

# 4 The Architecture of Speaking

ARDI ROELOFS AND VICTOR S. FERREIRA

Theorizing about the architecture of speaking has a long history (see Levelt, 2013), ranging from the classic work of Wernicke (1874), Lichtheim (1885), and Wundt (1900) in the 19th century to that of Garrett (1975) and Levelt (1989) in the modern era. Although Wernicke gave his cognitive account a neuroanatomical basis using lesion-deficit analyses, later architectures were often mainly based on behavioral evidence (but see Geschwind, 1970). Since the 1990s, structural and functional neuroimaging has increasingly informed theoretical views on the cognitive architecture of speaking (see Shallice & Cooper, 2011, for a defense of this approach; see Kemmerer, 2015, for a literature review). In this chapter, we outline a neuroanatomically grounded cognitive architecture for speaking that takes account of modern neuroimaging insights.

Speaking involves multiple component processes and representations. We describe current views on these processes and representations and discuss debates along the way. It is generally assumed that three major types of processes underlie speaking: conceptualization, formulation, and articulation (Garrett, 1975; Levelt, 1989). Conceptualization processes generate messages, which are conceptual representations to be verbally expressed. Formulation processes use a message to retrieve appropriate words and to build a syntactic structure, which is used to generate a phonetic plan for the utterance. Articulation processes execute the phonetic plan, which yields overt speech. During conceptualization, formulation, and articulation, a speaker monitors progress on these processes. Figure 4.1 outlines the major component processes and representations, largely following Levelt (1989). We first briefly describe these processes and representations and discuss them in greater detail in the remainder of this chapter.

Conceptualization processes generate a message, for example, based on visually perceived scenes (see Konopka & Brown-Schmidt, 2014, for a review). Assume that your friend's cat is lost and you discover that the animal is hidden under the kitchen table. To say this to your friend requires the construction of a message making this information explicit, for example, BE(CAT, UNDER(TABLE)). The intended utterance *the cat is under the table* is useful if your friend doesn't know this

fact yet (which is in your model of the addressee) and you are with her in the kitchen (as specified by your situation model; otherwise, you should make the exact table explicit by saying *kitchen table*). A controversial and unresolved issue is whether a message is specified in terms of nondecomposed lexical concepts (e.g., Roelofs, 1992), such as CAT and TABLE, or in terms of the conceptual features of these concepts (e.g., Bierwisch & Schreuder, 1992; Dell, Schwartz, Martin, Saffran, & Gagnon, 1997), such as FURRY, MEOWING, PURRING, FELINE, ANIMAL. Either way, messages are input to formulation processes.

According to a prominent view, the formulation of sentences is (for the most part) lexically driven (for reviews, see Ferreira, 2010; Bock & Ferreira, 2014). That is, concepts in the message activate corresponding words in memory, which make available the syntactic and morphophonological information that is needed for constructing a syntactic structure and phonetic plan. During syntactic encoding, syntactic constituents are assembled (e.g., the noun phrase *the cat*) and grammatical functions are assigned (e.g., *the cat* is made the subject of the sentence). During morphophonological encoding, the sound structure of the utterance is constructed, including its prosody. Some theorists assume that formulation processes draw on distinct declarative and procedural types of knowledge (e.g., Levelt, 1989), whereas other theorists do not make this distinction (e.g., Chang & Fitz, 2014). Declarative knowledge includes knowing that the word *cat* is a noun, and procedural knowledge includes knowing how to construct a word form and a noun phrase. According to a prominent theory (Levelt, Roelofs, & Meyer, 1999), a particular type of representation in declarative memory, called a lemma, links the conceptual, syntactic, and phonological representations of a word. For example, the lemma of the word *cat* links conceptual information (e.g., the lexical concept CAT, linked to other concepts such as FURRY, MEOWING, ANIMAL), syntactic properties such as word class and syntactic frame (e.g., specifying that a noun is the head of a noun phrase), morphological and phonological information (e.g., the morpheme <cat> and the phonemes /k/, /æ/, and /t/), and phonetic information (e.g., a motor program for [kæt]). In this theory, the phonetic plan makes explicit

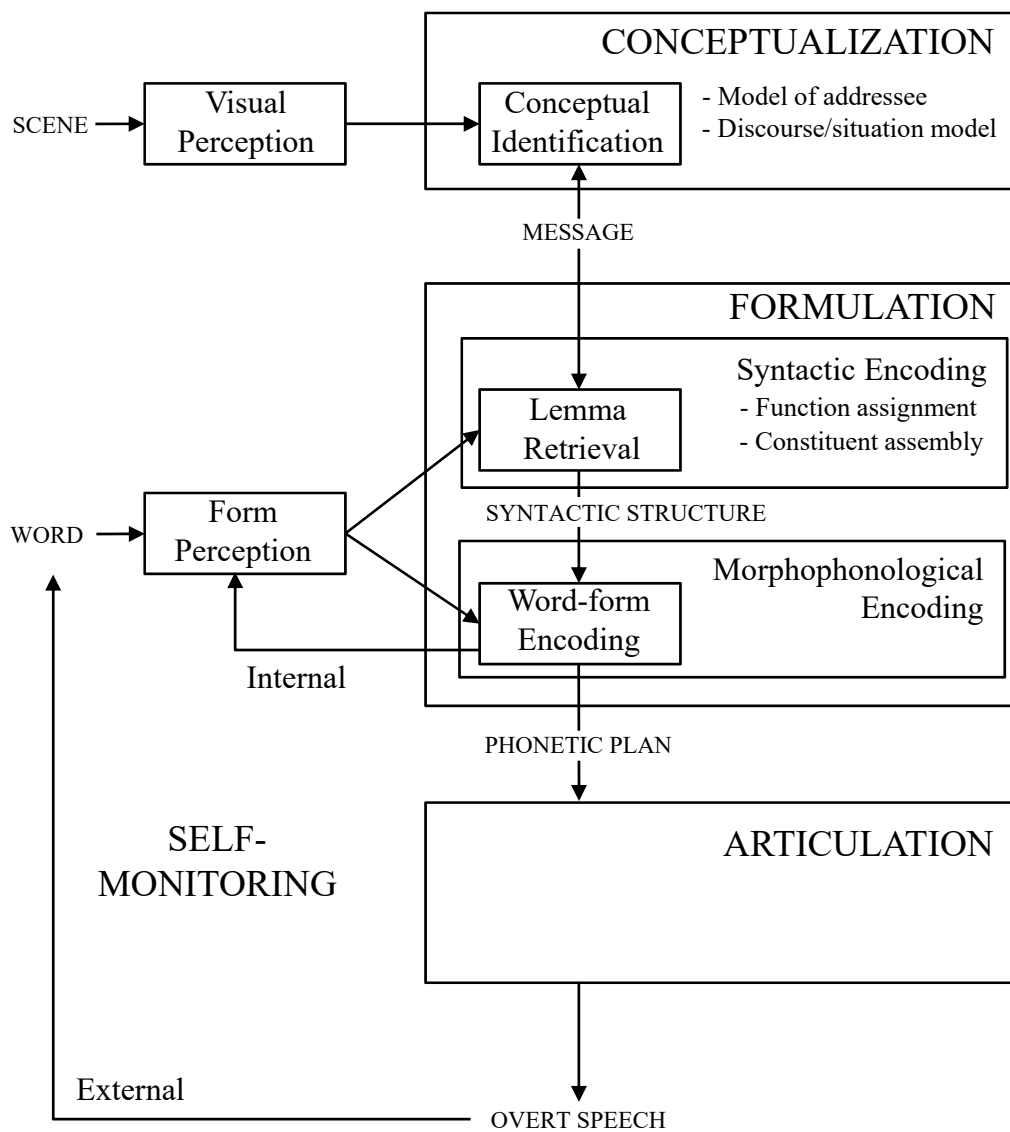


FIGURE 4.1 Cognitive architecture for speaking.

the motor programs for syllables, which are executed by the articulation process.

Through self-monitoring, a speaker assesses whether formulation and articulation are consistent with intent, and if not, this may trigger a correction (e.g., *the dog... uh... cat is under the table*). Errors may not only be detected when speakers hear their own overt speech, but also already before articulation onset. According to a comprehension-based account, the latter is accomplished by internal feedback through the speech perception system (Levelt, 1989; Levelt et al., 1999), whereas according to a production-based account, errors are detected inside the production system, for example, by monitoring for conflict (Nozari, Dell, & Schwartz, 2011).

In the remainder of this chapter, we provide further details on this cognitive architecture for speaking. We

discuss key evidence from behavioral studies measuring response time and accuracy, eye tracking, electrophysiological recordings, hemodynamic neuroimaging, tractography, disorders, and computational modeling. Converging evidence from multiple methods is discussed with the idea that if evidence from multiple sources is in agreement, the conclusion can be strong even if each method has its weaknesses (cf. Shallice & Cooper, 2011). Moreover, some major past and current controversies are evaluated.

### 1. Word Production

For word-level processes in speaking, a number of computational models, which agree but also differ in several respects, have been developed. Differences concern

the number and nature of the representation levels and the information flow between the levels. For example, the WEAVER++ model of Levelt et al. (1999) assumes concept, lemma, morpheme, phoneme, and motor program levels, whereas the model of Dell et al. (1997) assumes concept, lexical, and phoneme levels only (but see Dell, 1986). We use the more extensive model of Levelt et al. as a framework and discuss empirical evidence for and against the levels assumed by this model. The main processing stages in the model are lemma retrieval and word-form encoding, which is further subdivided into morphological, phonological, and phonetic encoding. Next, we discuss evidence on the nature of processing within each stage (i.e., whether lexical selection is a competitive process) and the information flow between levels (i.e., whether this is discrete, cascading, or interactive). Moreover, we discuss the relation between word planning and word comprehension, which supports self-monitoring in the model. The WEAVER++ model does not say much about phonetic encoding and articulation, which is addressed by several other models in the literature (e.g., Guenther & Vladusich, 2012; Hickok, 2012). We briefly discuss the most comprehensive of these models, namely the DIVA model of Guenther and colleagues, which fits well with WEAVER++.

1.1. LEMMA RETRIEVAL AND WORD-FORM ENCODING Figure 4.2 outlines a recent extension of WEAVER++, called WEAVER++/ARC, which synthesizes behavioral psycholinguistic, functional neuroimaging, tractography, and aphasiological evidence (Roelofs,

2014). The model makes a distinction between a declarative lexical network and a procedural rule system (cf. Levelt et al., 1999), following Levelt (1989), Ullman (2004), and others. There is considerable cognitive neuroscience evidence for distinct declarative and procedural memory systems in the brain (see, e.g., Eichenbaum, 2012, for review). The declarative lexical network is thought to be underpinned by temporal and inferior frontal areas of the human brain (including Wernicke's and Broca's areas), whereas the procedural system is assumed to be underpinned by, among others, the basal ganglia, thalamus, frontal cortex (including Broca's area), and cerebellum. The lexical network is accessed by spreading activation while procedural if-then rules select nodes among the activated lexical information depending on the task demands specified in working memory. Activation spreads continuously from level to level (Roelofs, 2008), whereby each node sends a proportion of its activation to connected nodes. After a message concept is selected, an if-then rule enhances its activation (thought to be achieved through a ventral white matter pathway running from frontal cortex, including Broca's area, to anterior-ventral temporal cortex, possibly including the inferior frontal-occipital fasciculus) to achieve speeded and accurate lemma retrieval and word-form encoding operations.

The lexical network consists of lexical concepts (e.g., CAT) thought to be represented in anterior-ventral temporal cortex, lemmas (e.g., *cat*) in the middle section of the left middle temporal gyrus (MTG), input phonemes (e.g., /k/, /æ/, and /t/) and input and output lexical forms or morphemes (e.g., <cat>) in left

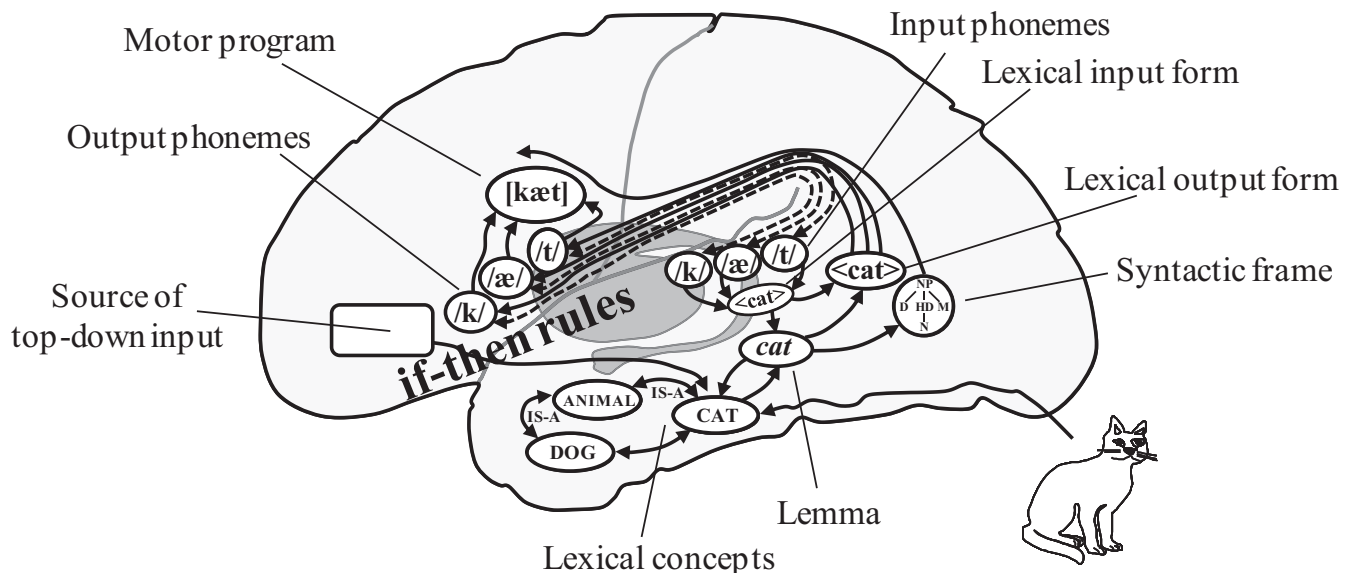


FIGURE 4.2 Schematic of the WEAVER++/ARC model of spoken word production and its relation with comprehension and repetition. D=determiner; HD=head; M=modifier; N=noun; NP=noun phrase.

posterior superior temporal gyrus (STG) and posterior MTG (Wernicke’s area), output phonemes (e.g., /k/, /æ/, and /t/) in left posterior inferior frontal gyrus (Broca’s area), and syllable motor programs (e.g., [kæt]) in ventral precentral gyrus. The functional localizations are based on an extensive meta-analysis of neuroimaging studies performed by Indefrey and Levelt (Indefrey & Levelt, 2004; Indefrey, 2011), with converging evidence from lesion-deficit analyses in aphasia (e.g., Schwartz, Faseyitan, Kim, & Coslett, 2012; Schwartz et al., 2009). Lexical concepts are assumed to be part of a hub of supramodal conceptual representations (in the anterior temporal lobes), which integrate modality-specific features that are represented in widespread brain areas for perception and action (Lambon Ralph, 2014; Patterson, Nestor, & Rogers, 2007; see Kemmerer, 2015, for a review). Lemmas specify the grammatical properties of words, which are thought to be represented in left posterior STG and MTG (Hagoort & Indefrey, 2014). These properties are propagated to Broca’s area for the encoding of phrases and sentences (see section 2.1 for further discussion), as suggested by evidence from neuropsychology (Sapolsky et al., 2010; Wilson et al., 2010) and neuroimaging (Haller, Radue, Erb, Grodd, & Kircher, 2005; Indefrey et al., 2001). For example, the lemma of *cat* specifies that the word is a noun (for languages such as Dutch, lemmas also specify grammatical gender) and also that it is the head of a noun phrase that may include a determiner and modifiers. Lemmas also allow for the specification of abstract morphosyntactic parameters, such as number (singular, plural) for nouns and number, person (first, second, third), and tense (past, present) for verbs, so that the appropriate lexical output forms may be retrieved (i.e., one or more morphemes, such as singular <cat>).

The connections from lexical output forms to output phonemes are thought to be underpinned by a dorsal white matter pathway in the brain made up by the left arcuate fasciculus. Also, connections from input to output phonemes, serving repetition (i.e., saying heard words or pseudowords), are assumed to be underpinned by this pathway. In vivo tractography and fiber microdissection of postmortem human brains suggest that there are two distinct subtracts within the left arcuate fasciculus (Brauer, Anwender, Perani, & Friederici, 2013; Fernandez-Miranda et al., 2015; Glasser & Rilling, 2008). One subtract interconnects left posterior MTG and Broca’s area, and the other subtract interconnects left posterior STG and Broca’s area. The MTG pathway is assumed to map lexical output forms for lexical-semantic representations onto output phonemes, whereas the STG pathway maps input phonemes onto output phonemes. Output phonemes also activate

input phonemes, followed by activation of lexical input forms, lemmas, and lexical concepts, which supports internal self-monitoring in the WEAVER++/ARC model (Levelt et al., 1999; Roelofs, 2004b).

In word production, activation traverses from lexical concepts to syllable motor programs. If-then rules select lexical concepts, lemmas, output lexical forms (i.e., morphemes), and output phonemes; syllabify the phonemes; and assign a stress pattern across syllables in polysyllabic words. The resulting phonological word representation is used to select the corresponding syllable motor programs, which are stored in a syllabary (Levelt et al., 1999). Suppose, for example, that the plural form of *cat* needs to be produced. The lemma of the word *cat* will be retrieved for the lexical concept CAT plus a plural specification. Next, during morphological encoding, the lemma with the number parameter set to *plural* will be mapped onto the stem <cat> and the plural suffix <-s>. During phonological encoding, phonemes are selected for these morphemes, namely /k/, /æ/, /t/ for the stem and /s/ (rather than /z/) for the suffix. The phonemes are then syllabified (and for polysyllabic words, a stress pattern is assigned), yielding a phonological word representation. Sahin, Pinker, Cash, Schomer, and Halgren (2009) obtained evidence from intracranial electrophysiology that morphological and phonological encoding are spatially and temporally distinct processes within Broca’s area. During phonetic encoding, the constructed phonological representation is used to select corresponding syllable motor programs (i.e., [kæts]). Evidence that these programs are retrieved from a syllabary comes from syllable frequency effects, which have been obtained for several languages (e.g., Cholin, Dell, & Levelt, 2011, for a review). Electrophysiological evidence suggests that the syllable frequency effect arises about 100–200 ms before articulation onset (Bürki, Pellet, & Laganaro, 2015), supporting the syllabary hypothesis.

Several of the processing and representational assumptions made by the WEAVER++ model are debated in the literature. Major controversies concern the distinction between lemmas and lexical forms (e.g., Caramazza & Miozzo, 1997), the activation flow between levels (e.g., Rapp & Goldrick, 2000), the nature of lexical selection (e.g., Dhooge & Hartsuiker, 2010), and the functional role of the arcuate fasciculus (e.g., Ueno, Saito, Rogers, & Lambon Ralph, 2011). We address these controversies in turn.

**1.2. LEVELS OF REPRESENTATION** Classic evidence for the distinction between lemmas and lexical forms comes from speech errors, in particular, the existence of two types of morpheme exchanges (Dell, 1986;

Garrett, 1980, 1988). The properties of these exchanges provide support for a lemma level with abstract morphosyntactic parameters that is distinct from a morphological level with concrete lexical form representations. 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 this concerns a morpheme error, since the plurality of *apple* is stranded, that is, it is realized on *pie*. The distributional properties of these morpheme exchanges are similar to those of whole-word exchanges, such as the exchange of *roof* and *list* in “we completely forgot to add the *list* to the *roof*” (from Garrett, 1980). 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 phoneme exchanges, such as “she is a real *rack pat*” for “*pack rat*” (from Garrett, 1988). This suggests that this second type of morpheme error and phoneme errors occur at the same level of processing, namely the level at which morphemes and phonemes are retrieved and the morphophonological form of the utterance is constructed. The errors occur when morphemes in a developing morphophonological structure trade places.

However, Caramazza and Miozzo (1997) argued against lemmas based on evidence from aphasic patients indicating that semantic errors and syntactic class deficits may be restricted to speaking or writing. For example, some patients make semantic substitution errors (e.g., *dog* for *cat*) in spoken but not in written naming, while the reverse holds true for other patients. Also, patients may have difficulty saying but not writing verbs, or vice versa. According to Caramazza and Miozzo, this suggests that conceptually driven and syntactically constrained lexical selection concerns modality-specific lexical forms rather than modality-neutral lemmas. On this account, brain damage in patients may differentially affect these conceptually and syntactically specified written and spoken forms. However, on an alternative account (Roelofs, Meyer, & Levelt, 1998), the brain damage may differentially affect the connections between lemmas and spoken or written forms (but see Caramazza & Miozzo, 1998, for a rejoinder).

**1.3. ACTIVATION FLOW** Another issue concerns the activation flow between the network levels. According to one view (Levelt et al., 1999), activation spreads from a lemma to its associated word form only upon

selection of the lemma. Moreover, within the form network, activation spreads forward from morphemes to phonemes and syllable programs, but not backward. Others (Dell, 1986; Dell et al., 1997; Rapp & Goldrick, 2000) hold that activation spreads forward from lemmas to forms and also backward from forms to lemmas.

Accumulating evidence has suggested that activation of word forms does not depend on lemma selection, contrary to what Levelt et al. (1999) maintained. For example, Morsella and Miozzo (2002) observed that in naming a target picture in green superimposed on a distractor picture in red, the naming time is shorter when the picture names share part of their form (e.g., *cat* and *cap*) than when they do not (e.g., *cat* and *pen*). This suggests that activation cascades from concepts (CAT, CAP, PEN) to lemmas (*cat*, *cap*, *pen*) and corresponding forms (/k/, /æ/, /t/, /p/, /e/, /n/), which speeds up the encoding of the target form (*cat*) when the distractor picture activates overlapping phonemes (i.e., /k/ and /æ/). Based on these and other findings, the discreteness assumption that was originally part of WEAVER++ (i.e., Levelt et al., 1999) has been dropped by Roelofs (2008).

Classic evidence that activation spreads backward comes from speech error biases (Dell et al., 1997), such as the statistical overrepresentation of mixed semantic-phonological errors (e.g., *rat* for *cat* vs. *dog* for *cat*) and the statistical overrepresentation of word errors among phonological errors (i.e., lexical bias). These error biases suggest that activation spreads back from forms to lemmas, yielding phonological influences on lemma retrieval and lexical influences on phoneme selection. However, it has remained unclear whether activation spreads back directly from output phonemes to output lexical forms and lemmas (Dell et al., 1997; Rapp & Goldrick, 2000) or via the speech perception network (i.e., via the connections between output and input phonemes). Moreover, it has been argued that the error biases are, at least partly, due to self-monitoring (e.g., Levelt et al., 1999). We refer to Rapp and Goldrick (2000, 2004) and Roelofs (2004a, 2004b) for an extensive discussion.

In meta-analyses of magnetoencephalography (MEG) studies on word production, Indefrey and Levelt (2004) and Munding, Dubarry, and Alario (2016) observed that it takes several hundred milliseconds for activity to progress from occipital-temporal via middle temporal to inferior frontal and precentral areas in picture naming. This evidence supports WEAVER++/ARC and the models of Dell et al. (1997) and Rapp and Goldrick (2000). We refer to Strijkers and Costa (2016) for a different view, which holds that lexical-semantic and



phonological representations are activated in parallel in temporal and frontal cortex (but see Indefrey, 2016, for a critical discussion).

**1.4. SELECTION BY COMPETITION** Researchers also have found no agreement on how lemma selection occurs, see Abdel Rahman and Melinger (2009) and Navarrete, Del Prato, Peressotti, and Mahon (2014) for reviews of work in this area and multiple perspectives about the selection issue. WEAVER++ implements a competition account (e.g., Roelofs, 1992), which holds that lemmas compete for selection. Evidence for competition comes from semantic interference in picture naming. In the picture-word interference paradigm, speakers name pictures while trying to ignore spoken or written distractor words. For example, they say “cat” to a pictured cat combined with the semantically related word *dog* or the unrelated word *pen*. The picture naming time is typically longer on semantically related than on unrelated trials. According to the competition account, the lemma of a semantically related distractor word (*dog*) receives activation from the target picture (cat) and is therefore a stronger competitor to the picture name than the lemma of an unrelated distractor word (*pen*), which is not activated by the picture. However, other researchers have argued that lemma selection does not happen by competition (e.g., Finkbeiner & Caramazza, 2006). According to this noncompetition account, a word is selected if its activation exceeds some threshold, but selection is assumed to be independent of the activation state of other words. Semantic interference is assumed to arise after word planning, reflecting the exclusion of an articulatory response to the distractor word from an output buffer. However, using MEG, Piai, Roelofs, Jensen, Schoffelen, and Bonnefond (2014) observed that the semantic interference effect in picture naming occurs in the time window of word planning, well before articulation onset. Moreover, the semantic interference effect arises in the left MTG, where lemmas are thought to be stored.

**1.5. DORSAL VERSUS VENTRAL PATHWAY** Controversy also exists about the white matter pathway that mediates the mapping of concepts onto motor-related form representations. According to the classic view (e.g., Geschwind, 1970), the arcuate fasciculus is the critical white matter fiber tract. This view has been adopted by WEAVER++/ARC (Roelofs, 2014). In the model, lemmas are linked to lexical output forms, which are thought to be connected by the arcuate fasciculus to output phonemes that are linked to motor programs (see figure 4.2). Dell, Schwartz, Nozari, Faseyitan, and

Coslett (2013) also adhere to a dorsal pathway view. However, according to an alternative account, computationally implemented in the Lichtheim 2 model (Ueno et al., 2011), concepts are mapped onto motor-related form representations primarily by a ventral pathway underpinned by the left uncinate fasciculus and fiber tracts passing through the left extreme capsule, which include the inferior fronto-occipital fasciculus and a branch of the middle longitudinal fasciculus. Computer simulations revealed that both Lichtheim 2 and WEAVER++/ARC can account for the typical patterns of impaired and spared language performance associated with classic abrupt-onset and progressive aphasias (Roelofs, 2014; Ueno et al., 2011). However, WEAVER++/ARC but not Lichtheim 2 accounts for the evidence that damage to the dorsal but not the ventral pathway yields impaired production performance (Marchina et al., 2011; Wang, Marchina, Norton, Wan, & Schlaug, 2013).

**1.6. ARTICULATION** In line with WEAVER++/ARC, the DIVA model (e.g., Guenther & Vladusich, 2012) assumes that motor programs for syllables play an important role in speech motor control. DIVA assumes that access codes for these programs (cf. the syllable program nodes in WEAVER++/ARC) are stored in left posterior inferior frontal gyrus and premotor cortex. These access codes activate representations that specify the articulatory tasks to be performed in pronouncing the syllables (such as closing the lips), thought to be represented in bilateral motor cortex. Initiation of the execution of the motor programs is controlled by the basal ganglia, whereas the temporal coordination of the motor execution is controlled by the cerebellum. Important for learning the motor programs or executing them under special circumstances (e.g., speaking while eating), the access codes for syllables also generate predictions about the auditory and somatosensory effects of producing them. These predictions may be compared with the actually perceived effects, and discrepancies may be used to adjust the motor programs.

## 2. Sentence Production

Sentences are collections of words, but producing a sentence involves much more than retrieving words (as described by the models) and stringing them one after another. This is because whereas individual words have the function of conveying specific entities (e.g., *cat*), states (*mangy*), and actions (*eat*), sentences have the function of conveying entire events (e.g., *the mangy cat is eating*). This greater expressive function brings with it

the need for a more complex set of production mechanisms.

**2.1. STAGES OF SENTENCE PRODUCTION** Like word production, sentence production is often viewed as proceeding through a sequence of processing stages, and ones that correspond to those involved in word production (see figure 4.1). A number of accounts have been put forth of this type (e.g., Bock & Levelt, 1994; Ferreira & Slevc, 2007). Here, we describe an account based largely on an influential proposal by Garrett (1975; see also Garrett, 1982, 1988), adapted to some extent by Ferreira and Slevc.

Any act of production begins with the speaker determining his or her intent for the utterance, what is often termed a *speech act*. Then, a process that is sometimes termed *message encoding* has the function of determining the features of meaning the speaker must express to be able to fulfill his or her intent. The message that serves as the input to sentence production includes as one of its components the same conceptual information that is used for word production. This component of a message can be termed *semantic meaning*—the who, what, and whom participating in an event. But additionally, the message for sentence production must represent *relational meaning*—who did what to whom in the event. Finally, the message must also include *perspective meaning*—what part of the event is more important and should be foregrounded, and what part is less important and should be backgrounded.

From this multifaceted message (consisting of semantic, relational, and perspective meaning), sentence production proceeds through two separate though interdependent processing streams. One stream, corresponding to the mechanisms of word production, has the task of retrieving the content words that are to be produced in the sentence. This *content stream* can be seen as taking the semantic meaning that is represented in the multifaceted message and engaging the process of lemma retrieval, that is, the retrieval of representations of the individual words to be spoken in the sentence, including those words' grammatical properties (e.g., form class, and if needed, grammatical gender), as described in section 1.1.

Meanwhile, the other stream has the task of determining the structural characteristics of the sentence to be produced. The first step within this *structure stream*, which operates in tandem and interdependently with lemma retrieval within the content stream, is often termed *function assignment* (Bock & Levelt, 1994; Bock, 1995; Ferreira & Slevc, 2007). Function assignment uses the relational and perspective components of the multifaceted message to determine the grammatical

functions that will be needed in a sentence. If the relational meaning represented in the discourse corresponds to an intransitive sentence such as *the cat sleeps*, function assignment must formulate grammatical functions for the subject and verb; for a transitive sentence like *the cat eats the food*, function assignment must additionally formulate a direct object; for a ditransitive sentence like *the cat bats the toy to its owner*, function assignment must additionally formulate an indirect object. Meanwhile, as lemma retrieval mechanisms retrieve lemmas for the appropriate content words, those lemmas must become *bound* to the corresponding grammatical functions, so that the subject noun is spoken as the grammatical subject, the object noun as the direct object, and so forth (and if these bindings become crossed up, a speech error of the type reviewed in section 1.2 results).

As lemmas are retrieved in the content stream and functions are assigned in the structure stream, the next stages of sentence production can unfold. On the content side, the above-described process proceeds: lemma retrieval feeds to word-form encoding, wherein the morphophonological specification of the words to be produced in the sentence is determined. On the structure side, function assignment proceeds to a process often termed *constituent assembly* (Ferreira & Slevc, 2007) or *positional processing* (Bock, 1995; Bock & Levelt, 1994; Garrett, 1975). Constituent assembly processes take the grammatical functions that were determined at function assignment and specify a hierarchical structure that represents the relations among the grammatical functions in a manner that allows that structure to be linearized into a sequence of words. To do so, constituent assembly processes must consult the language-internal grammatical rules or conventions that dictate how grammatical functions are to be linguistically encoded. As the parallel word-form encoding process (within the content stream) specifies the morphophonological features of words to be produced in the sentence, those features must be coordinated with (or bound to) the hierarchical structures that are built by constituent assembly processes, so that the syllables that compose the words of the sentence can be arranged into the required linear sequence. As the syllabic sequence is determined, the *prosody* of the sentence is set by prosodic mechanisms (see Ferreira, 1993; Levelt, 1989), which use the syntactic, morphological, and phonemic content of the to-be-produced sequence, along with independent prosodic principles, to determine the “rhythm and melody” of the to-be-produced sentence (for a review, see Shattuck-Hufnagel, 2014).

Although not commonly considered in traditional theories of sentence production, function assignment

and constituent assembly must also take into account the perspective meaning that is represented as part of the multifaceted message. Perspective meaning determines what is often termed the *information structure* of sentences (Lambrecht, 1994). Perspective meaning indicates that some aspects of the event represented in the message form the focus or topic to be expressed, whereas other aspects correspond to added information or a comment on the topic. Perspective meaning also represents the *discourse* properties of an event—that the participants in a conversation already know some aspects of the event to be expressed—*given* information—whereas other information is *new* information (as relevant to the model of the listener). The topic-comment and given-new aspects represented within the perspective meaning can then determine both which grammatical functions are used to express a sentence and the order that a particular set of grammatical functions is expressed with.

There is as of yet limited evidence regarding the neural basis of syntactic encoding. Evidence suggests that a particular form of aphasia due to neurodegenerative disease, progressive nonfluent/agrammatic aphasia, is associated with syntactic deficits, and often involves thinning of ventrolateral prefrontal regions, especially the pars opercularis (Sapolsky et al., 2010; Wilson et al., 2010; for review, see Kemmerer, 2015). These observations converge with brain-imaging evidence, suggesting that sentence production (as opposed to, for example, producing word lists) shows activity in and around Broca's area (e.g., Indefrey et al., 2001; Indefrey, Hellwig, Herzog, Seitz, & Hagoort, 2004). In a functional MRI study on healthy participants producing and comprehending sentences, Segaert, Menenti, Weber, Petersson, and Hagoort (2012) observed that repetition of syntactic structure (e.g., actives, passives), both within and between production and comprehension, reduced activity in Broca's area as well as in left posterior MTG. According to Segaert et al., syntactic frames are retrieved from posterior MTG and used to construct sentences in Broca's area in production and comprehension. In section 1.1 (see also figure 4.2), we indicated that Broca's area is involved in the top-down control of word retrieval and in word-form encoding, including morphological and phonological encoding. Here, we see that Broca's area is involved in syntactic encoding as well. Tractography evidence suggests that the arcuate fasciculus mediates the mapping of syntactic information between posterior MTG and Broca's area (e.g., Wilson et al., 2011), see also figure 4.2.

Evidence from stroke patients supports the distinction between function assignment and constituent assembly (see Kemmerer, 2015, for a review). Neuropsychological

evidence points to the possibility that function assignment may specifically be impaired, as at least some patients have shown language deficits where they tend to bind the wrong referents to grammatical functions (e.g., saying *the boy hits the ball* to describe a situation where a boy is hit by a ball; e.g., Schwartz, Linebarger, & Saffran, 1985). Evidence concerning the neural locus of constituent assembly comes from Thothathiri, Schwartz, and Thompson-Schill (2010). They tested stroke patients in a task that induced interference that is highly likely localized at a positional level of processing (i.e., constituent assembly). Patients saw a pair of pictures that they named in a conjoined phrase (*The eye and the pencil*). In critical trials, one of the pictures appeared in the ordinal position opposite of where it had appeared on the preceding two trials, inducing interference in ordering—positional interference. Two patients had unusual difficulty on such interference trials, and both had damage to a small amount of brain tissue at the junction of Brodmann's areas 44 and 6. This points to the possibility that this brain region plays a critical role in the ordering process that operates at the level of positional processing.

2.2. AN EXAMPLE OF THE SENTENCE PRODUCTION STAGES IN ACTION Imagine that a speaker wished to inform his or her listener of a particular fact about the best-known comet in our solar system, leading ultimately to the utterance, *Halley's comet orbits the sun every 75 years*. This act forms the beginning of the sentence-production process, namely, forming the speaker's intent for the utterance—here, to inform the listener of a particular astronomical fact. This intent then leads sentence-production mechanisms to specify a multifaceted message consisting of semantic, relational, and perspective meaning. The semantic meaning represents that the speaker needs to describe objects and actions, including particular celestial bodies (those we refer to as *Halley's comet* and *sun*), a particular duration (75 years), and a particular action (*orbiting*). The relational meaning represents that the comet orbits the sun (and not the other way around), and that the duration refers to a property of the orbiting process. Perspective meaning represents that, for example, the identity of the celestial bodies is foregrounded and the duration is added information, which will lead to *Halley's comet orbits the sun every 75 years* rather than *Every 75 years, Halley's comet orbits the sun*.

This multifaceted message serves as input to the processes within the content and structure streams. Lemma retrieval retrieves the content words in the sentence, including the lemmas corresponding to *comet* and *sun*, representing them as singular nouns, and *orbit*,



representing it as a verb. Function assignment retrieves a grammatical subject function for *Halley's comet* and a direct object function for *the sun*, as well as an adjunct function to express the information about the duration of the orbiting action. Lemma retrieval and function assignment must coordinate to bind lemmas to their corresponding functions, so that the lemmas of *Halley's comet* are assigned to the subject function and not the direct object function. As lemmas are retrieved, word-form encoding processes can determine the morphological structure of a complex form such as *orbits*, representing that it consists of a base verb form (*orbit*) plus a singular affix (-s). In tandem, constituent assembly processes determine hierarchical relations that allow *the sun* to be the direct object of *orbit* and *Halley's comet* as the subject of the entire verb phrase *orbits the sun*, as well as (based on the represented perspective meaning) the arrangement of the adjunct phrase *every 75 years* after the main clause rather than before. As the hierarchical structure is specified, constituent assembly determines the linear sequence of constituents, which must be specified before the parallel word-form encoding processes can construct syllables (because, for example, the third syllable of *Halley's comet* may be encoded as /skɑ/, with the /s/ from *Halley's* grouped into the same syllable as the /kɑ/ from *comet*). With the constituent structure and syllabic (and phonemic) content of the sentence determined, prosodic mechanisms can then compute pitch modulations, as well as the durations of individual phonemes and pauses between phonemes to determine the ultimate form the utterance will take before it is sent to articulatory processes.

2.3. BEYOND STAGED MODELS OF SENTENCE PRODUCTION The classic staged model of sentence production has provided a valuable service to the field since the modern era of sentence-production research began (in the 1970s, mainly from work published in Fromkin, 1971; Garrett, 1975; Shattuck-Hufnagel, 1979). It has organized much of the empirical work that has been undertaken in the decades since, and aspects of the theoretical framework remain viable today. However, some recent research efforts have moved the field to questions that are to some extent orthogonal to the claims of the classic framework. Here, we describe some of these efforts.

First, one of the most prominent phenomena in the field of sentence production, often termed *syntactic priming*, is not particularly usefully explained within the classic model. Syntactic priming, first reported by Levelt and Kelter (1982) and Bock (1986b), refers to the tendency whereby speakers will produce current

sentences (often called *target* sentences) with the same syntactic structure as they encountered in previous sentences (often called *prime* sentences). For example, if a speaker hears or says a prepositional dative sentence like *the governess made [a pot of tea] [for the princess]*, she or he is likely to describe a subsequent picture with another prepositional dative like *the girl is handing [the paintbrush] [to the boy]*, but instead, if the speaker hears or says a double-object dative like *the governess made [the princess] [a pot of tea]*, she or he is likely to describe a picture also with a double object, like *the girl is handing [the boy] [a paintbrush]* (Bock, 1986b).

Syntactic priming has proven extremely useful for exploring the basis of the production of sentence structure (for reviews, see Mahowald, James, Futrell, & Gibson, 2016; Pickering & Ferreira, 2008), which is squarely within the processing domain of the structure stream we described. The difficulty is that syntactic priming effects crosscut the distinctions posited within the structural stream by the classic model. For example, syntactic alternations that unambiguously involve differences in function assignment, such as the transitive alternation (wherein the active, the agent argument is assigned the grammatical subject and the patient the direct object, but in the passive, the reverse is true), cause structural priming effects (e.g., Bock, 1986b), suggesting that priming may operate at the level of function assignment. But syntactic alternations that do not involve differences in function assignment, such as the production of the optional complementizer *that* (as in “The coach knew [that] his starting quarterback missed practice”; Ferreira, 2003), also cause structural priming, suggesting that priming also operates at the level of constituent assembly. It is possible that priming operates at both levels, but to the extent that it does, the distinction posited by the classic model is not useful. Even more seriously, some evidence fails to support a prediction of the classic model: Syntactic alternations that have the same set of grammatical functions do not prime each other relative to other structures that do have such distinctions. For example, a *shifted* prepositional dative like *The driver showed to the mechanic the torn overalls* does not prime a regular prepositional dative like *The patient showed his injury to the doctor* any more than a neutral prime sentence like *The racing driver fainted*, even though the shifted preposition has the same grammatical function assignments as the (straight) prepositional dative does (Pickering, Branigan, & McLean, 2002). This raises the question of why distinctions in grammatical function assignment sometimes cause syntactic priming (as in the transitive alternation) and sometimes do not.

In fact, syntactic priming has played a part in promoting a different emphasis within sentence production,

one that has influenced language production (and the cognitive sciences outside of language research) more broadly: A focus on how linguistic knowledge is subject to learning and tuning as a function of ongoing experience, even in adulthood (in the sentence-production domain, see Chang, Dell, & Bock, 2006; in the lexical domain, see Oppenheim, Dell, & Schwartz, 2010; in the syllabic domain, see Dell, Reed, Adams, & Meyer, 2000; for comprehension effects, see Gaskell & Dumay, 2003, Leach & Samuel, 2007, and Norris, McQueen, & Cutler, 2003). This is a second research effort that is less usefully explored within the classic sentence-production approach, as the approach does not make claims about learning and tuning effects. Instead, syntactic priming (as well as many of the other noted learning and tuning effects) has been used to develop rich computational frameworks for understanding how we acquire and then adapt the mechanisms that are used to convert multifaceted prelinguistic messages into grammatical sequences of words—frameworks that do not appeal to the distinctions made within the structural stream of the classic model (e.g., Chang et al., 2006).

The exploration of learning and tuning effects in adult language production (and comprehension) has led to a third research effort that is less usefully explored within the classic framework: An emphasis on the rational or optimal nature of linguistic behavior (see Kleinschmidt & Jaeger, 2015; Jaeger, 2010; for a broad framework for such analyses, see Clark, 2013). In an important way, this approach represents a different emphasis for developing our understanding of sentence-production phenomena (as well as cognitive phenomena more generally). In particular, the emphasis in most approaches to sentence production (including the classic approach described) has been on the representations and processes that give rise to sentence-production behavior (what is sometimes called the *algorithmic* or *representational* level of analysis; Marr, 1982). Instead, recent research efforts that focus on the rational or optimal nature of linguistic behavior more strongly emphasize the *design* properties of the system (what is sometimes called the *computational* level of analysis; Marr).

The most prominent of these rational approaches to sentence-production behavior is *uniform-information density* (Jaeger, 2010). The idea underlying uniform-information density is that each overt feature of a sentence carries more or less information for an addressee, both in terms of meaning and in terms of grammatical and relational structure. Information-theoretically, it is optimal if information is transmitted through a channel (i.e., a sentence that is spoken by a speaker for an addressee) as uniformly as possible. “Peaks” of

information—highly informative short stretches of language—can overwhelm processing systems (both production and comprehension), whereas information troughs—weakly informative longer stretches of language—underutilize the channel’s bandwidth. According to uniform-information density, the communicative process operates optimally if such peaks and troughs can be smoothed out—and so that is what the sentence-formulation process does, either by adding content to stretch an information peak, or omitting content to condense an information trough.

For example, an optional *that* in a sentence like *The coach knew (that) his starting quarterback missed practice* carries useful information in that it signals that the material following the *that* is likely to be a complete clause (*his starting quarterback missed practice*) rather than a simple noun phrase (as in *The coach knew his starting quarterback well*). This raises a puzzle: If the optional *that* carries useful information, why is it optional? According to the uniform-information density framework, more information is not always better. In particular, if properties of a sentence other than the optional *that* already successfully indicate that a complete clause is upcoming, then the *that* can be unnecessary (from an information-theoretical perspective), and can even “overload” the information-processing channel. For example, if the main verb of a sentence is often followed by a complete clause, then the *that* is not as useful, and speakers should be more likely to omit it (for evidence of just such an effect, see Ferreira & Schotter, 2013; Jaeger, 2010).

Critically, the uniform-information density analysis is (strictly speaking) agnostic with respect to the specific representations or processes that give rise to behaviors that fit with the framework (though it is more compatible with some hypothetical mechanisms than others). One possibility is that the production system explicitly represents the information value of the different features of a sentence as it is produced, and decision-making processes within sentence production consult these information values to maintain uniform-information density. Or, uniform-information behaviors may emerge more implicitly, based on correlated cognitive factors that are otherwise independently needed for production. For example, a well-documented effect in production is the *accessibility effect* (see Bock, 1982, 1986a). Words that are easier to retrieve from memory (more accessible words) tend to be positioned in sentences sooner, whereas words that are harder to retrieve from memory (less accessible words) tend to be positioned later. If higher-information words tend to be harder to retrieve from memory (a reasonable conjecture) and so will tend to be mentioned later in a

sentence, and production mechanisms include other material (such as the optional *that*) to allow such later mention (Ferreira & Dell, 2000), then the information-density effect emerges. But important for present purposes, the precise nature of the relevant cognitive mechanisms is not as critical as the design principle that is explicated by the framework in the first place. This is the sense in which rational- or optimality-motivated approaches to sentence production represent a shift in the field—but one that can complement the traditional approach rather than replace it.

In the remainder, we return to a description of the cognitive mechanisms underlying speaking and their neural basis. We discuss the relation between speech production and comprehension, self-monitoring, and the attention demands of speaking.

### 3. *Interfacing Systems*

**3.1. RELATION TO SPEECH COMPREHENSION** According to Levelt et al. (1999), lemmas are shared between production and comprehension (see figure 4.2). Lemmas were localized to the middle section of the left MTG in a meta-analysis of neuroimaging studies on spoken word production by Indefrey and Levelt (2004). This evidence from normal healthy speakers has been corroborated by evidence from voxel-based lesion-deficit analyses concerning production and comprehension deficits in individuals with poststroke aphasia. The Montreal Neurological Institute (MNI) coordinates (a common brain atlas), of the middle MTG area identified by Indefrey and Levelt, are  $y = -6$  to  $-39$  mm. We provide MNI coordinates to make clear that the “anterior” temporal area for lemmas identified by lesion-deficit analyses on production (Schwartz et al., 2009) actually overlaps with the “posterior” temporal area identified for comprehension (Dronkers, Wilkins, Van Valin, Redfern, & Jaeger, 2004). Schwartz et al. (2009) found that damage to the left middle MTG (centered at MNI  $y = -18$  mm), taken to be part of the “left anterior temporal lobe,” was most highly associated with semantic error rate in picture naming (see also Walker et al., 2011). Moreover, Baldo, Arévalo, Patterson, and Dronkers (2013) observed that performance on the Boston Naming Test was also most critically dependent on damage to the left mid-posterior MTG (centered at MNI  $y = -24$  mm). A voxel-based lesion-deficit analysis of spoken sentence comprehension by Dronkers et al. (2004) indicated that performance was most affected by damage to the left mid-posterior MTG (approximately MNI  $y = -23$  to  $-63$  mm), which they referred to as the “posterior middle temporal gyrus.” Patients with damage to this area also performed

worse on tests of single-word comprehension and picture naming than patients whose lesions spared the area. In investigating resting functional and structural connectivity of the MTG region identified by Dronkers et al. (2004), Turken and Dronkers (2011) observed that its anterior part (MNI  $y = -18$  to  $-44$  mm) showed the strongest connectivity with other parts of the sentence comprehension network. To conclude, neuroimaging evidence from healthy adult individuals as well as voxel-based lesion-deficit analyses suggests that, during comprehension and production, lemmas located in the left middle MTG mediate between conceptual, syntactic, and phonological structures in the neighboring cortex.

However, the evidence on lemmas is not universally accepted. On some accounts, there are no lemmas, and the left middle MTG is assumed to play no role in language comprehension and production (Ueno et al., 2011). According to others (Hickock & Poeppel, 2007), a lexical interface exists in the brain but it resides bilaterally in the posterior MTG and posterior inferior temporal sulcus or gyrus.

**3.2. SELF-MONITORING** During planning and articulation, a speaker monitors progress and initiates a correction if needed. Errors may not only be detected in overt speech, but also already before articulation onset, which suggests monitoring of inner speech. Internal monitoring explains why error-to-interruption times may be as short as 100–150 ms (Blackmer & Mitton, 1991), which seems too short for error detection based on auditory perception. Lackner and Tuller (1979) elicited speech errors with or without masking noise and observed that errors were detected faster in the condition with noise. This suggests that internal monitoring takes place, which leads to faster error detection than external monitoring. Using an error elicitation procedure, Motley, Camden, and Baars (1982) observed that phoneme exchanges that led to taboo words (e.g., *shit head* for *hit shed*) were made less often than normal exchanges (e.g., *bad mack* for *mad back*), but they led to an elevated galvanic skin response, suggesting that the taboo words were internally detected and suppressed.

According to a comprehension-based account, internal monitoring is accomplished by internal feedback through the speech perception system and a comparison of production and comprehension representations (Levelt, 1989; Levelt et al., 1999). In contrast, according to a production-based account, internal monitoring is based on information inside the production system, such as conflict (Nozari et al., 2011). A comprehension-based account predicts perception-specific effects, like the perceptual uniqueness-point effect, in the

monitoring of internal speech. Özdemir, Roelofs, and Levelt (2007) tested this prediction using internal phoneme monitoring and picture naming tasks. They observed an effect of the perceptual uniqueness point of a word in internal phoneme monitoring but not in picture naming, which supports comprehension-based internal monitoring. However, self-monitoring and comprehension abilities may be uncorrelated in aphasia (Nozari et al., 2011). Moreover, the abilities may double dissociate. Some aphasic patients show poor error detection in spite of intact comprehension (e.g., Marshall, Robson, Pring, & Chiat, 1998), while other patients show the reverse (Marshall, Rappaport, & Garcia-Bunuel, 1985). However, under the comprehension-based account, good comprehension is neither necessary nor sufficient for error detection. Patients may have poor comprehension due to deficient processing of external speech, which may leave internal monitoring intact (perhaps as in Marshall et al., 1985). Also, patients may have good comprehension ability, but problems with error detection because of the comparison process required (perhaps as in Marshall et al., 1998).

According to the production-based account of self-monitoring proposed by Nozari et al. (2011), errors are detected by monitoring for conflict at lemma and phoneme levels within the production system. The conflict monitor is thought to be located in the anterior cingulate cortex, following a prominent theory about human performance monitoring of Botvinick, Braver, Barch, Carter, and Cohen (2001). Yeung, Botvinick, and Cohen (2004) provided evidence from a flanker task with manual responding that the N2 and error-related negativity (ERN) components of the event-related brain potential reflect response conflict and arise in the anterior cingulate cortex. Nozari et al. showed that their proposal explains the double dissociation between comprehension and self-monitoring in aphasia, and in particular, the observation that production but not comprehension parameters predict monitoring performance in patients. However, the account is challenged by neuroimaging evidence that anterior cingulate cortex activity reflects regulation rather than conflict monitoring (e.g., Aarts, Roelofs, & Van Turennout, 2008; Roelofs, Van Turennout, & Coles, 2006), and evidence that the magnitudes of the ERN and N2 are inversely related to the amount of conflict in manual responding (Burle, Roger, Allain, Vidal & Hasbroucq, 2008) and picture naming (Shao, Roelofs, Acheson, & Meyer, 2014).

**3.3. ATTENTION DEMANDS** Although speaking is one of our most highly exercised psychomotor skills, evidence suggests that it can have detrimental effects on unrelated tasks such as driving (Kubose et al., 2006)

and vice versa (Strayer & Drews, 2007). In particular, the selection of lemmas and lexical forms precludes simultaneous selection of responses in an unrelated task (Ferreira & Pashler, 2002). These findings suggest that speaking does not happen completely automatically, but requires some form of attention.

According to an influential theory proposed by Posner and colleagues (Posner, 2012, for a recent review), attention consists of alerting, orienting, and executive control. Alerting is the ability to achieve and maintain alertness, either briefly or sustained over an extended period of time. Orienting is the ability to shift the locus of processing toward a particular spatial position, either with or without corresponding eye movements. Executive control is the ability to regulate other processes to remain goal directed in the face of distraction. According to Miyake and colleagues (Miyake et al., 2000), executive control can be decomposed into updating (the ability to actively manipulate the contents of working memory), inhibiting (the ability to lower the activation of unwanted responses), and shifting (the ability to rapidly switch back and forth between tasks or mental sets). Evidence suggests that all these attentional abilities contribute to speaking. Speakers have to maintain alertness in producing utterances, and as a consequence, their sustained attention ability influences the speed of naming and describing pictures (Jongman, Roelofs, & Meyer, 2015). Also, speakers need to orient to perceptual information until the phonological form of the corresponding utterance has been encoded (e.g., Meyer, Sleiderink, & Levelt, 1998). For example, using eye tracking, it has been observed that speakers gaze longer at pictures with disyllabic than monosyllabic names (Meyer, Roelofs, & Levelt, 2003). Speakers must update the contents of working memory while engaging in conceptual and linguistic processes, and as a consequence, updating ability determines the speed of picture naming (Piai & Roelofs, 2013; Shao, Roelofs, & Meyer, 2012) and picture description (Sikora, Roelofs, Hermans, & Knoors, 2016). Moreover, the inhibiting ability is required to suppress incorrect names that are coactivated during lexical selection (Shao et al., 2014). Furthermore, the shifting ability is needed to switch between planning one type of phrase to another (Sikora, Roelofs, Hermans, & Knoors, 2016). Evidence from event-related brain potentials suggests that the shifting and inhibiting abilities are engaged before the updating ability in speaking (Sikora, Roelofs, & Hermans, 2016).

#### *4. Conclusion*

In the 1970s and 1980s, accounts of the cognitive architecture of speaking (i.e., Garrett, 1975; Levelt, 1989)



were mainly based on behavioral evidence, including analyses of speech errors and reaction times. During the past three decades, new accounts have been advanced building on this classic work. These accounts synthesize evidence from reaction times and errors as well as electrophysiological and hemodynamic neuroimaging, tractography, and lesion-deficit analyses. However, many controversies endure and there remains a lot to be learned. Whereas much is known about the neural basis of word production, comparable evidence on sentence production is lacking. Correspondingly, whereas comprehensive computationally implemented accounts of word production have been developed, comparable accounts of sentence production are lacking. We expect future cognitive neuroscience research and modeling to make progress in these areas.

#### REFERENCES

- Aarts, E., Roelofs, A., & Van Turennout, M. (2008). Anticipatory activity in anterior cingulate cortex can be independent of conflict and error likelihood. *Journal of Neuroscience*, *28*, 4671–4678.
- Abdel Rahman, R., & Melinger, A. (2009). Semantic context effects in language production: A swinging lexical network proposal and a review. *Language and Cognitive Processes*, *24*, 713–734.
- Baldo, J. V., Arévalo, A., Patterson, J. P., & Dronkers, N. F. (2013). Grey and white matter correlates of picture naming: Evidence from a voxel-based lesion analysis of the Boston Naming Test. *Cortex*, *49*, 658–667.
- Bierwisch, M., & Schreuder, R. (1992). From concepts to lexical items. *Cognition*, *42*, 23–60.
- Blackmer, E. R., & Mitton, J. L. (1991). Theories of monitoring and the timing of repairs in spontaneous speech. *Cognition*, *39*, 173–194.
- Bock, J. K. (1982). Toward a cognitive psychology of syntax: Information processing contributions to sentence formulation. *Psychological Review*, *89*, 1–47.
- Bock, J. K. (1986a). Meaning, sound, and syntax: Lexical priming in sentence production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *12*, 575–586.
- Bock, J. K. (1986b). Syntactic persistence in language production. *Cognitive Psychology*, *18*, 355–387.
- Bock, J. K. (1995). Sentence production: From mind to mouth. In J. L. Miller & P. D. Eimas (Eds.), *Handbook of perception and cognition. Vol II. Speech, language, and communication* (pp. 181–216). Orlando, FL: Academic Press.
- Bock, J. K., & Ferreira, V. S. (2014). Syntactically speaking. In M. Goldrick, V. Ferreira, & M. Miozzo (Eds.), *The Oxford handbook of language production* (pp. 21–46). Oxford: Oxford University Press.
- Bock, J. K., & Levelt, W. J. M. (1994). Language production: Grammatical encoding. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 945–984). San Diego: Academic Press.
- Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological Review*, *108*, 624–652.
- Brauer, J., Anwander, A., Perani, D., & Friederici, A. D. (2013). Dorsal and ventral pathways in language development. *Brain and Language*, *127*, 289–295.
- Bürki, A., Pellet, P., & Laganaro, M. (2015). Do speakers have access to a mental syllabary? ERP comparison of high frequency and novel syllable production. *Brain and Language*, *150*, 90–102.
- Burle, B., Roger, C., Allain, S., Vidal, F., & Hasbroucq, T. (2008). Error negativity does not reflect conflict: A reappraisal of conflict monitoring and anterior cingulate activity. *Journal of Cognitive Neuroscience*, *20*, 1637–1655.
- 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*, 309–343.
- Caramazza, A., & Miozzo, M. (1998). More is not always better: A response to Roelofs, Meyer, and Levelt. *Cognition*, *69*, 231–241.
- Chang, F., Dell, G. S., & Bock, K. (2006). Becoming syntactic. *Psychological Review*, *113*, 234–272.
- Chang, F., & Fitz, H. (2014). Computational models of sentence production: A dual-path approach. In M. Goldrick, V. Ferreira, & M. Miozzo (Eds.), *The Oxford handbook of language production* (pp. 70–87). Oxford: Oxford University Press.
- Cholin, J., Dell, G. S., & Levelt, W. J. M. (2011). Planning and articulation in incremental word production: Syllable-frequency effects in English. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*, 109–122.
- Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, *36*, 181–204.
- Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review*, *93*, 283–321.
- Dell, G. S., Reed, K. D., Adams, D. R., & Meyer, A. S. (2000). Speech errors, phonotactic constraints, and implicit learning: A study of the role of experience in language production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*, 1355–1367.
- 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*, 801–838.
- Dell, G. S., Schwartz, M. F., Nozari, N., Faseyitan, O., & Coslett, H. B. (2013). Voxel-based lesion-parameter mapping: Identifying the neural correlates of a computational model of word production. *Cognition*, *128*, 380–396.
- Dhooge, E., & Hartsuiker, R. J. (2010). The distractor frequency effect in picture-word interference: Evidence for response exclusion. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*, 878–891.
- Dronkers, N. F., Wilkins, D. P., Van Valin, R. D., Redfern, B. B., & Jaeger, J. J. (2004). Lesion analysis of the brain areas involved in language comprehension. *Cognition*, *92*, 145–177.
- Eichenbaum, H. (2012). *The cognitive neuroscience of memory: An introduction* (2nd ed.). Oxford: Oxford University Press.
- Fernandez-Miranda, J. C., Wang, Y., Pathak, S., Stefaneau, L., Verstynen, T., & Yeh, F.-C. (2015). Asymmetry, connectivity, and segmentation of the arcuate fascicle in the human brain. *Brain Structure and Function*, *220*, 1665–1680.
- Ferreira, F. (1993). Creation of prosody during sentence production. *Psychological Review*, *100*, 233–253.

- Ferreira, V. S. (2003). The persistence of optional complementizer mention: Why saying a “that” is not saying “that” at all. *Journal of Memory and Language*, *48*, 379–398.
- Ferreira, V. S. (2010). Language production. *Wiley Interdisciplinary Reviews: Cognitive Science*, *1*, 834–844.
- Ferreira, V. S., & Dell, G. S. (2000). Effect of ambiguity and lexical availability on syntactic and lexical production. *Cognitive Psychology*, *40*, 296–340.
- Ferreira, V. S., & Pashler, H. (2002). Central bottleneck influences on the processing stages of word production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*, 1187–1199.
- Ferreira, V. S., & Schotter, E. R. (2013). Do verb bias effects on sentence production reflect sensitivity to comprehension or production factors? *Quarterly Journal of Experimental Psychology*, *66*, 1548–1571.
- Ferreira, V. S., & Slevc, L. R. (2007). Grammatical encoding. In G. Gaskell (Ed.), *Oxford handbook of psycholinguistics* (pp. 453–470). Oxford: Oxford University Press.
- Finkbeiner, M., & Caramazza, A. (2006). Now you see it, now you don't: On turning semantic interference into facilitation in a Stroop-like task. *Cortex*, *42*, 790–796.
- Fromkin, V. A. (1971). The non-anomalous nature of anomalous utterances. *Language*, *47*, 27–52.
- Garrett, M. F. (1975). The analysis of sentence production. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 9, pp. 133–177). New York: Academic Press.
- Garrett, M. F. (1980). Levels of processing in sentence production. In B. Butterworth (Ed.), *Language production* (Vol. 1, pp. 177–220). London: Academic Press.
- Garrett, M. F. (1982). Production of speech: Observations from normal and pathological language use. In A. Ellis (Ed.), *Normality and pathology in cognitive functions* (pp. 19–76). London: Academic Press.
- Garrett, M. F. (1988). Processes in language production. In F. J. Newmeyer (Ed.), *Linguistics: The Cambridge survey* (Vol. 3, pp. 69–96). Cambridge, MA: Harvard University Press.
- Gaskell, M. G., & Dumay, N. (2003). Lexical competition and the acquisition of novel words. *Cognition*, *89*, 105–132.
- Geschwind, N. (1970). The organization of language and the brain. *Science*, *170*, 940–944.
- Glasser, M. F., & Rilling, J. K. (2008). DTI tractography of the human brain's language pathways. *Cerebral Cortex*, *18*, 2471–2482.
- Guenther, F. H., & Vladusich, T. (2012). A neural theory of speech acquisition and production. *Journal of Neurolinguistics*, *25*, 408–422.
- Hagoort, P., & Indefrey, P. (2014). The neurobiology of language beyond single words. *Annual Review of Neuroscience*, *37*, 347–362.
- Haller, S., Radue, E. W., Erb, M., Grodd, W., & Kircher, T. (2005). Overt sentence production in event-related fMRI. *Neuropsychologia*, *43*, 807–814.
- Hickok, G. (2012). Computational neuroanatomy of speech production. *Nature Reviews Neuroscience*, *13*, 135–145.
- Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience*, *8*, 393–402.
- Indefrey, P. (2011). The spatial and temporal signatures of word production components: A critical update. *Frontiers in Psychology*, *2*, article 255.
- Indefrey, P. (2016). On putative shortcomings and dangerous avenues: Response to Strijkers & Costa. *Language, Cognition and Neuroscience*, *31*, 517–520.
- Indefrey, P., Brown, C. M., Hellwig, F., Amunts, K., Herzog, H., Seitz, R. J., & Hagoort, P. (2001). A neural correlate of syntactic encoding during speech production. *Proceedings of the National Academy of Sciences*, *98*, 5933–5936.
- Indefrey, P., Hellwig, F., Herzog, H., Seitz, R. J., & Hagoort, P. (2004). Neural responses to the production and comprehension of syntax in identical utterances. *Brain and Language*, *89*, 312–319.
- Indefrey, P., & Levelt, W. J. M. (2004). The spatial and temporal signatures of word production components. *Cognition*, *92*, 101–144.
- Jaeger, T. F. (2010). Redundancy and reduction: Speakers manage syntactic information density. *Cognitive Psychology*, *61*, 23–62.
- Jongman, S. R., Roelofs, A., & Meyer, A. S. (2015). Sustained attention in language production: An individual differences investigation. *Quarterly Journal of Experimental Psychology*, *68*, 710–730.
- Kemmerer, D. (2015). *Cognitive neuroscience of language*. New York: Psychology Press.
- Kleinschmidt, D. F., & Jaeger, T. F. (2015). Robust speech perception: Recognize the familiar, generalize to the similar, and adapt to the novel. *Psychological Review*, *122*, 148–203.
- Konopka, A. E., & Brown-Schmidt, S. (2014). Message encoding. In M. Goldrick, V. Ferreira, & M. Miozzo (Eds.), *The Oxford handbook of language production* (pp. 3–20). Oxford: Oxford University Press.
- Kubose, T. T., Bock, K., Dell, G. S., Garnsey, S. M., Kramer, A. F., & Mayhugh, J. (2006). The effects of speech production and speech comprehension on simulated driving performance. *Applied Cognitive Psychology*, *20*, 43–63.
- Lackner, J. R., & Tuller, B. H. (1979). Role of efference monitoring in the detection of self-produced speech errors. In W. E. Cooper & E. C. T. Walker (Eds.), *Sentence processing* (pp. 281–294). Hillsdale, NJ: Erlbaum.
- Lambon Ralph, M. A. (2014). Neurocognitive insights on conceptual knowledge and its breakdown. *Philosophical Transactions of the Royal Society B*, *369*, 20120392.
- Lambrecht, K. (1994). *Information structure and sentence form*. Cambridge: Cambridge University Press.
- Leach, L., & Samuel, A. G. (2007). Lexical configuration and lexical engagement: When adults learn new words. *Cognitive Psychology*, *55*, 306–353.
- Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.
- Levelt, W. J. M. (2013). *A history of psycholinguistics: The pre-Chomskyan era*. Oxford: Oxford University Press.
- Levelt, W. J. M., & Kelter, S. (1982). Surface form and memory in question answering. *Cognitive Psychology*, *14*, 78–106.
- Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, *22*, 1–38.
- Lichtheim, L. (1885). On aphasia. *Brain*, *7*, 433–484.
- Mahowald, K., James, A., Futrell, R., & Gibson, E. (2016). A meta-analysis of syntactic priming in language production. *Journal of Memory and Language*, *91*, 5–27.
- Marchina, S., Zhu, L. L., Norton, A., Zipse, L., Wan, C. Y., & Schlaug, G. (2011). Impairment of speech production predicted by lesion load of the left arcuate fasciculus. *Stroke*, *42*, 2251–2256.
- Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. San Francisco: Freeman & Co.

- Marshall, J., Robson, J., Pring, T., & Chiat, S. (1998). Why does monitoring fail in jargon aphasia? Comprehension, judgment, and therapy evidence. *Brain and Language*, *63*, 79–107.
- Marshall, R. C., Rappaport, B. Z., & Garcia-Bunuel, L. (1985). Self-monitoring behavior in a case of severe auditory agnosia with aphasia. *Brain and Language*, *24*, 297–313.
- Meyer, A. S., Roelofs, A., & Levelt, W. J. M. (2003). Word length effects in object naming: The role of a response criterion. *Journal of Memory and Language*, *48*, 131–147.
- 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.
- Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. (2000). The unity and diversity of executive functions and their contributions to complex “frontal lobe” tasks: A latent variable analysis. *Cognitive Psychology*, *41*, 49–100.
- Morsella, E., & Miozzo, M. (2002). Evidence for a cascade model of lexical access in speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*, 555–563.
- Motley, M. T., Camden, C. T., & Baars, B. J. (1982). Covert formulation and editing of anomalies in speech production: Evidence from experimentally elicited slips of the tongue. *Journal of Verbal Learning and Verbal Behavior*, *21*, 578–594.
- Munding, D., Dubarry, A.-S., & Alario, F.-X. (2016). On the cortical dynamics of word production: A review of the MEG evidence. *Language, Cognition and Neuroscience*, *31*, 441–462.
- Navarrete, E., Del Prato, P., Peressotti, F., & Mahon, B. Z. (2014). Lexical selection is not by competition: Evidence from the blocked naming paradigm. *Journal of Memory and Language*, *76*, 253–272.
- Norris, D., McQueen, J. M., & Cutler, A. (2003). Perceptual learning in speech. *Cognitive Psychology*, *47*, 204–238.
- Nozari, N., Dell, G. S., & Schwartz, M. F. (2011). Is comprehension necessary for error detection? A conflict-based account of monitoring in speech production. *Cognitive Psychology*, *63*, 1–33.
- Oppenheim, G. M., Dell, G. S., & Schwartz, M. F. (2010). The dark side of incremental learning: A model of cumulative semantic interference during lexical access in speech production. *Cognition*, *114*, 227–252.
- Özdemir, R., Roelofs, A., & Levelt, W. J. M. (2007). Perceptual uniqueness point effects in monitoring internal speech. *Cognition*, *105*, 457–465.
- Patterson, K., Nestor, P. J., & Rogers, T. T. (2007). Where do you know what you know? The representation of semantic knowledge in the human brain. *Nature Reviews Neuroscience*, *8*, 976–987.
- Piai, V., & Roelofs, A. (2013). Working memory capacity and dual-task interference in picture naming. *Acta Psychologica*, *142*, 332–342.
- Piai, V., Roelofs, A., Jensen, O., Schoffelen, J.-M., & Bonfond, M. (2014). Distinct patterns of brain activity characterise lexical activation and competition in spoken word production. *PLOS ONE*, *9*, e88674.
- Pickering, M. J., Branigan, H. P., & McLean, J. F. (2002). Constituent structure is formulated in one stage. *Journal of Memory and Language*, *46*, 586–605.
- Pickering, M. J., & Ferreira, V. S. (2008). Structural priming: A critical review. *Psychological Bulletin*, *134*, 427–459.
- Posner, M. I. (2012). *Attention in a social world*. Oxford: Oxford University Press.
- Rapp, B., & Goldrick, M. (2000). Discreteness and interactivity in spoken word production. *Psychological Review*, *107*, 460–499.
- Rapp, B., & Goldrick, M. (2004). Feedback by any other name is still interactivity: A reply to Roelofs’ comment on Rapp & Goldrick (2000). *Psychological Review*, *111*, 573–578.
- Roelofs, A. (1992). A spreading-activation theory of lemma retrieval in speaking. *Cognition*, *42*, 107–142.
- Roelofs, A. (2004a). Comprehension-based versus production-internal feedback in planning spoken words: A rejoinder to Rapp and Goldrick (2004). *Psychological Review*, *111*, 579–580.
- Roelofs, A. (2004b). Error biases in spoken word planning and monitoring by aphasic and nonaphasic speakers: Comment on Rapp and Goldrick (2000). *Psychological Review*, *111*, 561–572.
- Roelofs, A. (2008). Tracing attention and the activation flow in spoken word planning using eye movements. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*, 353–368.
- Roelofs, A. (2014). A dorsal-pathway account of aphasic language production: The WEAVER++/ARC model. *Cortex*, *59*, 33–48.
- 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*, 219–230.
- Roelofs, A., Van Turenout, M., & Coles, M. G. H. (2006). Anterior cingulate cortex activity can be independent of response conflict in Stroop-like tasks. *Proceedings of the National Academy of Sciences USA*, *103*, 13884–13889.
- Sahin, N. T., Pinker, S., Cash, S. S., Schomer, D., & Halgren, E. (2009). Sequential processing of lexical, grammatical, and phonological information within Broca’s area. *Science*, *326*, 445–449.
- Sapolsky, D., Bakkour, A., Negreira, A., Nalipinski, P., Weintraub, S., Mesulam, M. M., ... Dickerson, B. C. (2010). Cortical neuroanatomic correlates of symptom severity in primary progressive aphasia. *Neurology*, *75*, 358–366.
- Schwartz, M. F., Faseyitan, O., Kim, J., & Coslett, H. B. (2012). The dorsal stream contribution to phonological retrieval in object naming. *Brain*, *135*, 3799–3814.
- Schwartz, M. F., Kimberg, D. Y., Walker, G. M., Faseyitan, O., Brecher, A., Dell, G. S., & Coslett, H. B. (2009). Anterior temporal involvement in semantic word retrieval: Voxel-based lesion-symptom mapping evidence from aphasia. *Brain*, *132*, 3411–3427.
- Schwartz, M. F., Linebarger, M. C., & Saffran, E. M. (1985). The status of the syntactic deficit theory of agrammatism. In M. L. Kean (Ed.) *Agrammatism* (pp. 83–124). New York: Academic Press.
- Segaert, K., Menenti, L., Weber, K., Petersson, K. M., & Hagoort, P. (2012). Shared syntax in language production and language comprehension—an fMRI study. *Cerebral Cortex*, *22*, 1662–1670.
- Shallice, T., & Cooper, R. (2011). *The organisation of mind*. Oxford: Oxford University Press.
- Shao, Z., Roelofs, A., Acheson, D. J., & Meyer, A. S. (2014). Electrophysiological evidence that inhibition supports lexical selection in picture naming. *Brain Research*, *1586*, 130–142.
- Shao, Z., Roelofs, A., & Meyer, A. S. (2012). Sources of individual differences in the speed of naming objects and



- actions: The contribution of executive control. *Quarterly Journal of Experimental Psychology*, 65, 1927–1944.
- Shattuck-Hufnagel, S. (1979). Speech errors as evidence for a serial-ordering mechanism in sentence production. In W. E. Cooper & E. C. T. Walker (Eds.), *Sentence processing: Psycholinguistic studies presented to Merrill Garrett* (pp. 295–342). Hillsdale, NJ: Erlbaum.
- Shattuck-Hufnagel, S. (2014). Phrase-level phonological and phonetic phenomena. In M. Goldrick, V. Ferreira, & M. Miozzo (Eds.), *The Oxford handbook of language production* (pp. 259–274). Oxford: Oxford University Press.
- Sikora, K., Roelofs, A., & Hermans, D. (2016). Electrophysiology of executive control in spoken noun-phrase production: Dynamics of updating, inhibiting, and shifting. *Neuropsychologia*, 84, 44–53.
- Sikora, K., Roelofs, A., Hermans, D., & Knoors, H. (2016). Executive control in spoken noun-phrase production: Contributions of updating, inhibiting, and shifting. *Quarterly Journal of Experimental Psychology*, 69, 1719–1740.
- Stemberger, J. P. (1985). An interactive activation model of language production. In A. Ellis (Ed.), *Progress in the psychology of language* (Vol. 1, pp. 143–186). London: Erlbaum.
- Strayer, D. L., & Drews, F. A. (2007). Cell-phone-induced driver distraction. *Current Directions in Psychological Science*, 16, 128–131.
- Strijkers, K., & Costa, A. (2016). The cortical dynamics of speaking: Present shortcomings and future avenues. *Language, Cognition and Neuroscience*, 31, 484–503.
- Thothathiri, M., Schwartz, M. F., & Thompson-Schill, S. L. (2010). Selection for position: The role of left ventrolateral prefrontal cortex in sequencing language. *Brain and Language*, 113, 28–38.
- Turken, A. U., & Dronkers, N. F. (2011). The neural architecture of the language comprehension network: Converging evidence from lesion and connectivity analyses. *Frontiers in Systems Neuroscience*, 5, article 1.
- Ueno, T., Saito, S., Rogers, T. T., & Lambon Ralph, M. A. (2011). Lichtheim 2: Synthesizing aphasia and the neural basis of language in a neurocomputational model of the dual dorsal-ventral language pathways. *Neuron*, 72, 385–396.
- Ullman, M. T. (2004). Contributions of memory circuits to language: The declarative/procedural model. *Cognition*, 92, 231–270.
- Walker, G. M., Schwartz, M. F., Kimberg, D. Y., Faseyitan, O., Brecher, A., Dell, G. S., & Coslett, H. B. (2011). Support for anterior temporal involvement in semantic error production in aphasia: New evidence from VLSM. *Brain and Language*, 117, 110–122.
- Wang, J., Marchina, S., Norton, A. C., Wan, C. Y., & Schlaug, G. (2013). Predicting speech fluency and naming abilities in aphasic patients. *Frontiers in Human Neuroscience*, 7, article 831.
- Wernicke, C. (1874). *Der aphasische Symptomencomplex. Eine psychologische Studie auf anatomischer Basis*. Breslau, Germany: Cohn und Weigert.
- Wilson, S. M., Galantucci, S., Tartaglia, M. C., Rising, K., Patterson, D. K., Henry, M. L., ... Gorno-Tempini, M. L. (2011). Syntactic processing depends on dorsal language tracts. *Neuron*, 72, 397–403.
- Wilson, S. M., Henry, M. L., Besbris, M., Ogar, J. M., Dronkers, N. F., Jarrold, W., ... Gorno-Tempini, M. L. (2010). Connected speech production in three variants of primary progressive aphasia. *Brain*, 133, 2069–2088.
- Wundt, W. (1900). *Die Sprache*. Leipzig: Verlag von Wilhelm Engelmann.
- Yeung, N., Botvinick, M. M., & Cohen, J. D. (2004). The neural basis of error detection: Conflict monitoring and the error-related negativity. *Psychological Review*, 111, 931–959.