73

A THEORY OF LEXICAL ACCESS IN SPEECH PRODUCTION

Willem J. M. Levelt, Ardi Roelofs and Antje S. Meyer

Source: *Behavioral and Brain Sciences* 22 (1999): 1–75.

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 feed-forward. 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.

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 phonological 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).

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.

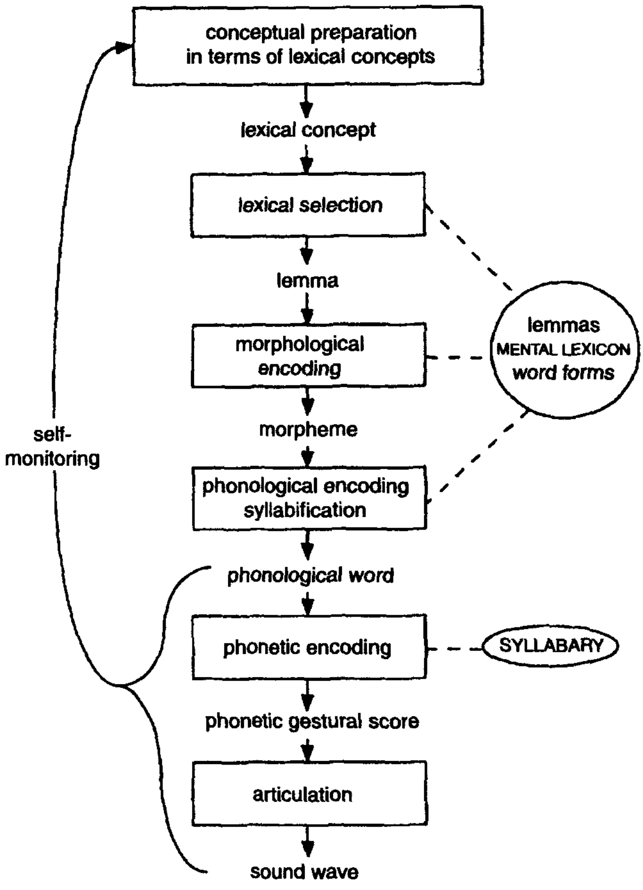


Figure 1 The theory in outline. Preparing a word proceeds through stages of conceptual preparation, lexical selection, morphological and phonological encoding, and phonetic encoding before articulation can be initiated. In parallel there occurs output monitoring involving the speaker's normal speech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

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 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

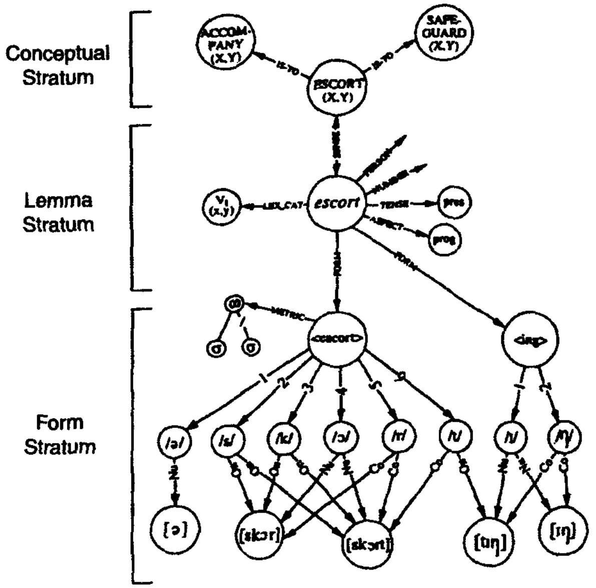


Figure 2 Fragment of the lexical network underlying lexical access. The feedforward activation spreading network has three strata. Nodes in the top, conceptual stratum represent lexical concepts. Nodes in the lemma stratum represent syntactic words or lemmas and their syntactic properties. Nodes in the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

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 lambic, 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>, σσ′, and for <ing> (or <us>), σ, will merge to form the trisyllabic template σσ′σ. 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], [skər], 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 feed-forward 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* concepts; 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 spellout 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; /skərt/, 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], [skər] and [tiŋ].

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 [piŋz] 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, [kiŋz] will be selected because the links in the network between [kiŋz] 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 [kiŋz] 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 "ese" 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 non-selected 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

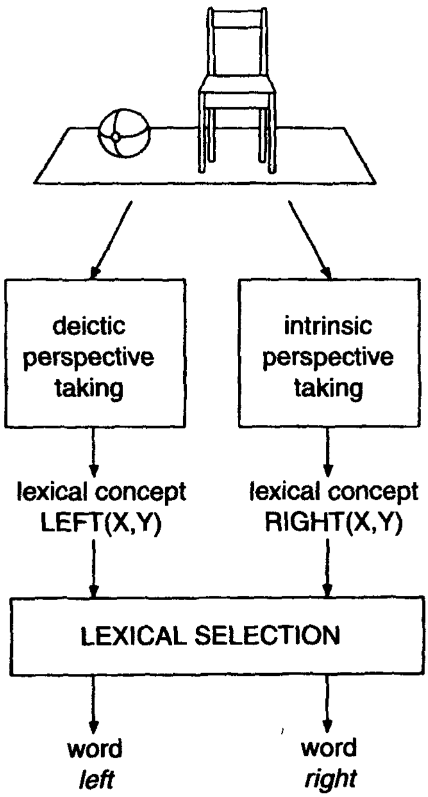


Figure 3 Illustration of perspective taking. Is the ball to the left of the chair or to the right of the chair?

taken. The expression *left of* arises when the speaker resorts to "deietic" 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 general property of all referring. It is even 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 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,

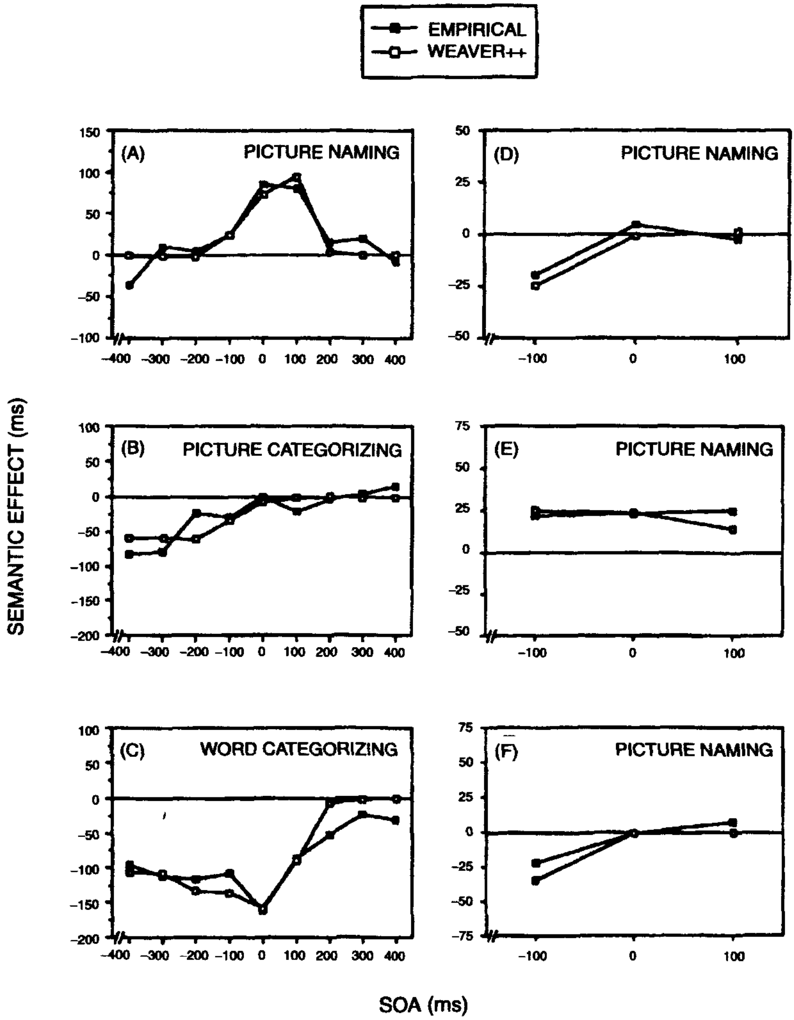


Figure 4 Effect of printed word distractors on picture naming and categorizing latencies. Degree of inhibition/facilitation as a function of stimulus onset asynchrony (SOA) between distractor and picture. A: Picture naming data; B: Picture categorizing data; C: Word categorization data from Glaser and 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).

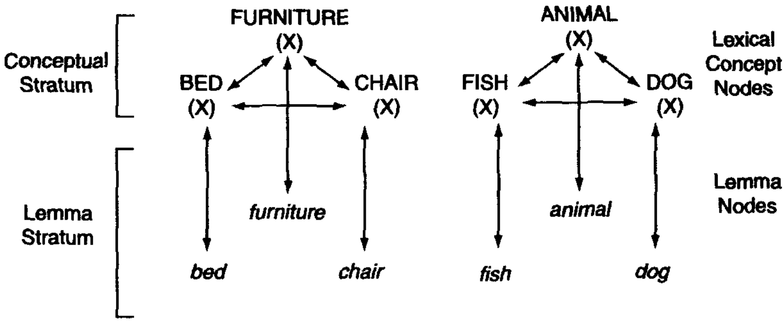


Figure 5 Miniature network illustrating how WEAVER++ accounts for the data in Figure 4.

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 conceptual 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,

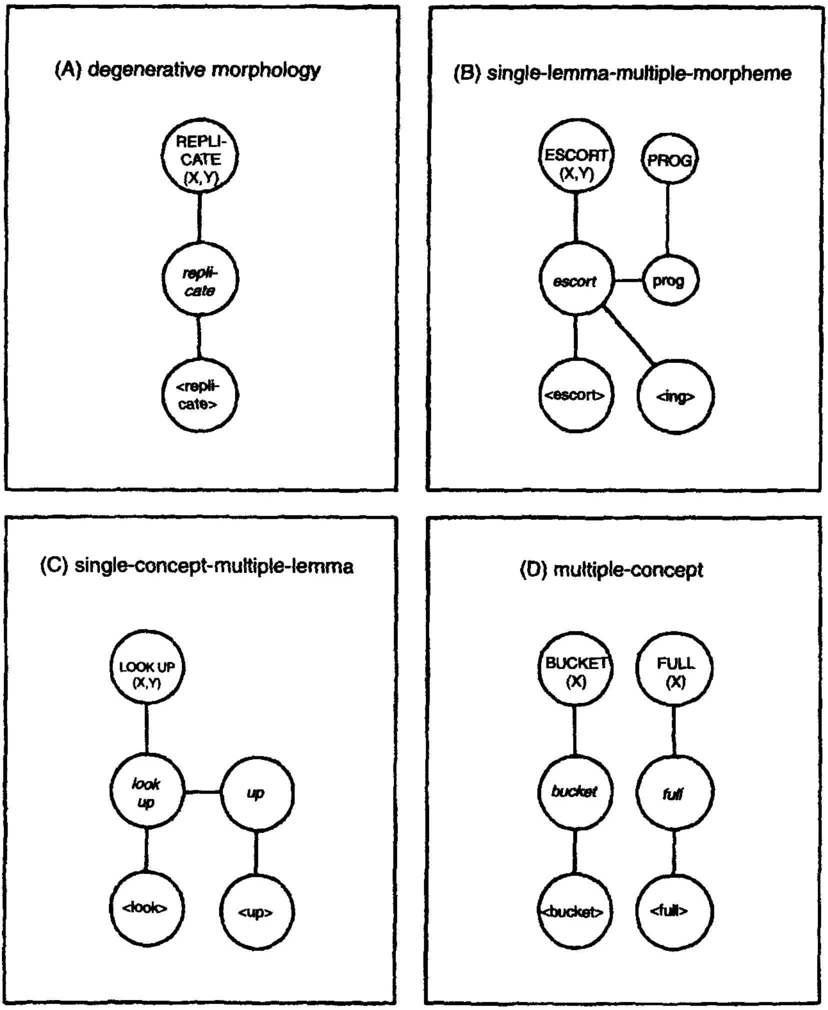


Figure 6 Four varieties of complex morphology in the theory.

re-pli-care). 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 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 Everaert 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, Lavelt, 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

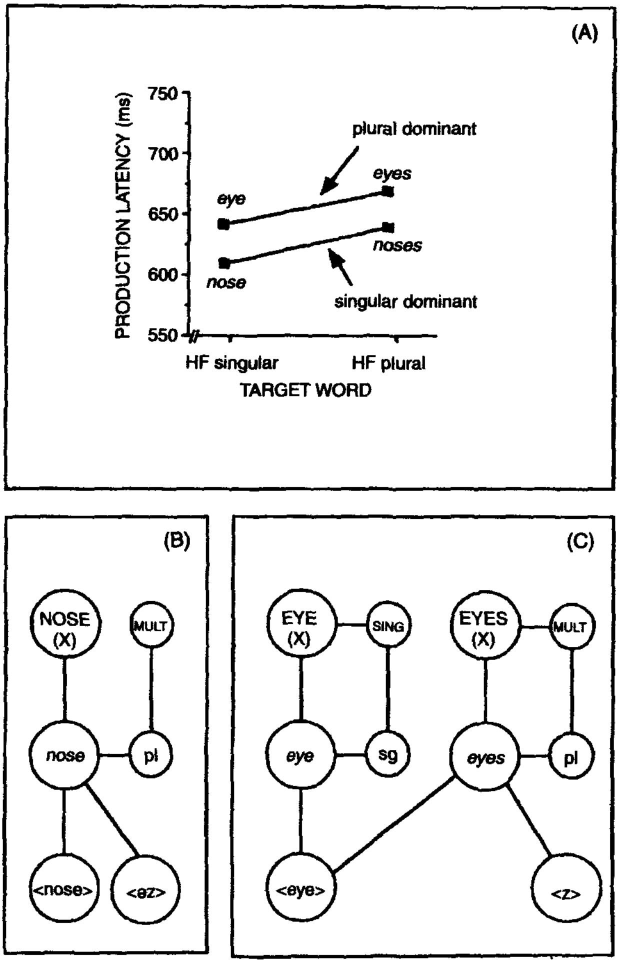


Figure 7 Naming latencies for pictures depicting one object or two identical objects (A). Plural names are slower than singular names, and both singular and plural names are slower for plural dominants (such as *eye*) than for singular dominants (such as *nose*). Possible representations of plural morphology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

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 related 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 morphene 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

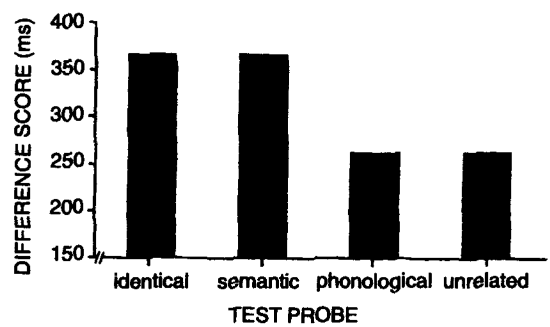


Figure 8 (Corrected) lexical decision latencies for auditory probes presented during picture naming. The y axis shows the lexical decision latency for a probe during picture naming minus the decision latency for the same auditory probe without concurrent picture naming. Probes could be identical to the 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).

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.

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 over-reaction, 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?

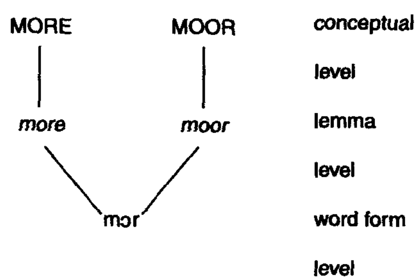
Preserving the accessing principle makes it theoretically impossible to adopt Dell's (1986; 1988) approach to the explanation of the often observed statistical overrepresentation of mixed errors (such as saying *rat* when the target is *cat*). That there is such a statistical overrepresentation is well 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, non-homophonous, 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

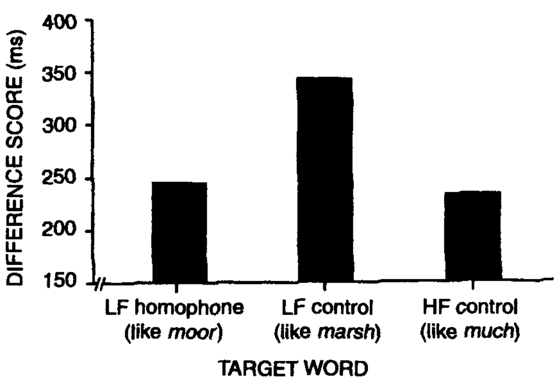


Figure 9 The homophone effect. (Corrected) naming latencies for low-frequency homophones (for example, *moor* as opposed to *more*) and nonhomophone controls that are frequency matched to the low-frequency homophone (e.g., *marsh*) or to the high-frequency twin (e.g., *much*). Data show that the low-frequency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

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.

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.

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

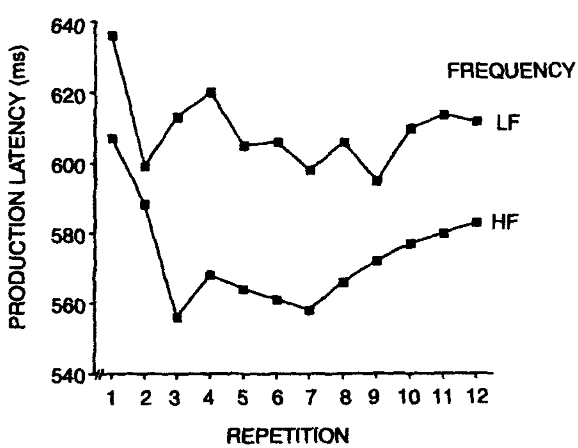
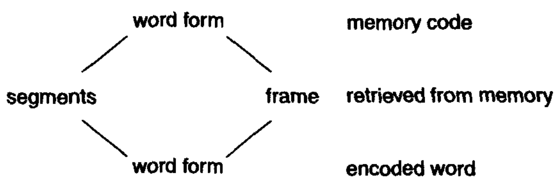


Figure 10 The robust word frequency effect. Naming latencies for 21 pictures with high-frequency names and 21 pictures with low-frequency names. The effect is stable over 12 repetitions.

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.

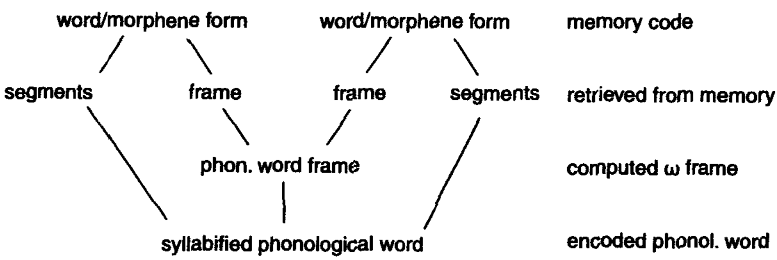


(1)

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-standing*, 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 $\sigma\sigma'$, 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 [tɪiŋ] 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 (Roclofs & 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 +10 msec). With both begin and 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 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

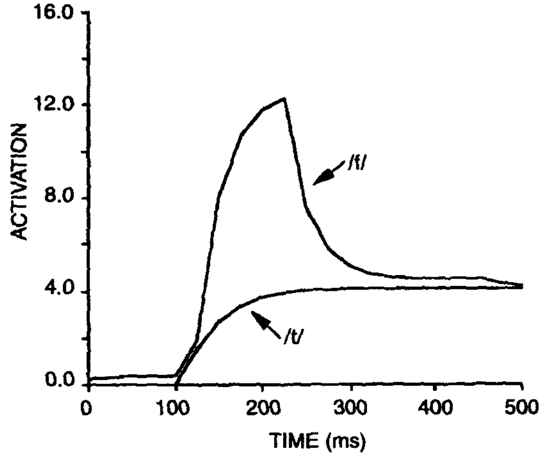


Figure 11 Activation curves for the /t/ and /f/ nodes in WEAVER++ during the encoding of *tafel*. Depicted is the aligned condition with the end-related distractor word *jofel* presented at SOA = 150 msec (after Roelofs 1997c).

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, *lower, 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 (/nɚ/), 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

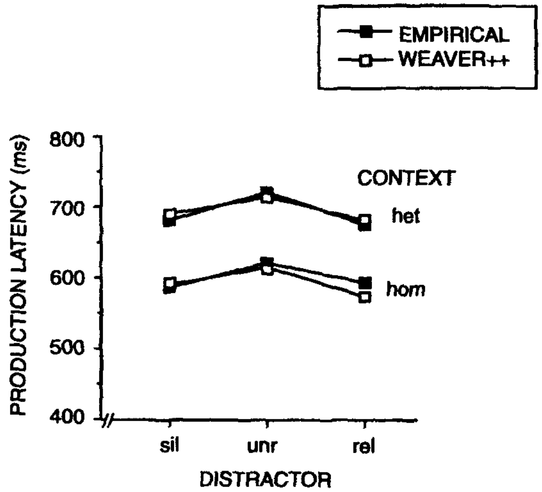


Figure 12 Combined effects of implicit and explicit priming. In the graph, "context" indicates the implicit variable "homogeneous" versus "heterogeneous" naming set; "distractor" denotes whether the auditory prime is phonologically related or unrelated to the second syllable of the target utterance.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

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 single-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 WEAVER++ about the generation of polymorphemic forms in simple phrasal constructions, namely, Dutch verb–particle combinations (Roelofs 1998). These are cases of

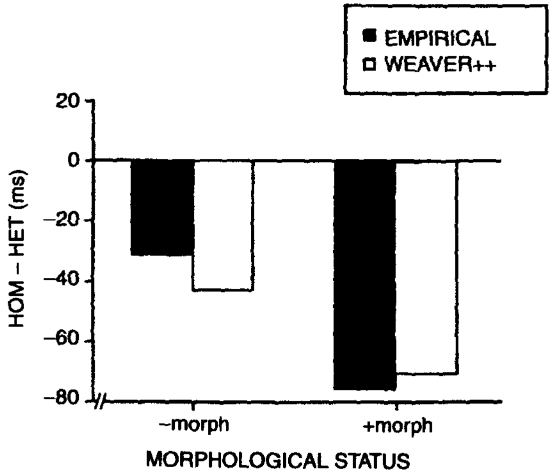


Figure 13 Implicit priming of a word's first syllable is more effective when that syllable is also a morpheme than when it is not. Experimental results and WEAVER++ simulation.

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 *op-zoeken* "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*, "decrust," "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," *makreel'*, "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

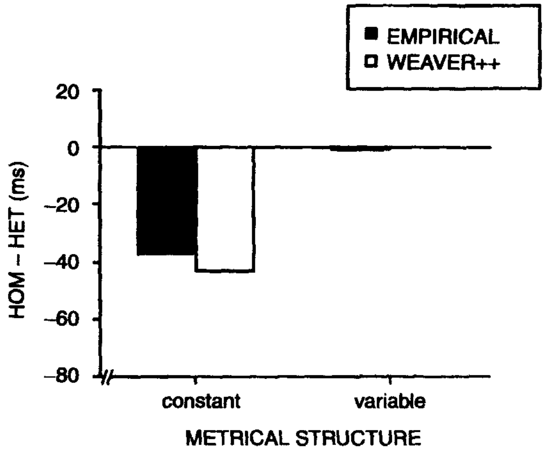


Figure 14 Implicit first-syllable priming for words with the same number of syllables versus words with a different number of syllables. Results show syllable priming in the former but not in the latter conditions. WEAVER++ predictions are also presented.

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. 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 *balade*, 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 non-masked 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, /l_{coda}/, but, in *ballade*, the /l/ will be encoded as onset segment, /l_{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 /bal/ in the input lexicon, and hence the segment /l_{coda}/. This segment, in turn, will spread its activation to *balcon*'s first syllable ([bal]) 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 (Culter 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 recording.

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 *un-der-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; Becasens 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/, /ðɪː/, /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.

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 frequen-

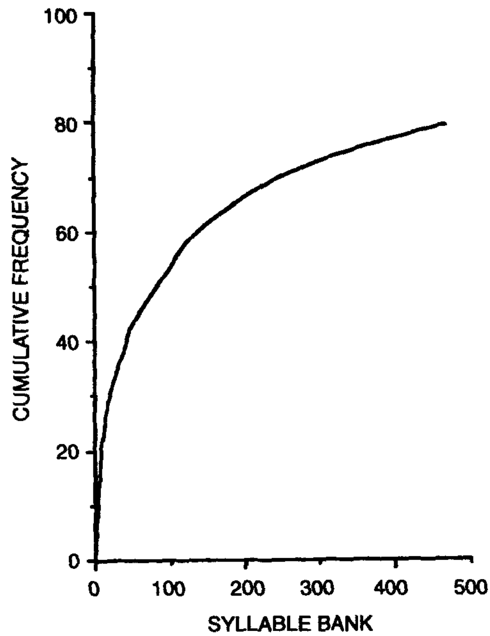


Figure 15 Cumulative frequency distribution for the 500 highest-ranked syllables in English. Derived from CELEX text-based statistics.

cies 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

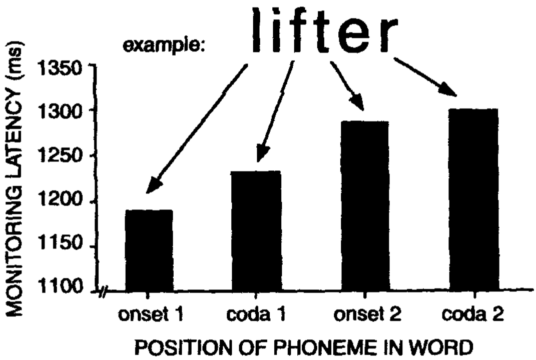


Figure 16 The self-monitoring task. Phoneme monitoring latencies for consonant targets in CVC–CVC words, such as *lifter*.

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 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

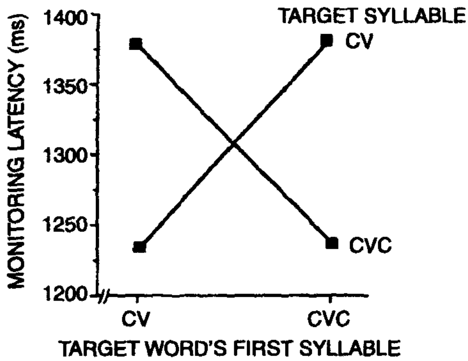


Figure 17 The syllable effect in self-monitoring. CV and CVC monitoring in target words that contain a CV or a CVC first syllable.

whether the steady increase in BT 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 errors 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* for target *cat* is more likely than *dog* for target *cat*), and lexical bias (i.e., errors tend to be real words rather than nonwords).

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 stimulation 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.

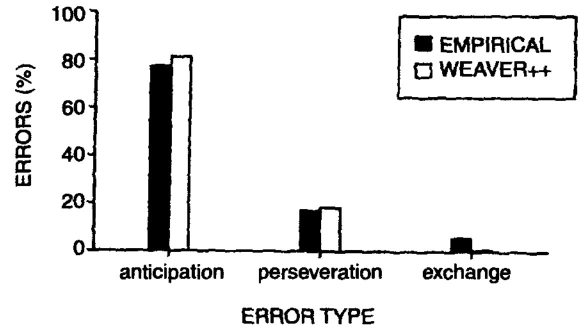


Figure 18 Frequency distribution of anticipation, perseveration, and exchange errors. Data from Nooteboom (1969) and WEAVER++ simulations.

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 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 *rat*. 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 $t + \Delta t$, whereas $a(n,t)$ 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

$$[h_1(s)V_1(s-1)] \sum_{j=0}^{s-1} [h_2(j)V_2(j-1)] +$$

$$[h_2(s)V_2(s-1)] \sum_{j=0}^{s-1} [h_1(j)V_1(j-1)] +$$

$$[h_1(s)V_1(s-1)]\,[h_2(s)V_2(s-1)]$$

$$= f_1(s) \sum_{j=0}^{s-1} f_2(s) + f_2(s) \sum_{j=0}^{s-1} f_1(s) + f_1(s)f_2(s),$$

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* and *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 underspecifi-

cation 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 data-bases 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. Theire 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.

References

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.

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

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.

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.

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

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

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.

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

—— (1989) Intersegmental cohesiveness. *Folia Linguistica* 23:245–80.

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

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

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

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

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

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

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

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

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

—— (1992) Articulatory phonology: An overview. *Phonetica* 49:155–80.

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

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

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

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

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

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

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.

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.

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

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

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.

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

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.

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

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.

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

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

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

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.

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.

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.*

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

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

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

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

Dell, G. S. (1986) A spreading-activation theory of retrieval in sentence production. *Psychological Review* 93:283–321.

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

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

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.

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.

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.

—— (1992) Stages of lexical access in language production. *Cognition* 42:287–314.

—— (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.

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.

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.

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

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

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

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

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

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.

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

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.

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

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

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.

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.

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

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

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.

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

Garcia-Albera, 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.

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

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

—— (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.

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.

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

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.

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

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

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

—— (1993) Phonological activation of semantic competitors during lexical access in speech production. *Language and Cognitive Processes* 8:291–309.

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

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

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

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

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.

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.

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

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

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

Jescheniak, J. D. (1994) *Word frequency effects in speech production*. Doctoral dissertation, Nijinegen University.

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.

—— (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.

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.

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

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

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

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.

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

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

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

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

—— (1989) *Speaking: From intention to articulation.* MIT Press.

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

—— (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.

—— (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.

Levelt, W. J. M. & Kelter, S. (1982) Surface form and memory in question answering. *Cognitive Psychology* 14:78–106.

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.

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.

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.

—— (1991b) Normal and deviant lexical processing: Reply to Dell and O'Seaghdha. *Psychological Review* 98:615–18.

Levelt, W. J. M. & Wheeldon, L. (1994) Do speakers have access to a mental syllabary? *Cognition* 50:239–69.

Liberman, A. (1996) *Speech: A special code.* MIT Press.

Lindblom, B. (1983) Economy of speech gestures. In: *The production of speech*, ed. P. F. MacNeilage. Springer.

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.

MacLeod, C. M. (1991) Half a century of research on the Stroop effect: An integrative review. *Psychological Bulletin* 109:163–203.

Marr, D. (1982) *Vision*, Freeman.

Marslen-Wilson, W. D. (1987) Functional parallelism in spoken word recognition. *Cognition* 25:71–102.

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.

Martin, A., Wiggs, C. L., Ungerleider, L. G. & Haxby, J. V. (1996) Neural correlates of category-specific knowledge. *Nature* 379:649–52.

McClelland, J. L. (1979) On the time relations of mental processes: An examination of systems of processes in cascade. *Psychological Review* 56:287–330.

McGuire, P. K., Silbersweig, D. A. & Frith, C. D. (1996) Functional neuroanatomy of verbal self-monitoring. *Brain* 119:101–11.

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.

Mehler, J., Dommergues, J. & Frauenfelder, U. (1981) The syllable's role in speech segmentation. *Journal of Verbal Learning and Verbal Behavior* 20:289–305.

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.

Meijer, P. J. A. (1994) Phonological encoding: The role of suprasegmental structures. Unpublished doctoral dissertation, Nijmegen University.

—— (1996) Suprasegmental structures in phonological encoding: The CV structure. *Journal of Memory and Language* 35:840–53.

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.

—— (1991) The time course of phonological encoding in language production: Phonological encoding inside a syllable. *Journal of Memory and Language* 30:69–89.

—— (1992) Investigation of phonological encoding through speech error analyses: Achievements, limitations, and alternatives. *Cognition* 42:181–211.

—— (1994) Timing in sentence production. *Journal of Memory and Language* 33:471–92.

—— (1996) Lexical access in phrase and sentence production: Results from picture-word interference experiments. *Journal of Memory and Language* 35:477–96.

—— (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.

Meyer, A. S. Roelofs, R. & Schiller, N. (in preparation) Prosodification of metrically regular and irregular words.

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.

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.

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.

—— (1991) *The science of words*. Scientific American Library.

Miller, G. A. & Johnson-Laird, P. N. (1976) *Language and perception*. Harvard University Press.

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.

Morton, J. (1969) The interaction of information in word recognition. *Psychological Review* 76:165–78.

Mowrey, R. A. & MacKay, I. R. A. (1990) Phonological primitives: Electromygraphic speech error evidence. *Journal of the Acoustical Society of America* 88:1299–312.

Munhall, K. G. & Löfqvist, A. (1992) Gestural aggregation in speech: Laryngeal gestures. *Journal of Phonetics* 20:111–26.

Nespor, M. & Vogel, I. (1986) *Prosodic phonology*. Foris.

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.

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.

—— (1994) Shortlist: A connectionist model of continuous speech recognition. *Cognition* 52:189–234.

Oldfield, R. C. & Wingfield, A. (1965) Response latencies in naming objects. *Quarterly Journal of Experimental Psychology* 17:273–81.

—— (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*.

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.

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.

Recasens, D. (1984) V-to-C coarticulation in Catalan VCV sequences: An articulatory and acoustic study. *Journal of Phonetics* 12:61–73.

—— (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.

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.

Roelofs, A. (1992a) A spreading-activation theory of lemma retrieval in speaking: *Cognition* 42:107–42.

—— (1992b) *Lemma retrieval in speaking: A theory, computer simulations, and empirical data.* Doctoral dissertation, NICI Technical Report 92–08, University of Nijmegen.

—— (1993) Testing a non-decompositional theory of lemma retrieval in speaking: Retrieval of verbs. *Cognition* 47:59–87.

—— (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.

—— (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.

—— (1996b) Serial order in planning the production of successive morphemes of a word. *Journal of Memory and Language*. 35:854–76.

—— (1996c) Morpheme frequency in speech production: Testing WEAVER. In: *Yearbook of morphology*, ed. G. E. Booij & J. van Marle. Kluwer Academic Press.

—— (1997a) A case for non-decomposition in conceptually driven word retrieval. *Journal of Psycholinguistic Research* 26:33–67.

—— (1997b) Syllabification in speech production: Evaluation of WEAVER. *Language and Cognitive Processes* 12:657–94.

—— (1997c) The WEAVER model of word-form encoding in speech production. *Cognition* 64:249–84.

—— (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.

—— (submitted a) Parallel activation and serial selection in planning linguistic performance: Evidence from implicit and explicit priming.

—— (submitted b) WEAVER++ and other computational models of lemma retrieval and word-form encoding.

—— (submitted c) Discrete or continuous access of lemmas and forms in speech production?

Roelofs, A. Baayen, H. & Van den Brink, D. (submitted) Semantic transparency in producing polymorphemic words.

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.

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.

—— (1998) A case for the lemma-lexeme distinction in models of speaking: Comment on Caramazza and Miozzo (1997). *Cognition* 69:219–30.

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.

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.

Rossi, M. & Defare, E. P. (1995) Lapsis linguae: Word errors or phonological errors? *International Journal of Psycholinguistics* 11:5–38.

Schade, U. (1996) *Konnektionistische Sprachproduktion [Connectionist language production]*. Habilitationsschrift, Bielefeld University.

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.

—— (1998) the effect of visually masked syllable primes on the naming latencies of words and pictures. *Journal of Memory and Language* 39:484–507.

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.

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.

Schmitt, B. (1997) Lexical access in the production of ellipsis and pronouns. Unpublished doctoral dissertation, Nijmegen University.

Schriefers, H. (1990) Lexical and conceptual factors in the naming of relations. *Cognitive Psychology* 22:111–42.

—— (1993) Syntactic processes in the production of noun phrases. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 19:841–50.

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.

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.

Seidenberg, M. S. & McClelland, J. L., (1989) A distributed, developmental model of word recognition and naming. *Psychological Review* 36:528–68.

Semenza, C., Mondini, S. & Cappelletti, M. (1997) The grammatical properties of mass nouns: An aphasia case study. *Neuropsychologica* 35:669–75.

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.

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.

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.

—— (1983) Sublexical units and suprasegmental structure in speech production planning. In: *The production of speech*, ed. P. F. MacNeilage. Springer.

—— (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.

—— (1987) The role of word-onset consonants in speech production planning: New evidence from speech error pattens. In: *Memory and sensory processes of language*, ed. E. Keller & M. Gopnik, Erlbaum.

—— (1992) The role of word structure in segmental serial ordering. *Cognition* 42:213–59.

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.

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.

—— (1996) From "thought and language" to "thinking for speaking." In: *Rethinking linguistic relativity*, ed., J. J. Gumperz & S. C. Levinson. Cambridge University Press.

—— (1998) Verbalized events: A dynamic approach to linguistic relativity and determinism. *Proceedings of the LAUD-Symposium*, Duisburg.

Snodgrass, J. G. & Yuditsky, T. (1996) Naming times for the Snodgrass and Vanderwart pictures. *Behavioral Research Methods, Instruments and Computers* 28:516–36.

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.

—— (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.

—— (1996b) Time-course analysis of semantic and orthographic context effects in picture naming. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 22:896–918.

Stemberger, J. P. (1983) *Speech errors and theoretical phonology: A review*. Indiana University Linguistics Club.

—— (1985) An interactive activation model of language production. In: *Progress in the psychology of language, vol. 1*, ed. A. W. Ellis. Erlbaum.

—— (1990) Wordshape errors in language production. *Cognition* 35:123–57.

—— (1991a) Radical underspecification in language production. *Phonology* 8:73–112.

—— (1991b) Apparent anti-frequency effects in language production: The addition bias and phonological underspecification. *Journal of Memory and Language* 30:161–85.

Stemberger, J. P. & MacWhitney, B. (1986) Form-oriented inflectional errors in language processing. *Cognitive Psychology* 18:329–54.

Stemberger, J. P. & Stoel-Gammon, C. (1991) The underspecification of coronals: Evidence from language acquisition and performance errors. *Phonetics and Phonology* 2:181–99.

Sternberg, S. (1969) The discovery of processing stages: Extensions of Donders' method. *Atterntion and Performance II*, ed. W. G. Koster. *Acta Psychologica* 30:276–315.

Thorpe, S., Fize, D. & Marlot, C. (1996) Speed of processing in the human visual system. *Nature* 381:520–22.

Touretzky, D. S. & Hinton, G. E. (1988) A distributed connectionist production system. *Cognitive Science* 12:423–66.

Turvey, M. T. (1990) Coordination. *American Psychologist* 45:938–53.

Van Berkum, J. J. A. (1996) The psycholinguistics of grammatical gender: Studies in language comprehension and production. Unpublished dissertation, Nijmergen University.

—— (1997) Syntactic processes in speech production. The retrieval of grammatical gender. *Cognition* 64:115–52.

Van Gelder, T. (1990) Compositionality: A connectionist variation on a classical theme. *Cognitive Science* 14:355–84.

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.

—— (1998) Brain activity during speaking: From syntax to phonology in 40 msec. *Science* 280:572–74.

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.

Vigliocco, G., Antonini, T. & Garrett, M. F. (1997) Grammatical gender is on the tip of Italian tongues. *Psychological Science* 8:314–17.

Wheeldon, L. R. & Lahiri, A. (1997) Prosodic units in speech production. *Journal of Memory and Language* 37:356–81.

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

Wingfield, A. (1968) Effects of frequency on identification and naming of objects. *American Journal of Psychology* 81:226–34.

Zwitserlood, P. (1989) The effects of sentential-semantic context in spoken-word processing. *Cognition* 32:25–64.