehow*. In other words, if we agree that, say, concatenation is an elemental operation, then the most helpful neurobiological research program would attempt to identify exactly how this operation is instantiated. Such an operation may or may not be localized in one place in the brain; that is, it may be distributed over different brain areas, or alternatively it may be instantiated multiple times in different places.

Insofar as linguistic operations are fractionated into generic formal operations on the appropriate level of abstraction, and insofar as neurobiological research identifies the basis for such formal operations, we have now created the possibility to state direct linking hypotheses. Linguistics (or any other area of the cognitive sciences) conceptualized on this level of abstraction now has a direct relationship to biological implementation.

(A note on free will: In the last several years, there has been extensive discussion on the relation between the neurobiological architecture of the human mind and the possibility of free will. This discussion has been – occasionally – interesting, often provocative, but to practicing cognitive neuroscientists practically incoherent. Prominent neurobiologists have argued in prominent media outlets that the neural basis of human cognition precludes free will. This point may or may not be true, and one might consult the extensive philosophical literature for arguments. However, given the discussion here, it should be readily apparent that it is not even clear that the issue can be coherently formulated in current cognitive neuroscience terminology. Free will is, surely, not a monolithic concept but comprised of numerous constitutive parts; every putative act of free will obviously includes an enormous range of information processing, drawing on memory, perception, etc. Since no single constituent activity underlying an act of free will is itself monolithic, free will cannot be a monolithic and elemental function. Therefore, its relation to neurobiological implementation is, at best, vague and tentative and, more realistically, unstateable. So, whether or not one consequence of the neural basis of behavior is the existence or absence of free will, the contemporary cognitive neurosciences have, as far as I can see, absolutely nothing to say about it.)

Finally, turning to the originary issue of localization, what is its status? In the view sketched here, how much can be gained by focusing on localization? What is the payoff for continuing to perform series of brain-imaging studies attempting to localize entire complex domains such as syntax or phonology? In my view, not much can be gained at this point. We might think of localization as the "homework problem", that is, an important intermediate step (the identification of areas that are implicated, certainly an important descriptive step), but ultimately uninteresting from the point of view of explanation. Is it possible to achieve a unification of linguistics and neurobiology by asking localization questions? No; rather, we need to develop the explicit linking hypotheses between linguistic computation on the right level of abstraction and well-characterized brain mechanisms. The *wrong* question is: where are syntax or phonology or semantics mediated? The *right* question might be something like: what kind of computations in the brain forms the basis for processing *elementary* linguistic functions? Characterizing the problem in the manner we have done here does away with questions such as "Is human language processing unique?" or "Are humans 'higher' or 'better' in an evolutionary sense than animals?" Rather we can turn to typical questions of the sciences such as, "How does this work?"

There is, of course, functional localization, but, in this view, what is localized is circuitry that executes specific computations such as concatenation or recursion or linearization. When we perform a brain imaging experiment and interpret the results – the localized colored "blobs" appearing on brains – it appears more sensible to attribute to each activation a specific computational function rather than some coarse domain. As a specific example, an apparently simple process, such as auditory word recognition, a process we do within hundreds of milliseconds over and over and without conscious effort, is comprised of various subroutines, and it is these subroutines that we ultimately want to identify and localize. Localization as an intermediate research goal survives, but the interpretation of localization is fundamentally recast.

From a historical perspective, we may characterize such a research program as third in a series of large-scale developments. Cognitive neuroscience research on brain and language began in a decidedly localizationist and phrenological stance. This "intuitive psychological phrenology" that built on the work of Franz Josef Gall is the wave of the past. The wave of the present is what we might call "cognitive psychological phrenology": guided by observations and theories from cognitive psychology, we attribute to certain brain areas relatively broad psychological functions. I argue here that the wave of the future must be "computational organology". The research program seeks to identify the elemental computations that form the basis for perception, cognition, and action and attempts to identify the neuronal basis of these elemental computations. In such a program of research we will no longer pursue misguided attempts to sell localization as explanation. Rather, we will have moved from a descriptive enterprise to the explanatory approach typical of the sciences.