3 Spoken word planning, comprehending, and self-monitoring: Evaluation of WEAVER++

Ardi Roelofs

Abstract

During conversation, speakers not only talk but they also monitor their speech for errors and they listen to their interlocutors. Although the interplay among speaking, self-monitoring, and listening stands at the heart of spoken conversation, it has not received much attention in models of language use. This chapter describes chronometric, error, and aphasic evidence on spoken word planning and its relations with self-monitoring and comprehending, and it uses the evidence to evaluate WEAVER++, which is a computational model of spoken word production that makes the relations explicit. The theoretical claims implemented in WEAVER++ are contrasted with other theoretical proposals.

Speaker as listener

Speakers not only talk but they also listen to their interlocutors' speech and they monitor their own speech for errors. This chapter describes empirical evidence on spoken word planning and its relationships with comprehending and self-monitoring. It uses the evidence to evaluate WEAVER++ (Levelt, Roelofs, & Meyer, 1999a; Roelofs, 1992, 1997a, 2003a), which is a computational model of spoken word production that makes the interplay among planning, comprehending, and monitoring explicit. The claims implemented in WEAVER++ are contrasted with other theoretical approaches. It is argued that the interplay among speaking, comprehending, and self-monitoring is not only of interest in its own right, but that it also illuminates classic issues in spoken word production.

This chapter consists of two parts. The first part reviews relevant empirical evidence and it explains what claims about spoken word planning, comprehending, and self-monitoring are implemented in WEAVER++. This model assumes that word planning is a staged process that traverses from conceptual preparation via lemma retrieval to word-form encoding. Comprehending spoken words traverses from forms to lemmas and meanings. Concepts and lemmas are shared between production and comprehension, whereas there

are separate input and output representations of word forms. After lemma retrieval, word planning is a strictly feedforward process. Following Levelt (1989; see also Hartsuiker, Kolk, & Lickley, this volume; Chapter 14), WEAVER++ assumes two self-monitoring routes, an internal and an external one, both operating via the speech comprehension system. Brain imaging studies also suggest that self-monitoring and speech comprehension are served by the same neural structures (e.g., McGuire, Silbersweig, & Frith, 1996; Paus, Perry, Zatorre, Worsley, & Evans, 1996). The external route involves listening to self-produced speech, whereas the internal route involves evaluating the speech plan. Self-monitoring requires cognitive operations in addition to speech comprehension. For example, lexical selection errors may be detected by verifying whether the lemma recognized in inner speech corresponds to the lexical concept prepared for production, which is an operation specific to self-monitoring. The self-monitoring through speech comprehension assumed by WEAVER++ is shown to be supported by a new analysis performed on the self-corrections and false starts in picture naming by 15 aphasic speakers reported by Nickels and Howard (1995).

The second part of the chapter applies WEAVER++ to findings that were seen as problematic for feedforward models (e.g., Damian & Martin, 1999; Rapp & Goldrick, 2000): The statistical overrepresentation of mixed semantic-phonological speech errors and the reduced latency effect of mixed distractors in picture naming. The mixed error bias is the finding that mixed semantic-phonological errors (e.g., the erroneous selection of *calf* for *cat*, which share the onset segment and, in American English, the vowel) are statistically overrepresented, both in natural speech-error corpora and in picture naming experiments with aphasic as well as nonaphasic speakers (Dell & Reich, 1981; Martin, Gagnon, Schwartz, Dell, & Saffran, 1996). The bias is also called the "phonological facilitation of semantic substitutions." Rapp and Goldrick (2000) observed that the presence of a mixed error bias depends on the impairment locus in aphasia. The bias occurs with a post-conceptual deficit (as observed with patients P.W. and R.G.B., who make semantic errors in word production only) but not with a conceptual deficit (as observed with patient K.E., who makes semantic errors in both word production and comprehension). The mixed-distractor latency effect is the finding that mixed semantic-phonological distractor words in picture naming (e.g., the spoken word CALF presented as a distractor in naming a pictured cat; hereafter, perceived words are referred to in uppercase) yield less interference than distractors that are semantically related only (distractor DOG), taking the facilitation from phonological relatedness per se (distractor CAP) into account (Damian & Martin, 1999; Starreveld & La Heij, 1996).

Elsewhere, it has been discussed how WEAVER++ deals with other speech error tendencies such as the bias towards word rather than non-word error outcomes (Roelofs, 2004a, 2004b), and with aphasic phenomena such as modality-specific grammatical class deficits and the finding that semantic errors may occur in speaking but not in writing, or vice versa (Roelofs, Meyer,

& Levelt, 1998). Nickels and Howard (2000) provide a general evaluation of the model in the light of a wide range of findings on aphasia. Here, I focus on the mixed-error bias, its dependence on the locus of damage in aphasia, and the mixed-distractor latency effect. I argue that these findings are not problematic for WEAVER++ but, on the contrary, support the claims about the relations among speaking, comprehending, and monitoring in the model. According to WEAVER++, mixed items (e.g., *calf* in naming a cat) are weaker lexical competitors than items that are semantically related only (e.g., *dog*), because they co-activate the target (*cat*) as a member of their speech comprehension cohort. Therefore, compared with items that are semantically related only, mixed items are more likely to remain unnoticed in error monitoring (yielding more mixed-speech errors) and, as distractors, they have a smaller effect on latencies (yielding less semantic interference). The simulations reported by Roelofs (2004a, 2004b) and reviewed here showed that WEAVER++ not only accounts for the mixed-distractor latency effect, but also for the mixed-error bias and the influence of the impairment locus in aphasia. The assignment of the mixed-distractor latency effect to properties of the speech comprehension system by the model is shown to be supported by recent chronometric studies, which revealed that there are semantic effects of word-initial cohort distractors in picture naming (Roelofs, submitted-a) and that there is no reduced latency effect for mixed rhyme distractors (Roelofs, submitted-b).

An outline of the WEAVER++ model

WEAVER++ distinguishes between conceptual preparation, lemma retrieval, and word-form encoding, with the encoding of forms further divided into morphological, phonological, and phonetic encoding (Roelofs, 1992, 1997a). During conceptual preparation, a lexical concept is selected and flagged as goal concept (e.g., the concept of a cat in naming a pictured cat). In lemma retrieval, a selected concept is used to activate and select a lemma from memory, which is a representation of the syntactic properties of a word, crucial for its use in sentences. For example, the lemma of the word *cat* says that it is a noun. Lemma retrieval makes these properties available for syntactic encoding. In word-form encoding, the selected lemma is used to activate and select form properties from memory. For example, for *cat*, the morpheme <cat> and the segments /k/, /æ/ and /t/ are activated and selected. Next, the segments are rightward incrementally syllabified, which yields a phonological word representation. Finally, a motor program for [kæt] is recovered. Articulation processes execute the motor program, which yields overt speech. Figure 3.1 illustrates the stages. Lemma retrieval and word-form encoding are discrete processes in that only the word form of a selected lemma becomes activated.

After lemma retrieval, word planning happens in a strictly feedforward fashion, with feedback only occurring via comprehension (Roelofs, 2004b).

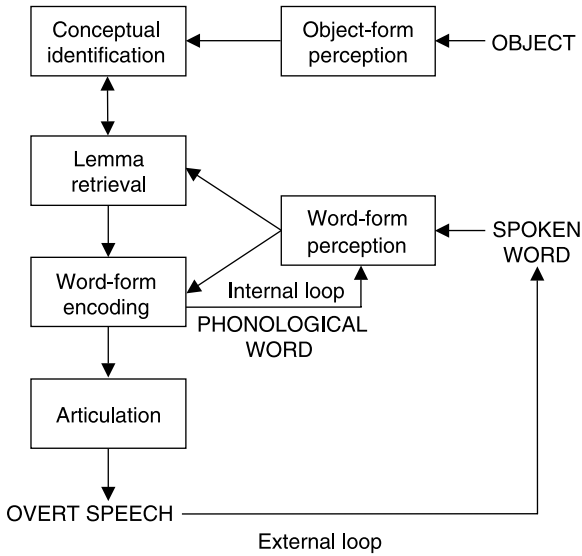

Figure 3.1 Flow of information in the WEAVER++ model during the planning, comprehending, and self-monitoring of spoken words.

Comprehending spoken words proceeds from word-form perception to lemma retrieval and conceptual identification. A perceived word activates not only its lemma, but also in parallel, its output form. Self-monitoring is achieved through the speech comprehension system. There exist internal and external self-monitoring routes, as illustrated in Figure 3.1. The external route involves listening to self-produced overt speech, whereas the internal route includes monitoring the speech plan by feeding a planned phonological word representation, specifying the syllables and stress pattern, back into the speech comprehension system.

Word planning involves retrieval of information from a lexical network. There are three network strata, shown in Figure 3.2. A conceptual stratum represents the concepts of words as nodes and links in a semantic network. A syntactic stratum contains lemma nodes for words, such as *cat*, which are connected to nodes for their syntactic class (e.g., *cat* is a noun, N). A word-form stratum represents the morphemes, segments, and syllable programs of words. The form of monosyllables such as *cat* presents the simplest case with one morpheme <cat>, segments such as /k/, /æ/, and /t/, and one syllable program [kæt]. Polysyllables such as *feline* have their segments connected to more than one syllable program; for *feline*, these program nodes are [fi:] and [laIn]. Polymorphemic words such as *catwalk* have one lemma connected to more than one morpheme; for *catwalk* these morphemes are <cat> and <walk>. For a motivation of these assumptions, I refer to Levelt (1989), Levelt et al. (1999a, 1999b), Roelofs (1992, 1993, 1996, 1997a, 1997b, 1997c,

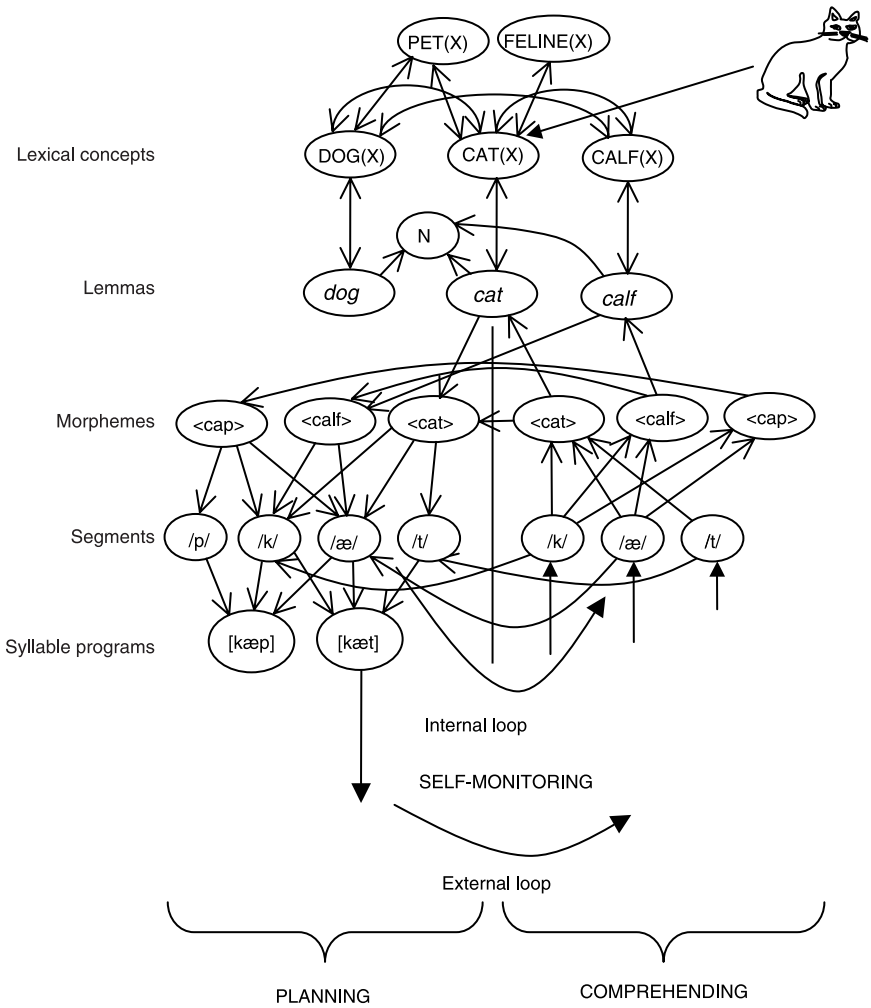

Figure 3.2 Fragment of the production and comprehension networks of the WEAVER++ model.

1998, 1999, 2003a), Roelofs and Meyer (1998), and Roelofs, Meyer, and Levelt (1996, 1998).

Information needed for word production is retrieved from the network by spreading activation. For example, a perceived object (e.g., a cat) activates the corresponding concept node (i.e., CAT(X); hereafter, propositional functions denote lexical concept nodes). Activation then spreads through the network following a linear activation rule with a decay factor. Each node sends a proportion of its activation to the nodes it is connected to. For example, CAT(X) sends activation to other concepts such as DOG(X) and to its lemma

node *cat*. Selection of nodes is accomplished by production rules. A production rule specifies a condition to be satisfied and an action to be taken when the condition is met. A lemma retrieval production rule selects a lemma after it has been verified that the connected concept is flagged as goal concept. For example, *cat* is selected for CAT(X) if it is the goal concept and *cat* has reached a critical difference in activation compared to other lemmas. The actual moment of firing of the production rule is determined by the ratio of activation of the lemma node and the sum of the activations of all the other lemma nodes. Thus, how fast a lemma node is selected depends on how active the other lemma nodes are.

A selected lemma is flagged as goal lemma. A morphological production rule selects the morpheme nodes that are connected to the selected lemma (<cat> is selected for *cat*). Phonological production rules select the segments that are connected to the selected morphemes (/k/, /æ/, and /t/ for <cat>) and incrementally syllabify the segments (e.g., /k/ is made syllable onset: onset(/k/)) to create a phonological word representation. Phonological words specify the syllable structure and, for polysyllabic words, the stress pattern across syllables. Finally, phonetic production rules select syllable-based motor programs that are appropriately connected to the syllabified segments (i.e., [kæt] is selected for onset(/k/), nucleus(/æ/) and coda(/t/)). The moment of selection of a program node is given by the ratio of activation of the target syllable-program node and the sum of the activations of all the other syllable-program nodes (thus, the selection ratio applies to lemmas and syllable programs).

To account for interference and facilitation effects from auditorily presented distractor words on picture naming latencies, Roelofs (1992, 1997a) assumed that information activated in a speech comprehension network activates compatible segment, morpheme, and lemma representations in the production network (see Figure 3.2). Covert self-monitoring includes feeding the incrementally constructed phonological word representation from the production into the comprehension system. An externally or internally perceived word activates a cohort of word candidates, including their forms, lemmas, and meanings.

Evidence for comprehension cohorts

The assumption implemented in WEAVER++ that a cohort of word candidates is activated during spoken word recognition is widely accepted in the comprehension literature. Cohort models of spoken word recognition such as the seminal model of Marslen-Wilson and Welsh (1978) claim that, on the basis of the first 150 milliseconds or so of the speech stream, all words that are compatible with this spoken fragment are activated in parallel in the mental lexicon. The activation concerns not only the forms but also the syntactic properties and concepts of the words. For example, when an American English listener hears the spoken word fragment CA, a cohort of words

including *cat, calf, captain* and *captive* becomes activated. Other models of spoken word recognition such as TRACE (McClelland & Elman, 1986) and Shortlist (Norris, 1994) make similar claims.

Evidence for the multiple activation of lexico-semantic representations of words during word recognition comes from cross-modal semantic priming experiments. For example, Zwitserlood (1989) asked participants to listen to spoken words (e.g., CAPTAIN) or fragments of these words (e.g., CAPT). The participants had to take lexical decisions to written probes that were presented at the offset of the spoken primes. The spoken fragments facilitated the lexical decision to target words that were semantically related to the complete word as well as to cohort competitors. For example, CAPT facilitated the response to SHIP (semantically related to *captain*) and also to GUARD (semantically related to *captive*).

Multiple activation appears to involve mainly cohort competitors. Several studies (e.g., Connine, Blasko, & Titone, 1993; Marslen-Wilson & Zwitserlood, 1989) have shown that when the first segments of a spoken non-word prime and the source word from which it is derived differ in more than two phonological features (such as place and manner of articulation, e.g., the prime ZANNER derived from MANNER), no priming is observed on the lexical decision to a visually presented probe (e.g., STYLE). Marslen-Wilson, Moss, and Van Halen (1996) observed that a difference of one phonological feature between the first segment of a word prime and its source word leads to no cross-modal semantic priming effect. In an eye-tracking study, Allopenna, Magnuson, and Tanenhaus (1998) observed that, for example, hearing the word COLLAR (a rhyme competitor of *dollar*) had much less effect than hearing DOLPHIN (a cohort competitor of *dollar*) on the probability of fixating a visually presented target dollar (a real object). Thus, the evidence suggests that in spoken word recognition there is activation of cohort competitors, whereas there is much less activation of rhyme competitors, even when they differ in only the initial segment from the actually presented spoken word or non-word.

Evidence for phonological words in inner speech

The assumption implemented in WEAVER++ that phonological words are monitored rather than, for example, articulatory programs (e.g., [kæt]) or strings of segments (e.g., /k/, /æ/, and /t/) was motivated by a study conducted by Wheeldon and Levelt (1995). The participants were native speakers of Dutch who spoke English fluently. They had to monitor for target speech segments in the Dutch translation equivalent of visually presented English words. For example, they had to indicate by means of a button press (yes/no) whether the segment /n/ is part of the Dutch translation equivalent of the English word WAITER. The Dutch word is *kelner*, which has /n/ as the onset of the second syllable, so requiring a positive response. All Dutch target words were disyllabic. The serial position of the critical segments in the Dutch words was manipulated. The segment could be the onset or coda of the first

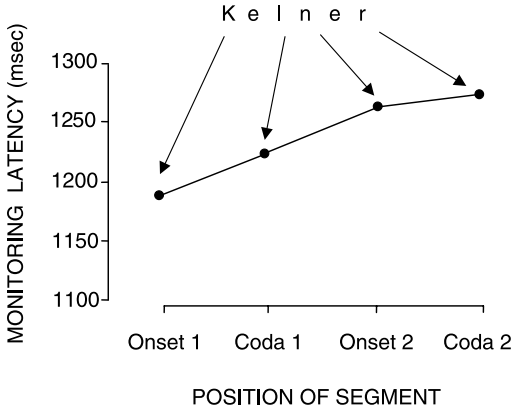

Figure 3.3 Mean monitoring latencies (in msec) as a function of the position of the target segment observed by Wheeldon and Levelt (1995).

syllable, or the onset or coda of the second syllable. If a rightward incrementally generated phonological word representation is consulted in performing self-monitoring, an effect of serial position is to be expected. Such an effect was indeed empirically obtained. Monitoring latencies increased with the serial position of the segments within the word, as illustrated in Figure 3.3.

In order to experimentally verify whether phonological words rather than phonetic motor programs are monitored, participants had to perform the segment-monitoring task while simultaneously counting aloud, which is known to suppress the maintaining of phonetic representations. The monitoring latencies were longer with the counting task, but the seriality effect was replicated. This suggests that monitoring involves a phonological rather than a phonetic representation. Finally, to assess whether monitoring involves a syllabified representation rather than a string of segments, participants had to monitor for target syllables. The target syllable corresponded to the first syllable of the Dutch word or it was larger or smaller. For example, the target syllable was CA or CAP and the first syllable of the Dutch word was CA or CAP. If phonological words are monitored, which are syllabified, then a syllable match effect should be obtained, whereas if a string of segments is monitored, the syllabic status of the segments should not matter. The experiment yielded a clear syllable match effect. Syllable targets were detected much faster when they exactly matched the first syllable of the words than when they were larger or smaller. This suggests that phonological words rather than strings of segments are monitored.

Evidence from self-corrections and false starts

Independent support for the involvement of planned phonological words in self-monitoring comes from a new analysis that I performed on the patient

data reported by Nickels and Howard (1995). The data consist of the responses in a picture-naming task from 15 aphasic individuals. Each aphasic speaker named 130 black and white drawings. As measures of self-monitoring, Nickels and Howard counted the attempted self-corrections and the false starts made by the patients in naming the pictures. False starts included all those responses where an initial portion of the target was correctly produced but with production being stopped before the word was completed (e.g., saying 'ka' when the target was *captain*). The presence of false starts was taken to specifically reflect internal feedback and monitoring. Nickels and Howard computed the correlations between the proportion of trials that included phonological errors, self-corrections and false starts, on the one hand, and the performance on three tasks involving speech perception (auditory synonym judgement, auditory lexical decision, and auditory minimal-pair discrimination), on the other hand, and found no significant correlations. They argued that this challenges the idea of self-monitoring through speech comprehension. However, the absence of significant correlations may also mean that the three speech perception tasks are not good indicators of the patients' self-monitoring abilities (see Nickels, 1997, for discussion). Most aphasic individuals performed close to ceiling on the three auditory input tasks. Furthermore, auditory synonym judgement and minimal-pair discrimination ask for input buffering of two perceived words, something that is not crucial for the type of monitoring proposed.

According to the proposed monitoring view, the capacity to feed back and evaluate planned phonological word representations via the comprehension system should be critical to self-monitoring. Interestingly, Nickels and Howard (1995) report the speakers' scores on a task that seems to tap into this capacity, namely homophone judgement from pictures, but they did not include this measure in their analyses. The homophone judgement task involves the selection of two homophone items from a triad of pictures (e.g., hair, hare, and steak). Performing this task minimally involves silently generating the sound form of the name of one of the pictures and then evaluating which of the other two pictures this sound form also correctly names. Thus, this task contains all the processing components that are presumed to be involved in the internal monitoring of a speech plan. To test whether this capacity is indeed involved in the speakers' self-monitoring, I computed the correlations between self-corrections and false starts and the performance on the homophone task.

I first confirmed that the total number of semantic errors made by the aphasic speakers in the picture-naming task was negatively correlated with their ability to perform auditory synonym judgements. On the synonym task, the speakers are presented with pairs of spoken words and they are required to judge whether the words are approximately synonymous. The correlations were indeed highly significant (high imageability words: $r = -.63$, $p = .01$; low imageability words: $r = -.91$, $p = .001$). Interestingly, the total number of semantic errors made by each speaker was also negatively correlated with

their performance on the homophone task ($r = -.77$, $p = .001$). The higher the score on this task, the lower the number of semantic errors. This suggests that the ability to evaluate a phonological representation is a precondition for (lexical/semantic) error detection. Moreover, there were positive correlations between performance on the homophone task and the proportion of phonological self-corrections and false starts. The correlation was higher for the false starts ($r = .64$, $p = .01$) than for the self-corrections ($r = .47$, $p = .08$), suggesting that the homophone task captures a capacity that is more heavily engaged in internal than in external monitoring. Thus, the capacity to silently generate word forms and to evaluate them with respect to their meaning is positively correlated with the number of false starts and, to a lesser extent, the number of self-corrections of the patients. This supports the idea that self-monitoring of speech may be performed by feeding back phonological word representations to the comprehension system and evaluating the corresponding meaning.

Accounting for mixed-error bias

When *cat* is intended, the substitution of *calf* for *cat* is more likely than the substitution of *dog* for *cat* (Dell & Reich, 1981; Martin et al., 1996), taking error opportunities into account. On the standard feedback account, the mixed-error bias arises because of production-internal feedback from segment nodes to lexical nodes within a lexical network. Semantic substitution errors are taken to be failures in lexical node selection. The word *calf* shares phonological segments with the target *cat*. So, the lexical node of *calf* receives feedback from these shared segments (i.e., /k/ and /æ/), whereas the lexical node of *dog* does not. Consequently, the lexical node of *calf* has a higher level of activation than the lexical node of *dog*, and *calf* is more likely involved in a selection error than *dog*.

The mixed-error bias does not uniquely support production-internal feedback, however (Rapp & Goldrick, 2000; 2004). Rapp and Goldrick (2000) demonstrated by means of computer simulation that the error bias may occur at the segment rather than the lexical level in a feedforward cascading network model. So, production-internal feedback is not critical. Likewise, Levelt et al. (1999a) argued that the mixed-error effect occurs in WEAVER++ when the lemma retrieval stage mistakenly selects two lemmas rather than a single one. In a cascading model, activation automatically spreads from one level to the other, whereas in a discrete multiple-output model the activation is restricted to the selected items. Both views predict a mixed error bias. The bias occurs during word planning in WEAVER++, because the sound form of a target like *cat* speeds up the encoding of the form of an intruder like *calf* but not of an intruder like *dog*. Therefore, the form of *calf* is completed faster than the form of *dog*, and *calf* has a higher probability than *dog* of being produced instead of *cat*. The assumption of multiple output underlying certain speech errors is independently supported by word blends, like a speaker's

integration of the near-synonyms *close* and *near* into the error "clear". Dell and Reich (1981) observed the mixed-error bias also for blends.

Levelt et al. (1999a) argued that a mixed-error bias is also inherent to self-monitoring. Monitoring requires attention and it is error prone. It has been estimated that speakers miss about 50 percent of the errors they make (Levelt, 1989). The more the error differs from the target, the better it should be noticeable. In planning to say "cat" and monitoring through comprehension, the lemma of the target *cat* is in the comprehension cohort of an error like *calf* (fed back through comprehension), whereas the lemma of the target *cat* is not in the cohort of the error *dog*. Consequently, if the lemma of *calf* is mistakenly selected for the goal concept CAT(X), there is a higher probability that the error remains undetected during self-monitoring than when the lemma of *dog* is mistakenly selected. Thus, the mixed error bias arises from the design properties of the internal self-monitoring loop.

Rapp and Goldrick (2000) rejected a self-monitoring account of the mixed error bias by arguing that "not only does it require seemingly needless reduplication of information, but because the specific nature of the mechanism has remained unclear, the proposal is overly powerful" (p. 468). However, on the account of self-monitoring through the speech comprehension system, as implemented in WEAVER++, there is no needless reduplication of information. The form representations in word production differ from those in comprehension, but this distinction is not needless because it serves production and comprehension functions. Furthermore, the reduplication is supported by the available latency evidence (see Roelofs, 2003b, for a review). Moreover, the distinction explains dissociations between production and comprehension capabilities in aphasia.

Under the assumption that word production and perception are accomplished via the same form network, one expects a strong correlation between production and comprehension accuracy, as verified through computer simulations by Dell, Schwartz, Martin, Saffran, Gagnon, (1997) and Nickels and Howard (1995). However, such correlations are not observed empirically for form errors (e.g., Dell et al., 1997; Nickels & Howard, 1995). Therefore, Dell et al. (1997) also made the assumption for their own model that form representations are not shared between production and perception, in spite of the presence of backward links in their production network, which might have served speech comprehension. Thus, the assumption implemented in WEAVER++ that form representations are not shared between word production and comprehension is well motivated.

Rapp and Goldrick (2000) argued that the mechanism achieving self-monitoring "has remained unclear" (p. 468). However, because the effect of spoken distractors has been simulated by WEAVER++ (Roelofs, 1997a) and self-monitoring is assumed to be accomplished through the speech comprehension system, the required mechanism is already computationally specified to some extent in the model. Technically speaking, self-monitoring in WEAVER++ is like comprehending a spoken distractor word presented at

a large post-exposure stimulus onset asynchrony (SOA), except that the spoken word is self-generated. In addition, self-monitoring via the speech comprehension system requires cognitive operations to detect discrepancies between selections made in production and comprehension. Lexical selection errors may be detected by verifying whether the lemma of the recognized word is linked to the target lexical concept in production. Errors in lexical concept selection may be detected by verifying whether the lexical concept of the recognized word is linked to the conceptual information derived from the to-be-named object. WEAVER++ implements such verification operations by means of condition-action production rules. Errors in planning and self-monitoring occur when production rules mistakenly fire. The probability of firing by mistake is a function of activation differences among nodes (cf. Roelofs, 1992, 1997a, 2003a).

WEAVER++ employs verification both in self-monitoring and in planning the production of spoken words. Verification in planning achieves that lemmas are selected for intended lexical concepts, morphemes for selected lemmas, segments for selected morphemes, and syllable programs for the syllabified segments. However, whereas verification in word planning happens automatically, verification that achieves self-monitoring is attention demanding. It is unlikely that in self-monitoring, the system can attend simultaneously to all aspects of the speech and at the same time equally well to the internal and external speech. Instead, if internal speech is monitored, external speech is monitored less well. This may explain dissociations between error and repair biases (cf. Nooteboom, this volume, Chapter 10). Thus, although verification in word planning may be seen as a kind of automatic monitoring, it should be distinguished from the operations involved in a speaker's self-monitoring through the speech comprehension system, which are attention demanding.

Computer simulations by Roelofs (2004a) demonstrated that self-monitoring in WEAVER++ suffices to explain the mixed error bias and its dependence on the functional locus of damage in aphasia. The simulations showed that when the lemma of *calf* is mistakenly selected and monitored by feeding its sound form into the speech comprehension system (the internal monitoring loop), the activation level of the lemma of *cat* is increased because of the form overlap with *calf*. However, when the sound form of *dog* is fed back, the activation of the lemma of *cat* is not increased. As a result, the difference in activation between the lemmas of *cat* and *calf* is greater than that between the lemmas of *cat* and *dog*. Consequently, a speaker is more likely to believe that the form of the target *cat* has correctly been prepared for production with the error *calf* than with the error *dog*. By contrast, the simulations showed that the activation of CAT(X) is not much affected by whether a form-related or unrelated item is fed back via the speech comprehension system. Thus, the mixed error bias in WEAVER++ arises at the level of lemmas but not at the level of lexical concepts.

Consequently, a wrong selection of a lexical concept node in naming a

picture because of a conceptual deficit (as observed with patient K.E.) has an equal probability of being caught when the wrong concept has a form-related (*calf*) or a form-unrelated (*dog*) name. In contrast, a wrong selection of a lemma for a correctly selected lexical concept because of a post-conceptual deficit (as observed with patients P.W. and R.G.B.) has a greater probability of being caught when the wrongly selected word has a form-unrelated (*dog*) than a form-related (*calf*) name. Thus, whether a mixed error bias occurs in WEAVER++ depends on the locus of the lesion: The bias occurs with a post-conceptual deficit but not with a conceptual deficit, in agreement with the observations by Rapp and Goldrick (2000).

Accounting for the mixed-distractor latency effect

In testing for production-internal feedback in spoken word production, Damian and Martin (1999) and Starreveld and La Heij (1996) looked at semantic and form effects of spoken and written distractor words in picture naming. Naming latency was the main dependent variable. They observed that the distractors yielded semantic and form effects on picture-naming latencies, and jointly, the effects interacted. For example, the naming of a picture of a cat was interfered with by the semantically related distractor DOG compared with an unrelated distractor, and the naming was facilitated by the phonologically related distractor CAP relative to an unrelated distractor. The semantic interference effect was smaller when target and distractor were phonologically related (distractor CALF versus distractor CAP) than when they were unrelated in form (distractor DOG versus distractor DOLL). This is the mixed-distractor latency effect. According to Damian and Martin (1999), the semantic relatedness and form relatedness of distractors influence successive word-planning stages, namely lexical selection and sound form retrieval (see also Levelt et al., 1999a). Therefore, according to Damian and Martin, the interaction between semantic relatedness and form relatedness of distractors in picture naming suggests that there exists production-internal feedback from sounds to lexical items. According to Starreveld and La Heij (1996), the interaction suggests that semantic and form relatedness influence the same stage in word production. In their view, lexical selection and sound retrieval are one and the same process.

However, in the light of the mixed error bias, the reduced-latency effect of mixed distractors raises an interesting problem. The latency findings suggest that there is less competition from mixed items than from items that are semantically related only (hence faster latencies), whereas on the standard production-internal feedback account, the speech error data suggest more competition for mixed items (hence the larger number of errors). On the feedback account, the mixed-error bias occurs because production-internal feedback of activation makes the lexical node *calf* a stronger competitor than *dog* in planning to say "cat," which is exactly opposite to what an explanation of the latency effect of mixed distractors would seem to require. The latency

data suggest that *calf* is a weaker competitor than *dog* in planning to say "cat." Thus, the challenge for models is to account for both the error and the latency findings.

Starreveld and La Heij (1996) proposed a new word production model without lemmas (cf. Caramazza, 1997) to account for the mixed-distractor latency finding. Their model consists of concept nodes directly connected to unitary phonological word-form nodes. Computer simulations by Starreveld and La Heij showed that their model could account for the mixed-distractor latency effect. Semantic relatedness and form relatedness both affect phonological word-form node selection in the model and therefore the effects interact. However, although the model can capture the latency effect, it fails on the mixed error bias. In planning to say "cat", the phonological word-form nodes of *calf* and *dog* also become active, but *calf* attains the same level of activation as *dog*. This is because there are no segment nodes in the model that are shared between *cat* and *calf*, and their phonological word-form nodes are not connected.

Furthermore, according to the model of Starreveld and La Heij (1996), the reduction of semantic interference and the pure phonological effect necessarily go together. Because semantic relatedness and form relatedness both have their effect through the activation of phonological word-form nodes, a reduction of semantic interference for mixed distractors is only observed in the context of pure form facilitation from phonologically related distractors. Similarly, on the production internal feedback account (Damian & Martin, 1999), semantic and form relatedness interact because activation of production forms spreads back to the level at which semantic effects arise, namely the level of lexical selection. Therefore, a reduction of semantic interference for mixed distractors should only be observed in the context of facilitation from form-related distractors. However, this is not supported empirically. Damian and Martin (1999) presented the spoken distractors at three SOAs. The onset of the spoken distractor was 150 msec before picture onset (SOA = –150 msec), simultaneously with picture onset, or 150 msec after picture onset. They observed semantic interference at the SOAs of –150 and 0 msec, and phonological facilitation at the SOAs of 0 and 150 msec. The mixed distractors yielded no semantic interference at SOA = –150 msec and facilitation at the later SOAs, exactly like the form-related distractors. Thus, the reduction of semantic interference for mixed distractors was already observed at an SOA (i.e., SOA = –150 msec) at which there was no pure phonological facilitation.

Compared to the unrelated distractors, the form effect at SOA = –150 msec was 5 msec, and the effect of form and mixed distractors combined was 2 msec. The point here is not that form related distractors may not yield facilitation at SOA = –150 msec (e.g., Meyer & Schriefers, 1991, and Starreveld, 2000, obtained such an early effect, whereas Damian & Martin, 1999, did not), but that there may be a temporal dissociation between mixed effects and phonological effects. This suggests that the mixed semantic-phonological

effect and the pure form effect are located at different word planning levels, namely the lemma and the word-form level, respectively, as argued by Roelofs et al. (1996).

The assignment of the semantic and form effects to different planning levels is independently supported by the finding that cohort and rhyme competitors yield differential effects in spoken word recognition tasks, whereas they yield similar form effects in picture naming (see Roelofs, 2003b, for an extensive discussion). Whereas form-based activation of cohort competitors in spoken word comprehension is observed (e.g., Zwitserlood, 1989), this does not hold for rhyme competitors when the first segment of the rhyme competitor is more than two phonological features different from the target (e.g., Allopenna et al., 1998; Connine et al., 1993; Marslen-Wilson & Zwitserlood, 1989). In contrast, cohort and rhyme distractors yield form effects of similar size in picture naming (Collins & Ellis, 1992; Meyer & Schriefers, 1991; Meyer & Van der Meulen, 2000), even when they are one complete syllable different from the target.

Meyer and Schriefers (1991) observed that when cohort and rhyme distractors are presented over headphones during the planning of monosyllabic picture names (e.g., the spoken distractors CAP or HAT presented during planning to say the target word "cat"), both distractors yield facilitation compared with unrelated distractors. Also, when cohort and rhyme distractors (e.g., METAL or VILLAIN) are auditory presented during the planning of disyllabic picture names (e.g., "melon"), both distractors yield facilitation too. When the difference in time between distractor and target presentation is manipulated, the SOA at which the facilitation is first detected differs between the two types of distractors. In particular, the onset of facilitation is at an earlier SOA for cohort than for rhyme distractors (i.e., respectively, SOA = –150 msec and SOA = 0 msec). At SOAs where both effects are present (i.e., 0 and 150 msec), the magnitude of the facilitation effect from cohort and rhyme distractors was the same in the study of Meyer and Schriefers (1991). Collins and Ellis (1992) and Meyer and Van der Meulen (2000) made similar observations.

To summarize, the evidence suggests that in spoken word recognition there is some but not much lexical activation of rhyme competitors differing in only the initial segment with a critical word. This contrasts with the findings from spoken distractors in picture naming, where cohort and rhyme distractors word yield comparable amounts of facilitation, even when the target and distractor are one syllable different (i.e., the spoken distractor VILLAIN facilitates the production of the target *melon*).

The difference between the findings from cross-modal priming studies in the spoken word recognition literature (Allopenna et al., 1998; Connine et al., 1993; Marslen-Wilson et al., 1996; Marslen-Wilson & Zwitserlood, 1989) and the findings from spoken distractors in picture naming (Collins & Ellis, 1992; Meyer & Schriefers, 1991; Meyer & Van der Meulen, 2000) is explained if one assumes that spoken distractor words do not activate rhyme competitors at

the lemma level but speech segments in the word-form production network. Roelofs (1997a) provided such an account, implemented in WEAVER++, and reported computer simulations of the effects. On this account, METAL and VILLAIN activate the segments that are shared with *melon* to the same extent (respectively, the segments of the first and second syllable), which explains the findings on picture naming of Meyer and Schriefers (1991). At the same time, METAL activates the lemma of *melon* whereas VILLAIN does not, which accounts for the findings on spoken word recognition of Connine et al. (1993), Marslen-Wilson and Zwitserlood (1989), and Allopenna et al. (1998). Cohort activation (because of begin relatedness) does not have to result in facilitation of lemma retrieval for production in the WEAVER++ model, unless there is also a semantic relationship involved (reducing the semantic interference from mixed distractors).

I argued that the mixed-error effect can at least partly be attributed to self-monitoring in WEAVER++. If in planning to say "cat", the lemma of *calf* is selected instead of the lemma of *cat* and the form of *calf* is fed back through the speech comprehension system, the lemma of *cat* is in the comprehension cohort of the error *calf*. However, if, in planning to say "cat", the lemma of *dog* is selected instead of the lemma of *cat*, then the lemma of the target *cat* is not in the cohort of the error *dog*. Hence, the lemma of *cat* is more active when activation from the word form of the error *calf* is fed back via the comprehension system than when activation from the form of the error *dog* is fed back, and the error *calf* for *cat* is more likely to remain unnoticed in self-monitoring than the error *dog* for *cat*. On this account, the lemma of *calf* is a weaker competitor than the lemma of *dog* in planning to say "cat". That *calf* is a weaker competitor than *dog* in planning to say "cat" also accounts for the mixed distractor latency effect in WEAVER++ (cf. Roelofs et al., 1996), except that the latency effect results from comprehension of the speech of others (i.e., spoken distractor words) rather than from self-monitoring.

The mixed distractor CALF yields less interference than the distractor DOG, because the lemma of the target *cat* is primed as a spoken cohort member during hearing the distractor CALF but not during hearing DOG. Thus, WEAVER++ explains why there is a reduction of semantic interference and an increased change of misselection for mixed items. In summary, according to WEAVER++, mixed items are weaker competitors rather than stronger competitors because of the activation dynamics. Therefore, selection failures concerning mixed items are more likely to remain unnoticed in error monitoring and mixed items have less effect as spoken distractors on latencies in WEAVER++, in agreement with the speech error data and the production latency findings.

Computer simulations by Roelofs (2004a) demonstrated that WEAVER++ exhibits the latency effect of mixed distractors. Damian and Martin (1999) observed empirically that at SOA = –150 msec, the semantically related distractor DOG yielded interference in planning to say "cat", but the mixed

distractor CALF yielded no interference, even though there was no pure form facilitation from CAP at this SOA. This was also the effect that the auditory distractors had on lemma retrieval in WEAVER++ simulations. At SOA = –150 msec in the simulations, the distractor DOG yielded interference in planning to say "cat" but the mixed distractor CALF yielded no interference, even though there was no pure form facilitation from CAP at this SOA. Thus, form relatedness affected lemma retrieval in case of semantic relatedness even when it yielded no pure form facilitation in lemma retrieval (cf. Levelt et al., 1999b). After both lemma retrieval and word-form encoding in the simulations, there were main effects of semantic and phonological relatedness, and jointly the effects interacted, as empirically observed. The results of the simulations were identical with and without comprehension-based feedback. Thus, self-monitoring in WEAVER++ does not affect latency fits of the model.

Cohort effects on mixed errors and production latencies

If the error biases arise during self-monitoring that is accomplished through the speech comprehension system, cohort effects on errors are to be expected. In particular, initial segment overlap between target and intruder should be critical. Dell and Reich (1981) indeed observed that the mixed error effect in their corpus of spontaneous speech errors was strongest for first segment overlap, and much less strong for second, third, or fourth segment overlap. Martin et al. (1996) replicated this seriality finding for picture naming, both with normal and aphasic speakers.

The claim that the mixed-distractor latency effect arises because of comprehension cohorts rather than because of activation feedback within the production system leads to a few new predictions concerning latencies. First, given the finding of Zwitserlood (1989) that word-initial fragments suffice to yield semantic effects in spoken word comprehension, initial fragments of spoken distractor words should yield semantic interference in picture naming, even when a fragment does not uniquely identify a word. In contrast, Damian and Martin (1999) and Starreveld (2000) argued that such effects of fragments cannot occur in a picture-word interference task. According to them, semantic effects of spoken distractors only occur when the spoken-word recognition system has "settled into a state in which only one candidate (corresponding to the distractor) is activated" (Starreveld, 2000, p. 518). Damian and Martin (1999) expressed the same view: "Semantic access in auditory word recognition critically depends on when lexical uniqueness is achieved" (p. 351). If lexical uniqueness is required to obtain a semantic effect, then the interaction between semantic and form effects cannot be a comprehension cohort effect: The cohort account assumes that CA suffices to activate the lemmas of *cat* and *calf*. Second, given that rhyme competitors are not much activated in speech comprehension (Allopenna et al., 1998; Connine et al., 1993), replicating the study of Damian and Martin (1999) with rhyme

competitors should yield additive rather than interactive effects of semantic and phonological relatedness.

The predictions about the spoken fragments and the rhyme distractors have been confirmed recently (Roelofs, submitted-a, -b). The upper panel of Figure 3.4. shows the semantic and phonological effects of word-initial fragments. Participants had to name, for example, a pictured tiger while hearing PU (the first syllable of the semantically related word *puma*), TI (phonologically related, the first syllable of the target *tiger*), or an unrelated syllable. The figure shows that fragments such as PU yielded semantic interference and that

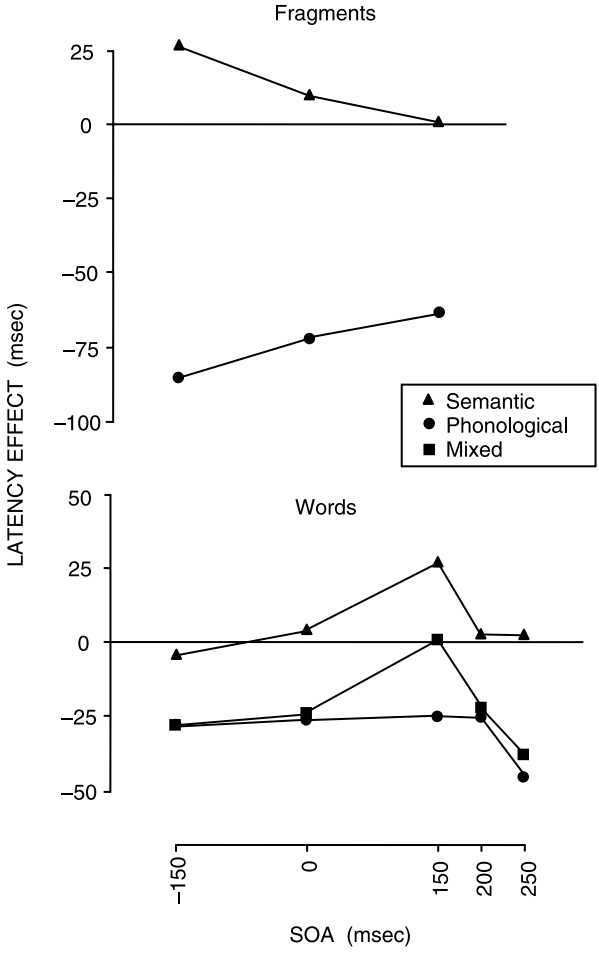

Figure 3.4 Latency effect (in msec) of semantically related, phonologically related, and mixed spoken distractors relative to unrelated distractors in picture naming as a function of SOA (in msec).

fragments such as TI produced phonological facilitation, the first finding supporting the cohort assumption.

The lower panel of Figure 3.4 shows the semantic, phonological, and mixed effects when the form-related and mixed items are rhyme competitors of the target. The mixed distractors were matched for semantic and phonological relatedness to the semantically related and the phonologically related distractors, respectively. Participants had to name, for example, a pictured dolphin while hearing SPARROW (semantically related), MUFFIN (phonologically related), ROBIN (mixed), or an unrelated word. The distractors yielded semantic and form effects, and together, the effects were additive, unlike the interactive effects from cohort distractors observed by Damian and Martin (1999). The semantic and form effects occurred at a positive SOA, presumably because the onset of the critical rhyme fragment determined the SOA in the experiment following Meyer and Schriefers (1991). By contrast, with the fragments and in the study of Damian and Martin (1999), the onset of the distractor determined the SOA. The additivity of the effects poses difficulty to models with production-internal feedback. Production-internal feedback of activation from the mixed rhyme distractor ROBIN should activate the target *dolphin*, just like the cohort distractor CALF activates the target *cat*. Therefore, the reduced latency effect should also be obtained for mixed rhyme distractors, contrary to the empirical findings. Importantly, the fact that the phonologically related rhyme distractor MUFFIN facilitated the production of *dolphin* suggests that the form overlap of rhyme distractors did have an effect, replicating earlier studies (Collins & Ellis, 1992; Meyer & Schriefers, 1991; Meyer & Van der Meulen, 2000).

Summary and conclusions

Aphasic and nonaphasic speakers listen to their own talking and they prevent and correct many of the errors made in the planning and actual production of speech. To account for this type of output control, models need to make assumptions about self-monitoring. I have explained the relations among planning, comprehending, and self-monitoring of spoken words in WEAVER++, a feedforward word-production model that assumes self-monitoring through comprehension-based feedback (Roelofs, 2004a, b). The self-monitoring through comprehension of the model was shown to be supported by a new analysis of the self-corrections and false starts in picture naming by 15 aphasic speakers. Furthermore, the model explained findings that seemingly require production-internal feedback: the mixed-error bias and its dependence on the locus of damage in aphasia, and the reduced latency effect of mixed distractors. Finally, the attribution of the mixed-distractor latency effect to comprehension cohorts by the model was shown to be supported by recent experimental research, which revealed semantic effects of word-initial cohort distractors and the absence of a reduced latency effect for mixed rhyme distractors. To conclude, the interplay among speaking,

comprehending, and self-monitoring is not only of interest in its own right, but it also illuminates classic issues in production.

References

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language*, *38*, 419–39.

Caramazza, A. (1997). How many levels of processing are there in lexical access? *Cognitive Neuropsychology*, *14*, 177–208.

Collins, A., & Ellis, A. (1992). Phonological priming of lexical retrieval in speech production. *British Journal of Psychology*, *83*, 375–88.

Connine, C. M., Blasko, D. G., & Titone, D. (1993). Do the beginnings of spoken words have a special status in auditory word recognition? *Journal of Memory and Language*, *32*, 193–210.

Damian, M. K., & Martin, R. C. (1999). Semantic and phonological codes interact in single word production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*, 345–61.

Dell, G. S., & Reich, P. A. (1981). Stages in sentence production: An analysis of speech error data. *Journal of Verbal Learning and Verbal Behavior*, *20*, 611–29.

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

Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999a). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, *22*, 1–38.

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999b). Multiple perspectives on word production. *Behavioral and Brain Sciences*, *22*, 61–75.

Marslen-Wilson, W., Moss, H. E., & van Halen, S. (1996). Perceptual distance and competition in lexical access. *Journal of Experimental Psychology: Human Perception and Performance*, *22*, 1376–92.

Marslen-Wilson, W. D., & Welsh, A. (1978). Processing interactions and lexical access during word recognition in continuous speech. *Cognitive Psychology*, *10*, 29–63.

Marslen-Wilson, W. D., & Zwitserlood, P. (1989). Accessing spoken words: The importance of word onsets. *Journal of Experimental Psychology: Human Perception and Performance*, *15*, 576–85.

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.

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

McGuire, P. K., Silbersweig, D. A., & Frith, C. D. (1996). Functional neuroanatomy of verbal self-monitoring. *Brain*, *119*, 907–17.

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: Learning, Memory, and Cognition*, *17*, 1146–60.

Meyer, A. S., & Van der Meulen, F. F. (2000). Phonological priming effects on speech onset latencies and viewing times in object naming. *Psychonomic Bulletin & Review*, *7*, 314–19.

Nickels, L. (1997). *Spoken word production and its breakdown in aphasia*. Hove: Psychology Press.

Nickels, L., & Howard, D. (1995). Phonological errors in aphasic naming: Comprehension, monitoring, and lexicality. *Cortex*, *31*, 209–37.

Nickels, L., & Howard, D. (2000). When the words won't come: Relating impairments and models of spoken word production. In L. Wheeldon (Ed.), *Aspects of language production* (pp. 115–42). Hove, UK: Psychology Press.

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

Paus, T., Perry, D. W., Zatorre, R. J., Worsley, K. J., & Evans, A. C. (1996). Modulation of cerebral blood flow in the human auditory cortex during speech: Role of motor-to-sensory discharges. *European Journal of Neuroscience*, *8*, 2236–46.

Rapp, B., & Goldrick, M. (2000). Discreteness and interactivity in spoken word production. *Psychological Review*, *107*, 460–99.

Rapp, B., & Goldrick, M. (2004). Feedback by any other name is still interactivity: A reply to Roelofs (2004). *Psychological Review*, *111*, 573–8.

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

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

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

Roelofs, A. (1997a). The WEAVER model of word-form encoding in speech production. *Cognition*, *64*, 249–84.

Roelofs, A. (1997b). A case for nondecomposition in conceptually driven word retrieval. *Journal of Psycholinguistic Research*, *26*, 33–67.

Roelofs, A. (1997c). Syllabification in speech production: Evaluation of WEAVER. *Language and Cognitive Processes*, *12*, 657–93.

Roelofs, A. (1998). Rightward incrementality in encoding simple phrasal forms in speech production: Verb-particle combinations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*, 904–21.

Roelofs, A. (1999). Phonological segments and features as planning units in speech production. *Language and Cognitive Processes*, *14*, 173–200.

Roelofs, A. (2003a). Goal-referenced selection of verbal action: Modeling attentional control in the Stroop task. *Psychological Review*, *110*, 88–125.

Roelofs, A. (2003b). Modeling the relation between the production and recognition of spoken word forms. In A. S. Meyer & N. O. Schiller (Eds.), *Phonetics and phonology in language comprehension and production: Differences and similarities* (pp. 115–158). Berlin: Mouton.

Roelofs, A. (2004a). Error biases in spoken word planning and monitoring by aphasic and nonaphasic speakers: Comment on Rapp and Goldrick (2000). *Psychological Review*, *111*, 561–72.

Roelofs, A. (2004b). Comprehension-based versus production-internal feedback in planning spoken words: A rejoinder to Rapp and Goldrick (2004). *Psychological Review*, *111*, 579–80.

Roelofs, A. (submitted-a). *Word-initial cohort effects of spoken distractors in picture naming*.

Roelofs, A. (submitted-b). *Autonomous stages in planning the production of spoken words: Evidence from semantic and form effects of rhyme distractors in picture naming*.

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.

Roelofs, A., Meyer, A. S., & Levelt, W. J. M. (1998). A case for the lemma-lexeme distinction in models of speaking: Comment on Caramazza and Miozzo (1997). *Cognition*, *69*, 219–30.

Starreveld, P. A. (2000). On the interpretation of context effects in word production. *Journal of Memory and Language*, *42*, 497–525.

Starreveld, P. A., & La Heij, W. (1996). Time-course analysis of semantic and orthographic context effects in picture naming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*, 896–918.

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

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