BEHAVIORAL AND BRAIN SCIENCES (1999) **22**, 1–75
Printed in the United States of America

A theory of lexical access in speech production

Willem J. M. Levelt
Max Planck Institute for Psycholinguistics,
6500 AH Nijmegen, The Netherlands
pim@mpi.nl www.mpi.nl

Ardi Roelofs
School of Psychology, University of Exeter,
Washington Singer Laboratories, Exeter EX4 4QG, United Kingdom
a.roelofs@exeter.ac.uk

Antje S. Meyer
Max Planck Institute for Psycholinguistics,
6500 AH Nijmegen, The Netherlands
asmeyer@mpi.nl

Abstract: Preparing words in speech production is normally a fast and accurate process. We generate them two or three per second in fluent conversation; and overtly naming a clear picture of an object can easily be initiated within 600 msec after picture onset. The underlying process, however, is exceedingly complex. The theory reviewed in this target article analyzes this process as staged and feedforward. After a first stage of conceptual preparation, word generation proceeds through lexical selection, morphological and phonological encoding, phonetic encoding, and articulation itself. In addition, the speaker exerts some degree of output control, by monitoring of self-produced internal and overt speech. The core of the theory, ranging from lexical selection to the initiation of phonetic encoding, is captured in a computational model, called WEAVER++. Both the theory and the computational model have been developed in interaction with reaction time experiments, particularly in picture naming or related word production paradigms, with the aim of accounting for the real-time processing in normal word production. A comprehensive review of theory, model, and experiments is presented. The model can handle some of the main observations in the domain of speech errors (the major empirical domain for most other theories of lexical access), and the theory opens new ways of approaching the cerebral organization of speech production by way of high-temporal-resolution imaging.

Keywords: articulation, brain imaging, conceptual preparation, lemma, lexical access, lexical concept, lexical selection, magnetic encephalography, morpheme, morphological encoding, phoneme, phonological encoding, readiness potential, self-monitoring, speaking, speech error, syllabification, WEAVER++

1. An ontogenetic introduction

Infants (from Latin *infans,* speechless) are human beings who cannot speak. It took most of us the whole first year of our lives to overcome this infancy and to produce our first few meaningful words, but we were not idle as infants. We worked, rather independently, on two basic ingredients of word production. On the one hand, we established our primary notions of agency, interactancy, the temporal and causal structures of events, object permanence and location. This provided us with a matrix for the creation of our first *lexical concepts,* concepts flagged by way of a verbal label. Initially, these word labels were exclusively auditory patterns, picked up from the environment. On the other hand, we created a repertoire of babbles, a set of syllabic articulatory gestures. These motor patterns normally spring up around the seventh month. The child carefully attends to their acoustic manifestations, leading to elaborate exercises in the repetition and concatenation of these syllabic patterns. In addition, these audiomotor patterns start resonating with real speech input, becoming more and more tuned to the mother tongue (De Boysson-Bardies & Vihman 1991; Elbers 1982). These exercises provided us with a *protosyllabary,* a core repository of speech motor patterns, which were, however, completely meaningless.

Real word production begins when the child starts connecting some particular babble (or a modification thereof) to some particular lexical concept. The privileged babble auditorily resembles the word label that the child has acquired perceptually. Hence word production emerges from a coupling of two initially independent systems, a conceptual system and an articulatory motor system.

This duality is never lost in the further maturation of our word production system. Between the ages of 1;6 and 2;6 the explosive growth of the lexicon soon overtaxes the protosyllabary. It is increasingly hard to keep all the relevant whole-word gestures apart. The child conquers this strain on the system by dismantling the word gestures through a process of *phonemization;* words become generatively represented as concatenations of phonological segments (Elbers & Wijnen 1992; C. Levelt 1994). As a consequence, phonetic encoding of words becomes supported by a system of phono-

0140-525X/99 $12.50

logical encoding. Adults produce words by spelling them out as a pattern of phonemes and as a metrical pattern. This more abstract representation in turn guides phonetic encoding, the creation of the appropriate articulatory gestures.

The other, conceptual root system becomes overtaxed as well. When the child begins to create multiword sentences, word order is entirely dictated by semantics, that is, by the prevailing relations between the relevant lexical concepts. One popular choice is "agent first"; another one is "location last." However, by the age of 2;6 this simple system starts foundering when increasingly complicated semantic structures present themselves for expression. Clearly driven by a genetic endowment, children restructure their system of lexical concepts by a process of *syntactization.* Lexical concepts acquire syntactic category and subcategorization features; verbs acquire specifications of how their semantic arguments (such as agent or recipient) are to be mapped onto syntactic relations (such as subject or object); nouns may acquire properties for the regulation of syntactic agreement, such as gender, and so forth. More technically speaking, the child develops a system of *lemmas,*[1] packages of syntactic information, one for each lexical concept. At the same time, the child quickly acquires a closed class vocabulary, a relatively small set of frequently used function words. These words mostly fulfill syntactic functions; they have elaborate lemmas but lean lexical concepts. This system of lemmas is largely up and running by the age of 4 years. From then on, producing a word always involves the selection of the appropriate lemma.

The original two-pronged system thus develops into a four-tiered processing device. In producing a content word, we, as adult speakers, first go from a lexical concept to its lemma. After retrieval of the lemma, we turn to the word's phonological code and use it to compute a phonetic-articulatory gesture. The major rift in the adult system still reflects the original duality in ontogenesis. It is between the lemma and the word form, that is, between the word's syntax and its phonology, as is apparent from a range of phenomena, such as the tip-of-the-tongue state (Levelt 1993).

left to right: Willem J. Levelt, Ardi Roelofs, and Antje S. Meyer.

Willem J. M. Levelt is founding director of the Max Planck Institute for Psycholinguistics, Nijmegen, and Professor of Psycholinguistics at Nijmegen University. He has published widely in psychophysics, mathematical psychology, linguistics and psycholinguistics. Among his monographs are *On Binocular Rivalry* (1968), *Formal Grammars in Linguistics and Psycholinguistics* (3 volumes, 1974) and *Speaking: From Intention to Articulation* (1989).

Ardi Roelofs, now a Lecturer at the School of Psychology of the University of Exeter, obtained his Ph.D. at Nijmegen University on an experimental/computational study of lexical selection in speech production, performed at the Nijmegen Institute for Cognition and Information and the Max Planck Institute. After a postdoctoral year at MIT, he returned to the Max Planck Institute as a member of the research staff to further develop the WEAVER computational model of lexical selection and phonological encoding.

Antje S. Meyer is a Senior Research Fellow of the Max Planck Institute for Psycholinguistics, Nijmegen. After completing her Ph.D. research at the Max Planck Institute, she held postdoctoral postions at the University of Rochester and the University of Arizona. Returning to Nijmegen, she continues her experimental research in word and sentence production at Nijmegen University and the Max Planck Institute. She is presently coordinating a project using eye tracking to study the allocation of attention in scene descriptions.

2. Scope of the theory

In the following, we will first outline this word producing system as we conceive it. We will then turn in more detail to the four levels of processing involved in the theory: the activation of lexical concepts, the selection of lemmas, the morphological and phonological encoding of a word in its prosodic context, and, finally, the word's phonetic encoding. In its present state, the theory does not cover the word's articulation. Its domain extends no further than the initiation of articulation. Although we have recently been extending the theory to cover aspects of lexical access in various syntactic contexts (Meyer 1996), the present paper is limited to the production of isolated prosodic words (see note 4).

Every informed reader will immediately see that the theory is heavily indebted to the pioneers of word production research, among them Vicky Fromkin, Merrill Garrett, Stephanie Shattuck-Hufnagel, and Gary Dell (see Levelt, 1989, for a comprehensive and therefore more balanced review of modern contributions to the theory of lexical access). It is probably in only one major respect that our approach is different from the classical studies. Rather than basing our theory on the evidence from speech errors, spontaneous or induced, we have developed and tested our notions almost exclusively by means of reaction time (RT) research. We believed this to be a necessary addition to existing methodology for a number of reasons. Models of lexical access have always been conceived as process models of normal speech production. Their ultimate test, we argued in Levelt et al. (1991b) and Meyer (1992), cannot lie in how they account for infrequent derailments of the process but rather must lie in how they deal with the normal process itself. RT studies, of object naming in particular, can bring us much closer to this ideal. First, object naming is a normal, everyday activity indeed, and roughly one-fourth of an adult's lexicon consists of names for objects. We admittedly start tampering with the natural process in the laboratory, but that hardly ever results in substantial derailments, such as naming errors or tip-of-the-tongue states. Second, reaction time measurement is still an ideal procedure for analyzing the time course of a mental process (with evoked potential methodology as a serious competitor). It invites the development of real-time process models, which not only predict the ultimate outcome of the process but also account for a reaction time as the result of critical component processes.

RT studies of word production began with the seminal studies of Oldfield and Wingfield (1965) and Wingfield (1968; see Glaser, 1992, for a review), and RT methodology is now widely used in studies of lexical access. Still, the theory to be presented here is unique in that its empirical scope is in the temporal domain. This has required a type of modeling rather different from customary modeling in the domain of error-based theories. It would be a misunderstanding, though, to consider our theory as neutral with respect to speech errors. Not only has our theory construction always taken inspiration from speech error analyses, but, ultimately, the theory *should* be able to account for error patterns as well as for production latencies. First efforts in that direction will be discussed in section 10.

Finally, we do not claim completeness for the theory. It is tentative in many respects and is in need of further development. We have, for example, a much better understanding of access to open class words than of access to closed class words. However, we do believe that the theory is productive in that it generates new, nontrivial, but testable predictions. In the following we will indicate such possible extensions when appropriate.

3. The theory in outline

3.1. Processing stages

The flow diagram presented in Figure 1 shows the theory in outline. The production of words is conceived as a staged process, leading from conceptual preparation to the initiation of articulation. Each stage produces its own characteristic output representation. These are, respectively, lexical concepts, lemmas, morphemes, phonological words, and phonetic gestural scores (which are executed during articulation). In the following it will be a recurring issue whether these stages overlap in time or are strictly sequential, but here we will restrict ourselves to a summary description of what each of these processing stages is supposed to achieve.

3.1.1. Conceptual preparation. All open class words and most closed class words are meaningful. The intentional[2] production of a meaningful word always involves the activation of its lexical concept. The process leading up to the activation of a lexical concept is called "conceptual preparation." However, there are many roads to Rome. In everyday language use, a lexical concept is often activated as part of a larger message that captures the speaker's communicative intention (Levelt 1989). If a speaker intends to refer to a female horse, he may effectively do so by producing the word "mare," which involves the activation of the lexical concept MARE(X). But if the intended referent is a female elephant, the English speaker will resort to a phrase, such as "female elephant," because there is no unitary lexical concept available for the expression of that notion. A major issue, therefore, is how the speaker gets from the notion/information to be expressed to a message that consists of lexical concepts (here *message* is the technical term for the conceptual structure that is ultimately going to be formulated). This is called the *verbalization problem,* and there is no simple one-to-one mapping of notions-to-be-expressed onto messages (Bierwisch & Schreuder 1992). Even if a single lexical concept is formulated, as is usually the case in object naming, this indeterminacy still holds, because there are multiple ways to refer to the same object. In picture naming, the same object may be called "animal," "horse," "mare," or what have you, depending on the set of alternatives and on the task. This is called *perspective taking.* There is no simple, hard-wired connection between percepts and lexical concepts. That transition is always mediated by pragmatic, context-dependent considerations. Our work on perspective taking has, until now, been limited to the lexical expression of spatial notions (Levelt 1996), but see E. Clark (1997) for a broader discussion.

Apart from these distal, pragmatic causes of lexical concept activation, our theory recognizes more proximal, semantic causes of activation. This part of the theory has been modeled by way of a conceptual network (Roelofs 1992a; 1992b), to which we will return in sections 3.2 and 4.1. The top level in Figure 2 represents a fragment of this network. It depicts a concept node, ESCORT(X, Y), which stands for the meaning of the verb "escort." It links to other concept nodes, such as ACCOMPANY (X, Y), and the links are labeled to express the character of the connection – in this case, IS-TO, because to ESCORT (X, Y) is to ACCOMPANY (X, Y). In this network concepts will spread their activation via such links to semantically related concepts. This mechanism is at the core

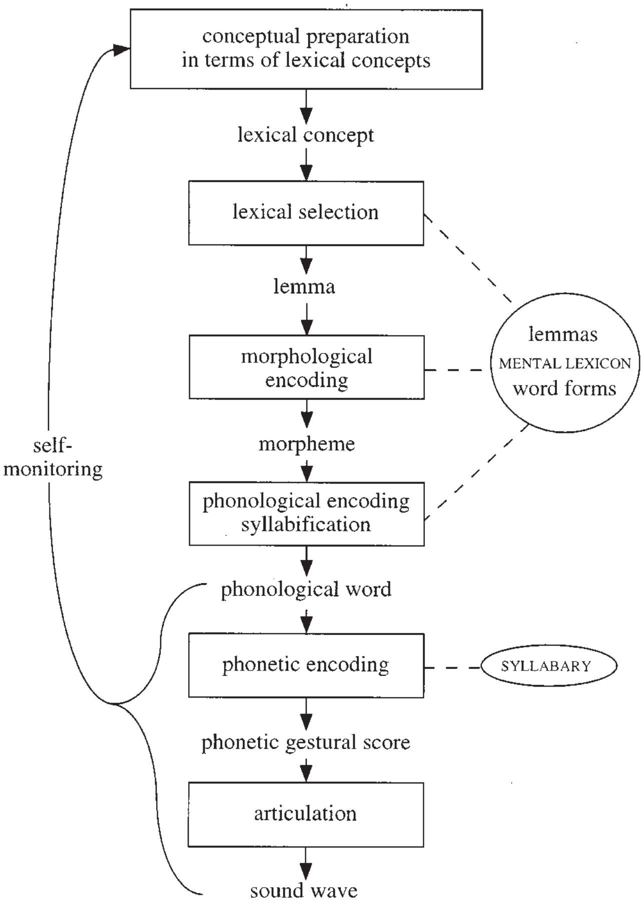


Figure 1. The theory in outline. Preparing a word proceeds through stages of conceptual preparation, lexical selection, morphological and phonological encoding, and phonetic encoding before articulation can be initiated. In parallel there occurs output monitoring involving the speaker's normal speech comprehension mechanism.

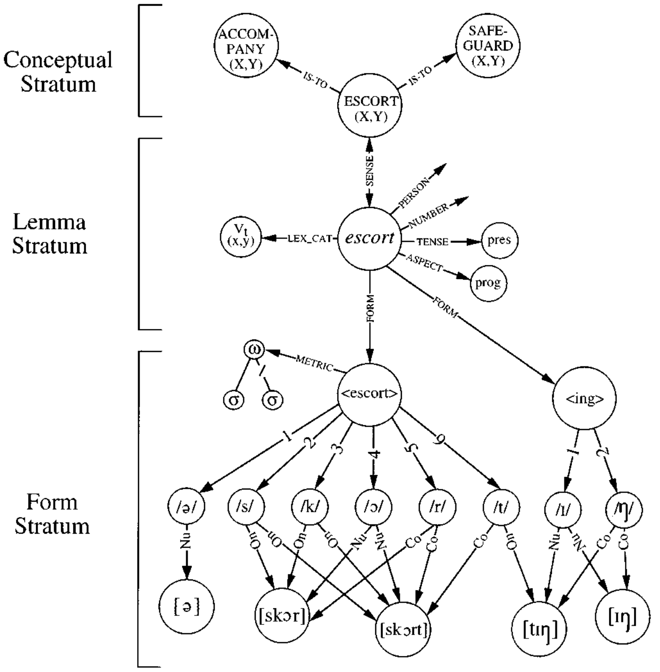


Figure 2. Fragment of the lexical network underlying lexical access. The feedforward activation spreading network has three strata. Nodes in the top, conceptual stratum represent lexical concepts. Nodes in the lemma stratum represent syntactic words or lemmas and their syntactic properties. Nodes in the form stratum represent morphemes and their phonemic segments. Also at this level there are syllable nodes.

of our theory of lexical selection, as developed by Roelofs (1992a). A basic trait of this theory is its nondecompositional character. Lexical concepts are not represented by sets of semantic features because that creates a host of counterintuitive problems for a theory of word production. One is what Levelt (1989) has called the *hyperonym problem.* When a word's semantic features are active, then, *per definition,* the feature sets for all of its hyperonyms or superordinates are active (they are subsets). Still, there is not the slightest evidence that speakers tend to produce hyperonyms of intended target words. Another problem is the nonexistence of a semantic complexity effect. It is not the case that words with more complex feature sets are harder to access in production than words with simpler feature sets (Levelt et al. 1978). These and similar problems vanish when lexical concepts are represented as undivided wholes.

The conceptual network's state of activation is also measurably sensitive to the speaker's auditory or visual word input (Levelt & Kelter 1982). This is, clearly, another source of lexical concept activation. This possibility has been exploited in many of our experiments, in which a visual or auditory distractor word is presented while the subject is naming a picture.

Finally, Dennett (1991) suggested a pandemonium-like spontaneous activation of words in the speaker's mind. Although we have not modeled this, there are three ways to implement such a mechanism. The first would be to add spontaneous, statistical activation to lexical concepts in the network. The second would be to do the same at the level of lemmas, whose activation can be spread back to the conceptual level (see below). The third would be to implement spontaneous activation of word forms; their resulting morphophonological encoding would then feed back as internal speech (see Fig. 1) and activate the corresponding lexical concepts.

3.1.2. Lexical selection. Lexical selection is retrieving a word, or more specifically a lemma, from the mental lexicon, given a lexical concept to be expressed. In normal speech, we retrieve some two or three words per second from a lexicon that contains tens of thousands of items. This high-speed process is surprisingly robust; errors of lexical selection occur in the one per one thousand range. Roelofs (1992a) modeled this process by attaching a layer of lemma nodes to the conceptual network, one lemma node for each lexical concept. An active lexical concept spreads some of its activation to "its" lemma node, and lemma selection is a statistical mechanism, which favors the selection of the highest activated lemma. Although this is the major selection mechanism, the theory does allow for the selection of function words on purely syntactic grounds (as in "John said *that* . . .", where the selection of *that* is not conceptually but syntactically driven). Upon selection of a lemma, its syntax becomes available for further grammatical encoding, that is, creating the appropriate syntactic environment for the word. For instance, retrieving the lemma *escort* will make available that this is a transitive verb [node $V_t(x, y)$ in Fig. 2] with two argument positions (x and y), corresponding to the semantic arguments X and Y, and so on.[3]

Many lemmas have so-called diacritic parameters that have to be set. For instance, in English, verb lemmas have features for number, person, tense, and mood (see Fig. 2). It is obligatory for further encoding that these features are valued. The lemma *escort,* for instance, will be phonologically realized as *escort, escorts, escorted, escorting,* depending on the values of its diacritic features. The values of these features will in part derive from the conceptual representation. For example, tense being an obligatory feature in English, the speaker will always have to check the relevant temporal properties of the state of affairs being expressed. Notice that this need not have any communicative function. Still, this extra bit of thinking has to be done in preparation of any tensed expression. Slobin (1987) usefully called this "thinking for speaking." For another part, these diacritic feature values will be set during grammatical encoding. A verb's number feature, for instance, is set by agreement, in dependence on the sentence subject's number feature. Here we must refrain from discussing these mechanisms of grammatical encoding (but see Bock & Levelt 1994; Bock & Miller 1991; and Levelt 1989 for details).

3.1.3. Morphophonological encoding and syllabification. After having selected the syntactic word or lemma, the speaker is about to cross the rift mentioned above, going from the conceptual/syntactic domain to the phonological/articulatory domain. The task is now to prepare the appropriate articulatory gestures for the selected word in its prosodic context, and the first step here is to retrieve the word's phonological shape from the mental lexicon. Crossing the rift is not an entirely trivial matter. The tip-of-the-tongue phenomenon is precisely the momentary inability to retrieve the word form, given a selected lemma. Levelt (1989) predicted that in a tip-of-the tongue state the word's syntactic features should be available in spite of the blockage, because they are lemma properties. In particular, a Dutch or an Italian speaker should know the grammatical gender of the target word. This has recently been experimentally demonstrated by Vigliocco et al. (1997) for Italian speakers. Similarly, certain types of anomia involve the same inability to cross this chasm. Badecker et al. (1995) showed this to be the case for an Italian anomic patient, who could hardly name *any* picture, but *always* knew the target word's grammatical gender. However, even if word form access is unhampered, production is much harder for infrequent words than for frequent words; the difference in naming latency easily amounts to 50–100 msec. Jescheniak and Levelt (1994) showed that word form access is the major, and probably unique, locus of the word frequency effect (discovered by Oldfield & Wingfield 1965).

According to the theory, accessing the word form means activation of three kinds of information, the word's morphological makeup, its metrical shape, and its segmental makeup. For example, if the lemma is *escort,* diacritically marked for progressive tense, the first step is to access the two morphemes <escort> and <ing> (see Fig. 2). Then, the metrical and segmental properties of these morphemes will be "spelled out." For *escort,* the metrical information involves that the morpheme is iambic, that is, that it is disyllabic and stress-final, and that it can be a phonological word[4] (ω) itself. For <ing> the spelled out metrical information is that it is a monosyllabic, unstressed morpheme, which cannot be an independent phonological word (i.e., it must become attached to a phonological head, which in this case will be *escort).* The segmental spell out for <escort> will be /ə/[5], /s/, /k/, /ɔ/, /r/, /t/, and for <ing> it will be /ɪ/, /ŋ/ (see Fig. 2). Notice that there are no syllables at this level. The syllabification of the phonological word *escort* is *e-scort,* but this is not stored in the mental lexicon. In the theory, syllabification is a late process, because it often depends on the word's phonological environment. In *escorting,* for instance, the syllabification is different: *e-scor-ting,* where the syllable *ting* straddles the two morphemes *escort* and *ing.* One might want to argue that the whole word form *escorting* is stored, including its syllabification. However, syllabification can also transcend lexical word boundaries. In the sentence *He'll escort us,* the syllabification will usually be *e-scor-tus.* It is highly unlikely that this cliticized form is stored in the mental lexicon. An essential part of the theory, then, is its account of the syllabification process. We have modeled this process by assuming that a morpheme's segments or phonemes become *simultaneously* available, but with labeled links indicating their correct ordering (see Fig. 2). The word's metrical template may stay as it is, or be modified in the context. In the generation of *escorting* (or *escort us,* for that matter), the "spelled out" metrical templates for <escort>, $\sigma\sigma'$, and for <ing> (or <us>), σ, will merge to form the trisyllabic template $\sigma\sigma'\sigma$. The spelled-out segments are successively inserted into the current metrical template, forming phonological syllables "on the fly": *e-scor-ting* (or *e-scor-tus).* This process follows quite universal rules of syllabification (such as maximization of onset and sonority gradation; see below) as well as language-specific rules. There can be no doubt that these rules are there to create maximally pronounceable syllables. The domain of syllabification is called the "phonological" or "prosodic word" (ω). *Escort, escorting,* and *escortus* can be phonological words, that is, domains of syllabification. Some of the phonological syllables in which *escort,* in different contexts, can participate are represented in Figure 2. If the current phonological word is *escorting,* the relevant phonological syllables, *e, scor,* and *ting,* with word accent on *scor,* will activate the phonetic syllable scores [ə], [skɔr], and [tɪŋ].

3.1.4. Phonetic encoding. The theory has an only partial account of phonetic encoding. The theoretical aim is to explain how a phonological word's gestural score is computed. It is a specification of the articulatory task that will produce the word, in the sense of Browman and Goldstein (1992).[6] This is a (still rather abstract) representation of the articulatory gestures to be performed at different articulatory tiers, a glottal tier, a nasal tier, and an oral tier. One task, for instance, on the oral tier would be to close the lips (as should be done in a word such as *apple*). The gestural score is abstract in that the way in which a task is performed is highly context dependent. Closing the lips after [æ], for instance, is a quite different gesture than closing the lips after rounded [u].

Our partial account involves the notion of a syllabary. We assume that a speaker has access to a repository of gestural scores for the frequently used syllables of the language. Many, though by no means all, of the coarticulatory properties of a word are syllable-internal. There is probably more gestural dependence within a word's syllables than between its syllables (Browman & Goldstein 1988; Byrd 1995; 1996). More importantly, as we will argue, speakers of English or Dutch – languages with huge numbers of syllables – do most of their talking with no more than a few hundred syllables. Hence, it would be functionally advantageous for a speaker to have direct access to these frequently used and probably internally coherent syllabic scores. In this theory they are highly overlearned gestural patterns, which need not be recomputed time and again. Rather, they are ready-made in the speaker's syllabary. In our computational model, these syllabic scores are activated by the segments of the phonological syllables. For instance, when the active /t/ is the onset of the phonological syllable /tiŋ/, it will activate all syllables in the syllabary that contain [t], and similarly for the other segments of /tiŋ/. A statistical procedure will now favor the selection of the gestural score [tiŋ] among all active gestural scores (see sect. 6.3), whereas selection failures are prevented by the model's binding-by-checking mechanism (sect. 3.2.3). As phonological syllables are successively composed (as discussed in the previous section), the corresponding gestural scores are successively retrieved. According to the present, partial, theory, the phonological word's articulation can be initiated as soon as all of its syllabic scores have been retrieved.

This, obviously, cannot be the full story. First, the speaker can compose entirely new syllables (e.g., in reading aloud a new word or nonword). It should be acknowledged, though, that it is a very rare occasion indeed when an adult speaker of English produces a new syllable. Second, there may be more phonetic interaction between adjacent syllables within a word than between the same adjacent syllables that cross a word boundary. Explaining this would either require larger, word-size stored gestural scores or an additional mechanism of phonetic composition (or both).

3.1.5. Articulation. The phonological word's gestural score is, finally, executed by the articulatory system. The functioning of this system is beyond our present theory. The articulatory system is, of course, not just the muscular machinery that controls lungs, larynx, and vocal tract; it is as much a computational neural system that controls the execution of abstract gestural scores by this highly complex motor system (see Levelt 1989, for a review of motor control theories of speech production and Jeannerod, 1994, for a neural control theory of motor action).

3.1.6. Self-monitoring. The person to whom we listen most is ourself. We can and do monitor our overt speech output. Just as we can detect trouble in our interlocutor's speech, we can discover errors, dysfluencies, or other problems of delivery in our own overt speech. This, obviously, involves our normal perceptual system (see Fig. 1). So far, this ability is irrelevant for the present purposes. Our theory extends to the initiation of articulation, not beyond. However, this is not the whole story. It is apparent from spontaneous self-repairs that we can also monitor our "internal speech" (Levelt 1983), that is, we can monitor some internal representation as it is produced during speech encoding. This might have some relevance for the latency of spoken word production because the process of self-monitoring can affect encoding duration. In particular, such self-monitoring processes may be more intense in experiments in which auditory distractors are presented to the subject. More important, though, is the possibility to exploit this internal self-monitoring ability to trace the process of phonological encoding itself. A crucial issue here is the nature of "internal speech." What kind of representation or code is it that we have access to when we monitor our "internal speech"? Levelt (1989) proposed that it is a phonetic representation, the output of phonetic encoding. Wheeldon and Levelt (1995), however, obtained experimental evidence for the speaker's ability also to monitor a slightly more abstract, phonological representation (in accordance with an earlier suggestion by Jackendoff 1987). If this is correct, it gives us an additional means of studying the speaker's syllabification process (see sect. 9), but it also forces us to modify the original theory of self-monitoring, which involved phonetic representations and overt speech.

3.2. General design properties

3.2.1. Network structure. As is already apparent from Figure 2, the theory is modeled in terms of an essentially feedforward activation-spreading network. In particular, Roelofs (1992a; 1993; 1994; 1996a; 1996b; 1997c) instantiated the basic assumptions of the theory in a computational model that covers the stages from lexical selection to syllabary access. The word-form encoding part of this computational model is called WEAVER (for Word-form Encoding by Activation and VERification; see Roelofs 1996a; 1996b; 1997c), whereas the full model, including lemma selection, is now called WEAVER++.

WEAVER++ integrates a spreading–activation-based network with a parallel object-oriented production system, in the tradition of Collins and Loftus (1975). The structure of lexical entries in WEAVER++ was already illustrated in Figure 2 for the word *escort.* There are three strata of nodes in the network. The first is a conceptual stratum, which contains concept nodes and labeled conceptual links. A subset of these concepts consists of *lexical c*oncepts; they have links to lemma nodes in the next stratum. Each lexical concept, for example ESCORT(X, Y), is represented by an independent node. The links specify conceptual relations, for example, between a concept and its superordinates, such as IS-TO-ACCOMPANY(X, Y). A word's meaning or, more precisely, *sense* is represented by the total of the lexical concept's labeled links to other concept nodes. Although the modeling of the conceptual stratum is highly specific to this model, no deep ontological claims about "network semantics" are intended. We need only a mechanism that ultimately provides us with a set of active, nondecomposed lexical concepts.

The second stratum contains lemma nodes, such as *escort;* syntactic property nodes, such as $V_t(x, y)$; and labeled links between them. Each word in the mental lexicon, simple or complex, content word or function word, is represented by a lemma node. The word's syntax is represented by the labeled links of its lemma to the syntax nodes. Lemma nodes have diacritics, which are slots for the specification of free parameters, such as person, number, mood, or tense, that are valued during the process of grammatical encoding. More generally, the lemma stratum is linked to a set of procedures for grammatical encoding (not discussed herein).

After a lemma's selection, its activation spreads to the third stratum, the word-form stratum. The word-form stratum contains morpheme nodes and segment nodes. Each morpheme node is linked to the relevant segment nodes. Notice that links to segments are numbered (see Fig. 2). The segments linked to *escort* are also involved in the spell-out of other word forms, for instance, *Cortes,* but then the links are numbered differently. The links between segments and syllable program nodes specify possible syllabifications. A morpheme node can also be specified for its prosody, the stress pattern across syllables. Related to this morpheme/segment stratum is a set of procedures that generate a phonological word's syllabification, given the syntactic/phonological context. There is no fixed syllabification for a word, as was discussed above. Figure 2 represents one possible syllabification of *escort,* but we could have chosen another; /skɔrt/, for instance would have been a syllable in the citation form of *escort.* The bottom nodes in this stratum represent the syllabary addresses. Each node corresponds to the gestural score of one particular syllable. For *escorting,* these are the phonetic syllables [ə], [skɔr] and [tɪŋ].

What is a "lexical entry" in this network structure? Keeping as close as possible to the definition given by Levelt (1989, p. 182), a lexical entry is an item in the mental lexicon, consisting of a lemma, its lexical concept (if any), and its morphemes (one or more) with their segmental and metrical properties.

3.2.2. Competition but no inhibition. There are no inhibitory links in the network, either within or between strata. That does not mean that node selection is not subject to competition within a stratum. At the lemma and syllable levels the state of activation of nontarget nodes does affect the latency of target node selection, following a simple mathematical rule (see Appendix).

3.2.3. Binding. Any theory of lexical access has to solve a *binding problem.* If the speaker is producing the sentence *Pages escort kings,* at some time the lemmas *page* and *king* will be selected. How to prevent the speaker from erroneously producing *Kings escort pages*? The selection mechanism should, in some way, bind a selected lemma to the appropriate concept. Similarly, at some later stage, the segments of the word forms <king> and <page> are spelled out. How to prevent the speaker from erroneously producing *Cages escort pings*? The system must keep track of /p/ belonging to *pages* and /k/ belonging to *kings.* In most existing models of word access (in particular that of Dell 1988 and Dell et al. 1993), the binding problem is solved by *timing.* The activation/deactivation properties of the lexical network guarantee that, usually, the "intended" element is the most activated one at the crucial moment. Exceptions precisely explain the occasional speech errors. Our solution (Roelofs 1992a; 1993; 1996b; 1997c) is a different one. It follows Bobrow and Winograd's (1977) "procedural attachment to nodes." Each node has a procedure attached to it that checks whether the node, when active, links up to the appropriate active node one level up. This mechanism will, for instance, discover that the activated syllable nodes [pɪŋz] and [kɛɪ] do not correspond to the word form nodes <kings> and <pages>, and hence should not be selected.[7] For example, in the phonological encoding of *kings,* the /k/ but not the /p/ will be selected and syllabified, because /k/ is linked to <king> in the network and /p/ is not. In phonetic encoding, [kɪŋz] will be selected, because the links in the network between [kɪŋz] and its segments correspond with the syllable positions assigned to these segments during phonological encoding. For instance, /k/ will be syllabified as onset, which corresponds to the link between /k/ and [kɪŋz] in the network. We will call this "binding-by-checking" as opposed to "binding-by-timing."

A major reason for implementing binding-by-checking is the recurrent finding that, during picture naming, distractor stimuli hardly ever induce systematic speech errors. When the speaker names the picture of a king, and simultaneously hears the distractor word *page,* he or she will produce neither the semantic error *page,* nor the phonological error *ping,* although both the lemma *page* and the phoneme /p/ are strongly activated by the distractor. This fact is more easily handled by binding-by-checking than through binding-by-timing. A perfect binding-by-checking mechanism will, of course, prevent *any* speech error. A systematic account of speech errors will require our theory to allow for lapses of binding, as in Shattuck-Hufnagel's (1979) "check off" approach.

3.2.4. Relations to the perceptual network. Though distractor stimuli do not induce speech errors, they are highly effective in modulating the speech production process. In fact, since the work by Schriefers et al. (1990), picture–word interference has been one of our main experimental methods. The effectiveness of word primes implicates the existence of activation relations between perceptual and production networks for speech. These relations have traditionally been an important issue in speech and language processing (see Liberman 1996): Are words produced and perceived by the same mechanism or by different mechanisms, and, if the mechanisms are different, how are they related? We will not take a position, except that the feedforward assumption for our form stratum implies that form perception and production cannot be achieved by the same network, because this would require both forward and backward links in the network. An account of the theoretical and empirical motivation of the distinction between form networks for perception and production can be found elsewhere (Roelofs et al. 1996). Interestingly, proponents of backward links in the form stratum for production (Dell et al. 1997b) have also argued for the position that the networks are (at least in part) different. Apart from adopting this latter position, we have only made some technical, though realistic, assumptions about the way in which distractor stimuli affect our production network (Roelofs et al. 1996). They are as follows.

Assumption 1 is that a distractor word, whether spoken or written, affects the corresponding morpheme node in the production network. This assumption finds support in evidence from the word perception literature. Spoken word recognition obviously involves phonological activation (McQueen et al. 1995). That visual word processing occurs along both visual and phonological pathways has time and again been argued (see, e.g., Coltheart et al. 1993; Seidenberg & McClelland 1989). It is irrelevant here whether one mediates the other; what matters is that there is phonological activation in visual word recognition. This phonological activation, we assume, directly affects the state of activation of phonologically related morpheme units in the form stratum of the production network.

Assumption 2 is that active phonological segments in the perceptual network can also directly affect the corresponding segment nodes in the production lexicon. This assumption is needed to account for phonological priming effects by nonword distractors (Roelofs, submitted a).

Assumption 3 is that a spoken or written distractor word can affect corresponding nodes at the lemma level. Because recognizing a word, whether spoken or written, involves accessing its syntactic potential, that is, the perceptual equivalent of the lemma, we assume activation of the corresponding lemma-level node. In fact, we will bypass this issue here by assuming that all production lemmas *are* perceptual lemmas; the perceptual and production networks coincide from the lemma level upwards. However, the lemma level is not affected by active units in the form stratum of the production network, whether or not their activation derives from input from the perceptual network; there is no feedback here.

A corollary of these assumptions is that one should expect cohort-like effects in picture-distractor interference. These effects are of different kinds. First, it follows from assumption 3 that there can be semantic cohort effects of the following type. When the word "accompany" is the distractor, it will activate the joint perception/production lemma *accompany* (see Fig. 2). This lemma will spread activation to the corresponding lexical concept node ACCOMPANY(X, Y) (as it always does in perception). In turn, the concept node will coactivate semantically related concept nodes, such as the ones for ESCORT(X, Y) and SAFEGUARD(X, Y). Second, there is the possibility of phonological cohort effects, both at the form level and at the lemma level. When the target word is "escort" there will be relative facilitation by presenting "escape" as a distractor. This comes about as follows. In the perceptual network "escape" initially activates a phonological cohort that includes the word form and lemma of "escort" (for evidence concerning form activation, see Brown 1990 and, for lemma activation, see Zwitserlood 1989). According to assumption 1, this will activate the word form node <escort> in the production network. Although there is the possibility that nonword distractors follow the same route (e.g., the distractor "esc" will produce the same initial cohort as "escape"), assumption 2 is needed to account for the facilitating effects of spoken distractors that correspond to a word-final stretch of the target word. Meyer and Schriefers (1991), for instance, obtained facilitation of naming words such as "hammer" by presenting a distractor such as "summer," which has the same word-final syllable. For all we know, this distractor hardly activates "hammer" in its perceptual cohort, but it will speed up the segmental spell-out of all words containing "mer" in the production network. In an as yet unpublished study, Roelofs and Meyer obtained the same facilitation effect when only the final syllable (i.e., "mer") was used as a distractor.

3.2.5. Ockham's razor. Both the design of our theory and the computational modeling have been guided by Ockham's methodological principle. The game has always been to work from a minimal set of assumptions. Processing stages are strictly serial: there is neither parallel processing nor feedback between lexical selection and form encoding (with the one, still restricted, exception of self-monitoring); there is no free cascading of activation through the lexical network; there are no inhibitory connections in the network; WEAVER++'s few parameters were fixed on the basis of initial data sets and then kept constant throughout all further work (as will be discussed in sects. 5.2 and 6.4). This minimalism did not emanate from an a priori conviction that our theory is right. It is, rather, entirely methodological. We wanted theory and model to be maximally vulnerable. For a theory to be empirically productive, it should *forbid* certain empirical outcomes to arise. In fact, a rich and sophisticated empirical search has been arising from our theory's ban on activation spreading from an active but nonselected lemma (see sect. 6.1.1) as well as from its ban on feedback from word form encoding to lexical selection (see sect. 6.1.2), to give just two examples. On the other hand, we have been careful not to claim superiority for our serial stage reaction time model compared to alternative architectures of word production on the basis of good old additive factors logic (Sternberg 1969). Additivity does not uniquely support serial stage models; nonserial explanations of additive effects are sometimes possible (McClelland 1979; Roberts & Sternberg 1993). Rather, we had to deal with the opposite problem. How can apparently interactive effects, such as semantic/phonological interaction in picture/word interference experiments (sect. 5.2.3) or the statistical overrepresentation of mixed semantic/phonological errors (sect. 6.1.2), still be handled in a serial stage model, without recourse to the extra assumption of a feedback mechanism?

4. Conceptual preparation

4.1. Lexical concepts as output

Whatever the speaker tends to express, it should ultimately be cast in terms of lexical concepts, that is, concepts for which there exist words in the target language. In this sense, lexical concepts form the terminal vocabulary of the speaker's message construction. That terminal vocabulary is, to some extent, language specific (Levelt 1989; Slobin 1987). From lifelong experience, speakers usually know what concepts are lexically expressible in their language. Our theory of lexical access is not well developed for this initial stage of conceptual preparation (but see sect. 4.2). In particular, the computational model does not cover this stage. However, in order to handle the subsequent stage of lexical selection, particular assumptions have to be made about the output of conceptual preparation. Why have we opted for lexical concepts as the terminal vocabulary of conceptual preparation?

It is a classical and controversial issue whether the terminal conceptual vocabulary is a set of lexical concepts or, rather, the set of primitive conceptual features that make up these lexical concepts. We assume that message elements make explicit the intended lexical concepts (see Fodor et al. 1980) rather than the primitive conceptual features that make up these concepts, as is traditionally assumed (see, e.g., Bierwisch & Schreuder 1992; Goldman 1975; Miller & Johnson-Laird 1976; Morton 1969). That is, we assume that there is an independent message element that says, for example, ESCORT(X, Y) instead of several elements that say something such as IS-TO-ACCOMPANY(X, Y) and IS-TO-SAFEGUARD(X, Y) and so forth. The representation ESCORT(X, Y) gives access to conceptual features in memory such as IS-TO-ACCOMPANY(X, Y) but does not contain them as proper parts (Roelofs 1997a). Van Gelder (1990) referred to such representations as "functionally decomposed." Such memory codes, that is, codes standing for more complex entities in memory, are traditionally called "chunks" (Miller 1956).

There are good theoretical and empirical arguments for this assumption of chunked retrieval in our theory, which have been reviewed extensively elsewhere (Roelofs 1992b; 1993; 1996a; and, especially, 1997a). In general, how information is represented greatly influences how easy it is to use (see Marr 1982). Any representation makes some information explicit at the expense of information that is left in the background. Chunked retrieval implies a message that indicates which lexical concepts have to be expressed, while leaving their featural composition in memory. Such a message provides the information needed for syntactic encoding and reduces the computational burden for both the message encoding process and the process of lexical access. Mapping thoughts onto chunked lexical concept representations in message encoding guarantees that the message is ultimately expressible in the target language, and mapping these representations onto lemmas prevents the hyperonym problem from arising (see Roelofs 1996a; 1997a).

4.2. Perspective taking

Any state of affairs can be expressed in many different ways. Take the scene represented at the top of Figure 3. Two possible descriptions, among many more, are: *I see a chair with a ball to the left of it* and *I see a chair with a ball to the right of it.* Hence one can use the converse terms *left* and *right* here to refer to the same spatial relation. Why? It all depends on the perspective taken. The expression *left of* arises when the speaker resorts to "deictic" perspective in mapping the spatial scene onto a conceptual representation, deictic perspective being a three-term relation between the speaker as origin, the relatum (chair), and the referent (ball). However, *right of* results when the speaker interprets the scene from an "intrinsic perspective," a two-term relation in which the relatum (chair) is the origin and the referent (ball) relates to the intrinsic right side of the referent. Depending on the perspective taken, the lexical concept LEFT or RIGHT is activated (see Fig. 3). Both lead to veridical descriptions. Hence, there is no hard-wired relation between the state of affairs and the appropriate lexical concept. Rather, the choice of perspective is free. Various aspects of the scene and the communicative situation make the speaker opt for one perspective or the other (see Levelt, 1989, 1996, for reviews and experimental data).

Perspective taking is not just a peculiar aspect of spatial description; rather, it is a general property of all referring. It is even

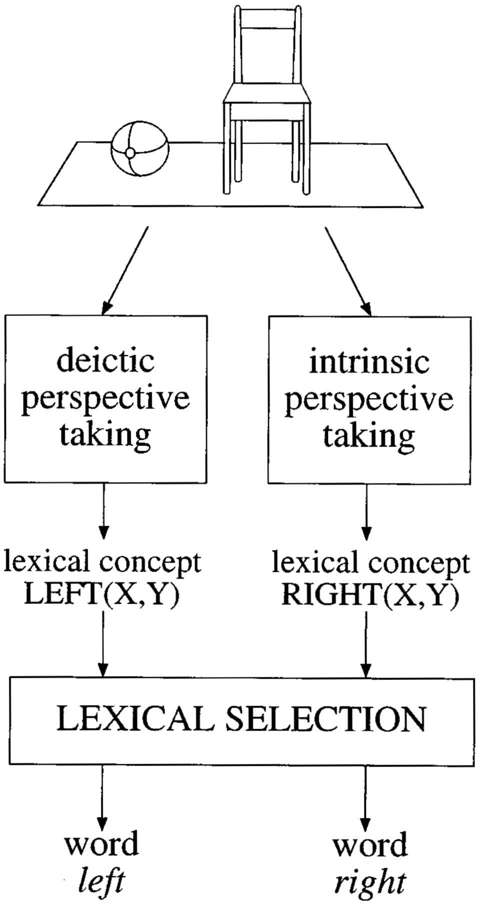


Figure 3. Illustration of perspective taking. Is the ball to the left of the chair or to the right of the chair?

an essential component in tasks as simple as picture naming. Should the object be referred to as *an animal, a horse,* or *a mare*? All can be veridical, but it depends on context which perspective is the most appropriate. It is a convenient illusion in the picture naming literature that an object has a fixed name, but there is no such thing. Usually, there is only the tacit agreement to use basic level terms (Rosch et al. 1976). Whatever the intricacies of conceptual preparation, the relevant output driving the subsequent steps in lexical access is the active lexical concept.

5. Lexical selection

5.1. Algorithm for lemma retrieval

The activation of a lexical concept is the proximal cause of lexical selection. How is a content word, or rather lemma (see sect. 3.1.2), selected from the mental lexicon, given an active lexical concept? A basic claim of our theory is that lemmas are retrieved in a conceptually nondecomposed way. For example, the verb *escort* is retrieved on the basis of the abstract representation or chunk ESCORT(X, Y) instead of features such as IS-TO-ACCOMPANY(X, Y) and IS-TO-SAFEGUARD(X, Y). Retrieval starts by enhancing the level of activation of the node of the target lexical concept. Activation then spreads through the network, each node sending a proportion of its activation to its direct neighbors. The most highly activated lemma node is selected when verification allows. For example, in verbalizing "escort," the activation level of the lexical concept node ESCORT(X, Y) is enhanced. Activation spreads through the conceptual network and down to the lemma stratum. As a consequence, the lemma nodes *escort* and *accompany* will be activated. The *escort* node will be the most highly activated node, because it receives a full proportion of ESCORT(X, Y)'s activation, whereas *accompany* and other lemma nodes receive only a proportion of a proportion. Upon verification of the link between the lemma node of *escort* and ESCORT(X, Y), this lemma node will be selected. The selection of function words also involves lemma selection; each function word has its own lemma, that is, its own syntactic specification. Various routes of lemma activation are open here. Many function words are selected in just the way described for selecting *escort,* because they can be used to express semantic content. That is often the case for the use of prepositions, such as *up* or *in.* However, the same prepositions can also function as parts of particle verbs (as in *look up,* or *believe in*). Here they have no obvious semantic content. In section 5.3 we will discuss how such particles are accessed in the theory. The lemmas of still other function words are activated as part of a syntactic procedure, for instance, *that* in the earlier example "John said *that*. . . ." Here we will not discuss this "indirect election" of lemmas (but see Levelt, 1989).

The equations that formalize WEAVER++ are given by Roelofs (1992a; 1992b; 1993; 1994; 1996b; 1997c), and the Appendix gives an overview of the selection algorithm. There are simple equations for the activation dynamics and the instantaneous selection probability of a lemma node, that is, the hazard rate of the lemma retrieval process. The basic idea is that, for any smallest time interval, given that the selection conditions are satisfied, the selection probability of a lemma node equals the ratio of its activation to that of all the other lemma nodes (the "Luce ratio"). Given the selection ratio, the expectation of the retrieval time can be computed.

5.2. Empirical RT support

5.2.1. SOA curves of semantic effects. The retrieval algorithm explains, among other things, the classical curves of the semantic effect of picture and word distractors in picture naming, picture categorizing, and word categorizing. The basic experimental situation for picture naming is as follows. Participants have to name pictured objects while trying to ignore written distractor words superimposed on the pictures or spoken distractor words. For example, they have to say "chair" to a pictured chair and ignore the distractor word "bed" (semantically related to target word "chair") or "fish" (semantically unrelated). In the experiment, one can vary the delay between picture onset and distractor onset, the so-called stimulus onset asynchrony (SOA). The distractor onset can be, typically, at 400, 300, 200, or 100 msec before picture onset (negative SOAs); simultaneous with picture onset; or at 100, 200, 300, or 400 msec after picture onset (positive SOAs). The classical finding is shown in Figure 4A; this is the SOA curve obtained by Glaser and Düngelhoff (1984), when the distractors were visually presented words. It shows a semantic effect (i.e., the difference between the naming latencies with semantically related and unrelated distractors) for different

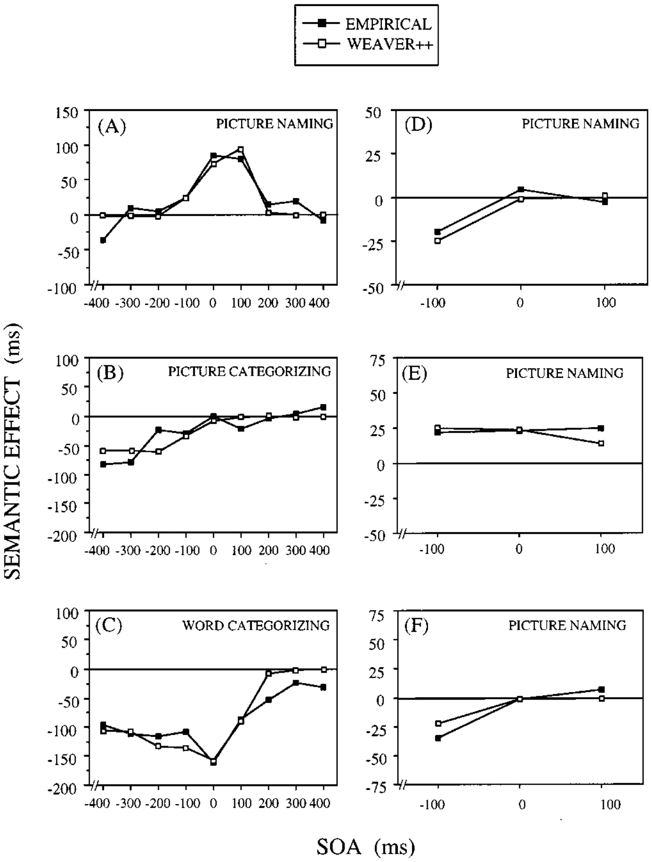


Figure 4. Effect of printed word distractors on picture naming and categorizing latencies. Degree of inhibition/facilitation as a function of stimulus onset asynchrony (SOA) between distractor and picture. A: Picture naming data; B: Picture categorizing data; C: Word categorization data from Glaser and Düngelhof (1984) (black dots) and WEAVER++ model fit (open dots). D: Picture naming with hyperonym, cohyponym, and hyponym distractors [black dots are data (means across these three distractor types) from Roelofs (1992a); open dots show WEAVER++ model fit]. E: Picture naming by verbs (e.g., "drink") with cohyponym verbs that are in the response set as distractor (e.g., "eat"). F: Picture naming by verbs (e.g., "eat") with hyponym verbs that are not in the response set (e.g., "booze") as distractor (black dots are data from Roelofs (1993); open dots are WEAVER++ model fits).

SOAs. Thus, a positive difference indicates a semantic inhibition effect. Semantic inhibition is obtained at SOA-100, 0, and 100 msec.

Before discussing these and the other data in Figure 4, we must present some necessary details about how WEAVER++ was fit to these data. The computer simulations of lemma retrieval in picture naming, picture categorizing, and word categorizing experiments were run with both small and larger lexical networks. The small network (see Fig. 5) included the nodes that were minimally needed to simulate the conditions in the experiments. To examine whether the size of the network influenced the outcomes, the simulations were run using larger networks of either 25 or 50 words that contained the small network as a proper part. The small and larger networks produced equivalent outcomes.

All simulations were run using a single set of seven parameters whose values were held constant across simulations: (1) a real-time value in milliseconds for the smallest time interval (time step) in the model, (2) values for the general spreading rate at the conceptual stratum, and (3) values for the general spreading rate at the lemma stratum, (4) decay rate, (5) strength of the distractor input to the network, (6) time interval during which this input was provided, and (7) a selection threshold. The parameter values were obtained by optimizing the goodness of fit between the model and a restricted number of data sets from the literature; other known data sets were subsequently used to test the model with these parameter values.

The data sets used to obtain the parameter values concerned the classical SOA curves of the inhibition and facilitation effects of distractors in picture naming, picture categorizing and word categorizing; they are all from Glaser and Düngelhoff (1984). Figure 4A–C presents these data sets (in total 27 data points) and the fit of the model. In estimating the seven parameters from these 27 data points, parameters 1–5 were constrained to be constant across tasks, whereas parameters 6 and 7 were allowed to differ between tasks to account for task changes (i.e., picture naming, picture categorizing, word categorizing). Thus, WEAVER++ has significantly fewer degrees of freedom than the data contain. A goodness of fit statistic adjusted for the number of estimated parameter values showed that the model fit the data. (The adjustment "punished" the model for the estimated parameters.)

After fitting of the model to the data of Glaser and Düngelhoff, the model was tested on other data sets in the literature and in new experiments specifically designed to test nontrivial predictions of the model. The parameter values of the model in these tests were identical to those in the fit of Glaser and Düngelhoff's data. Figure 4D–F presents some of these new data sets together with the predictions of the model. Note that WEAVER++ is not too powerful to be falsified by the data. In the graphs presented in Figure 4, there are 36 data points in total, 27 of which were simultaneously fit by WEAVER++ with only seven parameters; for the remainder, no further fitting was done, except that parameter 7 was fine-tuned between experiments. Hence there are substantially more empirical data points than there are parameters in the model. The fit of the model to the data is not trivial.

We will now discuss the findings in each of the panels of Figure 4 and indicate how WEAVER++ accounts for the data. As in any modeling enterprise, a distinction can be made between empirical phenomena that were specifically built into the model and phenomena that the model predicts but that had not been previously explored. For example, the effects of distractors are inhibitory in picture naming (Fig. 4A) but they are facilitatory in picture and word categorizing (Fig. 4B, C). This phenomenon was built into the model by restricting the response competition to permitted response words, which yields inhibition in naming but facilitation in categorizing, as we will explain below. Adopting this restriction led to predictions that had not been tested before. These predictions were tested in new experiments; the results for some are shown in Figure 4D–F. How does WEAVER++ explain the picture naming findings in Figure 4A? We will illustrate the explanation using the miniature network depicted in Figure 5 (larger networks yield the same outcomes), which illustrates the con-

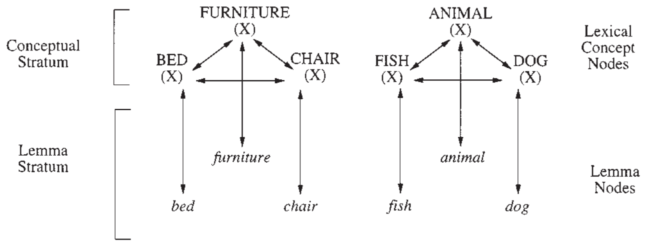


Figure 5. Miniature network illustrating how WEAVER++ accounts for the data in Figure 4.

ceptual stratum and the lemma stratum of two semantic fields, furniture and animals. Thus, there are lexical concept nodes and lemma nodes. It is assumed here that, in this task, presenting the picture activates the corresponding basic level concept (but see sect. 4.2). Following the assumptions outlined in section 3.2.4, we suppose that distractor words have direct access to the lemma stratum. Now assume that "chair" is the target. All distractors are names of other pictures in the experiment. In the case of a pictured chair and a distractor "bed," activation from the picture and the distractor word will converge on the lemma of the distractor "bed," owing to the connections at the conceptual stratum. In case of the unrelated distractor "fish," there will be no such convergence. Although the distractor "bed" will also activate the target lemma *chair* [via the concept nodes BED(X) and CHAIR(X)], the pictured chair will prime the distractor lemma *bed* more than the distractor word "bed" will prime the target lemma *chair.* This is due to network distances: three links versus four links [pictured chair → CHAIR(X) → BED(X) → *bed* vs. word "bed" → *bed* → BED(X) → CHAIR(X) → *chair*]. Consequently, it will take longer before the activation of *chair* exceeds that of *bed* than that of *fish.* Therefore, *bed* will be a stronger competitor than *fish,* which results in the semantic inhibition effect.

Let us now consider the results in Figure 4B. It is postulated in WEAVER++ that written distractors are only competitors when they are permitted responses in an experiment (i.e., when they are part of the response set). In the case of picture or word categorization, *furniture* and *animal* instead of *chair, bed,* or *fish* are the targets. Now the model predicts a semantic facilitation effect. For example, the distractor "bed" will prime the target *furniture,* but will not be a competitor itself because it is not a permitted response in the experiment. By contrast, "fish" on a pictured chair will prime *animal,* which is a competitor of the target *furniture.* Thus, semantic facilitation is predicted, and this is also what is empirically obtained. Figure 4B gives the results for picture categorizing (for example, when participants have to say "furniture" to the pictured bed and ignore the distractor word). Again, the semantic effect is plotted against SOA. A negative difference indicates a semantic facilitation effect. The data are again from Glaser and Düngelhoff (1984). WEAVER++ fits the data well.

By the same reasoning, the same prediction holds for *word* categorizing, for example, when participants have to say "furniture" when they see the printed word "bed" but have to ignore the picture behind it. Figure 4C gives the results for word categorizing. Again, WEAVER++ fits the data.

Still another variant is picture naming with hyperonym, cohyponym, and hyponym distractors superimposed. As long as these distractors are not part of the response set, they should facilitate naming relative to unrelated distractors. For example, in naming a pictured chair (the only picture of a piece of furniture in the experiment), the distractor words "furniture" (hyperonym), "bed" (cohyponym), or "throne" (hyponym) are superimposed. Semantic facilitation was indeed obtained in such an experiment (Roelofs 1992a; 1992b). Figure 4D plots the semantic facilitation against SOA. The semantic effect was the same for hyperonym, cohyponym, and hyponym distractors. The curves represent means across these types of word. The findings concerning the facilitation effect of hyponym distractors exclude one particular solution to the hyp(er)onymy problem in lemma retrieval. Bierwisch and Schreuder (1992) have proposed that the convergence problem is solved by inhibitory links between hyponyms and hyperonyms in a logogen-type system. However, this predicts semantic inhibition from hyponym distractors, but facilitation is what is obtained.

The WEAVER++ model is not restricted to the retrieval of noun lemmas. Thus, the same effects should be obtained in naming actions using verbs. For example, ask participants to say "drink" to the picture of a drinking person (notice the experimental induction of perspective taking) and to ignore the distractor words "eat" or "laugh" (names of other actions in the experiment). Indeed, semantic inhibition again is obtained in that experiment, as shown in Figure 4E (Roelofs 1993). Also, facilitation is again predicted for hyponym distractors that are not permitted responses in the experiment. For instance, the participants have to say "drink" to a drinking person and ignore "booze" or "whimper" (not permitted responses in the experiment) as distractors. Semantic facilitation is indeed obtained in this paradigm, as shown in Figure 4F (Roelofs 1993).

In summary, the predicted semantic effects have been obtained for nouns, verbs, and adjectives (e.g., color, related to the classical Stroop effect), not only in producing single words (see, e.g., Glaser & Glaser 1989; Roelofs 1992a; 1992b; 1993) but also for lexical access in producing phrases, as has been shown by Schriefers (1993). To study semantic (and phonological) priming in sentence production, Meyer (1996) used auditory primes and found semantic inhibition, although the distractors were not in the response set. In an as yet unpublished study, Roelofs obtained semantic facilitation from written distractor words but semantic inhibition when the same distractor words were presented auditorily. Why it is, time and again, hard to obtain semantic facilitation from auditory distractors is still unexplained.

5.2.2. Semantic versus conceptual interference. One could ask whether the semantic effects reported in the previous section could be explained by access to the conceptual stratum. In other words, are they properties of lexical access proper? They are; the semantic effects are obtained *only* when the task involves producing a verbal response. In a control experiment carried out by Schriefers et al. (1990), participants had to categorize pictures as "old" or "new" by pressing one of two buttons; that is, they were not naming the pictures. In a preview phase of the experiment, the participants had seen half of the pictures. Spoken distractor words were presented during the old/new categorization task. In contrast to the corresponding naming task, no semantic inhibition effect was obtained. This suggests that the semantic interference effect is due to lexical access rather than to accessing conceptual memory. Of course, these findings do not exclude interference effects at the conceptual level. Schriefers (1990) asked participants to refer to pairs of objects by saying whether an object marked by a cross was bigger or smaller than the other; that is, the subject produced the verbal response "bigger" or "smaller." However, there was an additional variable in the experiment: Both objects could be relatively large, or both could be relatively small. Hence not only relative size but also absolute size was varied. In this relation naming task, a congruency effect was obtained. Participants were faster in saying "smaller" when the absolute size of the objects was small than when it was big, and vice versa. In contrast to the semantic effect of distractors in picture naming, this congruency effect was a concept-level effect. The congruency effect remained when the participants had to press one button when the marked object was taller and another button when it was shorter.

5.2.3. Interaction between semantic and orthographic factors. Starreveld and La Heij (1995; see also Starreveld & La Heij 1996a) observed that the semantic inhibition effect in picture naming is reduced when there is an orthographic relationship between target and distractor. For example, in naming a picture of a cat, the semantic inhibition was less for distractor "calf" compared to "cap" (orthographically related to "cat") than for distractor "horse" compared to "house." According to Starreveld and La Heij, this interaction suggests that there is feedback from the word form level to the lemma level, that is, from word forms <calf> and <cap> to lemma *cat,* contrary to our claim that the word form network contains forward links only. However, as we have argued elsewhere (Roelofs et al. 1996; see also sect. 3.2.4), Starreveld and La Heij overlooked the fact that printed words activate their lemma nodes and word form nodes in parallel in our theory (see sect. 3.2.4). Thus, printed words may affect lemma retrieval directly, and there is no need for backward links from word form nodes to lemmas in the network. Computer simulations showed that WEAVER++ predicts that, in naming a pictured cat, the semantic inhibition will be less for distractor "calf" compared to "cap" than for distractor "horse"compared to "house," as is empirically observed.

5.3. Accessing morphologically complex words

There are different routes for a speaker to generate morphologically complex words, depending on the nature of the word. We distinguish four cases, depicted in Figure 6.

5.3.1. The degenerate case. Some words may linguistically count as morphologically complex, but are not complex psychologically. An example is *replicate,* which historically has a morpheme boundary between *re* and *plicate.* That this is not any more the case appears from the word's syllabification, *rep-li-cate* (which even violates maximization of onset). Normally, the head morpheme of a prefixed word will behave as a phonological word (ω) itself, so syllabification will respect its integrity. This is not the case for *replicate,* where *p* syllabifies with the prefix (note that it still is the case in *re-ply,* which has the same latinate origin, *re-plicare).* Such words are monomorphemic for all processing means and purposes (Fig. 6A).

5.3.2. The single-lemma-multiple-morpheme case. This is the case depicted in Figure 6B and in Figure 2. The word *escorting* is generated from a single lemma *escort* that is marked for +progressive. It is only at the word form level that two nodes are involved, one for <escort> and the other one for <ing>. Regular inflections are probably all of this type, but irregular verb inflections are not, usually. The lemma *go*+past will activate the one morpheme <went>. Although inflections for number will usually go with the regular verb inflections, there are probably exceptions here (see sect. 5.3.5). The case is more complicated for complex derivational morphology. Most of the frequently used compounds are of the type discussed here. For example, *blackboard, sunshine, hotdog,* and *offset* are most likely single lemma items, though *thirty-nine* and complex numbers in general (see Miller 1991) might not be. Words with bound derivational morphemes form a special case. These morphemes typically change the word's syntactic category. However, syntactic category is a lemma level property. The simplest story, therefore, is to consider them to be single-lemma cases, carrying the appropriate syntactic category. This will not work though for more productive derivation, to which we will shortly return.

5.3.3. The single-concept-multiple-lemma case. The situation shown in Figure 6C is best exemplified by the case of particle verbs. A verb such as "look up," is represented by two lemma nodes in our theory and computational model (Roelofs 1998). Particle verbs are not words but minimal verb projections (Booij 1995). Given that the semantic interpretation of particle verbs is often not simply a combination of the meanings of the particle and the base (hence they do not stem from multiple concepts), the verb–particle combinations have to be listed in the mental lexicon. In producing a verb–particle construction, the lexical concept

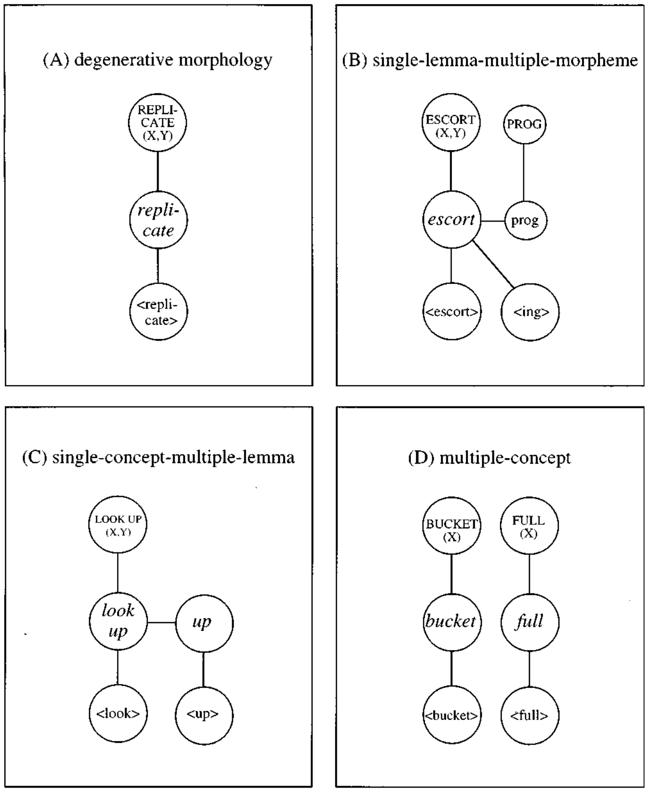


Figure 6. Four varieties of complex morphology in the theory.

selects for a *pair* of lemma nodes from memory and makes them available for syntactic encoding processes. Some experimental evidence on the encoding of particle verbs will be presented in section 6.4.4.

A very substantial category of this type is formed by idioms. The production of "kick the bucket" probably derives from activating a single, whole lexical concept, which in turn selects for multiple lemmas (see Everaerd et al. 1995).

5.3.4. The multiple-concept case. This case, represented in Figure 6D, includes all derivational new formations. Clearest here are newly formed compounds, the most obvious case being complex numbers. At the conceptual level the number 1,007 is probably a complex conceptualization, with the lexical concepts 1,000 and 7 as terminal elements. These, in turn, select for the lemmas *thousand* and *seven*, respectively. The same process is probably involved in generating other new compounds, for example, when a creative speaker produced the word *sitcom* for the first time. There are still other derivational new formations, those with bound morphology, that seem to fit this category. Take very-low-frequency *X-ful* words, such as bucketful. Here, the speaker may never have heard or used the word before and hence does not yet have a lemma for it. There are probably two active lexical concepts involved here, BUCKET and something like FULL, each selecting for its own lemma. Semantics is clearly compositional in such cases. Productive derivational uses of this type require the bound morpheme at the lemma level to determine the word's syntactic category during the generation process.

Do these four cases exhaust all possibilities in the generation of complex morphology? It does not seem so, as will appear in the following section.

5.3.5. Singular- and plural-dominant nouns. In an as yet unpublished study, Baayen, Levelt, and Haveman asked subjects to name pictures containing one or two identical objects, and to use singular or plural, respectively. The depicted objects were of two kinds. The first type, so-called *singular dominants,* were objects whose name was substantially more frequent in the singular than in the plural form. An example is "nose," for which *nose* is more frequent than *noses.* For the second type, the so-called *plural dominants,* the situation was reversed, the plural being more frequent than the singular. An example is "eye," with *eyes* more frequent than *eye.* The upper panel in Figure 7 presents the naming latencies for relatively high-frequency singular and plural dominant words.

These results display two properties, one of them remarkable. The first is a small but significant longer latency for plurals than for singulars. That was expected, because of greater morphological complexity. The remarkable finding is that both the plural dominant singulars (such as *eye)* and the plural dominant plurals (such as *eyes)* were significantly slower than their singular dominant colleagues, although the stem frequency was controlled to be the same for the plural and the singular dominants. Also, there was no interaction. This indicates, first, that there was no surface frequency effect: The relatively high-frequency plural dominant plurals had the longest naming latencies. Because the surface frequency effect originates at the word form level, as will be discussed in section 6.1.3, a word's singular and plural are likely to access the same morpheme node at the word form level. More enigmatic is why plural dominants are so slow. A possible explanation is depicted in Figure 7B and 7C. The "normal" case is singular dominants. In generating the plural of "nose," the speaker first activates the lexical concepts NOSE and something like MULTIPLE. Together, they select for the one lemma *nose,* with diacritic feature *"pl."* The lemma with its plural feature then activates the two morpheme nodes <nose> and <-əz>, following the single-lemma – multiple-morpheme case of section 5.3.2. However, the case may be quite different for plural dominants, such as "eye." Here there are probably two different lexical concepts involved in the singular and the plural. The word "eyes" is not just the plural of "eye," there is also some kind of meaning difference: "eyes" has the stronger connotation of "gaze." And similar shades of meaning variation exist between "ears" and "ear," "parents" and "parent," etc. This is depicted in Figure 7C. Accessing the plural word "eyes" begins by accessing the specific lexical concept EYES. This selects for its own lemma, *eyes* (with a diacritic plural feature). This in turn activates morphemes <eye> and <z> at the word form level. Singular "eye" is similarly generated from the specific lexical concept EYE. It selects for its own (singular) lemma *eye.* From here, activation converges on the morpheme <eye> at the word form level.

How do the diagrams shown in Figure 7B and 7C account for the experimental findings? For both the singular and the plural dominants, the singular and plural converge on the same morpheme at the word form level. This explains the lack of a surface frequency effect. That the plural dominants are relatively slow, for both the singular and the plural, follows from the main lemma selection rule, discussed in section 5.1. The semantically highly re-

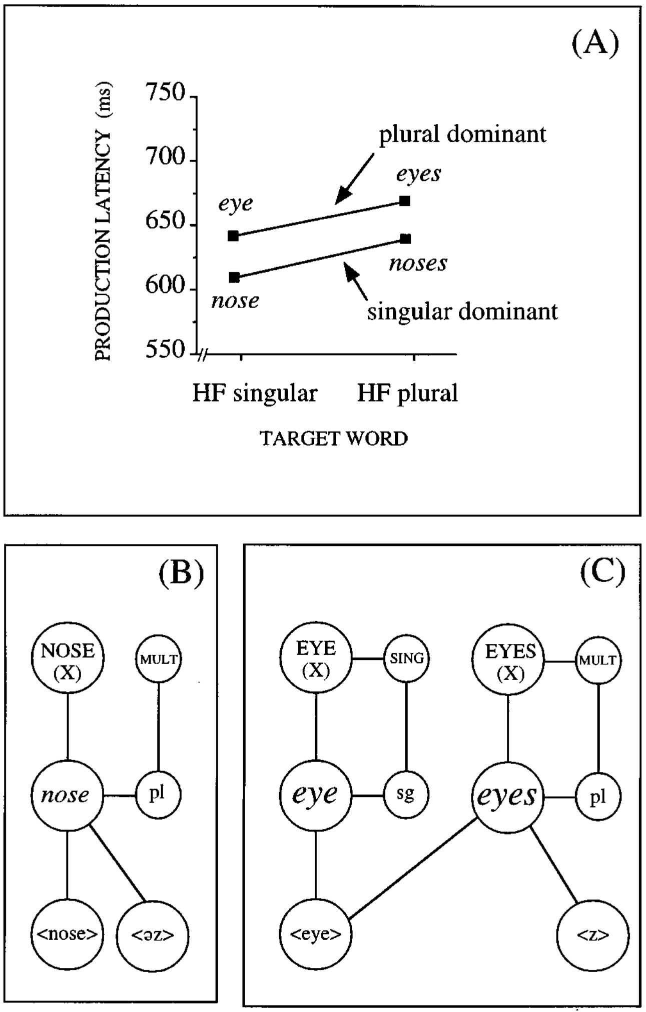


Figure 7. Naming latencies for pictures depicting one object or two identical objects (A). Plural names are slower than singular names, and both singular and plural names are slower for plural dominants (such as *eye)* than for singular dominants (such as *nose).* Possible representations of plural morphology for singular-dominant nouns (B) and for plural dominant nouns (C).

lated lexical concepts EYE and EYES will always be coactivated, whichever is the target. As a consequence, both lemmas *eye* and *eyes* will receive activation, whichever is the target. The lexical selection rule then predicts relatively long selection latencies for both the singular and the plural lemmas (following Luce's rule), because of competition between active lemmas. This is not the case for selecting *nose;* there is no competitor there.

In conclusion, the generation of complex morphology might involve various levels of processing, depending on the case at hand. It will always be an empirical issue to determine what route is followed by the speaker in any concrete instance.

5.4. Accessing lexical syntax and the indispensability of the lemma level

A core feature of the theory is that lexical selection is conceived of as selecting the syntactic word. What the speaker selects from the mental lexicon is an item that is just sufficiently specified to function in the developing syntax. To generate fluent speech incrementally, the first bit of lexical information needed is the word's syntax. Accessing word form information is less urgent in the process (see Levelt 1989), but what evidence do we have that lemma and word form access are really distinct operations?

5.4.1. Tip-of-the-tongue states. Recent evidence supporting the distinction between a lemma and form level of access comes from the tip-of-the-tongue phenomenon. As was mentioned above (sect. 3.1.3), Italian speakers in tip-of-the-tongue states most of the time know the grammatical gender of the word, a crucial syntactic property in the generation of utterances (Vigliocco et al. 1997). However, they know the form of the word only partially or not at all. The same has been shown for an Italian anomic patient (Badecker et al. 1995), confirming earlier evidence from French anomic patients (Henaff Gonon et al. 1989). This shows that lemma access can succeed where form access fails.

5.4.2. Agreement in producing phrases. A further argument for the existence of a distinct syntax accessing operation proceeds from gender priming studies. Schriefers (1993) asked Dutch participants to describe colored pictured objects using phrases. For example, they had to say *de groene tafel* ("the green table") or *groene tafel* ("green table"). In Dutch, the grammatical gender of the noun (non-neuter for *tafel,* "table") determines which definite article should be chosen (*de* for non-neuter and *het* for neuter) and also the inflection on the adjective (*groene* or *groen,* "green"). On the pictured objects, written distractor words were superimposed that were either gender congruent or gender incongruent with the target. For example, the distractor *muis* ("mouse") takes the same non-neutral gender as the target *tafel* ("table"), whereas distractor *hemd* ("shirt") takes neuter gender. Schriefers obtained a gender congruency effect, as predicted by WEAVER++. Smaller production latencies were obtained when the distractor noun had the same gender as the target noun compared to a distractor with a different gender (see also Van Berkum 1996; 1997). According to WEAVER++, this gender congruency effect should only be obtained when agreement has to be computed, that is, when the gender node has to be selected in order to choose the appropriate definite article or the gender marking on the adjective, but not when participants have to produce bare nouns, that is, in "pure" object naming. WEAVER++ makes a distinction between activation of the lexical network and the actual selection of nodes. All noun lemma nodes point to one of the grammatical gender nodes (two in Dutch), but there are no backward pointers. Thus, boosting the level of activation of the gender node by a gender-congruent distractor will not affect the level of activation of the target lemma node and therefore will not influence the selection of the lemma node. Consequently, priming a gender node will affect only lexical access when the gender node itself has to be selected. This is the case when the gender node is needed for computing agreement between adjective and noun. Thus, the gender congruency effect should be seen only in producing gender-marked utterances, not in producing bare nouns. This corresponds to what is empirically observed (Jescheniak 1994).

5.4.3. A short-lived frequency effect in accessing gender. A further argument for an independent lemma representation derives from experiments by Jescheniak and Levelt (1994; Jescheniak 1994). They demonstrated that, when lemma information such as grammatical gender is accessed, an idiosyncratic frequency effect is obtained. Dutch participants had to decide on the gender of a picture's name (e.g., they had to decide that the grammatical gender of *tafel,* "table" is non-neuter), which was done faster for high-frequency words than for low-frequency ones. The effect quickly disappeared over repetitions, contrary to a "robust" frequency effect obtained in naming the pictures (to be discussed in sect. 6.1.3). In spite of substantial experimental effort (van Berkum 1996; 1997), the source of this short-lived frequency effect has not been discovered. What matters here, however, is that gender and form properties of the word bear markedly different relations to word frequency.

5.4.4. Lateralized readiness potentials. Exciting new evidence for the lemma–word form distinction in lexical access stems from a series of experiments by van Turennout et al. (1997; 1998). The authors measured event related potentials in a situation in which the participants named pictures. On the critical trials, a gender/segment classification task was to be performed before naming, which made it possible to measure lateralized readiness potentials (LRPs; see Coles et al. 1988; Coles 1989). This classification task consisted of a conjunction of a push-button response with the left or right hand and a go–no go decision. In one condition, the decision whether to give a left- or right-hand response was determined by the grammatical gender of the picture name (e.g., respond with the left hand if the gender is non-neuter or with the right hand if it is neuter). The decision on whether to carry out the response was determined by the first segment of the picture name (e.g., respond if the first segment is /b/; otherwise do not respond). Hence, if the picture was one of a bear (Dutch "beer," with non-neutral gender) the participants responded with their left hand; if the picture was one of a wheel (Dutch "wiel," with neutral gender) they did not respond. The measured LRPs show whether the participants prepared for pushing the correct button not only on the go trials but also on the no-go trials. For example, the LRPs show whether there is response preparation for a picture whose name does not start with the critical phoneme. When gender determined the response hand and the segment determined whether to respond, the LRP showed preparation for the response hand on both the go and the no-go trials. However, under a condition in which the situation was reversed, that is, when the first segment determined the response hand and the gender determined whether to respond, the LRP showed preparation for the response hand on the go trials but not on the no-go trials.

These findings show that, in accessing lexical properties in production, you can access a lemma property, gender, and halt there before beginning to prepare a response to a word form property of the word, but the reverse is not possible. In this task you will have accessed gender before you access a form property of the word. Again, these findings support the notion that a word's lexical syntax and its phonology are distinct representations that can be accessed in this temporal order only. In other experiments, the authors showed that onsets of LRP preparation effects in monitoring word onset and word offset consonants (e.g., /b/ vs. /r/ in target *bear)* differed by 80 msec on average. This gives an indication of the speed of phonological encoding, to which we will return in section 9.

5.4.5. Evidence from speech errors. The findings discussed so far in this section support the notion that accessing lexical syntax is a distinct operation in word access. A lemma level of word encoding explains semantic interference effects in the picture–word interference paradigm, findings on tip-of-the-tongue states, gender congruency effects in computing agreement, specific frequency effects in accessing gender information, and event related potentials in accessing lexical properties of picture names.

Although our theory has (mostly) been built upon such latency data, this section would not be complete without referring to the classical empirical support for a distinction between lemma retrieval and word form encoding coming from speech errors. A lemma level of encoding explains the different distribution of word and segment exchanges. Word exchanges, such as the exchange of *roof* and *list* in *we completely forgot to add the list to the roof* (from Garrett 1980), typically concern elements from different phrases and of the same syntactic category (here, noun). By contrast, segment exchanges, such as *rack pat* for *pack rat* (from Garrett 1988), typically concern elements from the same phrase and do not respect syntactic category. This finding is readily explained by assuming lemma retrieval during syntactic encoding and segment retrieval during subsequent word form encoding.

Speech errors also provide support for a morphological level of form encoding that is distinct from a lemma level with morphosyntactic parameters. Some morphemic errors appear to concern the lemma level, whereas others involve the form level (see, e.g., Dell 1986; Garrett 1975; 1980; 1988). For example, in *how many pies does it take to make an apple?* (from Garrett 1988), the interacting stems belong to the same syntactic category (i.e., noun) and come from distinct phrases. Note that the plurality of *apple* is stranded, that is, it is realized on *pie.* Thus, the number parameter is set after the exchange. The distributional properties of these morpheme exchanges are similar to those of whole-word exchanges. This suggests that these morpheme errors and whole-word errors occur at the same level of processing, namely, when lemmas in a developing syntactic structure trade places. By contrast, the exchanging morphemes in an error such as *slicely thinned* (from Stemberger 1985) belong to different syntactic categories (adjective and verb) and come from the same phrase, which is also characteristic of segment exchanges. This suggests that this second type of morpheme error and segment errors occur at the same level of processing, namely, the level at which morphemes and segments are retrieved and the morphophonological form of the utterance is constructed. The errors occur when morphemes in a developing morphophonological structure trade places.

The sophisticated statistical analysis of lexical speech errors by Dell and colleagues (Dell 1986; 1988) has theoretically always involved a level of lemma access, distinct from a level of form access. Recently, Dell et al. (1997b) reported an extensive picture naming study on 23 aphasic patients and 60 matched normal controls, analyzing the spontaneous lexical errors produced in this task. For both normal individuals and patients, a perfect fit was obtained with a two-level spreading activation model, that is, one that distinguishes a level of lemma access. Although the model differs from WEAVER++ in other respects, there is no disagreement about the indispensability of a lemma stratum in the theory.

6. Morphological and phonological encoding

After having selected the appropriate lemma, the speaker is in the starting position to encode the word as a motor action. Here the functional perspective is quite different from the earlier move toward lexical selection. In lexical selection, the job is to select the one appropriate word from among tens of thousands of lexical alternatives, but in preparing an articulatory action, lexical alternatives are irrelevant; there is only one pertinent word form to be encoded. What counts is context. The task is to realize the word in its prosodic environment. The dual function here is for the prosody to be expressive of the constituency in which the word partakes and to optimize pronounceability. One aspect of expressing constituency is marking the word as a lexical head in its phrase. This is done through phonological phrase construction, which will not be discussed here (but see Levelt 1989). An aspect of optimizing pronounceability is syllabification in context. This is, in particular, achieved through phonological word formation, which we introduced in section 3.1.3. Phonological word formation is a central part of the present theory, to which we will shortly return. However, the first move in morphophonological encoding is to access the word's phonological specification in the mental lexicon.

6.1. Accessing word forms

6.1.1. The accessing mechanism. Given the function of word form encoding, it would appear counterproductive to activate the word forms of all active lemmas that are *not* selected.[8] After all, their activation can only interfere with the morphophonological encoding of the target, or, alternatively, there should be special, built-in mechanisms to prevent this – a curiously baroque design. In Levelt et al. (1991a), we therefore proposed the following principle: *Only selected lemmas will become phonologically activated.*

Whatever the face value of this principle, it is obviously an empirical issue. Levelt et al. (1991a) put this to a test in a picture naming experiment. Subjects were asked to name a series of pictures. On about one-third of the trials, an auditory probe was presented 73 msec after picture onset. The probe could be a spoken word or a nonword, and the subject had to make a lexical decision on the probe stimulus by pushing one of two buttons; the reaction time was measured. In the critical trials, the probe was a word and it could be an *identical,* a *semantic,* a *phonological,* or an *unrelated* probe. For example, if the picture was one of a sheep, the identical probe was the word *sheep* and the semantic probe was *goat.* The critical probe was the phonological one. In a preceding experiment, we had shown that, under the same experimental conditions, a phonological probe related to the target, such as *sheet* in the example, showed a strong latency effect in lexical decision, testifying to the phonological activation of the target word, the picture name *sheep.* In this experiment, however, we wanted to test whether a semantic alternative, such as *goat,* showed any phonological activation, so we now used a phonological probe related to that semantic alternative. In the example, that would be the word *goal,* which is phonologically related to *goat.* The *unrelated* probe, finally, had no semantic or phonological relation to the target or its semantic alternatives. Figure 8 shows the main findings of this experiment.

Both the identical and the semantic probes are significantly slower in lexical decision than the unrelated probes, but the phonological distractor, related to the (active) semantic alternative, shows not the slightest effect. This is in full agreement with the above-described activation principle. A nonselected semantic alternative remains phonologically inert. This case exemplifies the Ockham's razor approach discussed in section 3.2.5. The theory forbids something to happen, and that is put to the test. A positive outcome of this experiment would have falsified the theory.

There have been two kinds of reaction to the principle and to our empirical evidence in its support. The first was computational, the second experimental. The computational reaction, from Harley (1993), addressed the issue of whether this null result could be compatible with a connectionist architecture in which activation cascades, independently of lexical selection. We had, on various grounds, argued against such an architecture. The only serious argument in favor of interactive activation models had been their ability to account for a range of speech error phenomena, in particular the alleged statistical overrepresentation of so-called mixed errors, that is, errors that are both semantically and phonologically related to the target (e.g., a speaker happens to say *rat* instead of *cat).* In fact, Dell's (1986) original model was, in part, designed to explain precisely this fact in a simple and elegant way. Hence we concluded our paper with the remark that, maybe, it is possible to choose some connectionist model's parameters in such a way that it can both be reconciled with our negative findings and still account for the crucial speech error evidence. Harley (1993) took up that challenge and showed that his connectionist model (which differs rather substantially from Dell's, particularly in that it has inhibitory connections both within and between levels) can be parameterized in such a way that it produces our null effect and still accounts, in principle, for the crucial mixed errors. That is an existence proof, and we accept it, but it does not convince us that this is the way to proceed theoretically. The model precisely has the baroque properties mentioned above. It first activates the word forms of all semantic alternatives and then actively suppresses this activation by mutual inhibition. Again, the only serious reason to adopt such a design is the explanation of speech error statistics, and we will return to that argument below.

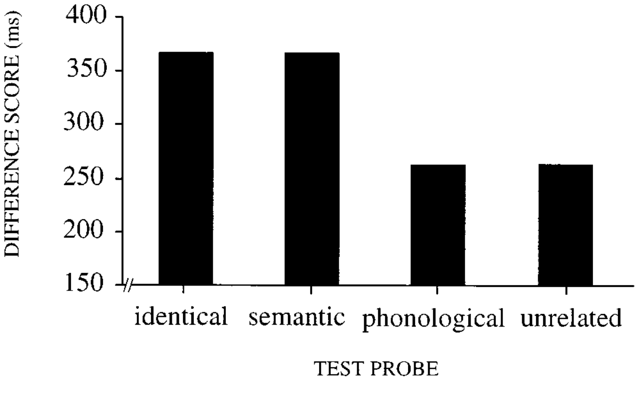


Figure 8. (Corrected) lexical decision latencies for auditory probes presented during picture naming. The y axis shows the lexical decision latency for a probe during picture naming minus the decision latency for the same auditory probe without concurrent picture naming. Probes could be identical to the target name (e.g., "sheep" for target *sheep),* semantically related to it ("goat"), phonologically related to the semantic alternative ("goal"), or wholly unrelated to target or semantic alternative ("house"). Data show that a semantically active semantic alternative is phonologically inert. Data from Levelt et al. (1991a).

The experimental reaction has been a head-on attack on the principle, i.e., to show that active semantic alternatives *are* phonologically activated. In a remarkable paper, Peterson and Savoy (1998) demonstrated this to be the case for a particular class of semantic alternatives, namely, (near-) synonyms. Peterson and Savoy's method was similar to ours from 1991, but they replaced lexical decision by word naming. Subjects were asked to name a series of pictures, but in half the cases they had to perform a secondary task. In these cases, a printed word appeared in the picture shortly after picture onset (at different SOAs), and the secondary task was to name that printed word. That distractor word could be semantically or phonologically related to the target picture name or phonologically related to a semantic alternative. There were controls as well, distractors that were neither semantically nor phonologically related to target or alternative. In a first set of experiments, Peterson and Savoy used synonyms as semantic alternatives. For instance, the subject would see a picture of a couch. Most subjects call this a *couch,* but a minority calls it a *sofa.* Hence, there is a dominant and a subordinate term for the same object. That was true for all 20 critical pictures in the experiment. On average, the dominant term was used 84% of the time. Would the subordinate term (*sofa* in the example) become phonologically active at all, maybe as active as the dominant term? To test this, Peterson and Savoy used distractors that were phonologically related to the subordinate term (e.g., *soda* for *sofa)* and compared their behavior to distractors related to the target (e.g., *count* for *couch).* The results were unequivocal. For SOAs ranging from 100 to 400 msec, the naming latencies for the two kinds of distractor were equally, and substantially, primed. Only at SOA = 600 msec did the subordinate's phonological priming disappear. This clearly violates the principle: Both synonyms are phonologically active, not just the preferred one (i.e., the one that the subject was probably preparing), and initially they are equally active.

In a second set of experiments, Peterson and Savoy tested the phonological activation of nonsynonymous semantic alternatives, such as *bed* for *couch* (here the phonological distractor would be *bet).* This, then, was a straight replication of our experiment. So were the results. There was not the slightest phonological activation of these semantic alternatives, just as we had found. Peterson and

Savoy's conclusion was that there was multiple phonological activation only of *actual* picture names. Still, as Peterson and Savoy argue, that finding alone is problematic for the above principle and supportive of cascading models.

Recently, Jescheniak and Schriefers (1998) independently tested the same idea in a picture–word interference task. When the subject was naming a picture (for instance, of a couch) and received a phonological distractor word related to a synonym (for instance, *soda)*, there was measurable interference with naming. The naming latency was longer in this case than when the distractor was unrelated to the target or its synonym (for instance, *figure)*. This supports Peterson and Savoy's findings.

What are we to make of this? Clearly, our theory has to be modified, but how? There are several ways to go. One is to give up the principle entirely, but that would be an overreaction, given the fact that multiple phonological activation has been shown to exist only for synonyms. Any other semantic alternative that is demonstrably *semantically* active has now been repeatedly shown to be phonologically entirely inert. One can argue that it is phonologically active nevertheless, as both Harley (1993) and Peterson and Savoy (1998) do, but unmeasurably so. Our preference is a different tack. In his account of word blends, Roelofs (1992a) suggested that

> "they might occur when two lemma nodes are activated to an equal level, and both get selected. . . . The selection criterion in spontaneous speech (i.e., select the highest activated lemma node of the appropriate syntactic category) is satisfied simultaneously by two lemma nodes. . . . This would explain why these blends mostly involve near-synonyms."

The same notion can be applied to the findings under discussion. In the case of near-synonyms, both lemmas often are activated to a virtually equal level. Especially under time pressure, the indecision will be solved by selecting both lemmas.[9] In following the above principle, this will then lead to activation of both word forms. If both lemmas are indeed about equally active (i.e., have about the same word frequency, as was indeed the case for Peterson and Savoy's materials), one would expect that, upon their joint selection, both word forms would be equally activated as well. This is exactly what Peterson and Savoy showed to be the case for their stimuli. Initially, for SOAs of 50–400 msec the dominant and subordinate word forms were indeed equally active. Only by SOA = 600 msec did the dominant word form take over.[10]

Is multiple selection necessarily restricted to near-synonyms? There is no good reason to suppose that it is. Peterson and Savoy talk about multiple activation of "actual picture names." We rather propose the notion "appropriate picture names." As was discussed in section 4.2, what is appropriate depends on the communicative context. There is no hard-wired connection between percepts and lexical concepts. It may, under certain circumstances, be equally appropriate to call an object either *flower* or *rose.* In that case, the two lemmas will compete for selection although they are not synonyms, and multiple selection may occur.

A final recent argument for activation spreading from nonselected lemmas stems from a study by Cutting and Ferreira (in press). In their experiment subjects named pictures of objects whose names were homophones, such as a (toy) ball. When an auditory distractor was presented with a semantic relation to the other meaning of the homophone, such as "dance" in the example, picture naming was facilitated. The authors' interpretation is that the distractor ("dance") activates the alternative (social event) *ball* lemma in the production network. This lemma, in turn, spreads activation to the shared word form <ball> and hence facilitates naming of the "ball" picture. In other words, not only the selected $ball_1$ lemma but also the nonselected $ball_2$ sends activation to the shared <ball> word form node. These nice findings, however, do not exclude another possible explanation. The distractor "dance" will semantically and phonologically coactivate its associate "ball" in the perceptual network. Given assumption 1 from section 3.2.4, this will directly activate the word form node in the production lexicon.

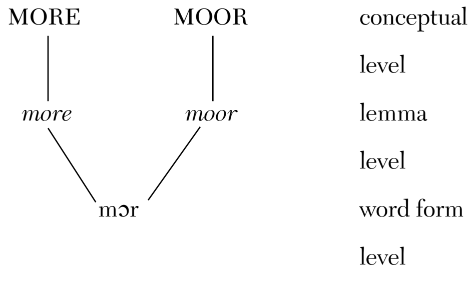
6.1.2. Do selected word forms feed back to the lemma level?

Preserving the accessing principle makes it theoretically impossible to adopt Dell's (1986; 1988) approach to the explanation of the often observed statistical overrepresentation of mixed errors (such as saying *rat* when the target is *cat)*. That there is such a statistical overrepresentation is well established by the recent paper of Martin et al. (1996). In that study 60 healthy controls and 29 aphasic speakers named a set of 175 pictures. Crucial here are the data for the former group. The authors carefully analyzed their occasional naming errors and found that when a semantic error was made there was an above-chance probability that the first or second phoneme of the error was shared with the target. This above-chance result could not be attributed to phonological similarities among semantically related words. In this study the old, often hotly debated factors such as perceiver bias, experimental induction, or set effects could not have produced the result. Clearly, the phenomenon is real and robust (see also Rossi & Defare 1995).

The crucial mechanism that Dell (1986; 1988), Dell et al. (1997b), and Martin et al. (1996) proposed for the statistical overrepresentation of mixed errors is feedback from the word form nodes to the lemma nodes. For instance, when the lemma *cat* is active, the morpheme <cat> and its segments /k/, /æ/, and /t/ become active. The latter two segments feed part of their activation back to the lemma *rat,* which may already be active because of its semantic relation to *cat.* This increases the probability of selecting *rat* instead of the target *cat.* For a word such as *dog,* there is no such phonological facilitation of a semantic substitution error, because the segments of *cat* will not feed back to the lemma of *dog.* Also, the effect will be stronger for *rat* than for a semantically neutral phonologically related word, such as *mat,* which is totally inactive from the start. This mechanism is ruled out by our activation principle, because form activation *follows* selection, so feedback cannot affect the selection process. We will not rehearse the elaborate discussions that this important issue has raised (Dell & O'Seaghdha 1991; 1992; Harley 1993; Levelt et al. 1991a; 1991b). Only two points are relevant here. The first is that, until now, there is no reaction time evidence for this proposed feedback mechanism. The second is that alternative explanations are possible for the statistical effects, in particular the case of mixed errors. Some of these were discussed by Levelt et al. (1991a). They were, essentially, self-monitoring explanations going back to the experiments by Baars et al. (1975), which showed that speakers can prevent the overt production of internally prepared indecent words, nonwords, or other output that violates general or task-specific criteria (more on this in sect. 10). However, in addition, it turns out that in WEAVER++, slightly modified to produce errors, mixed errors become overrepresented as well (see sect. 10) and this does not require feedback. Hence, although the mixed error case has now been empirically established beyond reasonable doubt, it cannot be a decisive argument for the existence of feedback from the form level to the lemma level.

6.1.3. The word frequency effect. One of the most robust findings in picture naming is the word frequency effect, discovered by Oldfield and Wingfield (1965). Producing an infrequent name (such as *broom)* is substantially slower than producing a frequent name (such as *boat).* From an extensive series of experiments (Jescheniak & Levelt 1994) it appeared that the effect arises at the level of accessing word forms. Demonstrating this required exclusion of all other levels of processing in the theory (see Fig. 1). This was relatively easy for pre- and postlexical levels of processing but was harder for the two major levels of lexical access, lemma selection and word form access. The prelexical level was excluded by using Wingfield's (1968) procedure. If the frequency effect arises in accessing the lexical concept, given the picture, it should also arise in a recognition task in which the subject is given a lexical concept (for instance, "boat") and has to verify the upcoming picture. There was a frequency effect neither in the "yes" nor in the "no" responses. This does not mean, of course, that infrequent objects are as easy to recognize as frequent objects but only that, for our pictures, where this was apparently well controlled, there is still a full-fledged word frequency effect.[11] Hence, that must arise at a different level. Similarly, a late level of phonetic-articulatory preparation could be excluded. The word frequency effect always disappeared in delayed naming tasks.

The main argument for attributing the word frequency effect to word form access rather than to lemma selection stemmed from an experiment in which subjects produced homophones. Homophones are different words that are pronounced the same. Take *more* and *moor,* which are homophones in various British-English dialects. In our theory they differ at the lexical concept level and at the lemma level, but they share their word form. In network representation:



The adjective *more* is a high-frequency word, whereas the noun *moor* is low-frequency. The crucial question now is whether low-frequency *moor* will behave like other, nonhomophonous, low-frequency words (such as *marsh),* or rather like other, nonhomophonous high-frequency words (such as *much).* If word frequency is coded at the lemma level, the low-frequency homophone *moor* should be as hard to access as the equally low-frequency nonhomophone *marsh.* If, however, the word frequency effect is due to accessing the word form, one should, paradoxically, predict that a low-frequency homophone such as *moor* will be accessed just as quickly as its high-frequency twin *more,* because they share the word form. Jescheniak and Levelt (1994) tested these alternatives in an experiment in which subjects produced low-frequency homophones (such as *moor),* as well as frequency-matched low-frequency nonhomophones (such as *marsh).* In addition, there were high-frequency nonhomophones, matched to the homophony twin (such as *much,* which is frequency matched to *more).* How can one have a subject produce a low-frequency homophone? This was done by means of a translation task. The Dutch subjects, with good mastery of English, were presented with the English translation equivalent of the Dutch low-frequency homophone. As soon as the word appeared on the screen, they were to produce the Dutch translation and the reaction time was measured. The same was done for the high- and low-frequency nonhomophonous controls. In this task, reaction times are also affected by the speed of recognizing the English word. This recognition speed was independently measured in an animateness decision task. All experimental items were inanimate terms, but an equal set of fillers consisted of animate words. The same subjects performed the push-button, animateness decision task on the English words one week after the main experiment. Our eventual data were the difference scores, naming latency (for the Dutch response word) minus semantic decision latency (for the English stimulus word). A summary of the findings is presented in Figure 9.

We obtained the paradoxical result. The low-frequency homophones (such as *moor)* were statistically as fast as the high-frequency controls (such as *much)* and substantially faster than the low-frequency controls (such as *marsh).* This shows that a low-frequency homophone inherits the fast access speed of its high-frequency partner. In other words, the frequency effect arises in accessing the word form rather than the lemma.

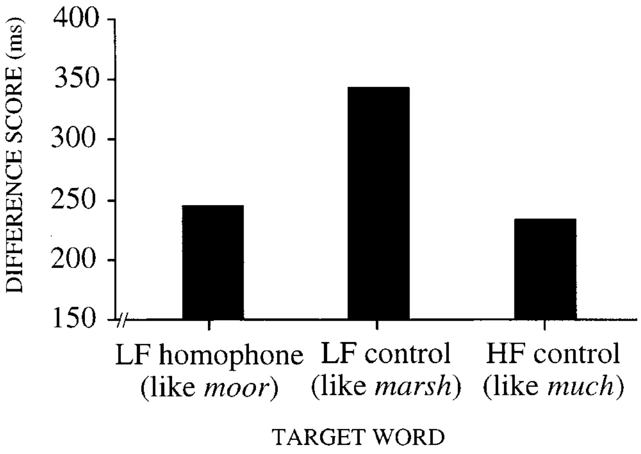


Figure 9. The homophone effect. (Corrected) naming latencies for low-frequency homophones (for example, *moor* as opposed to *more)* and nonhomophone controls that are frequency matched to the low-frequency homophone (e.g., *marsh)* or to the high-frequency twin (e.g., *much).* Data show that the low-frequency homophone inherits the accessibility of its high-frequency twin.

A related homophone effect has been obtained with speech errors. Earlier studies of sound-error corpora had already suggested that slips of the tongue occur more often on low-frequency words than on high-frequency ones (e.g., Stemberger & MacWhinney 1986). That is, segments of frequent words tend not to be misordered. Dell (1990) showed experimentally that low-frequency homophones adopt the relative invulnerability to errors of their high-frequency counterparts, completely in line with the above findings. Also in line with these results are Nickels's (1995) data from aphasic speakers. She observed an effect of frequency on phonological errors (i.e., errors in word form encoding) but no effect of frequency on semantic errors (i.e., errors in conceptually driven lemma retrieval). These findings suggest that the locus of the effect of frequency on speech errors is the form level.

There are, at least, two ways of modeling the effect, and we have no special preference. Jescheniak and Levelt (1994) proposed to interpret it as the word form's activation threshold, low for high-frequency words and high for low-frequency words. Roelofs (1997c) implemented the effect by varying the items' verification times as a function of frequency. Remember that, in the model, each selection must be licenced; this can take a varying amount of verification time.

Estimates of word frequency tend to correlate with estimates of age of acquisition of the words (see, e.g., Carroll & White 1973; Morrison et al. 1992; Snodgrass & Yuditsky 1996). Although some researchers found an effect of word frequency on the speed of object naming over and above the effect of age of acquisition, others have argued that it is age of acquisition alone that affects object naming time. In most studies, participants were asked to estimate at what age they first learned the word. It is not unlikely, however, that word frequency "contaminates" such judgments. When more objective measures of age of acquisition are used, however, it remains a major determinant of naming latencies. Still, some studies do find an independent contribution of word frequency (see, e.g., Brysbaert 1996). Probably, both factors contribute to naming latency. Morrison et al. (1992) compared object naming and categorization times and argued that the effect of age of acquisition arises during the retrieval of the phonological forms of the object names. This is, of course, exactly what we claim to be the case for word frequency. Pending more definite results, we will assume that both age of acquisition and word frequency affect picture naming latencies and that they affect the same processing step, that is, accessing the word form. Hence, in our theory they can be modeled in exactly the same way, either as activation thresholds or as verification times (see above). Because the independent variable in our experiments has always been CELEX word frequency,[12] we will continue to indicate the resulting effect by "word frequency effect." We do acknowledge, however, that the experimental effect is probably, in part, an age of acquisition effect.

The effect is quite robust, in that it is preserved over repeated namings of the same pictures. Jescheniak and Levelt (1994) showed this to be the case for three consecutive repetitions of the same pictures. In a recent study (Levelt et al. 1998), we tested the effect over 12 repetitions. The items tested were the 21 high-frequency and 21 low-frequency words from the original experiment that were monosyllabic. Figure 10 presents the results. The subjects had inspected the pictures and their names before the naming experiment began. The, on average, 31 msec word frequency effect was preserved over the full range of 12 repetitions.

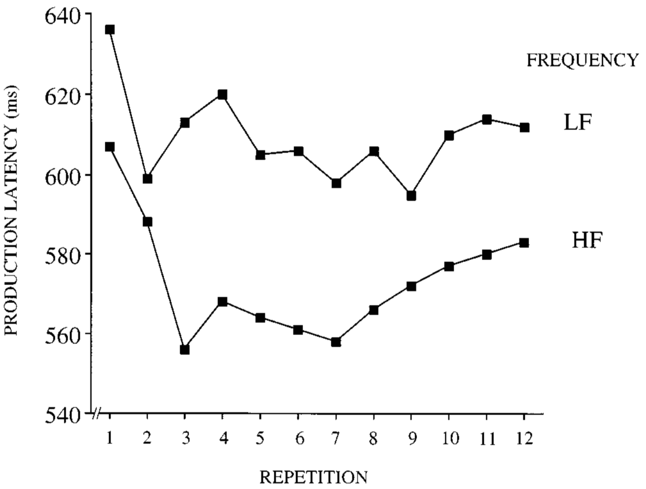


Figure 10. The robust word frequency effect. Naming latencies for 21 pictures with high-frequency names and 21 pictures with low-frequency names. The effect is stable over 12 repetitions.

6.2. Creating phonological words

The main task across the rift in our system is to generate the selected word's articulatory gestures in its phonological/phonetic context. This contextual aspect of word form encoding has long been ignored in production studies, which has led to a curious functional paradox.

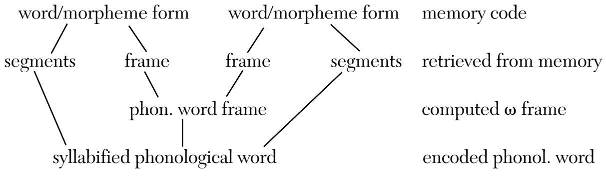
6.2.1. A functional paradox. All classical theories of phonological encoding have, in some way or another, adopted the notion that there are frames and fillers (Dell 1986; 1988; Fromkin 1971; Garrett 1975; Shattuck-Hufnagel 1979). The frames are metrical units, such as word or syllable frames; the fillers are phonemes or clusters of phonemes that are inserted into these frames during phonological encoding. Not only are there good linguistic reasons for such a distinction between structure and content, but speech error evidence seems to support the notion that constituency is usually respected in such errors. In *mell wade* (for *well made*) two word/syllable onsets are exchanged, in *bud beggs* (for *bed bugs*) two syllable nuclei are exchanged, and in *god to seen* (for *gone to seed*) two codas are exchanged (from Boomer & Laver 1968). This type of evidence has led to the conclusion that word forms are retrieved from the mental lexicon not as unanalyzed wholes but rather as sublexical and subsyllabic units, which are to be positioned in structures (such as word and syllable skeletons) that are independently available (Meyer, 1997, calls this the "standard model" in her review of the speech error evidence). Apparently, when accessing a word's form, the speaker retrieves both structural and segmental information. Subsequently, the segments are inserted in, or attached to, the structural frame which produces their correct serial ordering and constituent structure, somewhat like the following diagram.

word form — memory code
segments / frame — retrieved from memory (1)
word form — encoded word

Shattuck-Hufnagel, who was the first to propose a frame-filling processing mechanism (the "scan copier") that could account for much of the speech error evidence, right away noticed the paradox in her 1979 paper: "Perhaps its [the scan copier's] most puzzling aspect is the question of why a mechanism is proposed for the one-at-a-time serial ordering of phonemes when their order is already specified in the lexicon" (p. 338). Or, to put the paradox in more general terms, what could be the function of a mechanism that independently retrieves a word's metrical skeleton and its phonological segments from lexical memory and subsequently reunifies them during phonological encoding? It can hardly be to create the appropriate speech errors.

The paradox vanishes when the contextual aspect of phonological encoding is taken seriously. Speakers generate not lexical words but phonological words, and it is the phonological word, not the lexical word, that is the domain of syllabification (Nespor & Vogel 1986). For example, in *Peter doesn't understand it* the syllabification of the phrase *understand it* does not respect lexical boundaries, that is, it is not *un-der-stand-it.* Rather, it becomes *un-der-stan-dit,* where the last syllable, *dit,* straddles the lexical word boundary between *understand* and *it.* In other words, the segments are not inserted in a *lexical* word frame, as diagram (1) suggests, but in a larger *phonological* word frame. And what will become a phonological word frame is context dependent. The same lexical word *understand* will be syllabified as *un-der-stand* in the utterance *Peter doesn't understand.* Small, unstressed function words, such as *it, her, him,* and *on,* are pro- or encliticized to adjacent content words if syntax allows. Similarly, the addition of inflections or derivations creates phonological word frames that exceed stored lexical frames. In *understanding,* the lexical word-final *d* syllabifies with the inflection: *un-der-stan-ding;* the phonological word (ω) exceeds the lexical word. One could argue (as was done by Levelt 1989) that in such a case the whole inflected form is stored as a lexical word; but this is quite probably not the case for a rare derivation such as *understander,* which the speaker will unhesitantly syllabify as *un-der-stan-der.*

Given these and similar phonological facts, the functional significance of independently retrieving a lexical word's segmental and metrical information becomes apparent. The metrical information is retrieved for the construction of phonological word frames in context. This often involves combining the metrics of two or more lexical words or of a lexical word and an inflectional or derivational affix. Spelled-out segments are inserted not in retrieved *lexical* word frames, but in computed *phonological* word frames (but see sect. 6.2.4 for further qualifications). Hence, diagram (1) should be replaced by diagram (2).



(2)

In fact, the process can involve any number of stored lexical forms.

Although replacing diagram (1) with diagram (2) removes the functional paradox, it does not yet answer the question of why speakers do not simply concatenate fully syllabified lexical forms, that is, say things such as *un-der-stand-it* or *e-scort-us.* This would have the advantage for the listener that each morpheme boundary will surface as a syllable boundary, but speakers have different priorities. They are in the business of generating high-speed syllabic gestures. As we suggested in section 3.1.3, late, context-dependent syllabification contributes to the creation of maximally pronounceable syllables. In particular, there is a universal preference for allocating consonants to syllable onset positions, to build onset clusters that increase in sonority, and to produce codas of decreasing sonority (see, especially, Venneman 1988).

So far our treatment of phonological word formation has followed the standard theory, except that the domain of encoding is not the lexical word or morpheme but the phonological word, ω. The fact that this domain differs from the lexical domain in the standard theory resolves the paradox that always clung to it, but now we have to become more specific on segments, metrical frames, and the process of their association. It will become apparent that our theory of phonological encoding differs in two further important respects from the standard theory. The first difference concerns the nature of the metrical frames, and the second concerns the lexical specification of these frames. In particular we will argue that, different from the standard theory, metrical frames do not specify syllable-internal structure and that there are no lexically specified metrical frames for words adhering to the default metrics of the language, at least for stress-assigning languages such as Dutch and English. In the following we will first discuss the nature of the ingredients of phonological encoding, segments and frames, and then turn to the association process itself.

6.2.2. The segments. Our theory follows the standard model in that the stored word forms are decomposed into abstract phoneme-sized units. This assumption is based on the finding that segments are the most common error units in sound errors; 60–90% of all sound errors are single-segment errors (see, e.g., Berg 1988; Boomer & Laver 1968; Fromkin 1971; Nooteboom 1969; Shattuck-Hufnagel 1983; Shattuck-Hufnagel & Klatt 1979). This does not deny the fact that other types of error units are also observed. There are, on the one hand, consonant clusters that move as units in errors; about 10–30% of sound errors are of this sort. They almost always involve word onset clusters. Berg (1989) showed that such moving clusters tend to be phonologically coherent, in particular with respect to sonority. Hence it may be necessary to allow for unitary spell out of coherent word-onset clusters, as proposed by Dell (1986) and Levelt (1989). There is, on the other hand, evidence for the involvement of subsegmental phonological features in speech errors (Fromkin 1971) as in a slip such as *glear plue sky.* They are relatively rare, accounting for under 5% of the sound form errors. However, there is a much larger class of errors in which target and error differ in just one feature (e.g., *Baris* instead of *Paris).* Are they segment or feature errors? Shattuck-Hufnagel and Klatt (1979) and Shattuck-Hufnagel (1983) have argued that they should be considered as segment errors (but see Browman & Goldstein 1990; Meyer 1997). Is there any further reason to suppose that there is feature specification in the phonological spell out of segments? Yes, there is. First is the robust finding that targets and errors tend to share most of their features (Fromkin 1971; García-Albea et al. 1989; Garrett 1975; Nooteboom 1969). Second, Stemberger (1983; 1991a; 1991b), Stemberger and Stoel-Gammon (1991), and also

Berg (1991) have provided evidence for the notion that spelled-out segments are specified for some features but unspecified for others. Another way of putting this is that the segments figuring in phonological encoding are *abstract.* Stemberger et al.'s analyses show that asymmetries in segment interactions can be explained by reference to feature (under)specification. In particular, segments that are, on independent linguistic grounds, specified for a particular feature tend to replace segments that are unspecified for that feature. This is true even though the feature-unspecified segment is usually the more frequent one in the language. Stemberger views this as an "addition bias" in phonological encoding. We sympathize with Stemberger's notion that phonological encoding proceeds from spelling out rather abstract, not fully specified segments to a further, context-dependent filling in of features (see Meyer 1997), though we have not yet modeled it in any detail. This means at the same time that we do not agree with Mowrey and MacKay's (1990) conclusion that there are no discrete underlying segments in phonological encoding but only motor programs to be executed. If two such programs are active at the same time, all kinds of interaction can occur between them. Mowrey and MacKay's electromyographic (EMG) data indeed suggested that these are not whole-unit all-or-none effects, but, as the authors noted themselves, such data are still compatible with the standard model. Nothing in that model excludes the possibility that errors also arise at a late stage of motor execution. It will be quite another, and probably impracticable, thing to show that *all* sound error patterns can be explained in terms of motor pattern interactions.

6.2.3. The metrical frames. As was mentioned, our theory deviates from the standard model in terms of the nature of the metrical frames. The traditional story is based on the observation that interacting segments in sound errors typically stem from corresponding syllable positions: Onsets exchange with onsets, nuclei with nuclei, and codas with codas. This "syllable-position constraint" has been used to argue for the existence of syllable frames, that is, metrical frames that specify for syllable positions, onset, nucleus, and coda. Spelled-out segments are correspondingly marked with respect to the positions that they may take (onset, etc.). Segments that can appear in more than one syllable position (which is true for most English consonants) must be multiply represented with different position labels. The evidence from the observed syllable-position constraint, however, is not really compelling. Shattuck-Hufnagel (1985; 1987; 1992) has pointed out that more than 80% of the relevant cases in the English corpora that have been analyzed are errors involving word onsets (see also Garrett 1975; 1980). Hence, this seems to be a word-onset property in the first place, not a syllable-onset effect. English consonantal errors not involving word onsets are too rare to be analyzed for adherence to a positional constraint. That vowels tend to exchange with vowels must hold for the simple reason that usually no pronounceable string will result from a vowel → consonant replacement. Also, most of the positional effects other than word-onset effects follow from a general segment similarity constraint: Segments tend to interact with phonemically similar segments. In short, there is no compelling reason from the English sound error evidence to assume the existence of spelled-out syllabic frames. Moreover, such stored lexical syllable frames should be frequently broken up in the generation of connected speech, for the reasons discussed in section 6.2.1.

The situation may be different in other languages. Analyzing a German corpus, Berg (1989) found that word-onset consonants were far more likely to be involved in errors than word-internal syllable onsets, but in addition he found that word-internal errors preferentially arose in syllable-onset rather than coda positions. García-Albea et al. (1989) reported that, in their Spanish corpus, errors arose more frequently in word-internal than in word-initial syllable onset positions and that the syllable position constraint was honored in the large majority of cases. It is, however, not certain that these observations can be explained exclusively by assuming metrical frames with specified syllable positions. It is also possible that the described regularity arises, at least in part, because similar, rather than dissimilar, segments tend to interact with each other, because the phonotactic constraints of the language are generally honored (which excludes, for instance, the movement of many onset clusters into coda positions and vice versa), because syllables are more likely to have onsets than codas, or because onsets tend to be more variable than codas. In the present section, we treat the metrical frames of Dutch and English, and we will briefly discuss cross-linguistic differences in frame structures in section 6.4.7.

Because the parsing of phonological words into syllables is completely predictable on the basis of segmental information, we assume that syllable structure is not stored in the lexical entries but generated "on the fly," following universal and language-specific rules. Because some of these rules, in particular those involved in sonority gradient decisions, refer to features of segments, these must be visible to the processor. Hence, although features are not independently retrieved, the segments' internal composition must still be accessible to the phonological encoder.

What then is specified in the metrical frame, if it is not syllable-internal structure? For stress-assigning languages such as English and Dutch, we will make the following rather drastically minimal assumption:

> *Metrical frame:* The metrical frame specifies the lexical word's number of syllables and main stress position.

This is substantially less than what metrical phonology specifies for a word's metrical skeleton, but there is no conflict here. The issue for phonological encoding is what should be minimally specified in the mental lexicon for the speaker to build up, "from left to right," a metrically fully specified phonological word with complete specification of its phonological segments, their order, and the syllabification. Hence, the ultimate output of phonological word encoding should indeed comply with standard metrical phonology.

The metrical frame assumption is even weaker than what we have proposed in earlier publications (Levelt 1992; Levelt & Wheeldon 1994), where we assumed that syllable weight was also part of the metrical frame information (we distinguished between single and multiple mora syllables). However, syllable weight is better thought of as an emerging property of the syllabification process itself. Syllable weight is determined by the syllable's CV structure. In Dutch, for instance, any "closed" syllable (-VC, -VVC, -VCC) is heavy. Our experiments (sect. 6.4.7) have shown that a speaker cannot profit from experience with a target's CV pattern, whereas experience with its number of syllables/

stress pattern, together with segmental information, is an effective prime (Roelofs & Meyer 1998). We are aware of the fact that there is no unanimity in the literature regarding the independent representation of CV structure in the word's metrical frame. Stemberger (1990) has argued for the independent existence of CV-frame information from the higher probability of source/error pairs that share CV structure. The argument is weakened by the fact that this effect ignored the VV versus V structure of the vowels (i.e., long vs. short). Experimental evidence on the representation of CV structure is scarce. In our laboratory, Meijer (1994; 1996) used a translation task to prime a word's CV structure. Native speakers of Dutch with good knowledge of English saw an English word to be translated into Dutch. Shortly after the onset of the English word, they heard a Dutch distractor word that agreed or disagreed with the target in CV structure. In one experiment (Meijer 1996) a facilitatory effect of shared CV structure was obtained, but in another (Meijer 1994) this effect was not seen. Sevald et al. (1995) found that participants could pronounce more pairs of a mono- and a disyllabic target within a given response period when the monosyllable and the first syllable of the disyllabic target had the same CV structure (as in *kul – par.fen)* than when their CV structure differed (as in *kult – par.fen).* No further facilitation was obtained when the critical syllables consisted of the same segments (as in *par – par.fen).* This fine result shows that the CV structure of words is in some sense psychologically real; the facilitatory effect apparently had a fairly abstract basis. It does not imply, however, that CV structure is part of the metrical frame. The effect may arise because the same routines of syllabification were applied for the two syllables.[13] The CV priming effect obtained by Meijer (1994; 1996) might have the same basis. Alternatively, it could arise because primes and targets with the same CV structure are similar in their phonological features or because they activate syllable program nodes with similar addresses.

So far, our assumption is that speakers spell out utterly lean metrical word frames in their phonological encoding. For the verb *escort* it will be σσ′, for *Manhattan* it will be σσ′σ, et cetera. Here we deviate substantially from the standard model, but there is also a second departure from the standard model. It is this economy assumption:

> *Default metrics:* For a stress assigning language, no metrical frame is stored/spelled out for lexical items with regular default stress.

For these regular items, we assume, the phonological word is generated from its segmental information alone; the metrical pattern is assigned by default. What is "regular default stress"? For Dutch, as for English (Cutler & Norris 1988), it is the most frequent stress pattern of words, which follows this rule: "Stress the first syllable of the word with a full vowel." By default, closed class items are unstressed. Schiller (personal communication) has shown that this rule suffices to syllabify correctly 91% of all Dutch content word tokens in the CELEX database. Notice that this default assignment of stress does not follow the main stress rule in Dutch phonology, which states "stress the penultimate syllable of the word's rightmost foot," or a similar rule of English phonology. However, default metrics does not conflict with phonology. It is part of a phonological encoding procedure that will ultimately generate the correct metrical structure. Just as with the metrical frame assumption given above, default metrics are an empirical issue. Our experimental evidence so far (Meyer et al., in preparation) supports the default metrics assumption (cf. sect. 6.4.6). In short, in our theory the metrics for words such as the verb *escort* (σσ′) are stored and retrieved, but for words such as *father* (σ′σ) they are not.

6.2.4. Prosodification. Prosodification is the incremental generation of the phonological word, given the spelled-out segmental information and the retrieved or default metrical structure of its lexical components. In prosodification, successive segments are given syllable positions following the syllabification rules of the language. Basically, each vowel and diphthong is assigned to the nucleus position of a different syllable node, and consonants are treated as onsets unless phonotactically illegal onset clusters arise or there is no following vowel. Let us exemplify this from the *escort* example given in Figure 2. The first segment in the spell out of <escort> is the vowel /ə/ (remember that the order of segments is specified in the spell out). Being a vowel, it is made the nucleus of the first syllable. That syllable will be unstressed, following the retrieved metrical frame of <escort>, σσ′. The next segment, /s/, will be assigned to the onset of the next syllable. As was just mentioned, this is the default assignment for a consonant, but, of course, the encoder must know that there is indeed a following syllable. It can know this from two sources. One would be the retrieved metrical frame, which is bisyllabic. The other would be look ahead as far as the next vowel (i.e., /ɔ/). We have opted for the latter solution, because the encoder cannot rely on spelled out metrical frame information in the case of items with default metrics. On this ground, /s/ begins the syllable that will have /ɔ/ as its nucleus. The default assignment of the next segment /k/ is also to onset; then follows /ɔ/, which becomes nucleus. The remaining two segments, /r/ and /t/, cannot be assigned to the next syllable, because no further nucleus is spotted in look ahead. Hence, they are assigned to coda positions in the current syllable. Given the spelled-out metrical frame, this second syllable will receive word stress. The result is the syllabified phonological word /ə-skɔrt′/.

In planning polymorphemic phonological words, the structures of adjacent morphemes or words will be combined, as discussed in section 3.1.3. For instance, in the generation of *escorting*, two morphemes are activated and spelled out, <escort> and <ing>. The prevailing syntactic conditions will induce a phonological word boundary only after <ing>. The prosodification of this phonological word will proceed as described above. However, when /r/ is to be assigned to its position, the encoder will spot another vowel in its phonological word domain, namely, /ɪ/. It would now, normally, assign /r/ to the next syllable, but in this case that would produce the illegal onset cluster /rt/, which violates the sonority constraint. Hence, /r/ must be given a coda position, closing off the second syllable. The next segment /t/ will then become onset of the third syllable. This, in turn, is followed by insertion of nucleus /ɪ/ and coda /ŋ/ following rules already discussed. The final result is the phonological word /ə-skɔr′-tɪŋ/. The generation of the phrase *escort us* will follow the same pattern of operations. Here the prevailing syntactic conditions require cliticization of *us* to *escort;* hence, the phonological word boundary will be not after *escort* but after the clitic *us*. The resulting phonological word will be /ə-skɔr′-təs/.

Notice that in all these cases the word's syllabification and the internal structure of each syllable are generated on the fly. There are no prespecified syllable templates. For example, it depends only on the local context whether a syllable /-skɔr′/ or a syllable /- skɔrt′/ will arise.

Many, though not all, aspects of prosodification have been modeled in WEAVER++. Some main syllabification principles, such as maximization of onset (see Goldsmith 1990) have been implemented, but more is to be done. In particular, various aspects of derivational morphology still have to be handled. One example is stress shift in cases such as *expect* → *expectation.* The latter example shows that creating the metrical pattern of the phonological word may involve more than the mere blending of two spelled-out or default metrical structures. We will return shortly to some further theoretical aspects of syllabification in the experimental section 6.4. For now it suffices to conclude that the output of prosodification is a fully specified phonological word. All or most syllables in such representations are at the same time addresses of phonetic syllable programs in our hypothetical mental syllabary.

6.3. Word form encoding in WEAVER++

In our theory lemmas are mapped onto learned syllable-based articulatory programs by serially grouping the segments of morphemes into phonological syllables. These phonological syllables are then used to address the programs in a phonetic syllabary.

Let us once more return to Figure 2 in order to discuss some further details of WEAVER++'s implementation of the theory. The nonmetrical part of the form network consists of three layers of nodes: morpheme nodes, segment nodes, and syllable program nodes. Morpheme nodes stand for roots and affixes. Morpheme nodes are connected to the lemma and its parameters. The root verb stem <escort> is connected to the lemma *escort,* marked for "singular"or "plural." A morpheme node points to its metrical structure and to the segments that make up its underlying form. For storing metrical structures, WEAVER++ implements the economy assumption of default stress discussed above: for polysyllabic words that do not have main stress on the first stressable syllable, the metrical structure is stored as part of the lexical entry, but for monosyllabic words and for all other polysyllabic words it is not. At present, metrical structures in WEAVER++ still describe groupings of syllables into feet and of feet into phonological words. The latter is necessary because many lexical items have internal phonological word boundaries, as is, for instance, standard in compounds. With respect to feet, WEAVER++ is slightly more specific than the theory. It is an empirical issue whether a stored foot representation can be dispensed with. WEAVER++ follows the theory in that no CV patterns are specified.

The links between morpheme and segment nodes indicate the serial position of the segments within the morpheme. Possible syllable positions (onset, nucleus, coda) of the segments are specified by the links between segment nodes and syllable program nodes. For example, the network specifies that /t/ is the coda of syllable program [skɔrt] and the onset of syllable program [tɪŋ].

Encoding starts when a morpheme node receives activation from a selected lemma. Activation then spreads through the network in a forward fashion, and nodes are selected following simple rules (see Appendix). Attached to each node in the network is a procedure that verifies the label on the link between the node and a target node one level up. Hence, an active but inappropriate node cannot become selected. The procedures may run in parallel.

The morphological encoder selects the morpheme nodes that are linked to a selected lemma and its parameters. Thus, <escort> is selected for singular *escort.*

The phonological encoder selects the segments and, if available, the metrical structures that are linked to the selected morpheme nodes. Next, the segments are input to a prosodification process that associates the segments with the syllable nodes within the metrical structure (for metrically irregular words) or constructs metrical structures based on segmental information. The prosodification proceeds from the segment whose link is labeled first to the one labeled second, and so forth, precisely as described above, generating successive phonological syllables.

The phonetic encoder selects the syllable program nodes whose labeled links to the segments correspond to the phonological syllable positions assigned to the segments. For example, [skɔrt] is selected for the second phonological syllable of "escort," because the link between [skɔrt] and /k/ is labeled onset, between [skɔrt] and /ɔ/ nucleus, and between [skɔrt] and /r/ and /t/ coda. Similarly, the phonetic encoder selects [kɔr] and [tɪŋ] for the form "escorting." Finally, the phonetic encoder addresses the syllable programs in the syllabary, thereby making the programs available to the articulators for the control of the articulatory movements (following Levelt 1992; Levelt & Wheeldon 1994; see sect. 7.1). The phonetic encoder uses the metrical representation to set the parameters for loudness, pitch and duration. The hierarchical speech plan will then govern articulation (see, e.g., Rosenbaum et al. 1983).

The equations for form encoding are the same as those for lemma retrieval given above, except that the selection ratio now ranges over the syllable program nodes instead of the lemma nodes in the network. The equations for the expected encoding times of monosyllables and disyllables are given by Roelofs (1997c; submitted b); the Appendix gives an overview.

In sum, word form encoding is achieved by a spreading–activation-based network with labeled links that is combined with a parallel object-oriented production system. WEAVER++ also provides for a suspension/resumption mechanism that supports incremental or piecemeal generation of phonetic plans. Incremental production means that encoding processes can be triggered by a fragment of their characteristic input (Levelt 1989). The three processing stages compute aspects of a word form in parallel from the beginning of the word to its end. For example, syllabification can start on the initial segments of a word without having all of its segments. Only initial segments and, for some words, the metrical structure are needed to make a successful start. When given partial information, computations are completed as far as possible, after which they are put on hold. When given further information, the encoding processes continue from where they stopped.

6.4. Experimental evidence

In the following we will jointly discuss the experimental evidence collected in support of our theory of morphophonological encoding and its handling by WEAVER++. Together they make specific predictions about the time course of phonological priming, the incremental build-up of morphological and syllable structure, the modularity of morphological processing (in particular its independence of semantic transparancy), and the role of default versus spelled-out metrical structure in the generation of phonological words. One crucial issue here is how, in detail, the computer simulations were realized, and in particular how restrictive the parameter space was. This is discussed in an endnote,[14] which shows that the 48 data points referred to in section 6.4.1 below were fit with just six free parameters. These parameters, in turn, were kept fixed in the subsequent simulations depicted in Figures 12–14. Furthermore, size and content of the network have been shown not to affect the simulation outcomes.

In discussing the empirical evidence and its handling by WEAVER++, we will again make a distinction between empirical phenomena that were specifically built into the model and phenomena that the model predicts but had not been previously explored. For example, the assumption that the encoding proceeds from the beginning of a word to its end was motivated by the serial order effects in phonological encoding obtained by Meyer (1990; 1991), which we discuss below. The assumption led to the prediction of serial order effects in morphological encoding (Roelofs 1996a), which had not been tested before. Similarly, the assumption of on-line syllabification led to the prediction of effects of metrical structure (Roelofs & Meyer 1998) and morphological decomposition (Roelofs 1996a; 1996b).

6.4.1. SOA curves in form priming. The theory predicts that form encoding should be facilitated by presenting the speaker with an acoustic prime that is phonologically similar to the target word. Such a prime will activate the corresponding segments in the production network (which will speed up the target word's spell out) and also indirectly the syllable program nodes in the network (which will speed up their retrieval). These predictions depend, of course, on details of the further modeling.

Such a facilitatory effect of spoken distractor words on picture naming was first demonstrated by Schriefers et al. (1990) and was further explored by Meyer and Schriefers (1991). Their experiments were conducted in Dutch. The target and distractor words were either monomorphemic monosyllables or disyllables. The monosyllabic targets and distractors shared either the onset and nucleus (begin related) or the nucleus and coda (end related). For example, participants had to name a pictured bed (i.e., they had to say *bed,* [bɛt]), where the distractor was either *bek* ([bɛk]), "beak," which is begin related to target *bed,* or *pet* ([pɛt]), "cap," which is end related to [bɛt]; or there was no distractor (silence condition). The disyllabic targets and distractors shared either the first syllable (begin related) or the second syllable (end related). For example, the participants had to name a pictured table (i.e., they had to say *tafel,* [ta′.fəl]), where the distractor was *tapir* ([ta′.pir], "tapir," begin related to *tafel)* or *jofel* ([jo′.fəl], "pleasant," end related to *tafel).* Unrelated control conditions were created by recombining pictures and distractors. The distractor words were presented just before (i.e., −300 msec or −150 msec), simultaneously with, or right after (i.e., +150 msec) picture onset. Finally, there was a condition ("silence") without distractor.

The presentation of spoken distractors yielded longer object naming latencies compared to the situation without a distractor, but the naming latencies were less prolonged with related distractors than with unrelated ones. Thus a facilitatory effect was obtained from word form overlap relative to the nonoverlap situation. The difference between begin and end overlap for both the monosyllables and the disyllables was in the onset of the facilitatory effect. The onset of the effect in the begin related condition was at SOA = −150 msec, whereas the onset of the effect in the end condition occurred at SOA = 0 msec. With both begin and end overlap, the facilitatory effect was still present at the SOA of +150 msec.

Computer simulations showed that WEAVER++ accounts for the empirical findings (Roelofs 1997c). With begin overlap, the model predicts for SOA = −150 msec a facilitatory effect of −29 msec for the monosyllables (the real effect was −27 msec) and a facilitatory effect of −28 msec for the disyllables (real −31 msec). In contrast, with end overlap, the predicted effect for SOA = −150 msec was −3 msec for the monosyllables (real −12 msec) and −4 msec for the disyllables (real +10 msec). With both begin and end overlap the facilitatory effect was present at SOA 0 and +150 msec. Thus, the model captures the basic findings.

Figure 11 presents the WEAVER++ activation curves for the /t/ and the /f/ nodes during the encoding of *tafel* when *jofel* is presented as a distractor (i.e., the above disyllabic case with end overlap). Clearly, the activation of /f/ is greatly boosted by the distractor. In fact, it is always more active than /t/. Still, /t/ becomes appropriately selected in the target word's onset position. This is accomplished by WEAVER++'s verification procedure (see sect. 3.2.3).

6.4.2. Implicit priming. A basic premise of the theory is the incremental nature of morphophonological encoding. The phonological word is built up "from left to right," so to speak. The adoptation of rightward incrementality in the theory was initially motivated by Meyer's (1990;1991) findings and was

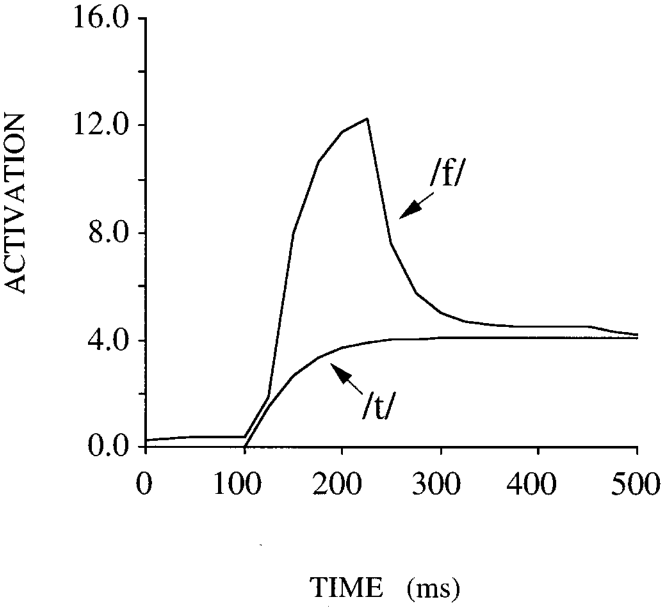


Figure 11. Activation curves for the /t/ and /f/ nodes in WEAVER++ during the encoding of *tafel.* Depicted is the aligned condition with the end-related distractor word *jofel* presented at SOA = 150 msec (after Roelofs 1997c).

further tested in new experiments. The implicit priming method involves producing words from learned paired associates. The big advantage of this paradigm compared to the more widely used picture–word interference (or "explicit priming") paradigm[15] is that the responses do not have to be names of depictable entities, which puts fewer constraints on the selection of materials. In Meyer's experiments, participants first learned small sets of word pairs such as *single–loner, place–local, fruit–lotus;* or *signal–beacon, priest–beadle, glass–beaker;* or *captain–major, cards–maker, tree–maple* (these are English examples for the Dutch materials used in the experiments). After learning a set, they had to produce the second word of a pair (e.g., *loner)* upon the visual presentation of the first word (*single*), the prompt. Thus, the second members of the pairs constitute the response set. The instruction was to respond as quickly as possible without making mistakes. The prompts in the set were repeatedly presented in random order, and the subjects' responses were recorded. The production latency (i.e., the interval between prompt onset and speech onset) was the main dependent variable. An experiment comprised homogeneous and heterogeneous response sets. In a homogeneous set, the response words shared part of their form and in a heterogeneous set they did not. For example, the responses could share the first syllable, as is the case in the above sets, *loner, local, lotus; beacon, beadle, beaker; major, maker, maple;* or they could share the second syllable as in *murder, ponder, boulder.* Heterogeneous sets in the experiments were created by regrouping the pairs from the homogeneous sets. For instance, regrouping the above homogeneous first syllable sets can create the new response sets *loner, beacon, major; local, beadle, maker;* and *lotus, beaker, maple.* Therefore, each word pair could be tested both under the homogeneous and under the heterogeneous condition, and all uncontrolled item effects were kept constant across these conditions.

Meyer found a facilitatory effect from homogeneity, but only when the overlap was from the beginning of the response words onward. Thus, a facilitatory effect was obtained for the set *loner, local, lotus* but not for the set *murder, ponder, boulder.* Furthermore, facilitation increased with the number of shared segments.

According to WEAVER++, this seriality phenomenon reflects the suspension–resumption mechanism that underlies the incremental planning of an utterance. Assume that the response set consists of *loner, local, lotus* (i.e., the first syllable is shared). Before the beginning of a trial, the morphological encoder can do nothing, the phonological encoder can construct the first phonological syllable (/ləʊ/), and the phonetic encoder can recover the first motor program [ləʊ]. When the prompt *single* is given, the morphological encoder will retrieve <loner>. Segmental spell out makes available the segments of this morpheme, which includes the segments of the second syllable. The phonological and phonetic encoders can start working on the second syllable. In the heterogeneous condition (*loner, beacon,* etc.), nothing can be prepared. There will be no morphological encoding, no phonological encoding, and no phonetic encoding. In the end-homogeneous condition (*murder, ponder,* etc.), nothing can be done either. Although the segments of the second syllable are known, the phonological word cannot be computed because the remaining segments are "to the left" of the suspension point. In WEAVER++, this means that the syllabification process has to go to the initial segments of the word, which amounts to restarting the whole process. Thus, a facilitatory effect will be obtained for the homogeneous condition relative to the heterogeneous condition for the begin condition only. Computer simulations of these experiments supported this theoretical analysis (Roelofs 1994; 1997c). Advance knowledge about a syllable was simulated by completing the segmental and phonetic encoding of the syllable before the production of the word. For the begin condition, the model yielded a facilitatory effect of −43 msec (real −49 msec), whereas for the end condition, it predicted an effect of 0 msec (real +5 msec). Thus, WEAVER++ captures the empirical phenomenon.

6.4.3. Priming versus preparation. The results of implicit and explicit priming are different in an interesting way. In implicit priming experiments, the production of a disyllabic word such as *loner* is speeded up by advance knowledge about the first syllable (/ləʊ/) but not by advance knowledge about the second syllable (/nɝ/), as shown by Meyer (1990; 1991). In contrast, when explicit first-syllable or second-syllable primes are presented during the production of a disyllabic word, both primes yield facilitation (Meyer & Schriefers 1991). As we saw, WEAVER++ resolves the discrepancy. According to the model, both first-syllable and second-syllable spoken primes yield facilitation, because they will activate segments of the target word in memory and therefore speed up its encoding. However, the effects of implicit priming originate at a different stage of processing, namely, in the rightward prosodification of the phonological word. Here, later segments or syllables cannot be prepared before earlier ones.

New experiments (Roelofs, submitted a) tested WEAVER++'s prediction that implicit and explicit primes should yield independent effects because they affect different stages of phonological encoding. In the experiments, there were homogeneous and heterogeneous response sets (the implicit primes) as well as form-related and form-unrelated spoken distractors (the explicit primes). Participants had to produce single words such as *tafel,* "table," simple imperative sentences such as *zoek op!,* "look up!," or cliticizations such as *zoek's op!,* "look up now!" where *'s* [əs] is a clitic attached to the base verb. In homogeneous sets, the responses shared the first syllable (e.g., *ta* in *tafel),* the base verb (e.g., *zoek,* "look" in *zoek op!),* or the base plus clitic (e.g., *zoek's* in *zoek's op!).* Spoken distractors could be related or unrelated to the target utterance. A related prime consisted of the final syllable of the utterance (e.g., *fel* for *tafel* or *op* for *zoek op!).* An unrelated prime was a syllable of another item in the response set. There was also a silence condition in which no distractor was presented. The homogeneity variable (called "context") and the distractor variable ("distractor") yielded main effects, and the effects were additive (see Fig. 12). Furthermore, as predicted by WEAVER++, the effects were the same for the production of single words, simple imperative sentences, and cliticizations, although these are quite different constructions. In particular, only in the single word case, the target consisted of a single phonological word. In the other two cases, the utterance consisted of two phonological words. We will return to this relevant fact in the next section.

6.4.4. Rightward incrementality and morphological decomposition. In section 5.3 we discussed the representation of morphology in the theory. There we saw that the sin-

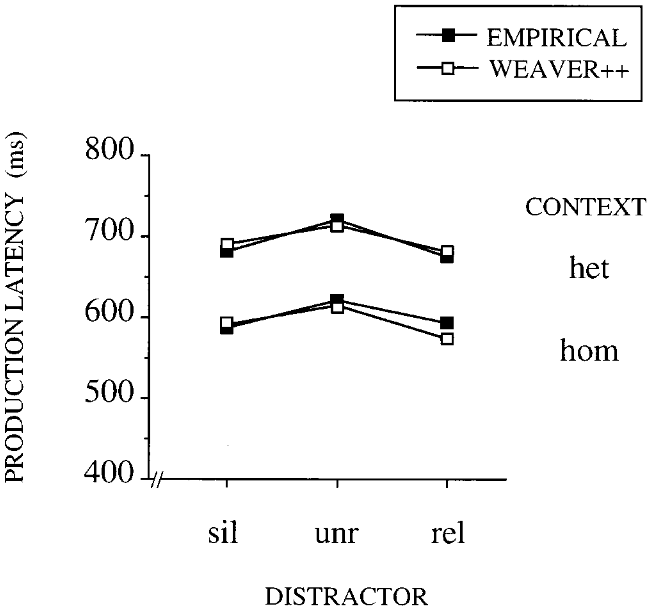


Figure 12. Combined effects of implicit and explicit priming. In the graph, "context" indicates the implicit variable "homogeneous" versus "heterogeneous" naming set; "distractor" denotes whether the auditory prime is phonologically related or unrelated to the second syllable of the target utterance. The two effects are additive, as they are in WEAVER++ simulation.

gle-lemma–multiple-morpheme case and the single-concept–multiple-lemma cases are the "normal" ones in complex morphology. Examples of the first type are prefixed words and most compounds; they are represented by a single lemma node at the syntactic level. An example of the latter type is particle verbs. In both cases, there are multiple morpheme nodes at the word form level, but only in case of the latter kind must two different lemmas be selected.

These cases of morphology are represented in WEAVER++'s encoding algorithm. It is characteristic of this algorithm not only to operate in a rightward incremental fashion but also that it requires morphologically decomposed form entries. Morphological structure is needed, because morphemes usually define domains of syllabification within lexical words (see Booij 1995). For example, without morphological structure, the second /p/ of *pop* in *popart* would be syllabified with *art,* following maximization of onset. This would incorrectly produce *po-part* (with the syllable-initial second *p* aspirated). The phonological word boundary at the beginning of the second morpheme *art* prevents that, leading to the syllabification *pop-art* (where the intervocalic /p/ is not aspirated because it is syllable-final).

Roelofs (1996a) tested effects of rightward incrementality and morphological decomposition using the implicit priming paradigm. WEAVER++ predicts that a larger facilitatory effect should be obtained when shared initial segments constitute a morpheme than when they do not. For example, the effect should be larger for sharing the syllable *by* (/baɪ/) in response sets including compounds such as *bystreet* (morphemes <by> and <street>) than for sharing the syllable /baɪ/ in sets including simple words such as *bible* (morpheme <bible>). Why would that be expected? When the monomorphemic word *bible* is produced in a homogeneous condition where the responses share the first syllable, the phonological syllable /baɪ/, and the motor program [baɪ] can be planned before the beginning of a trial. The morpheme <bible> and the second syllable /bəl/ will be planned during the trial itself. In a heterogeneous condition where the responses do not share part of their form, the whole monomorphemic word *bible* has to be planned during the trial. When the polymorphemic word *bystreet* is produced in a homogeneous condition where the responses share the first syllable, the first morpheme <by>, the phonological syllable (/baɪ/), and the motor program [baɪ] may be planned before the beginning of a trial. Thus, the second morpheme node <street> can be selected during the trial itself, and the second syllable /stri:t/ can be encoded at the phonological and the phonetic levels. In the heterogeneous condition, however, the initial morpheme node <by> has to be selected first, before the second morpheme node <street> and its segments can be selected so that the second syllable /stri:t/ can be encoded. Thus, in case of a polymorphemic word such as *bystreet,* additional morphological preparation is possible before the beginning of a trial. Consequently, extra facilitation should be obtained. Thus, the facilitatory effect for /baɪ/ in *bystreet* should be larger than the effect for /baɪ/ in *bible.*

The outcomes of the experiment confirmed these predictions. In producing disyllabic simple and compound nouns, a larger facilitatory effect was obtained when a shared initial syllable constituted a morpheme than when it did not (see Fig. 13).

The outcomes of further experiments supported WEAVER++'s claim that word forms are planned in a rightward fashion. In producing nominal compounds, no facilitation was obtained for noninitial morphemes. For example, no effect was obtained for <street> in *bystreet.* In producing prefixed verbs, a facilitatory effect was obtained for the prefix but not for the noninitial base. For example, a facilitatory effect was obtained for the Dutch prefix <be> of *behalen,* "to obtain," but not for the base <halen>.

Another series of experiments tested predictions of

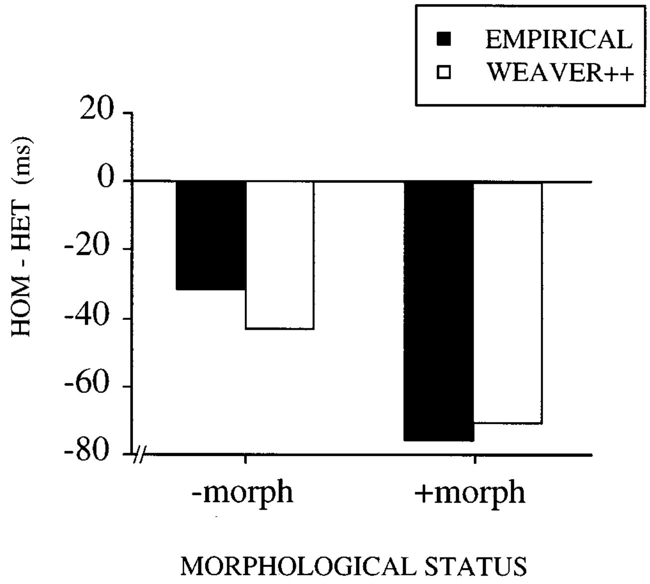


Figure 13. Implicit priming of a word's first syllable is more effective when that syllable is also a morpheme than when it is not. Experimental results and WEAVER++ simulation.

WEAVER++ about the generation of polymorphemic forms in simple phrasal constructions, namely, Dutch verb–particle combinations (Roelofs 1998). These are cases of single-concept–multiple-lemma morphology (sect. 5.3.3), given that the semantic interpretation of particle verbs is often not simply a combination of the meanings of the particle and the base. In producing a verb–particle construction, the lemma retriever recovers the two lemma nodes from memory and makes them available for syntactic encoding processes. In examining the production of particle verbs, again the implicit priming paradigm was used.

For particle–first infinitive forms, a facilitatory effect was obtained when the responses shared the particle but not when they shared the base. For example, in producing *opzoeken* "look up" (or rather "up look"), a facilitatory effect was obtained for the particle *op,* "up," but not for the base *zoeken,* "look." In Dutch particle verbs, the linear order of the major constituents can be reversed without creating another lexical item. That happens, for instance, in imperatives. For such base-first imperative forms, a facilitatory effect was obtained for the bases but not for the particles. For example, in producing *zoek op!,* "look up!," a facilitatory effect was obtained for *zoek,* "look," but not for *op,* "up." As was predicted by WEAVER++, the facilitatory effect was larger for the bases than for the particles (i.e., larger for *zoek* in *zoek op!* than for *op* in *opzoeken).* Bases such as *zoek* are longer and of lower frequency than particles such as *op.* Long fragments of low frequency take longer to encode than short fragments of high frequency, so the facilitatory effect from preparation will be higher in the former case. Subsequent experiments excluded the possibility that this difference in effect was due to the verb's mood or to the length of the nonoverlapping part and provided evidence for independent contributions of length and frequency (the latter following the mechanism discussed in sect. 6.1.3). This appeared from two findings. First, the facilitatory effect increased when the overlap (the implicit prime) became larger with frequency held constant. For example, the effect was larger for *door* (three segments) in *doorschieten,* "overshoot," than for *aan* (two segments) in *aanschieten,* "dart forward." Also, the effect was larger when the responses shared the particle and the first base syllable, such as *ople* in *opleven,* "revive," than when they shared the particle only, such as *op* in *opleven.* Second, bases of low frequency yielded larger facilitatory effects than bases of high frequency when length was held constant. For example, the effect was larger for *veeg,* "sweep" (low frequency), in *veeg op!,* "sweep up!," than for *geef,* "give," (high frequency) in *geef op!,* "give up!" A closely related result was obtained by Roelofs (1996c), but this time for compounds. When nominal compounds shared their initial morpheme, the facilitatory effect was larger when the morpheme was of low frequency (e.g., <schuim> in *schuimbad,* "bubble bath") than when it was high-frequency (e.g., <school> in *schoolbel,* "school bell"). This differential effect of frequency was stable over repetitions, which is compatible with the assumption that the locus of the effect is the form level rather than the lemma level (see sect. 6.1.3).

To return to the experiments with particle verbs, the results obtained with the items sharing the particle and the first base syllable (e.g., *ople* in *opleven)* are of special interest. The absence of a facilitatory effect for the bases and particles in second position (i.e., *zoeken* in *opzoeken* and *op* in *zoek op!)* in the earlier experiments does not imply that there was no preparation of these items. The particles and the bases in the first position of the utterances are independent phonological words. Articulation may have been initiated upon completion of (part of) this first phonological word in the utterance (i.e., after *op* in *opzoeken* and after *zoek* in *zoek op!).* If this was the case, then the speech onset latencies simply did not reflect the preparation of the second phonological word, even when such preparation might actually have occurred. The results for sharing *ople* in *opleven,* however, show that the facilitatory effect increases when the overlap crosses the first phonological word boundary. In producing particle verbs in a particle–first infinitive form, the facilitatory effect is larger when the responses share both the particle syllable and the first base syllable than when only the particle syllable is shared. This suggests that planning a critical part of the second phonological word, that is, the base verb, determined the initiation of articulation in the experiments rather than planning the first phonological word (the particle) alone. These results in morphological encoding give further support to a core feature of the theory, the incrementality of word form encoding in context.

6.4.5. Semantic transparency. The upshot of the previous section is that a word's morphology is always decomposed at the form level of representation, except for the occasional degenerate case (such as *replicate),* whether or not there is decomposition on the conceptual or lemma level. This crucial modularity claim was further tested in a study by Roelofs et al. (submitted), which examined the role of semantic transparency in planning the forms of polymorphemic words. According to WEAVER++, morphological complexity can play a role in form planning without having a synchronic semantic motivation.

There are good a priori reasons for the claim that morphological processing should not depend on semantic transparency. One major argument derives from the syllabification of complex words. Correct syllabification requires morpheme boundaries to be represented in semantically opaque words. In Dutch this holds true for a word such as *oogappel,* "dear child." The word's meaning is not transparent (though biblical, "apple of the eye"), but there should be a syllable boundary between *oog* and *appel,* that is, between the composing morphemes (if the word were treated as a single phonological word in prosodification, it would syllabify as *oo-gap-pel).* The reverse case also occurs. Dutch *aardappel,* "potato," literally "earth apple," is semantically rather transparent. However, syllabification does not respect the morpheme boundary; it is *aar-dap-pel.* In fact, *aardappel* falls in our "degenerate" category, which means that it is not decomposed at the form level. This double dissociation shows that semantic transparancy and morphological decomposition are not coupled. In WEAVER++, nontransparent *oogappel* is represented by two morpheme nodes <oog> and <appel>, whereas "transparent" *aardappel* is represented by one node <aardappel>. Other reasons for expecting independence of morphological processing are discussed by Roelofs et al. (submitted).

In WEAVER++, morphemes are planning units when they determine aspects of the form of words such as their syllabification, independent of transparency. Roelofs et al. (submitted) obtained morphological priming for compounds (e.g., *bystreet* and *byword)* but not for simple nouns (e.g., *bible),* and the size of the morphemic effect was identical for transparent compounds (*bystreet*) and opaque compounds (*byword*). In producing prefixed verbs, the priming effect of a shared prefix (e.g., *ont,* "de-") was the same for fully transparent prefixed verbs (*ontkorsten,* "de-crust," "remove crust"), opaque prefixed verbs with meaningful free bases (*ontbijten,* "to have breakfast," which has *bijten,* "to bite," as base), and opaque prefixed verbs with meaningless bound bases (*ontfermen,* "to take pity on"). In the production of simple and prefixed verbs, morphological priming for the prefixed verbs was obtained only when morphological decomposition was required for correct syllabification. That is, the preparation effect was larger for *ver-* in *vereren,* "to honor," which requires morpheme structure for correct syllabification (*ver-eren*), than for *ver-* in *verkopen,* "to sell," where morpheme structure is superfluous for syllabification (*ver-kopen*), because /rk/ is an illegal onset cluster in Dutch. The preparation effect for the latter type of word was equal to that of a morphologically simple word. These results suggest that morphemes may be planning units in producing complex words without making a semantic contribution. Instead, they are planning units when they are needed to compute the correct form of the word.

6.4.6. Metrical structure. Whereas incrementality has been a feature of the standard model all along, our theory is substantially different in its treatment of metrical frame information. Remember the two essential features. First, for a stress-assigning language, stored metrical information consists of number of syllables and position of main-stress syllable, no less, no more. Second, for a stress-assigning language, metrical information is stored and retrieved only for "nonregular" lexical items, that is, items that do not carry main stress on the first full vowel. These are strong claims. The present section discusses some of the experimental evidence we have obtained in support of these claims.

Roelofs and Meyer (1998) conducted a series of implicit priming experiments testing predictions of WEAVER++ about the role of metrical structure in the production of polysyllabic words that do not have main stress on the first stressable syllable. According to the model, the metrical structures of these words are stored in memory. The relevant issue now is whether the stored metrical information is indeed essential in the phonological encoding of the word, or, to put it differently, is a metrical frame at all required in the phonological encoding of words? (Béland et al., 1990, discuss a syllabification algorithm for French, which does not involve a metrical frame. At the same time, they suggest that speakers frequently access a stored, already syllabified representation of the word.)

As in previous implicit priming experiments, participants had to produce one Dutch word, out of a set of three or four, as quickly as possible. In homogeneous sets, the responses shared a number of word-initial segments, whereas in heterogeneous sets, they did not. The responses shared their metrical structure (the constant sets), or they did not (the variable sets). WEAVER++ computes phonological words for these types of words by integrating independently retrieved metrical structures and segments. Metrical structures in the model specify the number of syllables and the stress pattern but not the CV sequence.

WEAVER++'s view of syllabification implies that preparation for word-initial segments should be possible only for response words with identical metrical structure. This prediction was tested by comparing the effect of segmental overlap for response sets with a constant number of syllables such as {*ma-nier′*, "manner," *ma-tras′*, "mattress," *ma-kreel′*, "mackerel"} to that for sets having a variable number of syllables such as {*ma-joor′*, "major," *ma-te′-rie,* "matter," *ma-la′-ri-a,* "malaria"}, with two, three, and four syllables, respectively. In this example, the responses share the first syllable /ma/. Word stress was always on the second syllable. Figure 14 shows that, as predicted, facilitation (from sharing the first syllable) was obtained for the constant sets but not for the variable sets. This shows that, even in order to prepare the first syllable, the encoder must know the word's ultimate number of syllables.

What about the main stress position, the other feature of stored metrics in our theory? This was tested by comparing the effect of segmental overlap for response sets with a constant stress pattern versus sets with a variable stress pattern, but always with the same number of syllables (three). An example of a set with constant stress pattern is {*ma-ri′-ne,* "navy," *ma-te′-rie,* "matter," *ma-lai′-se,* "depression," *ma-don′-na,* "madonna"}, where all responses have stress on the second syllable. An example of a set with variable stress pattern is {*ma-ri′-ne,* "navy," *ma-nus-cript′*, "manuscript," *ma-te′-rie,* "matter," *ma-de-lief′*, "daisy"}, containing two items with second syllable stress and two items with third-syllable stress. Again, as predicted, facilitation was obtained for the constant sets but not for the variables sets. This shows that, in the phonological encoding of an "irregularly" stressed word, the availability of the stress information is indispensible, even for the encoding of the word's first syllable, which was unstressed in all cases. WEAVER++ accounts for the key empirical findings. In contrast, if metrical structures are not involved in advance planning or if metrical structures are computed on line on the basis of segments for these words, sharing metrical structure should be irrelevant for preparation. The present results contradict that claim.

In WEAVER++, metrical and segmental spell out occur in parallel and require about the same amount of time.

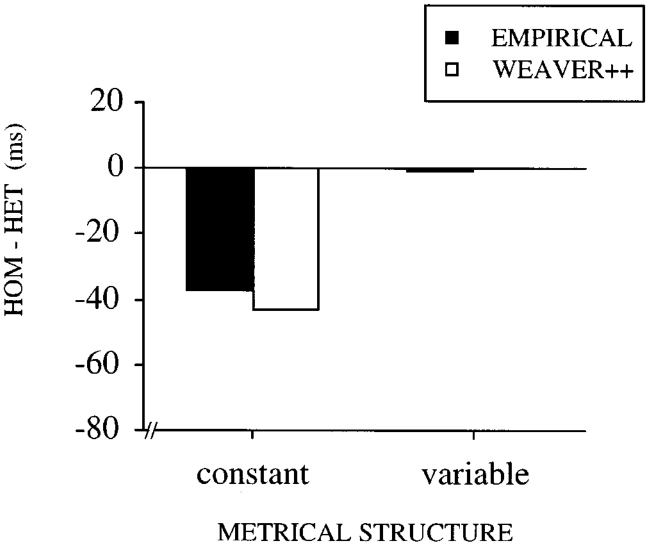


Figure 14. Implicit first-syllable priming for words with the same number of syllables versus words with a different number of syllables. Results show syllable priming in the former but not in the latter conditions. WEAVER++ predictions are also presented.

Consequently, sharing the number of syllables or stress pattern without segmental overlap should have no priming effect (this argument was first put forward by Meijer 1994). That is, pure metrical priming should not be obtained. If initial segments are shared but the metrical structure is variable, the system has to wait for metrical spell out, and no facilitation will be obtained (as shown in the experiments mentioned above). However, the reverse should also hold. If metrical spell out can take place beforehand, but there are no pregiven segments to associate to the frame, no facilitation should be obtained. This was tested in two new experiments. One experiment directly compared sets having a constant number of syllables such as {*ma-joor′*, "major," *si-gaar′*, "cigar," *de-tail′*, "detail"}, all disyllabic, to sets having a variable number of syllables such as {*si-gaar′*, "cigar," *ma-te′-rie,* "matter," *de-li′-ri-um,* "delirium"}, with two, three, and four syllables, respectively. Mean response times were not different between the two sets. In another experiment, sets with a constant stress pattern such as {*po′-di-um,* "podium," *ma′-ke-laar,* "broker," *re′-gi-o,* "region"}, all with stress on the first syllable, were directly compared to sets with a variable stress pattern such as {*po′-di-um,* "podium," *ma-don′-na,* "madonna," *re-sul-taat′*, "result"}, with stress on the first, second, and third syllables, respectively. Again, response latencies were statistically not different between the two sets. Hence knowing the target word's metrical structure in terms of number of syllables or stress pattern is in itself no advantage for phonological encoding. There must be shared initial segments as well in order to obtain an implicit priming effect. In summary, the data so far confirm the indispensability of retrieved metrical frames in phonological encoding.

The second feature of our theory, see section 6.2.3, argues against this indispensability, though for a subset of lexical items. It says that no retrieved metrical frame is required for the prosodification of words with default metrical structure. This prediction was made by Meyer et al. (in preparation) and implemented in WEAVER++. The experiments tested whether for these default words prosodification, including stress assignment, can go ahead without metrical preinformation. Implicit priming of initial segments should now be possible for both metrically constant and variable sets. This prediction was tested by comparing the effect of segmental overlap for response sets with a constant number of syllables, such as {*bor′-stel,* "brush," *bot′-sing,* "crash," *bo′-chel,* "hump," *bon′-je,* "rumpus"}, all disyllables stressed on the first syllable, to that for sets having a variable number of syllables such as {*bor′-stel,* "brush," *bot′-sing,* "crash," *bok′*, "goat," *bom′*, "bomb"}, with two disyllables stressed on the first syllable and two monosyllables, respectively. In the example, the responses share the onset and nucleus /bo/. As predicted, facilitation was obtained for both the constant and the variable sets. The same result is predicted for varying the number of syllables of polysyllabic words with an unstressable first syllable (i.e., schwa-initial words) and stress on the second syllable. This prediction was tested by comparing the effect of segmental overlap for response sets with a constant number of syllables such as {*ge-bit′*, "teeth," *ge-zin′*, "family," *ge-tal′*, "number," *ge-wei′*, "antlers"}, all disyllables having stress on the second syllable, to that for sets having a variable number of syllables such as {*ge-raam′-te,* "skeleton," *ge-tui′-ge,* "witness," *ge-bit′*, "teeth," *ge-zin′*, "family"}, with two disyllables stressed on the second syllable and two trisyllables stressed on the second syllable, respectively. As predicted, facilitation was obtained for both the constant and the variable sets.

6.4.7. Syllable priming. A core assumption of our theory is that there are no syllable representations in the form lexicon. Syllables are never "spelled out," that is, retrieved during phonological encoding. Rather, syllabification is a late process, taking place during prosodification; it strictly follows form retrieval from the lexicon.

Ferrand et al. (1996) recently obtained evidence for a late syllabification process in French. They conducted a series of word naming, nonword naming, picture naming, and lexical decision experiments using a masked priming paradigm. Participants had to produce French words such as *balcon,* "balcony," and *balade,* "ballad." Although the words *balcon* and *balade* share their first three segments, /b/, /a/, and /l/, their syllabic structure differs, such that *bal* is the first syllable of *bal-con* but more than the first syllable of *ba-lade,* whereas *ba* is the first syllable of *ba-lade* but less than the first syllable of *bal-con.* A first finding was that word naming latencies for both disyllabic and trisyllabic words were faster when preceded by written primes that corresponded to the first syllable (e.g., *bal* for *bal-con* and *ba* for *ba-lade)* than when preceded by primes that contained one letter (segment) more or one less than the first syllable of the target (e.g., *ba* for *bal-con* and *bal* for *ba-lade).* Second, these results were also obtained with disyllabic nonword targets in the word naming task. Third, the syllable priming effects were also obtained using pictures as targets. Finally, the syllable priming effects were not obtained with word and nonword targets in a lexical decision task.

The fact that the syllable priming effects were obtained for word, nonword, and picture naming but not for lexical decision suggests that the effects really are due to processes in speech production rather than to perceptual processes. Also, the finding that syllable priming was obtained for both word and nonword targets suggests that the effects are due to computed syllabifications rather than to the stored syllabifications that come with lexical items (i.e., different from the standard model, but in agreement with our theory). Syllabified nonwords, after all, are not part of the mental lexicon.

However, in spite of this, WEAVER++ does not predict syllable priming for Dutch or English (or even for French when no extra provisions are made). We will first discuss why that is so, and then contrast the Ferrand et al. (1996) findings for French with recent findings from our own laboratory for Dutch, findings that do not show any syllable priming.

Why does WEAVER++ not predict a syllable priming effect? When a prime provides segmental but no syllabic information, the on-line syllabification will be unaffected in the model. In producing a CVC.VC word, a CV prime will activate the corresponding first two segments and partly the CVC syllable program node for the first syllable, whereas a CVC prime will activate the first three segments and fully the syllable program node of the first CVC syllable. The longer CVC prime, which matches the first syllable of the word, will therefore be more effective than the shorter CV prime. In producing a CV.CVC word, a CV prime will activate the corresponding first two segments and the syllable program node for the first CV syllable, whereas a CVC prime will activate the first three segments, the full first CV syllable program node as well as partly the second syllable program node (via its syllable-initial C). Thus, again, the longer CVC prime, which now does not correspond to the first syllable of the word, will be more effective than the shorter CV prime, which does correspond to the first syllable. Thus, the model predicts an effect of prime length but no "crossover" syllabic effect. Without further provisions, therefore, Ferrand et al.'s findings are not predicted by our model. Before turning to that problem, let us consider the results for Dutch syllable priming obtained in our laboratory.

A first set of results stems from a study by Baumann (1995). In a range of elegant production experiments, she tested whether auditory syllable priming could be obtained. One crucial experiment was the following. The subject learned a small set of semantically related A–B pairs (such as *pijp–roken,* "pipe–smoke"). In the experiment the A word was presented on the screen and the subject had to produce the corresponding B word from memory; the response latency was measured. All B words were verbs, such as *roken,* "smoke." There were two production conditions. In one, the subject had to produce the verb in its infinitive form (in the example: *roken,* which is syllabified as *ro-ken).* In the other condition, the verb was to be produced in its past tense form (viz. *rookte,* syllabified as *rook-te).* This manipulation caused the first syllable of the target word to be either a CV or a CVC syllable (viz. /ro:/ vs. /ro:k/). At some SOA after presentation of the A word (−150, 0, 150, or 300 msec), an auditory prime was presented. It could be either the relevant CV (viz. [ro:]) or the relevant CVC (viz. [ro:k]), or a phonologically unrelated prime. The primes were obtained by splicing from spoken tokens of the experimental target verb forms. The main findings of this experiment were: (1) related primes, whatever their syllabic relation to the target word, facilitated the response; latencies on trials with related primes were shorter than latencies on trials with phonologically unrelated primes; in other words, the experimental procedure was sensitive enough to pick up phonological priming effects; (2) CVC primes were in all cases more effective than CV primes; hence, there is a prime length effect, as predicted by WEAVER++; and (3) there was no syllable priming effect whatsoever, again as predicted by WEAVER++.

Could the absence of syllable priming effects in Baumann's (1995) experiments be attributed to the use of *auditory* primes or to the fact that the subjects were aware of the prime? Schiller (1997; 1998) replicated Ferrand et al.'s visual masked priming procedure for Dutch. In the main picture naming experiment, the disyllabic target words began with a CV syllable (as in *fa-kir)* or with a CVC syllable (as in *fak-tor)* or the first syllable was ambisyllabic CV[C] (as in *fa[kk]el,* "torch"). The visual masked primes were the corresponding orthographic CV or CVC or a neutral prime (such as %&$). Here are the major findings of this experiment: (1) Related primes, whatever their syllabic relation to the target word, facilitated the response (i.e., compared to neutral primes); (2) CVC primes were in all cases more effective than CV primes; hence there is a prime length effect, as predicted by WEAVER++; and (3) there was no syllable priming effect whatsoever, again as predicted by WEAVER++. In short, this is a perfect replication of the Baumann (1995) results, which were produced with nonmasked auditory primes.

Hence the main problem for our model is to provide an explanation for the positive syllable priming effects that Ferrand et al. (1996) obtained for French. We believe it is to be sought in French phonology and its reflection in the French input lexicon. French is a syllable-timed language with rather clear syllable boundaries, whereas Dutch and English are stress-timed languages with substantial ambisyllabicity (see Schiller et al., 1997, for recent empirical evidence on Dutch). The classical syllable priming results of Cutler et al. (1986) demonstrate that this difference is reflected in the perceptual segmentation routines of native speakers. Whereas substantial syllable priming effects were obtained for French listeners listening to French, no syllable priming effects were obtained for English listeners listening to English. Also, for Dutch, the syllable is not used as a parsing unit in speech perception (Cutler, in press). Another way of putting this is that in French, but not in English or Dutch, input segments are assigned to syllable positions. For instance, in perceiving *balcon,* the French listener will encode /l/ as a syllable coda segment, /$\mathrm{l}_{\mathrm{coda}}$/, but, in *ballade,* the /l/ will be encoded as onset segment, /$\mathrm{l}_{\mathrm{onset}}$/. The English listener, however, will encode /l/ in both *balcony* and *ballad* as just /l/, that is, unspecified for syllable position (and similarly for the Dutch listener). Turning now to Ferrand et al.'s results, we assume that the orthographic masked prime activates a phonological syllable, with position-marked segments. These position-marked phonological segments in the perceptual network spread their activation to just those syllables in WEAVER++'s syllabary where the segment is in the corresponding position. For instance, the orthographic CVC prime BAL will activate the phonological syllable /bɑl/ in the input lexicon, and hence the segment /$\mathrm{l}_{\mathrm{coda}}$/. This segment, in turn, will spread its activation to *balcon's* first syllable ([bɑl]) in the syllabary, but not *ballade's* second syllable ([la:d]); it will, in fact, interfere because it will activate alternative second syllables, namely, those ending in [l]. As a consequence, CV prime BA will be more effective than CVC prime BAL as a facilitator of *ballade,* but CVC prime BAL will be more effective than CV prime BA as a facilitator of *balcon.* Notice that in this theory the longer prime (CVC) is, on average, not more effective than the shorter prime (CV). This is because the position-marked second C of the CVC prime has no facilitatory effect. This is exactly what Ferrand et al. (1996) found: they obtained no prime length effect. However, such a prime length effect should be found if the extra segment is not position marked, because it will facilitate the onset of the next syllable. That is what both Baumann and Schiller found in their experiments.

Two questions remain. The first is why, in a recent study, Ferrand et al. (1997) did obtain a syllable priming effect for English. That study, however, did not involve picture naming, but only word reading and so the effect could be entirely orthographic in nature. Schiller (personal communication) did not obtain the English-language syllable priming effect in a recent replication of the Ferrand et al. (1997) experiment, nor did he obtain the effect in a picture naming version of the experiment. The second question is why Ferrand et al. (1996) did not obtain a syllable priming effect in lexical decision (the authors used that finding to exclude a perceptual origin of their syllable priming effects). If the French orthographic prime activates a phonological input syllable, why does it not speed up lexical decision on a word beginning with that syllable? That question is even more pressing in view of the strong syllable priming effects arising in French spoken word perception (Cutler et al. 1986;

Mehler et al. 1981). Probably, orthographic lexical decision in French can largely follow a direct orthographic route, not or hardly involving phonological recoding.

6.4.8. "Resyllabification." The claim that syllabification is late and does not proceed from stored syllables forces us to consider some phenomena that traditionally fall under the heading of "resyllabification." There is "resyllabification" if the surface syllabification of a phonological word differs from the underlying lexical syllabification. In discussing the "functional paradox" (sect. 6.2.1), we mentioned the two major cases of "resyllabification": in cliticization and in the generation of complex inflectional and derivational morphology. An example of the first was the generation of *escort us,* where the surface syllabification becomes *e-scor-tus;* which differs from the syllabification of the two underlying lexical forms, *e-scort* and *us.* Examples of the latter were *un-der-stan-ding* and *un-der-stander,* where the syllabification differs from that of the base term *under-stand.* These examples are not problematic for our theory; they do not require two subsequent steps of syllabification, but other cases cause more concern. Baumann (1995) raised the following issue. Dutch has syllable-final devoicing. Hence, the word *hond,* "dog," is pronounced as /hɔnt/. The voicing reappears in the plural form *hon-den,* where /d/ is no longer syllable-final. Now, consider cliticization. In pronouncing the phrase *de hond en de kat,* "the dog and the cat," the speaker can cliticize *en,* "and," to *hond.* The bare form of our theory predicts that exactly the same syllabification will arise here, because in both cases one phonological word is created from exactly the same ordered set of segments. Hence, the cliticized case should be *hon-den.* But it is not. Careful measurements show that it is *hon-ten.*

Why do we get devoicing here in spite of the fact that /d/ is not syllable-final? The old story here is real resyllabification. The speaker first creates the syllabification of *hond,* devoicing the syllable-final consonant. The resulting *hont* is then resyllabified with the following *en,* with *hon-ten* as the outcome. Is this a necessary conclusion? We do not believe it is. Booij and Baayen (in progress) have proposed a different solution for this case and many related ones, which is to list phonological alternates of the same phoneme in the mental lexicon, with their context of applicability. For example, in Dutch there would be two lexical items, <hont> and <hond>, where only the latter is marked for productive inflection/derivation. The first allomorph is the default, unmarked case. In generating plural *hon-den,* the speaker must access the latter, marked allomorph <hond>. It contains the segment /d/, which will appear as voiced in syllable-initial position, but, in case of cliticization, where no inflection or derivation is required, the speaker accesses the unmarked form <hont>, which contains the unvoiced segment /t/. By the entirely regular syllabification process described in section 6.2.4, the correct form *hon-ten* will result. There are two points to notice. First, this solution is not intended to replace the mechanism of syllable-final devoicing in Dutch. It works generally. Any voiced dostruent ending up in a syllable-final position during prosodification will usually be devoiced. Second, the solution multiplies lexical representations, and phonologists abhor this. However, as Booij and Baayen are arguing, there is respectable independent phonological, historical speech error and acquisition evidence for listing phonological alternates of the same lexical item. Our provisional conclusion is that resyllabification is never a real-time process in phonological word generation, but this important issue deserves further experimental scrutiny.

These considerations conclude our remarks on phonological encoding. The output of morphophonological word encoding, a syllabically and metrically fully specified phonological word, forms the input to the next stage of processing, phonetic encoding.

7. Phonetic encoding

Producing words involves two major systems, as we have argued. The first is a conceptually driven system that ultimately selects the appropriate word from a large and ever-expanding mental lexicon. The second is a system that encodes the selected word in its context as a motor program. An evolutionary design feature of the latter system is that it can generate an infinite variety of mutually contrasting patterns, contrasting in both the articulatory and the auditory senses. For such a system to work, it requires an abstract calculus of gesture/sound units and their possible patternings. This is the phonology the young child builds up during the first 3 years of life. It is also this system that is involved in phonological encoding, as was discussed in the previous section.

However, more must be done in order to encode a word as a motor action. This is to generate a specification of the articulatory gestures that will produce the word as an overt acoustic event in time. This specification is called a phonetic representation. The need to postulate this step of phonetic encoding follows from the abstractness of the phonological representation (see note 6). In our theory of lexical access, as in linguistic theory, the phonological representation is composed of phonological segments, which are discrete (i.e., they do not overlap on an abstract time axis), static (i.e., the features defining them refer to states of the vocal tract or the acoustic signal), and context free (i.e., the features are the same for all contexts in which the segment appears). By contrast, the actions realizing consonants and vowels may overlap in time, the vocal tract is in continuous movement, and the way features are implemented is context-dependent.

What does the phonetic representation look like? Though speakers ultimately carry out movements of the articulators, the phonetic representation most likely does not specify movement trajectories or patterns of muscle activity but rather characterizes speech tasks to be achieved (see, e.g., Fowler et al. 1980; Levelt 1989). The main argument for this view is that speakers can realize a given linguistic unit in infinitely many ways. The sound /b/, for instance, can be produced by moving both lips, or only one lip, with or without jaw movement. Most speakers can almost without practice adapt to novel speech situations. For instance, Lindblom et al. (1979) showed that speakers can produce acoustically almost normal vowels while holding a bite block between their teeth, forcing their jaw into a fixed open position. Abbs and his colleagues (Abbs & Gracco 1984; Folkins & Abbs 1975) asked speakers to produce an utterance repeatedly (e.g., "aba" or "sapapple"). On a small number of trials, and unpredictably for the participants, the movement of an articulator (e.g., the lower lip) was mechanically hampered. In general, these perturbations were almost immediately (within 30 msec after movement onset) compensated for, such that the utterance was acoustically almost normal. One way to account for these findings is that the phonetic representation specifies speech tasks (e.g., to accomplish lip closure) and that there is a neuromuscular execution system that computes how the tasks are best carried out in a particular situation (see, e.g., Kelso et al., 1986 and Turvey, 1990, for a discussion of the properties of such systems). Thus, in the perturbation experiments, participants maintained constant task descriptions on all trials, and on each trial the execution system computed the best way to fulfill them. The distinction between a specification of speech tasks and the determination of movements is attractive because it entails that down to a low planning level the speech plan is the same for a given linguistic unit, even though the actual movements may vary. It also invites an empirical approach to the assignment of fast speech phenomena and feature specification, such as reduction and assimilation. Some will turn out to be properties of the speech plan, whereas others may arise only in motor execution (see Levelt, 1989, for a review).

7.1. A mental syllabary?

How are phonetic representations created? The phonological representation, i.e., the fully specified phonological word, can be viewed as an ordered set of pointers to speech tasks. The phonological units that independently refer to speech tasks could be features or segments or larger units, such as demisyllables or syllables. Levelt (1992; see also Levelt & Wheeldon 1994), following Crompton's (1982) suggestion, has proposed that in creating a phonetic representation speakers may access a mental syllabary, which is a store of complete gestural programs for at least the high-frequency syllables of the language. Thus high-frequency phonological syllables point to corresponding units in the mental syllabary. A word consisting of *n* such syllables can be phonetically encoded by retrieving *n* syllable programs from the syllabary. The phonetic forms of words composed of low-frequency syllables are assembled using the segmental and metrical information provided in the phonological representation. (The forms of high-frequency syllables can be generated in the same way, but usually retrieval from the syllabary will be faster.) Levelt's proposal is based on the assumption that the main domain of coarticulation is the syllable (as was proposed, e.g., by Fujimura & Lovins, 1978, and Lindblom, 1983). Coarticulatory effects that cross syllable boundaries (as discussed, e.g., by Farnetani 1990; Kiritani & Sawashima 1987; Recasens 1984; 1987) are attributed to the motor execution system.

The obvious advantage of a syllabary is that it greatly reduces the programming load relative to segment-by-segment assembly of phonetic forms, in particular because as the syllables of a language differ strongly in frequency. But how many syllable gestures should be stored in such a hypothetical syllabary? That depends on the language. A syllabary would be most profitable for languages with a very small number of syllables, such as Japanese and Chinese. For languages such as English or Dutch, the situation might be different. Both languages have over 12,000 different syllables (on a CELEX count[12]). Will a speaker have all of these gestural patterns in store? Although this should not be excluded in principle (after all, speakers store many more lexical items in their mental lexicon), there is a good statistical argument to support the syllabary notion even for such languages.

Figure 15 presents the cumulative frequency of use for the 500 highest ranked syllables in English (the first 10 are /eɪ/, /ðiː/, /tuː/, /ʌv/, /ɪn/, /ænd/, /aɪ/, /lɪ/, /ə/, and /rɪ/). It appears from the curve that speakers can handle 50% of their speech with no more than 80 different syllables, and 500 syllables suffice to produce 80% of all speech.[16] The number is 85% for Dutch, as Schiller et al. (1996) have shown. Hence it would certainly be profitable for an English or Dutch speaker to keep the few hundred highest ranking syllables in store.

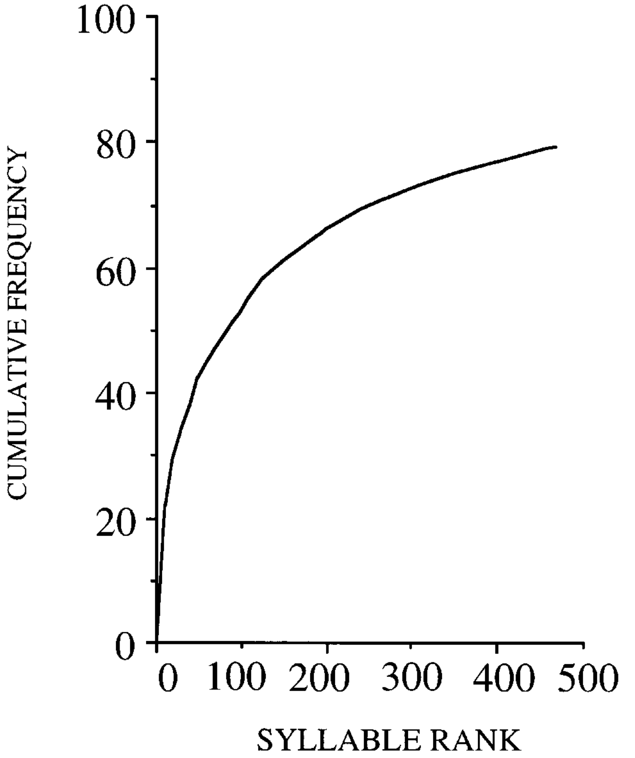


Figure 15. Cumulative frequency distribution for the 500 highest-ranked syllables in English. Derived from CELEX text-based statistics.

Experimental evidence compatible with this proposal comes from a study by Levelt and Wheeldon (1994), in which a syllable frequency effect was found that was independent of word frequency. Participants first learned to associate symbols with response words (e.g., /// = *apple*). On each trial of the following test phase, one of the learned symbols was presented (e.g., ///), and the participant produced the corresponding response word ("apple" in the example) as quickly as possible. In one experiment, speech onset latencies were found to be faster for disyllabic words that ended in a high-frequency syllable than for comparable disyllabic words that ended in a low-frequency syllable. This suggests that high-frequency syllables were accessed faster than low frequency ones, which implies the existence of syllabic units. However, in some of Levelt and Wheeldon's experiments, syllable and segment frequencies were correlated. In recent experiments by Levelt and Meyer (reported in Hendriks & McQueen 1996), in which a large number of possible confounding factors were controlled for, neither syllable nor segment frequency effects were obtained. These results obviously do not rule out that speakers retrieve syllables, or segments for that matter; they show only that the speed of access to these units does not strongly depend on their frequency. Other ways must be developed to approach the syllabary notion experimentally.

7.2. Accessing gestural scores in WEAVER++

The domain of our computational model WEAVER++ (Roelofs 1997c) ranges precisely to syllabary access, that is, the hinge between phonological and phonetic encoding in our theory. The mechanism was described in section 6.3. It should be added that WEAVER++ also accesses other, nonsyllabic speech tasks, namely, phonemic gestural scores. These are, supposedly, active in the generation of new or infrequent syllables.

7.3. The course of phonetic encoding

As far as the theory goes, phonetic encoding should consist of computing whole-word gestural scores from retrieved scores for syllables and segments. Much is still to be done. First, even if whole syllable gestural scores are retrieved, it must be specified for a phonological word how these articulatory tasks should be aligned in time. Also, still free parameters of these gestural scores, such as for loudness, pitch, and duration, have to be set (see Levelt 1989). Second, syllables in a word coarticulate. It may suffice to leave this to the articulatory-motor system, that is, it will execute both tasks at the right moments and the two patterns of motor instructions will simply add where there is overlap in time (Fowler & Saltzman 1993). However, maybe more is involved, especially when the two gestures involve the same articulators. Munhall and Löfquist (1992) call this *gestural aggregation*. Third, one should consider mechanisms for generating gestural scores for words from units smaller or larger than the syllable. Infrequent syllables must be generated from smaller units, such as demisyllables (Fujimura 1990) or segments. There might also well be a store of high-frequency, overused whole-word gestural scores, which still has no place in our theory. In its present state, our theory has nothing new to offer on any of these matters.

8. Articulation

There are, at least, two core theoretical aspects to articulation, its initiation and its execution (see Levelt, 1989, for a review). As far as initiation is concerned, some studies (Levelt & Wheeldon 1994; Schriefers et al., in press ; Wheeldon & Lahiri 1998) suggest that the articulation of a phonological word will be initiated only after all of its syllables have been phonetically encoded. This, then, puts a lower limit on incrementality in speech production, because a speaker cannot proceed syllable by syllable. The evidence, however, is so far insufficient to make this a strong claim. As far as execution of articulation is concerned, our theory has nothing to offer yet.

9. Self-monitoring

It is a property of performing any complex action that the actor exerts some degree of output monitoring. This holds true for the action of speaking (see Levelt, 1989, for a review). In self-monitoring a speaker will occasionally detect an ill-formedness or an all-out error. If these are deemed to be disruptive for realizing the current conversational intention, the speaker may decide to self-interrupt and make a correction. What is the output monitored? Let us consider the following two examples of spontaneous self-correction.

entrance to yellow . . . er, to gray

we can go straight to the ye- . . . to the orange dot

In both cases the trouble word was *yellow,* but there is an important difference. In the former example *yellow* was fully pronounced before the speaker self-interrupted. Hence the speaker could have heard the spoken error word and judged it erroneous. If so, the output monitored was overt speech. This is less likely for the second case; here the speaker self-interrupted while articulating *yellow.* To interrupt right after its first syllable, the error must already have been detected a bit earlier, probably before the onset of articulation. Hence some other representation was being monitored by the speaker. In Levelt (1989) this representation was identified with "internal speech." This is phenomenologically satisfying, because we know from introspection that indeed we can monitor our internal voice and often just prevent the embarrassment of producing an overt error. But what is internal speech? Levelt (1989) suggested that it was the "phonetic plan" or, in the present terminology, the gestural score for the word. However, Jackendoff (1987) proposed that the monitored representation is of a more abstract, phonological kind. Data in support of either position were lacking.

Wheeldon and Levelt (1995) set out to approach this question experimentally, guided by the theory outlined in this paper. There are, essentially, three[17] candidate representations that could be monitored in "internal speech." The first is the initial level of spell-out, in particular the string of phonological segments activated in word form access. The second is the incrementally produced phonological word, that is, the representation generated during prosodification. The third is the phonetic level of gestural scores, that is, the representation that ultimately drives articulation.

To distinguish between these three levels of representation, we developed a self-monitoring task of the following kind. The Dutch subjects, with a good understanding of English, were first given a translation task. They would hear an English word, such as *hitchhiker,* and had to produce the Dutch translation equivalent, for this example, *lifter.* After some exercise the experimental task was introduced. The participant would be given a target phoneme, for instance, /f/. Upon hearing the English word, the task was to detect whether the Dutch translation equivalent contained the target phoneme. That is the case for our example, *lifter.* The subject had to push a "yes" button in the positive case, and the reaction time was measured. Figure 16 presents the result for monitoring disyllabic CVC.CVC words, such as *lifter.* All four consonants were targets during different phases of the experiment.

It should be noticed that reaction times steadily increase for later targets in the word. This either expresses the time course of target segments becoming available in the production process or it is due to some "left-to-right" scanning pattern over an already existing representation. We will shortly return to this issue.

How can this method be used to sort out the three candidate levels of representation? Let us consider the latest representation first, the word's gestural score. We decided to wipe it out and check whether basically the same results would be obtained. If so, then that representation could not be the critical one. The subjects were given the same phoneme detection task, but there was an additional independent variable. In one condition the subject counted aloud during the monitoring task, whereas the other condition was without such a secondary task. This task is known to suppress the "articulatory code" (see, e.g., Baddeley et al. 1984). Participants monitored for the two syllable onset consonants (i.e., for /l/ or /t/ in the *lifter* example). Under both conditions the data in Figure 16 were replicated. Monitoring was, not surprisingly, somewhat slower during counting, and the RT difference between a word's two targets was a tiny bit less, but the difference was still substantial and significant. Hence the mechanism was not wiped out by this manipulation. Apparently the subjects could self-monitor without access to a phonetic–articulatory plan.

Which of the two earlier representations was involved? In our theory, the first level, initial segmental spell-out, is not yet syllabified, but the second level, the phonological word, is. Hence we tested whether self-monitoring is sensitive to syllable structure. Subjects were asked to monitor not for a target segment but for a CV or CVC target. The following English example illustrates the procedure. In one session the target would be /ta/ and in another session it would be /tal/. Among the test words in both cases were *talon* and *talcum.* The target /ta/ is the first syllable of *talon,* but not of *talcum,* whereas the target /tal/ is the first syllable of *talcum,* but not of *talon.* Would monitoring latencies reflect this interaction with syllable structure?

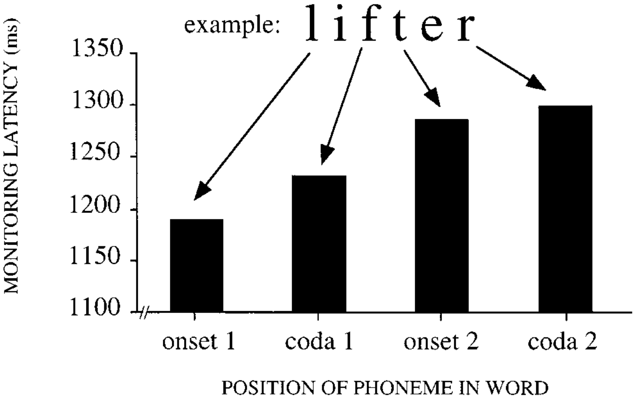


Figure 16. The self-monitoring task. Phoneme monitoring latencies for consonant targets in CVC – CVC words, such as *lifter.*

Figure 17 presents the results, showing a classical crossover effect. Subjects are always fastest on a target that is the word's first syllable, and slowest on the other target. Hence self-monitoring is sensitive to syllable structure. This indicates that it is the phonological word level that is being monitored, in agreement with Jackendoff's (1987) suggestion. The remaining question is whether the steady increase in RT that appears from the Figure 16 results is due to incremental creation of the phonological word, as discussed in section 6.4.4, or rather to the "left-to-right" nature of the monitoring process that scans a whole, already existing representation. We cannot tell from the data but prefer the former solution. In that case the latencies in Figure 16 tell us something about the speed of phonological word construction in "internal speech." The RT difference, for instance, between the onset and the offset of the first CVC syllable was 55 msec, and between the two syllable onset consonants it was 111 msec. That would mean that a syllable's internal phonological encoding takes less than half the time of its articulatory execution, because for the same words in overt articulation we measured a delay between the two onset consonants of 210 msec on average. This agrees nicely with the LRP findings by van Turennout et al., discussed in section 5.4.4. Their task was also a self-monitoring task, and they found an 80 msec LRP effect difference between monitoring for a word's onset and its offset, just about 50% more than the 55 msec mentioned above. Their experimental targets were, on average, 1.5 syllables long, that is, 50% longer than the present ones. This would mean, then, that the upper limits on speech rate are not set by phonological encoding, but by the "inertia" of overt articulation. This agrees with findings in the speech perception literature, where phonological *decoding* still functions well at triple to quadruple rates in listening to compressed speech (Mehler et al. 1993). These are, however, matters for future research.

10. Speech errors

A final issue we promised to address is speech errors. As was mentioned at the outset, our theory is primarily based on latency data, most of them obtained in naming experiments of one kind or another. However, traditionally, models of lexical access were largely based on the analysis of speech error data. Ultimately, these approaches should converge. Although speech errors have never been our main target of explanation, the theory seems to be on speaking terms with some of the major observations in the error literature. To argue this, we once more turn to WEAVER++. Below (see also Roelofs 1997c; Roelofs & Meyer 1998), we will show that the model is compatible with key findings such as the relative frequencies of segmental substitution errors (e.g., the anticipation error *sed sock* for *red sock* is more likely than the perseveration *red rock,* which is in its turn more likely than the exchange *sed rock)*, effects of speech rate on error probabilities (e.g., more errors at higher speech rates), the phonological facilitation of semantic substitution errors (e.g., *rat f*or target *cat* is more likely than *dog* for target *cat)*, and lexical bias (i.e., errors tend to be real words rather than nonwords).

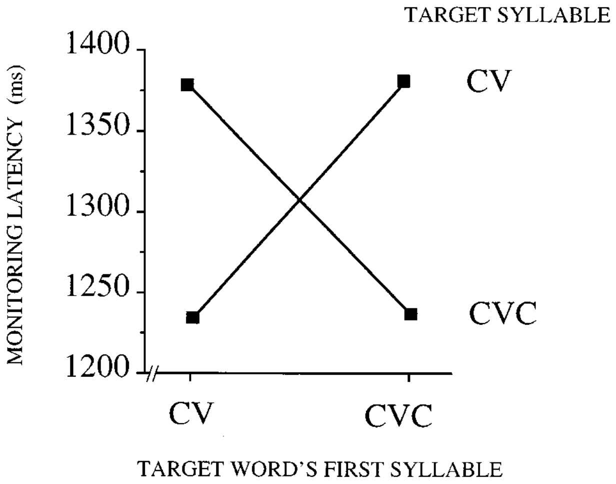


Figure 17. The syllable effect in self-monitoring. CV and CVC monitoring in target words that contain a CV or a CVC first syllable.

In its native state, WEAVER++ does not make errors at all. Its essential feature of "binding-by-checking" (see sect. 3.2.3) will prevent any production of errors. But precisely this feature invites a natural way of modeling speech errors. It is to allow for occasional binding failures, that is, somewhat reminiscent of Shattuck-Hufnagel's (1979) "check off" failures. In particular, many errors can be explained by indexing failures in accessing the syllable nodes. For example, in the planning of *red sock,* the selection procedure of [sɛd] might find its selection conditions satisfied. It requires an onset /s/, a nucleus /ɛ/, and a coda /d/, which are present in the phonological representation. The error is of course that the /s/ is in the wrong phonological syllable. If the procedure of [rɛd] does its job well, there will be a race between [rɛd] and [sɛd] to become the first syllable in the articulatory program for the utterance. If [sɛd] wins the race, the speaker will make an anticipation error. If this indexing error occurs, instead, for the second syllable, a perseveration error will be made, and if the error is made both for the first syllable and the second one, an exchange error will be made. Errors may also occur when WEAVER++ skips verification to gain speed in order to obtain a higher speech rate. Thus, more errors are to be expected at high speech rates.

Figure 18 gives some simulation results concerning segmental anticipations, perseverations, and exchanges. The real data are from the Dutch error corpus of Nooteboom (1969). As can be seen, WEAVER++ captures some of the basic findings about the relative frequency of these types of substitution errors in spontaneous speech. The anticipation error *sed sock* for *red sock* is more likely than the perseveration *red rock,* which is in turn more likely than the exchange *sed rock.* The model predicts almost no exchanges, which is, of course, a weakness. In the simulations, the verification failures for the two error locations were assumed to be independent, but this is not a necessary assumption of WEAVER++'s approach to errors. An anticipatory failure may increase the likelihood of a perseveratory failure, such that the absolute number of exchanges increases.

Lexical bias has traditionally been taken as an argument for backward form → lemma links in a lexical network, but such backward links are absent in WEAVER++. Segmental errors tend to create words rather than nonwords. For example, in producing *cat,* the error /h/ for /k/, producing the word *hat,* is more likely than /j/ for /k/, producing the nonword *yat.* In a model with backward links, this bias is due to feedback from shared segment nodes to morpheme nodes (e.g., from /æ/ and /t/ to <cat> and <hat>) and from these morpheme nodes to other segment nodes (i.e., from <cat> to /k/ and from <hat> to /h/). This will not occur for nonwords, because there are no morpheme nodes for nonwords (i.e., there is no node <yat> to activate /j/). Typically, errors are assumed to occur when, owing to noise in the system, a node other than the target node is the most highly activated one and is erroneously selected. Because of the feedback, /h/ will have a higher level of activation than /j/, and it is more likely to be involved in a segment selection error. Reverberation of activation in

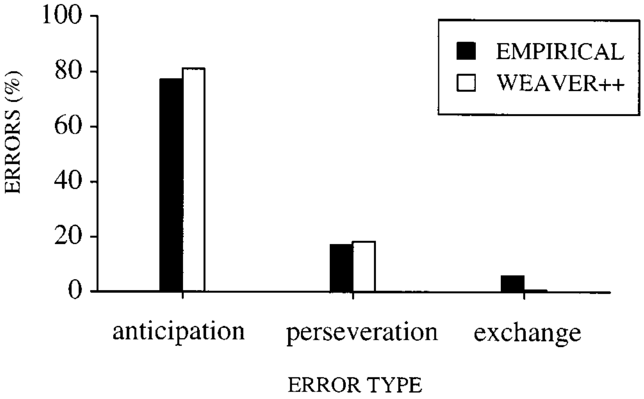


Figure 18. Frequency distribution of anticipation, perseveration, and exchange errors. Data from Nooteboom (1969) and WEAVER++ simulations.

the network requires time, so lexical influences on errors require time to develop, as is empirically observed (Dell 1986).

The classical account of lexical bias, however, meets with a difficulty. In this view, lexical bias is an automatic effect. The seminal study of Baars et al. (1975), however, showed that lexical bias is not a necessary effect. When all the target and filler items in an error-elicitation experiment are nonwords, there is no lexical bias. Only when some real words are included as filler items does the lexical bias appear. The account of Baars et al. of lexical bias was in terms of speech monitoring by speakers. Just before articulation, speakers monitor their internal speech for errors. If an experimental task deals exclusively with nonwords, speakers do not bother to attend to the lexical status of their phonetic plan. Levelt (1983) proposed that the monitoring may be achieved by feeding the phonetic plan to the speech comprehension system (see also sect. 9). On this account, there is no direct feedback in the output form lexicon, only indirect feedback via the speech comprehension system. Feedback via the comprehension system takes time, so lexical influences on errors take time to develop.

Similarly, the phonological facilitation of semantic substitutions may be a monitoring effect. The substitution of *rat* for *cat* is more likely than that of *dog* for *cat.* Semantic substitution errors are taken to be failures in lemma node selection. The word *rat* shares segments with the target *cat.* Thus, in a model with backward links, the lemma node of *rat* receives feedback from these shared segments (i.e., /æ/, /t/), whereas the lemma node of *dog* does not. Consequently, the lemma node of *rat* will have a higher level of activation than the lemma node of *dog,* and it is more likely to be involved in a lemma selection error (Dell & Reich 1980). In our theory, the semantic bias may be a monitoring effect. The target *cat* and the error *rat* are perceptually closer than the target *cat* and the error *dog.* Consequently, it is more likely that *rat* will pass the monitor than that *dog* will.

Another potential error source exists within a forward model such as WEAVER++. Occasionally, the lemma retriever may erroneously select two lemmas instead of one, the target and an intruder. This assumption is independently motivated by the occurrence of blends such as *clear* combining *close* and *near* (Roelofs 1992a) and by the experimental results of Peterson and Savoy (1998) and Jescheniak and Schriefers (1998) discussed in section 6.1.1. In WEAVER++, the selection of two lemmas instead of one will lead to the parallel encoding of two word forms instead of one. The encoding time is a random variable, whereby the word form that is ready first will control articulation. In the model, it is more likely that the intruder wins the form race when there is phonological overlap between target and intruder than when there is no phonological relation (i.e., when the form of the target primes the intruder). Thus WEAVER++ predicts that the substitution *rat* for *cat* is more likely than *dog* for *cat,* which is the phonological facilitation of semantic substitution errors. The selection of two lemmas also explains the syntactic category constraint on substitution errors. As in word exchanges, in substitution errors the target and the intruder are typically of the same syntactic category.

Although these simulations guide our expectation that speech error-based and reaction time-based theorizing will ultimately converge, much work is still to be done. A major issue, for instance, is the word onset bias in phonological errors (discussed in sect. 6.2.3). There is still no adequate account for this effect in either theoretical framework. Another issue is what we will coin "Dell's law" (Dell et al. 1997a), which says that with increasing error rate (regardless of its cause) the rate of anticipations to perseverations decreases. In its present state, our model takes no account of that law.

11. Prospects for brain imaging

Nothing is more useful for cognitive brain imaging, that is, relating functional processing components to anatomically distinct brain structures, than a detailed processing model of the experimental task at hand. The present theory provides such a tool and has in fact been used in imaging studies (Caramazza 1996; Damasio et al. 1996; Indefrey & Levelt, 1998; McGuire et al. 1996). A detailed timing model of lexical access can, in particular, inspire the use of high-temporal-resolution imaging methods such as ERP and MEG. Here are three possibilities.

First, using ERP methods, one can study the temporal successiveness of stages, as well as potentially the time windows within stages, by analyzing readiness potentials in the preparation of a naming response. This approach (Van Turennout et al. 1997; 1998) was discussed in sections 5.4.4 and 9.

Second, one can relate the temporal stratification of the stage model to the spatiotemporal course of cortical activation during lexical encoding. Levelt et al. (1998) did so in an MEG study of picture naming. Based on a meta-analysis of our own experimental data, other crucial data in the literature (such as those from Potter 1983; Thorpe et al. 1996), and parameter estimates from our own model, we estimated the time windows for the successive stages of visual-to-concept mapping, lexical selection, phonological encoding, and phonetic encoding. These windows were then related to the peak activity of dipole sources in the individual magnetic response patterns of the eight subjects in the experiment. All sources peaking during the first time window (visual-to-concept mapping) were located in the occipital lobes. The dipole sources with peak activity in the time window of lemma selection were largely located in the occipital and parietal areas. Left hemispherical sources peaking in the time window of phonological encoding showed remarkable clustering in Wernicke's area, whereas the right hemispheric sources were quite scattered over parietal and temporal areas. Sources peaking during the temporal window of phonetic encoding, finally, were also quite scattered over both perisylvian and rolandic areas, but with the largest concentration in the sensory-motor cortex (in particular, the vicinity of the face area). Jacobs and Carr (1995) suggested that anatomic decomposability is supportive for models with functionally isolable subsystems. Our admittedly preliminary findings support the distinctness of initial visual/conceptual processing (occipital), of phonological encoding (Wernicke's area), and of phonetic encoding (sensory/motor area). Still, this type of analysis also has serious drawbacks. One is, as Jacobs and Carr (1995) correctly remark, that most models make predictions about the total time for a system to reach the end state, the overt response time, but not about the temporal dynamics of the intermediate processing stages. Another is that stage-to-brain activation linkage breaks down when stages are not strictly successive. That is, for instance, true for the operations of self-monitoring in our theory. As was discussed in section 9, self-monitoring can be initiated during phonological encoding and it can certainly overlap with phonetic encoding. Hence, we cannot decide whether a dipole source whose activation is peaking during the stage of phonetic encoding is functionally involved in phonetic encoding or in self-monitoring. The more parallel a process model, the more serious this latter drawback.

Third, such drawbacks can (in principle) be circumvented by using the processing model in still another way. Levelt et al. (1998) called this the "single factors method." Whether or not a functionally decomposed processing model is serial, one will usually succeed in isolating an independent variable that affects the timing of one processing component but of none of the others. An example for our own theory is the word frequency variable, which (in a well-designed experiment) affects solely the duration of morphophonological encoding (as discussed in sect. 6.1.3). Any concomitant variation in the spatiotemporal course of cerebral activation must then be due to the functioning of that one processing component. It is theoretically irrelevant for this approach whether the processing components function serially or in parallel, as long as they function independently. But interactiveness in a processing model will also undermine this third approach, because no "single-component variables" can be defined for such models.

12. Conclusions

The purpose of this target article was to give a comprehensive overview of the theory of lexical access in speech production which we have developed in recent years, together with many colleagues and students. We discussed three aspects of this work. The first is the theory itself, which considers the generation of words as a dual process, both in ontogenesis and in actual speech production. There is, on the one hand, a conceptually driven system whose purpose it is to select words ("lemmas") from the mental lexicon that appropriately express the speaker's intention. There is, on the other hand, a system that prepares the articulatory gestures for these selected words in their utterance contexts. There is also a somewhat fragile link between these systems. Each of these systems is itself staged. Hence, the theory views speech as a feedforward, staged process, ranging from conceptual preparation to the initiation of articulation. The second aspect is the computational model WEAVER++, developed by one of us, Ardi Roelofs. It covers the stages from lexical selection to phonological encoding, including access to the mental syllabary. This model incorporates the feedforward nature of the theory but has many important additional features, among them a binding-by-checking property, which differs from the current binding-by-timing architectures. In contrast to other existing models of lexical access, its primary empirical domain is normal word production latencies. The third aspect is the experimental support for theory and model. Over the years, it has covered all stages from conceptual preparation to self-monitoring, with the exception of articulation. If articulation *had* been included, the more appropriate heading for the theory would have been "lexical generation in speech production." Given the current state of the theory, however, "lexical access" is still the more appropriate term. Most experimental effort was spent on the core stages of lexical selection and morphophonological encoding, that is, precisely those covered by the computational model, but recent brain imaging work suggests that the theory has a new, and we believe unique, potential to approach the cerebral architecture of speech production by means of high-temporal-resolution imaging.

Finally, what we do not claim is completeness for theory or computational model. Both the theory and the modeling have been in a permanent state of flux for as long as we have been developing them. The only realistic prediction is that this state of flux will continue in the years to come. One much needed extension of the theory is the inclusion of different kinds of languages. Details of lexical access, in particular those concerning morphological and phonological encoding, will certainly differ between languages in interesting ways. Still, we would expect the range of variation to be limited and within the general stratification of the system as presented here. Only a concerted effort to study real-time aspects of word production in different languages can lead to significant advances in our understanding of the process and its neurological implementation.

APPENDIX

We summarize here the mathematical characteristics of WEAVER++. The equations for the spreading of activation and the selection ratio are as follows (see Roelofs 1992a; 1993; 1994; 1996b; 1997c). Activation spreads according to

$$a(k,t+\Delta t) = a(k,t)(1-d) + \sum_n r\,a(n,t),$$

where $a(k,t)$ is the activation level of node k at point in time t, d is a decay rate ($0 < d < 1$), and Δt is the duration of a time step (in msec). The rightmost term denotes the amount of activation that k receives between t and $t + \Delta t$, where $a(n,t)$ is the output of node n directly connected to k (the output of n is equal to its level of activation). The factor r indicates the spreading rate.

The probability that a target node m will be selected at $t < T \leq t + \Delta t$ given that it has not been selected at $T \leq t$, and provided that the selection conditions for a node are met, is given by the ratio

$$\frac{a(m,t)}{\sum_i a(i,t)}.$$

For lemma retrieval, the index i ranges over the lemma nodes in the network. The selection ratio equals the hazard rate $h_m(s)$ of the retrieval of lemma m at time step s, where $t = (s-1)\Delta t$, and $s = 1, 2. \ldots$ The expected latency of lemma retrieval, $E(T)$, is

$$E(T) = \sum_{s=1}^{\infty} h_m(s)\{\prod_{j=0}^{s-1}[1-h_m(j)]\}s\Delta t.$$

For word form encoding, the index i in the selection ratio ranges over the syllable program nodes in the network. The selection ratio then equals the hazard rate $h_m(s)$ of the process of the encoding of syllable m (up to the access of the syllabary) at time step s. The equation expressing the expected latency of word form encoding for monosyllables is the same as that for lemma retrieval. In encoding the form of a disyllabic word, there are two target syllable program nodes, syllable 1 and syllable 2. The probability *p(word form encoding completes at s)* for a disyllabic word equals

$$\begin{aligned}&\left[h_1(s)V_1(s-1)\right]\sum_{j=0}^{s-1}\left[h_2(j)V_2(j-1)\right]+\\&\left[h_2(s)V_2(s-1)\right]\sum_{j=0}^{s-1}\left[h_1(j)V_1(j-1)\right]+\\&\left[h_1(s)V_1(s-1)\right]\left[h_2(s)V_2(s-1)\right]\\&= f_1(s)\sum_{j=0}^{s-1} f_2(s) + f_2(s)\sum_{j=0}^{s-1} f_1(s) + f_1(s)f_2(s),\end{aligned}$$

where $h_1(s)$ and $h_2(s)$ are the hazard rates of the encoding of syllable 1 and 2, respectively, $V_1(s)$ and $V_2(s)$ the corresponding cumulative survivor functions, and $f_1(s)$ and $f_2(s)$ the probability mass functions. For the expectation of T holds

$$E(T) = \sum_{s=1}^{\infty} \{f_1(s) \sum_{j=0}^{s-1} f_2(s) + f_2(s) \sum_{j=0}^{s-1} f_1(s) + f_1(s) f_2(s)\} s\Delta t.$$

The estimates for the parameters in these equations were as follows. The spreading rate r within the conceptual, lemma, and form strata was 0.0101, 0.0074, and 0.0120 msec^{-1}, respectively, and the overall decay rate d was 0.0240 msec^{-1}. The duration of basic events such as the time for the activation to cross a link, the latency of a verification procedure, and the syllabification time per syllable equalled $\Delta t = 25$ msec. For details of the simulations, we refer the reader to the original publications (Roelofs 1992a; 1993; 1994; 1996b; 1997c).

ACKNOWLEDGMENTS

We are grateful for critical and most helpful comments on our manuscript by Pat O'Seaghdha, Niels Schiller, Laura Walsh-Dickey, and five anonymous BBS reviewers.

NOTES

1. Kempen and Hoenkamp (1987; but already cited in Kempen & Hiybers 1983) introduced the term *lemma* to denote the word as a semantic/syntactic entity (as opposed to the term *lexeme,* which denotes the word's phonological features) and Levelt (1983; 1989) adopted this terminology. As the theory of lexical access developed, the term *lemma* acquired the more limited denotation used here, that is, the word's syntax (especially in Roelofs 1992a). This required an equally explicit denotation of the word's semantics. That role is now played by the technical term *lexical concept.* None of this, however, is a change of theory. In fact, even in our own writings we regularly use the term *lemma* in its original sense, in particular if the semantics/syntax distinction is not at issue. In the present paper, though, we will use the term *lemma* exclusively in its restricted, syntactic sense. Although we have occasionally used the term *lexeme* for the word *form,* this term has led to much confusion, because traditionally *lexeme* means a separate dictionary entry. Here we will follow the practice used by Levelt (1989) and speak of "morphemes" and their phonological properties.

2. By "intentional production" we mean that, for the speaker, the word's meaning has *relevance* for the speech act. This is often not the case in recitation, song, reading aloud, and so on.

3. The syntactic representation of *escort* in Figure 2 is, admittedly, quite simplified.

4. A *phonological* or *prosodic* word is the domain of syllabification. It can be smaller than a lexical word, as is the case in most compound words, or it can be larger, as is the case in cliticization (in *Peter gave it,* the syllabification *ga-vit* is over *gave it,* not over *gave a*nd *it* independently).

5. There are dialectal variations of the initial vowel; the *Collins English Dictionary,* for instance, gives /ɪ/ instead of /ə/. Stress shift will turn /ə/ into a full vowel.

6. It should be noted, though, that in the Browman/Goldstein theory (and different from ours) not only the word's phonetics are gestural but also all of the word's phonology. In other words, word form representations in the mental lexicon are gestural to start with. We are sympathetic to this view, given the signaled duality in the word production system. Across the "rift," the system's sole aim is to prepare the appropriate articulatory gestures for a word in its context. However, the stored form representations are likely to be too abstract to determine pronunciation. Stemberger (1991), for instance, provides evidence for phonological underspecification of retrieved word forms. Also, the same underlying word form will surface in rather drastically different ways, depending on the morphonological context (as in *period/periodic* or *divine/divinity),* a core issue in modern phonology. These and other phenomena (see sect. 6.2.2) require rather abstract underlying form representations. Gestural phonology is not yet sufficiently equipped to cope with these issues. Hence we will follow current phonological approaches, by distinguishing between phonological encoding involving all-or-none operations on discrete phonological codes and phonetic encoding involving more gradual, gestural representations.

7. The arcs in Figure 2, and also in Figures 6 and 7, represent the labeled activation routes between nodes. Checking involves the same labeled relations. Details of the checking procedures, though, are not always apparent from the figures. For instance, to check the appropriateness of <ing> in Figure 2, the procedure will test whether the lemma *escort* has the correct aspectual diacritic "prog."

8. Note that we are not considering here the flow of information between the visual network and the lemma nodes. Humphreys et al. (1988; 1995) have argued for a cascading architecture there. At that level our model is a cascading network as well, although the visual network itself is outside our theory.

9. This potential relation between time pressure and multiple selection was suggested to us by Schriefers. This is an empirical issue; it predicts that the subordinate term will not be phonologically activated in highly relaxed naming conditions.

10. This "taking over" is still unexplained in any model, but Peterson and Savoy make some useful suggestions to be further explored. The fact that one form tends to take over eventually is consonant with the observation that blends are exceedingly rare, even in the case of near-synonyms.

11. Humphreys et al. (1988) reported a picture naming study in which the effect of name frequency interacted with that of stuctural similarity. (Structural similarity is the degree to which members of a semantic category look alike. For instance, animals or vegetables are structurally similar categories, whereas categories such as tools or furniture are structurally dissimilar categories.) A significant word frequency effect was obtained only for the members of structurally dissimilar categories. Clearly, our serial stage model does not predict this interaction of a conceptual variable with name frequency. However, it is possible that in the materials used by Humphreys et al. word frequency was confounded with object familiarity, a variable likely to have visual and/or conceptual processes as its origin and which we would expect to interact with structural similarity. Moreover, Snodgrass and Yuditsky (1996), who tested a much larger set of pictures, failed to replicate the interaction reported by Humphreys et al.

12. CELEX, the Center for Lexical Information, based at the Nijmegen Max Planck Institute, develops and provides lexical-statistical information on English, German, and Dutch. The databases are available on CD-ROM: Baayen, R. H., Piepenbrock, R., & van Rijn, H. (1993) *The CELEX lexical database.* Philadelphia: Linguistic Data Consortium, University of Pennsylvania.

13. We thank Gary Dell for pointing this out to us.

14. The computer simulations of the word-form encoding experiments were run using both small and larger networks. The small network included the morpheme, segment, and syllable program nodes of the words minimally needed to simulate the conditions of the experiments. For example, the small network in the simulations of the experiments on picture naming with spoken distractor words of Meyer and Schriefers (1991) comprised 12 words. Figure 2 illustrates the structure of the forms in the network for the word "escort." To examine whether the size and the scope of the network influenced the outcomes, the simulations were run using larger networks. These networks contained the words from the small network plus either (1) the forms of the 50 nouns with the highest frequency in the Dutch part of the CELEX lexical database (see note 12) or (2) the forms of 50 nouns randomly selected from CELEX. The outcomes for the small and the two larger networks were the same (Roelofs 1997c). The simulations of word-form encoding were run using a single set of nine parameters, but three of these parameters were fixed to the values set in the lemma retrieval simulations depicted in Figure 4. They were the three parameters shared between the simulations: decay rate, smallest time interval, and size of the distractor input to the network. Hence there were six free parameters; they included values for the general spreading rate at the form stratum, the size of the time interval during which spoken distractors provided input to the network, and a selection threshold. Their values were held fixed across simulations. The parameter values were obtained by optimizing the fit of the model to a restricted number of data sets from the literature. Other known data sets were subsequently used to test the model with these parameter values. The data sets used to obtain the parameters concerned the SOA curves of facilitation and inhibition effects of form-based priming in picture naming that were obtained by Meyer and Schriefers (1991). The data sets of Meyer and Schriefers comprised 48 data points, which were simultaneously fit by WEAVER++ with only six free parameters. Thus WEAVER++ has significantly fewer degrees of freedom than the data contain, that is, the fit of the model to the data is not trivial. WEAVER++ could have been falsified by the data. After fitting the model to the data sets of Meyer and Schriefers (1991), the model was tested on other data sets known from the literature (e.g., Meyer 1990) and in new experiments that were specifically designed to test nontrivial predictions of the model. Figures 12–14 present some of these new data sets together with the predictions of the model. The parameter values in these tests were identical to those in the fit of the model to the data of Meyer and Schriefers.

15. The terms "explicit" versus "implicit" priming are purely technical terms here, referring to the experimental method used. A priming technique is called "explicit" if an auditory or visual distractor stimulus is presented at some stage during the speaker's generation of the target word. A priming technique is called "implicit" if the target words in an experimental set share some linguistic property (such as their first syllable, their last morpheme, or their accent structure); that property is called the "implicit prime." The terms do not denote anything beyond this, such as whether the subject is conscious or unconscious of the prime.

16. These data are derived, though, from a database of written, not spoken, text. There is good reason for taking them seriously nevertheless. Schiller et al. (1996) find that, if such a text base is "resyllabified" by applying rules of connected speech, there is basically no change in the frequency distribution of the high-ranking syllables.

17. We are considering here only representations that could be subject to monitoring at the word form (i.e., output) level. This does not exclude the possibility of "input" monitoring. Levelt (1989), for instance, argues that there is also self-monitoring at the conceptual level of message formation.

Open Peer Commentary

Commentary submitted by the qualified professional readership of this journal will be considered for publication in a later issue as Continuing Commentary on this article. Integrative overviews and syntheses are especially encouraged.

Grossberg and colleagues solved the hyperonym problem over a decade ago

Jeffrey S. Bowers
Department of Experimental Psychology, University of Bristol, Bristol BS8 1TN, England. **j.bowers@bris.ac.uk**

Abstract: Levelt et al. describe a model of speech production in which lemma access is achieved via input from nondecompositional conceptual representations. They claim that existing decompositional theories are unable to account for lexical retrieval because of the so-called hyperonym problem. However, existing decompositional models have solved a formally equivalent problem.

An important theoretical claim by Levelt et al. is that conceptual knowledge is coded in an nondecompositional format; that is, concepts for the various morphemes in a language are primitive (sect. 3.1.1). They claim that existing decompositional accounts have been unable to solve the so-called hyperonym problem, which Levelt (1992, p. 6) defined as follows:

> When lemma A's meaning entails lemma B's meaning, B is a hyperonym of A. If A's conceptual conditions are met, then B's are necessarily also satisfied. Hence, if A is the correct lemma, B will (also) be retrieved.

For example, when a speaker wants to express the concept CAT, all the conceptual conditions for retrieving the lemma for ANIMAL are satisfied as well, since the meaning of cat entails the meaning of animal. Thus, both lemmas would be retrieved, contrary to what in fact occurs.

In this commentary, I simply want to point out that one well-developed network model has in fact solved a formally equivalent problem maintaining a decompositional approach; namely, the *adaptive resonance theory* (ART) of Grossberg and colleagues (e.g., Carpenter & Grossberg 1987; Grossberg 1976b).

The Carpenter and Grossberg solution. Carpenter and Grossberg (1987) addressed the question as to how a network can correctly categorize subset and superset visual patterns. For illustrative purposes, consider the case of visual word recognition in which the words *my* and *myself* are presented. The problem is to develop a model that can correctly categorize both patterns, because as the authors note, the superset pattern *myself* contains all the features of the subset *my,* so that the presentation of *myself* might be expected to lead to the full activation of the lexical orthographic codes of *my* as well as *myself.* Similarly, if the features of *my* are sufficient to access the subset pattern, it might be expected that the superset pattern *myself* would access the subset pattern as well. Thus, how does the network decide between the two inputs? The hyperonym problem revisited.

Their solution was embedded within an ART network that contains two fields of nodes, input and output layers, called F1 and F2, respectively. The nodes in F1 each refer to a feature (a letter in the above example), so that the nodes that become active in response to the input *my* are a subset of the active nodes in response to *myself.* The active nodes generate excitatory signals along pathways to the target nodes in F2, which are modified by the long term memory traces (LTMs) that connect F1 and F2. Each target node in F2 sums up all of the incoming signals, and transforms this pattern of activation based on the in-

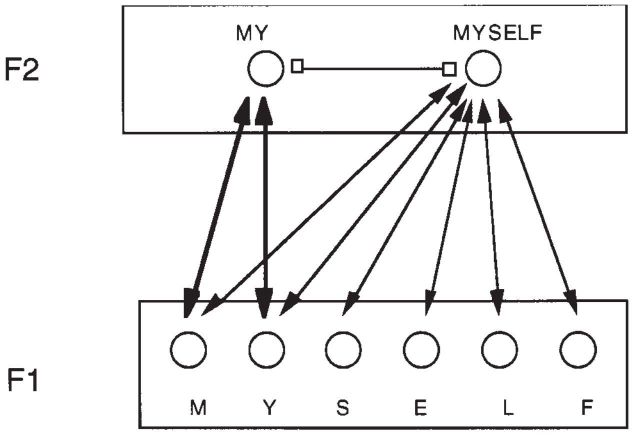


Figure 1 (Bowers). Direct connections between levels F1 and F2 of ART network. The solid arrows indicate excitatory connections, and the squares indicates inhibitory connections.

teractions among the nodes of F2, resulting in a single active node at F2 (see Fig. 1).

For present purposes, consider the case in which v1 and v2 refer to the two nodes in F2 that code for the subset and superset patterns in F1, respectively. Thus, the subset pattern at F1 must selectively activate v1, and the superset pattern at F1 must selectively activate v2. In order to achieve this result, Carpenter and Grossberg (1987) incorporated learning rules that followed the *Weber* and the *Associative Decay* principles. Very briefly, according to the Weber rule, there is an inverse relationship between LTM strength and input pattern scale, so that the LTM traces connecting the subset pattern at F1 to v1 are stronger than the LTM traces connecting this same subset pattern to v2 (otherwise, the superset could activate v1). And according to the Associative Decay rule, LTM weights decay towards 0 during learning when nodes at F1 and F2 are not coactive. In particular, LTM weights decay to 0 between inactive nodes in F1 that are part of the superset pattern to v1. Together, these rules accomplish the goal: since the superset pattern includes more active nodes than the subset, the Weber Law Rule insures that the LTM traces in the pathways to v2 do not grow as large as the LTM traces to v1. On the other hand, after learning occurs, more positive LTM traces project to v2 from F1, which combine to produce larger activation to v2 than to v1 when the superset is presented. Thus, there is a trade-off between the individual sizes of the LTM traces and the number of traces, which allows direct access to both subset and superset representations.

Carpenter and Grossberg's proof that the Weber and Associative Decay Rules achieve subset and superset access. As learning of an input pattern takes place, the bottom-up LTM traces that join F1 and F2 approach an asymptote of the form:

$$\partial/(\beta + |I|) \tag{1}$$

where ∂ and β are positive constants, and $|I|$ equals the number of nodes active in a pattern I in F1. From (1), larger $|I|$ values imply smaller positive LTM traces in the pathways encoding I.

Direct access to the subset and superset patterns can now be understood as follows. By (1), the positive LTM traces that code the subset pattern have the size

$$\partial/(\beta + |\text{subset}|) \tag{2}$$

and the positive LTM traces that code the superset have the size

$$\partial/(\beta + |\text{superset}|) \tag{3}$$

When the *subset* is presented at F1, |subset| nodes in F1 are active. The total input to v1 (T_{11}) is proportional to

$$T_{11} = \partial|\text{subset}|/(\beta + |\text{subset}|) \tag{4}$$

And the total input to v2 (T_{12}) is proportional to

$$T_{12} = \partial|\text{subset}|/(\beta + |\text{superset}|) \tag{5}$$

Because (1) defines a decreasing function of $|I|$ and because |subset| < |superset|, it follows that $T_{11} > T_{12}$. Thus, the subset pattern activates v1 instead of v2.

When the *superset* is presented to F1, |superset| nodes are active. Thus the total input to v2 (T_{22}) is proportional to

$$T_{22} = \partial|\text{superset}|/(\beta + |\text{superset}|) \tag{6}$$

Now, the Associative Decay Rule is critical, because only those F1 nodes in the superset that are also activated by the subset project LTM traces to v1. Thus, the total input to v1 is proportional to

$$T_{21} = \partial|\text{subset}|/(\beta + |\text{subset}|). \tag{7}$$

Both T22 and T21 are expressed in terms of the Weber function

$$W(|I|) = \partial|I|/(\beta + |I|), \tag{8}$$

which is an *increasing* function of $|I|$. Since |subset| is smaller than |superset|, $T_{22} > T_{21}$. Thus, the superset activates v2 rather than v1. In summary, direct access to subsets and supersets can be traced to the opposite monotonic behavior of the functions (1) and (8).

In conclusion, the hyperonym problem can be solved within a decompositional framework. Whether or not a general theory of lexical retrieval can be accomplished using this framework, however, remains to be seen.

How does WEAVER pay attention?

Thomas H. Carr

Department of Psychology, Michigan State University, East Lansing, MI 48824-1117. **carrt@pilot.msu.edu** **psychology.msu.edu/carr.htm**

Abstract: Though WEAVER has knowledge that gets activated by words and pictures, it is incapable of responding appropriately to these words and pictures as task demands are varied. This is because it has a most severe case of attention deficit disorder. Indeed, it has no attention at all. I discuss the very complex attention demands of the tasks given to WEAVER.

In their fourth figure, Levelt, Roelofs & Meyer show data from six different primed picture and word processing tasks. In all tasks participants see a picture of an object with the printed name of a different object superimposed. The two stimuli appear either simultaneously (reproducing the Stroop-like "picture–word interference task"), or staggered in time so that the picture appears a bit before the word or vice versa. Four tasks require naming the picture, one requires naming the picture's superordinate category, and one requires naming the word's superordinate category.

In all these tasks, the critical manipulation is semantic relatedness between picture and word. The resulting priming effects are a major object of theoretical explanation for Levelt et al.'s model of speech production, WEAVER++. WEAVER's performance is presented in Figure 4 together with the priming effects produced by the participants. Though error bars are absent, the discrepancies appear small and the correspondence between model and human being is obviously close. Levelt et al. tout WEAVER's success.

The correspondence is impressive. But there is a problem. It is literally impossible for WEAVER to have produced the data in Figure 4. This is because WEAVER has a severe case of attention deficit disorder (ADD). Indeed, it has no attention at all. A simple dose of Ritalin won't help – we must implant an entire attention system. Without it, there is no way for WEAVER to adapt to the varying demands of the six tasks shown in Figure 4: sometimes responding to the picture, sometimes to the word, sometimes saying a basic-level name, sometimes a category name. In its current architectural incarnation, the response produced by the model will be the first word form that gets fully activated, given the picture and word presented and the timing of their appearance. Period. Without a *deus ex machina* intervention from outside the model itself, this will happen regardless of whether WEAVER is *supposed* to act only on the picture or only on the word and regardless of whether WEAVER is *supposed* to say a basic-level name or a superordinate-level name. Thus if WEAVER does the right thing, it is by the lucky chance that what it was going to do anyway happened to match the task demands in effect at that particular instant (as so often seems the case with someone who has severe ADD!).

Perhaps this is unfair criticism. Levelt et al. are psycholinguists, not attention theorists. The conventions of our business give them the right to take for granted the processes of selective attention that they currently enact themselves by deciding which activation values they are going to extract from the representational nodes in the model, track over time, and turn into response latencies and priming effects. But if we are to make progress toward integrated theories of human performance, rather than remain at the level of theories of the component processes from which that performance is somehow put together, we must build selection and control operations into the model. We must address questions that require psycholinguists to become attention theorists, attention theorists to become psycholinguists, and so forth. In what follows, I want simply to say a few words about the varieties of attention that might ultimately need to be implanted in order to fix WEAVER's attention deficit disorder.

To a student of attention, the universe of tasks examined by Levelt et al. is like Disneyland to an 8-year-old. Things begin simply enough. Meeting Levelt et al.'s task demands requires selective attention to one stimulus rather than the other. The picture and the word both provide visual information, but only one of these sources of information will support a correct response. How can this selection be accomplished?

It could be spatial. Picture and word do not occupy exactly the same points in space, so implanting a "posterior attention system" (Posner & Raichle 1994; Rafal & Henik 1994) that selects inputs from particular spatial locations might do the job. However, simply picking a location and increasing the gain on all information coming from that location will not be sufficient, especially when it is the picture that must be selected. Because the word is superimposed on the picture, selecting the region of space in which the picture can be found will get the word, too – just what the participant needs to avoid.

Perhaps the problem could be solved by *not* being spatial. Under some circumstances, the visual system appears to select *stimuli* rather than points or regions of space. This is called "object-based attention" (e.g., Duncan 1984; Kanwisher & Driver 1992; Valdes-Sosa et al. 1998), and there is debate about the conditions under which visual selection will be object-based rather than spatial. However, much of the evidence for object-based attention involves facilitated responding to secondary stimuli that happen to appear inside the boundaries of the stimulus that has been selected – as if selecting a stimulus fills its boundaries with spatial attention, facilitating *all* information from the resulting spatial region. Thus this type of selection will also fail to meet the task demands imposed by Levelt et al., at least when the stimulus that must govern responding is the picture. If selecting the picture fills it with spatial attention, then the word will be selected, too.

Therefore we must consider attentional processes that select on something more like the ontological status of the visual entities available to perception – what *kinds* of objects are they? – rather than their spatial locations or physical boundaries. Though we know that people can do this, we don't know much about how they do it. Neuroimaging evidence on the color–word Stroop task suggests that midline frontal regions centered in the anterior cingulate cortex might play an important role in this type of selection, perhaps interacting with various regions in lateral prefrontal cortex, each of which "knows" how to interact with more specialized posterior perceptual and memorial structures that handle a particular class of information. This network of interacting structures centered on cingulate cortex has been called the "anterior attention system" (Posner & Raichle 1994) and has been hypothesized to be the "executive control" component of "working memory" (e.g., Carr 1992; Carr & Posner 1995).

In addition to spatial, object-based, and ontological considerations, selecting the right stimulus for Levelt et al.'s tasks also has a temporal aspect. A large literature on sequential interference effects such as the "psychological refractory period" and the "attentional blink" indicates that when two stimuli appear in close temporal succession, as in Levelt et al.'s tasks, one or the other or both suffer interference and are processed more slowly (and are more likely to be missed entirely, if presentation is brief) than if they had been separated by a longer interval. Such interference is quite common if the first stimulus requires an overt response (this is the "Psychological Refractory Period" or PRP; see, e.g., Pashler 1994), and it can also occur if the first stimulus requires attention and decision or memory storage even if it does not require an immediate response (this is the "attentional blink" or AB; see, e.g., Arnell & Jolicoeur 1997; Chun & Potter 1995; Shapiro & Raymond 1994). But so what? If the interference simply adds main-effect increments to all target response latencies, then no harm is done to Levelt et al.'s arguments about priming effects, which are differences among these latencies that would remain unchanged. However, there is evidence that the AB interacts with priming effects, at least when stimulus presentation is brief and correct responding is the dependent measure. Related items suffer less interference than unrelated items (Maki et al. 1997). Attention-based interference between stimuli occurring in close temporal succession may be something Levelt et al. can't ignore.

Finally, meeting Levelt et al.'s task demands requires yet another type of attention. Once the right stream of information processing has been selected – the one representing the picture or the one representing the word – participants must select the right code or level of processing within that stream from which to construct their response. Sometimes this is the basic level, when the task is object naming, and sometimes it is the superordinate level, when the task is categorization. We are now clearly dealing with executive control and memory retrieval operations commonly attributed to "working memory" and the "anterior attention system," rather than to input selection operations of the "posterior attention system" (Rafal & Henik 1994). Added complications arise when the word to which these representations correspond is newly learned and hence relatively unfamiliar and weakly established in conceptual and lexical memory. It appears that a special class of inhibitory retrieval operations are brought to bear when such a weak code must be retrieved to meet task demands, and the consequences of its operation for related codes in conceptual and lexical memory is inhibitory. Dagenbach and I (1994) have conceived of this inhibition as arising from a center-surround attentional mechanism that centers on the weak code and suppresses stronger competitors in nearby semantic space. This contrasts with the facilitative effects of semantic relatedness observed with better-learned materials that gets translated into automatic spreading activation in WEAVER. Thus taking account of differences in degree of learning among the lexical items being processed may add a final layer of complexity to the task of fixing WEAVER's attention deficit.

Sharpening Ockham's razor

Anne Cutler[a] and Dennis Norris[b]
[a]*Max-Planck-Institute for Psycholinguistics, 6500 AH Nijmegen, The Netherlands.* **anne.cutler@mpi.nl** [b]*MRC Cognition and Brain Sciences Unit, Cambridge CB2 2EF, United Kingdom.*
dennis.norris@mrc-cbu.cam.ac.uk

Abstract: Language production and comprehension are intimately interrelated; and models of production and comprehension should, we argue, be constrained by common architectural guidelines. Levelt et al.'s target article adopts as guiding principle Ockham's razor: the best model of production is the simplest one. We recommend adoption of the same principle in comprehension, with consequent simplification of some well-known types of models.

Levelt, Roelofs & Meyer propose an account of lexical access, which forms part of the overall enterprise of Levelt and his colleagues – beginning with Levelt (1989) – to model the process of speaking. As clearly stated in the present paper and elsewhere, their enterprise is guided by the principle of Ockham's razor: Choose the simplest model that will explain the relevant data. Thus their theory currently proposes that flow of information between processing levels is unidirectional, that activation may be facilitatory but not inhibitory, and so on (sect. 3.2.5).

As comprehension researchers, we laud Levelt et al.'s methodological stringency, which we consider all too unusual, especially within a connectionist modelling framework. In this commentary we recommend that the same razor could be used to trim excess growths in the modelling of comprehension processes as well.

In research on language comprehension, connectionist models are now not at all scarce; McClelland and Rumelhart (1981; Rumelhart & McClelland 1982) changed the way of life for reading theoreticians, and McClelland and Elman (1986), with TRACE, changed it for researchers in spoken-language understanding. The models McClelland and his colleagues proposed were interactive activation networks with topdown connections between levels of processing, and this feature of the architecture was invoked to explain important experimental findings. In each case, however, it turned out that the topdown connections were *not* necessary to explain the observed effects. We will describe one case from each area: the word superiority effect in reading, and the apparent lexical mediation of compensation for coarticulation in spoken-language understanding.

The "word superiority effect" is the finding that letters can be more easily identified in words (e.g., BRAT) than in nonwords (e.g., TRAB). This was easily simulated in the interactive activation model of McClelland and Rumelhart (1981); feedback connections from the word level to the letter level in the model allowed activation to flow topdown and increase the availability of the target letter nodes. Norris (1990), however, built a simple network with no feedback connections and with separate, unconnected, output nodes for words and letters; these two sets of outputs were separately trained and, after training, the model identified letters in words better than letters in nonwords. Thus the availability of feedback connections was not necessary for the appearance of the word superiority effect.

Compensation for coarticulation is a shift in the category boundary for a particular phoneme distinction as a function of the preceding phonetic context. Elman and McClelland (1988) apparently induced such compensation from lexical information; the preceding phonetic context supplied in their experiment was in fact a constant token ambiguous between [s] and [S], but it occurred at the end of Christma* versus fooli*. Listeners' responses to the phoneme following this constant token were shifted in the same direction as would have been found with the really different phonemes at the end of Christmas and foolish. Subsequently, it has been shown that this result is dependent on listeners' knowledge of transitional probabilities rather than on individual lexical items (Pitt & McQueen 1998). However, Elman and McClelland simulated their result in TRACE and attributed it to TRACE's feedback connections between the lexical and the phoneme level. Norris (1993), though, managed to simulate these experimental findings in a network that did not make use of feedback connections during recognition. The simulation used a recurrent network with interconnected hidden units. In the critical case, the network was only ever trained to identify phonemes and did not learn to identify words. The network developed a bias towards identifying ambiguous phonetic tokens consistently with the words it had been trained on, and this bias exercised an effect on identification of following phonemes in just the way that the network had been trained to achieve with unambiguous tokens. Importantly, the availability of feedback connections was again not necessary for the appearance of the lexically sensitive compensation for coarticulation effect.

Thus the incorporation of topdown connections in these influential models of comprehension was in no way motivated by a need to explain empirical observations. The models' architects simply chose a more complex structure than was necessary to explain the data.

Further, as Norris et al. (submitted) have argued, topdown feedback in a model of comprehension is not even necessarily able to deliver on the promises, concerning a general improvement in recognition performance, in which its proponents seem to have trusted. Consider feedback from the lexical level to prelexical processing stages. The best word-recognition performance is achieved by selection of the best lexical match(es) to whatever prelexical representation has been computed. Adding feedback from the lexical level to the prelexical level does not improve the lexical level's performance: either the match selected is the best one or it is not. Feedback can certainly result in changed performance at prelexical levels. For instance, if the output of prelexical processing is a string of phonetic representations corresponding to "feed*ack" where the * represents some unclear portion, topdown activation from the lexicon might change the prelexical decision from uncertainty to certainty that there had been a [b]. But suppose that there had not in fact been a [b]? Suppose that the speaker had actually made a slip of the tongue and said feedpack, or feedfack? In that case, the topdown information flow would, strictly speaking, have led to poorer performance by the prelexical processor, since it would have caused a wrong decision to be made about the phonetic structure of the input. (In many listening situations this would of course matter very little, but it might be disastrous for a language production researcher who was interested in collecting slips of the tongue!)

Thus topdown connections can clear up ambiguity in prelexical processing, but they do so at a potential cost. more importantly, they do not result in an improvement of word recognition accuracy. Simulations with TRACE, for instance, have shown that the overall accuracy of the model is neither better nor worse if the topdown connections that the model normally contains are removed (Frauenfelder & Peeters 1998).

Ockham's razor clearly applies: models without topdown connections can explain the currently available comprehension data as well as the models with topdown connections, so in the absence of any superiority of performance to force a choice between the models, selection must be motivated on architectural grounds alone. We propose, with Levelt et al., that the simpler model should always be the first choice.

Levelt and his colleagues at present confine their modelling to the process of speaking. But in principle one would surely want to see models of comprehension and of production that were interdependent, as well as, of course, architecturally constrained in a similar fashion. The processes themselves are interdependent, after all – especially the ability to speak depends on the ability to perceive. Therefore it cannot in the long run be the case that the model of speaking is unconstrained by perceptual processes. The Grand Unified Theory of language perception and production is some time in the future. But we would argue that the success so far of the Levelt et al. enterprise, with its highly constrained and pared-down theory, suggests that Ockham's razor should be kept honed for general psycholinguistic use.

Binding, attention, and exchanges

Gary S. Dell,[a] Victor S. Ferreira,[b] and Kathryn Bock[a]
[a]*Beckman Institute and Department of Psychology, University of Illinois, Urbana, IL 61801.* [b]*Department of Psychology, University of California San Diego, La Jolla, CA 92093.* **gdell@s.psych.uiuc.edu**
kbock@s.psych.uiuc.edu **ferreira@psy.ucsd.edu**

Abstract: Levelt, Roelofs & Meyer present a comprehensive and sophisticated theory of lexical access in production, but we question its reliance on binding-by-checking as opposed to binding-by-timing and we discuss how the timing of retrieval events is a major factor in both correct and errorful production.

There's an old joke: "What is the most important ingredient in humor? TIMING!" (The word "timing" is said too quickly after "humor." Therein lies the joke.) One of us once heard this joke delivered as "What's the most important ingredient in timing? HUMOR!" This kind of slip – assuming that it *was* a slip rather than another very subtle joke – is an exchange. We think that exchanges and other speech errors occur because of variation in, well, timing. In general, we believe that the timing of retrieval events is a major factor in both correct and errorful production, and that this influence is more naturally expressed by "binding-by-timing" than "binding-by-checking" mechanisms. Levelt, Roelofs & Meyer hold a different view, and that is the subject of our commentary.

Levelt et al.'s is undoubtedly the most complete and sophisticated theory of lexical access in production. It offers a precise account of diverse data and situates that account in an original ontogenetic framework. We particularly like the emphasis on the natural rift between sound and meaning in the theory, and the way that the rift is associated with frequency effects, the tip-of-the-tongue phenomenon, and the grammatical features of words. Although we question Levelt et al.'s claim that there is no interaction across the rift, we agree that the arbitrary mapping between words' semantic (and syntactic) properties and their phonological forms has profound effects on the production system.

We disagree with Levelt et al.'s view of binding-by-timing, however. They are rightly impressed that people can name a picture accurately even when they hear a distractor word. (For example, in Cutting & Ferreira, in press, the distractor was erroneously spoken instead of the picture name a mere 14 out of 6,642 trials). Binding-by-checking insures that the system does not select the distractor instead of the picture's name. In binding-by-timing theories (e.g., Dell 1986) what is selected is determined by activation. Retrieval mechanisms must make sure that the right things are activated at the right time. Thus a binding-by-timing theory would have to claim that the distractors are activated enough to slow things down, but not activated enough to generate more than a handful of errors. This seems implausible to Levelt et al.

We think Levelt et al. are confusing the need for attentional control with a need for binding-by-checking over binding-by-timing. The picture-word interference experiments are a kind of Stroop task, where a response based on relevant information must be executed while the irrelevant information is ignored. The Stroop literature is vast, and the mechanisms proposed to account for the effect of distracting information on the target are numerous, but at least some of these mechanisms involve control of the activation of distracting information through attention (e.g., Cohen et al. 1990). Thus, in these theories, binding is ultimately expressed in terms of activation, rather than binding-by-checking. In fact, a model of picture-word interference without binding-by-checking can be specified so that the semantic distractor's activation causes a slow-down without inducing errors (Cutting & Ferreira, in press). So we are not persuaded that the low error rate in these tasks favors binding-by-checking over binding-by-timing.

The timing of retrieval events determines all sorts of things in production. According to Levelt et al., binding-by-checking is needed to account for why speakers don't say "kings escort pages" instead of "pages escort kings" when the timing of "kings" happens to be premature. However, much evidence suggests that timing is very influential in assignment of phrases to grammatical functions. Factors that cause a word to be retrieved more rapidly (e.g., semantic priming) also cause the corresponding phrase to be bound to an earlier sentence position (Bock 1982; 1986). In these cases, though, timing affects relative placement without causing errors because of syntactic flexibility (Ferreira 1996). Rather than the accelerated timing of "kings" leading to an error like "kings escort pages," it can lead to an alternate structure such as "kings are escorted by pages." In fact, this highlights how binding-by-timing, coupled with a principle of incrementality, can account succinctly for both errorless order of mention effects and word exchange errors. These kinds of generalizations are not as neatly captured by binding-by-checking.

Levelt et al. make a cogent observation when they identify speech errors with binding failures. However, their binding-by-checking mechanism leads them to an unusual (and we think, erroneous) account of speech errors. Their model does not naturally produce phonological exchanges, such as "sed rock" for "red sock." Many models based on binding-by-timing do. First, a hyperactivated unit, [s], is selected over the correct one, [r]. The selected unit is then inhibited and hence is less likely to be selected for the next syllable. But the replaced unit [r] is still available for that syllable because it was not inhibited. Exchanges are not a necessary consequence of binding-by-timing, but at least they mesh well with theories that emphasize the need to control when the elements of a sequence are activated. In Levelt et al.'s account, any tendency for exchanges would have to be stipulated.

In general, viewing phonological exchanges as temporal binding errors may be very useful. Just as feature migrations in visual search tasks occur when spatial attention is too widely distributed to bind features to objects (e.g., Treisman & Gelade 1980), speech errors may occur when attention is too widely distributed over the relevant "objects," specifically phonological words or syllables. In both visual search and speaking, errors reflect the similarity of the objects and how they are arranged in space or time. The only difference is that in speech production attention is distributed over time rather than space.

In conclusion, Levelt et al. have presented a ground-breaking theory of lexical access in production, one that makes far greater use of experimental data than any previous account. Their binding-by-checking mechanism may be a convenient way of dealing with the myriad binding problems in speaking, but ultimately, theorists will have to explain how attentional mechanisms control the timing of the activation of representational units. In speaking, just as in humor, timing really is the most important ingredient.

ACKNOWLEDGMENT

Preparation of this comment was supported by NIH grants HD21011 and DC00191.

Applying Ockham's chainsaw in modeling speech production

Ludovic Ferrand

Laboratoire de Psychologie Expérimentale, CNRS and Université René Descartes, 75006 Paris, France. **ludovic.ferrand@psycho.univ-paris5.fr**

> Things should be made as simple as possible, but not simpler.
>
> –Albert Einstein

Abstract: The theory of lexical access in speech production reported by Levelt, Roelofs & Meyer is exciting, well-described and well-organized, but because it relies mainly on the principle of simplicity (Ockham's razor), I argue that it might not be true. In particular, I suggest that over-applying this principle is wrong.

Levelt, Roelofs & Meyer have done a good job in summarizing their exciting work. They provide us with an authoritative, well-presented review of their speech production model and related experimental studies. However, I feel that their assumptions about the speech production model rest on a number of theoretical preconceptions that are not necessarily valid. My main criticism concerns their reliance on Ockham's razor, which I believe they overapply.

Levelt et al. apply Ockham's razor in modeling lexical access in speech production. This principle states: "Plurality should not be posited without necessity" (William of Ockham, ca. 1285–1349). In other words, one should not make more assumptions than the minimum needed. Ockham's razor, also called the principle of parsimony or simplicity, underlies almost all scientific modeling and theory building. Levelt et al. adopt its strong version, requiring that their model – and models in general – be as simple as possible. But why should a theory of lexical access in speech production be simple? Any real problem may be complex and may hence call for a complex model.

Part of the problem is also knowing what counts as necessary. To modularists like Levelt et al., interactionists multiply entities unnecessarily (feedback connections, inhibitory mechanisms, parallel activation, cascade processing). To interactionists, positing interactivity (e.g., Dell et al. 1997; Humphreys et al. 1997) and inhibition (e.g., Stemberger 1985) is necessary. In the end, Ockham's razor says little more than that interactivity and inhibition are superfluous for modularists but not for interactionists.

Levelt et al.'s working hypothesis is that a strictly serial, feed-forward, semantically holistic, localist model, with morphological decomposition and no inhibition is considerably simpler to describe, comprehend, and test than a cascaded, interactive, semantically componential, distributed model with inhibition and without morphological composition (see also Jacobs & Grainger, 1994, for a similar claim applied to visual word recognition). But since every model is, by definition, incomplete (and a fortiori, so are simple models such as the one defended by Levelt et al.), it is hardly surprising that a more complex set of complementary models is generally needed to describe lexical access in speech production more adequately.

I will consider two features that are not needed in Levelt et al.'s model, although experimental evidence favor their inclusion in a (more complex) model of speech production. Consider first the "inhibitory" feature. Levelt et al. did not include any inhibition, but their model is unable to explain the semantic inhibition effects in picture naming observed by Wheeldon and Monsell (1994) who reported unambiguous evidence that the lexicalization process during picture naming is inhibited when a word likely to be a competitor has been primed by a recent production. Naming a pictured object (e.g., SHARK) was retarded when a competing word (e.g., WHALE) had been recently elicited by a definition. Berg and Schade (1992; Schade & Berg 1992; see also Dell & O'Seaghdha 1994) have summarized other empirical evidence in favor of a model that includes both excitatory and inhibitory mechanisms. These data are hard to reconcile with the purely activation-based model of Levelt et al.

Consider now the "interactivity" feature. A thorough examination of the cognitive neuropsychology literature reveals evidence for the existence of interactive processes between lexical selection and phonological encoding (Dell et al. 1997; Laine & Martin 1996; Martin et al. 1996). For example, Laine and Martin (1996) studied an anomic patient who suffered from a partial functional disconnection between lexical-semantic and lexical-phonological levels. Systematic manipulation of semantic and phonological relatedness between the to-be-named targets indicated that this patient's word error patterns were sensitive to both types of lexical relatedness. Levelt et al.'s discrete theory of lexical access is unable to explain the observed effects of semantic and phonological relatedness. However, these results are consistent with an interactive activation model.

In sum, there is a great deal to disagree with in the target article. The moral of this story is that Ockham's razor should not be wielded blindly.

ACKNOWLEDGMENT
I am grateful to Art Jacobs for his insightful comments.

Prosody and word production

Fernanda Ferreira

Department of Psychology, Michigan State University, East Lansing, MI 48824-1117. **fernanda@eyelab.msu.edu**
eyelab.msu.edu/people/ferreira/fernanda.html

Abstract: Any complete theory of lexical access in production must address how words are produced in prosodic contexts. Levelt, Roelofs & Meyer make some progress on this point: for example, they discuss resyllabification in multiword utterances. I present work demonstrating that word articulation takes into account overall prosodic context. This research supports Levelt et al.'s hypothesized separation between metrical and segmental information.

Levelt, Roelofs & Meyer have done an enormous service to the field of psycholinguistics by articulating an explicit theory of lexical access in production. The model is motivated by almost the entire range of relevant empirical data: speech errors, patterns of reaction times in distractor tasks, and so on. My goal in this commentary will be to amplify their main points by bringing in issues concerning word production in multiword utterances.

As Levelt et al. note, words are produced in phonological contexts. In Ferreira (1993), I examined how prosodic constituents created from syntactic structure might influence the characteristics of spoken words, especially their durations. In two separate experiments I asked speakers to perform the following task: They were to read a sentence on a computer monitor, commit the sentence to immediate memory, and produce the sentence upon receipt of a cue (the question "What happened?"). Two factors were varied in each experiment: the segmental content of a critical word within the sentence and the prosodic context in which that word occurred. The dependent measures were the duration of the critical word, the duration of any following pause, and the sum of those two durations (total duration).

In the first experiment, the critical word was either phrase-medial or phrase-final and had either a short or long intrinsic duration. Thus, the following four conditions were tested:

> The chauffeur thought he could *stop* (short duration) / *drive* (long duration) the car (phrase-medial position).
>
> Even though the chauffeur thought he could *stop* / *drive*, the passengers were worried (phrase-final condition).

The theory I proposed assumed that segmental content and position within a prosodic constituent would both influence duration, but independently. The results supported the prediction: word durations were longer for long words and for words in the phrase-final environments, and the two effects were additive. More important, total duration was affected only by position, indicating that pause time was used to compensate for intrinsic duration to allow the prosodic interval specified by the prosodic constituent structure to be filled out. I concluded that when words are produced in sentences, the production system assigns an abstract duration to each word based on its placement in a prosodic constituent structure. Words at the ends of prosodic constituents would be assigned longer durations than those in other locations (Selkirk 1984); word lengthening and pausing are used to fill out the interval. In addition, I argued that when the intervals are created the production system does not yet know about the segmental content of the words that will occupy them.

The second experiment used the same logic but varied prosodic constituency in a different way: the critical word was either spoken with contrastive prominence or without. The conditions are illustrated as follows:

> The *cat* (short duration) / *mouse* (long duration) crossed the busy street.
>
> The CAT / MOUSE crossed the busy street.

The discourse context preceding each item established the need for contrastive prominence. Again, it was expected that this prosodic variable and segmental content would both affect durations, but independently. The prediction was again confirmed, and furthermore the pattern of data was identical to that found in the experiment manipulating location within prosodic constituents: total durations were virtually identical in the two segmental conditions but longer when the word was spoken with contrastive prominence. This finding reinforces the point that abstract intervals are specified for words based on the properties of the prosodic context in which they occur. A word with contrastive prominence would be assigned a longer interval before the production system had any knowledge of the segmental content of the word that would end up occupying that interval. When the speaker actually produces the sentence, he has to assign a duration that maintains the size of the interval; thus, if the word's segments give it a short intrinsic duration, a longer pause will be necessary.

These experiments make two relevant points. First, they support Levelt et al.'s assumption that words have metrical frames that specify a word's number of syllables and main stress position, but which do not include segmental content. Before elaborating further on this point and going on to the second, it is necessary to consider a study conducted by Meyer (1994), which is apparently inconsistent with this conclusion. Meyer's experiments considered whether the conclusions I drew in my 1993 paper would hold up in Dutch. Meyer varied certain characteristics of the critical words, such as whether vowels were short or long, and she found only partial support for my conclusions: pause compensation occurred but was only partial, so that the words with longer syllables had longer total durations. Two possible counterarguments can be made: one is that there simply are cross-linguistic differences in these phenomena (this is the conclusion Meyer draws). The other is that when duration intervals are created they ignore segmental content but not CV (consonant-vowel) structure. In other words, the difference between a word with a short vowel such as *tak* (branch) and a long vowel such as *taak* (task) might be represented as CVC versus CVVC, and that information might be available independently of segmental content. A critical experiment, then, would be to vary Dutch syllables differing only in whether a vowel is tense or lax; the model I laid out predicts that pause compensation will be total, while Meyer's arguments would lead one to expect any compensation to be only partial.

Second, regardless of when segmental content is accessed during the production of words in sentences, it is clear that word production depends critically on prosodic context. Words are produced differently when they occur at the edge of a major prosodic constituent compared with any other position, and their characteristics differ when they are semantically prominent rather than more discourse-neutral. The latter point suggests that the conceptual information that makes its way through the word production system includes not just stored knowledge but dynamically specified information as well.

Naming versus referring in the selection of words

Peter C. Gordon

Department of Psychology, The University of North Carolina at Chapel Hill, Chapel Hill, NC 27599-3270. **pcg@email.unc.edu**
www.unc.edu/depts/cogpsych/gordon

Abstract: The theory of lexical selection presented by Levelt, Roelofs & Meyer addresses the mechanisms of semantic activation that lead to the selection of isolated words. The theory does not appear to extend naturally to the referential use of words (particularly pronouns) in coherent discourse. A more complete theory of lexical selection has to consider the semantics of discourse as well as lexical semantics.

A successful theory of lexical selection in speech production must explain a huge number of phenomena, ranging from the semantic makeup of words to how the articulatory form of a phonetic segment accommodates neighboring segments. The theory presented by Levelt, Roelofs & Meyer tackles an impressive number of these phenomena with substantial conceptual analysis and empirical evidence, but it does not tackle all of them. In this commentary I will analyze issues associated with one particular lexical phenomenon, the use of pronouns, that Levelt et al. do not address and that provide some challenges to the framework that they present.

Pronouns are very common; "he" is the tenth most frequent word in English and other pronouns ("it," "his," and "I") are not far behind (Kuçera & Francis 1967). It seems likely that this class of words is frequent because it is very useful. Researchers from a number of disciplines (e.g., Brennan 1995; Fletcher 1984; Grosz et al. 1995; Marslen-Wilson et al. 1982) have focused on how pronouns (and other reduced expressions) contribute to the coherence of discourse by implicitly marking the semantic entities that are central to a discourse. Do the mechanisms described by Levelt et al. provide a way of accounting for this use of pronouns?

Levelt et al.'s theory has been developed primarily to provide a detailed account of speakers' performance in the picture-naming task. A speaker who repeatedly exclaimed "it" in response to the stimulus pictures would justifiably be regarded as uncooperative, so it is not surprising that the task provides little evidence on pronoun selection. Yet during an ordinary conversation a speaker would be expected to use "it" frequently to refer to the kinds of objects represented in the stimulus pictures. Consider the following sentence, which Levelt et al. offered as a possible way that a speaker might describe a picture.

> I see a chair and a ball to the right of it.

Here the expressions "a chair" and "it" are coreferential; they refer to the same thing. Presumably Levelt et al. would say that the word "chair" is selected because the conceptual and semantic preconditions stimulated by the picture lead to its being the most highly activated lemma. But then why is the same entity subsequently referred to with "it"? The semantic system used by Levelt et al. was selected to avoid the hyperonym problem, that is, in a compositional representation the activation of the semantic features of a concept will also activate any superordinates of that concept. From this perspective on semantics, pronouns might be seen as the ultimate hyperonyms, those specifying only minimal semantic features such as gender, number, and animacy. Levelt et al.'s use of a semantic system designed to avoid inadvertent intrusion of hyperonyms does not immediately suggest a reason for the widespread, systematic use of a particular class of hyperonyms.

The example sentence points to the powerful ways in which sentential and discourse context can influence lexical selection, a topic I have worked on with my colleague Randall Hendrick. Building on the work of others, we (Gordon & Hendrick 1997; in press) have argued that a discourse model consists of a set of semantic entities and a series of predications, of which the entities are arguments. With respect to lexical selection, we have argued that the primary purpose of names (and other unreduced referring expressions) is to introduce semantic entities into a model of discourse, whereas the primary purpose of pronouns (and other reduced expressions) is to refer directly to entities that are prominent in the discourse model. On this view, results from a task like picture naming may provide evidence about the mechanisms that a speaker uses to initially introduce an entity into the discourse. Competitive selection based on activation levels is an attractive mechanism for accomplishing this, one that builds on the substantial explanatory use of such mechanisms within cognitive psychology to account for a variety of semantic and lexical processes. However, this appealingly straightforward mechanism may have to be substantially augmented if it is to account for the selection of words that refer to entities in an evolving discourse model as well as words generated from long-term memory. For the example sentence, we would argue that the selection of the pronoun reflects the prominence of the semantic entity CHAIR in the discourse model. To the extent that we are right, a theory of lexical selection must not only address the semantic organization of lexical concepts in long-term memory (as Levelt et al.'s theory does); it must also address the semantics of models created to represent the interrelations of entities and concepts in a discourse. The principles governing lexical selection from these two domains may be quite different.

Will one stage and no feedback suffice in lexicalization?

Trevor A. Harley
Department of Psychology, University of Dundee, Dundee DD1 4HN, Scotland. **t.a.harley@dundee.ac.uk**
www.dundee.ac.uk/psychology/staff.htm#harley

Abstract: I examine four core aspects of WEAVER++. The necessity for lemmas is often overstated. A model can incorporate interaction between levels without feedback connections between them. There is some evidence supporting the absence of inhibition in the model. Connectionist modelling avoids the necessity of a nondecompositional semantics apparently required by the hypernym problem.

Four aspects of the WEAVER++ model are important for the earlier stages of lexical access in speech production: the existence of lemmas, the absence of feedback connections, the absence of inhibition, and the existence of the hypernym problem to motivate a nondecompositional semantic representation. I wish to question the necessity of the first aspect, half agree with the second, agree with the third, and disagree with the fourth.

According to WEAVER++, lexicalization occurs in two stages. Before we can access phonological forms (morphemes) we must first access an intermediate level of syntactically and semantically specified lemmas. The evidence for lemmas is, however, often overstated. Much of the evidence frequently adduced to motivate the existence of two stages of lexical access only motivate the existence of two levels of processing: a semantic–conceptual level and a morpheme–phonological level. ("Stratum," with its implication of sublevels, is indeed a better name here than "level.") The ambiguous evidence includes two types of whole word substitution – semantic and phonological – in normal speech, a distinction between semantic and phonological anomia, and two types of processes in picture naming.

As is increasingly being recognised, the only evidence that directly addresses the existence of lemmas as syntactic–semantic packages is fairly recent, and is currently mainly restricted to findings that noun gender can be retrieved independently of phonological information in the tip-of-the-tongue state in Italian (Vigliocco et al. 1997) and by the anomic Dante (Badecker et al. 1995). Even these findings have an alternative interpretation: it is not clear that we must access an intermediate stage to access the phonological forms (Caramazza 1997). Hence at present the data may be accounted for more parsimoniously without lemmas.

A second important aspect of WEAVER++ is that there is no interaction between levels in the model; in particular, there are no feedback connections between phonological forms and lemmas. On the one hand, the evidence for feedback connections does look decidedly shaky, particularly given the reaction time work of, for example, Levelt et al. (1991a). On the other hand, speech error evidence (in particular the existence of mixed word substitution errors; see Dell & Reich 1981; Harley 1984; Shallice & McGill 1978) suggests that there is interaction between the phonological and semantic levels. How can we observe interaction without feedback? Hinton and Shallice (1991) showed how mixed errors in reading can arise in a feedforward network through the presence of semantic attractors. Applying this architecture in reverse, we will find mixed errors in lexicalization if phonological forms lie in a basin of attraction which we might coin a "phonological attractor." Of course, we still have the problem of whether or not overlap in lexicalization actually occurs: Levelt et al. (1991a) failed to find experimental evidence for cascading activation (but see Dell & O'Seaghdha 1991; Harley 1993).

Third, there is no inhibition in WEAVER. The presence of inhibition is notoriously difficult to test. Some evidence comes from a tip-of-the-tongue study by Harley and Brown (1998), who found that tips-of-the-tongue are least likely to occur on words that have many phonological neighbours. This is suggestive of the absence of within-level inhibition (at the phonological level at least) rather than conclusive, but is consistent with WEAVER++.

The fourth aspect of the model is that semantic representations are unitary: there is no semantic decomposition into semantic features, but instead each lemma stands in a one-to-one relationship with a lexical concept. The motivation for this is the hypernym problem (Levelt 1989): if we posit a decompositional semantics, accessing a word (e.g., *dog*) should automatically access its hypernym (e.g., *animal*), as the truth conditions are equally satisfied for both. Connectionist models of semantics (e.g., Hinton & Shallice 1991) provide a resolution to this paradox: it is possible to have a decompositional semantics yet avoid talking in hypernyms (see also Caramazza 1997).

Thus I suggest that it is worth investigating a one-stage, feedforward, inhibition-free, decompositional model as an alternative to the two-stage, inhibition-free, and noninteractive model of Levelt et al., and indeed to the two-stage, feedback models (some containing inhibition) that have been set up as the main alternative (Dell 1986; Dell et al. 1997; Harley 1993). The advantage that Levelt et al. have over the simpler alternative is that they can point to an enormous body of modelling and experimental work that supports their model. Whether the simpler alternative can do as well is as yet unknown.

What exactly are lexical concepts?

Graeme Hirst
Department of Computer Science, University of Toronto, Toronto, Ontario, Canada M5S 3G4. **gh@cs.toronto.edu**
www.cs.toronto.edu/dcs/people/faculty/gh.html

Abstract: The use of lexical concepts in Levelt et al.'s model requires further refinement with regard to syntactic factors in lexical choice, the prevention of pleonasm, and the representation of near-synonyms within and across languages.

It is an appealing idea in Levelt et al.'s model that there are distinct lexical concepts that mediate between purely semantic features or concepts on the one side and lemmas on the other. This would indeed simplify a number of matters in lexical choice. It is unclear, however, whether Levelt et al. see lexical concept nodes as ordinary concept nodes that just happen to have a lemma node associated with them or as a separate and qualitatively different kind of node. Are they different at least to the extent that the conceptual preparation process uses their special status to recognize whether or not the message is fully "covered" by lexical concepts? Or does coverage just "happen" as an automatic result of activation of the concepts in the message, with any failure of coverage resulting in an error condition (such as a temporizing word like *um*)? Can syntactic operations on the lemmas feed back to the lexical concepts?

These questions arise because in Levelt et al.'s description, to activate a lexical concept is, in effect, to choose a word that is eventually to appear in the utterance; yet this lexical choice is blind with respect to syntax, as syntactic information is given only in the corresponding yet-to-be-activated lemma. While lexical concepts might guarantee the expressibility of the content of the message as words (sect. 4.1), they do not guarantee the syntactic realizability of those words in the current context. For example, selecting no more than the lexical concepts that lead to the lemmas for two verbs and a noun is insufficient if a sentence is what is to be uttered, but it is too much if a noun phrase is what is required.

Thus it must sometimes happen that lexical concepts and their lemmas are selected but cannot be used. Either there must be some syntactic feedback that influences a subsequent search for more-suitable lexical alternatives (notwithstanding Levelt et al.'s remarks in sect. 3.2.5), or lexical selection must be syntactically redundant in the first place so that syntactic processes have a high enough probability of finding what they need among the active lemmas, or both. Consider, for example, the English noun *lie* ("untrue utterance") and the verb *lie* ("utter an untruth"). These words are usually considered to be "counterparts" in some sense – noun and verb forms of "the same concept"; we might therefore posit some special relationship between them in the model. Yet there must nonetheless be two distinct lemmas here, because their syntax is different, and indeed there must be two distinct lexical concepts, because the act is not the product; *John's lie was incredible* and *John's lying was incredible* are distinct in meaning and both must be producible. However, both *lie* concepts would normally be activated at the same time, because there is no lie without lying, no lying without a lie. But selecting just one of the two *lie* lemmas would risk the possibility of syntactic ill-formedness in context, requiring the time-consuming selection of some lexical or syntactic alternative. In this case, the alternative could be either the other lemma or a completely different one or a change in syntactic structure (e.g., using the noun lemma by finding a suitable verb, as in *tell a lie*). Simultaneous selection of both *lie* lemmas would allow syntactic processes to create either, say, the sentence *John lied* or the noun phrase *John's lie,* as the context necessitates – but *John's lying* must be prevented if that is not the sense required, and so must the syntactically ill-formed **John lied a lie.*

Levelt et al.'s "hyperonym problem" is really just a special case of what we might call *the pleonasm problem* – the need to prevent the selection of redundant words, whether because of simple hyperonymy or because one includes the other in its meaning. For example, one can say *John removed the cork from the bottle* or *John uncorked the bottle;* but **John uncorked the cork from the bottle* is ill-formed because of pleonasm, even though the concept of corks will surely be active whenever that for cork-removal is. The more general problem is not solved simply by the idea of lexical concepts (nor heavy-duty selectional restrictions on verbs), or by shifting the problem to the arena of "convergence in message encoding" (Roelofs 1997a); meaning-inclusion relationships can be complex and arbitrary. What is needed is a psychological version of computational models such as that of Stede (1996), in which pleonasm is prevented by distinguishing between the meaning of a word and its coverage. Stede's model can produce both of the well-formed *cork* sentences above from the same semantic representation, but will not produce the ill-formed sentence.

Lexical concepts are an unparsimonious representation, as they require a distinct node for each sense of each word in each language that the speaker knows. But in any language, there are many groups of near-synonyms – words whose meanings overlap conceptually and yet are distinct in their nuances, emphases, or implicatures (Hirst 1995), as evinced by the many books of synonym-differentiation that are available (e.g., Gove 1984; Hayakawa 1968). For example, *error, mistake, blunder, slip,* and *lapse* are (part of) such a cluster in English. German and French have similar but different clusters (*Fehler, Versehen, Mißgriff, Fehlgriff, . . . ; impair, bévue, faux pas, bavure, bêtise, gaffe, . . .*). Levelt et al.'s model predicts that such near-synonyms within and across languages will all have distinct lexical concept nodes. But it is in fact quite difficult to differentiate near-synonyms in a concept hierarchy in strictly positive terms (Hirst 1995). The multilingual speaker is at a particular disadvantage. While Levelt et al. suggest that lexical concepts would be "to some extent language specific" (sect. 4.1), in practice we would expect them to be almost fully language specific, for even words that are superficially equivalent in two languages rarely turn out to be conceptually identical. For example, the English *warm* and *hot* do not precisely correspond to the German *warm* and *heiss,* as *hot water* (for washing) in English is *warmes Wasser* in German (Collinson 1939); and in any pair of languages, such cases are more the rule than the exception (for many examples, see Collinson 1939 and Hervey & Higgins 1992). In each such case, the lexical concepts must be language specific. So in this model, code-switching by a multilingual speaker must involve more than making lexicons available or unavailable *en masse* – it must involve making many individual elements of the conceptual system available or unavailable as well. (Alternatively, one might posit that lexical concepts for languages not presently in use *do* remain available but their lemmas do not. If this were so, we would expect wrong-language errors to be much more frequent for multilingual speakers than they actually are.)

Edmonds and Hirst (Edmonds 1999; Hirst 1995) have developed an alternative model of the lexicon in which groups of near-synonyms (in one language or several) share a single concept node, and the lexical entries contain explicit information as to their semantic and pragmatic differentiation. Edmonds and Hirst make no claims for psychological reality of this model – it is based solely on computational considerations – but the results of Peterson and Savoy (1998), which Levelt et al. acknowledge, offer tantalizing possibilities in this regard.

ACKNOWLEDGMENTS
This commentary was written while I was a guest of the Department of Computer Science, Monash University, with facilities generously provided by that department, and was supported by a grant from the Natural Sciences and Engineering Research Council of Canada. I am grateful to Chrysanne DiMarco, Philip Edmonds, Manfred Stede, and Nadia Talent for helpful comments on an earlier version.

Modeling a theory without a model theory, or, computational modeling "after Feyerabend"

Arthur M. Jacobs[a] and Jonathan Grainger[b]
[a]*Fachbereich Psychologie, Philipps-University Marburg, D-35032 Marburg, Germany;* [b]*CREPCO-CNRS, Université de Provence, 13621 Aix-en-Provence, France.* **jacobsa@mailer.uni-marburg.de**
staff-www.uni-marburg.de/~agjacobs/ **grainger@newsup.univ-mrs.fr**

> It is the principle: anything goes
>
> (Feyerabend 1975 p. 19)

Abstract: Levelt et al. attempt to "model their theory" with WEAVER++. Modeling theories requires a model theory. The time is ripe for a methodology for building, testing, and evaluating computational models. We propose a tentative, five-step framework for tackling this problem, within which we discuss the potential strengths and weaknesses of Levelt et al.'s modeling approach.

The cognitive science literature provides no generally accepted policy for building, testing, and evaluating computational models such as WEAVER++, which Levelt et al. use to "model their theory." The situation is anarchic, resembling Feyerabend's (1975) "anything goes" principle, and badly needs a model theory or methodology to organize the plethora of computational models (Jacobs & Grainger 1994). Whereas mathematical learning theory provides a wealth of testing principles, computational modelers today seem little concerned with this issue (Prechelt 1996).

Here, we discuss the strengths and weaknesses of Levelt et al.'s modeling approach in a tentative five-step framework for building, testing, and evaluating computational models inspired by psychometrics and test theory (Jacobs et al. 1998). Although we do not claim that this framework is optimal, such an explicit approach can be constructively criticized and thus advances the enterprise of computational modeling.

Step 1. Parameter-tuning studies check the global appropriateness of the architectural and parametric assumptions during the initial phase of model construction to uncover fundamental flaws (e.g., parameter configurations producing catastrophic behavior). In Levelt et al.'s original 12-node/7-parameter WEAVER, this step is not made transparent. Although the model is said to yield "equivalent results" when scaled up to a 25- or 50-node version, parameter-tuning studies may become necessary when the toy-WEAVER++ is scaled up to become seriously comparable with potential competitor models of word recognition and production that have artificial lexica of several thousand words, such as the multiple read-out model with phonology (MROM-p; Jacobs et al. 1998).

Step 2. Estimator set studies estimate model parameters from "important" data such that the point in parameter space is identified for which the model is "true." This raises several problems, only partly made explicit by Levelt et al., who fit WEAVER++ to Glaser and Düngelhoff's data using the Simplex method and a chi-square criterion.

First, the importance of estimator data, optimality of estimation method, and the choice of evaluation criterion have to be discussed. Levelt et al. say nothing about the consistency and/or bias of Glaser and Düngelhoff's data or the sufficiency of their estimation method. The simplex method is suboptimal (slow and inaccurate), as can be demonstrated using Rosenbrock's ("banana") function. Moreover, using chi-square analysis is suboptimal because, with enough data points, every model is eventually rejected. Apart from these potential weaknesses, not specific to their own approach, Levelt et al. leave readers in ignorance about two related problems: parameter independence and uniqueness of solution. Only with respect to the problem of identifiability do LRM become explicit: WEAVER++ indeed fares well, since its seven parameters were estimated from 27 data points.

Step 3. Criterion set studies test model predictions with parameters fixed (in Step 2) against fresh data, using an explicit criterion. One problem is which criterion to use: descriptive accuracy, the criterion chosen by Levelt et al., is one possibility.

Another problem Levelt et al. overlook is that finding that one model fits data better than competing models does not establish the best-fitting model as the probable source of the data (Collyer 1985). Levelt et al. choose a relatively strong test, using data concerning the same effect as the one used in Step 2 (i.e., stimulus onset asynchrony effects), but from different empirical studies and tasks. A stronger test would use a different effect from a different study (Jacobs et al. 1998).

A third problem is that Levelt et al. fit their model to averaged data. Whether this is a fair test for complex models such as WEAVER++ can be questioned. Fine-grained, item- and participant-specific predictions, or predictions of response (time) distributions and speed-accuracy trade-off functions seem more appropriate for powerful simulation models. Models that predict only mean data can be contradicted by further distributional analyses of the same data and cannot be reasonably discriminated from other models (Grainger & Jacobs 1996; Ratcliff & Murdock 1976).

Step 4. Strong inferential studies involve formal, criterion-guided comparisons of alternative models against the same data sets to select the best model. Although reasonably comparable models are on the market (Dell 1988; Schade 1996), Levelt et al. omit this important step at the risk of confirmation bias: there is little doubt that strong inference is the best falsification stratagem against modelers' "built-in" tendency to avoid falsifying their own model (Platt 1964). Of course, strong inference requires more than criteria of descriptive accuracy, for example, generality and simplicity/falsifiability. Perhaps Levelt et al. omit this step because they think there is no consensus on evaluation criteria for computational models. However, promising attempts exist in the field of mathematical modeling (Myung & Pitt 1998), as well as for computational models of word recognition (Jacobs & Grainger 1994).

Step 5 (model refinement or replacement). The final step depends on the outcome of Step 4. As firm believers in, but practically mild (nondogmatic, non-naive) users of, a Popperian theory building approach, we would personally continue with a process of model refinement (after which one reiterates back to Step 1 as long as a model is only mildly discredited and no better alternative is available. Levelt et al. seem to share this view. One must nevertheless keep confirmation bias in mind. Step 4 is a good guard against this. Another is to define a strongest falsificator: an effect that the model excludes under all initial conditions. An exemplary strength of WEAVER++ is that it has such strongest falsificators (sects. 3.2.5 and 6.1.1).

If one accepts this tentative five-step framework for building, testing, and evaluating computational models, Levelt et al.'s approach fares well overall. We have pointed out several unsolved problems, not specific to Levelt et al.'s approach, and conclude that computational modelers need a methodology or model-theory if they want to optimize "modeling their theories." This involves agreement on two things: (1) a set of criteria for model comparison and evaluation and a standard way of applying them, and (2) a set of standard effects (from standard tasks) that any model of word recognition and production should be able to predict in a way that allows strong inference competition.

ACKNOWLEDGMENT
This research was supported by a grant from the Deutsche Forschungsgemeinschaft to A. Jacobs (Teilprojekt 7 der Forschergruppe Dynamik kognitiver Repräsentationen).

Strictly discrete serial stages and contextual appropriateness

J. D. Jescheniak[a] and H. Schriefers[b]
[a]*Max-Planck-Institute of Cognitive Neuroscience, D-04109 Leipzig, Germany and Centre of Cognitive Science, University of Leipzig, D-04109 Leipzig, Germany;* [b]*Nijmegen Institute for Cognition and Information, Nijmegen University, NL-6500 HE Nijmegen, The Netherlands.* **jeschen@cns.mpg.de schriefers@nici.kun.nl**

Abstract: According to the theory of lexical access presented by Levelt, Roelofs & Meyer, processing of semantic–syntactic information (i.e., lemma information) and phonological information (i.e., lexeme information) proceeds in a strictly discrete, serial manner. We will evaluate this claim in light of recent evidence from the literature and unpublished findings from our laboratory.

In their target article, Levelt, Roelofs & Meyer present a theory of lexical access in speech production that is unique in that it tries to cover most of the subprocesses involved in lexical access in speaking. In our commentary, we will focus on one central aspect of the theory, namely, the assumption of strictly discrete serial stages in lexical access.

According to this assumption, semantic–syntactic (i.e., lemma) processing and phonological (i.e., lexeme) processing proceed in two nonoverlapping stages with phonological activation being contingent on the selection of the corresponding lemma. While semantic competitors might well become activated at the lemma level early in lexicalization, their word forms remain inactive. Phonological coactivation, that is, the simultaneous activation of word forms other than the target word form, is explicitly excluded.

This is at least our reading of the strict principle proposed by Levelt et al. (1991a). In the present version of the theory, this principle has been somewhat relaxed. It now allows for one exception: If two lexical items are contextually equally appropriate to express a given concept, both lemmas might be selected and each of them will consequently undergo phonological activation (hereafter, *contextual appropriateness account*). This relaxation of the principle is a reaction to recent experimental work demonstrating that near-synonymous nontarget competitors (e.g., couch–sofa) become phonologically active, whereas the findings for nonsynonymous semantic competitors are less conclusive (Cutting & Ferreira, in press; Jescheniak & Schriefers 1997; 1998; Levelt et al. 1991a; Peterson & Savoy 1998).

The contextual appropriateness account accommodates these experimental findings yet maintains the spirit of the original principle. However, an alternative explanation exists, namely that semantic competition must be sufficiently strong in order to yield measurable phonological coactivation effects (hereafter *strength of competition account*). In studies investigating nonsynonymous semantic competitors, this competition might have been too small (e.g., Dell & O'Seaghdha 1992; Harley 1993; O'Seaghdha &

Marin 1997). These two competing accounts can be put to an empirical test by investigating competitors that differ semantically only to a minimal extent from the target but cannot replace it, given the communicative context.

The strength of competition account predicts measurable phonological coactivation of these competitors, whereas the contextual appropriateness account excludes it. Possible candidate items for such a test are pairs of antonymous adjectives (e.g., *rich – poor*). They are assumed to differ only with respect to a single semantic–conceptual feature (cf. H. H. Clark 1973), but this minimal difference makes the semantic–conceptual preconditions for their usage clearly incompatible. This leads to the following predictions: if a speaker intends to produce the word *"rich,"* the strength of competition account might well predict phonological coactivation of its antonym *"poor."*

By contrast, the contextual appropriateness account does not allow for such coactivation. We tested these predictions in a translation paradigm (Jescheniak & Schriefers, in preparation); native speakers of German translated visually presented English probe words into German target words while ignoring auditorily presented German distractor words occurring at different points in time (SOAs). If the distractors were presented early, production latencies were affected by distractors that were phonologically related to the (German) target or to its (German) antonym. Importantly, at later SOAs the effects dissociated. While the distractors phonologically related to the target were still effective, the other effect disappeared. This pattern contrasts with our observations from near-synonyms. Using the related cross-modal picture–word interference paradigm, no such dissociation of effects from distractors phonologically related either to the target or its near-synonym occurred at late SOAs (Jescheniak & Schriefers 1998). That is, near-synonyms and antonyms behave differently.

Elsewhere (Jescheniak & Schriefers 1998) we have argued that in the cross-modal interference paradigms we used, the dissociation versus nondissociation of two effects at late SOAs is most decisive in distinguishing between discrete and cascaded processes – with nondissociation being in conflict with discrete serial processing. If this is correct, the combined data from near-synonyms and antonyms support Levelt et al.'s contextual appropriateness account, according to which deviations from strictly discrete serial processing should be restricted to situations in which two lexical items are equally appropriate given the communicative setting.

Recent evidence on the production of pronouns (Jescheniak et al., submitted) also supports the assumption of strictly discrete serial stages. In cross-modal interference experiments, speakers of German named pictures by either a noun or the corresponding gender-marked pronoun. Distractors that were semantically related to the picture name affected picture naming latency in both noun and pronoun production, whereas distractors that were phonologically related to the picture name only affected noun production. This suggests that, in pronoun production, speakers can access the noun lemma (providing information about the noun's grammatical gender) without necessarily accessing the noun's phonological form, in line with the assumption of discrete serial stages. With respect to the contextual appropriateness account, these results also indicate that during pronoun production speakers did not consider the noun as an equally appropriate alternative, at least in our experiments.

ACKNOWLEDGMENT
Our research mentioned in this commentary was supported by a grant from the German Research Council (DFG Je229/2-1).

Incremental encoding and incremental articulation in speech production: Evidence based on response latency and initial segment duration

Alan H. Kawamoto
Department of Psychology, University of California at Santa Cruz, Santa Cruz, CA 95064. **ahk@cats.ucsc.edu** **zzyx.ucsc.edu/psych/psych.html**

Abstract: The WEAVER++ model discussed by Levelt et al. assumes incremental encoding and articulation following complete encoding. However, many of the response latency results can also be accounted for by assuming incremental articulation. Another temporal variable, initial segment duration, can distinguish WEAVER++'s incremental encoding account from the incremental articulation account.

Incremental encoding plays a key role in how the WEAVER++ model accounts for a number of response latency predictions. For example, rightward incrementality accounts for the facilitative effect in the implicit priming task when the primed segments occur at the beginning of the word but not when they occur elsewhere in the word (sect. 6.4.2).

However, the effect of the location of the implicitly primed segments on response latency could also be accounted for by assuming incremental articulation instead of articulation being initiated only after all the segments and syllables of a word have been phonologically and phonetically encoded (sects. 6.3, 8). That is, if articulation is initiated as soon as the first segment has been encoded, then articulation can be initiated prior to lexical selection when primed segments occur at the beginning of words, but articulation cannot begin until some time after lexical selection when there are no primed segments or when the primed segments occur at locations other than the beginning of the word. According to the incremental articulation account, sequential effects arise because articulation must necessarily be sequential, not because encoding is sequential.

It turns out that the incremental encoding and incremental articulation accounts can be distinguished empirically. For example, the incremental encoding account predicts that the facilitatory effect of the prime on response latency increases as the number of word-initial segments increases, but the incremental articulation account predicts no effect. The results from two experiments (Meyer 1991, expts. 5 and 6) provide support for the incremental encoding account; the amount of priming increases as the number of word-initial segments increases.

Despite the demonstration that the priming effect increases as the number of word-initial segments increases, this result does not rule out the incremental articulation account. To make the comparison of the two accounts more equitable, the relationship of the encoding and articulation accounts to the null and alternative hypotheses must be reversed. Such a reversal arises for predictions of the initial segment duration when a word-initial consonant is the only segment primed. The incremental articulation account predicts a longer initial segment duration in the homogenous compared with the heterogenous condition because articulation of the initial segment is initiated before lexical selection and continues until the second segment is encoded during the phonological and phonetic encoding stage. However, the initial segment duration is short in the heterogenous condition because both the first and second segments are encoded in quick succession during the phonological and phonetic encoding stages. By contrast, the incremental encoding account predicts no effect on initial segment duration because articulation is initiated only after all the syllables of a phonological word have been encoded.

Although the initial segment duration has not been measured directly, indirect measurements of initial segment duration can be obtained by exploiting differences in the articulatory and acoustic characteristics of word-initial segments (Kawamoto et al. 1998). In particular, naming latency for words beginning with plosives conflates response latency and initial segment duration because the onset of acoustic energy coincides with the end of the segment (i.e., the release of the plosive). By contrast, naming latency for words beginning with nonplosives corresponds closely to the actual response latency. Thus, the effect of the prime on initial segment duration corresponds to the interaction of plosivity and condition – the difference in the implicit priming effect for plosives compared to nonplosives.

To determine the initial segment duration, the summary results from experiments in which the implicit prime corresponded to the onset of the first syllable (Meyer 1991; expts. 1, 3, and 7) were combined and reanalyzed with plosivity as a factor. The implicit priming effect for plosives and nonplosives was 26 msec and 37 msec, respectively. Although the 11 msec difference was sizeable relative to segment durations, the interaction was not significant ($F(1,11) = 2.11, p > .15$). There are two reasons why the results may not have reached significance. First, N was very small because the analysis used summary results for specific onsets instead of specific words. Second, the analysis included words beginning with /h/, a segment that generates very little acoustic energy even compared to other voiceless fricatives. If voice keys do not detect the /h/, then /h/ is similar to a plosive. When onsets beginning with /h/ were not included in the analysis, the interaction was significant ($F(1,10) = 5.65, p < .05$).

The incremental articulation account also predicts no initial segment duration effect when the implicit prime includes the onset and nucleus of the initial syllable because the segment following the onset is already known when articulation is initiated. To test this prediction, the results from Meyer (1991; expts. 5–8) in which the homogenous condition included only the onset and either the nucleus or rime of the first syllable were analyzed with plosivity as another factor. There was no significant interaction whether /h/ was or was not included in the analysis ($F(1,13) < 1$, and $F(1,11) < 1$, respectively).

In conclusion, response latency effects provide additional insight into the timecourse of speech production. However, this is just the first step; segment duration must also be considered to understand temporal effects that arise after articulation has been initiated. These additional temporal effects, if substantiated, have important theoretical consequences.

Indirect representation of grammatical class at the lexeme level

Michael H. Kelly

Department of Psychology, University of Pennsylvania, Philadelphia, PA 19104-6196. **kelly@psych.upenn.edu** **www.sas.upenn.edu/~kellym**

Abstract: Lexemes supposedly represent phonological but not grammatical information. Phonological word substitutions pose problems for this account because the target and error almost always come from the same grammatical class. This grammatical congruency effect can be explained within the Levelt et al. lexicon given that (1) lexemes are organized according to phonological similarity and (2) lexemes from the same grammatical category share phonological properties.

Levelt et al.'s target article is an elaboration of Levelt (1989), in which the mental lexicon is divided into lemma and lexeme components. The lemma contains semantic and grammatical information about a word, whereas the lexeme contains its phonological representation. The relationship between these two levels in speech production is purported to be modular and feedforward. Competition among lemma alternatives is first resolved independently of phonological considerations, and the winner sends activation to the lexeme level. Lexeme selection then proceeds independently of lemma level factors. In other words, activity here is blind to semantic and grammatical information.

Levelt et al. summarize extensive evidence to support the lemma/lexeme distinction. However, certain types of word substitutions seem to contradict the strict separation of the levels. I will focus here on phonological interlopers, such as saying "apartment" for "appointment." The target and error in such "malapropisms" do not seem to be related semantically, but are strikingly similar phonologically. Indeed, detailed analysis has confirmed that they are likely to have the same stress pattern, constituent phonemes, and number of syllables (Fay & Cutler 1977). One might attribute these errors to incorrect lexeme selection, and the lack of semantic relatedness fits with Levelt et al.'s modular theory. However, the target and error almost always come from the same grammatical class. Thus, nouns replace nouns, verbs replace verbs, and so on (Fay & Cutler 1977). This grammatical agreement phenomenon seems to contradict Levelt et al.'s view of the lemma/lexeme distinction, since grammatical information is not represented at the lexeme level.

Levelt et al. could address this problem by invoking a grammatical binding principle that ensures the insertion of grammatically appropriate words into a developing syntactic frame. Dell (1986), in fact, has proposed such a binding system in his models of speech production. However, this mechanism may not be needed if the grammatical similarity effect in malapropisms can be seen as a simple byproduct of lexeme organization. In particular, it is generally agreed that the lexicon is organized according to phonological similarity. Hence, phonological neighbors of a target word could intrude on its successful selection. Suppose, in addition, that grammatical categories in a language are also distinguished by phonological properties. If so, then lexemes organized according to phonological similarity will, de facto, also be organized according to grammatical similarity. Even though grammatical information is not explicitly represented at the lexeme level, word substitutions driven by phonological similarity would carry grammatical similarity as free baggage.

An objection to this account would invoke the arbitrariness assumption in linguistic theory and claim that relations between word form and function are too rare for the latter to hitch wholesale rides on the phonology wagon. At least for English, this objection is false, as grammatical categories are correlated with many phonological variables. Furthermore, these correlations involve thousands of words and so permeate the entire lexicon (see Kelly, 1992, for review). These differences can be observed in both segmental and supersegmental phonology. At the segmental level, for example, back vowels occur more often in nouns, whereas front vowels occur more often in verbs (Sereno 1994). At the prosodic level, disyllabic English nouns are overwhelmingly likely to have stress on the first syllable, whereas disyllabic verbs favor stress on the second syllable (Kelly & Bock 1988; Sherman 1975). Nouns also tend to contain more syllables than verbs (Cassidy & Kelly 1991).

To determine just how accurately grammatical class can be inferred from the total phonological structure of a word, I recently developed a connectionist model that learned to classify English words as nouns or verbs based solely on their phonological representations. The model did quite well in generalizing its knowledge to novel cases, achieving 75% accuracy. This suggests that phonology is highly correlated with noun and verbhood in English, so a lexicon organized according to phonological similarity would also be organized according to grammatical similarity. Hence, targets and errors in malapropisms will be grammatically similar by virtue of their phonological similarity.

This approach can be pushed further by proposing the existence of word substitution patterns that have not been previously identified. I will assume that the intruder in malapropisms was selected for production because it had a higher activation level than the target. Grammatical class could be indirectly involved in determining these activation levels in the following way. Phonological neighborhoods can be envisioned as a core of quite similar items spreading outward to a more diffuse periphery where neighborhoods overlap. Suppose that a selected lemma sends activation to a lexeme that is relatively peripheral in its neighborhood. This activation will spread to phonologically similar neighbors. Some of these neighbors will be closer to the core, whereas others will be further toward the periphery. Activation will reverberate more strongly among the core alternatives than the more peripheral alternatives because of their greater phonological similarity. Hence, interlopers should be closer phonologically to the core of the neighborhood than to the periphery. Given the correlations between phonology and grammatical class in English, these phonological neighborhoods can be described grammatically. Thus, word substitutions should sound more typical phonologically for their grammatical class than the targets.

I tested this hypothesis using Fromkin's (1971) and Fay and Cutler's (1977) corpora of malapropisms. Malapropisms were classified as noun substitutions (e.g., "remission" replaced with "recession") or verb substitutions ("diverge" replaced with "deserve"). The targets and errors were then coded for their phonological structure and submitted to my connectionist model, with activation scores on the model's noun and verb nodes serving as dependent variables. Since the correlation between noun and verb activations was highly negative, results will be discussed in terms of noun activations.

Among the noun malapropisms, the mean error noun activation was higher than the mean target noun activation (.86 versus .82). In contrast, the errors in verb malapropisms had lower noun activations (and, correspondingly, higher verb activations) than the targets (.51 versus .49). It is not surprising that the sizes of these differences are small. The high phonological similarity between targets and errors guarantees that they will have similar noun activation scores. For instance, the scores for target "confusion" and error "conclusion" were .93 and .96 respectively. Likewise, the noun score for verb target "bought" was .35, whereas the score for its phonologically similar replacement "built" was .25. Despite these small differences, the errors move toward greater phonological typicality for their grammatical class. The pattern was statistically significant for nouns ($t(65) = 2.06, p < .05$), though not for verbs ($t(47) = -0.81, p > .40$).

In sum, then, the tendency for targets and errors in malapropisms to come from the same grammatical class does not pose a problem for Levelt et al.'s theory of the modular lexicon. In particular, the grammatical effect need not be attributed directly to grammatical representations at all. It arises from the process of phonological activation among lexemes coupled with correlations between phonology and grammatical class that effectively organize the lexeme level according to grammatical class. One implication of this account is that malapropisms should show grammatical congruency effects only in languages that have significant correlations between phonology and grammatical class.

The lexicon from a neurophysiological view

Horst M. Müller

Experimental Neurolinguistics Group, Faculty of Linguistics, University of Bielefeld, 33501 Bielefeld, Germany. **hmm@bionext.uni-bielefeld.de**
www.lili.uni-bielefeld.de/~enl/mueller.html

Abstract: (1) Reaction time (RT) studies give only a partial picture of language processing, hence it may be risky to use the output of the computational model to inspire neurophysiological investigations instead of seeking further neurophysiological data to adjust the RT based theory. (2) There is neurophysiological evidence for differences in the cortical representation of different word categories; this could be integrated into a future version of the Levelt model. (3) EEG/MEG coherence analysis allows the monitoring of synchronous electrical activity in large groups of neurons in the cortex; this is especially interesting for activation based network models.

Understanding the anatomical substrates and physiological functions underlying language processing in the brain is one of the most difficult challenges within cognitive neuroscience. In addition to empirical findings of psycholinguistics (e.g., reaction time [RT] studies), neurophysiological techniques offer a further source of evidence for understanding the functional organization of language. Although there has been remarkable development in noninvasive brain mapping and imaging techniques during the last decade, which has led to an impressive increase in facts and complex insights, the functioning of higher cognitive brain processes remains unclear. However, even though the present neurophysiological findings cannot explain language processing, they provide empirical evidence and allow new hypotheses within the triad of psycholinguistics, neurophysiology, and computational modeling. This is especially true for detail problems of language processing such as the representation of the lexicon in the brain.

Since the major part of the present knowledge of language processing is based on empirical findings with RT studies, Levelt, Roelofs & Meyer accentuate the role of the RT methodology and ground their theory and computational model mainly on it. Despite its advantages, this method has at least two disadvantages: (1) It is only a very indirect method to study the time course of language processing, where the times for unrelated motor processes (e.g., pressing buttons) are disproportionately longer than those for certain cognitive processes. This may miss differences in the millisecond range, which can be assumed for language processing. (2) There are cognitive processes that could be detected with neurophysiological techniques (e.g., electroencephalogram [EEG] or magnet-encephalogramm [MEG]), but not by the behavioral task because the subject is unaware of these processes.

For historically technical reasons, the present knowledge about language processing is somewhat distorted by the dominance of RT findings. Almost at the end of the "decade of the brain," the present neurophysiological findings and those still expected will counteract this and contribute to a view a bit closer to the cognitive reality. For this reason, at least at the moment, it seems critical already to use the output of the computational model to inspire neurophysiological investigations (sect. 11), instead of using neurophysiological findings to adjust the theory. This is to ensure that no self-fulfilling prophecies are produced. Based on the progress in neurophysiological techniques, the next step in evaluating Levelt et al.'s theory could be a stronger incorporation of present findings and those to be expected in cognitive neuroscience. This may be especially true for analysing the time course of language processing with electroencephalography (e.g., Kutas 1997; Müller et al. 1997), as well as for cortical representation of the lexicon or its components with brain imaging techniques such as functional magnetic resonance imaging (fMRI) (e.g., Spitzer et al. 1995) or positron emission tomography (PET) (e.g., Beauregard et al. 1997).

There is no need to mention the obvious elegance of the language production theory, or that Levelt et al. do not claim that it is complete. From a neurophysiological point of view, however, two points may provide additional support for the theory presented by the authors:

(1) There is empirical evidence for differences in the cortical representation of different word categories. The lexicon does not seem to be one functional unit within a distinct cortical arena. More probably it is organized by some categorical principles and separated into distinct subareas of the cortex that are characterized by different kinds of lexical concepts. Our knowledge about the physiology of the lexicon begins with electrophysiologically grounded processing differences between closed class and open class words, which can be described as function words (nouns, verbs, and adjectives) and content words (e.g., conjunctions, prepositions, and determiners), respectively (Kutas & Hillyard 1983). The next differentiations are much more sensitive and reveal several subclasses of nouns. Neuropsychological findings with brain damaged patients support differences within the lexicon (e.g., Damasio et al. 1996), as do electrophysiological studies (e.g., Müller & Kutas 1996). For example, the physiological reality of the distinction between concrete and abstract nouns is shown in several studies (Beauregard et al. 1997; Paller et al. 1992; Weiss & Rappelsberger 1996). In addition, there is evidence for differ-

ences between verbs and nouns (Brandeis & Lehmann 1979; Kutas 1997; Preissl et al. 1995) and for a special status for proper names within the category of concrete nouns (Müller & Kutas 1996).

Proper names (*Nomina propria*), as used by Levelt et al. in the example "*John said that . . .*" (sect. 3.1.2) are very special lexical entries. The authors state, "the activation of a lexical concept is the proximal cause of lexical selection" (sect. 5.1). Many see proper names as nouns without a lexical concept or at least with a very special one, and there is indeed electrophysiological evidence for a special status for proper names (Müller & Kutas 1996; Weiss et al. in press). In attempting to understand language processing, it would be of interest at a future stage in the language production theory to take these physiological findings about different word classes into consideration. There seem to be clearly differentiable neuronal generators for classes of lexical entries that we do not yet know how to classify. First results suggest that biologically grounded classes such as proper names, *tools* vs. *edible,* or *concrete* vs. *abstract* exist as real subsystems whose incorporation could make the present theory even stronger.

(2) One basic assumption of Levelt et al.'s theory is anchored in the physiological reality of language processing is supported by EEG coherence findings for the cortical activity network during single word comprehension. By means of a spectral-analytic analysis of the EEG (coherence analysis), it is possible to show the increase or decrease of synchronous electrical activity in certain frequency bands of large neuronal groups in the cortex. Weiss and Rappelsberger (1996; 1998) showed that hearing and reading of concrete nouns lead to a widespread cortical activity, while hearing and reading of abstract nouns lead to an activity mainly in the classical language related areas. Levelt et al. give the following explanation for this finding: Concrete nouns activate percepts in all sensory modalities, whereas abstract nouns are not related to such feelings and do not usually bring about such percepts, which seems to belong to the lexical concept. For example, hearing the word *cat* would evoke auditory, tactile, visual, and olfactoric percepts, which can be detected from an analysis of EEG coherence. Because of the widespread activation in a greater network induced by concrete nouns, these are not affected as easily by misfunction of some neurons as abstract nouns. This could explain the greater vulnerability of abstract nouns in patients (Weiss & Rappelsberger 1996). These findings provide additional support, especially for activation-based network models.

Parsimonious feedback

Padraig G. O'Seaghdha
Department of Psychology, Lehigh University, Bethlehem, PA 18015.
pgo0@lehigh.edu www.lehigh.edu/~pgo0/pgo0.html

Abstract: The insistence on strict seriality and the proscription of feedback in phonological encoding place counterproductive limitations on the theory and WEAVER++ model. Parsimony cannot be stipulated as a property of the language system itself. Lifting the methodological prohibition on feedback would allow free exploration of its functionality in relation to the substantial content of this major research program.

Levelt et al.'s target article provides a commanding overview of a comprehensive, ramified, and fecund theoretical, experimental, and computational program that is a touchstone for researchers in language production and connected issues. However, the exclusion of feedback from the model is an unnecessary limitation. Though historical fealty is undoubtedly at work, Ockham's famous admonition to parsimony is the only explicit motivation presented by Levelt et al. However, parsimony cannot be assumed to be a property of the language system; it is only something to which accounts of its underlying principles aspire. Moreover, the prohibition of feedback may be antiparsimonious, because it is at least partly responsible for the proliferation of supporting corollaries (e.g., between-level verification, monitoring, multiple lemma selection, storage of variant word forms, a suspension-resumption mechanism, a syllabary) that undermine the goal of simplicity in the theory. Levelt et al.'s willingness to entertain these other complications makes the doctrinaire rejection of feedback difficult to understand. I suggest that little of real value in the theory would be lost, but much would be gained, if the rejection of feedback were dropped from the model.

Feedback is endemic in biological systems, most pertinently in cortical systems that implement perceptual and motor processes. Therefore, without compelling evidence to the contrary, we should assume it is intimately involved in language production. Feedback may play an important if somewhat restricted role in lexical access, considered in isolation (Dell & O'Seaghdha 1991). In utterance generation, it probably plays a significant role in regulating complex horizontal and vertical planning and coordination requirements. If so, the prohibition of feedback by Levelt et al. can only delay the exploration of its actual role in language production. The more parsimonious scientific *strategy* is to allow feedback to be a "free parameter" and to let experiments and models discover its actual functionality.

Though the raison d'etre of feedback may lie in utterance planning and coordination, several lines of evidence converge on the conclusion that it has at least a limited influence on lexical access (sect. 6.1.1). First, it has been shown computationally that feedback is entirely compatible with the intrinsically multistaged nature of production (Dell & O'Seaghdha 1991; Harley 1993). Second, Martin et al. (1996) have made an incontrovertible case for semantic–phonological convergence in speech errors during single word production. Third, several forthcoming studies, cited by Levelt et al. point strongly to the conclusion that near-synonyms are phonologically coactivated at early moments in picture naming (Jescheniak & Schriefers 1998; Peterson & Savoy 1998), thus contradicting strict seriality.

Levelt et al. acknowledge the first two points but hold that they are not conclusive. With regard to the phonological coactivation evidence, they propose that near-synonyms are a special case where double lemma selection is likely, but that there is no evidence of more general phonological coactivation of activated lemmas. However, Peterson and Savoy's study is ambiguous on this point. Given a picture of a couch, the naming of *bet* (mediated by *bed*) was facilitated 5 msec in the relevant conditions (SOA > 50 msec), relative to 25 msec for the direct phonological relative *count.* This ratio is reasonable, given the simulations of Dell and O'Seaghdha (1991), but the weak coactivation effect requires considerably more power to arbitrate its reality (see O'Seaghdha & Marin, 1997, for relevant calculations and arguments). In addition, O'Seaghdha and Marin showed significant though small phonological coactivation (or semantic–phonological mediation) in equivalent conditions of a visual word-priming paradigm. This evidence is relevant to production because the issue is the phonological status of indirectly activated words, not the source of their activation; in neither case is there an intention to produce the semantic mediator (O'Seaghdha & Marin 1997, p. 249). Note that this claim is consistent with the higher-level perception–production unity in WEAVER++ (sect. 3.2.4, Assumption 3).

In short, it is reasonable to conclude from the existing evidence that experiments of sufficient power will show small but reliable phonological coactivation of all substantially activated lemmas, as well as strong phonological coactivation of near-synonyms. Of course, this configuration of outcomes is not a demonstration of feedback per se, but it removes the cornerstone of Levelt et al.'s argument against its existence.

The proscription of feedback also limits WEAVER++ by forcing it to rely heavily on a verification process that in comparison with interactive activation and parallel distributed processing (PDP) approaches is much too declarative for processes of word-form encoding. Verification in WEAVER++ appears to serve at least one of the functions of feedback in interactive models by supporting the selection of representations that are consistent across adjacent levels. But the lack of interactive dynamics is responsible for failures such as the absence of segment exchange errors (sect. 10, para. 3).

WEAVER++'s domain of application is also too closely tied to the picture–word paradigm in which form-related distractors yield facilitation (sect. 6.4). The model handles these cases easily, but will have difficulty in explaining how form facilitation modulates to competition in other production tasks. For example, naming a *tafel* is facilitated by the "distractor" *tapir* (Levelt et al.), but saying the word pair *taper–table* repeatedly tends to be inhibitory (O'Seaghdha & Marin, in press). A model with segment to word or word-form feedback explains segmental speech errors and form-related competition more easily in a unified framework (Dell 1986; O'Seaghdha & Marin, in press; Peterson et al. 1989; Sevald & Dell 1994). Of course, Levelt et al.'s detailed expositions of syllabification, metrical stress, morphological encoding, and many other matters in turn present new and interesting challenges to all existing models of language production. Given this embarrassment of riches, it is not surprising that no existing model can parsimoniously account for all the evidence.

Lexical access as a brain mechanism*

Friedemann Pulvermüller

Department of Psychology, University of Konstanz, 78434 Konstanz, Germany. **friedemann.pulvermueller@uni-konstanz.de**
www.clinical.psychology.uni-konstanz.de

*This commentary will be responded to in the Continuing Commentary section of a forthcoming issue.

Abstract: The following questions are addressed concerning how a theory of lexical access can be realized in the brain: (1) Can a brainlike device function without inhibitory mechanisms? (2) Where in the brain can one expect to find processes underlying access to word semantics, syntactic word properties, phonological word forms, and their phonetic gestures? (3) If large neuron ensembles are the basis of such processes, how can one expect these populations to be connected? (4) In particular, how could one-way, reciprocal, and numbered connections be realized? and, (5) How can a neuroscientific approach for multiple access to the same word in the course of the production of a sentence?

A processing model of lexical access such as the one described in detail in the target article is not necessarily a theory about brain mechanisms. Nevertheless, it may be fruitful to ask how the model can be translated into the language of neurons.

Feedback regulation is necessary! The brain is a device with extreme plasticity. Early in ontogenesis, neurons rapidly grow thousands of synapses through which they influence their neighbors and, in turn, receive influence from other neurons. These synaptic links become stronger with repeated use. Therefore, a particular brain-internal data highway that initially consists of a few fibers, may later include thousands or millions of cables with weak synaptic links, and may finally exhibit a comparably large number of high-impact connections. In this case, the same input to the system will lead early on to a minimal wave of activity, but finally lead to a disastrous breaker. A system with such an enormous variation of activity levels requires a regulation mechanism in order to function properly (Braitenberg 1978). The task of this mechanism would be to enhance or depress the global level of activity to keep it within the limits of optimal neuronal functioning.

A simple way to regulate activity in a neuronal system is to monitor activity levels of all neurons, calculate their sum, and provide an additional input to the system that is excitatory if this sum is small (to prevent extinction of excitation), but inhibitory if it is large (to prevent overactivation). Thus, a mechanism of inhibition (or disfacilitation) appears necessary in any brainlike model.

Levelt, Roelofs & Meyer state that their model does not include inhibition (sect. 3.2.2) and the fact that it does not may be interpreted as one of their minimal assumptions – evidencing a research strategy guided by Ockham's razor. Certainly, looking at the theory in abstract space, the assumption of inhibitory links would be an additional postulate that made it less economical and therefore less attractive. However, considering the brain with its well-known intracortical and striatal inhibitory neurons that are likely to be the basis of feedback regulation (Braitenberg & Schüz 1991; Fuster 1994; Wickens 1993), it does not seem desirable to propose that inhibitory mechanisms are absent in a model meant to mirror brain functioning.

Would this mean that the theory proposed by Levelt et al. is unrealistic from the perspective of brain theory? Certainly not. Although such mechanisms are not explicitly postulated or wired into the network (and therefore do not affect activation spreading), they kick in at the level of node *selection* where Luce ratios are calculated to obtain the probability with which a preactivated representation is selected (fully activated, so to speak). The probability that a particular node is selected depends upon its actual activity value *divided by* the sum of the activation values in a particular layer. Because the calculation performed is very similar to what a regulation device would do, one may want to call this implicit, rather than explicit, inhibition (or regulation). To make it explicit in the network architecture, an addition device in series with numerous intra-layer inhibitory links would have to be introduced. Thus, the model includes inhibition – although on a rather abstract level – and this makes it more realistic from the neurobiological perspective.

Brain loci of lexical access. Where in the brain would one expect the proposed computation of lexical concept, lexical syntax (lemma), word form, and phonetics? Most likely, phonological plans and articulatory gestures are wired in primary motor and premotor cortices in the inferior frontal lobe. The percepts and motor programs to which words can refer probably correspond to activity patterns in various sensory and motor cortices and thus may involve the entire cortex or even the forebrain. More specificity is desirable here; for example, words referring to movements of one's own body are likely to have their lexical concept representations localized in motor cortices and their vicinity, while lexical concepts of words referring to objects that one usually perceives visually should probably be searched for in visual cortices in occipital and inferior temporal lobes (Pulvermüller 1996; Warrington & McCarthy 1987).

Between phonetic-phonological and lexical-semantic representations the model postulates lemmas whose purpose can be considered to be three-fold: (1) not only do they glue together the meaning and form representations of a word, but, in addition, (2) they bind information about the word's articulation pattern and its sound image. Furthermore, (3) lemmas are envisaged to store syntactic knowledge associated with a word.

Intermediary neuronal units mediating between word form and semantics – the possible counterparts of lemmas have been proposed to be housed in the inferior temporal cortex (Damasio et al. 1996). The present theory would predict that lesions in the "lemma area" lead to a deficit in accessing syntactic knowledge about words (in addition to a deficit in naming). However, lesions in inferior temporal areas can lead to a category-specific naming deficit while syntactic knowledge is usually spared. Hence, it appears unlikely that lemmas are housed in the inferior temporal lobe. Is there an alternative to Damasio's suggestion?

One of the jobs of a lemma is to link the production network to the perception network (sect. 3.2.4). On the receptive side, sound waves and features of speech sounds activate neurons in the auditory cortex in the temporal lobe, and in order to store the many, many relation between acoustic phonetic features and articulatory phonetic features, it would have advantages to couple the respective neuron populations in auditory and motor cortices. Such coupling, however, is not trivial, because, for example, direct neuroanatomical connections between the primary motor and auditory cortices are rare, if they exist at all. Therefore, the connection can only be indirect and the detour to take on the articulatory-acoustic path would probably lead through more anterior frontal and additional superior temporal areas (Pulvermüller 1992). In contrast to the primary areas, these areas are connected to each other, as can be inferred from neuroanatomical studies in macaca (Deacon 1992; Pandya & Yeterian 1985). The coupling of acoustic and articulatory information will, therefore, involve additional neurons that primarily serve the purpose of binding linguistic information. Thus, the physical basis of lemmas may be distributed neuron populations including at least neurons in inferior prefrontal areas (in Brodmann's terminology areas 44, 45, and perhaps 46) and in the superior temporal lobe (anterior and posterior parts of area 22 and perhaps area 40). These neurons may not only be the basis of the binding of information about production and perception of language, they may well be the targets of connections linking the word form to its meaning, and, most important, their mutual connections may store syntactic information about a word. This proposal is consistent with the neurological observation that lesions in anterior or posterior perisylvian sites (but not in the inferior temporal lobe) frequently lead to a syntactic deficit called agrammatism (Pulvermüller 1995; Vanier & Caplan 1990).

Reciprocal, one-way, and numbered connections. Statistics of cortico-cortical connections suggest that two individual pyramidal neurons located side by side have a moderate (1–2%) probability of exhibiting one direct synaptic link, and only a very low probability of having two or more links (Braitenberg & Schüz 1991). Because synaptic connections are always one-way, a model for interaction of individual neurons may, therefore, favor one-way connections. Language mechanisms, however, are probably related to interactions of large neuronal populations, and if such ensembles include several thousands of neurons, chances are high that two ensembles exhibit numerous connections in both directions (Braitenberg & Schüz 1991). Hence the zero-assumption should probably be reciprocal connections between neuronal representations of cognitive entities such as lemmas, word forms, and their meanings.

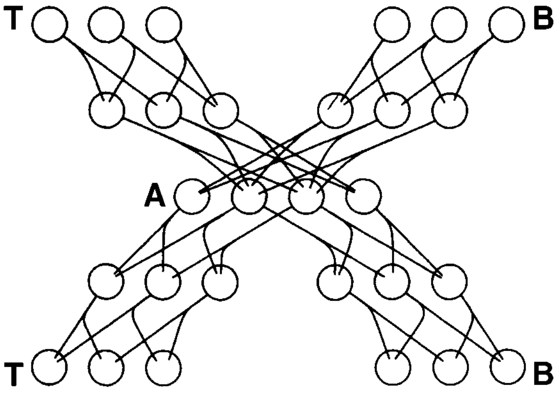


Figure 1 (Pulvermüller). Synfire chains possibly underlying the serial order of phonemes in the words "tab" and "bat." Circles represent neurons and lines their connections (the penetrated neurons being the ones that receive activation). The A (or /æ/-sound) is shared by the two words, but its neuronal counterparts are not identical but have overlapping representations, the non-overlapping neurons (leftmost and rightmost neurons in middle row) processing information about the sequence in which the phoneme occurs ("context-sensitive neurons"). Also the syllable-initial and syllable-final phonemes have distinct representations. If all neurons have a threshold of 3 and receive 1 input from their respective lemma node, selection of one of the word-initial phoneme representations (uppermost triplets) leads to a well-defined activation sequence spreading through the respective chain (but not through the competitor chain). Modified from Braitenberg and Pulvermüller (1992).

The model postulates reciprocal connections between semantic and syntactic representations (Fig. 2) and for some within-layer links (see, for example, Figs. 4, 6, and 7). Lemmas and word forms are connected through unidirectional links and within the form stratum there are directed and numbered links.

How could two large cortical neuron populations be connected in one direction, but lack the reciprocal link? Here are two possibilities: first, one-way connections could involve directed subcortical links (for example, through the striatum). As an alternative, connections could in fact (that is, neuroanatomically) be reciprocal, but activity flow during processing could be primarily in one direction. According to the present theory, the processes of naming include activity spreading from the conceptual to the lemma stratum, and from there to the form stratum whence backward flow of activity to the lemmas is prohibited. This could, for example, be due to early termination of the computation if the appropriate lemma is already selected before upward activity from the form stratum can influence the computation. Conceptualizing the process of selection as the full activation (ignition) of a lemma representation that leads instantaneously to equally strong activation of both conceptual and form representations of the same word, it could be stated that such ultimate activation makes additional activity flow impossible or irrelevant. A regulation mechanism (as detailed above) can be envisaged to suppress all activity in competing nodes as soon as selection has taken place (therefore accounting, for example, for the lack of priming of phonological relatives of words semantically related to the target (Levelt et al. 1991a if lemma selection occurs early). Activity flow primarily in one direction can still be accounted for in a system based on the economical assumption of reciprocal connections between large neuronal populations.

Numbered directed connections are proposed to link morpheme and phoneme nodes. Here, the brain-basis of the numbering needs to be specified. Again, there are (at least) two possibilities: first, different axonal conduction delays could cause sequential activation of phoneme nodes. This option has the disadvantage that differences in the delays would be hardwired in the network making it difficult to account for variations between speaking fast and speaking slow. The second alternative would suggest a slight modification of Levelt et al.'s model: phoneme nodes may receive input from morpheme nodes, but their sequence would be determined by connections between phoneme representations. Here, Abeles's (1991) concept of synfire chains comes to mind. A synfire chain is a collection of neurons consisting of subgroups A, B, C . . . with directed links from A to B, B to C, and so on. Each subgroup includes a small number n of neurons, $7 < n < 100$, and therefore, the assumption of one-way connections appears consistent with the statistics of cortical connectivity (Braitenberg & Schüz 1991).

Because phonemes can occur in variable contexts, it is not sufficient to assume that phoneme representations are the elements corresponding to the neuronal subgroups of the synfire chains in the phonological machinery (Lashley 1951). In order to distinguish the phonemes in "bat" and "tab," it is necessary to postulate that not phonemes, but phonemes-in-context are the elements of representation. Thus, the representation of a /æ/ following a /b/ and followed by a /t/ would be distinct from that of an /æ/ followed by a /b/ and preceded by a /t/ (cf. Wickelgren 1969). In addition, it has advantages to distinguish syllable-initial, central, and syllable-final phonemes, as suggested in the target article. The two /æ/s occurring in the words /tæb/ and /bæt/ could be neuronally organized as sketched in Figure 1. The selection of one of the chains would be determined (1) by activating input to all to-be-selected context-sensitive phonemes and (2) by strong input to the first neuronal element that initializes the chain. This proposal opens the possibility of determining the speed with which activity runs through the synfire chain by the amount of activation from the lemma and morpheme representations to context-sensitive phoneme representations.

Predictions about neurobiological mechanisms of language may be helpful for planning experiments in cognitive neuroscience and for interpreting their results. It is needless to say, however, that these considerations are at present necessarily preliminary, as pointed out in the target article, not only because the proposals may be falsified by future research, but also because they leave so many questions unanswered. For example, how is it possible to model multiple occurrences of a particular word (same form, same syntax, same meaning) in a given sentence? A not so attractive possibility would be that there are multiple representations for every word type in the processing model or its neurobiological counterpart. Other solutions may make the models much more complicated. Although it is clear that we can, at present, only scratch the surface of lexical processes in the brain, Levelt, Roelofs & Meyer's target article clearly evidences that the insights obtained so far are worth the scientific enterprise.

ACKNOWLEDGMENTS
This work is supported by grants from the Deutsche Forschungsgemeinschaft. I am grateful to Valentino Braitenberg and to Bettina Mohs for comments on an earlier version of the text.

Decontextualised data IN, decontextualised theory OUT

Benjamin Roberts,[a] Mike Kalish,[a] Kathryn Hird,[b] and Kim Kirsner[a]
[a]*Psychology Department, University of Western Australia, Perth 6907, Australia;* [b]*School of Speech and Hearing Science, Curtin University of Technology, Perth 6845, Australia.* **benjamin,kalish,kim@psy.uwa.edu.au**
www.psy.uwa.edu.au **thirdkm@alpha1.curtin.edu.au**
www.curtin.edu.au/curtin/dept/shs

Abstract: We discuss our concerns associated with three assumptions upon which the model of Levelt, Roelofs & Meyer is based: assumed generalisability of decontextualised experimental programs, assumed highly modular architecture of the language production systems, and assumed symbolic computations within the language production system. We suggest that these assumptions are problematic and require further justification.

The model proposed by Levelt, Roelofs & Meyer depends critically on the proposition that the language production system is composed of autonomous processing components whose internal operations are not influenced by the internal operations of other components. This assumption was explicit in Levelt (1989) and is implicit in the article under commentary. The target article is concerned with one such component, lexical access. The first point in our analysis concerns the intrinsic limits of decontextualised data. Implementation of naming and lexical decision experiments involving isolated words can only yield evidence about the way in which isolated words are produced in response to impoverished experimental conditions. If the way in which a given word is accessed and uttered is sensitive to pragmatic and other contextual variables, single-word test procedures will not reveal this because the critical variables are not active under these conditions. A person who is faced with a stimulus such as *heavy* does not have any communicative intent, experimental compliance excepted. A second problem with decontextualised presentation is that if the management of planning and production processes depends on the distribution of limited cognitive resources or some other form of coordination, it will be utterly masked.

It is our contention that Levelt et al. have overgeneralised the (in many ways attractive) notion of modularity. It is as if language production were a pure output process where, following the establishment of intent, later or lower processes are implemented without influence from, or reference to, the central system. But there is an alternative point of view. According to one alternative, intention and subsidiary processes such as lexical selection and premotor planning occur in parallel, and intention is superimposed on the utterance as the detailed motor program is formed. Models of this type cannot even be tested by experiments such as those described by Levelt et al., an omission that is strange given Levelt et al.'s established interest in planning (Levelt 1989).

For example, if it is granted that communicative intention including topic selection defines the input for lexical selection, and it is also assumed that input to the lexical process is restricted to conceptual information, how are mechanisms such as resetting and declination implemented in language production? They cannot be treated as part of the conceptual input to the lexicon, yet they must influence production through variation in one or more speech parameters – fundamental frequency, amplitude, or duration – in ways that influence the articulatory code. It is our contention that a system in which input into the lexical system is restricted to conceptual information and output is restricted to an articulatory code specifying the phonological form of the selected lexeme is untenable without provision for moderation from higher processes. Put in other words, transformation of the prosodic form of a given word must occur while the articulatory form of the word is being generated. The processes controlling higher order processes cannot simply provide conceptual information; they must intrude during the process of lexical formation to define the correct prosodic form.

Regarding Levelt et al.'s model as it stands, it is not clear how the information from the level of conceptualisation is integrated with lexical selection prior to articulation. Evidence from studies of brain damaged subjects, involving both higher order processes as well as all levels of the motor pathway suggest that the articulatory code for each lexical item is influenced by its significance in communicating the intention of the speaker, and therefore its position in an intentional planning unit. For example, Hird and Kirsner (1993) found that although subjects with damage to the right cerebral hemisphere were able to articulate the subtle acoustic difference required to differentiate noun phrases and noun compounds at the end of a carrier phrase, acoustic analysis revealed that they produced significantly shorter durations for all targets in comparison to normal, ataxic dysarthric, and hyperkinetic dysarthric subjects. The authors concluded that the right-hemisphere-damaged subjects, in keeping with their reported difficulty in processing pragmatic information, were not using pragmatic prosody to communicate to the examiner the importance of the target stimuli. That is, the other subjects increased the overall duration of both the noun phrase and the noun compounds in order to highlight their importance in the utterance. This example highlights the possibility that information derived from the level of conceptualisation concerning the intention of the speaker has direct influence on the way in which the articulatory code for each lexical item is specified. It may be suggested, therefore, that while Levelt et al.'s model might provide a valid description of the decontextualised left hemisphere, solving specific lexical retrieval requests on demand, it cannot explain pragmatic control as it is observed in the undamaged population.

A final concern amplifies the problem of decontextualisation. Levelt et al. have adopted a strictly formal theory of language, combined with a neo-Piagetian model of strictly discontinuous development. The likelihood of either of these extreme views being true in isolation is small; the likelihood of both is vanishingly small. Formal syntaxes can be developed for grammatical sentences, but this does not mean that a child necessarily uses one. Words have syntactic roles, but this does not mean that lemmas exist independently of lexemes. If we treat psychology seriously as a science of natural computation, then we should search for evolutionally plausible explanations. The structure Levelt et al. develop is impressive: a towering construction of symbols cemented together with rules and grammars. But this construction is artificial: it is a building, not a biological system. An explanation of language production must deal explicitly with growth (see Locke 1997) and with behaviours that exist in time and space (Elman 1995). Cognition may be simple as Levelt et al.'s model implies, and it may involve the mechanical application of rules to symbols, but it is our contention that cognition has passed the point where formal models can proceed without justifying their assumptions.

Constraining production theories: Principled motivation, consistency, homunculi, underspecification, failed predictions, and contrary data

Julio Santiago[a] and Donald G. MacKay[b]

[a]*Departamento de Psicología Experimental y Fisiología del Comportamiento, Universidad de Granada, 18071 Granada, Spain;* [b]*Psychology Department, University of California at Los Angeles, Los Angeles, CA 90095-1563.*
mackay@psych.ucla.edu santiago@platon.ugr.es

Abstract: Despite several positive features, such as extensive theoretical and empirical scope, aspects of Levelt, Roelofs & Meyer's theory can be challenged on theoretical grounds (inconsistent principles for phonetic versus phonological syllables, use of sophisticated homunculi, underspecification, and lack of principled motivation) and empirical grounds (failed predictions regarding effects of syllable frequency and incompatibility with observed effects of syllable structure).

Unlike many language production theories, Levelt, Roelofs & Meyer's theory addresses reaction time as well as speech-error data, and focuses on rarely examined theoretical issues such as the form of the morphology–phonology interface and phenomena such as the resyllabification across word boundaries that occurs during informal speech. This makes Levelt et al.'s theory one of the most comprehensive computational models available for understanding intentional word production. However, the present commentary will focus on some theoretical and empirical problems in hopes of stimulating further development of the models.

Theoretical problems. Four types of theoretical problems stand out: inconsistent principles, lack of principled motivation, sophisticated homunculi, and underspecification.

1. *Inconsistent principles.* Consistency is important in logical devices (see MacKay 1993), but Levelt et al.'s theory seems to harbor several inconsistent principles. The behavior of phonetic versus phonological syllables provides an example. Under the theory, phonetic syllables are stored as rote units in a "syllabary" within the internal lexicon because a small number of frequently used syllables suffices for generating most speech in most languages. However, this "frequency argument" applies with equal or greater force to phonological syllables, which are even more abstract and invariant or uniformly practiced than phonetic syllables. So why are phonetic syllables called up from a rote store, whereas phonological syllables are computed on-the-fly using ordered, language-specific syllabification rules? What principle prevents phonological syllables from becoming unitized or chunked in the same way as lexical concepts at the message level and phonetic syllables at the phonetic level? In the absence of some new (compensatory) theoretical principle, consistency requires the same unitizing process throughout the model.

2. *Lack of principled motivation.* A related problem is that Levelt et al.'s theoretical mechanisms often seem to lack principled motivation. An example is the way spreading-activation combines with binding-by-checking mechanisms in the model. The main function of spreading-activation mechanisms is to automatically integrate multiple sources of information prior to competition and selection (see Mackay 1987, pp. 127–32), and the binding-by-checking mechanisms is said to serve two additional functions: to check whether a selected node links to the right node(s) one level up and to prevent erroneous selection of a wrong (nonintended) node. However, these binding-by-checking functions render the integration and competition aspects of spreading activation superfluous: if the binding-by-checking mechanism knows the correct node all along, why doesn't it just select it in the first place without the complex spreading-activation, competition, and subsequent checking process? And if binding-by-checking can prevent erroneous selection, what is the purpose of the output monitor, a second type of checking mechanism postulated by Levelt et al.? We need some principle in the theory that limits how many checking mechanisms check checking mechanisms.

The issue of principled motivation also applies to what the authors take as support for their theory. For example, the hypothesis that abstract-syllable structures are stored in the brain independently of particular phonological units has received a substantial body of recent support from a variety of tasks (e.g., Meijer 1994; 1996; Romani 1992; Sevald & Dell 1994; Sevald et al. 1995), but not from implicit priming tasks (Roelofs & Meyer 1998). Because Levelt et al. assume that implicit priming results provide the true picture, they need to provide principled reasons for ignoring the wide variety of other relevant tasks and results in the literature.

3. *Sophisticated homunculi.* Sophisticated homunculi abound in Levelt et al.'s theory. For example, their binding-by-checking mechanism resembles the watchful little homunculus that scrutinized the phonological output for speech errors prior to articulation in prearticulatory editor theories (e.g., Baars et al. 1975). The main difference is that binding-by-checking involves a large number of watchful little homunculi, one for each node in each system (see MacKay 1987, pp. 167–74).

4. *Underspecification.* Levelt et al.'s theory suffers from underspecification in many important areas. One of these concerns the nature and processing characteristics of morphological, phonological, and phonetic output buffers. This is a curious omission because these buffers are said to play an essential role in many aspects of language production, for example, temporal coordination of different production stages, support for the suspension/resumption and look-ahead mechanisms, and the generation of appropriate (and inappropriate) serial order (an extremely general problem in itself; see MacKay 1987, pp. 39–89). These buffers are even said to be important for understanding why at least one full phonological word must be prepared before onset of pronunciation can begin. Without further specification, however, this amounts to an assumption for explaining an assumption. Unless the nature and characteristics of these output buffers are spelled out in detail, Levelt et al.'s theory lacks explanatory adequacy.

Empirical problems: Failed predictions and contrary effects. Levelt et al.'s theory suffers from two types of empirical problems: failed predictions and incompatibility with already observed effects. An example of failed prediction concerns the absence of syllable frequency effects in current data. These missing syllable frequency effects are problematic for the assumption that frequency used phonetic syllables become unitized in a mental syllabary. Even if we assume that unitization precludes further increases in processing speed with practice, between-subject differences in degree of use should cause a frequency-related gradient in production latencies because between-subject variation and syllable frequency variation will determine which subjects in an experiment will exhibit asymptotic production latency for any given syllable. Consequently, Levelt et al. predict longer production latencies for lower frequency syllables (stored in the syllabaries of relatively few subjects) than for higher frequency syllables (stored in the syllabaries of most subjects), contrary to present results.

The effects of syllable structure on vocal production times in Levelt et al.'s theory illustrate a second type of empirical problem: incompatibility with already observed effects. Santiago (1997; Santiago et al., submitted) showed that words starting with an onset cluster (e.g., *bread*) have longer production latencies than otherwise similar words starting with a singleton (e.g., *bulb*), a phonological encoding effect that seems incompatible with Levelt et al.'s assumption that an entire phonological word must be prepared before onset of pronunciation can. To account for these and other data (e.g., Costa & Sebastian 1998), Levelt et al. must either drop this assumption and postulate hierarchic syllable structures (in which consonant clusters constitute a complex processing unit) or suggest some other way in which consonant clusters may increase the time to begin to say an about-to-be pronounced word.

Lemma theory and aphasiology

Carlo Semenza,[a] Claudio Luzzatti,[b] and Sara Mondini[c]

[a]*Department of Psychology, University of Trieste, 34123 Trieste, Italy;* [b]*Department of Psychology, School of Medicine, University of Milan, 20134 Milan, Italy;* [c]*Department of Psychology, University of Padua, 35100 Padua, Italy.* **semenza@uts.univ.trieste.it** **luzz@imiucca.csi.unimi.it** **mondini@ux1.unipd.it**

Abstract: Recent aphasiological findings, not mentioned in the target article, have been accounted for by Levelt et al.'s theory and have, in turn, provided it with empirical support and new leads. This interaction is especially promising in the domain of complex word retrieval. Examples of particular categories of compounds are discussed.

Levelt, Roelofs & Meyer list all the empirical sources supporting their model and in particular the lemma/lexeme dichotomy (sect. 5.4.5). However, late in the paper (sect. 11), even though they mention the relevance of their theory for neuropsychological purposes (claiming that their model should be and has been used in functional correlative brain imaging studies), they do not seem to consider neuropsychological findings as prima facie relevant evidence for the validity of the model itself. Yet neuropsychological cases not only provide a unique opportunity to test the validity of information processing models but might have intrinsic heuristic value, since counterintuitive findings may force crucial developments of a model. This is particularly true in the domain of word retrieval, where a number of aphasiological studies have been directly inspired by Levelt et al.'s theory and, in turn, provide even more compelling evidence than error studies in normal subjects. Furthermore, hypotheses based on particular instances of word retrieval that are only schematically proposed by Levelt et al. have recently been supported by empirical evidence from aphasiological studies and may provide grounds for some theoretical advancement.

Many aphasic phenomena can be and have been explicitly interpreted as the result of a disconnection between the lemma and the lexeme level (e.g., Badecker et al. 1995, the only paper that Levelt et al. actually mention; Hittmair-Delazer et al. 1994; Zingeser & Berndt 1990; see Semenza, in press, for a review). It is in the domain of processing morphologically complex words, however, that the interaction between Levelt et al.'s model and neuropsychology seems more promising.

An observation repeatedly reported in aphasiology is the dissociation between impaired access to the phonological form of a compound word and preserved knowledge of the fact that the inaccessible word is a compound (Hittmair-Delazer et al. 1994; Semenza et al. 1997). This preserved morpholexical knowledge can only be explained by a preservation of the lemma and indicates that the lemma also contains morpholexical specifications of a word.

Another interesting finding was reported by Semenza et al. (1997): Broca's aphasics, a subgroup of patients known for their inability to retrieve function words, and often verbs, but who can normally retrieve nouns, often drop the verb component of verb–noun compounds (the most productive type of compound nouns in Italian). This has been interpreted as evidence of decomposition at some point of processing before the lexeme but after the lemma level, which must be undecomposed because most compounds are opaque in meaning.

Delazer and Semenza (1998) have also described a patient with a specific difficulty for naming compound and not simple words. The patient tended to name two-word-compound targets by substituting one or both components (for example *pescecane,* shark, (literally, fish-dog), with *pescetigre,* fish tiger, a neologism). The phenomenon was interpreted as reflecting an inability in retrieving two lemmas with a single lexical entry (corresponding to a single concept). This interpretation was put forward to account for the fact that there were no comprehension deficits and that the wrong answer was nevertheless an acceptable description of the target concept and could not apply properly to any of the two components separately.

The retrieval of a further type of Italian compound, the prepositional compound (PC), has recently been studied by Mondini et al. (1997). This is a typical and very productive noun–noun modification in Italian, where a head noun is modified by a prepositional phrase (e.g., *mulino a vento,* wind-mill). In these compounds the linking preposition is usually syntactically and semantically opaque (e.g., *film* in *bianco e nero,* black and white movie, vs. *film* a *colori,* colour movie). This is a compelling reason to consider PCs as fully lexicalized items, likely to be processed as a unit. Mondini et al.'s agrammatic patient, when tested over a series of tasks including picture naming, naming on definition, repetition, reading aloud, and writing on dictation, committed a significant number of omissions and substitutions on the linking preposition of PCs. If PCs were processed as units, agrammatism would not affect their retrieval. The following events are therefore proposed for retrieval of PCs: (1) a single lexical entry corresponding to the PC is activated after the conceptual level; but (2) the activation concerns the syntactic function of the lemma corresponding to the whole PC and, at the same time, the independent lemma representations bearing the lexical and syntactic aspects of each component of the compound.

A final interesting case supporting Levelt et al.'s theory has been reported outside the word retrieval domain. A patient described by Semenza et al. (1997) showed a selective impairment in using the grammatical properties of mass nouns. This case demonstrates how the grammatical rules said to be stored at the lemma level are indeed independently represented and accessible.

What about phonological facilitation, response-set membership, and phonological coactivation?

Peter A. Starreveld and Wido La Heij

Unit of Experimental and Theoretical Psychology, 2333 AK Leiden, The Netherlands. **starreve@rulfsw.leidenuniv.nl** **lahey@rulfsw.leidenuniv.nl** **www.fsw.leidenuniv.nl**

Abstract: We discuss two inconsistencies between Levelt et al.'s model and existing theory about, and data from, picture-word experiments. These inconsistencies show that the empirical support for WEAVER++ is much weaker than Levelt et al. suggest. We close with a discussion of the proposed explanation for phonological coactivation of near synonyms of a target word.

Starreveld and La Heij (1995; 1996b) showed that semantic and phonological context effects interact in picture naming and argued that this finding is at variance with strict-serial models of language production like the one that is presented in Levelt et al.'s target article. In a comment, Roelofs et al. (1996) argued that this conclusion is not warranted because "existing conceptions of serial access" (p. 246) assume that in the picture-word task phonological similarity has an effect at both the word-form and lemma levels. In the target article, Levelt et al. repeat this argument (paras. 3.2.4 and 5.2.3), although, again, no empirical evidence from picture-word studies is presented that substantiates this claim. Here we elaborate on this discussion and present some other troubling aspects of WEAVER++.

To make predictions about word-context effects in picture naming, serial-stage models of language production have to make assumptions about the processing of the distractor word.

Schriefers et al. (1990) made the explicit assumption that semantic similarity between picture and distractor affects only the lemma level and that phonological similarity affects only the word-form level. In accordance with a strict-serial model, their results showed an early semantic effect and a late phonological effect. This finding is one of the highlights in the language production literature: It is widely cited as empirical evidence for (1) the lemma/word-form distinction and (2) the strict-serial nature of language processing.

The assumption that phonological similarity affects only the word-form level was made not only by Schriefers et al. (1990), but also by Meyer (1996), Meyer and Schriefers (1991), Roelofs (1992b; 1997c), and Zwitserlood (1994). As a result, Roelofs et al.'s (1996) claim (as repeated by Levelt et al.) – that phonological similarity also has a substantial effect at the lemma level – has profound consequences for both the interpretation of these studies and an evaluation of WEAVER++.

First, the position of the Schriefers et al. (1990) study becomes very awkward: (1) apparently, the rationale behind that paper was flawed and (2) the results – the absence of concurrent phonological and semantic effects – clearly contradict Levelt et al.'s present claim. Exactly the same problems arise for Meyer's (1996) study. She also assumed that phonological similarity affects only word-form encoding and her results and interpretation with respect to the production of a second noun in short expressions paralleled those of Schriefers et al. (1990). Third, Meyer and Schriefers (1991) used the assumption that phonological similarity affects only the word-form level to test models of phonological encoding. However, if phonological similarity also affects lemma retrieval, then these tests were invalid. Finally, if phonological similarity also affects lemma retrieval, part of the results reported by Zwitserlood (1994) are in need of reinterpretation. So, if Levelt et al. maintain this claim, the findings obtained in these studies clearly need to be reexamined and explained.

What are the implications for WEAVER++? In paragraph 6.4.1, Levelt et al. report simulations of the phonological effects reported by Meyer and Schriefers (1991). Given Levelt et al.'s claim that phonological similarity also affects lemma retrieval, the lemma-level part of the model should have been used in these simulations. However, despite the fact that the simulated phonological effects spanned a broad SOA range (from −150 msec to 150 msec), only the word-form level part of the model was used (see also Roelofs 1997c). Given this omission, the results of these simulations cannot be viewed as evidence in favor of WEAVER++. Furthermore, because these simulations were used to calibrate the model, the parameter values used in the other simulations that involved the word-form level most probably have to be adjusted.

We conclude from the above that, with respect to their claim that phonological similarity affects lemma selection, Levelt et al. (1) deviate from claims about the locus of phonological facilitation in the studies mentioned above, (2) fail to discuss the profound implications of this deviation for these earlier studies, some of which had seemed to provide support for strict-serial models, and (3) have not incorporated their claim in the simulations of WEAVER++. As things stand, the only function of the author's lemma-level assumption seems to be to account for the interaction reported by Starreveld and La Heij (1995; 1996b). In our view, these observations seriously undermine the credibility of Levelt et al.'s present proposal.

Our second issue concerns Levelt et al.'s assumption that no semantic effects should be found in the picture-word task when distractors are used that do not belong to the response set (Roelofs 1992a; but see La Heij & van den Hof 1995, for an alternative account of Roelofs's findings). In the lemma-level simulations of WEAVER++, Levelt et al. implemented this response-set assumption: lemmas that were part of the response set were flagged, and only flagged lemmas were competitors of the target node (para. 5.2.1). However, several studies in which the distractors were not part of the response set showed clear semantic interference effects (e.g., La Heij 1988; Schriefers et al. 1990; Starreveld & La Heij 1995; 1996b). Roelofs (1992b) presented tentative explanations for such findings in terms of mediated priming effects when several response-set members belong to the same semantic category and/or in terms of relative weight-changes in the Luce ratio of the lemma retriever's selection rule when the ratio of semantically related and unrelated stimulus pairs deviates from 1:1.

However, we argued earlier (Starreveld & La Heij 1996b) that an explanation of these findings in terms of mediated effects (a non-response-set lemma primes a response-set lemma via their corresponding conceptual nodes, i.e., via three links) is incorrect, because, given the parameter values for the links in WEAVER++, the resulting additional activation for the response-set lemma is negligible. Hence, the semantic effects reported by Starreveld and La Heij (1995; 1996b) – who used the same number of semantically related and semantically unrelated distractors – cannot be explained this way, and, consequently, these semantic effects present clear experimental evidence against Levelt et al.'s response-set assumption. In addition, La Heij and van den Hof (1995) and La Heij et al. (1990) also reported semantic interference effects that are difficult to account for by the tentative explanations proposed by Roelofs (1992b).

We conclude that Levelt et al.'s response-set assumption is invalid. Hence no simulations that depend heavily on this assumption should be taken as evidence in favor of WEAVER++. These simulations are the ones presented in Figure 4, in panels B and C and in part of panel D. Again, the reported evidence for WEAVER++ is much weaker than the authors suggest.

Finally, we discuss Levelt et al.'s proposal with respect to recent findings of phonological coactivation of nontarget words. The authors review evidence that, in addition to the target word, near synonyms of the target word get phonologically encoded as well. To incorporate these findings in their strict-serial model, the authors propose a new mechanism: When two lemma nodes are activated to a virtually equal level, both get selected and phonologically encoded (para. 6.1.1 and sect. 10).[1] Although at first sight this might seem a reasonable assumption, it undermines the essence of strict-serial models. In addition, implementing the assumption in WEAVER++ necessitates ad hoc modifications both at the lemma level and at the word-form level in the model. In the present version, the model uses not only the spreading activation process and signaling activation to select a target node at the lemma level, but it requires three additional mechanisms. First, the source of the activation (picture or word) is indicated by *tags*; second, members of the response-set receive *flags*; and third, only one node can be a target node (the first flagged node that receives picture-tagged activation). To allow multiple lemma selection in the case of a near synonym of the target, at least two modifications of this already very complex machinery are necessary. Now a node that is not a member of the response-set (i.e., the node of the near synonym) should also be flagged, and the selection rule should allow two nodes to be target nodes. In addition, these modifications should only operate when a target has a near synonym, because they would disrupt "normal" processing.

Furthermore, participants produce the target name quickly and accurately, including when a target has a near synonym. So if multiple lemmas are selected and phonologically encoded then the word-form part of the model must perform two additional tasks. It has to ensure that no blends of the target and the near synonym result (although the simultaneous selection of two lemmas is the proposed mechanism for the occurrence of blends; sect. 6.1.1 and sect. 10), and it has to select the target while preventing the phonologically encoded near-synonym from being articulated. The latter additional task especially opposes the essence of a strict-serial (modular) model. The advantage of a modular model is that it allows for the clear separation of jobs, so that each module can specialize in the performance of a particular task: the lemma module selects the targets and the phonological module encodes the targets. This functional separation breaks down if the phonological module must select targets too.

In conclusion, many additional mechanisms, both at the lemma and word-form levels are necessary to allow the phonological coactivation of near synonyms. In our view, it is questionable whether a model that depends on such a large number of ad hoc mechanisms really provides a clear insight in the processes that underlie language production.

ACKNOWLEDGMENT
In writing this commentary, Peter A. Starreveld received support from the Dutch Organization for Scientific Research (Grant 575-21-006).

NOTE

1. It is interesting to note that in the picture–word task, precisely this situation arises: two lemmas, that of the picture's name and that of the distractor, are highly activated.

Contact points between lexical retrieval and sentence production

Gabriella Vigliocco[a] and Marco Zorzi[b]
[a]*Department of Psychology, University of Wisconsin–Madison, Madison, WI 53706;* [b]*Department of Psychology, University of Trieste, 34123 Trieste, Italy.*
gviglioc@facstaff.wisc.edu
psych.wisc.edu/faculty/pages/gvigliocco/GV.html
zorzi@uts.univ.trieste.it
www.psychol.ucl.ac.uk/marco.zorzi/marco.html

Abstract: Speakers retrieve words to use them in sentences. Errors in incorporating words into sentential frames are revealing with respect to the lexical units as well as the lexical retrieval mechanism; hence they constrain theories of lexical access. We present a reanalysis of a corpus of spontaneously occurring lexical exchange errors that highlights the contact points between lexical and sentential processes.

We retrieve words in order to communicate and we communicate using sentences, not (or very rarely) words in isolation. The lexical retrieval mechanism and sentence building machinery need to be coordinated. Errors in assigning lexical entries in a sentential frame can be revealing with respect to the lexicalization process and vice versa: errors in lexical retrieval may suggest ways in which sentence construction is controlled by lexical structures.

Levelt, Roelofs & Meyer discuss how lexical retrieval affects sentence construction (sect. 5.4); however, they do not draw the corresponding implications with respect to how sentence construction constrains lexical retrieval. The danger of this approach is explaining phenomena that are paradigmatically related to sentence level processes solely in terms of lexical properties. The goal of this commentary is to fill this gap, providing constraints for Levelt et al.'s theory on the basis of observations of slips of the tongue. We present a reanalysis of a corpus of speech errors in Spanish (del Viso et al. 1987) focusing on *lexical exchanges.* An example of an exchange error in English (from the MIT corpus; Garrett 1980) is reported below.

> Error: How many *pie*s does it take to make an *apple*?
> Intended: How many apples does it take to make a pie?

In the example, "apple" and "pie" swapped position. The plural marking on the target "apple" does not move with the lexical stem; it is stranded. These errors reflect misassignment of lexical elements into sentential frames.

The rationale for considering Spanish is that Spanish is a highly inflected language that allows us to better assess the morphological status of the exchanged units. Exchange errors come in five different flavors in the Spanish corpus (total number of errors considered = 134). Examples in the different categories are provided in Table 1. Errors, such as those in categories 1–4, involve nouns; our argument is built around what happens to features shared by the nouns and the determiners (such as number and gender) when the exchange occurs. Errors in category 5 instead involve units from different grammatical categories (more precisely an adjective and an adverb). Let us consider the units involved in the different exchanges.

In *phrasal exchanges,* both the nouns and the determiners (i.e., the whole noun phrases) move together. In *word exchanges,* the nouns with their bound inflectional morphology exchange, leaving behind the determiners fully inflected for the targets. In *stem exchanges,* only the word stems move, leaving behind bound inflections (in the example, "number" does not move with the lexical stem). *Ambiguous exchanges* could be described as phrasal, word, or stem exchanges. In examples such as the reported one, the two nouns share number and gender, making a more precise classification impossible. Finally, while for phrasal, word, and stem exchanges the two exchanging elements share the same grammatical category and belong to separate syntactic phrases, in *morpheme exchanges* the two units do not share grammatical cat-

Table 1 (Vigliocco & Zorzi). *Examples of different exchange errors in Spanish*

Error	Target
1. Phrasal Exchange	
. . . *las chicas* de *la cara* estan . . .	. . . la cara de las chicas esta . . .
(the-F,P girl-F,P of the-F,S face-F,S are)	(the-F,S face-F,S of the-F,P girl-F,P is)
2. Word Exchange	
. . . la *corte* del *imagen* Ingles	. . . la imagen del corte ingles
(the-F,S cut-M,S of the-M,S image-F,S English)	(the-F,S image-F,S of the-M,S cut-M,S English)
3. Stem Exchange	
Pasame las *tortillas* para la *patata*	. . . las patatas para la tortilla
(Pass me the-F,P omelette-F,P for the-F,S potato-F,S)	(. . . the-F,P potato-F,P for the-F,S omelette-F,S)
4. Ambiguous	
. . . le han dedicado *periodicos* sus *editoriales*	. . . le han dedicado editoriales sus periodicos
(to-her have devoted periodicals-M,P on editorials-M,P)	(. . . editorials-M,P on periodicals-M,P)
5. Morpheme Exchange	
. . . un efecto *significativa*mente *estadistico*	. . . un efecto estadisticamente significativo
(an-M,S effect-M,S significantly statisticant-M,S)	(an-M,S effect-M,S statistically significant-M,S)

Note: In the English translations, F = feminine; M = masculine; S = singular; P = plural.

egory ("estadisticamente" is an adverb; "significativo" is an adjective) and are in the same phrase. Note that stem and morpheme exchanges involve the same units (a word stem) but differ in terms of their syntactic environment.

There were 134 relevant errors in the corpus. Of these, 21 were phrasal exchanges; 1 was a word exchange; 41 were stem exchanges; 49 were ambiguous cases; and finally 22 were morpheme exchanges.

A very important observation is that *word* exchanges are virtually absent. The example reported above is the only one we found in the corpus. This suggests that fully inflected words are not lexical units of encoding.

What are the implications of these observations for the model of lexical retrieval proposed by Levelt et al.? In their discussion of exchange errors, Levelt et al. (sect. 5.4.5) assume a distinction between word and morpheme exchanges and attribute the first to lemma level and the second to word form level. We agree that errors reflect both lemma and word form level representations. Furthermore, we agree with their analysis of morpheme exchanges. However, Levelt et al. conflate under "word exchanges" three types of errors: phrasal, word, and stem exchanges. We believe it is important to separate them.

Both phrasal and stem exchanges involve lemmas, but they occur during two different processes. Phrasal exchanges would arise because of a misassignment of grammatical functions to lemmas. In the example in Table 1, "cara" and "chica" are "subject" and "modifier," respectively, in the target utterance, but their functions are swapped in the error (for a similar treatment see Bock & Levelt 1994). Hence these errors reflect the information flow from a "message level" sentential representation to lemmas. Stem exchanges, instead, would reflect a mistaken insertion of lemmas into frames that specify inflectional morphology.

The contrast between phrasal and stem exchanges suggests a system in which grammatical functions are first assigned to lemmas. Next, a syntactic frame would be initiated on the basis of these functions. The frame would specify inflectional morphology on the basis of the message (for conceptual features such as number) as well as on the basis of the specific lemmas (for lexical features such as gender).[1] Lemmas would finally be inserted in these frames. Hence features such as number and gender for nouns would be specified during the construction of the corresponding noun phrases, not at a lexical level. The absence of true word exchanges in the corpus strengthens this hypothesis. If these features were specified at a lexical level, we should have observed a larger number of word exchanges. In this account, inflections would be assigned to a developing syntactic frame when lemmas are selected and "tagged" for grammatical functions *before* stem exchanges occur, not *after* as Levelt et al. argue.

To sum up, our analysis highlights further important sentential constraints on lexical retrieval that Levelt et al. need to take into account to model lexical retrieval during connected speech.

ACKNOWLEDGMENT

Support was provided by a NSF grant (SBR 9729118) to the first author. We would like to thank Ines Anton-Mendes, Merrill Garrett, Rob Hartsuiker, and Jenny Saffran for their comments.

NOTE

[1] The picture may actually be more complex, and conceptually motivated features may be treated differently from lexically motivated features. Interestingly, there are a few (= 3) cases in the corpus when number is stranded while grammatical gender moves with the word (and the determiners agree with the gender of the error, not the target). However, a substantially larger number of cases would be necessary to see whether conceptual and lexical features behave differently.

Competitive processes during word-form encoding

Linda R. Wheeldon

School of Psychology, University of Birmingham, Birmingham B15 2TT, United Kingdom **l.r.wheeldon@bham.ac.uk**

Abstract: A highly constrained model of phonological encoding, which is minimally affected by the activation levels of alternative morphemes and phonemes, is proposed. In particular, WEAVER++ predicts only facilitatory effects on word production of the prior activation of form related words. However, inhibitory effects of form priming that suggest that a more strongly competitive system is required exist in the literature.

Despite superficial similarities, the approach taken by Levelt, Roelofs & Meyer to the problem of word-form encoding differs fundamentally from previous attempts to model this process. This difference hinges on the attitude taken to noise in the system. The encoding of most words occurs within a sea of activated alternative morphemes and their constituent sublexical representations. Some of this activation comes from external sources such as the speech of others or written language. Some activation will be generated directly by the utterance to be produced (i.e., the residual activation of the morphemes just articulated and the anticipatory activation of those to be generated). There is also activation of words that do not form part of the intended utterance but are related in form to a word to be produced (e.g., malapropisms: saying "apartment" when you intended to say "appointment").

Traditionally, models of word-form encoding have used this riot of background activation to explain how speech errors can occur by incorporating selection mechanisms that are sensitive to the activation levels of alternative representations. In Dell's (1986) model, the selection of a given phoneme is determined by its level of activation during an ordinal selection process. Alternatively, inhibitory or competitive mechanisms have been proposed. Stemberger (1985) postulates inhibition between form-related lexical representations activated by feedback from shared phonemes. Sevald and Dell (1994) propose competition between activated phonemes during a left-to-right assignment process of segments to frames.

In contrast, Levelt et al. postulate no competitive selection process at the morpheme level and no feedback of activation from phonemes to morphemes. In addition, the selection by checking device employed in WEAVER++ divorces the selection and encoding of phonemes from their level of activation, thereby ensuring that encoding runs correctly regardless of the activation levels of alternative phonemes. WEAVER++ restricts competition to phonetic encoding processes but even here the activation of form related morphemes will speed encoding of the target by increasing the numerator of the access ratio for the target syllables compared with unrelated syllable programs.

The strict limitations on activation spreading and competition imposed by Levelt et al. allow them to successfully model the facilitatory effects of form similarity in the picture-word interference task and in the implicit priming paradigm. The cost, as they freely admit, is an unrealistically error-free encoding system; they suggest several additional error-generating mechanisms compatible with the model (sect. 10).

However, the evidence for competitive processes during word form encoding is not restricted to speech error data. A number of experimental paradigms have demonstrated inhibitory effects of form priming on speech production. In replanning studies, participants prepare a target utterance but on a small number of critical trials are cued to produce an alternative utterance. When the alternative utterance is related in form to the target, performance is slower and more error prone (O'Seaghdha et al. 1992; Yaniv et al. 1990). In tasks requiring speeded recitation of syllables, the number of correct repetitions is increased by the repetition of final consonants but decreased by the repetition of initial consonants (Sevald & Dell 1994). It could be argued that these tasks involve elements designed to create difficulty and that the inhibitory effects could be attributed to task-specific demands. However, inhibitory effects of form priming have also been observed in more straightforward speech production tasks. Bock (1987) found that when participants produce declarative sentence descriptions of simple pictures, they placed form primed words later in the sentence and were less likely to use them at all. And Wheeldon (submitted) has demonstrated a short-lived inhibitory effect of form priming on spoken word production (in Dutch) using a simple single word production task in which prime words are elicited by definitions and targets by pictures (e.g., Wheeldon & Monsell 1994). The prior production of a form-related word (e.g., HOED-HOND, hat-dog) slowed picture naming latency compared with the prior production of an unrelated word. This inhibitory priming was observed only when onset segments were shared (e.g., BLOED-BLOEM, blood-flower). Shared offset segments yield facilitation (e.g., KURK-JURK, cork-dress). No priming was observed when the prime and probe words had no mismatching segments (e.g., OOM-BOOM, uncle-tree; HEK-HEKS, fence-witch).

In its present formulation, WEAVER++ cannot account for the inhibitory form priming effects discussed above. Thus the strict limitation of competition that enables the simulation of facilitatory form priming effects has consequences for WEAVER++'s ability to model reaction time data as well as speech-error generation. It remains to be seen whether a future formulation of WEAVER++ can model both facilitative and inhibitory effects with equal success.

Compositional semantics and the lemma dilemma

Marco Zorzi[a] and Gabriella Vigliocco[b]

[a]*Department of Psychology, University of Trieste, 34123 Trieste, Italy;*
[b]*Department of Psychology, University of Wisconsin–Madison, Madison, WI 53706.* **zorzi@univ.trieste.it**
www.psychol.cul.ac.uk/marco.zorzi/marco.html
gviglioc@facstaff.wisc.edu
psych.wisc.edu/faculty/pages/gvigliocco/FV/html

Abstract: We discuss two key assumptions of Levelt et al.'s model of lexical retrieval: (1) the nondecompositional character of concepts and (2) lemmas as purely syntactic representations. These assumptions fail to capture the broader role of lemmas, which we propose as that of lexical–semantic representations binding (compositional) semantics with phonology (or orthography).

Theories of speech production distinguish between two different levels of lexical representations: *lemmas* (abstract representations of the words that *also* contain syntactic information) and *lexemes* (or word forms that specify morphological and phonological information) (Butterworth 1989; Dell 1986; Garrett 1980; Levelt 1989; Levelt et al.'s target article 1998; but see Caramazza 1997). A basic assumption of the theory developed by Levelt et al. is that concepts are represented as undivided wholes, rather than by sets of semantic features. This is implemented in their model by means of *concept nodes,* which are interconnected through labeled links to form a semantic network. Each node in the conceptual network is linked to a single lemma node. The role of the lemma nodes is to connect to syntactic information necessary for grammatical encoding. In this view, lemma nodes are a copy of the conceptual nodes, so that the only architectural distinction between the two levels of representation is the within-level connectivity.[1]

The objective of our commentary is to evaluate the claims by Levelt et al. with respect to the specification of conceptual and lemma level representations. We will present some arguments for compositional semantics and briefly sketch a view in which lemma-level representations would specify lexical semantic information in addition to being connected to syntactic features.

Compositional semantics. Levelt et al. argue that the nondecompositional character of conceptual knowledge in their model overcomes problems such as the so-called "hyponym–hyperonym" problem (sect. 3.1.1). If concepts are represented by sets of semantic features, the active features for a given concept (e.g., "chair") will include the feature set for all of its hyperonyms or superordinates (e.g., "furniture"). The inverse reasoning applies to the hyponym problem (i.e., the erroneous selection of subordinates). This problem (if it is a problem at all) is not peculiar to language production, but also arises in other domains. In visual word recognition, word forms are accessed from active sets of letters (just as lemmas are accessed from active sets of semantic features). When a word like "mentor" is presented, what prevents readers from accessing "men" or "me," which are formed from subsets of the active letters? Connectionist models of reading (e.g., Zorzi et al. 1998) present computational solutions. Even in a localist framework, the problem can be solved using networks trained with algorithms such as competitive learning (Grossberg 1976a; Kohonen 1984; Rumelhart & Zipser 1985). In competitive learning, the weights (w) of each output node are normalized, that is, all connection weights to a given node must add up to a constant value. This takes the selected node (winner) to be the node closest to the input vector x in the l_1 norm sense, that is

$$\left|\overline{w}_{i^0} - \overline{x}\right| \le \left|\overline{w} - \overline{x}\right|$$

where i^0 is the winning node. For the winning node, the weight vector is displaced towards the input pattern. Several distance metrics can be used, although the Euclidean is more robust. Therefore, the activation of the features for "animal" will not be sufficient for the node "dog" to win, because the links to "dog" will have smaller values than those to "animal" (assuming that the concept "dog" entails more semantic features than "animal"). Conversely, a concept like "chair" cannot activate the superordinate "furniture," because the number of active features (and hence the length of the input vector) for the two concepts is different (a similar solution to the problem is proposed by Carpenter & Grossberg [1987] in the domain of pattern recognition).

What is a lemma? If concepts are represented by sets of semantic features, these features must be "bound" to represent a lexical unit before any other kind of representation can be properly accessed. That is, an intermediate level of representation must exist between semantic features and the phonological (or orthographic) form of the word, because the mapping between meaning and phonology is largely arbitrary. This issue is generally known as the problem of *linear separability* (e.g., Minsky & Papert 1969; Rumelhart et al. 1986). In neural network models, the problem is typically solved using a further layer of nodes (e.g., hidden units) lying between input and output. It is important to note that nothing (except from the choice of learning algorithm) prevents the intermediate layer from developing localist representations. Lemmas provide exactly this kind of intermediate level of representation. If the lemma level has the role of binding semantic and phonological information, then lemmas have a definite content that is best described as lexical–semantic.

The organization of the lemma level will be largely dictated by the conceptual level (semantic features). For instance, the use of an unsupervised, self-organizing learning method (e.g., Kohonen 1984) will result in the lemma nodes being topographically organized to form clusters corresponding to semantic categories (Erba et al. 1998). Evidence compatible with this idea comes from the observation of "semantic field effects" in word substitution errors (see Garrett 1992). Further evidence comes from a study by Damasio et al. (1996). They reported converging evidence from a neuropsychological study on a large group of anomic patients and from a neuroimaging study on normal subjects that an intermediate level of representation, which they describe precisely as "binding" semantic and phonological information, is anatomically localized in the left temporal lobe. Crucially, they found that different categories (e.g., animates vs. artifacts) are clustered in distinct (but contiguous) cortical regions.

In this view, lemmas would also have a syntactic role. It is clear that syntactic properties cannot be directly attached to concepts, because semantic features do not directly map onto syntactic features. The syntactic properties could be attached to the phonological or orthographic word forms; however, this is computationally inefficient because syntactic information is modality-independent (but see Caramazza 1997). Therefore, the intermediate lemma level is the most adequate for accessing syntactic information.

Lexical concepts acquire syntactic properties relatively late in development (between the ages of 2.6 and 4 years; see Levelt et al., sect. 1). This process is termed *syntactization* by Levelt et al. and refers to the development of a system of lemmas. However, the explosive growth of the lexicon takes place between the ages of 1.6 and 2.6. This means that an efficient mapping between concepts and phonological word forms is already established at that onset of the syntactization process. Within the architecture of Levelt et al.'s model, such mapping would presumably involve conceptual nodes and word forms, thus bypassing the yet-to-be-developed lemmas. Therefore, the later development of the lemma level would mean a massive rewiring of the lexical system. We believe that such a process is truly unlikely (both from the neural and computational standpoints). By contrast, if lemmas develop as a necessary component of the mapping between meaning and phonology, syntactization is simply the process of linking syntactic features to the existing lemma representation.

ACKNOWLEDGMENT
Preparation of this paper was supported by NSF grant SBR 9729118 to Gabriella Vigliocco and by NATO grant CRG 971002 to Gabriella Vigliocco and Marco Zorzi.

NOTE

[1] This very restricted notion of lemma is what led Caramazza (1997) to argue that lemma nodes are contentless representations (the "empty lemma"), and as such they are dispensed with in his model of lexical access.

Authors' Response

Multiple perspectives on word production

Willem J. M. Levelt,[a] Ardi Roelofs,[b] and Antje S. Meyer[a]
[a]*Max Planck Institute for Psycholinguistics, 6500 AH Nijmegen, The Netherlands;* [b]*School of Psychology, University of Exeter, Washington Singer Laboratories, Exeter EX4 4QG, United Kingdom.* **pim@mpi.nl**
www.mpi.nl **a.roelofs@exeter.ac.uk** **asmeyer@mpi.nl**

Abstract: The commentaries provide a multitude of perspectives on the theory of lexical access presented in our target article. We respond, on the one hand, to criticisms that concern the embeddings of our model in the larger theoretical frameworks of human performance and of a speaker's multiword sentence and discourse generation. These embeddings, we argue, are either already there or naturally forgeable. On the other hand, we reply to a host of theory-internal issues concerning the abstract properties of our feedforward spreading activation model, which functions without the usual cascading, feedback, and inhibitory connections. These issues also concern the concrete stratification in terms of lexical concepts, syntactic lemmas, and morphophonology. Our response stresses the parsimony of our modeling in the light of its substantial empirical coverage. We elaborate its usefulness for neuroimaging and aphasiology and suggest further cross-linguistic extensions of the model.

R1. The larger context

R1.1. Lexical access in utterance generation

The stated aim of our target article was to present a theory of lexical access covering "the production of isolated prosodic words." Several commentators didn't like this straitjacket and preferred to consider lexical access in the larger pragmatic context of utterance generation. We share this preference, and a historical note may clarify our position. The larger framework was developed by Levelt (1989; henceforth *Speaking*), which outlines a theory of speaking, the skill that gets us from communicative intentions within ever-changing pragmatic settings to articulated utterances. A major observation in reviewing the relevant literatures was that a core aspect of this skill, lexical access, was theoretically deeply underdeveloped. Our team at the Max Planck Institute set out to fill this gap, and the target article reports on a decade of research dedicated to unraveling the process of normal lexical access. This required, among other things, the invention (by us and by others) of appropriate reaction time paradigms. Of course, the pragmatic context is limited in the laboratory (although none of our subjects ever said "this is not the real world"), but we always had the larger, explicit theoretical framework at hand. Several comments on our target article can be handled by referring to that framework.

Gordon addressed the issue of pronoun generation, indeed something we hardly touched on in the target article. He suggested that the primary purpose of unreduced referring expressions is "to introduce semantic entities into a model of discourse, whereas the primary purpose of pronouns (and other reduced expressions) is to refer directly to entities that are prominent in the discourse model" – almost a citation from *Speaking* (sect. 4.5.2). Meanwhile Schmitt (1997) in her Max Planck dissertation project, elaborated this notion in terms of a processing model and performed the relevant reaction time experiments.

Hirst's well-taken discussion of language-dependent conceptualization, as it may occur in bilinguals, is foreshadowed in section 3.6 of *Speaking,* referred to in section 3.1.2 of the target article. Unlike Hirst, though, we are not worried by the thought that many, or even most, lexical concepts are language-specific. The empirical evidence for such a notion is rapidly increasing (see, e.g., Slobin 1987; 1996; 1998). We agree that this is a special challenge for modeling the bilingual lexicon.

Ferreira correctly notes (and has demonstrated experimentally) that a word's prosody will vary from context to context. For instance, a word tends to be relatively long in phrase-final position or when it is contrastively stressed. Chapter 10 of *Speaking* outlines an architecture for the modulation of phonetics by such higher-level types of information. In that architecture there is indeed *parallel* planning of prosody and lexical access, as **Roberts et al.** would have it. These commentators could have read in section 10.3.1 ("Declination") of *Speaking* how declination and resetting are conceived within that framework. Indeed, these phenomena are not handled through conceptual input, and neither are the setting of amplitude and duration. It is curious to be confronted with a caricature of one's theoretical framework and subsequently to be accused of not justifying one's assumptions.

Another issue directly related to the larger sentential context of lexical access came up in **Vigliocco**'s **& Zorzi** commentary. They argued that stranding errors in a Spanish corpus suggest that the diacritics of lemmas are set before lemma exchanges occur, not afterwards, as we proposed in section 5.4.5 of the target article. We disagree. In our theory, the diacritics at the lemma level ensure that appropriately inflected forms are generated during word-form encoding. In word exchanges and in one of two types of stem exchanges, lemmas trade places. Whether a word or stem exchange occurs depends on the relative timing of filling the diacritical slots of the lemmas. When the diacritical slots are filled after the exchange, a stranding error such as "how many pies does it take to make an apple" (from Garrett 1988) occurs. That is, the diacritical slots get the wrong fillers. When the diacritics are set before the exchange, however, a word exchange occurs. Vigliocco & Zorzi argued that word exchanges were virtually absent in their Spanish corpus. However, because there were so many ambiguous errors, one cannot exclude that the actual number of word exchanges was much higher. Note that this controversy over the moment of the setting of diacritics does not affect our basic argument (laid out in sect. 5.4.5), which is that, besides word exchanges, there are two types of morpheme errors, supporting our distinction between a lemma and a form level of representation.

R1.2. Issues of attention and short-term storage

Carr is concerned that WEAVER++ suffers from a debilitating attentional disorder. We indeed do not include a general theory of attention covering how attention is directed towards a particular stimulus or location, how it is disengaged, and which cortical regions subserve it. However, contrary to Carr's claims, WEAVER++ is capable of directing attention to a word or picture in a picture–word interference experiment and to select a particular type of response, such as basic-level terms in naming and their hyperonyms in categorization. It is, in particular, not the case that the model selects the first word that gets fully activated. As explained by Roelofs (1992a; 1993), WEAVER++ keeps track of the external sources of activation (picture and word) in line with what Carr calls selection on the basis of the "ontological status" of the visual entities available for perception. In addition to tracing the activation sources, WEAVER++ marks in memory which words are possible responses. If there are no a priori constraints, the whole lexicon constitutes the response set for naming responses. Selection is based on the intersection of the activation from the target source (e.g., the picture in picture naming) and one of the marked responses. This selective attention mechanism generates the appropriate response in all task varieties discussed: naming pictures (or categorizing them using hyperonyms) while ignoring distractor words and categorizing words while ignoring distractor pictures.

Roelofs (1992a) showed how, with this attention mechanism, WEAVER++ explains not only correct performance in these tasks but also the intricate patterns of inhibition and facilitation obtained. We know of no other implemented model of picture–word interference that has been able to do the same. As was pointed out by **Dell et al.** the model proposed by Cohen et al. (1990) realizes attention as a selective modulation of activation. The same is true for Starreveld and La Heij's (1996b) model. However, these models have been applied only to picture naming and not to categorization tasks. Furthermore, Starreveld and La Heij's model predicts only semantic inhibition, but facilitation is also often observed, for example, in categorization tasks (for an extensive discussion, see Roelofs 1992a; 1993). In addition, these other models do less well than WEAVER++ in explaining the time course of the effects. For example, the model of Cohen et al. (1990) incorrectly predicts that most inhibition of written distractors occurs at negative SOAs (−300 and −400 msec) rather than around SOA = 0 msec as is empirically observed (see, e.g., Glaser & Düngelhoff 1984; Glaser & Glaser 1982; 1989). Cohen et al. argue that their model, rather than the data, reflects the real effect (pp. 344–45). However, it has been found time and again that inhibition centers around the SOA of 0 msec rather than at the longer negative SOAs (for reviews, see Glaser 1992; MacLeod 1991). Thus, existing implemented models of attention in picture–word interference account less well for the empirical findings than does WEAVER++. Nevertheless, as was argued by **Starreveld & La Heij** and in our target article (sect. 5.2.1), there are reasons to assume that the response set principle in WEAVER++ is too narrow. It explains why semantic facilitation is obtained in categorization tasks: subordinate distractors (such as *dog*) help rather than compete with superordinate targets (such as *animal*) because the subordinates are not permitted responses. However, in picture naming, semantic inhibition or facilitation can be obtained. Distractors in the response set always yield semantic inhibition, but sometimes distractors that are outside the response set yield inhibition, too. Perhaps a response set is only marked in memory when the number of responses is small and can be kept in short-term memory (La Heij & van den Hof 1995). Thus, when the response set is large, or when there are no a priori constraints on the responses, semantic inhibition is predicted, which is indeed obtained. Similarly, without a priori restrictions, the lemmas of all possible names of a target (e.g., two synonyms) may be selected and our analysis of the experiments of Peterson and Savoy (1998) applies, contrary to the claim of Starreveld and La Heij. Another reason for assuming that our response set principle is too narrow was mentioned in our target article: the same distractors may yield semantic inhibition in the spoken domain but facilitation in the written domain (Roelofs, submitted c), or inhibition in both domains (Damian & Martin 1999). Clearly, it is not fully understood exactly under what conditions inhibition and facilitation are obtained, and this requires further investigation.

Santiago & MacKay note that we have not treated morphological, phonological, and phonetic output buffers in our theory. Again, we can refer to the embedding of the theory in the larger framework of *Speaking.* The functioning of various buffers in utterance generation, in particular working memory, the syntactic buffer, and the articulatory buffer, is extensively discussed there, but none of these buffers is specific to lexical processing.

In short, the theory reported in the target article does not function in a vacuum. On the one hand, it is embedded in the larger theoretical framework developed in *Speaking.* On the other hand, it already incorporates major attainments of human performance theory, with detailed solutions for aspects of selective attention and response selection.

R2. Modeling lexical access

R2.1. Abstract properties of the model

R2.1.1. Symbolic computation, spreading activation, and homunculi. WEAVER++ integrates a spreading-activation network with a parallel system of production rules. Some commentators reject such an approach a priori because it is said not to embody "natural computation" (**Roberts et al.**), or "in comparison with interactive activation and parallel distributed processing (PDP) approaches is much too declarative for processes of word-form encoding" (**O'Seaghdha**), or, curiously enough, because production rules introduce "homunculi" (**Santiago & MacKay**). These objections are mistaken in that they fail to acknowledge that the relation between symbolic and lower level constructs is simply one of levels of analysis. For example, symbolic production rules probably can be realized connectionistically (Touretzky & Hinton 1988) and, more importantly, in real neural networks (see, e.g., Shastri & Ajjanagadde 1993). Whether a model is at the right level of analysis can only be determined by the model's empirical success, and here WEAVER++ fares rather well, so far. Spreading activation gives a simple account of the computational problem of retrieval of information from memory and it naturally explains graded priming effects. Production rules provide a natural account for the computational problem of selection and permit the system to be generative, that is, to create, for instance, novel morphological and phonological structures. Also, the marriage of activation and production rules provides a natural account of priming without loss of binding. We know of no other implemented model of lexical access that achieves the same by simpler means.

R2.1.2. Binding by checking versus binding by timing. According to **Dell et al.,** binding by timing should be preferred over binding by checking. However, this may depend on whether one sets out to explain speech errors or speech latencies. Paradoxically, models using binding by timing have difficulty accounting for data on the timing of lexical access, whereas models using binding by checking need extra provisions to account for checking failures that cause speech errors. In models using binding by timing, such as Dell's (1986; 1988), selection of a node occurs on an ordinal basis, as the most highly activated node is selected. Selection of nodes takes place at a constant time interval after the onset of their activation. Therefore, priming affects the levels of activation but does not affect latencies. At the predetermined moment of selection either the target node is the most highly activated node and is selected or it is not the most highly activated node and an incorrect selection occurs. Thus, priming may increase the error rate but will not decrease production latencies. In short, priming interferes with binding without having any temporal effect. Note that were priming to influence the moment of selection by increasing or decreasing the interval between activation onset and selection this would interfere with binding, too.

WEAVER++ has been designed to account primarily for latency data, not for speech errors. **Dell et al.** and **O'Seaghdha** argue that the model has difficulty in explaining exchanges. In other models an anticipation heightens the chance of a perseveration, whereas in WEAVER++ it does not. In the simulation (depicted in Fig. 18 of the target article) the verification failures for the two error locations were assumed to be independent, which underestimates the exchange rate. However, this is not a necessary assumption of our theory. To make WEAVER++ generate more exchange errors, one might take up Shattuck-Hufnagel's (1979) proposal and implement a "check-off" mechanism that marks the segments that have been used. If, after access of a syllable motor program, the corresponding segments within the phonological word representation are checked off, an anticipatory failure would increase the likelihood of a perseveratory failure, thereby increasing the absolute number of exchanges. This is because, after checking off an anticipated segment, the original segment would remain available. This is a possible, though admittedly post hoc, extension of WEAVER++. The crucial point is that the low rate of exchanges in WEAVER++ is not due to a principled feature of our approach to errors, unlike what has been shown by Dell et al. (1993) for certain PDP approaches to errors. Nevertheless, we agree that, in further development of WEAVER++, its error mechanism deserves much more attention.

Dell et al. also doubt whether binding by checking can handle the evidence for conceptual prominence in syntactic planning. Of course, that is not yet part of WEAVER++. However, the philosophy of binding by checking is to secure the proper associations between lexical concepts and lemmas, and that mechanism may well be involved in the assignment of lemmas to grammatical functions in the generation of syntax. Thus, when the grammatical function of subject is to be filled in, creating the utterance *pages escort kings,* the binding-by-checking mechanism may test whether the selected lemma plus diacritic *page+pl* is eligible for this role by determining whether the concept PAGES was indeed specified as the agent in the conceptual structure. Dell et al. refer to the classical evidence that the order in which words become activated may affect syntactic structure. Those results reflect on the interplay between lexical retrieval and syntactic structure-building processes but do not specifically support a binding-by-timing mechanism.

R2.1.3. Competition without network inhibition. In WEAVER++ information is transmitted from level to level via spreading activation. There are no inhibitory links within or between levels. **Wheeldon** and **Ferrand** were concerned that WEAVER++ would, therefore, be unable to account for the semantic and phonological inhibition effects reported in the literature. This concern is unfounded. Of course, inhibitory effects (slower responses in a related than in an unrelated condition of a priming experiment) can be due to inhibitory links between units, but other accounts are possible and may be preferred (see, e.g., Dell & O'Seaghdha 1994). As was pointed out by **Harley,** inhibitory effects often occur in production, but there is little evidence for underlying inhibitory links.

In Roelofs (1992a; 1997c) and in the target article (sects. 3.2.2, 5.1, and 6.3), we explained that WEAVER++'s selection mechanism weighs activation of response candidates against each other, using Luce's ratio, which yields a competition-sensitive response mechanism. Competition in WEAVER++ can occur at two levels: between lemma nodes in lemma retrieval and between syllable program nodes in word-form encoding. We showed that the Luce ratio allowed WEAVER++ to account for both facilitation and inhibition effects in lemma retrieval in the picture– word interference task (sect. 5.2.1). Thus, semantic inhibition effects come as no surprise in the model. However, WEAVER++ predicts only phonological facilitation in this task (sect. 6.4). It should be noted, however, that phonological inhibition occurs in tasks other than the picture–word task. The tasks used by Wheeldon (submitted) and by O'Seaghdha and Marin (in press; see **O'Seaghdha**) concern the inhibitory influence of the production of a word on the subsequent production of a related word. Insofar as the target article was concerned only with the generation of isolated phonological words, WEAVER++'s handling of interference between subsequent words was not discussed. The still unpublished mechanism is this: without a recent production history, all candidate motor programs are weighed equally by WEAVER++. However, if motor programs have been recently used, they are more heavily weighed in the Luce ratio in a subsequent encoding of a form, which may cause phonological inhibition. For example, when WEAVER++ produces *book* after having produced *booze* (phonologically related) or *lane* (phonologically unrelated) and one increases the weight of the activation of the syllable program nodes for *booze* and *lane* in the Luce ratio by 5.0, an inhibitory effect of 65 msec from phonological similarity is obtained. In planning *book,* the syllable program node of *booze* will become reactivated to some degree owing to the shared segments, whereas the syllable program node of *lane* will not be coactivated. It should be stressed that this example is meant not as a full account of phonological inhibition effects in production but only to illustrate that inhibitory effects at the behavioral level are fully compatible with WEAVER++ even though the model has no inhibitory links. Laine and Martin (1996), cited in **Ferrand**'s commentary, made the following prediction from our theory: repeated naming of pictures that have sound-related names should not affect naming latency compared to when names are not sound related. However, in our published work, that case of multiple sequential access had never been modeled. The above analysis predicts interference for this experimental case.

Wheeldon points out that in several form-priming studies different effects have been observed for begin-related and rhyming word pairs. In the picture–word interference paradigm both types of relatedness yield facilitatory effects, and, in implicit priming experiments, only begin-relatedness yields facilitation, whereas rhyming words are produced no faster than unrelated ones. Both results are predicted by our model. However, in other paradigms in which participants pronounce both words, begin-relatedness yields inhibition, but end-relatedness yields facilitation. The present version of WEAVER++ does not include an account of this pattern of results.

R2.2. Ockham's razor and the discreteness issue

As we argued in section 3.2.5 and is stressed by **Cutler & Norris**, a serial model should be one's first choice. Seriality of lemma retrieval and form encoding excludes cascading (only the form of the selected lemma is activated) as well as direct form-to-lemma feedback (i.e., there are no backward links from forms to their lemmas). However, contrary to what **O'Seaghdha** claims, we do not deny the relevance of feedback for speech production or for skilled action in general. First, our theory assumes that production is guided by indirect feedback via the speech-comprehension system (see target article Fig. 1 and sect. 9). This indirect feedback serves to detect errors. Speakers monitor their own phonological plans and external speech to discover and repair errors, dysfluencies, or other problems. Second, we assume that lemmas are shared between production and perception, which implies feedback from lemmas to concepts (see sect. 3.2.4). O'Seaghdha mistakenly believes that the exclusion of feedback is partly responsible for our introducing "between-level verification, monitoring, multiple lemma selection, storage of variant word forms, a suspension-resumption mechanism, a syllabary." However, only monitoring bears on the feedback issue. Multiple lemma selection bears on the issue of cascading, and the other devices have to do with neither feedback nor cascading.

Evidence that seemingly argues in favor of nondiscreteness (coactivation of near-synonyms and malapropisms) may in fact require only a refinement of a discrete model. **Jescheniak & Schriefers** present new data confirming our proposal of one such refinement, namely, our account of coactivation of the forms of synonyms, which others have taken to support cascading (e.g., Peterson & Savoy 1998). Also, **Kelly** proposes a discrete analysis of malapropisms, which are often taken as evidence for interactivity. In line with his proposal, we would like to point to another possibility, which reveals the similarity between malapropisms and tip-of-the tongue states (TOTs) (see also Roelofs et al. 1998). A malapropism may occur when a speaker can generate only an incomplete form representation of the intended word (as in a TOT). This incomplete form is fed back to the conceptual system via the comprehension system, which leads to the activation of several lemmas that are phonologically related to the target. These lemmas typically will be semantically unrelated to the target. If one of the form-related lemmas of the appropriate grammatical class is selected, a malapropism will occur. This explains why both in TOTs and in malapropisms, the substituted words typically resemble the intended words in number of syllables, stress pattern, and initial segments (Brown 1991; Fay & Cutler 1977), and it explains the syntactic class constraint on malapropisms.

Starreveld & La Heij refute serial access on the basis of an interaction between semantic and orthographic similarity in picture–word interference experiments (Starreveld & La Heij 1995; 1996b). We have argued that the interaction can be readily explained by assuming that form similarity affects both lemma retrieval and form encoding in parallel. Our explanation assumed that perceived words activate a cohort of lemmas and forms in the mental lexicon. There is much evidence for phonological cohorts in spoken word recognition (for reviews, see, e.g., Elman & McClelland 1986; Marslen-Wilson 1987; Norris 1994). Furthermore, there is evidence for orthographic cohorts in written word perception as well. Roelofs (submitted c; MPI Annual Report 1996) showed that written and spoken word-fragment distractors yield semantic effects. For example, the syllable *ta* yields a semantic effect in the naming of a desk (*ta* is the first syllable of *table,* a semantic competitor of *desk*). Furthermore, Starreveld and La Heij mistakenly believe that, in our model, part of the orthographic facilitation effect itself is located at the lemma level. This, however, is not the case. We have explained that "the locus of the main effect of orthographic relatedness is the level of phonological output codes" (Roelofs et al. 1996, p. 249). In the relevant simulations, written words activated an orthographic cohort, but orthographic overlap between target and distractor influenced lemma retrieval for production only when there was also a semantic relation between target and distractor. Orthographic overlap had an effect on lemma retrieval only in the case of semantic inhibition, that is, when lemma retrieval was delayed owing to a semantic relationship. Thus, our claim is that the locus of the orthographic facilitation effect per se is the level of word-form encoding. The interaction with semantic similarity occurs at the lemma level, but there is no pure form facilitation at this level. Because our earlier studies referred to by Starreveld & La Heij did not use mixed distractors (i.e., distractors that were both orthographically and semantically related to the target), there is no need to reinterpret their results or reestimate parameters of WEAVER++.

R2.3. General issues of model construction

Jacobs & Grainger criticize us for not making explicit all aspects of our methodology in building, testing, and evaluating WEAVER++, but that was not the primary goal of our target article. It presents a theory of lexical access, not a full description of the computational model WEAVER++, which has already been given elsewhere (Roelofs 1992a; 1993; 1996b; 1997c). The authors' commentary invites us to make these issues explicit here.

R2.3.1. Scalability. We agree that, for the lemma retrieval component, the scalability has not been seriously tested. Testing whether this model component scales up to larger lexicons would require detailed assumptions about the structure of the semantic network. However, WEAVER++ is a theory not of semantic memory but of lexical memory. For the word-encoding component, the issue of scalability has been explicitly dealt with. WEAVER++ makes detailed claims about the structure of the form lexicon. As was mentioned in the target article, to examine whether the size and content of the network influence the outcomes, simulations were run using both small and larger networks. The small networks contained the nodes that were minimally needed to simulate all the experimental conditions. The large networks embedded the small networks in networks representing the forms of either 50 words randomly selected from CELEX or 50 words of highest frequency. The small and larger networks produced virtually identical outcomes, as shown by a test statistic and evident from Figure 4 of Roelofs (1997c). We agree that our large lexica are still small compared to real lexica of some 30,000 words. However, it is important to keep in mind the goal of a modeling exercise. Some connectionist models use thousands of words in their network (those published for language production, however, do not). The required size of a network typically depends on the particular theoretical claim that is made. WEAVER++ claims that the effects from picture–word interference experiments can be explained by reference to the memory representations of the target and the distractor only. The effects do not arise from the massive influence of all words in the mental lexicon, as is the core claim of many connectionist models. The simulations run with WEAVER++ suffice to demonstrate this.

R2.3.2. Parameter estimation. As was mentioned by Roelofs (1992a), the data obtained by Glaser and Düngelhoff (1984) were taken to estimate parameters because this study provides several of the most important findings on the time course of the semantic effects in picture–word interference. The same materials were used in picture naming, picture categorizing, and word categorizing, and all SOAs were tested within participants. This makes the Glaser and Düngelhoff study, compared to other studies published in the literature, particularly attractive for unbiased parameter estimation. **Jacobs & Grainger** argue that Simplex estimation is suboptimal, but this is a rather sweeping statement. Whether an estimator is optimal depends on the structure of the parameter space, and modellers typically try several estimators, as we did in the case of WEAVER++.

R2.3.3. Comparison with other models and measures. **Jacobs & Grainger** wonder why we have made no comparisons to other models in the literature, even though comparable models are said to be available, namely, those of Dell (1988) and Schade (1996). A problem is, however, that these other models are designed to account for speech errors but not for production latencies. WEAVER++'s unique contribution is that it makes quantitative predictions about latencies. Thus, the ground for making detailed comparisons to these other models is missing. A further complication is that other models have often not been formalized. For example, Jacobs and Grainger refer to Dell (1988), but this model has not been implemented. Jacobs & Grainger argue that, in order to discriminate WEAVER++ from other models, we should have tested item- and participant-specific predictions and predictions about response time distributions rather than about mean production latencies only. However, given that other models cannot even predict mean latencies to start with, it seems premature to test such more fine-grained predictions from WEAVER++. Furthermore, testing the model on response latency distributions is not as straightforward as it may seem. The empirical distributions of naming times reflect not only lemma retrieval and word-form encoding but perceptual and motor processes as well, but WEAVER++ does not include these other processes.

R3. Lexical selection

R3.1. Why lemmas?

The notion of "lemma" was introduced by Kempen and Huijbers (1983) to denote a lexical entry's semantic/syntactic makeup. In addition, they coined the term "lexeme" to denote the item's morphological/phonological makeup. The distinction proved most useful in theories of speaking, for instance, in Dell (1986) and in Levelt (1989). Lemmas are crucial in grammatical encoding, the mapping of conceptual representations (or "messages") onto surface structures. Lexemes are crucial in morphophonological encoding, the generation of segments, syllables, phonological words and phrases, intonational phrases, and ultimately the entire utterance. Garrett's (1975) classical analysis of speech errors in utterance generation showed that grammatical and morphophonological encoding (in his terms "functional" and "positional" encoding) are relatively independent processes. The lemma–lexeme distinction provides the lexical basis or "input" for this two-stage process of formulation.

The development of the theory presented in our target article necessitated a further refinement. Roelofs (1992a) handled both the hyperonym problem and the naming-latency data reported in section 5.2 by introducing a spreading activation network in which the original "lemma" information is presented in a pair of nodes. One node (plus its connectivity in the conceptual stratum of the network), now called the "lexical concept," represents the item's semantics. The other node (plus its connectivity in the syntactic stratum), now called "lemma," represents the item's syntax. Together, they precisely cover the "old" lemma. It is no wonder that this change of terminology led to occasional confusion. **Zorzi & Vigliocco** argue that "if concepts are represented by sets of semantic features, these features must be bound to represent a lexical unit before any other kind of representation can be properly accessed" (such as the item's form or lexeme information). That is entirely correct. However, in our model this function is performed *not* by the (new) lemma but rather by the lexical concept node. It is the lexical concepts that are organized into semantic fields. It is activation spreading through the conceptual network that is responsible for semantic substitution errors. Zorzi and Vigliocco's analysis does hold for the "old" lemma though. **Harley** proposes to eliminate the (new) lemma level entirely. We are all for parsimony, and this is worth considering. Harley's main argument is that the evidence for the existence of lemmas is quite restricted, namely, the finding that in TOT states and in cases of anomia the target noun's gender can be accessible without its phonology being accessible (see sect. 5.4.1 of the target article). Harley joins Caramazza and Miozzo (1997) in arguing that these findings can be handled without introducing an intermediate lemma stratum in the network. There are, however, two reasons for not moving too fast here. First, it is not the case that the evidence is currently mainly restricted to these gender findings. In fact, we go to substantial lengths in the target article to list several more arguments in support of a lemma level (sects. 5.4.1 through 5.4.4); all that evidence is conveniently ignored by Harley. Second, as we argue in detail in our reply to Caramazza and Miozzo (Roelofs et al. 1998), their solution to connect the item's syntactic information (such as gender) to the word form node (or lexeme) creates a host of problems for the explanation of a range of experimental findings, which are quite naturally accounted for in terms of a lemma level in the network. One example is the homophone word-frequency effect reported in section 6.1.3 of the target article. The homonyms *more* and *moor* are syntactically different (adjective and noun, respectively). Therefore, in WEAVER++ they have different lemmas (see diagram in sect. 6.1.3), but they share the word form (/mɔr/) and hence the word form (or lexeme) node. Jescheniak and Levelt (1994) showed that the low-frequency homophone (*moor* in the example) is produced with the fast latency of a high-frequency word. The reason is that the shared word form (/mɔr/)is high-frequency owing to the high frequency of *more.* In Caramazza and Miozzo's (1997) solution this cannot be accounted for, because the two items have two distinct lexemes; one represents the phonology plus syntax of *moor* and the other the phonology plus syntax of *more.* But then one would predict relatively slow access to the low-frequency homonym *moor,* contrary to the experimental findings. There are, in addition, important linguistic arguments for maintaining the syntactic lemma level in the representation of a lexical item.

R3.2. Solutions to the hyperonym problem

Bowers, Harley, and **Zorzi & Vigliocco** point to possible decomposed mechanisms for solving the hyperonym problem. As we have argued elsewhere (Levelt 1989; Roelofs 1997a), a decompositional solution is probably possible, but we have independent reasons for preferring our nondecompositional position. First, the hyperonym problem is part of a much larger class of convergence problems, including the dissection problem, the problem of disjunctive concepts, and the word-to-phrase synonymy problem (see, e.g., Bierwisch & Schreuder 1992; Levelt 1992; Roelofs 1992a; 1992b; 1997a). Nondecomposition provides a principled solution to this class of problems as a whole rather than a solution to one of its subproblems. The proposed decompositional mechanisms solve one problem but not the others. Second, as we have argued in several places (e.g., Roelofs 1992a; 1992b; 1993), the challenge is to solve this class of convergence problems as a whole *and* to account for relevant empirical data. So far, the proposed decompositional mechanisms do not meet this challenge. These arguments concern nondecomposition in a theory of language production, but there are more general arguments in favor of nondecomposition as well (see, e.g., Fodor et al. 1980).

R4. Word-form encoding and articulation

R4.1. Sequentiality of prosodification

In our model, prosodification is a sequential process proceeding segment by segment from the beginning of each phonological word to its end. The articulation of a phonological word may be initiated as soon as its prosodification has been completed. Important evidence in support of this view comes from the implicit priming experiments reviewed in the target article (sect. 6.4.2). A central finding is that speakers are faster to initiate word production when all words of a test block share one or more word-initial segments than when their beginnings are different. By contrast, shared word-final segments do not facilitate word production. **Kawamoto** proposes that this pattern does not show, as we argue, that prosodification is a serial process but shows that articulatory preparation is serial and that articulation can begin as soon as a word's first segment has been selected.

Perhaps articulatory preparation contributes to the facilitatory effects observed in implicit priming experiments. However, there are a number of observations for which the articulatory hypothesis by itself cannot account. First, the size of the preparation effect increases with the number of shared word-initial segments (Meyer 1990; 1991; Roelofs 1998) as predicted by our prosodification hypothesis. According to the articulatory account, the speaker can initiate the utterance as soon as its first segment has been selected. Hence, the benefits of one or several shared word-initial segments should be equal. Second, Roelofs (1996b; 1996c; 1998) obtained stronger facilitatory effects when the shared segments corresponded to complete morphemes in morphologically complex targets than when they corresponded to syllables. In addition, the size of the facilitatory effects depended on the frequencies of the shared morphemes. These results would not be expected under the articulatory hypothesis. Third, as Kawamoto explains, the articulatory hypothesis predicts weaker preparation effects from initial plosives than from nonplosives, but his reanalyses of Meyer's data do not support this claim. Fourth, the measurements of segment durations proposed by Kawamoto have, in fact, been carried out, albeit only for a sample of the responses of two experiments reported by Meyer (1990). In those experiments, the implicit primes included the first syllable, or the first and second syllables of the targets. Contrary to Kawamoto's prediction, these syllables were slightly shorter, rather than longer, in the homogeneous than in the heterogeneous condition. Fifth, Wheeldon and Lahiri (1997), using a different paradigm, have provided additional support for the assumption that speakers usually generate the phonological representation of an entire phonological word before initiating articulation. Finally, there is the important theoretical argument against Kawamoto's view that it does not account for anticipatory coarticulation, that is, for the fact that the way in which a segment is realized may depend on properties of following segments. In short, then, we still believe that prosodification is a sequential process and that the utterance of a phonological word is usually initiated only after it has been completely prosodified.

R4.2. Representation of CV structure

In our model a word's metrical structure specifies its number of syllables and stress pattern, but not its CV structure. How can we account for the effects of CV structure reported in the literature (e.g., Meijer 1994; 1996; Sevald et al. 1995; Stemberger 1990; see also **Santiago & MacKay**'s commentary)? As was explained in the target article (sects. 6.2.2 and 6.2.3) we do not claim that the CV structure of words is not represented at all. It is, in fact, captured in several ways and there are several possible loci for effects of CV structure. First, during the segmental spell-out of a word, information about its phonological features becomes available. This information includes a specification of the consonantal or vocalic character of each segment. This is demonstrated by the strong tendency of the segments interacting in sound errors to overlap in more features than expected on the basis of chance estimates, in particular the tendency of vowels almost exclusively to replace other vowels and of consonants to replace other consonants. There is also an important theoretical reason to assume that phonological features are represented: the retrieved string of segments must be syllabified, and the syllabification rules refer to the segments' phonological features. Thus, some of the alleged effects of CV structure may be effects of featural similarity. Second, effects of CV structure can arise because similar rules are applied to words with similar CV structure during the syllabification process. Finally, words are phonetically encoded by selecting syllable program nodes. The addresses for these nodes include specifications of the number and types of segments (consonantal or vocalic) and their order. Thus, CV structure is represented in the addresses to syllable program nodes. Words with similar CV structure have similar addresses, and some effects of CV structure may arise during the selection of syllable program nodes. In sum, the model captures CV structure in several ways and different accounts of the observed effects of CV structure are possible.

Santiago's as yet unpublished finding that words beginning in consonant clusters are initiated more slowly than words beginning in singletons is interesting. As was noted in the target article (sect. 6.2.2), word onsets are clearly different from other segments; they are, for instance, far more likely to be involved in errors but also to be recalled in TOT states and anomia. Onset clusters are particularly problematic in that in speech errors they sometimes behave like one unit and sometimes like two segments. To account for Santiago's finding we could postulate that clusters necessitate an extra processing step during segmental spell-out (as proposed in *Speaking*) or during the assembly of the addresses to the syllable program nodes, but these solutions are clearly ad hoc stipulations.

Ferreira proposes that speakers may refer to the CV structure when planning the temporal structure of utterances. Accordingly, a longer time slot could be given to a Dutch word such as *taak* (*task*), which includes a long vowel that occupies two vocalic positions, than to a word such as *tak* (*branch*) with a short vowel, which occupies only one vocalic position. In Meyer's (1994) experiments speakers indeed allocated longer time slots to words with long vowels than to words with short vowels. However, exactly the same difference was found for word pairs differing in the phonetic length of their single-onset or coda consonant. This strongly suggests that phonetic length, not CV structure, determined the length of the timing slots.

R4.3. Phonological versus phonetic syllables

An important claim in our theory is that the parsing of words into syllables is not part of their lexical entries but is generated online. This is more parsimonious than first retrieving stored syllables of the citation forms of words and subsequently modifying them during a resyllabification process. The assumption that phonological syllables are generated, while phonetic syllables are stored, is not a theoretical inconsistency as was suggested by **Santiago & MacKay** but is dictated by our striving for a parsimonious theory and by the fact that the syllabification of a word in context depends, in many cases, on properties of the preceding or following words. As was explained in the target article (sect. 7.1), most of the articulatory work is done by a small set of syllables, and it would therefore appear advantageous for the speaker to store them in a syllabary and retrieve them as units instead of assembling them afresh each time they are needed.

R5. The neural correlates of lexical access

R5.1. Electrophysiology and brain imaging

There is no reason to claim superiority for our dominant methodology of reaction time measurement, and we do not. In his commentary **Müller** correctly notices possible advantages of using electrophysiological measurements such as event-related potentials (ERP). We do not agree, though, that subjects must be *aware* of the process in order to produce a behavioral reaction time effect. An example in our target paper is the LRP study from our laboratory by van Turennout et al. (1998), which showed that, during noun phrase production, syntactic (gender) information is accessed some 40 msec before word-form information, even if the task would make the reverse ordering more efficient. Müller mentions a host of studies, including his own work in cooperation with Kutas, showing word-class effects, such as closed versus open class, concrete versus abstract words, verbs versus nouns. The electrophysiological studies among them (i.e., those that provide timing information) are all word-comprehension studies. ERP production studies are rare, because of the myoelectric interference arising in overt articulation tasks. Some authors (e.g., Abdullaev & Posner 1997) resort to "silent naming" tasks, and there are other ways out, such as that in the just-mentioned van Turennout et al. study. Still, ERP measurement is not the most obvious alternative to RT studies of production. MEG fares better. The MEG study of picture naming (Levelt et al. 1998) referred to in section 11 of the target article used the timing information from picture-naming RT experiments to localize dipole sources involved in subsequent stages of word production. A major finding here was left Wernicke's area involvement in accessing the word's phonological code. Ultimately, brain imaging results in word production, whether MEG, PET or fMRI, can be interpreted *only* from a process analysis of the production task used. Indefrey and Levelt (in press) used the componential analysis of Figure 1 in the target article to make a taxonomy of word-production tasks used in imaging studies (such as picture naming, verb generation, word repetition) and used it in a meta-analysis of 58 published imaging studies of word production. It showed that the word-production network is strictly left-lateralized, consisting of the posterior inferior frontal gyrus (Broca's area), the midsuperior and middle temporal gyri, the posterior superior and middle temporal gyri (Wernicke's area), and the left thalamus. Often component processes (such as phonological encoding) correlated with activation in one or more of these cerebral areas. Müller refers to one production imaging study (Damasio et al. 1996) suggesting that different semantic classes (such as tools and animals) have locally distinctive representations in the brain. He could have added the paper by Martin et al. (1996). Meanwhile this has become a hot issue in several laboratories.

R5.2. Neuropsychology

Although our theory of lexical access was not developed to account for neuropsychological disorders in word production, it is satisfying to see that the theory can inspire neuropsychological research in word production. In our view, the inspiration should mainly be to become more explicit about the component processes of word production when testing aphasic patients and to apply experimental paradigms developed in our research. However, we feel it as a bridge too far to expect a patient's behavior to conform to our theory. **Ferrand** mentioned the study by Laine and Martin (1996) of an anomic patient with substantial problems in picture naming and other word-production tasks. The authors derived from our theory the prediction that the patient should not show evidence of phonological interference if a set of pictures with phonologically related names is to be named repeatedly. We mentioned above (sect. R2.1.3) that our theory does not make this prediction. Hence, Ferrand's conclusion that "LRM's discrete theory of lexical access is unable to explain the observed effects" is incorrect. Here, however, we want to make the point that even if our theory did make the prediction for normal speakers, anomic patients may well behave differently. There is little a priori reason to suppose that an impaired system performs according to an intact theory. The real theoretical challenge is to create and test a theory of the impaired system (as is done, e.g., by Dell et al. 1997b).

That is the approach in the commentary of **Semenza et al.,** who use theoretical distinctions from our theory to interpret the functioning of the damaged system in various types of aphasic patients. The commentary concentrates on the generation of morphologically complex words. In section 5.3.2 we argued that most lexicalized compounds, such as *blackboard,* have a single lemma node, but multiple form nodes, one for each constituent morpheme. Semenza and colleagues (Delazer & Semenza 1998; Hittmair-Delazer et al. 1994; Semenza et al. 1997) observed in careful group and single-case studies that naming errors produced by aphasic patients often preserved the target compound morphology. Here are three examples: (1) *portafoglio* (wallet) → *portalettere* (postman); (2) *portarifiuti* (dustbin; "carry-rubbish") → *spazzarifiuti* (neologism; "sweep rubbish"); (3) *pescecane* (shark; "fish-dog") → *pescetigre* (neologism; fish-tiger). In example (1), one existing compound is replaced by another existing compound, but they share the first morphological component. This may have been generated as follows: the correct single lemma is selected, activation spreads to both morpheme nodes, but only the first one reaches suprathreshold activation and is phonologically encoded as a whole phonological word (see sect. 6.4.4). It becomes available as "internal speech" (cf. sect. 9).

The phonological code of the second morpheme remains inaccessible. To overcome this blockage, the patient resorts to an alternative existing compound that is activated by self-perceiving the encoded first morpheme (*porta*); it happens to be *portalettere.* This, then, is a secondary process, involving the "internal loop" (Levelt 1989). The prediction is that the response is at least relatively slow and probably hesitant. We could not check this because none of the cited papers present latency or dysfluency data. Whatever the value of this explanation, it cannot handle the neologisms (2) and (3). The initial steps (selecting the single compound lemma and phonologically encoding of one of its morphemes) could be the same, but how does the patient respond to blockage of the other morpheme? In example 2, self-perceiving *fiutti* ("rubbish") may have activated a rubbish-sweeping scene, leading to the selection and encoding of *spazzari* in the still open slot. In example (3), however, no credible perceptual loop story can be told. The patient correctly generates *pesce* ("fish"), but how can self-perceiving this morpheme elicit the generation of *tigre* ("tiger") instead of *cane* ("dog")? Semenza et al. (1997) and Delazer and Semenza (1998) draw the obvious conclusion that the *cane* component in the compound must have been semantically and syntactically available to the patient. In other words, the "normal" single-lemma–multiple-morpheme networking that we propose for compounds (sect. 5.3.2) doesn't hold for this patient (or at least not for this compound in the patient's lexicon).

The authors propose a two-lemma–two-morpheme solution and the apparent transparency of the semantics suggests that there are two active lexical concepts as well. This production mechanism corresponds to Figure 6D in the target article. It is a case of productive morphology and presupposes the intact availability of morphological production rules, as Semenza et al. note. The *cane* → *tigre* replacement would then be a "normal" selectional error, generated by a mechanism discussed in section 10 of the target article. If indeed the patient's lexical network is partially reorganized in this way, errors of types (1) and (2) may also find their explanation in these terms. Several examples in the commentary and the relevant papers suggest a reorganization of the lexicon towards "fixed expressions." These are stored "lexical" items with complex lemma structures. Very little is known about their production.

R6. Suggestions for future research

Several commentators have alerted us to issues that we have not treated yet. One of these is the embedding of our theory in a comprehensive theory of development (**Roberts et al.; Zorzi & Vigliocco**). Here we respond to those homework assignments that may be manageable within another decade of research. We do, for instance, whole-heartedly agree with **Cutler & Norris** that comprehension and production researchers should join forces to create a parsimonious model to account for both capacities. In section 3.2.4 of the target article we proposed (Assumption 3) that, from the lemma level up, the two networks are identical; however, we have argued for one-way connections at lower processing levels. Cutler and Norris similarly argue for one-way bottom-up connections in the perceptual network. Hence the two networks are arguably different at that level of processing. What is in urgent need of development is a theory of their interconnections. Carr (personal communication) proposes to focus on the word-frequency effect. It arises in both production (see our sect. 6.1.3) and perception, and in both cases it involves access to the phonological code. Can there be a unified theoretical account? **Hirst** wonders whether syntactic operations on the lemmas can feed back to the lexical concepts. Assumption 3 invites the answer "yes." Chapter 7 of *Speaking* presents examples of syntactic operations checking conceptual argument structure (see also sect. R2.1.2). But Hirst's point for future research is whether speakers will adapt their lexical choice if the developing syntactic structure turns out to have no solution for the current set of lemmas. We would expect this to happen, but an experimental demonstration will not be easy. In addition, Hirst invites a psychological account of how pleonasm is (usually) prevented in our lexical choices. Theoretical solutions to both issues may naturally emerge from Kempen's (1997) recent work. It already provides a principled theoretical solution to the conceptual primacy effects in the generation of syntax, as referred to in **Dell et al.**'s commentary. **Gordon**'s homework assignment is to further relate lexical choice to the constraints arising in discourse. This obviously involves pronominalization (extensively discussed in *Speaking*) and other reduced or alternating forms of reference. The experimental work is on, both in our own laboratory and other laboratories (see Schmitt 1997 and **Jescheniak & Schriefers**'s commentary). Another important area of investigation is the use of what Clark (1998) calls "communal lexicons."

Dell et al. close their commentary by recommending the study of attentional mechanisms that control the timing of the activation of conceptual and linguistic units. We have already taken up this challenge. Assuming that there is a close relationship between gaze and visual attention, we have started to register speakers' eye movements during the description of pictures in utterances such as *the ball is next to the chair* (Meyer et al. 1998). In a first series of experiments, we found that speakers have a strong tendency to fixate on each object they name and that the order of looking at the objects corresponded almost perfectly to the order of naming. Most importantly, we found that speakers fixated on each object until they had retrieved the phonological form of its name. This suggests that at least these simple descriptions are generated in a far more sequential way than one might have expected. Whether more complex utterances are generated in the same sequential manner must be further explored.

Some homework assignments failed to come forth. We were somewhat surprised to find that two core issues of word-form encoding, the generation of morphology and the generation of metrical structure, invited very little reaction. The experimental findings are nontrivial, and in fact the first of their kind in language-production research; they cry out for cross-linguistic comparisons. It is by no means obvious that the generation of morphology involves the same mechanisms in languages with limited morphological productivity (such as Dutch or English) and languages whose generativity largely resides in morphology (such as Turkish). The storage/computation issue that we addressed for the generation of a word's metrical phonology will most certainly be resolved differently for stress-assigning languages (such as Dutch or English) than for languages with other rhythmic structures. The production mechanisms will probably vary as much as the corresponding comprehension mechanisms (see Cutler et al. 1997 for a review).

References

Letters "a" and "r" appearing before authors' initials refer to target article and response, respectively.

Abbs, J. H. & Gracco, V. L. (1984) Control of complex motor gestures: Orofacial muscle responses to load perturbations of lip during speech. *Journal of Neurophysiology* 51:705–23. [aWJML]

Abdullaev, Y. G. & Posner, M. I. (1997) Time course of activating brain areas in generating verbal associations. *Psychological Science* 8:56–59. [rWJML]

Abeles, M. (1991) *Corticonics - Neural circuits of the cerebral cortex.* Cambridge University Press. [FP]

Arnell, K. & Jolicoeur, P. (1997) The attentional blink across stimulus modalities: Evidence for central processing limitations. *Journal of Experimental Psychology: Human Perception and Performance.* 23:999–1013. [THC]

Baars, B. J., Motley, M. T. & MacKay, D. G. (1975) Output editing for lexical status from artificially elicited slips of the tongue. *Journal of Verbal Learning and Verbal Behavior* 14:382–91. [aWJML, JS]

Badecker, W., Miozzo, M. & Zanuttini, R. (1995) The two-stage model of lexical retrieval: Evidence from a case of anomia with selective preservation of grammatical gender. *Cognition* 57:193–216. [TAH, aWJML, CS]

Baddeley, A., Lewis, V. & Vallar, G. (1984) Exploring the articulatory loop. *Quarterly Journal of Experimental Psychology* 36A:233–52. [aWJML]

Baumann, M. (1995) The production of syllables in connected speech. Unpublished doctoral dissertation, Nijmegen University. [aWJML]

Beauregard, M., Chertkow, H., Bub, D., Murtha, S., Dixon, R. & Evans, A. (1997) The neural substrate for concrete, abstract, and emotional word lexica: A positron emission tomography study. *Journal of Cognitive Neuroscience* 9:441–61. [HMM]

Béland, R., Caplan, D. & Nespoulous, J.-L. (1990) The role of abstract phonological representations in word production: Evidence from phonemic paraphasias. *Journal of Neurolinguistics* 5:125–64. [aWJML]

Berg, T. (1988) *Die Abbildung des Sprachproduktionsprozesses in einem Aktivationsflußmodell: Untersuchungen an deutschen und englischen Versprechern.* Niemeyer. [aWJML]

(1989) Intersegmental cohesiveness. *Folia Linguistica* 23:245–80. [aWJML]

(1991) Redundant-feature coding in the mental lexicon. *Linguistics* 29:903–25. [aWJML]

Berg, T. & Schade, U. (1992) The role of inhibition in a spreading-activation model of language production: I. The psycholinguistic perspective. *Journal of Psycholinguistic Research* 21:405–34. [LF]

Bierwisch, M. & Schreuder, R. (1992) From concepts to lexical items. *Cognition* 42:23–60. [arWJML]

Bobrow, D. G. & Winograd, T. (1977) An overview of KRL, a knowledge representation language. *Cognitive Science* 1:3–46. [aWJML]

Bock, J. K. (1982) Toward a cognitive psychology of syntax: Information processing contributions to sentence formulation. *Psychological Review* 89:1–47. [GSD]

(1986) Meaning, sound, and syntax: Lexical priming in sentence production. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 23:575–86. [GSD]

(1987) An effect of the accessibility of word forms on sentence structures. *Journal of Memory and Language* 26:119–37. [LRW]

Bock, K. & Levelt, W. (1994) Language production. Grammatical encoding. In: *Handbook of psycholinguistics,* ed. M. A. Gernsbacher. Academic Press. [aWJML, GV]

Bock, K. & Miller, C. A. (1991) Broken agreement. *Cognitive Psychology* 23:45–93. [aWJML]

Booij, G. (1995) *The phonology of Dutch.* Oxford University Press. [aWJML]

Boomer, D. S. & Laver, J. D. M. (1968) Slips of the tongue. *British Journal of Disorders of Communication* 3:2–12. [aWJML]

Braitenberg, V. (1978) Cell assemblies in the cerebral cortex. In: *Theoretical approaches to complex systems (Lecture notes in biomathematics, vol. 21),* ed. R. Heim & G. Palm. Springer. [FP]

Braitenberg, V. & Pulvermüller, F. (1992) Entwurf einer neurologischen Theorie der Sprache. *Naturwissenschaften* 79:103–17. [FP]

Braitenberg, V. & Schüz, A. (1991) *Anatomy of the cortex. Statistics and geometry.* Springer. (2nd edition in press). [FP]

Brandeis, D. & Lehmann, D. (1979) Verb and noun meaning of homophone words activate different cortical generators: A topographic study of evoked potential fields. *Experimental Brain Research* 2:159–68. [HMM]

Brennan, S. (1995) Centering attention in discourse. *Language and Cognitive Processes* 10:137–67. [PCG]

Browman, C. P. & Goldstein, L. (1988) Some notes on syllable structure in articulatory phonology. *Phonetica* 45:140–55. [aWJML]

(1990) Representation and reality: Physical systems and phonological structure. *Journal of Phonetics* 18:411–24. [aWJML]

(1992) Articulatory phonology: An overview. *Phonetica* 49:155–80. [aWJML]

Brown, A. S. (1991) A review of the tip-of-the-tongue experience. *Psychological Bulletin* 109:204–23. [rWJML]

Brown, C. (1990) Spoken word processing in context. Unpublished doctoral dissertation, Nijmegen University. [aWJML]

Brysbaert, M. (1996) Word frequency affects naming latency in Dutch when age of acquisition is controlled. *European Journal of Cognitive Psychology* 8:185–94. [aWJML]

Butterworth, B. (1989) Lexical access in speech production. In: *Lexical representation and process,* ed. W. Marslen-Wilson. MIT Press. [MZ]

Byrd, D. (1995) C-centers revisited. *Phonetica* 52:285–306. [aWJML]

(1996) Influences on articulatory timing in consonant sequences. *Journal of Phonetics* 24:209–44. [aWJML]

Caramazza, A. (1996) The brain's dictionary. *Nature* 380:485–86. [aWJML]

(1997) How many levels of processing are there in lexical access? *Cognitive Neuropsychology* 14:177–208. [TAH, MZ]

Caramazza, A. & Miozzo, M. (1997) The relation between syntactic and phonological knowledge in lexical access: Evidence from the 'tip-of-the-tongue' phenomenon. *Cognition* 64:309–43. [rWJML]

Carpenter, G. & Grossberg, S. (1987) A massively parallel architecture for a self-organizing neural recognition machine. *Computer Vision, Graphics, and Image Processing* 37:54–115. [JSB, MZ]

Carr, T. H. (1992) Automaticity and cognitive anatomy: Is word recognition "automatic"? *American Journal of Psychology* 105:201–37. [THC]

Carr, T. H. & Posner, M. I. (1995) The impact of learning to read on the functional anatomy of language processing. In: *Speech and reading: A comparative approach,* ed. B. de Gelder & J. Morais. Erlbaum. [THC]

Carroll, J. B. & White, M. N. (1973) Word frequency and age-of-acquisition as determiners of picture naming latency. *Quarterly Journal of Experimental Psychology* 25:85–95. [aWJML]

Cassidy, K. W. & Kelly, M. H. (1991) Phonological information for grammatical category assignments. *Journal of Memory and Language* 30:348–69. [MHK]

Chun, M. M. & Potter, M. C. (1995) A two-stage model for multiple target detection in RSVP. *Journal of Experimental Psychology: Human Perception and Performance* 21:109–27. [THC]

Clark, E. (1997) Conceptual perspective and lexical choice in acquisition. *Cognition* 64:1–37. [aWJML]

Clark, H. H. (1973) Space, time, semantics, and the child. In: *Cognitive development and the acquisition of language,* ed. T. Moore. Academic Press. [JDJ]

(1998) Communal lexicons. In: *Context in learning and language understanding,* ed. K. Malmkjaer & J. Williams. Cambridge University Press. [rWJML]

Cohen, J. D., Dunbar, K. & McClelland, J. L. (1990) On the control of automatic processes: A parallel distributed processing account of the Stroop effect. *Psychological Review* 97:332–61. [GSD, rWJML]

Coles, M. G. H. (1989) Modern mind-brain reading: Psychophysiology, physiology and cognition. *Psychophysiology* 26:251–69. [aWJML]

Coles, M. G. H., Gratton, G. & Donchin, E. (1988) Detecting early communication: Using measures of movement-related potentials to illuminate human information processing. *Biological Psychology* 26:69–89. [aWJML]

Collins, A. M. & Loftus, E. F. (1975) A spreading-activation theory of semantic processing. *Psychological Review* 82:407–28. [aWJML]

Collinson, W. E. (1939) Comparative synonymics: Some principles and illustrations. *Transactions of the Philological Society* 54–77. [GH]

Collyer, C. E. (1985) Comparing strong and weak models by fitting them to computer- generated data. *Perception and Psychophysics* 38:476–81. [AMJ]

Coltheart, M., Curtis, B., Atkins, P. & Haller, M. (1993) Models of reading aloud: Dual-route and parallel-distributed processing approaches. *Psychological Review* 100:589–608. [aWJML]

Costa, A. & Sebastian, N. (1998) Abstract syllabic structure in language production: Evidence from Spanish studies. *Journal of Experimental Psychology: Learning, Memory and Cognition* 24:886–903. [JS]

Crompton, A. (1982) Syllables and segments in speech production. In: *Slips of the tongue and language production,* ed. A. Cutler. Mouton. [aWJML]

Cutler, A. (in press) The syllable's role in the segmentation of stress languages. *Language and Cognitive Processes.* [aWJML]

Cutler, A., Dahan, D. & van Donselaar, W. (1997) Prosody in the comprehension of spoken language: A literature review. *Language and Speech* 40:141–201. [rWJML]

Cutler, A., Mehler, J., Norris, D. G. & Segui, J. (1986) The syllable's differing role in the segmentation of French and English. *Journal of Memory and Language* 25:385–400. [aWJML]

Cutler, A. & Norris, D. (1988) The role of strong syllables in segmentation for lexical access. *Journal of Experimental Psychology: Human Perception and Performance* 14:113–21. [aWJML]

Cutting, C. J. & Ferreira, V. S. (in press) Semantic and phonological information flow in the production lexicon. *Journal of Experimental Psychology: Learning, Memory, and Cognition.* [GSD, JDJ, aWJML]

Dagenbach, D. & Carr, T. H. (1994) Inhibitory processes in perceptual recognition: Evidence for a center-surround attentional mechanism. In: *Inhibitory processes in attention, memory, and language,* ed. D. Dagenbach & T. H. Carr. Academic Press. [THC]

Damasio, H., Grabowski, T. J., Tranel, D., Hichwa, R. D. & Damasio, A. R. (1996) A neural basis for lexical retrieval. *Nature* 380:499–505. [aWJML, HMM, FP, MZ]

Damian, M. F. & Martin, R. C. (1999) Semantic and phonological factors interact in single word production. *Journal of Experimental Psychology: Learning, Memory, and Cognition.* [rWJML]

Deacon, T. W. (1992) Cortical connections of the inferior arcuate sulcus cortex in the macaque brain. *Brain Research* 573:8–26. [FP]

De Boysson-Bardies, B. & Vihman, M. M. (1991) Adaptation to language: Evidence from babbling and first words in four languages. *Language* 67:297–318. [aWJML]

Delazer, M. & Semenza, C. (1998) The processing of compound words: A study in aphasia. *Brain and Language* 61:54–62. [rWJML, CS]

Dell, G. S. (1986) A spreading-activation theory of retrieval in sentence production. *Psychological Review* 93:283–321. [GSD, TAH, MHK, arWJML, LRW, MZ]

(1988) The retrieval of phonological forms in production: Tests of predictions from a connectionist model. *Journal of Memory and Language* 27:124–42. [AMJ, arWJML]

(1990) Effects of frequency and vocabulary type on phonological speech errors. *Language and Cognitive Processes* 5:313–49. [aWJML]

Dell, G. S., Burger, L. K. & Svec, W. R. (1997a) Language production and serial order: A functional analysis and a model. *Psychological Review* 104:123–44. [aWJML]

Dell, G. S., Juliano, C. & Govindjee, A. (1993) Structure and content in language production. A theory of frame constraints in phonological speech errors. *Cognitive Science* 17:149–95. [arWJML]

Dell, G. S. & O'Sheaghdha, P. G. (1991) Mediated and convergent lexical priming in language production: A comment on Levelt et al. (1991). *Psychological Review* 98:604–14. [TAH, aWJML, PGO]

(1992) Stages of lexical access in language production. *Cognition* 42:287–314. [JDJ, aWJML]

(1994) Inhibition in interactive activation models of linguistic selection and sequencing. In: *Inhibitory processes in attention, memory, and language,* ed. D. Dagenbach & T. H. Carr. Academic Press. [LF, rWJML]

Dell, G. S. & Reich, P. A. (1980) Toward a unified model of slips of the tongue. In: *Errors in linguistic performance: Slips of the tongue, ear, pen, and hand,* ed. V. A. Fromkin. Academic Press. [aWJML]

(1981) Stages in sentence production: An analysis of speech error data. *Journal of Verbal Learning and Verbal Behavior* 20:611–29. [TAH]

Dell, G. S., Schwartz, M. F., Martin, N., Saffran, E. M. & Gagnon, D. A. (1997b) Lexical access in aphasic and nonaphasic speakers. *Psychological Review* 104:801–38. [LF, TAH, aWJML]

Del Viso, S., Igoa, J. M. & Garcia-Albea, J. E. (1987) Corpus of spontaneous slips of the tongue in Spanish. Unpublished manuscript. [GV]

Dennett, D. (1991) *Consciousness explained.* Little, Brown. [aWJML]

Duncan, J. (1984) Selective attention and the organization of visual information. *Journal of Experimental Psychology: General* 113:501–17. [THC]

Edmonds, P. (1999) *Semantic representations of near-synonyms for automatic lexical choice.* Doctoral dissertation, Department of Computer Science, University of Toronto. [GH]

Elbers, L. (1982) Operating principles in repetitive babbling: A cognitive continuity approach. *Cognition* 12:45–63. [aWJML]

Elbers, L. & Wijnen, F. (1992) Effort, production skill, and language learning. In: *Phonological development: Models, research, implications,* ed. C. A. Ferguson & C. Stoel-Gammon. York Press. [aWJML]

Elman, J. (1995) Language as a dynamical system. In: *Mind as motion,* ed. R. Port & T. van Gelder. MIT Press. [BR]

Elman, J. & McClelland, J. (1986) The TRACE model of speech perception. *Cognitive Psychology* 18:1–86. [rWJML]

(1988) Cognitive penetration of the mechanisms of perception: Compensation for coarticulation of lexically restored phonemes. *Journal of Memory and Language* 27:143–65. [AC]

Erba, A., Zorzi, M. & Umiltà, C. (1998) Categorizzazione percettiva e semantica in soggetti umani e reti neurali. *Giornale Italiano di Psicologia* 25:123–52. [MZ]

Everaert, M., van der Linden, E. J., Schenk, A. & Schreuder, R., eds. (1995) *Idioms: Structural and psychological perspectives.* Erlbaum. [aWJML]

Farnetani, E. (1990) V-C-V lingual coarticulation and its spatiotemporal domain. In: *Speech production and speech modelling,* ed. W. J. Hardcastle & A. Marchal. Kluwer. [aWJML]

Fay, D. & Cutler, A. (1977) Malapropisms and the structure of the mental lexicon. *Linguistic Inquiry* 8:505–20. [MHK, rWJML]

Ferrand, L., Segui, J. & Grainger, J. (1996) Masked priming of word and picture naming: The role of syllable units. *Journal of Memory and Language* 35:708–23. [aWJML]

Ferrand, L., Segui, J. & Humphreys, G. W. (1997) The syllable's role in word naming. *Memory and Cognition* 25:458–70. [aWJML]

Ferreira, F. (1993) Creation of prosody during sentence production. *Psychological Review* 100:233–53. [FF]

Ferreira, V. S. (1996) Is it better to give than to donate? Syntactic flexibility in language production. *Journal of Memory and Language* 35:724–55. [GSD]

Feyerbrand, P. (1975) *Against method.* Verso. [AMJ]

Fletcher, C. R. (1984) Markedness and topic continuity in discourse processing. *Journal of Verbal Learning and Verbal Behavior* 23:487–93. [PCG]

Fodor, J. A., Garrett, M. F., Walker, E. C. T. & Parkes, C. H. (1980) Against definitions. *Cognition* 8:263–367. [arWJML]

Folkins, J. W. & Abbs, J. H. (1975) Lip and jaw motor control during speech: Responses to resistive loading of the jaw. *Journal of Speech and Hearing Research* 18:207–20. [aWJML]

Fowler, C. A., Rubin, P., Remez, R. E. & Turvey, M. T. (1980) Implications for speech production of a general theory of action. In: *Language production: Vol. I. Speech and talk,* ed. B. Butterworth. Academic Press. [aWJML]

Fowler, C. A. & Saltzman, E. (1993) Coordination and coarticulation in speech production. *Language and Speech* 36:171–95. [aWJML]

Frauenfelder, U. H. & Peeters, G. (1998) Simulating the time-course of spoken word recognition: An analysis of lexical competition in TRACE. In: *Localist connectionist approaches to human cognition,* ed. J. Grainger & A. M. Jacobs. Erlbaum. [AC]

Fromkin, V. A. (1971) The non-anomalous nature of anomalous utterances. *Language* 47:27–52. [MHK, aWJML]

Fujimura, O. (1990) Demisyllables as sets of features: Comments on Clement's paper. In: *Papers in laboratory phonology I. Between the grammar and physics of speech,* ed. J. Kingston & M. E. Beckman. Cambridge University Press. [aWJML]

Fujimura, O. & Lovins, J. B. (1978) Syllables as concatenative phonetic units. In: *Syllables and segments,* ed. A. Bell & J. B. Hooper. North-Holland. [aWJML]

Fuster, J. M. (1994) *Memory in the cerebral cortex. An empirical approach to neural networks in the human and nonhuman primate.* MIT Press. [FP]

Garcia-Albea, J. E., del Viso, S. & Igoa, J. M. (1989) Movement errors and levels of processing in sentence production. *Journal of Psycholinguistic Research* 18:145–61. [aWJML]

Garrett, M. F. (1975) The analysis of sentence production. In: *The psychology of learning and motivation: Vol. 9,* ed. G. H. Bower. Academic Press. [arWJML]

(1980) Levels of processing in sentence production. In: *Language production: Vol. I. Speech and talk,* ed. B. Butterworth. Academic Press. [aWJML, GV, MZ]

(1988) Processes in language production. In: *Linguistics: The Cambridge survey. Vol. III. Biological and psychological aspects of language,* ed. F. J. Nieuwmeyer. Harvard University Press. [arWJML]

(1992) Lexical retrieval processes: Semantic field effects. In: *Frames, fields and contrasts: New essays in semantic and lexical organization,* ed. E. Kittay & A. Lehrer. Erlbaum. [MZ]

Glaser, M. O. & Glaser, W. R. (1982) Time course analysis of the Stroop phenomenon. *Journal of Experimental Psychology: Human Perception and Performance* 8:875–94. [rWJML]

Glaser, W. R. (1992) Picture naming. *Cognition* 42:61–105. [arWJML]

Glaser, W. R. & Düngelhoff, F.-J. (1984) The time course of picture-word interference. *Journal of Experimental Psychology: Human Perception and Performance* 10:640–54. [arWJML]

Glaser, W. R. & Glaser, M. O. (1989) Context effects in Stroop-like word and picture processing. *Journal of Experimental Psychology: General* 118:13–42. [arWJML]

Goldman, N. (1975) Conceptual generation. In: *Conceptual information processing,* ed. R. Schank. North-Holland. [aWJML]

Goldsmith, J. A. (1990) *Autosegmental and metrical phonology.* Basil Blackwell. [aWJML]

Gordon, P. C. & Hendrick, R. (in press) The representation and processing of coreference in discourse. *Cognitive Science.* [PCG]

(1997) Intuitive knowledge of linguistic co-reference. *Cognition* 62:325–70. [PCG]

Gove, P., ed. (1984) *Webster's new dictionary of synonyms.* Merriam-Webster. [GH]

Grainger, J. & Jacobs, A. M. (1996) Orthographic processing in visual word recognition: A multiple read-out model. *Psychological Review* 103:518–65. [AMJ]

(1998) On localist connectionism and psychological science. In: *Localist connectionist approaches to human cognition,* ed. J. Grainger & A. M. Jacobs. Erlbaum. [AMJ]

Grossberg, S. (1976a) Adaptive pattern classification and universal recoding: 1. Parallel development and coding of neural feature detectors. *Biological Cybernetics* 23:121–34. [MZ]

(1976b) Adaptive pattern classification and universal recoding: 2. Feedback, expectation, olfaction, and illusions. *Biological Cybernetics* 23:187–203. [JSB]

Grosz, B. J., Joshi, A. K. & Weinstein, S. (1995) Centering: A framework for modelling the local coherence of discourse. *Computational Linguistics* 21:203–26. [PCG]

Harley, T. A. (1984) A critique of top-down independent levels models of speech production: Evidence from non-plan-internal speech errors. *Cognitive Science* 8:191–219. [TAH]

(1993) Phonological activation of semantic competitors during lexical access in speech production. *Language and Cognitive Processes* 8:291–309. [TAH, JDJ, aWJML, PGO]

Harley, T. A. & Brown, H. (1998) What causes tip-of-the-tongue states? *British Journal of Psychology* 89:151–74. [TAH]

Hayakawa, S. I., ed. (1987) *Choose the right word: A modern guide to synonyms.* Harper and Row. [GH]

Henaff Gonon, M. A., Bruckert, R. & Michel, F. (1989) Lexicalization in an anomic patient. *Neuropsychologia* 27:391–407. [aWJML]

Hendriks, H. & McQueen, J., eds. (1996) *Annual Report 1995.* Max Planck Institute for Psycholinguistics, Nijmegen. [aWJML]

Hervey, S. & Higgins, I. (1992) *Thinking translation: A course in translation method.* Routledge. [GH]

Hinton, G. E. & Shallice, T. (1991) Lesioning an attractor network: Investigations of acquired dyslexia. *Psychological Review* 99:74–95. [TAH]

Hird, K. & Kirsner, K. (1993) Dysprosody following acquired neurogenic impairment. *Brain and Language* 45:46–60. [BR]

Hirst, G. (1995) Near-synonymy and the structure of lexical knowledge. *Working notes, AAAI symposium on representation and acquisition of lexical knowledge: Polysemy, ambiguity, and generativity.* Stanford University. [GH]

Hittmair-Delazer, M., Semenza, C., De Bleser, R. & Benke, T. (1994) Naming by German compounds. *Journal of Neurolinguistics* 8:27–41. [rWJML, CS]

Humphreys, G. W., Riddoch, M. J. & Price, C. J. (1997) Top-down processes in object identification : Evidence from experimental psychology, neuropsychology and functional anatomy. *Philosophical Transactions of the Royal Society of London B* 352:1275–82. [LF]

Humphreys, G. W., Riddoch, M. J. & Quinlan, P. T. (1988) Cascade processes in picture identification. *Cognitive Neuropsychology* 5:67–103. [aWJML]

Humphreys, G. W., Lamote, C. & Lloyd-Jones, T. J. (1995) An interactive activation approach to object processing: Effects of structural similarity, name frequency and task in normality and pathology. *Memory* 3:535–86. [aWJML]

Indefrey, P. G. & Levelt, W. J. M. (in press) The neural correlates of language production. In: *The cognitive neurosciences, 2nd edition,* ed. M. Gazzaniga. MIT Press. [aWJML]

Jackendoff, R. (1987) *Consciousness and the computational mind.* MIT Press. [aWJML]

Jacobs, A. M. & Carr, T. H. (1995) Mind mappers and cognitive modelers: Toward cross- fertilization. *Behavioral and Brain Sciences* 18:362–63. [aWJML]

Jacobs, A. M. & Grainger, J. (1994) Models of visual word recognition: Sampling the state of the art. *Journal of Experimental Psychology: Human Perception and Performance* 20:1311–34. [LF, AMJ]

Jacobs, A. M., Rey, A., Zeigler, J. C. & Grainger, J. (1998) MROM-P: An interactive activation, multiple read-out model of orthographic and phonological processes in visual word recognition. In: *Localist connectionist approaches to human cognition,* ed. J. Grainger & A. M. Jacobs. Erlbaum. [AMJ]

Jeannerod, M. (1994) The representing brain: Neural correlates of motor intention and imagery. *Behavioral and Brain Sciences* 17:187–245. [aWJML]

Jescheniak, J. D. (1994) *Word frequency effects in speech production.* Doctoral dissertation, Nijmegen University. [aWJML]

Jescheniak, J. D. & Levelt, W. J. M. (1994) Word frequency effects in speech production: Retrieval of syntactic information and of phonological form. *Journal of Experimental Psychology: Language, Memory, and Cognition* 20:824–43. [arWJML]

Jescheniak, J. D. & Schriefers, H. (1997) Lexical access in speech production: Serial or cascaded processing? *Language and Cognitive Processes* 12:847–52. [JDJ]

(1998) Discrete serial versus cascaded processing in lexical access in speech production: Further evidence from the co-activation of near-synonyms. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 24:1256–74. [JDJ, aWJML, PGO]

(in preparation) Lexical access in speech production: The case of antonyms. [JDJ]

Jescheniak, J. D., Schriefers, H. & Hantsch, A. (submitted) Semantic and phonological activation in noun and pronoun production. [JDJ]

Kanwisher, N. & Driver, J. (1993) Objects, attributes, and visual attention: Which, what, and where. *Current Directions in Psychological Science* 1:26–31. [THC]

Kawamoto, A. H., Kello, C. T., Jones, R. M. & Bame, K. A. (1998) Initial phoneme versus whole word criterion to initiate pronunciation: Evidence based on response latency and initial phoneme duration. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 24:862–85. [AHK]

Kelly, M. H. (1992) Using sound to solve syntactic problems: The role of phonology in grammatical category assignments. *Psychological Review* 99:349–64. [MHK]

Kelly, M. H. & Bock, J. K. (1988) Stress in time. *Journal of Experimental Psychology: Human Perception and Performance* 14:389–403. [MHK]

Kelso, J. A. S., Saltzman, E. L. & Tuller, B. (1986) The dynamical perspective on speech production: Data and theory. *Journal of Phonetics* 14:29–59. [aWJML]

Kempen, G. (1997) *Grammatical performance in human sentence production and comprehension.* Report Leiden University, Psychology Department. [rWJML]

Kempen, G. & Hoenkamp, E. (1987) An incremental procedural grammar for sentence formulation. *Cognitive Science* 11:201–58. [aWJML]

Kempen, G. & Huijbers, P. (1983) The lexicalization process in sentence production and naming: Indirect election of words. *Cognition* 14:185–209. [aWJML]

Kiritani, S. & Sawashima, M. (1987) The temporal relationship between articulations of consonants and adjacent vowels. In: *In honor of Ilse Lehiste,* ed. R. Channon & L. Shockey. Foris. [aWJML]

Kohonen, T. (1984) *Self-organization and associative memory.* Springer. [MZ]

Kuçera, H. & Francis, W. N. (1967) *Computational analysis of present day English.* Brown University Press. [PCG]

Kutas, M. (1997) Views on how the electrical activity that the brain generates reflects the functions of different language structures. *Psychophysiology* 34:383–98. [HMM]

Kutas, M. & Hillyard, S. A. (1983) Event-related brain potentials to grammatical errors and semantic anomalies. *Memory and Cognition* 11:539–50. [HMM]

La Heij, W. (1988) Components of Stroop-like interference in picture naming. *Memory and Cognition* 16:400–10. [PAS]

La Heij, W., Dirkx, J. & Kramer, P. (1990) Categorical interference and associative priming in picture naming. *British Journal of Psychology* 81:511–25. [PAS]

La Heij, W. & van den Hof, E. (1995) Picture-word interference increases with target-set size. *Psychological Research* 58:119–33. [rWJML, PAS]

Laine, M. & Martin, N. (1996) Lexical retrieval deficit in picture-naming: Implications for word-production models. *Brain and Language* 53:283–314. [LF, rWJML]

Lashley, K. S. (1951) The problem of serial order in behavior. In: *Cerebral mechanisms in behavior. The Hixxon symposium,* ed. L. A. Jeffress. Wiley. [FP]

Levelt, C. C. (1994) *The acquisition of place.* Holland Institute of Generative Linguistics Publications. [aWJML]

Levelt, W. J. M. (1983) Monitoring and self-repair in speech. *Cognition* 14:41–104. [aWJML]

(1989) *Speaking: From intention to articulation.* MIT Press. [AC, TAH, MHK, arWJML, BR, MZ]

(1992) Accessing words in speech production: Stages, processes and representations. *Cognition* 42:1–22. [JSB, rWJML]

(1993) Lexical selection, or how to bridge the major rift in language processing. In: *Theorie und Praxis des Lexikons,* ed. F. Beckmann & G. Heyer. De Gruyter. [aWJML]

(1996) Perspective taking and ellipsis in spatial descriptions. In: *Language and space,* ed. P. Bloom, M. A. Peterson, L. Nadel & M. F. Garrett. MIT Press. [aWJML]

Levelt, W. J. M. & Kelter, S. (1982) Surface form and memory in question answering. *Cognitive Psychology* 14:78–106. [aWJML]

Levelt, W. J. M., Praamstra, P., Meyer, A. S., Helenius, P. & Salmelin, R. (1998) A MEG study of picture naming. *Journal of Cognitive Neuroscience* 10:553–67. [arWJML]

Levelt, W. J. M., Schreuder, R. & Hoenkamp, E. (1978) Structure and use of verbs of motion. In: *Recent advances in the psychology of language,* ed. R. N. Campbell & P. T. Smith. Plenum. [aWJML]

Levelt, W. J. M., Schriefers, H., Vorberg, D., Meyer, A. S., Pechmann, T. & Havinga, J. (1991a) The time course of lexical access in speech production: A study of picture naming. *Psychological Review* 98:122–42. [TAH, JDJ, aWJML, FP]

(1991b) Normal and deviant lexical processing: Reply to Dell and O'Seaghdha. *Psychological Review* 98:615–18. [aWJML]

Levelt, W. J. M. & Wheeldon, L. (1994) Do speakers have access to a mental syllabary? *Cognition* 50:239–69. [aWJML]

Liberman, A. (1996) *Speech: A special code.* MIT Press. [aWJML]

Lindblom, B. (1983) Economy of speech gestures. In: *The production of speech,* ed. P. F. MacNeilage. Springer. [aWJML]

Lindblom, B., Lubker, J. & Gay, T. (1979) Formant frequencies of some fixed-mandible vowels and a model of speech motor programming by predictive simulation. *Journal of Phonetics* 7:147–61. [aWJML]

Locke, J. (1997) A theory of neurolinguistic development. *Brain and Language* 58:265–326. [BR]

MacKay, D. G. (1987) *The organization of perception and action: A theory for language and other cognitive skills.* Springer-Verlag. [JS]

(1993) The theoretical epistemology: A new perspective on some long-standing methodological issues in psychology. In: *Methodological and quantitative issues in the analysis of psychological data,* ed. G. Keren & C. Lewis. Erlbaum. [JS]

MacLeod, C. M. (1991) Half a century of research on the Stroop effect: An integrative review. *Psychological Bulletin* 109:163–203. [rWJML]

Maki, W. S., Frigen, K. & Paulson, K. (1997) Associative priming by targets and distractors during rapid serial visual presentation: Does word meaning survive the attentional blink? *Journal of Experimental Psychology: Human Perception and Performance* 23:1014–34. [THC]

Marr, D. (1982) *Vision.* Freeman. [aWJML]

Marslen-Wilson, W. D. (1987) Functional parallelism in spoken word recognition. *Cognition* 25:71–102. [rWJML]

Marslen-Wilson, W. D., Levy, W. & Tyler, L. K. (1982) Producing interpretable discourse: The establishment and maintenance of reference. In: *Speech, place, and action,* ed. R. Jarvella & W. Klein. Wiley. [PCG]

Martin, N., Gagnon, D. A., Schwartz, M. F., Dell, G. S. & Saffran, E. M. (1996) Phonological facilitation of semantic errors in normal and aphasic speakers. *Language and Cognitive Processes* 11:257–82. [LF, aWJML, PGO]

Martin, A., Wiggs, C. L., Ungerleider, L. G. & Haxby, J. V. (1996) Neural correlates of category-specific knowledge. *Nature* 379:649–52. [rWJML]

McClelland, J. L. (1979) On the time relations of mental processes: An examination of systems of processes in cascade. *Psychological Review* 56:287–330. [aWJML]

McClelland, J. L. & Elman, J. L. (1986) The TRACE model of speech perception. *Cognitive Psychology* 18:1–86. [AC]

McClelland, J. L. & Rumelhart, D. E. (1981) An interactive activation model of context effects in letter perception: 1. An account of the basic findings. *Psychological Review* 88:357–407. [AC]

McGuire, P. K., Silbersweig, D. A. & Frith, C. D. (1996) Functional neuroanatomy of verbal self-monitoring. *Brain* 119:101–11. [aWJML]

McQueen, J. M., Cutler, A., Briscoe, T. & Norris, D. (1995) Models of continuous speech recognition and the contents of the vocabulary. *Language and Cognitive Processes* 10:309–31. [aWJML]

Mehler, J., Dommergues, J. & Frauenfelder, U. (1981) The syllable's role in speech segmentation. *Journal of Verbal Learning and Verbal Behavior* 20:289–305. [aWJML]

Mehler, J., Altmann, G., Sebastian, N., Dupoux, E., Christophe, A. & Pallier, C. (1993) Understanding compressed sentences: The role of rhythm and meaning. *Annals of the New York Academy of Science* 682:272–82. [aWJML]

Meijer, P. J. A. (1994) Phonological encoding: The role of suprasegmental structures. Unpublished doctoral dissertation, Nijmegen University. [aWJML, JS]

(1996) Suprasegmental structures in phonological encoding: The CV structure. *Journal of Memory and Language* 35:840–53. [aWJML, JS]

Meyer, A. S. (1990) The time course of phonological encoding in language production: The encoding of successive syllables of a word. *Journal of Memory and Language* 29:524–45. [aWJML]

(1991) The time course of phonological encoding in language production: Phonological encoding inside a syllable. *Journal of Memory and Language* 30:69–89. [AHK, aWJML]

(1992) Investigation of phonological encoding through speech error analyses: Achievements, limitations, and alternatives. *Cognition* 42:181–211. [aWJML]

(1994) Timing in sentence production. *Journal of Memory and Language* 33:471–92. [FF, rWJML]

(1996) Lexical access in phrase and sentence production: Results from picture-word interference experiments. *Journal of Memory and Language* 35:477–96. [aWJML, PAS]

(1997) Word form generation in language production. In: *Speech production: Motor control, brain research and fluency disorders,* ed. W. Hulstijn, H. F. M. Peters & P. H. H. M. van Lieshout. Elsevier. [aWJML]

Meyer, A. S., Roelofs, R. & Schiller, N. (in preparation) Prosodification of metrically regular and irregular words. [aWJML]

Meyer, A. S. & Schriefers, H. (1991) Phonological facilitation in picture-word interference experiments: Effects of stimulus onset asynchrony and types of interfering stimuli. *Journal of Experimental Psychology: Language, Memory, and Cognition* 17:1146–60. [aWJML, PAS]

Meyer, A. S., Sleiderink, A. M. & Levelt, W. J. M. (1998) Viewing and naming objects: Eye movements during noun phrase production. *Cognition* 66:B25-B33. [rWJML]

Miller, G. A. (1956) The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review* 63:81–97. [aWJML]

(1991) *The science of words.* Scientific American Library. [aWJML]

Miller, G. A. & Johnson-Laird, P. N. (1976) *Language and perception.* Harvard University Press. [aWJML]

Miller, J. & Weinert, R. (1998) *Spontaneous spoken language: Syntax and discourse.* Clarendon Press. [MS]

Minsky, M. & Papert, S. (1969) *Perceptrons.* MIT Press. [MZ]

Mondini, S., Luzzatti, C., Semenza, C. & Calza, A. (1997) Prepositional compounds are sensitive to agrammatism: Consequences for models of lexical retrieval. *Brain and Language* 60:78–80. [CS]

Morrison, C. M., Ellis, A. W. & Quinlan, P. T. (1992) Age of acquisition, not word frequency, affects object naming, not object recognition. *Memory and Cognition* 20:705–14. [aWJML]

Morton, J. (1969) The interaction of information in word recognition. *Psychological Review* 76:165–78. [aWJML]

Mowrey, R. A. & MacKay, I. R. A. (1990) Phonological primitives: Electromygraphic speech error evidence. *Journal of the Acoustical Society of America* 88:1299–312. [aWJML]

Müller, H. M., King, J. W. & Kutas, M. (1997) Event related potentials elicited by spoken relative clauses. *Cognitive Brain Research* 5:193–203. [HMM]

Müller, H. M. & Kutas, M. (1996) What's in a name? Electrophysiological differences between spoken nouns, proper names, and one's own name. *NeuroReport* 8:221–25. [HMM]

Munhall, K. G. & Löfqvist, A. (1992) Gestural aggregation in speech: Laryngeal gestures. *Journal of Phonetics* 20:111–26. [aWJML]

Myung, I. J. & Pitt, M. A. (1998) Issues in selecting mathematical models of cognition. In: *Localist connectionist approaches to human cognition,* ed. J. Grainger & A. M. Jacobs. Erlbaum. [AMJ]

Nespor, M. & Vogel, I. (1986) *Prosodic phonology.* Foris. [aWJML]

Nickels, L. (1995) Getting it right? Using aphasic naming errors to evaluate theoretical models of spoken word recognition. *Language and Cognitive Processes* 10:13–45. [aWJML]

Nooteboom, S. (1969) The tongue slips into patterns. In: *Nomen: Leyden studies in linguistics and phonetics,* ed. A. G. Sciarone, A. J. van Essen & A. A. van Rood. Mouton. [aWJML]

Norris, D. (1990) Connectionism: A case for modularity. In: *Comprehension processes in reading,* ed. D. A. Balota, G. B. Flores d'Arcais & K. Rayner. Erlbaum. [AC]

(1993) Bottom-up connectionist models of 'interaction.' In: *Cognitive models of speech processing,* ed. G. Altmann & R. Shillcock. Erlbaum. [AC]

(1994) Shortlist: A connectionist model of continuous speech recognition. *Cognition* 52:189–234. [rWJML]

Norris, D., McQueen, J. M. & Cutler, A. (submitted) Merging phonetic and lexical information in phonetic decision-making. [AC]

Oldfield, R. C. & Wingfield, A. (1965) Response latencies in naming objects. *Quarterly Journal of Experimental Psychology* 17:273–81. [aWJML]

O'Seaghdha, P. G., Dell, G. S., Peterson, R. R. & Juliano, C. (1992) Modelling form- related priming effects in comprehension and production. In: *Connectionist approaches to language processing, vol. 1,* ed. R. Reilly & N. E. Sharkey. Erlbaum. [LRW]

O'Seaghdha, P. G. & Marin, J. W. (1997) Mediated semantic-phonological priming: Calling distant relatives. *Journal of Memory and Language* 36:226–52. [JDJ, PGO]

(in press) Phonological competition and cooperation in form-related priming: Sequential and nonsequential processes in word production. *Journal of Experimental Psychology: Human Perception and Performance.* [PGO, rWJML]

Paller, K. A., Kutas, M., Shimamura, A. P. & Squire, L. R. (1992) Brain responses to concrete and abstract words reflect processes that correlate with later performance on a test of stem-completion priming. *Electroencephalography and Clinical Neurophysiology (Supplement)* 40:360–65. [HMM]

Pandya, D. N. & Yeterian, E. H. (1985) Architecture and connections of cortical association areas. In: *Cerebral cortex. Vol. 4: Association and auditory cortices,* ed. A. Peters & E. G. Jones. Plenum Press. [FP]

Pashler, H. (1994) Dual-task interference in simple tasks: Data and theory. *Psychological Bulletin* 116:220–44. [THC]

Peterson, R. R., Dell, G. S. & O'Sheaghdha, P. G. (1989) A connectionist model of form- related priming effects. In: *Proceedings of the Eleventh Annual Conference of the Cognitive Science Society,* 196–203. Erlbaum. [PGO]

Peterson, R. R. & Savoy, P. (1998) Lexical selection and phonological encoding during language production: Evidence for cascaded processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 24:539–57. [GH, JDJ, arWJML, PGO]

Pitt, M. A. & McQueen, J. M. (1998) Is compensation for coarticulation mediated by the lexicon? *Journal of Memory and Language.* 39:347–70. [AC]

Platt, J. R. (1964) Strong inference. *Science* 146:347–53. [AMJ]

Posner, M. I. & Raichle, M. E. (1994) *Images of mind.* Scientific American Library. [THC]

Potter, M. C. (1983) Representational buffers: The eye-mind hypothesis in picture perception, reading, and visual search. In: *Eye movements in reading: Perceptual and language processes,* ed. K. Rayner. Academic Press. [aWJML]

Prechelt, L. (1996) A quantitative study of experimental evaluations of neural network learning algorithms: Current research practice. *Neural Networks* 9:457–62. [AMJ]

Preissl, H., Pulvermüller, F., Lutzenberger, W. & Birbaumer, N. (1995) Evoked potentials distinguish between nouns and verbs. *Neuroscience Letters* 197:81–83. [HMM]

Pulvermüller, F. (1992) Constituents of a neurological theory of language. *Concepts in Neuroscience* 3:157–200. [FP]

(1995) Agrammatism: Behavioral description and neurobiological explanation. *Journal of Cognitive Neuroscience* 7:165–81. [FP]

(1996) Hebb's concept of cell assemblies and the psychophysiology of word processing. *Psychophysiology* 33:317–33. [FP]

Rafal, R. & Henik, A. (1994) The neurology of inhibition: Integrating controlled and automatic processes. In: *Inhibitory processes in attention, memory, and language,* ed. D. Dogenbach & T. H. Carr. Academic Press. [THC]

Ratcliff, R. & Murdock, B. B., Jr. (1976) Retrieval processes in recognition memory. *Psychological Review* 83:190–214. [AMJ]

Recasens, D. (1984) V-to-C coarticulation in Catalan VCV sequences: An articulatory and acoustic study. *Journal of Phonetics* 12:61–73. [aWJML]

(1987) An acoustic analysis of V-to-C and V-to-V coarticulatory effects in Catalan and Spanish VCV sequences. *Journal of Phonetics* 15:299–312. [aWJML]

Roberts, S. & Sternberg, S. (1993) The meaning of additive reaction-time effects: Tests of three alternatives. In: *Attention and performance XIV,* ed. D. E. Meyer & S. Kornblum. MIT Press. [aWJML]

Roelofs, A. (1992a) A spreading-activation theory of lemma retrieval in speaking. *Cognition* 42:107–42. [arWJML, PAS]

(1992b) *Lemma retrieval in speaking: A theory, computer simulations, and empirical data.* Doctoral dissertation, NICI Technical Report 92–08, University of Nijmegen. [arWJML, PAS]

(1993) Testing a non-decompositional theory of lemma retrieval in speaking: Retrieval of verbs. *Cognition* 47:59–87. [arWJML]

(1994) On-line versus off-line priming of word-form encoding in spoken word production. In: *Proceedings of the Sixteenth Annual Conference of the Cognitive Science Society,* ed. A. Ram & K. Eiselt. Erlbaum. [aWJML]

(1996a) Computational models of lemma retrieval. In: *Computational psycholinguistics: AI and connectionist models of human language processing,* ed. T. Dijkstra & K. De Smedt. Taylor and Francis. [aWJML]

(1996b) Serial order in planning the production of successive morphemes of a word. *Journal of Memory and Language* 35:854–76. [arWJML]

(1996c) Morpheme frequency in speech production: Testing WEAVER. In: *Yearbook of morphology,* ed. G. E. Booij & J. van Marle. Kluwer Academic Press. [aWJML]

(1997a) A case for non-decomposition in conceptually driven word retrieval. *Journal of Psycholinguistic Research* 26:33–67. [GH, aWJML]

(1997b) Syllabification in speech production: Evaluation of WEAVER. *Language and Cognitive Processes* 12:657–94. [aWJML]

(1997c) The WEAVER model of word-form encoding in speech production. *Cognition* 64:249–84. [arWJML, PAS]

(1998) Rightward incrementality in encoding simple phrasal forms in speech production: Verb-particle combinations. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 24:904–19. [aWJML]

(submitted a) Parallel activation and serial selection in planning linguistic performance: Evidence from implicit and explicit priming. [aWJML]

(submitted b) WEAVER++ and other computational models of lemma retrieval and word-form encoding. [aWJML]

(submitted c) Discrete or continuous access of lemmas and forms in speech production? [rWJML]

Roelofs, A., Baayen, H. & Van den Brink, D. (submitted) Semantic transparency in producing polymorphemic words. [aWJML]

Roelofs, A. & Meyer, A. S. (1998) Metrical structure in planning the production of spoken words. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 24:922–39. [aWJML, JS]

Roelofs, A., Meyer, A. S. & Levelt, W. J. M. (1996) Interaction between semantic and orthographic factors in conceptually driven naming: Comment on Starreveld and La Heij (1995). *Journal of Experimental Psychology: Learning, Memory, and Cognition* 22:246–51. [arWJML, PAS]

(1998) A case for the lemma-lexeme distinction in models of speaking: Comment on Caramazza and Miozzo (1997). *Cognition* 69:219–30. [rWJML]

Romani, C. (1992) The representation of prosodic and syllabic structure in speech production. Unpublished doctoral dissertation, Johns Hopkins University. [JS]

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M. & Boyes-Braem, P. (1976) Basic objects in natural categories. *Cognitive Psychology* 8:382–439. [aWJML]

Rosenbaum, D. A., Kenny, S. B. & Derr, M. A. (1983) Hierarchical control of rapid movement sequences. *Journal of Experimental Psychology: Human Perception and Performance* 9:86–102. [aWJML]

Rossi, M. & Defare, E. P. (1995) Lapsis linguae: Word errors or phonological errors? *International Journal of Psycholinguistics* 11:5–38. [aWJML]

Rumelhart, D. E. & McClelland, J. L. (1982) An interactive activation model of context effects in letter perception: 2. The contextual enhancement effect and some tests and extensions of the model. *Psychological Review* 89:60–94. [AC]

Rumelhart, D. E., Hinton, G. E. & Williams, R. J. (1986) Learning internal representations by error propagation. In: *Parallel distributed processing: Explorations in the microstructure of cognition: Vol. 1. Foundations,* ed. D. E. Rumelhart & J. L. McClelland. MIT Press. [MZ]

Rumelhart, D. E. & Zipser, D. (1985) Feature discovery by competitive learning. *Cognitive Science* 9:75–112. [MZ]

Santiago, J. (1997) Estructura silábica y activation secuencial en produccion del lenguaje. [Syllable structure and sequential activation in language production]. Unpublished doctoral dissertation, Department Psicologia Experimental y Fisiologia del Comportamiento, Universidad de Granada, Spain. [JS]

Santiago, J., MacKay, D. G., Palma, A. & Rho, C. (submitted) Picture naming latencies vary with sequential activation processes in hierarchic syllable and word structures. [JS]

Schade, U. (1996) *Konnektionistische Sprachproduktion [Connectionist language production].* Habilitationsschrift, Bielefeld University. [AMJ, rWJML]

Schade, U. & Berg, T. (1992) The role of inhibition in a spreading-activation model of language production: II. The simulational perspective. *Journal of Psycholinguistic Research* 21:435–62. [LF]

Schiller, N. (1997) The role of the syllable in speech production. Evidence from lexical statistics, metalinguistics, masked priming, and electromagnetic midsagittal articulography. Unpublished doctoral dissertation, Nijmegen University. [aWJML]

(1998) The effect of visually masked syllable primes on the naming latencies of words and pictures. *Journal of Memory and Language* 39:484–507. [aWJML]

Schiller, N., Meyer, A. S., Baayen, R. H. & Levelt, W. J. M. (1966) A comparison of lexeme and speech syllables in Dutch. *Journal of Quantitative Linguistics* 3:8–28. [aWJML]

Schiller, N., Meyer, A. S. & Levelt, W. J. M. (1997) The syllabic structure of spoken words: Evidence from the syllabification of intervocalic consonants. *Language and Speech* 40:101–39. [aWJML]

Schmitt, B. (1997) Lexical access in the production of ellipsis and pronouns. Unpublished doctoral dissertation, Nijmegen University. [rWJML]

Schriefers, H. (1990) Lexical and conceptual factors in the naming of relations. *Cognitive Psychology* 22:111–42. [aWJML]

(1993) Syntactic processes in the production of noun phrases. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 19:841–50. [aWJML]

Schriefers, H., Meyer, A. S. & Levelt, W. J. M. (1990) Exploring the time course of lexical access in speech production: Picture-word interference studies. *Journal of Memory and Language* 29:86–102. [aWJML, PAS]

Schriefers, H., Teruel, E. & Meinshausen, R. M. (1998) Producing simple sentences: Results from picture-word interference experiments. *Journal of Memory and Language* 39:609–32. [aWJML]

Seidenberg, M. S. & McClelland, J. L. (1989) A distributed, developmental model of word recognition and naming. *Psychological Review* 96:523–68. [aWJML]

Selkirk, E. (1984) *Phonology and syntax.* MIT Press. [FF]

Semenza, C. (1998) Lexical semantic disorders in aphasia. In: *Handbook of Neuropsychology,* ed. G. Denes & L. Pizzamiglio. Erlbaum. [CS]

Semenza, C., Luzzatti, C. & Carabelli, S. (1997) Naming disorders for compounds in aphasia: A study in Italian. *Journal of Neurolinguistics* 10:33–43. [CS]

Semenza, C., Mondini, S. & Cappelletti, M. (1997) The grammatical properties of mass nouns: An aphasia case study. *Neuropsychologica* 35:669–75. [rWJML, CS]

Sereno, J. (1994) Phonosyntactics. In: *Sound symbolism,* ed. L. Hinton, J. Nichols & J. J. Ohala. Cambridge University Press. [MHK]

Sevald, C. A. & Dell, G. S. (1994) The sequential cuing effect in speech production. *Cognition* 53:91–127. [PGO, JS, LRW]

Sevald, C. A., Dell, G. S. & Cole, J. S. (1995) Syllable structure in speech production: Are syllables chunks or schemas? *Journal of Memory and Language* 34:807–20. [aWJML, JS]

Shallice, T. & McGill, J. (1978) The origins of mixed errors. In: *Attention and performance VII,* ed. J. Requin. Erlbaum. [TAH]

Shapiro, K. L. & Raymond, J. E. (1994) Temporal allocation of visual attention: Inhibition or interference? In: *Inhibitory processes in attention, memory, and language,* ed. D. Dagenbach & T. H. Carr. Academic Press. [THC]

Shastri, L. & Ajjanagadde, V. (1993) From simple associations to systematic reasoning: A connectionist representation of rules, variables and dynamic bindings using temporal synchrony. *Behavioral and Brain Sciences* 16:417–94. [rWJML]

Shattuck-Hufnagel, S. (1979) Speech errors as evidence for a serial-ordering mechanism in sentence production. In: *Sentence processing: Psycholinguistic studies presented to Merrill Garrett,* ed. W. E. Cooper & E. C. T. Walker. Erlbaum. [arWJML]

(1983) Sublexical units and suprasegmental structure in speech production planning. In: *The production of speech,* ed. P. F. MacNeilage. Springer. [aWJML]

(1985) Context similarity constraints on segmental speech errors: An experimental investigation of the role of word position and lexical stress. In: *On the planning and production of speech in normal and hearing-impaired individuals: A seminar in honor of S. Richard Siverman,* ed. J. Lauter. ASHA Reports, No. 15. [aWJML]

(1987) The role of word-onset consonants in speech production planning: New evidence from speech error patterns. In: *Memory and sensory processes of language,* ed. E. Keller & M. Gopnik. Erlbaum. [aWJML]

(1992) The role of word structure in segmental serial ordering. *Cognition* 42:213–59. [aWJML]

Shattuck-Hufnagel, S. & Klatt, D. H. (1979) The limited use of distinctive features and markedness in speech production: Evidence from speech error data. *Journal of Verbal Learning and Verbal Behavior* 18:41–55. [aWJML]

Sherman, D. (1975) Noun-verb stress alternation: An example of lexical diffusion of sound change. *Linguistics* 159:43–71. [MHK]

Slobin, D. I. (1987) Thinking for speaking. In: *Berkeley Linguistics Society: Proceedings of the Thirteenth Annual Meeting,* ed. J. Aske, N. Beery, L. Michaelis & H. Filip. Berkeley Linguistics Society. [arWJML]

(1996) From "thought and language" to "thinking for speaking." In: *Rethinking linguistic relativity,* ed.. J. J. Gumperz & S. C. Levinson. Cambridge University Press. [rWJML]

(1998) Verbalized events: A dynamic approach to linguistic relativity and determinism. *Proceedings of the LAUD-Symposium,* Duisburg. [rWJML]

Snodgrass, J. G. & Yuditsky, T. (1996) Naming times for the Snodgrass and Vanderwart pictures. *Behavioral Research Methods, Instruments and Computers* 28:516–36. [aWJML]

Sptizer, M., Kwong, K. K., Kennedy, W., Rosen, B. R. & Belliveau, J. W. (1995) Category-specific brain activation in fMRI during picture naming. *NeuroReport* 6:2109–112. [HMM]

Starreveld, P. A. & La Heij, W. (1995) Semantic interference, orthographic facilitation and their interaction in naming tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 21:686–98. [arWJML, PAS]

(1996a) The locus of orthographic-phonological facilitation: A reply to Roelofs, Meyer, and Levelt (1996). *Journal of Experimental Psychology: Learning, Memory, and Cognition* 22:252–55. [aWJML]

(1996b) Time-course analysis of semantic and orthographic context effects in picture naming. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 22:896–918. [rWJML, PAS]

Stede, M. (1996) Lexical paraphrases in multilingual sentence generation. *Machine Translation* 11:75–107. [GH]

Stemberger, J. P. (1983) *Speech errors and theoretical phonology: A review.* Indiana University Linguistics Club. [aWJML]

(1985) An interactive activation model of language production. In: *Progress in the psychology of language, vol. I,* ed. A. W. Ellis. Erlbaum. [LF, aWJML, LRW]

(1990) Wordshape errors in language production. *Cognition* 35:123–57. [aWJML]

(1991a) Radical underspecification in language production. *Phonology* 8:73–112. [aWJML]

(1991b) Apparent anti-frequency effects in language production: The addition bias and phonological underspecification. *Journal of Memory and Language* 30:161–85. [aWJML]

Stemberger, J. P. & MacWhitney, B. (1986) Form-oriented inflectional errors in language processing. *Cognitive Psychology* 18:329–54. [aWJML]

Stemberger, J. P. & Stoel-Gammon, C. (1991) The underspecification of coronals: Evidence from language acquisition and performance errors. *Phonetics and Phonology* 2:181–99. [aWJML]

Sternberg, S. (1969) The discovery of processing stages: Extensions of Donders' method. *Attention and Performance II,* ed. W. G. Koster. *Acta Psychologica* 30:276–315. [aWJML]

Thorpe, S., Fize, D. & Marlot, C. (1996) Speed of processing in the human visual system. *Nature* 381:520–22. [aWJML]

Touretzky, D. S. & Hinton, G. E. (1988) A distributed connectionist production system. *Cognitive Science* 12:423–66. [rWJML]

Treisman, A. M. & Gelade, G. (1980) A feature integration theory of attention. *Cognitive Psychology* 12:97–136. [GSD]

Turvey, M. T. (1990) Coordination. *American Psychologist* 45:938–53. [aWJML]

Valdes-Sosa, M., Bobes, M. A., Rodriguez, V. & Pinilla, T. (1998) Switching attention without shifting the spotlight: Object-based attentional modulation of brain potentials. *Journal of Cognitive Neuroscience* 10:137–51. [THC]

Van Berkum, J. J. A. (1996) The psycholinguistics of grammatical gender: Studies in language comprehension and production. Unpublished dissertation, Nijmegen University. [aWJML]

(1997) Syntactic processes in speech production: The retrieval of grammatical gender. *Cognition* 64:115–52. [aWJML]

Van Gelder, T. (1990) Compositionality: A connectionist variation on a classical theme. *Cognitive Science* 14:355–84. [aWJML]

Vanier, M. & Caplan, D. (1990) CT-scan correlates of agrammatism. In: *Agrammatic aphasia. A cross-language narrative sourcebook, vol. 1,* ed. L. Menn & L. K. Obler. John Benjamins. [FP]

Van Turennout, M., Hagoort, P. & Brown, C. M. (1997) Electrophysiological evidence on the time course of semantic and phonological processes in speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 23:787–806. [aWJML]

(1998) Brain activity during speaking: From syntax to phonology in 40 msec. *Science* 280:572–74. [arWJML]

Venneman, T. (1988) *Preference laws for syllable structure and the explanation of sound change. With special reference to German, Germanic, Italian, and Latin.* Mouton de Gruyter. [aWJML]

Vigliocco, G., Antonini, T. & Garrett, M. F. (1997) Grammatical gender is on the tip of Italian tongues. *Psychological Science* 8:314–17. [TAH, aWJML]

Warrington, E. K. & McCarthy, R. A. (1987) Categories of knowledge: Further fractionations and an attempted integration. *Brain* 110:1273–96. [FP]

Weiss, S. & Rappelsberger, P. (1996) EEG coherence within the 13–18 Hz band as a correlate of a distinct lexical organization of concrete and abstract nouns in humans. *Neuroscience Letters* 209:17–20. [HMM]

Weiss, S., Müller, H. M. & Rappelsberger, P. (1998) Left frontal EEG coherence reflects modality independent language processes. *Brain Topography* 11(1). (in press). [HMM]

(in press) EEG-coherence analysis of spoken names and nouns processing. *Abstracts of the 27th Annual Meeting of the Society for Neuroscience.* [HMM]

Wheeldon, L. R. (submitted) Inhibitory form priming of spoken word production. [LRW]

Wheeldon, L. R. & Lahiri, A. (1997) Prosodic units in speech production. *Journal of Memory and Language* 37:356–81. [aWJML]

Wheeldon, L. R. & Levelt, W. J. M. (1995) Monitoring the time course of phonological encoding. *Journal of Memory and Language* 34:311–34. [aWJML]

Wheeldon, L. R. & Monsell, S. (1994) Inhibition of spoken word production by priming a semantic competitor. *Journal of Memory and Language* 33:332–56. [LF, LRW]

Wickelgren, W. A. (1969) Context-sensitive coding, associative memory, and serial order in (speech) behavior. *Psychological Review* 76:1–15. [FP]

Wickens, J. R. (1993) *A theory of the striatum.* Pergamon Press. [FP]

Wingfield, A. (1968) Effects of frequency on identification and naming of objects. *American Journal of Psychology* 81:226–34. [aWJML]

Yaniv, I., Meyer, D. E., Gordon, P. C., Huff, C. A. & Sevald, C. A. (1990) Vowel similarity, connectionist models, and syllable structure in motor programming of speech. *Journal of Memory and Language* 29:1–26. [LRW]

Zingeser, L. B. & Berndt, R. S. (1990) Retrieval of nouns and verbs in agrammatism and anomia. *Brain and Language* 39:14–32. [CS]

Zorzi, M., Houghton, G. & Butterworth, B. (1998) Two routes or one in reading aloud? A connectionist dual-process model. *Journal of Experimental Psychology: Human Perception and Performance* 24:1131–61. [MZ]

Zwitserlood, P. (1989) The effects of sentential-semantic context in spoken-word processing. *Cognition* 32:25–64. [aWJML]

(1994) Access to phonological-form representations in language comprehension and production. In: *Perspectives on sentence processing,* ed. C. Clifton, Jr., L. Frazier & K. Rayner. Erlbaum. [PAS]

FIVE IMPORTANT BBS ANNOUNCEMENTS

(1) There have been some extremely important developments in the area of Web archiving of scientific papers very recently.

Please see:

Science:
http://www.cogsci.soton.ac.uk/~harnad/science.html

Nature:
http://www.cogsci.soton.ac.uk/~harnad/nature.html

American Scientist:
http://www.cogsci.soton.ac.uk/~harnad/amlet.html

Chronicle of Higher Education:
http://www.chronicle.com/free/v45/i04/04a02901.htm

---

(2) All authors in the biobehavioral and cognitive sciences are strongly encouraged to archive all their papers (on their Home-Servers as well as) on CogPrints:

http://cogprints.soton.ac.uk/

It is extremely simple to do so and will make all of our papers available to all of us everywhere at no cost to anyone.

---

(3) BBS has a new policy of accepting submissions electronically.

Authors can specify whether they would like their submissions archived publicly during refereeing in the BBS under-refereeing Archive, or in a referees-only, non-public archive.

Upon acceptance, preprints of final drafts are moved to the public BBS Archive:

ftp://ftp.princeton.edu/pub/harnad/BBS/.WWW/index.html
http://www.cogsci.soton.ac.uk/bbs/Archive/

---

(4) BBS has expanded its annual page quota and is now appearing bimonthly, so the service of Open Peer Commentary can now be offered to more target articles. The BBS refereeing procedure is also going to be considerably faster with the new electronic submission and processing procedures. Authors are invited to submit papers to:

Email: bbs@cogsci.soton.ac.uk

Web: http://cogprints.soton.ac.uk
http://bbs.cogsci.soton.ac.uk/

INSTRUCTIONS FOR AUTHORS:
http://www.princeton.edu/~harnad/bbs/instructions.for.authors.html
http://www.cogsci.soton.ac.uk/bbs/instructions.for.authors.html

---

(5) Call for Book Nominations for BBS Multiple Book Review

In the past, Behavioral and Brain Sciences (BBS) journal had only been able to do 1–2 BBS multiple book treatments per year, because of our limited annual page quota. BBS's new expanded page quota will make it possible for us to increase the number of books we treat per year, so this is an excellent time for BBS Associates and biobehavioral/cognitive scientists in general to nominate books you would like to see accorded BBS multiple book review.

(Authors may self-nominate, but books can only be selected on the basis of multiple nominations.) It would be very helpful if you indicated in what way a BBS Multiple Book Review of the book(s) you nominate would be useful to the field (and of course a rich list of potential reviewers would be the best evidence of its potential impact!).