

INTRODUCTION



Thirty years of *Speaking*: An introduction to the Special Issue

Antje S. Meyer^a, Ardi Roelofs^b and Laurel Brehm^a

^aMax Planck Institute for Psycholinguistics, Nijmegen, The Netherlands; ^bDonders Institute for Brain, Cognition and Behaviour, Centre for Cognition, Radboud University, Nijmegen, Netherland

ABSTRACT

Thirty years ago, Pim Levelt published *Speaking*. During the 10th International Workshop on Language Production held at the Max Planck Institute for Psycholinguistics in Nijmegen in July 2018, researchers reflected on the impact of the book in the field, developments since its publication, and current research trends. The contributions in this Special Issue are closely related to the presentations given at the workshop. In this editorial, we sketch the research agenda set by *Speaking*, review how different aspects of this agenda are taken up in the papers in this volume and outline directions for further research.

ARTICLE HISTORY

Received 24 July 2019
Accepted 29 July 2019

1. Introduction

1.1 Overview

In 1989, Pim Levelt published his book *Speaking - From Intention to Articulation*. In this book he reviewed what psycholinguists knew about speaking and, far more importantly, presented his own comprehensive model of speaking: the Blueprint for the speaker. Garrett (1990) presciently predicted “This work will serve the field of language research extremely well in the coming decade” (p. 290). Garrett’s prediction was borne out: *Speaking* has shaped the research agendas in many labs, and has done so for much longer than a decade, continuing its influence and relevance for the last 30 years. To illustrate, *Speaking* has been cited more than 60,000 times in the Web of Science catalogue, and more than 3000 times in the last five years alone. A glance at these citations shows that *Speaking* has been cited by scholars on all continents apart from Antarctica and in a wide range of contexts. It has had a lasting impact across psychology, influencing fields as diverse as psycholinguistics, cognitive psychology, neuroscience, developmental psychology and social psychology, a lasting impact on linguistics, influencing phonology, semantics, syntax, morphology, prosody and pragmatics, as well as a lasting impact on applied or translational work on second language acquisition, bilingualism, reading, education and speech disorders. This is because the Blueprint is broad in scope, encompassing the components of speaking from intention to articulation, as the title of the book says, but it is also specific, offering testable hypotheses about many of the

representations and processes involved in speaking, with, as Garrett (1990) says, “exceptional breadth and theoretical utility” (p. 273).

In July 2018, the 10th International Workshop on Language Production was held at the Max Planck Institute for Psycholinguistics in Nijmegen. The aim of the workshop was to celebrate *Speaking*, honour its author and survey how language production research had developed over the past 30 years. Among our distinguished guests were, in addition to Pim Levelt himself, many of his closest colleagues and pioneers in the field: Kay Bock, Herb and Eve Clark, Gary Dell, Merrill Garrett, Stephanie Shattuck-Hufnagel and Sieb Nooteboom. It was a privilege to host this festive and stimulating event.

The keynote speakers of the workshop were invited to contribute a paper to the current Special Issue of *Language, Cognition and Neuroscience*, which appears 30 years after the publication of *Speaking*. In this editorial, we provide a brief sketch of the Blueprint and mention some of the developments in the broader field of cognitive (neuro)science that facilitated testing and refining Levelt’s model, before providing an overview of the contributions in the volume and proposing some avenues for future work.

1.2. Assumptions and architecture of *Speaking*

The architecture of Levelt’s Blueprint for the speaker is shown in Figure 1. The main processing components are the conceptualizer, which generates a preverbal message, the formulator, which generates a phonetic plan for the utterance and the articulator, which generates

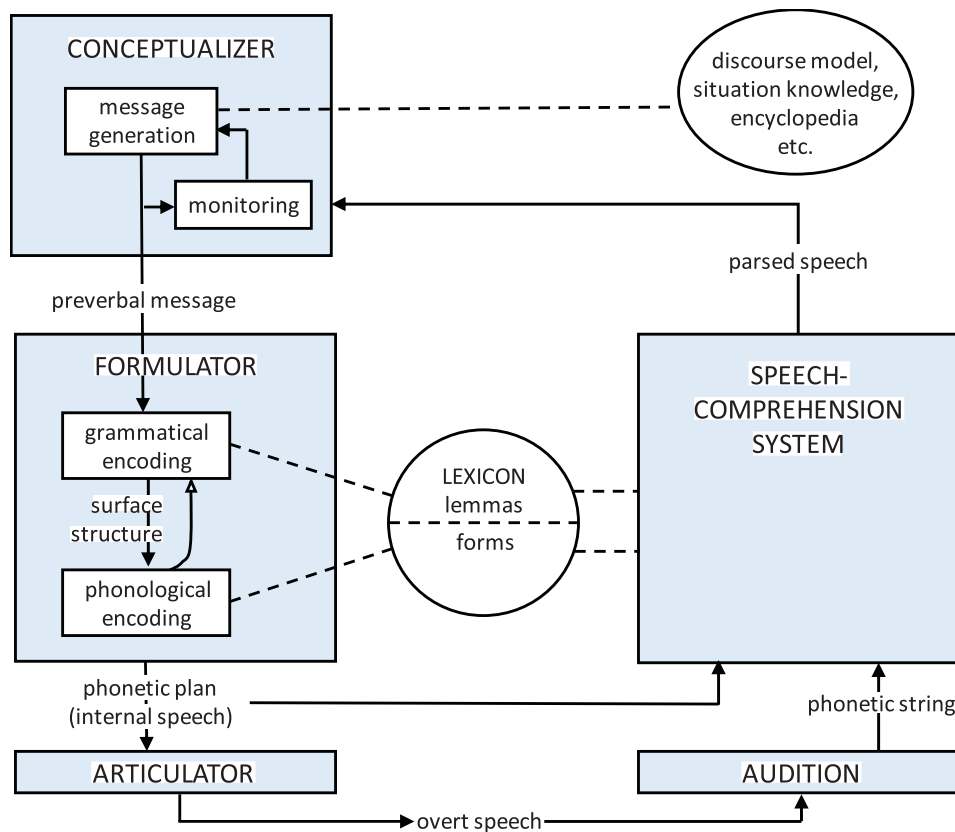


Figure 1. A blueprint for the speaker. Boxes represent processing components; circle and ellipse represent knowledge stores (After Figure 1.1., p. 9 in Levelt, 1989).

overt speech. The conceptualizer's activities are informed by the speaker's general world knowledge and awareness of the current communicative situation. The formulator receives input from the lexicon, which features separate representations of the semantic and syntactic properties of words (their lemmas) and their forms. The articulator temporarily stores pieces of phonetic plans as they are incrementally created, and generates the correct sequence to implement the plan. This whole speech production system is linked to the comprehension system via two feedback loops involving the phonetic plan and the overt speech, respectively, and via access to the mental lexicon, which is shared between comprehension and production. The architecture assumes independence: levels are separated and accessed serially without interaction, and incrementality: pieces of an utterance can move forward through the system once they are ready to be planned.

In part, the immense impact of the Blueprint in the field may stem from its intuitive appeal. The model simply makes sense. It is sensible that there is a distinction between thinking and speaking, which means that there must be processes to turn thoughts into representations that can instruct the articulators, and these processes must be fed by knowledge of the language, such

as described in a lexicon. That speakers can hear themselves, thereby engaging the comprehension system, is also quite sensible. This leads to the lasting impact of *Speaking*: By endorsing plausible assumptions about the properties of the language system and making their architecture explicit, the Blueprint offers a solid basis for theoretical and empirical work, enabling a set of challenges and an agenda for the research community. This is evident from the substantial body of work testing the key assumptions laid out in *Speaking*. For example, the modular architecture of the Blueprint isn't the only conceivable architecture of the mind: an alternate proposal allows more interactivity (e.g. Dell, 1986; Rapp & Goldrick, 2000). Likewise, the proposal that lemmas are major processing units has been challenged (e.g. Caramazza, 1997). In both cases, years of lively debate have helped the field advance, refining our knowledge about what is essential for the language system.

1.3. Methodological developments in cognitive (neuro)science

Speaking makes strong assumptions about the information flow and sequence of processes involved in

utterance planning. In the 1980s, these assumptions were hard to assess. In fact, the framework laid out in *Speaking* was almost exclusively based on theoretical considerations, evidence from analyses of speech errors and pauses and simple elicitation experiments. Chronometric work on speech planning had only just begun (e.g. Glaser & Dünghoff, 1984; Oldfield & Wingfield, 1965). In the following decades, many new research tools were developed in psychology, cognitive science and cognitive neuroscience, which also became widely used in studies of speaking (see de Groot & Hagoort, 2018, for a comprehensive review). Many fruitful paradigms have also been developed specifically for studying speaking. The use of these tools is amply illustrated in the current volume. They include:

*Development of chronometric paradigms that are suitable for assessing the timing of each stage of production. The main workhorse has been picture naming, often used in conjunction with interference paradigms. Pioneering work was done by Glaser and Dünghoff (1984; see also Schriefers, Meyer, & Levelt, 1990). Abdel Rahman and Melinger's article in this volume offers a review of interference studies examining lexical selection.

*Eye-tracking. Following developments in research on reading and spoken language comprehension, eye tracking became widely used to study the mapping of visual information onto conceptual and linguistic representations, shedding light on conceptualisation and linearisation. Papafragou and Grigoroğlu's article (this volume) shows how this paradigm can be used to study talking about the manner and path of events. Ferreira and Rehrig (this volume) demonstrate how it can be brought to a new level of sophistication by using photographs of real scenes, rather than line drawings, and state-of-the-art analyses to link overt attention and speech planning.

*Neurobiological methods. When *Speaking* was written, there was only very basic knowledge of the brain circuits involved in speaking, mostly coming from aphasia. Electrophysiological and hemodynamic neuroimaging tools for use in psycholinguistics had only just begun to be explored, primarily in reading and auditory language comprehension (e.g. Kutas & Hillyard, 1980; Petersen, Fox, Posner, Mintun, & Raichle, 1988). Early EEG studies avoided overt speech because of movement artefacts (e.g. van Turennout, Hagoort, & Brown, 1998). Now innovations in recording and analysis techniques can help researchers account for movement artefacts, and language production can be examined with techniques such as fMRI, MEG, EEG and ECoG (for reviews, see de Zubicaray & Piai, 2019; Kemmerer, 2014). Moreover, the impact of brain damage can be studied in much detail by voxel-based lesion-symptom mapping.

In this volume, the benefit of using these tools is illustrated in Kemmerer's article discussing the neurobiological support for the Blueprint and in Kearny and Guenther's article on the DIVA model.

*Computational methods. Advances in computational power have impacted all areas of production research. Data analysis has moved from classical analyses of variance (often featuring $\text{min}F'$, Clark, 1973) to mixed effect modelling (e.g. Baayen, Davidson, & Bates, 2008; Jaeger, 2008) and Bayesian analyses (e.g. Vasisht, Nicenboim, Beckman, Li, & Kong, 2018).

Statistical and computational modelling have improved our understanding of many facets of production. The Blueprint drew heavily upon analyses of speech errors; new methods have allowed the field to refine our understanding of the elicitation of errors in patients (e.g. Dell, Schwartz, Martin, Saffran, & Gagnon, 1997; Schwartz & Dell, 2010) and healthy speakers (e.g. work by Dell and colleagues in the current volume). Improved computational power has also afforded statistical modelling on large samples of data, which has provided further insights on the relationship between language experience and production (e.g. Jaeger, 2010; MacDonald, 2013). In this volume, Kempen and Harbusch illustrate how new statistical techniques advance corpus analysis.

Finally, researchers have used computational modelling to bring together behavioural, neurobiological and computational research in connectionist (e.g. Chang, Dell, & Bock, 2006; Dell, 1986; Levelt, Roelofs, & Meyer, 1999; Roelofs, 1992, 2014) and Bayesian models (e.g. Frank & Goodman, 2012). In this volume, the virtues of computational modelling are illustrated in the articles by Ferreira and Rehrig, Kearney and Guenther, and Dell and colleagues.

With the development of these tools, researchers have addressed many of the questions raised in *Speaking*. In the following sections, we review which issues were taken up in the articles in this volume.

2. Overview of the volume

2.1. Kemmerer: *Placing the Blueprint in the brain* (*Speaking Chapter 1*)

As reviewed above, little was known about the way the speech production system is represented in the brain when *Speaking* was written. This has changed dramatically with the advent and development of new neurobiological methods. *Speaking* makes strong claims about the processing components involved in speaking and the time course of their engagement. The challenge for the research community is to assess whether these components can indeed be separated from each other, and,

if so, whether they are orchestrated in the way proposed in the Blueprint. Kemmerer takes up this challenge, reviewing the neurobiological evidence for the separability of the main components of the Blueprint and for the time course of their involvement in speech planning. Drawing on evidence from neuroimaging (PET, fMRI and MEG), lesion-symptom mapping and TMS/tDCS, he concludes that there is good support for the view that different processing stages are implemented by broadly different brain circuits that are activated serially. The question of how the components of the blueprint can be empirically separated from each other is also taken up by Dell and colleagues, who used a learning paradigm to identify how linguistic constraints are represented and “bundled” in components, and it is discussed in Laganaro’s contribution in terms of the distinction between phonology and phonetics.

A key concept in *Speaking* is the notion of the lemma, initially seen as semantic-syntactic units and later defined as purely syntactic representations of words (Levelt et al., 1999; also discussed further in this volume by Abdel Rahman and Melinger). Kemmerer’s discussion focusses specifically on the neurobiological support for the existence of lemmas as separate from lexical concepts. He reviews evidence that several areas in the left anterior temporal lobe (ATL) support lexical concept selection, while the left middle temporal gyrus (left MTG) specifically supports lemma retrieval. This suggests that these processing stages are neurobiologically separable. Conversely, Kemmerer also highlights the commonalities between some of the stages that are separated in *Speaking*. For example, he demonstrates that the left inferior frontal gyrus (IFG) supports both syllabification and phonetic encoding, with BA44 – a segment of the left IFG – playing a key role in both processes. This suggests that these processes rely upon similar neural circuits. Behavioural evidence concerning the relationship between syllabification and phonetic encoding is discussed in Laganaro’s article.

2.2. Papafragou and Grigoroğlu: The conceptualisation of events (Speaking Chapters 3 and 4)

Before speakers can formulate an utterance, they need to decide what to say, i.e. to generate a representation of the utterance content. This pre-verbal message is a special kind of thought – not linguistic yet, but suitable for expression in the speaker’s language and tailored to the speaker’s audience. Slobin (1996) famously called the generation of preverbal messages “thinking for speaking”. Studying thinking for speaking (or other tasks) is challenging, not only because of the (alleged) non-linguistic

nature of the representations, but also because it is very hard to control a participant’s thoughts and communicative intentions. In order to guide participants’ thoughts, researchers have often used event or scene description tasks, which constrain to some extent what the speakers ought to think about. In many recent studies, participants’ eye movements have been recorded while viewing and describing these displays. This is helpful because eye gaze is tightly linked to visual attention (e.g. Rayner, 1998). Thus, by recording which parts of a display are inspected, in which order and for how long, researchers can track how speakers compose preverbal messages.

Papafragou and Grigoroğlu discuss a series of studies using this approach to examine how speakers generate preverbal messages about events. One important finding is that viewers need very little time (often less than 100 ms) to extract the gist of an event from a photograph. Thus, the “raw materials” for message generation are rapidly available to the viewer. The further visual exploration of the scenes was found to be affected by properties of the speakers’ native languages, but only when they were actually asked to speak about the scenes. This is consistent with the view that there is “thinking for speaking”. Other work in Papafragou and Grigoroğlu’s lab has shown how the communicative situation (specifically the listeners’ goals, and whether or not they interacted with the speakers) affected how children and adults described events. These studies illustrate how message generation can be traced by asking participants to describe carefully constructed displays in different communicative situations, and they begin to shed light on some of the cognitive, pragmatic, and linguistic factors that govern how speakers allocate their attention during message generation and decide what to include in their messages.

2.3. Ferreira and Rehrig: Meaning and visual salience effects on linearisation (Speaking Chapter 4)

Whereas Papafragou and Grigoroğlu consider the generation of preverbal messages capturing individual events, Ferreira and Rehrig discuss how speakers generate the macrostructure of multi-sentence descriptions, for instance of a room or building. As highlighted in *Speaking* (and earlier in Levelt, 1981), in order to deliver a description, speakers need to transform their visual impressions or memory of a complex visual scene into a single sequence of utterances. Ferreira and Rehrig ask to what extent this linearisation is driven by the visual salience of different regions of a scene or by the meaningfulness of the regions for the speaker. The results of their studies indicate that, from the earliest moments

of perceiving a scene, the viewer's eye movements are driven much more by scene meaning than visual salience. This dovetails with Papafragou and Grigoroglou's observation that viewers can rapidly understand the gist of scenes.

Ferreira and Rehrig introduce a novel research programme that is grounded in insights and techniques from both psycholinguistics and vision science, using sophisticated analyses of the physical properties of scenes and painstaking norming studies of scene meaning to generate experimental materials. These new techniques allow researchers to address in earnest some of the key issues behind utterance generation, such as how the uptake of visual-conceptual information is coordinated in time with the ongoing linguistic formulation process and how visual variables, higher-level cognitive processes and linguistic variables jointly direct a speaker's visual attention and utterance planning. Using these techniques, researchers can go beyond debating whether or not language processing is modular and generate detailed theories about the ways different cognitive components are jointly recruited to accomplish specific tasks.

2.4. Kempen and Harbusch: Frequency in linearisation (Speaking Chapter 3)

The process of linearisation is addressed from a different angle by Kempen and Harbusch. They present results of corpus analyses of German, Dutch and English that concern the distribution of verbs differing in frequencies across main and subordinate clauses. They observe that high frequency verbs are overrepresented in main clauses (relative to their overall frequency), and that this pattern is accentuated in German and Dutch compared to English. Kempen and Harbusch relate this pattern of frequency distributions to speakers' tendency to place high frequency verbs early in the sentence, following an "easy first" processing principle. Thus, speakers may be particularly likely to select high frequency words for the initial part of their utterance, when the pressure to begin to speak is high. The accentuation of the pattern in German and Dutch arises because in these languages, the verb must appear earlier in main clauses than in subordinate clauses. The implication is that the "easy first" principle has consequences for sentence-level grammatical encoding, specifically for the placement of materials in main or subordinate clauses. The analyses reported in this article offer a glimpse on the complex interplay of lexical and grammatical encoding and on the interplay of domain-general cognitive processing strategies and language specific constraints while simultaneously highlighting the potential of novel methods in corpus analysis.

2.5. Arnold and Zerkle: Selecting pronouns (Speaking Chapters 2, 4 and 7)

Message generation is also considered in Arnold and Zerkle's paper, which, rather than considering the broad organisation of utterances, zooms in on the selection of pronouns. Pronominalisation is an important and challenging issue for theories of language production because the use of pronouns is governed by many pragmatic, cognitive and linguistic variables. In their comprehensive review, the authors describe the central claims of two broad classes of theories of pronoun use and the relevant empirical evidence for each. Pragmatic selection accounts focus on the way pragmatic rules control the mapping of conceptual structures onto linguistic form. An important principle is that the more accessible or salient an entity is in the interlocutors' common ground, the more likely it should be to be referred to with a pronoun. By contrast, rational accounts focus on processing constraints, specifically the speaker's need to be efficient (minimising their own processing costs) and sufficiently informative to describe the appropriate referent (minimising the listener's processing costs).

As the authors discuss, neither type of approach is entirely satisfactory. For instance, pragmatic selection accounts can adequately describe conceptual and discourse-level linguistic constraints on pronoun use, but are rarely couched in explicit mechanistic terms. Furthermore, pragmatic selection accounts often assume that there is a single dimension of accessibility or salience, which is problematic as the criteria restricting pronoun use appear to differ across languages. Rational models explain variability across multiple levels of the production system and are easily couched in terms of Bayesian inference – an explicit mathematical implementation – but linking this implementation to psychological mechanisms is not straightforward. Moreover, rational models hinge on the assumption that using pronouns is relatively easy for the speaker compared to using nouns, which is not necessarily true. Finally, neither pragmatic selection nor rational accounts completely capture the effects of predictability of content, nor do they explain how speakers decide which form of reduction to use. The authors point to these issues as important directions for future work.

2.6. Abdel Rahman and Melinger: Lexical selection in the swinging network (Speaking Chapter 6)

In *Speaking*, the formulation of utterances is lexically driven: Utterances are composed by selecting and combining individual lexical units called lemmas. This is a key property of the Blueprint. Because of the centrality of

the lemma notion, Kemmerer refers to Levelt's model as the lemma model. But how are lemmas selected? Much empirical work on this issue has used picture naming, often in interference paradigms (picture-word interference, blocked-cyclic naming and continuous naming). This research has yielded an array of findings that is bewildering at first sight, with similar manipulations yielding contradictory effects. Most prominently, semantic relatedness between a target picture and a distractor word or picture can lead to interference (as compared to unrelated trials), facilitation or no effect. However, as Abdel Rahman and Melinger discuss, their Swinging Network model offers a parsimonious account of (almost) all of these findings. A key assumption of this model is that processing initially occurs in parallel at the conceptual and lexical levels, allowing activation to "reverberate" and making the network "swing". The model explains the direction of relatedness effects as due to the complex interplay of mainly facilitatory effects of relatedness at the conceptual level and mainly interference effects at the lexical level. Lexical selection is a competitive process, and the strength of the observed effects depends on properties of the sets of competitors, including the size of the set and the distance to the target.

2.7. Laganaro: Refining phonetic encoding (Speaking Chapters 8 to 10)

Whereas the papers discussed so far concerned the conceptual and semantic processing of utterances, the following three papers discuss the encoding of utterance form. In their articles, Laganaro and Shattuck-Hufnagel comment on the processes establishing the transition from meaning to phonetic form and highlight that this essential part of speech planning has not yet received due attention. In the Blueprint and many other models, generating utterance forms is a two-step process, involving the generation of abstract phonological and context-specific phonetic representations. One key issue is whether it is indeed useful to distinguish phonological and phonetic representations. Laganaro argues that this distinction is well supported by evidence from a variety of sources including sophisticated acoustic analyses of both correct speech and errors. Importantly, the evidence points to cascading information flow from phonological to phonetic representations. This is fully in line with the view that cascaded processing also occurs across other components of the cognitive system supporting speech planning.

Another key issue concerns the processing units at each level. For the phonetic level, syllable-sized units have been favoured in many frameworks, including

Speaking and later work on the mental syllabary (e.g. Bürki, Pellet Cheneval, & Laganaro, 2015; Cholin, Dell, & Levelt, 2011). As Laganaro argues, work done since *Speaking* has provided strong support for stored syllabic representations for high-frequency syllables, whereas articulatory programs for low-frequency and novel syllables are likely generated out of smaller components.

A final important point Laganaro makes is that the time required for phonetic encoding is often underestimated. Recent studies combining phonetic analyses of correct speech and errors with EEG recordings suggest that phonetic encoding takes considerably longer than the 150 ms or so commonly allowed for it in models of speaking. This long processing time points to the complexity of the planning processes occurring during this phase.

2.8. Shattuck-Hufnagel: Reduction and prosody (Speaking Chapter 10)

Shattuck-Hufnagel also discusses the complexity of phonetic encoding processes and the need to account for them in a complete model of speaking. As she points out, there have been many important discoveries in speech sciences in recent years indicating that the generation of connected speech is far more complex than sketched in most psycholinguistic models. This is perhaps illustrated most vividly by reduction phenomena, which are highly systematic and often so radical to call into question the notion that speakers generate word forms by selecting and combining stored programs for entire syllables. In line with these observations, Turk and Shattuck-Hufnagel have proposed an alternative feature-cue based framework, which is forthcoming.

Another important development in the speech sciences has been the refinement of theories of utterance level prosody. Utterance level prosody is partly governed by the thematic structure of the entire utterance. This means that some aspects of this structure must be in place before utterance onset, as discussed in Keating and Shattuck-Hufnagel's (2002) "prosody first" model. One might speculate that this early prosodic encoding happens during the gist extraction phase of planning, which is discussed by Papafragou and Grigorolou and by Ferreira and Rehrig in their articles. An important challenge for the field is to explore how prosodic information is generated and carried forward towards articulation.

2.9. Kearney and Guenther: The beauty of DIVA (Speaking Chapter 11)

Kearney and Guenther's article concerns the final component of the speech production system, articulation.

They provide an overview of their DIVA model of the planning and control of speech movements, which was developed soon after the publication of *Speaking*. The particular strength of the DIVA model derives from two features: the proposed fine-grained one-to-one mapping between postulated processing components and neuronal populations and the detailed computational implementation. These two features allow for stringent tests of the model against behavioural and neurobiology data.

The DIVA model itself has three main components – a motor reference frame, an auditory reference frame and a somatosensory reference frame. The output of the articulators is used for feedback to all three frames, allowing DIVA to adjust the motor plan and somatosensory reference according to the desired acoustic target. To illustrate this, the authors describe experiments where the feedback that participants received of their own speech was altered and the compensatory responses were recorded. DIVA captured these alterations of the participants' speech extremely well.

The fact that the DIVA model maps directly to neural circuitry has allowed the authors to test it directly on neurobiological data. For example, in experiments with altered feedback, additional activity was recorded on perturbation trials in brain areas in the posterior superior temporal cortex and ventral premotor cortex that are predicted by DIVA to be involved in processing auditory error signals and correcting the motor plan. Furthermore, DIVA has also been used to construct a brain-computer interface (BCI) to generate synthesised speech from the neural signals of a patient with locked-in syndrome.

An important challenge for future research is to link this model (or a similar one) to psycholinguistic models of the preceding planning processes and to determine how the requirements of motor implementation might affect planning at higher levels (see also Shattuck-Hufnagel's article).

2.10. Nozari, Martin and McCloskey: Semi-automatic repairs (Speaking Chapter 12)

When *Speaking* was written, analyses of speech errors and repairs were the most important sources of evidence about speech planning. In the following decades, other types of evidence became more important, but the issue of how speech errors are detected and repaired is still topical. One key issue is whether generating repairs for errors is an automatic or a controlled process. Nozari, Martin and McCloskey report a behavioural study inspired by Levelt (1983), which found that repair rates increased with error rates. At first glance, this may appear to be a counter-intuitive outcome. Shouldn't a

well-functioning system, which makes few errors, also to be able to detect most of the remaining errors, and conversely, shouldn't a less efficient system make more errors and detect fewer of them? As the authors argue, these expectations have been borne out in comparisons of systems that differ broadly in efficiency (as a trait), for instance healthy and impaired systems, but they do not apply to moment-by-moment changes in a single system's state. When, in a well-functioning system, processing becomes difficult, i.e. when high response conflict is detected, additional control mechanisms can be recruited which increase conflict resolution abilities. This can lead to a reduction of the error rates or an increase of repair rates for any errors that are still committed. As such, the authors argue for a hybrid system, where an automatic monitoring and repair system is supplemented by a controlled process.

2.11. Dell, Kelley, Bian, and Holmes: Speaking and learning (Speaking Chapter 1)

Speaking concerns the adult language production system, a snapshot of a grown-up at a particular time in their life. At the time of writing, language use and language learning were quite separate fields of study. Since then, the fields have approached each other: Work on language development has become more process-oriented, and the adult language processing literature has taken increasing interest in short-term adaptation and longer-term learning.

In their article, Dell, Kelley, Bian and Holmes discuss studies exploring the learnability of novel phonotactic constraints and the malleability of syntactic biases. As they demonstrate, evidence from learning studies can provide novel insights into the architecture of the linguistic processing system. The authors review phonotactic studies that showed that (1) adults could readily learn some second-order phonotactic constraints (i.e. rules that refer to conjunctions of conditions, such as "/f/ is an onset if the vowel is /ae/"); (2) some constraint learning required overnight consolidation or lengthy training in adults, but not in children; (3) some constraints were not learned at all, such as the conjunction of a positional constraint with speech rate ("/f/ is an onset in fast speech but a coda in slow speech"). These findings are interesting in their own right. But perhaps more importantly, Dell and colleagues link the differences in learnability to the state of the learner's linguistic system, where some constraints are "foregrounded" (play an important role in the currently used linguistic system) while others are "backgrounded" (were "tested" during language acquisition, but eventually discarded from the adult system), or don't exist at all. In computational terms these

differences correspond to differences in the availability of hidden nodes representing the conjunctions of constraints. Moreover, Dell and colleagues propose that these differences in the learnability of the constraints, and the availability of hidden units, map on the component structure of the processing system: Constraints that map together on a joint hidden unit refer to processes within a shared component, whereas constraints for which no joint hidden unit is available in the mature system are not part of a common component any more. For instance, speech rate and phonotactics are governed by different components, making rules referring to both simultaneously unlearnable.

In another set of studies, Dell and colleagues considered syntactic learning and showed that similar principles apply. They found that verb bias (the likelihood of a specific verb to take dative or prepositional object structure) can readily be modified and reversed through implicit learning processes. By contrast, changing participants' bias to use certain verbs in an active or passive sentence structure turned out to be impossible. While different accounts are possible for this pattern, an attractive option is that the binding of verbs to active or passive structures was not learned because the overall sentence structure is determined strongly by cues from the message level, and within the time course of sentence generation the structure has long been set before the specific verb is selected. Put differently, the broad structure of the sentence and the selection of a specific verb are part of different stages of the production system, and a rule that links a verb to a broad structure is therefore very difficult to learn implicitly.

3. Avenues for future research

The articles in this volume illustrate how many of the issues raised in *Speaking* have been taken up in subsequent theoretical and empirical work. Though there are still many highly contentious open issues, there is, in our view, wide consensus in the field about some key properties of the cognitive system that supports speaking: Most researchers would probably agree that it is useful to think of the language production system as consisting of different levels or stages of processing, which are engaged in sequence when utterances are planned. We also think that there is very strong support for incrementality – building utterances in a piecemeal fashion. Both of these assumptions are pivotal architectural properties of the Blueprint. Deviating from the spirit of *Speaking*, research in the past decades has revealed that the language production system is less modular than originally envisioned, both with respect to its internal components and with respect to its

relationships to other components of the cognitive system. Whereas the Blueprint primarily portrayed the language production system in isolation, recent work has often described the relationship of the system to other parts of the cognitive system and to neurobiology, as highlighted in the contributions to this volume from Kemmerer and from Kearney and Guenther.

What's next? Although substantial progress has been made over the last thirty years, we can't claim that the process of speaking is now well understood. Even for the production of the simplest of utterances, for instance naming an object using a monomorphemic noun, there are still open issues concerning the time course of speech planning, the processes involved in lexical selection, and the interfaces between linguistic planning and motor execution, to name a few.

While the utterances we produce to communicate with each other can be single words (e.g. "hmm", "yes" or even "horses!"), many are multi-word utterances. Our knowledge of the production of such utterances is relatively sparse. In addition to the work described or referred to in this volume, seminal work has been done by the Urbana-Champaign group led by Kay Bock and Gary Dell. One important line of work has concerned subject-verb agreement production, our knowledge of which has become fairly complete (e.g. Bock & Miller, 1991; Eberhard, Cutting, & Bock, 2005). Another important research line has concerned the mapping of conceptual information onto grammatical surface structures, in particular for declarative sentences featuring actives or passives and for ditransitive prepositional or direct object constructions. Here the main workhorse has been the syntactic priming paradigm (e.g. Bock, 1986; Mahowald, James, Futrell, & Gibson, 2016; the special issue of the *Journal of Memory and Language* edited by Dell & Ferreira, 2016). Based on this work, an influential model of syntactic processing was proposed by Chang, Dell, and Bock (2006). Other work has focused on the incremental planning of noun phrases and short sentences (Konopka, 2012; Konopka & Meyer, 2014; Smith & Wheeldon, 1999). We know less about planning other types of utterances, such as relative clauses, wh-phrases and negation, or linguistic phenomena such as the expression of aspect, tense or focus structure. As highlighted in Shattuck-Hufnagel's contribution to this issue, a particularly thorny issue is the generation of sentence-level prosody. Speakers have no problem generating prosodic structures; yet how they do this is largely unknown. To make matters worse, our understanding of sentence planning is based largely on studies of a small set of structurally similar languages, in particular English and Dutch, and we can only speculate about sentence planning in structurally different languages.

Why has there been relatively little work on sentence production? To an important extent this may be due to the difficulty of eliciting specific complex utterances and the inflation of effort in testing and transcribing needed to empirically examine small effects. New methods such as the innovations proposed by Ferreira and Rehrig in this volume and modern speech analysis systems can help us with these problems, and new tools for on-line testing and crowd sourcing are available now to reach speakers of many languages (e.g. Hartshorne, Tenenbaum, & Pinker, 2018).

Another important reason for our patchy knowledge of sentence production is that the issues to be addressed are theoretically challenging. Studying sentence production is hard because the underlying cognitive processes are complex and simultaneously involve all of the cognitive components of the Blueprint. To generate processing models of how speakers generate grammatically and prosodically well-formed sentences that express their intended meaning, each of the component processes and their interfaces need to be specified. This in turn involves solving a large number of specific problems, among them, for instance, understanding the visual, attentional and auditory systems, finding ways of expressing concepts and linguistic representations, and describing linguistic and motor representations in compatible formats. Addressing these issues requires, apart from determination and patience, interdisciplinary work between linguists and psychologists.

As concerns the neurobiological underpinning of sentence production, there is still a clear lack of experimental effort. In a meta-analysis of hemodynamic studies of syntactic processing, Indefrey (2012) could include 79 experiments on sentence comprehension, but only six on sentence production. In reviewing the literature, Kemmerer (2014) notes that “The daunting challenge of experimentally controlling all the different variables that influence sentence production has impeded the use of hemodynamic methods to investigate this topic. In fact, so far only a handful of PET and fMRI studies have sought to isolate the cortical mechanisms that underlie syntactic encoding” (p. 412). Clearly, there is a need for more neuroimaging studies on the neurobiological basis of sentence production (see also Roelofs & Ferreira, 2019). Researchers should have the courage to take the methodological hurdles, as they did in studying single word production using neuroimaging (for discussion see de Zubicaray & Piai, 2019).

Still, studying isolated words or sentences in many ways fails to address the essence of speaking – that it is used for communication with others. There are no obvious reasons why basic processes, such as the retrieval of individual words from the mental lexicon or the

linearisation of words in a phrase, would proceed in fundamentally different ways in conversational and lab contexts. After all, participants carrying out lab tasks build upon the language skills used in other contexts. Nonetheless, speaking in conversational contexts evidently differs in many important ways from speaking in the lab. Most obviously, in conversation, speakers should produce utterances that are relevant to the conversational context and do so shortly after the end of the interlocutor’s turn (e.g. Levinson & Torreira, 2015). This requires distribution of attention across the processes of speech planning and listening, as well as coordination of the conceptual and linguistic encoding processes required for speaking with the processes required for listening. Despite this additional effort, talking in conversation typically does not feel like a particularly complex task. It appears that conversation is easier than we might expect from examining speech production in the lab. One reason for this is that interlocutors support each other in many ways through mutual priming; another reason is that in natural conversations, speakers can choose what to say (or not to say anything at all), and very simple utterances (“*hmm*”, “*yes*”) can be timely and relevant. Such freedom of expression is typically not granted in laboratory contexts.

Studies of the cognitive processes that support conversation have only just begun. Corpora of conversational speech provide rich descriptive evidence of how conversation works, and there is a growing body of relevant experimental evidence. Herb Clark and his collaborators have been pioneers in this domain; their work has highlighted the bilateral nature of conversation, demonstrating in a variety of research paradigms that speaking and listening are not separate in conversation, but need to be studied together to understand how individuals build reference and meaning in discourse models (e.g. Bangerter & Clark, 2003; Brennan & Clark, 1996; Clark & Krych, 2004; Wilkes-Gibbs & Clark, 1992). Functional models of speaking and listening in conversation have begun to be developed, most notably by the group led by Pickering and Garrod (2004, 2013), who stress the similarity of the processes underlying listening and speaking, and the importance of production-like mechanisms in listening.

Yet, much more theoretical and empirical work is required to understand how utterances are produced in real-life contexts. How might one generate a systematic research programme on speaking in conversation? Given the complexity of the issue, there are many promising starting points. An obvious option is to follow the path already taken in much of the experimental literature and ask how speakers solve specific tasks that occur in everyday speech. An example of this would be to

provide a referential utterance in response to a question. The choice of the tasks to study with priority might be supported by descriptive work on conversation. A functional model would have to specify which cognitive tools interlocutors use in order to fulfill the task, how they orchestrate the use of these tools, and how the properties of the tools restrict the ways the task can be solved.

A research programme on word or sentence production or on speaking in conversation must always focus on specific tasks, such as the production of nouns, declarative sentences or indirect requests. But the empirical work needs to be driven by broader theoretical considerations, otherwise researchers engage in an activity dubbed “stamp collection” by an esteemed colleague (Hagoort, pers. communication). The most interesting and challenging theoretical questions, in our view, are those that concern the embedding of the speech production system in the broader cognitive system and the neurobiological underpinnings of this system. The embedding includes the relationship between the linguistic and the social cognition, conceptual, attention and motor systems, as well as the relationship between word and rule knowledge (the lexicon and the grammar). It also concerns the relationship between the production and comprehension system (see the special issue of the *Journal of Memory and Language* edited by Meyer, Huettig, & Levelt, 2016) and people’s ability to learn from linguistic input. To make progress, we need to build integrative theories that discover processing principles that apply across domains and reflect broader cognitive and neurobiological properties.

Finally, we specifically highlight two important unifying mechanisms that might provide an anchor for future work. One is the role of cognitive capacity and executive function in production. Deciding what to say – or not to say – specifically requires attention, as highlighted in this volume by Ferreira and Rehrig and by Nozari and colleagues, and in the WEAVER++ model (e.g. Roelofs, 2003; Roelofs & Piai, 2011); the role of attention in production may be especially critical in bilingual contexts (e.g. Costa & Santesteban, 2004). Considering how attention interfaces with production in the brain may allow us to leverage literature on the neural mechanisms of attention (e.g. Corbetta & Shulman, 2002; Desimone & Duncan, 1995; for a review see Posner, 2012). A second unifying theme is adaptation with experience. As Dell and colleagues highlight in their paper, speakers are flexible, with the capacity to learn new words and structures, but the architecture of the system can only change so much. We now understand priming as evidence of the tuning of the system (e.g. Chang et al., 2006), and we understand the importance of individual variability in

experience and language use. Again, it will be critical to link these findings to neurobiology. Assessing how the system *changes* will disclose important properties of how it is implemented in the brain. Studying these issues is important for understanding the architecture of the cognitive system as a whole and the way people adapt to each other and shape each other’s language system.

Acknowledgements

The workshop and this Special Issue would not have materialised without the kind support of many colleagues and friends. We are extremely grateful to Nicolas Wade for allowing us to display his portrait of Pim Levelt on the cover of this volume. We thank the scientific committee of the International Workshop for Language Production, F-Xavier Alario, Adam Buchwald, Albert Costa (farewell Albert!), Victor Ferreira, and Alissa Melinger, for their support and advice before and during the workshop, and the Max Planck Society, the Language in Interaction Consortium (NWO) and the Donders Institute of Radboud University for financial support. We thank the reviewers for their constructive feedback and Billi Randall at the Language, Cognition and Neuroscience editorial office for superb—and patient—technical and editorial assistance. Most importantly, we thank Pim Levelt for all his kindness, his never-ending support of our endeavours and 30 years of inspiration.

Disclosure statement

No potential conflict of interest was reported by the authors.

References

- Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59(4), 390–412. doi:10.1016/j.jml.2007.12.005
- Bangerter, A., & Clark, H. H. (2003). Navigating joint projects with dialogue. *Cognitive Science*, 27, 195–225. doi:10.1207/s15516709cog2702_3
- Bock, J. K. (1986). Syntactic persistence in language production. *Cognitive Psychology*, 18(3), 355–387. doi:10.1016/0010-0285(86)90004-6
- Bock, K., & Miller, C. A. (1991). Broken agreement. *Cognitive Psychology*, 23(1), 45–93. doi:10.1016/0010-0285(91)90003-7
- Brennan, S. E., & Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22(6), 1482–1493. doi:10.1037//0278-7393.22.6.1482
- Bürki, A., Pellet Cheneval, 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. doi:10.1016/j.bandl.2015.08.006
- Caramazza, A. (1997). How many levels of processing are there in lexical access? *Cognitive Neuropsychology*, 14(1), 177–208. doi:10.1080/026432997381664

- Chang, F., Dell, G. S., & Bock, K. (2006). Becoming syntactic. *Psychological Review*, 113(2), 234–272. doi:10.1037/0033-295X.113.2.234
- 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(1), 109–122. doi:10.1037/a0021322
- Clark, H. H. (1973). The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior*, 12(4), 335–359. doi:10.1016/S0022-5371(73)80014-3
- Clark, H. H., & Krych, M. A. (2004). Speaking while monitoring addressees for understanding. *Journal of Memory and Language*, 50(1), 62–81. doi:10.1016/j.jml.2003.08.004
- Corbetta, M., & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nature Reviews Neuroscience*, 3(3), 201–215. doi:10.1038/nrn755
- Costa, A., & Santesteban, M. (2004). Lexical access in bilingual speech production: Evidence from language switching in highly proficient bilinguals and L2 learners. *Journal of Memory and Language*, 50(4), 491–511. doi:10.1016/j.jml.2004.02.002
- de Groot, A. M., & Hagoort, P. (Eds.). (2018). *Research methods in psycholinguistics and the neurobiology of language: A practical guide* (Vol. 9). John Wiley & Sons.
- Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review*, 93(3), 283–321. doi:10.1037/0033-295X.93.3.283
- Dell, G. S., & Ferreira, V. S. (Eds.). (2016). Thirty years of structural priming: An introduction to the special issue. *Journal of Memory and Language*, 91, 1–4. doi:10.1016/j.jml.2016.05.005
- 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(4), 801–838. doi:10.1037/0033-295X.104.4.801
- Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience*, 18(1), 193–222. doi:10.1146/annurev.ne.18.030195.001205
- de Zubicaray, G. I., & Piai, V. (2019). Investigating the spatial and temporal components of speech production. In G. I. de Zubicaray & N. O. Schiller (Eds.), *The Oxford Handbook of Neurolinguistics* (pp. 471–497). Oxford: Oxford University Press. doi:10.1093/oxfordhb/9780190672027.013.19
- Eberhard, K. M., Cutting, J. C., & Bock, K. (2005). Making syntax of sense: Number agreement in sentence production. *Psychological Review*, 112(3), 531–559. doi:10.1037/0033-295X.112.3.531
- Frank, M. C., & Goodman, N. D. (2012). Predicting pragmatic reasoning in language games. *Science*, 336(6084), 998–998. doi:10.1126/science.1218633
- Garrett, M. (1990). Review of Levelt: Speaking. *Language and Speech*, 33, 273–291. doi:10.1177/002383099003300303
- Glaser, W. R., & Dünghoff, F. J. (1984). The time course of picture-word interference. *Journal of Experimental Psychology: Human Perception and Performance*, 10(5), 640–654. doi:10.1037/0096-1523.10.5.640
- Hartshorne, J. K., Tenenbaum, J. B., & Pinker, S. (2018). A critical period for second language acquisition: Evidence from 2/3 million English speakers. *Cognition*, 177, 263–277. doi:10.1016/j.cognition.2018.04.007
- Indefrey, P. (2012). Hemodynamic studies of syntactic processing. In M. Faust (Ed.), *The Handbook of the Neuropsychology of language* (Vol. 1, pp. 209–228). Malden, MA: Blackwell Publishing. doi:10.1002/9781118432501.ch11
- Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, 59(4), 434–446. doi:10.1016/j.jml.2007.11.007
- Jaeger, T. F. (2010). Redundancy and reduction: Speakers manage syntactic information density. *Cognitive Psychology*, 61(1), 23–62. doi:10.1016/j.cogpsych.2010.02.002
- Keating, P., & Shattuck-Hufnagel, S. (2002). A prosodic view of word form encoding for speech production. *UCLA working papers in phonetics*, 112–156.
- Kemmerer, D. (2014). *Cognitive neuroscience of language*. New York: Psychology Press.
- Konopka, A. E. (2012). Planning ahead: How recent experience with structures and words changes the scope of linguistic planning. *Journal of Memory and Language*, 66(1), 143–162. doi:10.1016/j.jml.2011.08.003
- Konopka, A. E., & Meyer, A. S. (2014). Priming sentence planning. *Cognitive Psychology*, 73, 1–40. doi:10.1016/j.cogpsych.2014.04.001
- Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*, 207(4427), 203–205. doi:10.1126/science.7350657
- Levelt, W. J. (1981). The speaker's linearization problem. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences*, 295(1077), 305–315. doi:10.1098/rstb.1981.0142
- Levelt, W. J. (1983). Monitoring and self-repair in speech. *Cognition*, 14(1), 41–104. doi:10.1016/0010-0277(83)90026-4
- Levelt, W. J. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT press.
- Levelt, W. J., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, 22(1), 1–38. doi:10.1017/S0140525X99001776
- Levinson, S. C., & Torreira, F. (2015). Timing in turn-taking and its implications for processing models of language. *Frontiers in Psychology*, 6, 731. doi:10.3389/fpsyg.2015.00731
- MacDonald, M. C. (2013). How language production shapes language form and comprehension. *Frontiers in Psychology*, 4, 226. doi:10.3389/fpsyg.2013.00226
- 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. doi:10.1016/j.jml.2016.03.009
- Meyer, A. S., Huettig, F., & Levelt, W. J. (Ed.) (2016). Same, different, or closely related: What is the relationship between language production and comprehension? [Special issue]. *Journal of Memory and Language*, 89, 1–7. doi:10.1016/j.jml.2016.03.002
- Oldfield, R. C., & Wingfield, A. (1965). Response latencies in naming objects. *Quarterly Journal of Experimental Psychology*, 17(4), 273–281. doi:10.1080/17470216508416445
- Petersen, S. E., Fox, P. T., Posner, M. I., Mintun, M., & Raichle, M. E. (1988). Positron emission tomographic studies of the cortical anatomy of single-word processing. *Nature*, 331(6157), 585–589. doi:10.1038/331585a0

- Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, 27(2), 169–190. doi:10.1017/S0140525X04000056
- Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences*, 36(4), 329–347. doi:10.1017/S0140525X12001495
- Posner, M. I. (2012). *Attention in a social world*. Oxford: Oxford University Press. doi:10.1093/acprof:oso/9780199791217.001.0001
- Rapp, B., & Goldrick, M. (2000). Discreteness and interactivity in spoken word production. *Psychological Review*, 107(3), 460–499. doi:10.1037/0033-295X.107.3.460
- Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, 124(3), 372–422. doi:10.1037/0033-2909.124.3.372
- Roelofs, A. (1992). A spreading-activation theory of lemma retrieval in speaking. *Cognition*, 42, 107–142. doi:10.1016/0010-0277(92)90041-F
- Roelofs, A. (2003). Goal-referenced selection of verbal action: Modeling attentional control in the Stroop task. *Psychological Review*, 110(1), 88. doi:10.1037/0033-295X.110.1.88
- Roelofs, A. (2014). A dorsal-pathway account of aphasic language production: The WEAVER++/ARC model. *Cortex*, 59, 33–48. doi:10.1016/j.cortex.2014.07.001
- Roelofs, A., & Ferreira, V. S. (2019). The architecture of speaking. In P. Hagoort (Ed.), *Human language: From genes and brains to behavior* (pp. 35–50). Cambridge: MIT Press.
- Roelofs, A., & Piai, V. (2011). Attention demands of spoken word planning: A review. *Frontiers in Psychology*, 2, 307. doi:10.3389/fpsyg.2011.00307
- Schriefers, H., Meyer, A. S., & Levelt, W. J. (1990). Exploring the time course of lexical access in language production: Picture-word interference studies. *Journal of Memory and Language*, 29(1), 86–102. doi:10.1016/0749-596X(90)90011-N
- Schwartz, M. F., & Dell, G. S. (2010). Case series investigations in cognitive neuropsychology. *Cognitive Neuropsychology*, 27(6), 477–494. doi:10.1080/02643294.2011.574111
- Slobin, D. I. (1996). From “thought and language” to “thinking for speaking”. In J. J. Gumperz & S. C. Levinson (Eds.), *Rethinking linguistic relativity* (pp. 70–96). Cambridge: Cambridge University Press.
- Smith, M., & Wheeldon, L. (1999). High level processing scope in spoken sentence production. *Cognition*, 73(3), 205–246. doi:10.1016/S0010-0277(99)00053-0
- van Turenhout, M., Hagoort, P., & Brown, C. M. (1998). Brain activity during speaking: From syntax to phonology in 40 milliseconds. *Science*, 280(5363), 572–574. doi:10.1126/science.280.5363.572
- Vasishth, S., Nicenboim, B., Beckman, M. E., Li, F., & Kong, E. J. (2018). Bayesian data analysis in the phonetic sciences: A tutorial introduction. *Journal of Phonetics*, 71, 147–161. doi:10.1016/j.wocn.2018.07.008
- Wilkes-Gibbs, D., & Clark, H. H. (1992). Coordinating beliefs in conversation. *Journal of Memory and Language*, 31(2), 183–194. doi:10.1016/0749-596X(92)90010-U