

Chapter 4

The Speaking Mind/Brain: Where do Spoken Words Come From?

Willem J. M. Levelt and Peter Indefrey

The scientific study of how we speak saw substantial progress in the late twentieth century, but it has not been a major information source for functional brain-imaging studies of language production. First, these studies have been largely restricted to the production of isolated words. Second, even with this limitation in mind, the choice of experimental and control tasks has not been much informed by psycholinguistic theorizing on how we generate a phonological word from a conceptual base. The present chapter begins with a synopsis of such a theory of word production. It involves various processing components that function largely in staged succession, such as conceptually driven lexical selection and phonological encoding. That theory is used to reanalyze the experimental tasks used in a set of fifty-eight reported brain-imaging experiments on word production, tasks such as picture naming or spoken-word repetition. This task analysis is then used in a meta-analysis of the brain-activation data reported in all these studies. The exercise reveals that the cerebral network subserving the core processes in the production of words is almost strictly left-lateralized, with the exception of the bilateral involvement of sensorimotor and cerebellar areas. The midpart of the middle temporal gyrus has an involvement in conceptually driven lexical selection, Wernicke's area in the retrieval of phonological codes, Broca's and adjacent areas in phonological encoding (such as syllabification), the sensorimotor areas in phonetic encoding and articulation, and the midsuperior temporal lobe in phonological encoding and/or self-monitoring. There is no involvement of occipital, parietal, and prefrontal brain regions in the core processes of word production.

Nothing is more useful in cognitive neuroscience than a well-founded processing theory. This is particularly true for the study of the brain's ability to produce and understand language. Each linguistic task, such as understanding a relative clause or naming a picture, involves various component processes that run in some temporal relation to each other. These component processes perform particular linguistic functions—phonological, morphological, syntactic, semantic, and so on—making use of knowledge of various kinds. A sophisticated processing theory specifies the compu-

tations performed by these component processes, the representations they generate, and the information exchange between them, as well as the time course of their functioning. Given such a theory, the cognitive neuroscientist can experimentally vary linguistic tasks in such a way that a particular component is more or less involved and then register the concomitant variation in brain-activation patterns. If the theory is specific enough about the time course of the process, one can in addition relate the time course of the activation pattern to the independently established time course of the component processes..

In the initial phase of linguistic brain-imaging research, experimental subjects were perhaps inevitably presented with linguistic tasks whose componential structure is opaque. What, for instance, does one do if asked to generate as many words as possible beginning with "t"? Nobody has ever analyzed that task to find out. It involves some form of lexical access, to be sure, but does it also involve visual word imaging, and if so, is the visual activation necessary for lexical retrieval? How does the task relate to "normal" generation of language? Is it mostly a metalinguistic task? More often than not the choice of linguistic experimental and control tasks in PET and fMRI research has been based on global, ad hoc task analyses. The composition of a new task has hardly ever been independently analyzed as a prerequisite to the imaging study. The situation has, however, been much better in ERP research, where there is a "human performance" tradition of careful task analysis.

In this chapter we will report some core findings of a more extensive study (Indefrey and Levelt 2000) in which we reanalyzed the imaging data obtained in fiftyeight word-production experiments. This meta-analysis was guided by a detailed processing theory of word production, a theory that helped us decompose the main tasks used in the imaging studies. The theory, moreover, has sufficient temporal detail to allow for a further analysis of some activation timing data in the literature. We will first outline the processing theory, which is followed by the task analysis we based on it. In the third section we present some of the main findings of the componential and temporal analyses.

4.1 A Theory of Lexical Access in Speech Production

Most of the utterances we normally produce are multiword utterances. Words are typically generated as parts of phrases, clauses, and sentences. Any theory of word generation should be embedded in a theory of speaking—that is, in an account of how we formulate utterances in a larger discourse context. That holds for the theory of lexical access adopted here. It figures in the theory of speaking outlined in Levelt (1989) and has been further developed since (see Levelt 1999). We will not review that theory here, but rather focus on the lexical access aspect of it, making occasional reference to the way it is embedded in the larger framework. Also, we will

be relatively concise with respect to the lexical access theory itself, because a comprehensive statement of the theory, with all its experimental and computational detail, is available (Levelt, Roelofs, and Meyer 1999).

A first step in the generation of any utterance is *conceptual preparation*. There are always multiple ways to reveal a communicative intention to an interlocutor or audience. As speakers we continually make rhetorical decisions on what information to express for what purposes (Clark 1996). One aspect of this process is "perspective taking"—the way we decide to refer to some state of affairs (Levelt 1989,1996; Clark 1997). I can refer to the same person as *the woman*, *the phonologist*, *my daughter* and so on. The lexical concept¹ I select for the expression (woman, phonologist, daughter) depends on the discourse context. If I am aware that my interlocutor knows the person as a phonologist but is not privy to her being my daughter, opting for "phonologist" rather than for "daughter" would be logical. Choosing a particular lexical concept for expression is the beginning of lexical access. This first stage is depicted in figure 4.1 as *conceptual preparation*.

Whatever the information we select for expression, it must become encoded grammatically. This involves two major operations. The first is to select the appropriate words from the mental lexicon, among them one for each lexical concept to be expressed. The second is to generate a syntactic representation in which these words participate. There is good experimental evidence now that what we select in this first stage are not yet fully specified words, but rather *lemmas*—syntactic words. The activation of the words' phonological properties is a subsequent step. The lemma specifies a word's syntax, its syntactic category and subcategory (mass noun, transitive verb, and so on), and the way it maps argument structure onto syntactic relations. For instance, the lemma for the mental verb *hate* specifies how the argument structure *somebody hates something* should be syntactically realized: the *somebody* (or experiencer) argument should end up in subject position and the *something* (or stimulus) argument in object position. This mapping is verb specific. Other mental verbs, such as *scare*, do it the other way around (*something scares somebody*). On retrieval of the lemma, such lexically specified syntactic information will drive further grammatical encoding: the construction of phrase and clause structure and the linear ordering and inflection of selected lexical material. The lexical part of grammatical encoding is called *lexical selection*; it is depicted as the second stage in figure 4.1. Roelofs (1992) developed a detailed computational model of lexical selection. The model, which predicts the time course of lexical selection, has survived extensive experimental testing.

Once grammatical encoding is under way, by and large generating the syntactic pattern incrementally (Kempen 1997)—that is "from left to right"—*phonological code retrieval* follows as closely as possible. For each word entered into the budding syntactic frame, the phonological code is retrieved from the mental lexicon. The delay

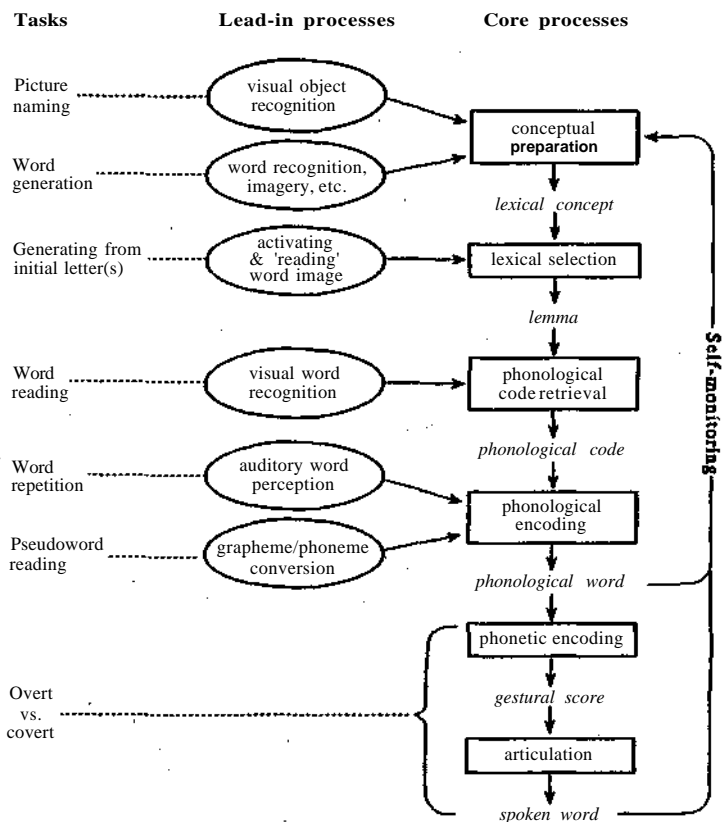


Figure 4.1

Processing stages involved in the generation of words (right column) and "lead-in" processes for various experimental word production tasks (left columns).

between retrieving the lemma and retrieving its phonological code may well be on the order of a mere 40 ms (van Turenout, Hagoort, and Brown 1998). The word's phonological code specifies (among other things) the phonological segments of which it is composed and their ordering. For instance, the phonological code of *hate* contains the string of segments /h/, /ei/, and /t/. Experimental measurements support the notion that a word's segments are retrieved simultaneously (though "strung together" successively—see below). Phonological codes are morphologically "packaged." The code of a monomorphemic word, such as *hale*, is a single package. However, when we retrieve a multimorphemic word, such as *blackboard*, we access the phonological code for *black* and another code for *board*. The speed of accessing these morpheme-size packages depends on the frequency of usage of the morpheme (which, in turn, is highly correlated with age of acquisition). For more detail on the process of morphological composition, see Levelt, Roelofs, and Meyer (1999).

As phonological codes become available, they become the ingredients of phonological encoding. This is the generation of the utterance form, from syllabic and phonological word composition to metrical phrasing and intonation. From the point of view of word production, the core process is the construction of phonological words. Phonological words are the domains of syllabification, and these often do not coincide with lexical words. For instance, in uttering the sentence *they hate us*, *hate* and *us* will blend into a single phonological word: a speaker will cliticize *us* to *hate*, which leads to the syllabification *ha-tus*. Here the last syllable *tus* straddles the lexical boundary between verb and pronoun. In *they hate ugliness*, however, *hate* keeps its integrity as a phonological word and becomes a whole syllable. In other words, the way a word syllabifies depends on the context in which it appears, which makes it unlikely that a word's syllables are specified in its stored phonological code. They are, rather, generated on the fly in the process of phonological encoding. There is good experimental evidence that this syllabification-in-context proceeds incrementally, from the first to the last segment of a phonological word (Levelt, Roelofs, and Meyer 1999).

Another aspect of encoding a phonological word is assigning word stress. The way this is done depends on the language. In French, for instance, it is always the last syllable that gets stressed. In English it tends to be the first full-voweled syllable. Syllabification and word stress are both aspects of word-form encoding that follow the retrieval of phonological word codes. In figure 4.1 this processing stage is labeled *phonological encoding*. As far as word formation is concerned, its output is a syllabified phonological word. In the larger utterance context it will be a constituent of a phonological phrase and of an intonational phrase.

The next processing component deals with *phonetic encoding*: the specification of articulatory gestures for syllables and phonological words in the larger utterance context. A major function of phonological encoding is to prepare a pronounceable

utterance. The context dependency of syllabification, for instance, serves to create syllables of moderate complexity. There is always a sonorous nucleus (mostly a vowel) in a syllable, and consonants are as much as possible grouped "to the left" of it, with increasing sonority from syllable onset to syllable nucleus. These are properties that facilitate the articulatory gesture (MacNeilage 1998). Although the details of syllabification still vary substantially among languages (some languages tolerate much more syllable complexity than others), it is rare within a language that an entirely new, never-used syllable arises during phonological encoding. Some languages, such as Chinese or Japanese, make use of only a few hundred different syllables. When we talk for an hour, we produce some 15,000 syllable tokens. Hence, most syllables will be heavily overused in such a language. But that also holds for languages such as English or Dutch that have far more than 10,000 different syllables. Statistics show that speakers of these languages also produce most of their speech by means of only a few hundred different high-frequency syllables (data from Schiller, reported in Levelt, Roelofs, and Meyer 1999). It is likely, therefore, that in most cases the phonetic encoding of a syllable consists of accessing an overlearned articulatory gesture, a "gestural score" that specifies the sequence of articulatory targets (such as lip closing or making a glottal stop) to be realized in articulation. The repository of these syllabic gestural scores has been called the speaker's *syllabary* (Levelt 1992). In addition, phonetic encoding involves setting some free parameters for these gestures (such as force and rate), the gestural planning of very low-frequency syllables or new syllables outside the syllabary, the smoothing of syllabic junctures within a phonological word, and so on. The output of phonetic encoding is the articulatory or gestural score (see figure 4.1).

The final step in producing an utterance is *articulation*. The high-speed execution of gestural scores by the laryngeal and supralaryngeal articulatory musculature is the most complicated motor behavior we can produce. To generate some ten to fifteen speech sounds per second, the articulators (such as tongue tip, tongue body, velum, and lips) are simultaneously active in approaching their target positions; it is co-articulation that makes our astonishing speech rate possible (Lieberman 1996). Articulation is, moreover, a sensorimotor event. Target positions can be approached in multiple ways (such as jaw vs. lip movement in lip closure). The articulators follow a least-effort solution, given the physical contingencies. The way a target position is going to be approached not only depends on the starting position of the relevant articulator (i.e., on the outcome of the previous gesture), but also on more external conditions, such as having food in the mouth or talking while resting the chin on the hand. Fast sensory feedback loops exist for the fine tuning of articulation.

Conceptual preparation, lexical selection, phonological code retrieval, phonological encoding, phonetic encoding, and articulation are the main stages in the generation of a spoken word. We have called these the *core processes*. The experimental evidence allows for rough estimations of the time windows in which these component

processes operate (see Levelt et al. 1998). In fast picture naming, for instance, visual object recognition plus conceptual preparation takes about 150 msec, lexical selection another 125 msec, accessing the phonological code and the phonological encoding of a word again take about 125 msec, and phonetic encoding until the initiation of articulation lasts about 200 msec, for a total of some 600 msec.

As is the case for any complicated motor action, producing speech involves some degree of self-monitoring. We cannot help hearing our own voice while speaking, and when we attend to it, we may detect errors or other infelicities that may interfere with our intentions. We can then decide to stop and make repairs (Levelt 1983). There is good evidence for the assumption that we can also monitor the internal phonological score—that is, the incremental process of phonological syllabification (Wheeldon and Levelt 1995). This allows us to intercept potential trouble before it is given phonetic or articulatory shape. Both external and internal feedback loops (see figure 4.1) involve our normal speech understanding system (see McGuire, Silbersweig, and Frith 1996 for brain-imaging evidence).

4.2 An Analysis of Tasks Used in Brain-Imaging Studies of Word Production

Many different tasks have been used in brain-imaging studies of word production. Two of them, picture naming and verb/noun generation, involve all the core processes discussed in the previous section, but others, such as auditory word repetition, involve only the later stages of word production. To clarify the task structure involved in these different experimental methods, Indefrey and Levelt (2000) distinguished between core processes and so-called lead-in processes. Each task has a way of entering the word production mechanism; Picture naming, for example, is initiated through visual object recognition. That triggers the activation of one or more lexical concepts; from there the core processes take their further course. Here visual object recognition is the lead-in process. As another example, in pseudoword reading (that is, reading nonwords, such as *virsof*), the lead-in process is visual letter recognition and grapheme-to-phoneme conversion. That leads into the core process of phonological encoding—the speaker will do normal syllabification on the activated phoneme string (the example nonword will be syllabified as *vir-sof* with stress on the first syllable). In this task, therefore, the initial core processes of conceptual preparation, lexical selection, and phonological code retrieval are not involved.²

For most tasks used in brain-imaging studies, the lead-in processes have been left unanalyzed. Earlier we gave an example of generating words from initial letters, but similar uncertainties exist for most other experimental tasks. A happy exception is the often-used task of picture naming. The lead-in process of visual object recognition is unproblematic. It is, moreover, an entirely natural lead-in process. Naming objects is a case of quite normal language use. Word reading is another natural case in our culture, and the word reading process is well understood. Other tasks, however, are

far from natural. Widely used is the task of verb generation. Here the subject is visually presented with a noun, such as *hammer*, and instructed to produce a verb that expresses a use of the object (for instance, *hit*). The subject will probably imagine some action scene involving a hammer and then pick out the action for conceptual preparation, but this is a mere guess. In noun generation the subject is asked to generate exemplars of a semantic category, such as *tools*. Both tasks are so-called word-fluency tasks, and together we will label them *word generation*. For our analysis it is not essential to understand the lead-in process in much detail. It is necessary only to make a sophisticated guess about which of the core processes is the entry point for a given task. Despite the obscure lead-in process in verb generation, it presumably triggers conceptual preparation. In this respect verb generation is not different from picture naming, and the same holds for noun generation. The left side of figure 4.2 (also plate 1) presents our view of the lead-in processes and of their entry points into the core processes of word production. Picture naming and word generation (both verb and noun) enter at conceptual preparation. Generating from initial letter(s), which may involve the recognition of imagined visual words (see Friedman et al. 1998 for a recent task analysis), probably enters at the level of lexical selection (the subject must select one word rather than another). In word reading there will, normally, be activation of the word's phonological code. Though one cannot exclude the involvement of higher-level semantic processes in word reading, it certainly does not involve (normal) conceptually driven lexical selection. The word repetition task probably enters at the level of phonological encoding, although the phonological code is not necessarily retrieved. Pseudoword reading, as we have already mentioned, involves phonological encoding not based on a retrieved phonological code.

A final, major task variable in the literature is overt versus silent word production. We will assume that phonetic encoding and articulation are involved in overt word production tasks, but not, or much less so, in "silent" production tasks since one cannot exclude the possibility that an experimental subject engages in some "covert mumbling" when instructed to silently generate the target word.

This task analysis allows us to perform critical contrasts among functional brain-imaging data obtained with different word production tasks. For instance, the operation of phonological code retrieval is involved in picture naming, verb generation, generating from initial letters, and word reading, but not (or substantially less so) in pseudoword reading. Any brain region involved in all the former tasks but not in the latter is a potential site involved in phonological code retrieval.

4.3 A Componential Analysis of Cerebral Localizations: Some Main Findings

The full data analysis (Indefrey and Levelt 2000) involved fifty-eight different functional brain-imaging experiments in thirty-five studies of word production (see

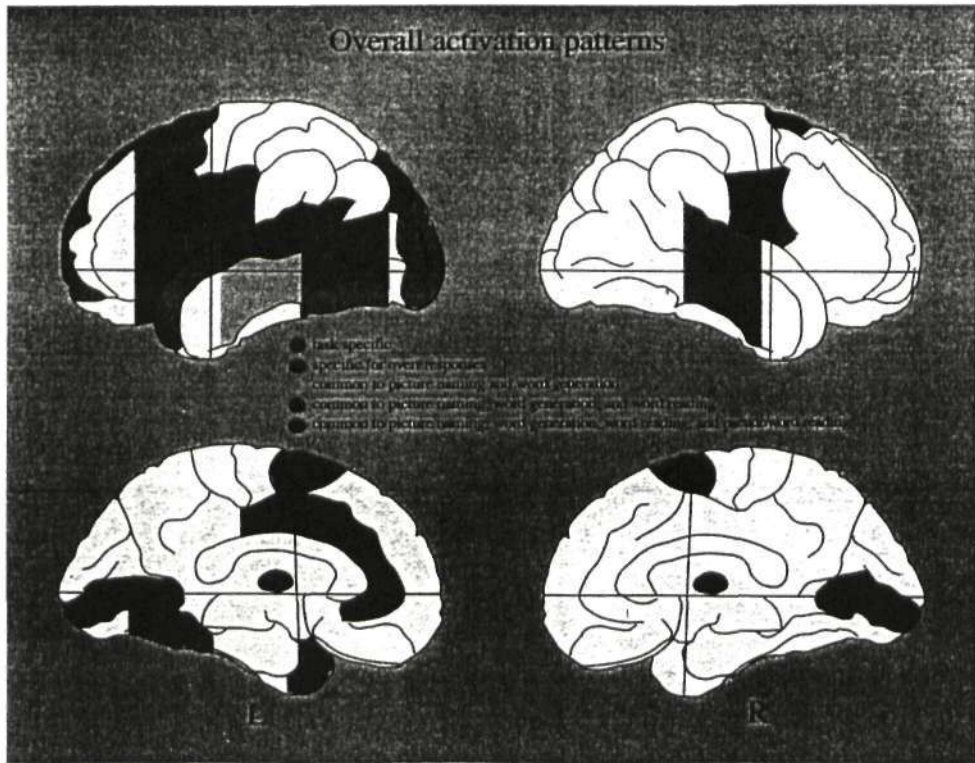


Figure 4.2

Overall left and right hemisphere activation patterns for various tasks in the meta-analysis. Pink, red, and blue colored regions (see plate 1) are involved with the core processes of word production.

Appendix), the majority of which were PET or fMRI studies. Our descriptive reference system for the coding of active loci in the brain was based on the stereotactic atlas of Talairach and Tournoux (1988). In terms of Talairach coordinates, we defined three domains in the frontal lobes— anterior ($y > 34$), posterior ($34 > y > 0$), and "motor" ($y < 0$); three domains in the temporal lobes— anterior ($y > -7$), mid ($-7 > y > -38$), and posterior ($y < -38$); three domains in the parietal lobes— "sensory" ($y > -23$), anterior ($-23 > y > -48$), and posterior ($y < -48$); and two domains in the occipital lobes— medial ($x > |25|$) and lateral ($x < |25|$). In most cases the reported activation data allowed for more fine-grained distinctions in terms of gyri and subcortical structures. Where possible, we also differentiated between anterior, mid, and posterior cingulum, medial and lateral cerebellum, as well as anterior and posterior insula. This gave us a total of 104 regions.

We also needed criteria to mark loci as "involved" in a particular set of critical experimental data. There were, on average, 8.8 activation sites reported per experiment. In other words, one could expect 1 in about 10 of the 104 possible loci to be active in a particular experiment. We expressed the agreement among studies about the activation of a particular brain region in terms of a binomial distribution. If, for any task contrast we made, the number of reports mentioning a particular active region exceeded a chance probability of 10%, we took the region to be "involved" in the critical component process.

In the first step of our analysis, we focused on distinguishing the core processes from the lead-in processes. The tasks of picture naming and of word generation involve the same core processes; they differ only in their lead-ins. Moreover, if word production is overt, both tasks involve *all* the core processes distinguished in the previous section. Hence, the activations shared in these two tasks should give us an initial picture of the brain regions involved in the core processes of word production. However, in many cases the studies used silent generation tasks, especially in word generation; hence, we risked losing the regions involved in phonetic encoding and articulation. To get the complete picture, we additionally opposed the overt and silent tasks, whatever they had been. The results of these contrasts are shown in figure 4.2 (plate 1).

The green-colored areas are involved in task-specific—that is, lead-in—processing. The nongreen areas are the ones involved in the core processes of word production. These areas are strictly left-lateralized, except for the sensorimotor areas specifically active in overt responding (marked in blue). The lateralized word-production network consists of Broca's area (the posterior inferior frontal gyrus), the midsuperior and middle temporal gyri, Wernicke's area (the posterior superior and middle temporal gyri), and the (left) thalamus.

The next step in the analysis was to partition this network in more detail. Beginning at the top component in figure 4.1 (conceptual preparation), we needed to contrast

picture naming and word generation, on the one hand, with generation from initial letter(s) on the other hand. However, we could not reach criterion here because there were only two experiments with the initial letter task in the reported literature. Hence, we could only package conceptual preparation and lexical selection (henceforth *conceptually driven lexical selection*) together by opposing word reading to picture naming and verb generation. The distinguishing region here turned out to be the midsegment of the left middle temporal gyrus (dark pink in plate 1). Vandenberghe et al. (1996) indeed found this area to be part of a "common semantic system" involved in both word and object processing. However, in the concluding section of this chapter, we will argue that this is most probably a serious underestimation of the regions involved in conceptually driven lexical selection.

The stage of phonological code retrieval can be focused on by contrasting picture naming, word generation, and word reading with pseudoword reading, as discussed in the previous section. This contrast yields a clear result: the regions involved are Wernicke's area and the left thalamus. The finding that Wernicke's area is involved in phonological code retrieval is in excellent agreement with the results of an MEG study by Levelt et al. (1998) (see below).

Phonological encoding is the processing stage that all tasks, down to pseudoword reading (see figure 4.1) have in common. The regions most involved in all these tasks are the left posterior inferior frontal gyrus, including Broca's area and operculum, and the midsuperior temporal gyrus (both marked dark red in figure 4.2, plate 1). However, the small number of studies that used word generation from initial letters, word repetition, and pseudoword reading made it impossible to reach criterion in all cases. If the involvement of these regions can be reconfirmed in further research, the emerging picture for the phonological encoding network is that Wernicke's area provides access to lexical phonological codes (called *Wortvorstellungen* by Wernicke); they are transmitted to the midsuperior temporal lobe and, via the arcuate fasciculus, to Broca's area for further postlexical syllabification. Damage to the arcuate fasciculus indeed tends to block all speech output, except for a few recurring utterances (Nina Dronkers, University of California, Davis, personal communication). As will be argued later, the midsuperior temporal lobe may also be involved in a speaker's self-monitoring.

A comparison of all overt speech production tasks (without overt controls) to all silent tasks showed the bilateral involvement of sensorimotor areas (the blue regions in Figure 4.2, plate 1), and that is how it should be for the component processes of phonetic encoding and articulation. In addition, there was significant involvement of the left anterior superior temporal gyrus. This, however, only occurred in the picture-naming task; hence it would be wrong to say that this region is generally involved in overt word production. Similarly, the involvement of the right supplementary motor area (SMA) also reached our statistical criterion, but it was practically absent in

picture naming tasks. Hence, we cannot claim general involvement in overt word production for this region either. Finally, both the left and right lateral and medial cerebellum are involved in the execution of overt word production tasks.

Turning now to the few published time course studies of word generation, we can use the time window estimates for the successive component processes involved in picture naming, presented above: roughly 150 msec for visual processing and conceptual preparation, 125 msec for lexical selection, another 125 msec for phonological encoding (including code retrieval), and some 200 msec for phonetic preparation and the initiation of articulation. Salmelin et al. (1994) and Levelt et al. (1998), both using MEG, found medial and increasingly lateral occipital activation corresponding to the first time window and extending into the second time window. In the study of picture naming by Levelt and associates, the time window for phonological encoding corresponded to a clustering of active dipole sources in Wernicke's area, which may signal the process of phonological code retrieval. The phonetic-articulatory time window, finally, corresponded to bilateral sensorimotor activation in the Salmelin et al. and Levelt et al. studies, but there was activation in other areas as well, in particular bilateral parietal in the Levelt et al. study. Preoperative subdural grid data by Crone et al. (1994) give information about the late time windows, after 300 ms postpicture onset. In picture-naming and overt word-repetition tasks, activation was measured in the left posterior and midtemporal lobes (there was no recording from the right hemisphere). Taken together, the timing data are consistent with the visual lead-in activation findings and the phonological encoding data from the PET and fMRI studies, but they are not yet very helpful in focusing on regions involved in conceptual and lemma processing.

4.4 Conclusions

The meta-analysis reported by Indefrey and Levelt (2000) and summarized in the present chapter is based on a detailed theory of lexical access in speaking, comprehensively reviewed in Levelt, Roelofs, and Meyer 1999. That theory was not in the minds of most of the research teams when they performed the experimental studies that we analyzed. On the contrary, their analyses of the experimental tasks used were in general quite minimal or ad hoc, mostly involving some interpretation of differences between experimental and control tasks as "semantic," "phonological," or the like. It is not at all self-evident that a post hoc analysis of the same experimental findings in terms of a different and more detailed processing model should yield any consistent result, but it does. This may plead for the processing model to start with, but it testifies as well to the quality and reliability of the reported experimental data.

Figure 4.3 (plate 2) summarizes the core findings of our meta-analysis. The network subserving the production of words is almost completely left-lateralized, exceptions being the bilateral sensorimotor regions involved in the phonetic encoding and/or

Summary

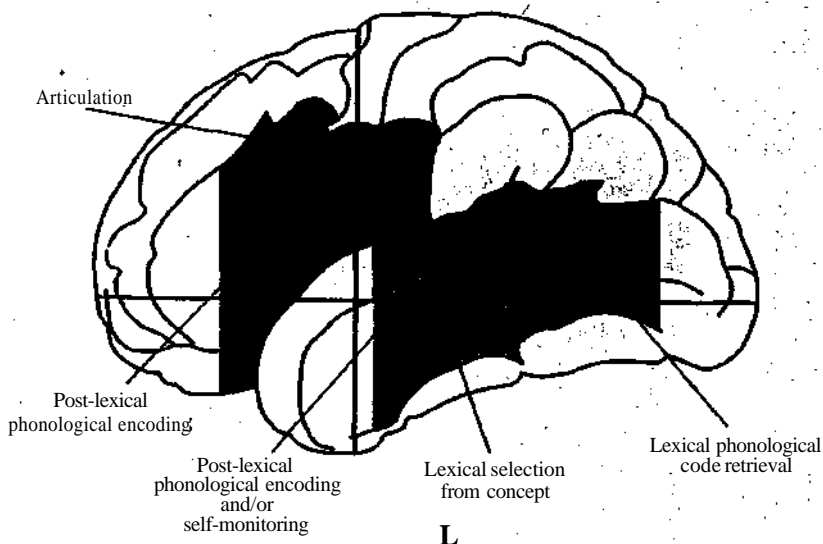


Figure 4.3

Summary of Regions activated by various core processes in word production.

articulation of the target words and the bilateral cerebellar involvement in overt word production. The lead-in processes of different tasks can involve quite divergent cortical areas, such as the left prefrontal lobe in word generation or the occipital lobes in picture naming. However, the core processes of word generation appear to involve a stable network encompassing temporal and frontal lobe areas, sensorimotor areas, the left thalamus, and the cerebellum bilaterally. Under the statistical criteria we used, there is no evidence for the involvement of occipital lobes, parietal lobes, or prefrontal lobes in that basic network.

The processing theory provides a tentative interpretation of the regions involved in the network, as shown in figure 4.3 (plate 2). For the first steps in lexical access, conceptually driven lexical selection, the only area that met our criteria is the midpart of the middle temporal gyrus. Although we referred to independent evidence for an involvement of this area in semantic word processing, it is also likely that this will not be the full story. There is increasing evidence that the semantic processing involved in the generation of words from different semantic domains, such as tools, animals, or vegetables, at least partly involves different, category-specific brain regions (Damasio

et al. 1996; Martin et al. 1996). Such brain regions will be obscured in our type of analysis.

The next processing step, retrieving a word's phonological code, appears to involve Wernicke's area. The convergence of evidence for this interpretation is encouraging. Not only do timing data from an MEG study support it, but there is also good evidence for Wernicke's area involvement in auditory word processing (Price et al. 1996). Further phonological encoding, such as rapid, incremental syllabification, involves Broca's and/or neighboring areas and the midsuperior temporal gyrus, according to our meta-analysis. The latter has not been noticed before as playing a part in phonological encoding, but the convergence of evidence is on first view substantial: the area was reported as active in thirty-eight out of fifty-seven relevant experimental word production studies that involved phonological encoding. However, care is necessary here. There is one other process, apart from phonological encoding, that is involved in all word production tasks: self-monitoring. If we are right in stipulating that self-monitoring implicates the speech comprehension system, one should expect to find midsuperior temporal lobe activation in most word production tasks, in accordance with the findings by McGuire, Silbersweig, and Frith (1996). In other words, in the present analysis we cannot really distinguish between a phonological encoding function of the superior temporal lobe and its potential role in the self-monitoring of the output of phonological encoding.

The meta-analysis indicates that phonetic encoding and articulation involve bilateral sensorimotor areas, a finding that confirms the reliability of our analysis procedure. Still, there may be more involved in phonetic encoding. If indeed overlearned motor programs, such as patterns of syllabic gestures, are activated during the execution of speech, one would expect the involvement of premotor areas (Rizzolatti and Gentilucci 1988) and/or supplementary motor areas. Our gross division of three domains in the frontal lobes does not allow us to distinguish BA 6 from motor area BA 4, and often no more detail was presented in the individual studies either. Hence, the involvement of premotor cortex is, so far, not at all excluded.

If a post hoc application of a detailed processing model to data gathered from quite different perspectives already shows so much convergence, how much more can be achieved if the processing theory is used beforehand to guide the planning of functional brain-imaging studies of language?

Notes

1. A *lexical concept* is a concept for which a word exists in the target language.
2. There is evidence though that nonwords, whether spoken or written, do momentarily activate real words that are similar (see Norris and Wise 2000 for a warning of this kind). The printed pseudoword *virsof*, for instance, may to some extent activate a word such as *virtue*. Here we will assume that nontarget word activation will normally be of a lesser degree than target word activation.

References

- Clark, E. V. 1997. Conceptual perspective and lexical choice in language acquisition. *Cognition* 64:309-343.
- Clark, H. 1996. *Using language*. Cambridge, England: Cambridge University Press.
- Crone, N. E., J. Hart, Jr., D. Boatman, R. P. Lesser, and B. Gordon. 1994. Regional cortical activation during language and related tasks identified by direct cortical electrical recording. Paper presented at the annual meeting of the Academy of Aphasia.
- Damasio, H., T. J. Grabowski, D. Tranel, R. D. Hichwa, and A. R. Damasio. 1996. A neural basis for lexical retrieval. *Nature* 380:499-505.
- Friedman, L., J. T. Kenny, A. L. Wise, D. Wu, T. A. Stuve, D. A. Miller, J. A. Jesberger, and J. S. Lewin. 1998. Brain activation during silent word generation evaluated with functional MRI. *Brain and Language* 64:231-256.
- Indefrey, P., and W. J. M. Levelt. 2000. The neural correlates of language production. In *The new cognitive neurosciences* (2nd ed.), M. Gazzaniga, ed. Cambridge, Mass.: MIT Press.
- Kempen, G. 1997. Grammatical performance in human sentence production and comprehension. Ms., University of Leiden.
- Levelt, W. J. M. 1983. Monitoring and self-repair in speech. *Cognition* 14:41-104.
- Levelt, W. J. M. 1989. *Speaking: From intention to articulation*. Cambridge, Mass.: MIT Press.
- Levelt, W. J. M. 1992. Accessing words in speech production: Stages, processes, and representations. *Cognition* 42:1-22.
- Levelt, W. J. M. 1996. Perspective taking and ellipsis in spatial descriptions. In *Language and space*, P. Bloom, M. A. Peterson, L. Nadel, and M. F. Garrett, eds., 77-107. Cambridge, Mass.: MIT Press.
- Levelt, W. J. M. 1999. Language production: A blueprint of the speaker. In *Neurocognition of language*, C. Brown and P. Hagoort, eds. Oxford, England: Oxford University Press.
- Levelt, W. J. M., P. Praamstra, A. S. Meyer, P. Helenius, and R. Salmelin. 1998. An MEG study of picture naming. *Journal of Cognitive Neuroscience* 10:553-567.
- Levelt, W. J. M., A. Roelofs, and A. S. Meyer. 1999. A theory of lexical access in speech production. *Behavior and Brain Sciences* 22:1-37.
- Lieberman, A. 1996. *Speech: A special code*. Cambridge, Mass.: MIT Press.
- MacNeilage, P. F. 1998. The frame/content theory of evolution of speech production. *Behavioral and Brain Sciences*, 21:499-511.
- Martin, A., C. L. Wiggs, L. G. Ungerleider, and J. V. Haxby. 1996. Neural correlates of category-specific knowledge. *Nature* 379:649-652.
- McGuire, P. K., D. A. Silbersweig, and C. D. Frith. 1996. Functional neuroanatomy of verbal self-monitoring. *Brain* 119:101-111.
- Norris, D., and R. Wise. 2000. The study of prelexical and lexical processes in comprehension: Psycholinguistic and functional neuroimaging. In *The cognitive neurosciences* (2nd ed.), M. Gazzaniga, ed. Cambridge, Mass.: MIT Press.
- Price, C. J., R. J. S. Wise, E. A. Warburton, C. J. Moore, D. Howard, K. Patterson, R. S. J. Frackowiak, and K. J. Friston. 1996. Hearing and saying: The functional neuro-anatomy of auditory word processing. *Brain* 119:919-931.

Rizolatti, G., and M. Gentilucci. 1988. Motor and visual-motor functions of the premotor cortex. In *Neurobiology of motor cortex*, P. Rakic and W. Singer, eds. Chichester, England: Wiley.

Roelofs, A. 1992. A spreading-activation theory of lemma retrieval in speaking. *Cognition* 42:107-142.

Salmelin, R., R. Hari, O. V. Lounasmaa, and M. Sams. 1994. Dynamics of brain activation during picture naming. *Nature* 368:463-465.

Talairach, J., and P. Tournoux. 1988. *Co-planar stereotaxic atlas of the human brain*. Stuttgart: Georg Thieme Verlag.

Vandenberghe, R., C. Price, R. Wise, O. Josephs, and R. S. J. Frackowiak. 1996. Functional anatomy of a common semantic system for words and pictures. *Nature* 383:254-256.

van Turenhout, M., P. Hagoort, and C. M. Brown, 1998. Brain activity during speaking: From syntax to phonology in 40 milliseconds. *Science* 280:572-574.

Wheeldon, L. R., and W. J. M. Levelt. 1995. Monitoring the time course of phonological encoding. *Journal of Memory and Language* 34:311-334.

Appendix

Publications involved in the meta-analysis and their experimental word production tasks

<u>Task</u>	<u>Publication</u>	
Picture naming aloud	Ojemann 1983	Ojemann et al. 1989
	<i>Beh Brain. Sc.</i> 2:189-230	<i>J. Neurosurg.</i> 71:316-326
	Schäffler et al. 1993	Crone et al. 1994
	<i>Brain</i> 116:695-715	<i>Paper Acad Aphasia</i>
	Haglund et al. 1994	Salmelin et al. 1994
	<i>Neurosurgery</i> 34:567-516	<i>Nature</i> 368:463-465
	Abdullaev and Melnichuk 1995	Bookheimer et al. 1995
Techn. Rep. 95-09 U. Oregon	<i>Hum. Brain Map.</i> 3:93-106	
Picture naming silent	Damasio et al. 1996	Levelt et al. (1998)
	<i>Nature</i> 380:499-505	<i>J. Cog. Neurosc.</i> 10:553-567
	Bookheimer et al. 1995	Martin et al. 1996
Word generation silent, verbs	<i>Hum. Brain Map.</i> 3:93-105	<i>Nature</i> 379:649-652
	Price, Moore, Humphreys et al. 1996	
	<i>Proc R. Soc. Lond B</i> 263:1501-1507	
Word generation silent, nouns	Wise et al. 1991	Crivello et al. 1995
	<i>Brain</i> 114:1803-1817	<i>Neuroimage</i> 2:253-263
	Poline et al. 1996	Warburton et al. 1996
Generation from initial lette(s)	<i>Neuroimage</i> 4:34-54	<i>Brain</i> 119:159-179
	Warburton et al. 1996	Paulesu et al. 1997
Generation from initial lette(s)	<i>Brain</i> 119:159-179	<i>NeuroReport</i> 8:2011-2016
	Buckner et al. 1995	Paulesu et al. 1997
	<i>J. Neurophysiol.</i> 74:2163-2173	<i>NeuroReport</i> 8:2011-2016

Appendix (continued)

<u>Task</u>	<u>Publication</u>	
Word reading aloud	Ojemann 1983 <i>Beh. Brain. Sc.</i> 2:189-230	Howard et al. 1992 <i>Brain</i> 115:1769-1782
	Sakurai et al. 1992 <i>NeuroReport</i> 4:327-330	Sakurai et al. 1993 <i>NeuroReport</i> 3:445-448
	Price et al. 1994 <i>Brain</i> 117:1255-1269	Bookheimer et al. 1995 <i>Hum. Brain Map.</i> 3:93-106
	Price, Moore, and Frackowiak 1996 <i>Neuroimage</i> 3:40-52	
	Gordon et al. 1997 W. Hulstijn (ed.) Speech Production. Elsevier	Herbster et al. 1997 <i>Hum. Brain Map.</i> 5:84-92
	Rumsey et al. 1997 <i>Brain</i> 120:739-759	
Word reading silent	Petersen et al. 1989 <i>J. Cog. Neurosc.</i> 7:153-170	Petersen et al. 1990 <i>Science</i> 249:1041-1044
	Bookheimer et al. 1995 <i>Hum. Brain Map.</i> 5:93-105	Menard et al. 1996 <i>Neuropsycholog</i> 34:185-194
	Price et al. 1996 <i>Neuroimage</i> 3:40-52	Beauregard et al. 1997 <i>J. Cog. Neurosci</i> 9:441-461
	Hagoort et al. (submitted)	
Pseudoword reading aloud	Sakurai et al. 1993 <i>NeuroReport</i> 4:327-330	Indefrey et al. 1996 <i>Neuroimage</i> 3:S442
	Herbster et al. 1997 <i>Hum. Brain Map.</i> 5:84-92	Rumsey et al. 1997 <i>Brain</i> 120:739-759
Pseudoword reading silent	Petersen et al. 1990 <i>Science</i> 249:1041-1044	Fujimaki et al. 1996 Hashimoto (ed.) <i>Visualization.</i> Elsevier
	Hagoort et al. (submitted)	
Word repetition aloud	Petersen et al. 1989 <i>J. Cog. Neurosc.</i> 1:153-170	Howard et al. 1992 <i>Brain</i> 115:1769-1782
	Crone et al. 1994 <i>Paper Acad. Aphasia</i>	Price et al. 1996 <i>Brain</i> 119:919-931
	Gordon et al. 1997 W. Hulstijn (ed.) Speech Production. Elsevier	
Pseudoword repetition silent	Warburton et al. 1996 <i>Brain</i> 119:159-179	

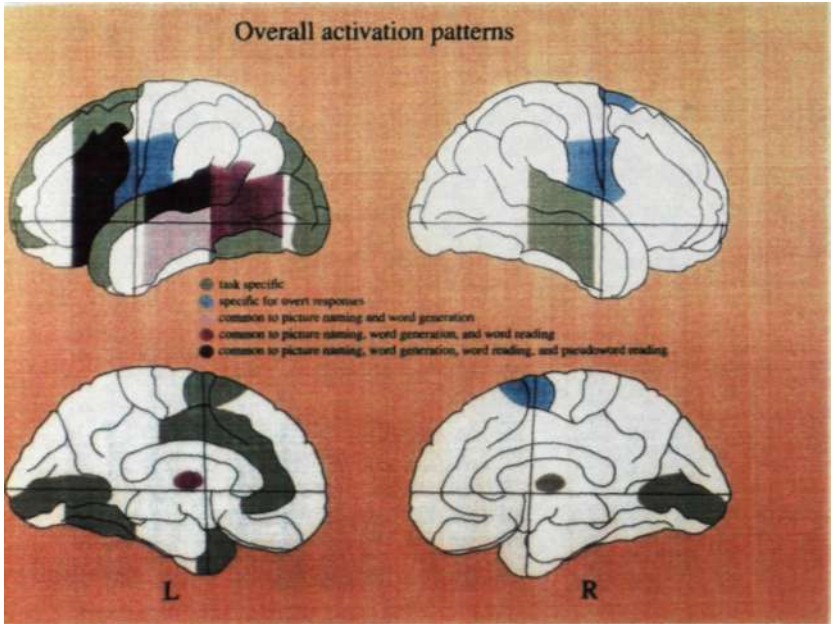


Plate 1

Overall left and right hemisphere activation patterns for various tasks in the meta-analysis. Pink, red, and blue colored regions are involved with the core processes of word production.

Summary

