THE UNIQUELY HUMAN CAPACITY FOR LANGUAGE COMMUNICATION: FROM *POPE* TO [po:p] IN HALF A SECOND

Peter Hagoort

1 Introduction

No other species besides *Homo sapiens* has developed in the course of its evolutionary history a system of communication in which a finite set of symbols together with a series of principles for their combination allows an infinite set of expressions to be generated. This system of natural language enables members of our species to externalize and exchange thoughts within the social group, and, through the invention of writing systems, within society at large. Speech and language are effective means for the management of social cohesion in societies where the group size no longer allows this to be done by grooming, which is the preferred way of bonding in our genetic neighbors, the old world primates.[1]

The generative power of the human language system rests on its tripartite architecture.[2] In this architecture, language-relevant information is encoded in at least three distinct representational formats: one for meaning, one for syntax, and one for the sound structures of words and utterances. Through the process of mapping these representational structures onto each other, the conceptual structures that specify the content of the speaker's message are expressed as a linear sequence of speech sounds (speaking). Alternatively, during listening to speech conceptual structures are derived from a linear string of speech sounds. In this mapping process, combining units into hierarchical phrase structures (syntax) is the necessary link between conceptual structures and phonological (or sound) structures.

In the division of labor between the sciences investigating the human language faculty, it is the task of the linguist to specify the representational structures involved; and it is the task of the psycholinguist to investigate how these structures are accessed and exploited during listening and speaking. Finally, the cognitive neuroscientist is faced with the challenge of specifying how the brain enables human language, and of determining the spatio-temporal profile of neurophysiological activity underlying speaking (writing) and listening (reading).

In order to get a more precise picture of what these different sciences reveal about the human language faculty, we first have to specify what the overarching term "human language faculty" stands for. It refers to the collection of the following set of complex, related but at the same time distinct skills: speaking and listening, reading and writing, and in communities of the deaf, using sign language. Each of these skills requires that distinct representational structures in memory are accessed and exploited in real time. The cognitive architectures for these skills specify which

---

[1] Robin Dunbar, *Grooming, Gossip, and the Evolution of Language* (London: Faber and Faber, 1996); Willem J.M. Levelt, "Producing Spoken Language: A Blueprint of the Speaker," in *Neurocognition of Language*, Colin M. Brown and Peter Hagoort, eds. (Oxford: Oxford University Press, 1999), 83–122.

[2] Ray Jackendoff, *The Architecture of the Language Faculty* (Cambridge: MIT Press, 1997); idem, "The Representational Structures of the Language Faculty and Their Interactions," in *Neurocognition of Language*, Brown and Hagoort, eds., 37–81.

representational structures are involved and how these are operated on in real time. The neural architectures specify the ways in which these skills are instantiated in the wetware of the human brain. However, one should keep in mind that the distinction between the cognitive architecture and the neural architecture is an idealization. As a first approximation, it is useful to make a distinction between computations in symbolic terms (cognitive architecture) and in neurophysiological terms (neural architecture), but in a complete cognitive neuroscience of language these levels should be brought together.

In the remainder of this essay I will first discuss the cognitive architecture for one of these skills in more detail. I will then discuss the neural architecture. Finally, I draw some implications for a theory of the person.

2 The Cognitive Architecture

A central component of our language skills is the mental lexicon. The mental lexicon is the part of declarative memory specifying the knowledge that a language user has about the words of his or her native language. It is estimated that speakers of a language have an active vocabulary of at least 40,000 words.[3] For these words, speakers know what they mean and how they sound. In addition, they know the syntactic properties of words such as word class (noun, verb, etc.). All this lexical information is retrieved very quickly from memory. On average a speaker produces two or three words per second. This requires not only the retrieval of different sources of word information, but also the coordinated activation of a large ensemble of muscles involved in articulating speech. About 100 muscles are involved in speaking, whose innervation has to be coordinated with millisecond precision. Despite the complexity and speed of this cognitive activity, speakers are very accurate, and, on average, make less than one error in 1000 words. The occasional speech errors are nevertheless very informative about the architecture of the speech process. For instance, sounds can be exchanged between different words as in "*h*eft *l*emisphere" (instead of "left hemisphere"),[4] or can be produced too early as in anticipations ("it's a *m*eal mystery" instead of "it's a real mystery").[5] What these examples of speech errors illustrate is that words are not stored in memory as units, but have to be assembled from the constituting phonemes every time we produce a word. This assembly process occasionally goes wrong, resulting in sounds ending up in the environment of the wrong word.

Figure 1 presents a blueprint for speaking single words. It specifies what happens between the moment that we recognize a particular retinal image as, say, John Paul II, and the actual articulation of the sound stream "pope."

Speaking starts with specifying the conceptual content of the utterance. This specification can be determined by visual input as in our example. But in many cases the conceptual specification is determined by internal input, for example, the speaker's intention to express a certain idea. Whatever triggers the conceptual specification of the utterance, the speaker has to select a particular concept or series of concepts from the knowledge base in memory, and s/he has to select and decide

---

[3] For a general introduction to the mental lexicon, see Jean Aitchison, *Words in the Mind: An Introduction to the Mental Lexicon* (Oxford: Basil Blackwell, 1987).

[4] Victoria A. Fromkin, ed., *Speech Errors as Linguistic Evidence* (The Hague: Mouton, 1973).

[5] Ibid

about the way of expressing (for example, a message can be expressed as a statement, but also as a question, with or without irony, etc.). For instance, in our example the speaker can decide to say "John Paul the Second" or "the pope," to mention just two possible alternatives.

The conceptual selection and specification process precedes the actual formulation process in which preverbal conceptual structures trigger the retrieval of linguistic structures necessary to express the idea as a series of speech sounds. Here, two completely different types of linguistic information need to be retrieved, one specifying the characteristics of the word sound, the other concerning the grammatical properties of a word.

Each word form in the mental lexicon is associated with syntactic word information.[6] This latter type of information is referred to as lemma information. Lemmas specify the syntactic properties of words, such as their word class (noun, verb, adjective, etc.). For nouns in gender-marked languages their grammatical gender is specified as well (e.g., *horse* in French has masculine gender, in Dutch it has neuter gender). Verb lemmas contain information on syntactic frames (the argument structures), and on the thematic roles of the syntactic arguments (the thematic structure). For instance, the lemma for the verb *donate* specifies that it requires a noun-phrase (NP) as the grammatical subject, an NP as the grammatical object, with the optional addition of a prepositional phrase (PP) as the indirect object (e.g., *John* <subject-NP> *donates a book* <direct object-NP> *to the library* <optional indirect object-PP>). In addition, the mapping of this syntactic frame onto the thematic roles is specified. For *donate* the subject is the *actor*, the direct object the *theme*, and the indirect object the *goal* of the action expressed by the predicate.

In the next phase of the formulation process, the selection of the appropriate lemmas trigger the retrieval of sound pattern of the utterance (see figure 1). During this phase, the speech sounds (phonemes) of the word become available. In addition to the phonemes, a word's metrical information is retrieved, specifying the number of syllables and the stress pattern (not shown in figure 1). In a processing step known as phonological encoding, the phonemes are assigned to their syllable positions in a left-to-right order. The outcome of phonological encoding is a "phonological word," containing the word-sound information as a sequence of syllables with the right stress pattern. Syllables are the codes that form the basis for the articulatory movements of the vocal cords, the velum, the tongue, the jaw and the lips. These are abstract codes, since they are independent of the starting positions of, for instance, the lips and the tongue. Speaking with or without a pipe in the mouth results in different articulation movement trajectories, which are nevertheless instructed by the same abstract syllable codes. The final outcome of this whole cascade of retrieval and activation processes is an acoustic signal that the listener uses to derive the intended message.

Apart from the experimental evidence for a distinction between the retrieval of lemma- and word-form information, we are all familiar with a phenomenon supporting this distinction. This is the so-called tip-of-the-tongue state, referring to the often embarrassing situation in which we know that we know the word, we even

[6] Willem J.M. Levelt, *Speaking: From Intention to Articulation* (Cambridge: MIT Press, 1989); idem, "Producing Spoken Language: A Blueprint of the Speaker"; Ardi Roelofs, "A Spreading-activation Theory of Lemma Retrieval in Speaking," *Cognition* 42 (1992): 107–42; idem, "Testing a Non-decompositional Theory of Lemma Retrieval in Speaking: Retrieval of Verbs," *Cognition* 47 (1993): 59–87

know that it is, say, a noun with a particular grammatical gender (for example, neuter), but for some reason the retrieval of the sound form is hampered. The fact that we can access some aspect of word information but fail to retrieve others illustrates the idea that different aspects of word information are differentially stored

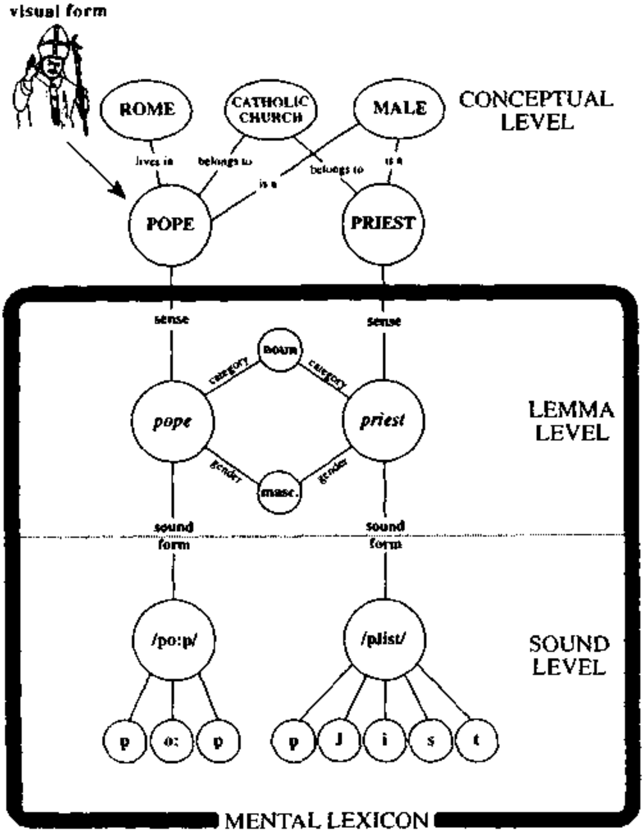


Figure 1. The Levelt and Roelofs model for speaking single words [see Willem J.M. Levelt, *Speaking From Intention to Articulation* (Cambridge: MIT Press, 1989); Ardi Roelofs, "A Spreading-activation Theory of Lemma Retrieval in Speaking," *Cognition* 42 (1992)]. Concept nodes (POPE) are activated on the basis of sensory and/or conceptual input. Activation from a concept node spreads to its lemma node (*pope*) in the mental lexicon. In addition, activation of POPE results in increased activation of related concepts in semantic memory, such as for instance PRIEST. Each concept node is linked to exactly one lemma in the lexicon. At the lemma level the syntactic word information is specified, such as grammatical gender and word category. For instance, in Italian the gender of the lemma *pope* is masculine (*il papa*), whereas the gender of, for example, the Italian lemma for church is feminine (*la chiesa*). After the lemma has been selected, word form information is retrieved and prepared for articulation.

and retrieved. The simplistic idea that words are units to be found somewhere in the brain is simply wrong. "Word" is just nothing more than a handy catch phrase for an orchestra of information with players of different instruments. Despite differences in the details of various models of speaking, there is general agreement among researchers of language processing that it requires the temporally orchestrated retrieval of the different types of information discussed above.

The way in which I discussed the process of speaking thus far is as a feedforward process from intention to articulation.[7] However, introspectively we often have the feeling that our way of expressing our thoughts sharpens and molds our intentions. That is, speech starts with an intention, but intentions are also (partly) derived from speech. This intuition has led to criticisms of an account that sees speech as the information flow from intention to articulation.[8] However, in my view this criticism can be dealt with easily if we realize that as speakers we are also at the same time listeners. That is, we listen to our own speech, using the same machinery that analyses the speech of others. In the listening process we derive the intention from the speech sounds that hit our ears. Speaking is a highly incremental process, which means that we have not specified all the details of our preverbal message before we start the formulation process. The incremental nature of speech planning opens a window of opportunity in which listening to our own speech can further shape and mold our intentions, and via this route influence the ongoing formulation process. Given that our cognitive machinery of language includes both production and comprehension, it instantiates a continuous internal dialogue between "speaker" and "listener" resulting in the introspective feeling that intentions are not only the source but also the by-product of speaking.

In the example above I have given the rough outlines of the cognitive architecture of speaking, mainly restricted to speaking single words. A full-blown model of speaking specifies additionally how words are combined into longer utterances, how the intonational contours of multi-word utterances are determined, etc. Similar blueprints can be made for listening, reading, and writing.[9] In all these cases, establishing the details of the cognitive architectures for the different language skills is based on a combination of conceptual analysis, computational modeling, and clever experimentation. With the cognitive architectures in hand we can ask sensible questions about the neural instantiation of the different language skills. Without such explicit models, the study of the neural mechanisms of language is doomed to fail.

As an example of science in action, I will discuss one simple experiment in some detail. If lexical concepts such as POPE and PRIEST are stored in a network-like way as shown above in figure 1, and if the activation of a particular concept partly spreads to nearby concept nodes in the network, this would predict certain processing consequences for words that are preceded by semantically related words. This is tested in the following way. Participants in the experiment see word pairs that are flashed on a computer screen. First, one word is flashed on the screen for half a second, followed by a few hundred milliseconds blank screen. Then the target word is flashed on the screen for half a second. The participants are instructed to read aloud this second word as soon as it appears on the screen. The presentation of the

---

[7] Levelt, *Speaking*.

[8] Cf. Daniel C. Dennett, *Consciousness Explained* (Boston, Mass.: Little, Brown and Company, 1991).

[9] For detailed examples, see Brown and Hagoort, eds., *Neurocognition of Language*

second word starts a clock that is stopped by the verbal response of the participant. As soon as the participant starts to read aloud the word on the screen, the clock stops. This allows the measurement of the participant's reaction time. In one condition of the experiment participants see the target word (for example, "priest") preceded in time by a semantically related word ("pope"). In the other condition the first word is unrelated in meaning to the second word (for example, "horse"). The prediction of the network model is that seeing the word "pope" results in partial activation of the word "priest," through the connection between the concepts POPE and PRIEST. If "priest" appears on the screen immediately after "pope," the reading of "priest" should be faster than in isolation, since it was already partly activated due to the preceding word "pope." However, the word "horse" does not spread part of its activation to "priest," since the concepts HORSE and PRIEST are too far apart in the semantic network space to influence each other. So if "priest" is read immediately after "horse" this should not speed up the reading process. The results of this type of experiment are in line with the predictions from the network model. Subjects are a few tenths of a millisecond faster in reading "priest" preceded by "pope" than in reading "priest" preceded by "horse." This so-called *priming effect* suggests that information about the meaning of words is stored in memory as a network of connected pieces of information, and not as isolated packages of individual word meanings.

3 The Neural Architecture

The neural architecture specifies the spatio-temporal dynamics of the brain processes that convert the retinal image of, say, John Paul II into the speech sound [po:p]. That is, we have to specify which areas of the brain are activated during the processes involved, and how the concomitant activations are temporally orchestrated.

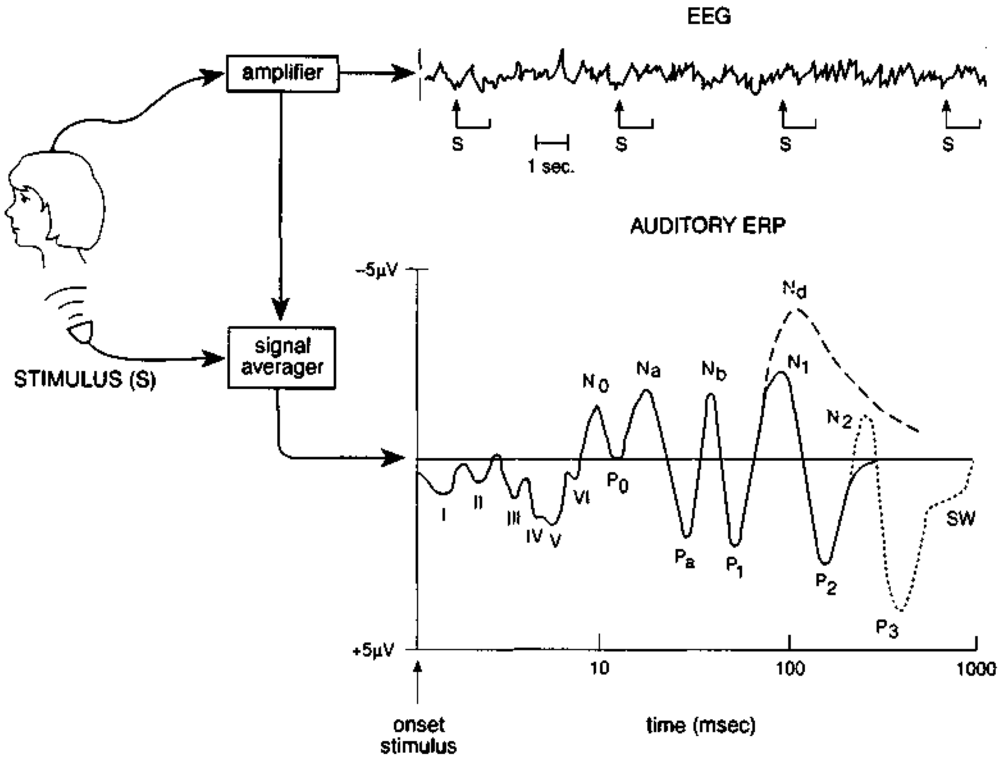
Recently, the recording of the electrical activity of the brain has resulted in a fairly fine-grained estimation of the time course of the different processes involved. Electrical brain activity is usually recorded from a series of electrodes at the scalp. If these recordings are done time-locked to sensory, motor, or cognitive processes, scalp-recorded event-related brain potentials (ERPs) result, reflecting the sum of simultaneous post-synaptic activity of a large ensemble of neurons. The ERPs have a high temporal resolution, in the order of a few milliseconds. Based on the latency of these components, relevant information can be obtained about the time course of the underlying processes (see figure 2).

Although for certain reasons[10] ERPs have been mainly recorded by language researchers in relation to aspects of language comprehension, recent research has applied this techniques successfully in studying the production of speech.[11] Without going into the details, on the basis of these and some other studies an educated guess can be made about the temporal dynamics of converting a visual image into an artic-

[10] Cf. Peter Hagoort and M. van Turennout, "The Electrophysiology of Speaking: Possibilities of Event-related Potential Research on Speech Production," in *Speech Production: Motor Control, Brain Research and Fluency Disorders*, Wouter Hulstijn, Herman F.M. Peters, and Pascal H.H.M. Van Lieshout, eds. (Amsterdam: Elsevier, 1997), 351–61.

[11] Miranda van Turennout, P. Hagoort, & C.M. Brown, "Electrophysiological Evidence on the Time Course of Semantic and Phonological Processes in Speech Production," *Journal of Experimental Psychology: Learning, Memory, and Cognition* 23 (1997): 787–806; idem, "Brain Activity During Speaking: From Syntax to Phonology in 40 Milliseconds," *Science* 280 (1998): 572–74.

Figure 2. (After Steven A. Hillyard and M. Kutas, "Electrophysiology of Cognitive Processing," *Annual Review of Psychology* 34 [1983]: 33–61.) Idealized waveform of a series of ERP components that become visible after averaging the EEG to repeated presentations of a short auditory stimulus. In this figure, the EEG is recorded from one electrode, placed at a central midline site on the scalp. Usually, averaging over a number of stimulus tokens is required to get an adequate signal-to-noise ratio. Along the logarithmic time axis the early brainstem potentials (Waves I–VI), the midlatency components (N_o, P_o, N_a, P_a, N_b), the largely exogenous components (P_1, N_1, P_2), and the endogenous, cognitive ERP components (N_d, N_2, P_3, Slow Wave) are shown. The components with a negative polarity are plotted upwards; the components with a positive polarity are plotted downwards. The exogenous components mainly reflect the physical stimulus characteristics (e.g., intensity, size, duration). The endogenous components reflect in particular the cognitive information processing consequences of the stimulation of (one of) the sensory systems.



ulated sequence of speech sounds.[12] In our example, it will take at least 150 milliseconds to perceive and categorize the retinal image as John Paul II.[13] The following activation of the concept POPE takes less than 200 milliseconds. It is followed by a cascade of retrieval processes. Activation of the syntactic features of *pope* (the lemma information) precedes the retrieval of the onset phoneme of the word [po:p] by about 40 milliseconds.[14] Importantly, the information about a word's phonological form is not available at once, but accrues in a left-to-right order. For words of 3 phonemes (/p//o://p/), it takes maximally 80 milliseconds to retrieve the remaining segments once the word-initial phoneme is available.[15] Since it takes about 600 milliseconds before articulation of the word [po:p] starts, the remaining time is necessary for preparing (and partly executing) the articulatory motor program on the basis of the phonological information.

Apart from answering the question about the time course, we need to specify the brain areas involved in the cascade of processing operations involved in speaking. For this we have to rely either on evidence from lesion data or on measurements of brain activity with the help of modern brain imaging techniques such as Positron Emission Tomography (PET) and functional Magnetic Resonance Imaging (fMRI). Lesion data come from patients who suffered from a stroke or brain tumor resulting in a language impairment. A precise analysis of the site and the size of the lesion on the one hand, and of the specific nature of the language impairment on the other, are used for making inferences about which areas of the brain subserve a particular aspect of language processing. For instance, lesions in the frontal cortex involving Broca's area often result in an impairment in producing the correct sound pattern for the intended words, which suggests that this brain area is normally involved in, among other things, the assembling of a word's sound pattern. In this way, relating lesion site to impairment symptoms is used for assigning language functions to brain structures.

PET and fMRI measure hemodynamic signals. They enable the detection and visualization of functionally induced local blood flow changes (PET), or changes in blood oxygenation (fMRI), which are assumed to be correlated with the activation of nearby neural tissue.[16] Roughly speaking, in this way the locus of neural activity related to cognitive processes is detected through a vascular filter. This implies that the temporal resolution of these techniques is inherently limited by the temporal dynamics of changes in blood flow or blood oxygenation, which is on the order of hundreds of milliseconds to a few seconds. This contrasts with the electrophysiological recordings that are directly related to neural activity, and have a temporal resolution on the order of milliseconds. However, the ERP measurements suffer from the so-called inverse problem, which makes it difficult to determine the localization of the electrical potentials that are picked up at the scalp. For the time being, only PET and fMRI provide measurements with the required spatial resolution.

[12] For more details, see Hagoort and van Turennout, "The Electrophysiology of Speaking."

[13] Simon J. Thorpe, D. Fize, & C. Marlot, "Speed of Processing in the Human Visual System," *Nature* 381 (1996): 520–22.

[14] Van Turennout, Hagoort, & Brown, "Brain Activity During Speaking."

[15] Van Turennout, Hagoort, & Brown, "Electrophysiological Evidence on the Time Course of Semantic and Phonological Processes in Speech Production."

[16] For a general introduction, see Marcus E. Raichle, "Visualizing the Mind," *Scientific American*, April 1994, 36–42.

Recent years have seen an increasing number of PET and fMRI studies on language processing. In the absence of an animal model for language, we are strongly dependent on these new brain imaging techniques to see the brain in action during language tasks. The following logic underlies most brain imaging studies on language. The patterns of brain activation associated with tasks that tap a specific step (for example, the retrieval of lemma information) in the cascade of processes involved in speaking are compared with activation patterns associated with tasks in which this particular process is not involved. Through this comparison one can determine the brain areas that are more strongly activated during this step (lemma retrieval) in the overall process. The areas that are more strongly activated are assumed to be the areas that are particularly involved in this aspect of speaking. For instance, one can ask participants to read aloud words and pseudo-words. The latter are phonotactically legal letter strings, which do not happen to be existing words in, say, English. An example is the letter string *floke*. Everyone can read this word, but no one knows what it means. If one compares the English word *smoke* with the pseudo-word *floke*, the following two differences arise in the processes between seeing it written and saying it aloud. One difference is that for *smoke* the meaning gets accessed in the course of the process, whereas for *floke* we don't have a semantic representation in memory. In addition, we cannot retrieve a phonological code from memory for the pseudo-word *floke*. Instead, we have to assemble such a phonological code by converting the individual graphemes into the corresponding phonemes. If we measure the patterns of brain activity associated with reading aloud words and pseudo-words, the differences between the brain activity associated with words and pseudo-words are due to the retrieval of word meaning and phonology and/or the assembly of the phonological code for pseudo-words. To further segregate the brain activations related to meaning and phonology, other task comparisons are needed. In this way one can figure out which areas of the brain are differentially activated during the different steps in the process of speaking. Of course, there are also areas that are crucial to both tasks, since in addition to the differences there are also commonalities between reading words and pseudo-words. For instance, in both cases one sees activation in the primary visual cortex, since the whole reading process starts with the analysis of the visual patterns that fall on the retina.

In a recent meta-analysis of more than fifty brain imaging studies on single word production, Peter Indefrey and W. Levelt summarized the current understanding of the neural circuitry underlying the cognitive activity that I described above.[17] All core steps in the speaking process are subserved by areas in the left hemisphere, which is the language dominant hemisphere in the large majority of people. Selecting the appropriate concept for speaking (POPE) seems to involve the left middle temporal gyrus (see figure 3 for an overview). From there the activation spreads to Wernicke's area, which is pivotal in retrieving the phonological code of a word stored in memory. Wernicke's area plays a crucial role in the whole network of language processing by linking the lexical aspects of a word form to the widely distributed associations that define its meaning. This role is played by Wernicke's area in both language production and language comprehension.[18] The lexical word form information is relayed to Broca's area in the left frontal cortex and/or the middle part

[17] Peter Indefrey and W.J.M. Levelt, "The Neural Correlates of Language Production," in *The Cognitive Neurosciences*, 2nd edition, M. Gazzaniga, ed. (Cambridge: MIT Press, 2000).

[18] M-Marsel Mesulam, "From Sensation to Cognition," *Brain* 121 (1998): 1013–52.

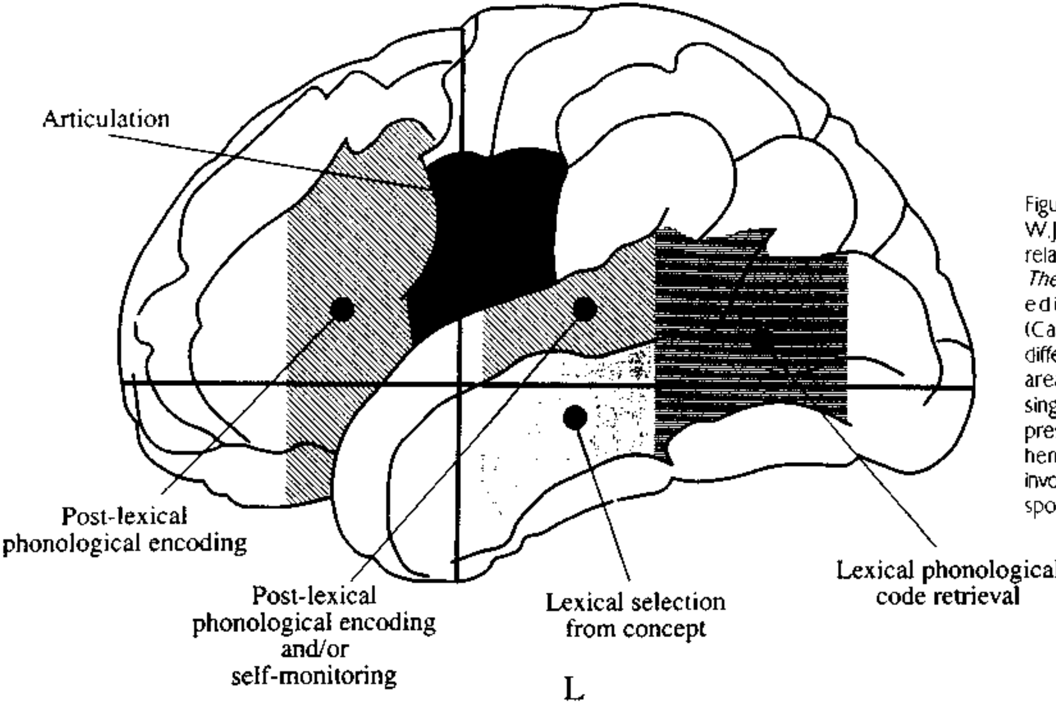


Figure 3. (After Peter Indefrey and W.J.M. Levelt, "The Neural Correlates of Language Production," in *The Cognitive Neurosciences*, 2nd edition, M. Gazzaniga, ed. (Cambridge: MIT Press, 2000.) The different functional roles of cortical areas of the left hemisphere in single word production. The figure presents a lateral view of the left hemisphere. Different areas are involved in different aspects of the spoken-word production process.

of the superior temporal lobe in the left hemisphere. These areas play a role in the conversion of the phonological codes in memory into phonological words from which the abstract articulatory program is derived. In the final phase of preparing for articulation and execution of articulation sensorimotor areas become activated, with the possible additional contribution of the supplementary motor area and the cerebellum (the latter two areas are not shown in figure 3).

The conceptual knowledge that we have about the words of our language seems to be distributed more widely than the lexical lemma and form aspects, and is not restricted to the left hemisphere. Moreover, brain imaging and lesion studies on the semantics of concrete nouns indicate that perceptual and functional attributes of word meaning might be accessed through different parts of the brain with the perceptual attributes closer to the primary sensory areas and the functional attributes closer to the motor cortex.[19] The transmodal cortex, the midtemporal cortex, and Wernicke's area are convergence zones or critical gateways for accessing relevant information that is represented in a distributed way.[20] On the basis of the overall organization of the cortex, these areas are well suited to gate our distributed conceptual knowledge into one word form. That all we know about John Paul II converges onto the single word form {po.p} requires the involvement of brain areas that are specialized for binding distributed fragments of knowledge into a single output, in this case a single word form.

Combining knowledge about the overall organization of the brain with information of specific patterns of language-related brain activity and its temporal dynamics allows us to gain insight into the neural organization of the uniquely human capacity for communication by means of natural language. Understanding this highly complex communication system requires a lot of skillful experimental research on detailed issues. Ultimately the understanding that we gain by doing this has wider ramifications for central questions concerning the human person.

4 Language and the Theory of the Person

A full theory of the person requires a specification of the ways in which the signals from such very different functional systems as language, memory, emotion, motor action, etc. with their own dedicated neural circuitry give rise to the sense of self and personhood. How exactly this happens is still largely unknown. How the brain solves the problem of binding the signals of these different systems into the sense of a unified self with continuity from past to future is an almost complete *terra incognita* for current cognitive neuroscience. Despite successful models of different cognitive systems, cognitive neuroscience still lacks an overarching theory of the person. However, even in the absence of a theory of the human person, it is a reasonable guess that such a theory would look quite different if we lacked language. Within the context of his schema theory, Michael Arbib (in this volume) argues that the self is a schema encyclopedia containing hundreds of thousands of schemas that a person uses to interpret and add new information to memory. The schema encyclopedia is used to "tell a story" to fit the new data. Not only are the metaphors used to describe

[19] For an overview, see Eleanor M. Saffran and A. Sholl, "Clues to the Functional and Neural Architecture of Word Meaning," in *Neurocognition of Language*, Brown and Hagoort, eds., 241–72.

[20] Mesulam, "From Sensation to Cognition."

the sense of self and person very often derived from the domain of language;[21] it is also clear that language allows us to increase the size of our schema encyclopedia at an amazing rate. Although a cognitive neuroscience of language does not explain the content of our schema encyclopedia, it is indispensable in explaining the machinery that allows us to build up this large schema encyclopedia. No doubt, in the absence of our language capacities our sense of self and person would be substantially more limited and boring. In this regard language is a key component of a theory of the person, for which the input from a cognitive neuroscience of language is thus much needed.

A similar story can be told about the relation between language and awareness. Although consciousness has also not been explained satisfactorily in terms of cognitive neuroscience, despite claims to the contrary,[22] awareness stands a better chance. Awareness is related to our ability to give phenomenal judgments and verbal reports about our sensations. Just as PET and fMRI allow a view of neural activity through a hemodynamic filter, awareness allows a view of consciousness through a linguistic filter. Understanding the characteristics of the filter is also in this case of crucial importance to a better understanding of the central but still mostly evanescent phenomenon of consciousness.

Acknowledgment. I am grateful to Michael Arbib for his comments on an earlier version of this essay.

[21] See also the proposal for the interpreter in Michael S. Gazzaniga, *Nature's Mind: The Biological Roots of Thinking, Emotions, Sexuality, Language, and Intelligence* (New York: BasicBooks, 1992).

[22] See David J. Chalmers, *The Conscious Mind: In Search of a Fundamental Theory* (Oxford: Oxford University Press, 1996), for a thought-provoking account.