7 Models Linking Production and Comprehension

CHIARA GAMBI[1, 2] AND
MARTIN J. PICKERING[2]

[1] Saarland University
[2] University of Edinburgh

Introduction

Production and comprehension have traditionally been studied in separate subfields of psycholinguistics, despite the fact that many psycholinguistic tasks involve both. For example, in the picture-word interference paradigm, participants name target pictures while ignoring auditorily presented or superimposed written distractor words (e.g., Schriefers, Meyer, & Levelt, 1990). Therefore, this paradigm measures the effect of comprehension processes on production processes. This disciplinary division notwithstanding, there is general consensus that production and comprehension are linked. However, opinions diverge with regard to the nature of this link. It is not easy to characterise scholarly disagreement on this issue, as the issue is often not discussed explicitly, and even when it is, it is rarely treated as a key element of the model being proposed.

There are, however, some exceptions to this trend. Below, we focus on models that make explicit claims about the links between production and comprehension. It is noteworthy that these models tend to emphasize what is shared between comprehension and production, rather than what is not shared. Not surprisingly, frameworks that emphasize sharing tend to be concerned with explaining language learning, acquisition, and change (see Dell & Chang, 2014; MacDonald, 2013) or dialogue (see Pickering & Garrod, 2013), rather than isolated acts of production or comprehension. In fact, the simple observation that the primary site of language use is dialogue constitutes, in itself, a strong motivation for positing links between production and comprehension (Garrod & Pickering, 2004; see section on *Dialogue*, this chapter). And theories of how linguistic representations are acquired and

The Handbook of Psycholinguistics, First Edition. Edited by Eva M. Fernández and Helen Smith Cairns.

develop over time must naturally take into account how linguistic input shapes the output of the language system (and *vice versa*; see *Frameworks that posit linked preferences*, this chapter).

Traditionally, a number of arguments are used to claim that a separation between comprehension and production should be maintained. First, dissociations between the comprehension and production abilities of patients with brain lesions appear to support the notion that the neural substrates of production and comprehension are separate. For example, Kim and Thompson (2000) reported that agrammatic patients showed intact comprehension of verbs but were impaired in verb naming. Second, asymmetries exist between the development of language comprehension and language production in infants (see Bates, 1993), and similarly, between the rate of decay of comprehension and production abilities in older adults (e.g., McKay, Abrams, & Pedroza, 1999). These findings indicate that it is not possible to fully equate production and comprehension. A comprehensive review of these and other arguments is beyond the scope of this chapter. Here, we just note that these arguments do not necessarily imply that comprehension and production are completely separate. Rather, dissociations and asymmetries are in principle compatible with some degree of sharing, as long as there are subcomponents of production that are not used in comprehension and *vice versa*.

When discussing the issue of what is shared between production and comprehension, it is useful to bear in mind the distinction between linguistic *representations* and *processes* acting on those representations. Linguistic representations are the components of memory that store information about linguistic units (e.g., phonemes, words, syntactic rules, concepts). Comprehension and production processes are the cognitive operations that can be applied to linguistic representations (e.g., retrieval, spreading activation, inhibition), as well as the operations that map from abstract representations to articulation and from acoustics to abstract representations. Processes are directional: so, for example, the process that retrieves a phonological representation given an activated semantic representation is not the same as the process that retrieves a semantic representation given an activated phonological representation. Below, we first discuss sharing with regard to representations (*Representational parity*), and then we turn to theories that posit common processes (*Linked processes*).

Representational parity

Most theorists assume that some of the representations that are accessed during production are the same as the representations accessed during comprehension. At the single word level, the most influential theory of lexical access in production assumes conceptual (semantic), lemma (syntactic), and word-form (sound-based) representations, and proposes the network for production and the network for comprehension "coincide from the lemma level upwards" (i.e., concept and lemma nodes are shared between production and comprehension; Levelt, Roelofs, & Meyer, 1999, p. 7). Pickering and Garrod (2004) also assume shared representations (which they term the *parity hypothesis*), but extend the scope of this assumption by positing that representations used in comprehension are the same as representations used in production at all linguistic levels, from the situation model (i.e., a representation of the situation being discussed, including time, space, causal relations, intentionality, and individuals involved; Zwaan & Radvansky, 1998) to phonology and phonetics. Shared phonological representations are also part of the Node Structure Theory of MacKay (1987). Moreover, parity at the sound level is the central assumption of the Motor Theory of speech perception (Liberman & Whalen, 2000; see Galantucci, Fowler, & Turvey, 2006), and of the Episodic Theory of speech perception (Goldinger, 1998).

Parity at the semantic and lexico-syntactic levels has been confirmed by several studies. The strongest evidence comes from findings of immediate effects of comprehension on production. First, silent reading of words semantically and associatively related to the name of a target picture affects naming times for the picture (e.g., Schriefers *et al.*, 1990; Alario, Segui, & Ferrand, 2000). Second, silent reading of sentences such as *A rock star sold an undercover agent some cocaine* (double object, DO) or *A rock start sold some cocaine to an undercover agent* (prepositional object, PO) influences what sentence structure (DO or PO) is used to describe a target scene (depicting an unrelated event, such as a man reading a book to a boy; Bock, Dell, Chang, & Onishi, 2007). And, similarly, the syntactic choices of their interlocutor influence speakers' syntactic choices in dialogue (Branigan, Pickering, & Cleland, 2000; Levelt & Kelter, 1982; see Pickering & Ferreira, 2008 for a review). In addition, interlocutors align their lexical choices (Brennan & Clark, 1996; Garrod & Anderson, 1987).

As well as behavioral evidence, there is growing evidence for lexico-syntactic parity at the neural level. Two fMRI studies showed that the same neural populations are recruited during comprehension and production of sentences (Menenti, Gierhan, Segaert, & Hagoort, 2011; Segaert, Menenti, Weber, Petersson, & Hagoort, 2012). These studies identified brain areas that were activated less while producing (or comprehending) a given sentence structure, when the participant had just processed a sentence with the same structure, compared to a sentence with a different structure. This phenomenon is called repetition suppression, and it is used to localize neural areas that are sensitive to a given property of the stimulus (in this case, structure). Importantly, the areas identified were the same regardless of whether prior processing of the same structure had taken place in production or comprehension.

Regarding parity at the phonological level, it is known that silent reading of, or passive listening to, distractor words phonologically related to the name of a target picture speeds up naming times for the picture (e.g., Schriefers *et al.*, 1990; Damian & Martin, 1999). This is usually taken as evidence that comprehension of distractor words pre-activates phonological representations they share with the target, so that those representations are subsequently easier to access in production. Moreover, Kerzel and Bekkering (2000) found that participants pronounced a printed syllable more quickly while watching a video of a mouth producing the same syllable compared to when the mouth produced a different syllable (see also

Jarick & Jones, 2008). In addition, Galantucci, Fowler, and Goldstein (2009) showed that speakers are faster to produce a syllable (e.g.,/ba/) if they have just listened to the same syllable than to a syllable with a different onset (e.g.,/da/) (see also Fowler, Brown, Sabadini, & Weihing, 2003).

There is also much evidence demonstrating motor activation during the perception of speech. Such evidence suggests that speech production and speech perception make use of overlapping neural representations. Several studies have found activation of motor areas in the brain during audiovisual speech perception (e.g., Skipper, Nusbaum, & Small, 2005; Skipper, van Wassenhove, Nusbaum, & Small, 2007), and also during passive listening to speech (e.g., Wilson, Saygin, Sereno, & Iacoboni, 2004). Importantly, motor activation during passive listening is articulator-specific: for example, listening to labial consonants is associated with activation of the lip representation area in motor cortex (Pulvermüller *et al.*, 2006; see Pulvermüller & Fadiga, 2010 for a review). Further, listening to speech modulates the excitability of the speech muscles that are involved in the production of the perceived sound (e.g., Fadiga, Craighero, Buccino, & Rizzolatti, 2002). Accordingly, listening to a phoneme also affects concurrent articulation of a different phoneme; for example the palatal sound/k/was produced with greater contact between the tip of the tongue and the alveolar ridge when participants were listening to the alveolar sound/t/(Yuen, Davis, Brysbaert, & Rastle, 2010).

This body of evidence supports the assumption of shared representations at the phonological level at least. However, recent findings suggest that motor activation during speech perception might occur primarily or exclusively when the task is difficult (e.g., when speech is degraded; Adank, 2012; D'Ausilio, Bufalari, Salmas, & Fadiga, 2012). Moreover, it may be that motor activation under normal listening conditions reflects listeners' tracking of the speaker's speech rate and preparation for speech in anticipation of the end of the speaker's utterance, rather than retrieval of shared content-specific phonological representations (S. K. Scott, McGettigan, & Eisner, 2009).

Figure 7.1a is intended as a schematic summary of the overview of the literature on representational parity presented in this section. It depicts what we take to be the consesus view on the issue of representational parity across levels of linguistic representations. First of all, it illustrates the fact that there is substantial consensus on parity at the semantic and syntactic level (gray *sem* and *syn* representations). Phonological representations, instead, are depicted as overlapping but not identical (partially superimposed black and white *phon* representations) to account for the fact that evidence for parity at this level might be restricted to particular tasks (as just discussed). Finally, representations at the phonetic level are labelled as speech percepts (comprehension) and speech motor commands (production) and are assumed to be separate (i.e., white speech percepts are separate from black motor command representations).

Some researchers assume substantial overlap at the phonetic level as well, but parity at this level is particularly controversial. On one hand, speakers can converge toward a model speaker at the level of low-level phonetic features that do not imply phonological distinctions (e.g., vowel duration, F0), therefore providing

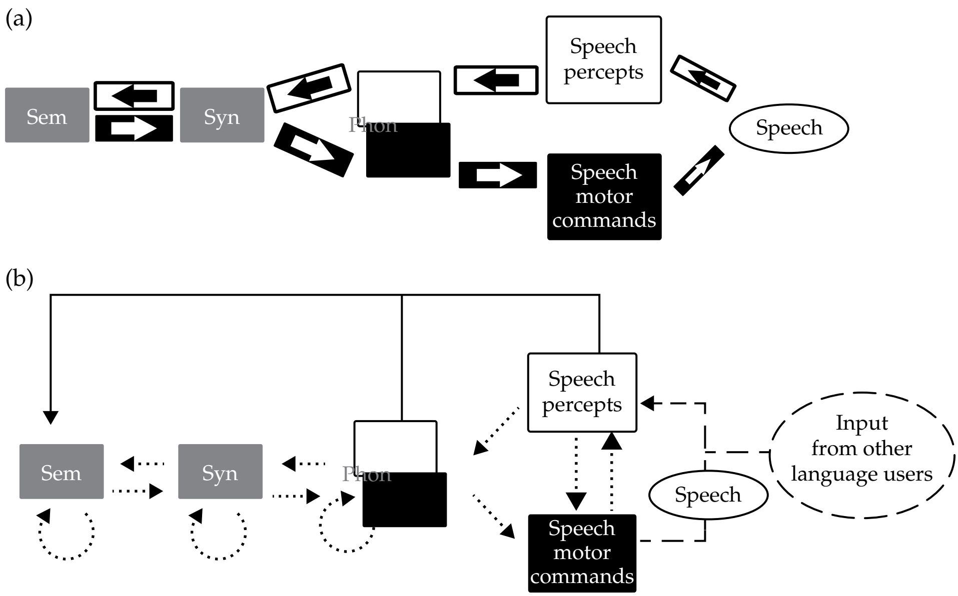

Figure 7.1 **Panel** (a) shows production processes (white arrows on black background) and comprehension processes (black arrows on white background); **Panel** (b) shows loops: the solid line is Levelt's (1989) Perceptual Loop; the dotted arrows are Pickering and Garrod's (2013) fast loops between and within levels; the dashed arrows are external loops that include long-term effects of exposure to distributional regularities in linguistic input (see *The P-chain framework*, and *Frameworks that posit linked preferences*).

support for the hypothesis that phonetic representations are shared between comprehension and production (e.g., Goldinger, 1998; Pardo, 2006; see Chapter 6 in this volume). However, other studies failed to find evidence that listening to a phonetic variant (e.g., alveolar/r/) facilitates subsequent production of the same variant compared to listening to an alternative variant (e.g., uvular/r/) that is phonologically equivalent (Mitterer & Ernestus, 2008; Mitterer & Müsseler, 2013).

Moreover, despite some indication that adaptation to an accent in comprehension correlates with adaptation to the same accent in production at the level of an individual speaker (Evans & Iverson, 2007), there are also asymmetries between people's production and comprehension abilities in the processing of regional variability in speech. For example, Sumner and Samuel (2009) argued that listeners who have long-term perceptual experience with a dialectal phonetic variant can achieve native-like perception of that variant despite lacking native-like production representations (see also Kralijc, Brennan, & Samuel, 2008). To sum up, some findings in the literature suggest comprehension and production phonetic representations are shared or overlapping, but others support the notion of distinct comprehension and production representations at this level. Overall, the evidence for separation is stronger at the phonetic than at the phonological level. In

Figure 7.1a, we capture this by drawing separate (rather than partially overlapping) phonetic representations, but it is still unclear how much separation should be assumed at this level.

Linked processes

Figure 7.1 does not only depict representations, but also processes. Traditionally, any process that takes place during an act of comprehension (i.e., while reading or listening) has been considered a comprehension process and, conversely, any process that takes place during an act of production (i.e., while writing or speaking), a production process. Within this tradition, the debate has focussed on whether processes that take place during production always map from semantics to syntax and from syntax to phonology (i.e., white arrows on black background, flowing from left to right in Figure 7.1a), or whether they can also map in the opposite direction (i.e., black arrows on white background, flowing from right to left in Figure 7.1a). Processes that go “backward” have been termed feedback processes and models that incorporate them (e.g., Dell, 1986) have been labelled as interactive (as opposed to purely feed-forward models; Levelt *et al.*, 1999). Similar issues have been discussed in the comprehension literature, where left-to-right processes (see Figure 7.1a) have been called top-down and right-to-left processes have been called bottom-up (e.g., Marslen-Wilson, 1987). Within the traditional view, researchers do not typically ask whether production processes can take place during comprehension or whether comprehension processes can take place during production. Instead, they tend to label any process that takes place during production as a “production process” and any process that takes place during comprehension as a “comprehension process,” regardless of the direction in which it flows.

But on an alternative view, a process is classed as a production process if it maps from representations that are higher in the linguistic hierarchy (e.g., semantics) to representations that are lower (e.g., phonology), that is from left to right in Figure 7.1a. Conversely, a process is classed as a comprehension process if it maps from lower to higher representations (e.g., from phonology to semantics), that is from right to left in Figure 7.1a. According to this definition, which we will follow in the remainder of this chapter, both production and comprehension processes could potentially be employed during any act of production, as they could during any act of comprehension (Pickering & Garrod, 2013).

Note that this difference is not merely terminological. Rather, it reflects a substantial theoretical distinction between the traditional and the alternative view, which could be tested experimentally. Should we conceptualize a processing flow that goes from phonology to syntax during an act of production as a production process (because it takes place during an act of production, as per the traditional view), or as a comprehension process (because it flows from right to left, as per the alternative view)? One way of answering this question would be by examining the neural pathways involved. For example, if we could show that the same neural pathway is involved during an act of production that implies reliance on feedback flow of information, as well as during an act of comprehension, then we would have empirical evidence for the alternative view. But if feedback during production and comprehension engage separate pathways despite a common direction in the flow of information, then the evidence would be more compatible with the traditional view.

Unfortunately, progress toward establishing how much overlap there is between production and comprehension processes has been hindered by the division between sub-disciplines within psycholinguistics. Comprehension researchers often use a production measure (e.g., word naming time) as their dependent variable, but do not interpret the production processes themselves (the "mind-in-the-mouth" assumption; Bock, 1996). Similarly, production researchers presumably realize that a task such as picture-word interference involves comprehension processes but tend to ignore them.

For example, production researchers concerned with picture-word interference have assumed that word form and phoneme representations activated by distractor words (in comprehension) send activation to related word forms and phonemes in the production network (Levelt *et al.*, 1999). But they do not ask what processes are involved in comprehension of the distractor words, focussing instead only on the end result of those processes (i.e., that some representation gets activated). For example, Damian and Martin (1999) showed that the time course of semantic and phonological effects in picture-word interference differed for visually presented distractors compared to auditorily presented distractor words, and concluded this was because of the longer presentation times used for visual distractors (in previous PWI studies). They did not consider the possibility that lexical co-activation (in comprehension) might have a different time course depending on modality (visual or auditory), and instead assumed that different presentation times led to different effects because the distractors interacted with different stages of production of the target picture name.

However, production-comprehension links have not been entirely overlooked. In fact, both feed-forward and feedback links have long been identified as a crucial component of neuro-computational theories of speech motor control (Tourville & Guenther, 2011; Hickok, 2012) and learning (Guenther & Vladusich, 2012; see also Plaut & Kello, 1999). Separately, the psycholinguistic literature on language production has also given ample consideration to this topic, albeit under the specific heading of self-monitoring. A long-standing psycholinguistic account of self-monitoring, the Perceptual Loop Theory, equates the self-monitoring system (active during acts of production) with the comprehension system (Levelt, 1983, 1989).

We first briefly present two neuro-computational models of speech motor control (Tourville & Guenther, 2011; Hickok, 2012). Then, we describe the Perceptual Loop Theory and discuss some criticisms of it. In the subsequent section, we introduce an integrated framework for language comprehension and language production (Pickering & Garrod, 2013), which includes an alternative account of self-monitoring (Pickering & Garrod, 2014), and also posits that production processes take place during acts of comprehension (and not just that comprehension processes take place during acts of production, as in self-monitoring). We then consider the P-chain framework (Dell & Chang, 2014), which makes a related proposal. Finally, we discuss proposals that place the link between production and comprehension outside specific acts of comprehension or production, in the long-term experience that speakers and listeners have with language (e.g., MacDonald, 2013).

Neuro-computational models of speech motor control

In neuro-computational models of speech motor control (the Directions Into Velocities of the Articulators, or DIVA, model, see Tourville & Guenther, 2011; the Hierarchical State Feedback Control, or HSFC, model, see Hickok, 2012, 2014), forward models map from motor commands sent to the articulators to the sensory (i.e., auditory or somatosensory) consequences of executing those commands (the upward dotted vertical arrow from motor commands to speech percepts in Figure 7.1b). Forward models therefore instantiate a relatively low-level mapping between production and comprehension representations and can be considered as internal models of the language system (that is, models of the processes that cause the articulation of speech sounds during acts of production). The inverse mapping corresponds to feedback-based correction of speech movements (downward dotted vertical arrow from speech percepts to motor commands in Figure 7.1b).

Evidence that forward models are implicated in speech production comes from the finding that auditory responses to speech sounds are suppressed during speaking compared to listening (e.g., M100 suppression, as reported using magneto-encephalography in the study of Houde, Nagarajan, Sekihara, & Merzenich, 2002). This is thought to occur because forward models can be used during production to anticipate sensory stimulation and cancel it out (in a way that could be useful in distinguishing between self-generated and externally-generated sounds). In support of this claim, auditory responses are suppressed in a stimulus-specific manner: suppression occurs during covert rehearsal compared with a control task, but only when the rehearsed stimulus matches (part of) the perceived stimulus (Ylinen *et al.*, 2014). In addition, enhancement rather than suppression takes place when auditory feedback is altered unexpectedly during speaking, for example by shifting pitch upwards or downward in real-time, so that the predicted stimulation ceases to match actual stimulation (e.g., Chang, Niziolek, Knight, Nagarajan, & Houde, 2013).

All of these studies are compatible with forward-model predictions operating at the level of fine-grained phonetic features. But there is also some indication that such predictions might be phonological in nature. Niziolek, Nagarjan, and Houde (2013) showed the degree of suppression is larger for sounds that are closer to a given speaker's median productions, suggesting that predictions could be computed on the basis of somewhat abstract representations. In other words, when the speaker selects a motor command to execute, the anticipated sensory consequences might correspond to an abstract phonological target (i.e., what it should sound like) rather than to a detailed phonetic target (i.e., what it is going to sound like on this particular instance).

This finding provides support for the HSFC model (Hickok, 2012, 2014). In this model, forward model predictions operate at two hierarchically organized levels: phonemes and syllables. Motor programs (corresponding to planned syllables and planned phonemes) inhibit sensory areas where perceptual targets are represented. These same areas are activated via feedback from the movements of the vocal tract and from the resulting speech output (i.e., when the speaker perceives her own productions). In addition, they can receive activation from concepts and lemmas. The discrepancy between the expected activation (propagated in the form of inhibitory connections from the motor targets) and the actual activation constitutes a prediction error, which is propagated back to the motor target areas and used for online corrections of motor programs (and learning of more accurate motor-to-sensory mappings).

The inhibitory connections therefore implement a form of prediction that maps from motor commands to expected sensory consequences of executing those commands. The excitatory backward connections, instead, implement a form of inverse correction, which maps from sensory prediction errors to changes in the motor commands needed to compensate for those errors. Evidence for a fast-cycling loop at the phonetic level, in which motor representations are rapidly mapped onto sensory representations and *vice versa*, comes from several demonstrations that speakers compensate very quickly for perturbed auditory feedback (e.g., Houde & Jordan, 1998; Jones & Munhall, 2002; Tourville, Reilly, & Guenther, 2008).

The HSFC model is closely related to the DIVA/GODIVA model proposed by Guenther and colleagues (see Tourville & Guenther, 2011). This model also incorporates the notion that somatosensory and auditory target areas are activated via forward-model predictions as well as via processing of sensory input, and that prediction errors are used for online correction (as well as for learning the mappings between movements and their sensory consequences). Importantly, both models incorporate what we might call an account of self-monitoring. They assume that a process that maps from motor areas to sensory areas (and *vice versa*) is essential for online control during speech production (see Plaut & Kello, 1999 for another computational model that instantiates this idea).

Crucially, unlike psycholinguistic theories of self-monitoring (see below), these neuro-computational models focus on sound-level representations and processes, and say very little about other linguistic levels. Hickok (2012, 2014) argued for the importance of integrating psycholinguistic theories of language production with models of speech production, and integrated lemma and conceptual representations within his HSFC. However, he did not explicitly extend the forward-model architecture to these levels. In terms of Figure 7.1b, his model assumes that dotted arrows flow in both directions between all levels, indicating that both comprehension and production processes take place during language production. But the HSFC model includes a fast-cycling within-level loop (recursive dotted arrows), which is responsible for fast error correction during production, at the phonological and phonetic levels only. No such loop is explicitly assumed at the lemma and conceptual levels.

An early model linking production and comprehension: The Perceptual Loop Theory of self-monitoring

According to the Perceptual Loop Theory (Levelt, 1983, 1989), production errors are detected via comprehension of speech output (the external loop), and also via comprehension of phonological representations (the internal loop). The comparison process takes place at the level of communicative intentions (messages): If the message reconstructed by the comprehension system does not match the message originally intended, the monitor flags up an error. This comprehension-production loop is depicted in Figure 7.1b using a solid black line. This loop links comprehension representations at the phonological (phonological comprehension representations) and phonetic level (the speech percepts formed during comprehension of the speech output) to semantic representations (that are shared between comprehension and production).

Crucially, the Perceptual Loop Theory is open to criticism because it posits a relatively slow-cycling loop. First, the speech signal must be analyzed by comprehension processes to recover speech percepts (if using the external loop), or phonological representations retrieved by production processes have to be analyzed by comprehension processes to activate corresponding phonological representations in the comprehension network (if using the internal loop). Additional comprehension processes then map from sound-based comprehension representations to a semantic comprehension representation. Finally, the activated semantic representation in the comprehension network is compared to the semantic representation that was originally activated in the production network (the latter process is facilitated by shared representations at the semantic level). In addition, if a discrepancy is detected, production of the current utterance must be stopped before production of a replacement can start.

Oomen and Poostma (2002) found that having participants engage in a concurrent task while speaking caused them to stop speaking more quickly after the onset of an error (Oomen & Postma, 2002). But if the time it takes to stop were attributable to a comprehension-based loop, then one would have expected that drawing attention away from the speech signal (as in a dual task condition) would have led to longer, not shorter stopping times.

Hartsuiker and Kolk (2001) criticized the Perceptual Loop Theory assumption that production of the erroneous utterance must stop before planning of the replacement begins. Instead, they proposed that stopping the current utterance and preparing a replacement can proceed in parallel. Indeed, speakers can often resume very quickly following an interruption (in less than 100ms; Blackmer & Mitton, 1991), which suggests that they start planning the replacements before they stop articulation. More direct evidence comes from Hartsuiker, Catchpole, de Jong, and Pickering (2008; see also Gambi, Cop, & Pickering, 2015), who showed that the time it takes to stop a word depends on how difficult it is to prepare a replacement word.

However, Hartsuiker and Kolk (2001) also assume that the monitor needs to detect an error in the phonological representation before sending a signal to stop production. Interestingly, there are also psycholinguistic theories of self-monitoring that posit a purely production-based monitor (see Postma, 2000 for discussion). For example, Nozari, Dell, and Schwartz (2011) proposed that error detection is based on the amount of noise associated with production processes. One argument in favor of production-based accounts is that they allow for very rapid error detection at all levels of the linguistic hierarchy. Below, we introduce Pickering and Garrod's (2013, 2014) comprehension-based theory of self-monitoring theory, which addresses this issue by allowing the monitor to compare expected and actual comprehension representations at all linguistic levels, as soon as these representations become available.

Prediction during production and comprehension: The integrated theory of language production and comprehension

Pickering and Garrod (2013) described an integrated theory of language production and comprehension that is based on the notion of forward models. This notion is derived from the motor control literature (e.g., Wolpert, 1997) and is also part of the models reviewed in *Neuro-computational models of speech motor control*. Crucially, Pickering and Garrod generalize it, by making the assumption that forward models are computed at all levels of the linguistic hierarchy, and that they are involved not only in language production but also in comprehension. Below, we first describe how forward models are implicated in self-monitoring, and then how they are implicated in prediction during comprehension.

During an act of production, the speaker forms a communicative intention (production command), which corresponds to the pre-linguistic message that the speaker intends to convey. The production command is sent to the production system (or, in Pickering and Garrod's terminology, the production implementer), and it triggers the retrieval of a set of production representations (semantics, syntax, and phonology). For example, if a speaker sees a kite and forms the intention to name this object, production processes would cause the retrieval of the corresponding concept (KITE), lemma (*kite*) and phonological form (/kaIt/). Importantly, it takes several hundred milliseconds to retrieve such representations (see Indefrey & Levelt, 2004, for estimates). Once production representations have been retrieved, they can be processed by the comprehension system (or, in Pickering and Garrod's terminology, the comprehension implementer).[1] Crucially, the theory assumes that the comprehension implementer has immediate access to production representations at all levels; so, for example, the semantic representation retrieved during production can be immediately comprehended, even before a phonological representation is built. In this respect, the proposal differs from the Perceptual Loop Theory of self-monitoring (Levelt, 1983).

In addition to retrieval of representations within the production and comprehension implementers, during an act of production a copy of the production command is sent to a forward model, which maps from the production command to the predicted comprehension representations that are about to be retrieved as a consequence of executing that production command. To return to our example of a speaker intending to name the picture of a kite, a forward model of this process could compute a prediction of aspects of the semantics (it's a flyable object), of the syntax (it's a noun),[2] and the phonology (it starts with a consonant), *before* the corresponding production representations are retrieved from memory. Therefore, predicted representations are typically ready before actual (implemented) representations. The process of self-monitoring constitutes the comparison between predicted and actual comprehension representations within the comparator (at any linguistic level). The resulting difference (the prediction error in motor control terms; Wolpert, 1997) can be used to drive online corrections (and learning), just as it can in the models described in section *Neuro-computational models of speech motor control*.

In sum, the account of self-monitoring proposed by Pickering and Garrod (2013, 2014) differs from the Perceptual Loop Theory (Levelt, 1983) in that it posits loops between production and comprehension at all levels of the linguistic hierarchy, and both within and between levels (not just from phonetics and phonology to semantics; see dotted arrows in Figure 7.1b). Moreover, such loops are faster than the loops assumed by the Perceptual Loop Theory, because they are based on comparisons between predicted and actual comprehension representations, with predicted comprehension representations being the outcome of production processes (i.e., left-to-right dotted arrows in Figure 7.1b).

With regard to acts of comprehension, Pickering and Garrod (2007) proposed that the collection of cognitive mechanisms underlying prediction during language comprehension coincides with the language production system. Federmeier (2007) made a similar proposal based on evidence that the left hemisphere is more sensitive to predictability of upcoming words than the right hemisphere (and the neural substrate for language production is predominantly left-lateralized in Broca's area), and Dell and Chang (2014) have recently reinstated this idea as the core principle of their P-chain framework (see *The P-chain framework*, this chapter).

An earlier proposal (Kempen, 2000, 2014) argued that grammatical encoding (production) and grammatical decoding (comprehension) are performed by the same processing architecture. Parallels between sentence comprehension and sentence production (e.g., similar patterns of errors occur during subject-verb agreement in both production and comprehension; Bock & Miller, 1991; Pearlmutter, Garnsey, & Bock, 1999), as well as evidence from structural priming from comprehension to production (Bock *et al.*, 2007; Branigan *et al.*, 2000) are consistent with this proposal. However, such evidence is indirect and can be explained by shared representations without shared processes. In addition, while it is clear how production and comprehension could share some processes (e.g., retrieving syntactic frames), it also necessary to assume some processing differences in order to explain the different start- and end-points. Note that the proposal that production processes are related to only a subset of comprehension processes, namely those involved in prediction during acts of comprehension, is not subject to this criticism.

But is there evidence for this hypothesis? Some evidence that production processes might be used during syntactic aspects of comprehension comes from a study by Kempen, Olsthoorn, and Sprenger (2012). They asked Dutch participants to paraphrase sentences from direct (e.g., *De lottowinnaar/zei:/"Ik/heb besloten/een rode auto/te kopen/voor mezelf*," The lottery winner said: "I have decided to buy a red car for myself") into indirect speech (e.g., *De lottowinnaar zei dat hij had besloten een rode auto te kopen voor zichzelf*, The lottery winner said that he had decided to buy a red car for himself) as they read them (i.e., fragment by fragment, as marked in the example). When the sentence contained an ungrammatical reflexive pronoun (e.g., the third-person reflexive pronoun in the sentence *De lottowinnaar zei: "Ik heb besloten een rode auto te kopen voor zichzelf"*), participants were faster producing the paraphrase (which contained the same third-person pronoun) than when the sentence contained a grammatical reflexive (i.e., *mezelf*), despite the fact that the ungrammaticality should have led to a processing delay in comprehension. One interpretation of these findings is that participants' expectations generated during the comprehension of the input sentences were replaced, on-line, by the expectations generated during concurrent encoding of the paraphrase (in which the same pronoun was grammatical).

Moreover, Federmeier, Kutas, and Schul (2010) showed that a late prefrontal positivity induced by plausible but unexpected nouns (which is thought to index the updating of disconfirmed predictions after an unexpected word has been encountered; Federmeier, Wlotko, De Ochoa-Dewald, & Kutas, 2007) is greatly reduced in older adults (compared to younger adults) and, importantly, the magnitude of this component in the older group correlated with production measures of verbal fluency. Similarly, Mani and Huettig (2012) found that two-year-olds with larger production (but not comprehension) vocabularies were more likely to predict upcoming referents (as indexed by looks to corresponding pictures in the so-called visual world paradigm; cf. Altmann & Kamide, 1999). More recently, a similar correlation between verbal fluency and prediction abilities during language comprehension (again, measured using the visual world paradigm) was reported for young adults as well, but only when listeners could preview pictures in the visual display (and presumably started retrieving their names; Hintz, Meyer, & Huettig, 2014). These studies suggest that the ability to predict during language comprehension is correlated with language production abilities at least in some task contexts.

In accordance with this and related evidence, Pickering and Garrod (2013) proposed that production processes underline comprehenders' ability to predict what another is about to produce. This route to comprehension is termed prediction-by-simulation. It starts with the comprehender covertly imitating the producer: this means that, based on the initial part of the producer's utterance, or on contextual information (i.e., what he assumes about the producer from previous interactions, or from background knowledge), the comprehender recovers the most likely intention (production command) underlying the utterance at time *t*. He can then run this command through the production implementer. If he does that, he will end up imitating the producer. This mechanism therefore explains alignment (Pickering & Garrod, 2004; see *Representational parity*).

In addition, the comprehender can run ahead the command that he recovered at time t, thus predicting what he would be likely to utter next if he were in the producer's shoes. If he runs this new command through the production implementer, he will be able to complete the producer's utterance (see *Dialogue*). If he runs this new command through the forward production model, the comprehender may generate the predicted semantics, syntax, and phonology at time $t+1$. When the producer continues his utterance, the comprehender builds comprehension representations for the actual utterance at $t+1$ and can compare them to the representations he had predicted. He can then use the resulting discrepancy to adjust the recovered production command, thus revising his understanding of the intention underlying the producer's utterance.

In addition to the correlational evidence cited above, one study established a causal link between production processes and prediction during language comprehension (Lesage, Morgan, Olson, Meyer, & Miall, 2012). This study applied repetitive Transcranial Magnetic Stimulation (rTMS) to the right cerebellum. In rTMS, as in other types of TMS, a magnetic coil is used to induce small electric currents in a particular area of the brain; in particular, several pulses are delivered at a low frequency, which is known to suppress neural activity for some time after stimulation has ended. Importantly, evidence suggests that forward model computations related to motor execution take place in the cerebellum (e.g., Wolpert, Miall, & Kawato, 1998), and some have linked it to the computation of internal models in general (e.g., Ito, 2008). Disrupting activity in the right cerebellum caused participants in this study to delay their eye-movements to predictable visual referents during sentence comprehension (but not to unpredictable visual referents). Crucially, other conditions with no stimulation or stimulation to a control site did not show the same selective effect. Therefore, this study suggests that forward model computations might support prediction during comprehension.

Note that, because the forward model is functionally distinct from the production implementer, Pickering and Garrod's (2013) account does not claim that full activation of the production implementer will be observed whenever prediction-by-simulation is used. For example, the account does not predict that activation of language production areas in the brain will be always observed during language comprehension. Rather, such activation is most likely to occur under conditions in which production is relied upon more. According to Pickering and Garrod, this is the case when comprehension is difficult (e.g., in a noisy environment; see Adank, 2012).

But in addition, there is another route to prediction available in comprehension, which they termed prediction-by-association. This route does not involve production processes (i.e., forward models or the production implementer), and could be used whenever covert imitation fails. This route to prediction in comprehension makes use of regularities in the input to the process of comprehension. Unlike prediction-by-simulation, it does not rely on knowledge of how we produce language. Instead, it relies on our ability to learn regular patterns of perceptual events, which applies equally to domains in which we have the ability to generate the patterns through action and domains in which we lack this ability (e.g., predicting the sound of leaves moved by the wind).

In sum, Pickering and Garrod's (2013) account allows for the existence of processes that are not (usually) shared between acts of comprehension and acts of production. In other words, there are some production processes that do not always operate during acts of comprehension (e.g., processes involved in retrieving articulatory programs; prediction-by-simulation via forward model computations), and there are comprehension processes that do not operate during acts of production (e.g., the prediction-by-association route).

The P-chain framework

The P-chain framework (Dell & Chang, 2014) claims that the process responsible for prediction during acts of comprehension is the same process that is used during acts of production. Therefore, although it does not claim that all processes are shared between comprehension and production, it posits that there is a single cognitive architecture subserving both tasks.

This assumption stems from the architecture of Chang, Dell, and Bock's (2006) Dual Path model of sentence acquisition (in children) and structural priming (in children and adults). This model is a recurrent neural network; during training, the model learns to predict the next word in the input (as in Elman, 1990). Prediction errors are generated by comparing the predicted with the actual comprehended input, and are used to change the weights between units in the network, so that the model learns to correctly predict grammatical word sequences in the future. Sometimes the model uses meaning (inferred from context) to help in this prediction process and the ability to do prediction from meaning is the same mechanism that the model uses for production.[3]

Dell and Chang (2014) proposed that the model's architecture instantiates a set of principles that govern the functioning of the cognitive system for language. They termed this collection of principles the *P-chain framework*. The framework highlights the tight links existing between language comprehension (which they term *processing*) and *production*. Such links are organized in a chain, or loop, of cause-effect relations. In a nutshell, the use of production-based prediction mechanisms during language comprehension generates prediction errors, which in turn drive changes in the language system (i.e., they make it more likely to generate predictions in line with previous input), thus providing an explanation for structural priming effects, that is for the fact that comprehending a given sentence structure makes it more likely that the same structure will be selected in a subsequent act of production. When these changes build up over time, they can explain acquisition of structural representations for different languages (English: Twomey, Chang, & Ambridge, 2014; Japanese: Chang, 2009; German: Chang, Baumann, Pappert, & Fitz, 2015). Finally, input regularities, on which comprehension is tuned, are themselves the output of the mechanism responsible for language production (in other speakers); therefore, there is also a slow-cycling loop through which the production processes of other speakers provide the input that trains comprehension processes. This final link is inspired by the Production-Distribution-Comprehension account (MacDonald, 2013) that is examined in the next section.

Frameworks that posit linked preferences

Several theorists have appealed to the idea that regular patterns in language use emerge as a consequence of the constraints imposed by communication, and that such patterns in turn affect how speakers and listeners process language. This approach corresponds to positing a long-term loop (dashed arrow in Figure 7.1b) that is external to the language production and comprehension architecture (i.e., outside the mind/brain of individual language users).

One version of this idea is that speaker choices in production are constrained by ease-of-comprehension principles. For example, the Hyper- And Hypo-Articulation model of speech production (Lindblom, 1990) claims that speakers' tendency to reduce articulation (and therefore their own effort) is constrained by the necessity for listeners to recover the intended message. This leads speakers to counteract the tendency to hypo-articulate (i.e., to produce forms that are reduced in duration and/or intensity), precisely in those contexts in which the listener cannot draw on other sources of information (i.e., other than the speech signal) to infer meaning. There are other versions of this proposal in phonetics (e.g., the Smooth Signal Redundancy hypothesis, Aylett, & Turk, 2004).

Similarly, the Uniform Information Density hypothesis (UID; Levy & Jaeger, 2007; Jaeger, 2010) claims that producers strive to keep information transfer rate within the range of the comprehender's processing rate (i.e., channel capacity); in this way, they avoid conveying too much information or too little information per unit (word, phoneme, etc.). This hypothesis can also be phrased in terms of surprisal, which is the predictability of a unit in context. High surprisal means that a unit is unpredictable based on previous context, and therefore adds information in that context, whereas low surprisal means that a unit is highly predictable and adds little information. UID claims that when given a choice speakers will tend to use a structure that keeps surprisal relatively constant across units. Production preferences are therefore explained as stemming from the limitations of the comprehension system. Interestingly, these preferences may be shaped by learning: Jaeger and Snider (2013) provided evidence that the higher the surprisal of a structural alternative in comprehension, the more likely that alternative is to be subsequently preferred in production.

The idea that production preferences can be explained with reference to comprehension is also related to the Audience Design Hypothesis: the notion that producers take into account their addressee's knowledge when planning their utterances (e.g., Brown-Schmidt, 2009; Clark, 1996). A review of this literature is beyond the scope of this chapter. However, we note that experimental evidence for the extent to which audience design affects production is mixed. In particular, there is controversy over the rapidity with which the addressee's knowledge can affect production. Keysar and colleagues have argued that producers are egocentric (e.g., Horton & Keysar, 1996), and take into account what their addressee can or cannot know only when given sufficient time and during a relatively late stage of production. In addition, it has been suggested that while speakers might adapt at the level of lexical choices, they in fact do not do so at the phonetic level (Bard *et al.*, 2000; but see Galati & Brennan, 2010 for criticism, and Arnold, Kahn, & Pancani, 2012 for some evidence that speakers might adapt at least at the level of production speed). Finally, addressees are clearly facilitated when speakers adopt previously established referential labels, but we do not know whether such facilitation occurs because listeners are sensitive to mutual knowledge between them and the speaker (Brown-Schmidt, 2009) or simply because repetition increases availability (Barr & Keysar, 2002).

Another related view is the Production-Distribution-Comprehension account (MacDonald, 2013), which assumes that the relevant constraints shaping patterns of language use over time relate to production rather than to comprehension. In particular, it assumes three principles: (i) Easy First: Words that are more easily retrieved are produced first (e.g., Bock & Warren, 1985); (ii) Plan Reuse: Utterance plans that have been recently used tend to be reused (e.g., Bock, 1986); and (iii) Reduce Interference: Elements that are more similar to one another (and therefore tend to interfere in memory) are placed farther apart (Gennari, Mirković, & MacDonald, 2012 for semantic similarity; Jaeger, Furth, & Hilliard, 2012 for phonological similarity). The claim, then, is that these constraints on production lead to distributional regularities to which comprehenders adapt as they accumulate linguistic experience. Therefore, linguistic forms that are easier to produce become easier to comprehend as well. This account thus explains both why certain structures are easier to comprehend than others and cross-linguistic patterns of language variation (i.e., typology). We refer the reader to Chapter 3 in this volume for an in-depth discussion of the evidence in favor and against the frameworks briefly introduced in this section.

Dialogue

As mentioned in the Introduction, the nature of language use in dialogue (i.e., conversation) is one key motivation for positing links between comprehension and production. First, each participant in a dialogue regularly has to switch between acts of production and acts of comprehension. Such switches do not only occur between dialogue turns, but also within a turn, as listeners produce backchannels (e.g., *Yes, OK*, or *eh?*) to provide continuous feedback to the speaker. Moreover, such switches occur rapidly, as long intervals between turns are rare (Stivers *et al.*, 2009). Finally, such switches can occur at any point within an utterance, with listeners taking over from speakers even after single words or incomplete constituents, and sometimes producing grammatical and pragmatically appropriate completions to these fragments (e.g., Clark & Wilkes-Gibbs, 1986; Lerner, 1991).

These phenomena demonstrate that the output of comprehension processes can rapidly affect production processes. For example, understanding the speaker's utterance leads to rapid backchannel responses from the listener and such backchannels can be quickly acted upon by the speaker. This suggests that loops must exist between comprehension processes and production processes, and that these loops must operate at a relatively fast rate. Moreover, the phenomenon of collaborative turn completions also suggests continuity between production and comprehension processes. Take the excerpt below (from Kurtić, Brown, & Wells, 2013, p. 726). B produces his utterance as a completion to the first part of A's utterance (*so I'm not sure whether they'll still be so willing to volunteer but I'll*) and it is so well-timed that it overlaps with the end of A's own turn (brackets indicate the start and end of speech produced in overlap by A and B). In turn, A shows evidence of having understood (and accepting) B's completion, despite it overlapping with her own turn; in fact, she goes as far as repeating B's utterance word by word.

A: …so I'm not sure whether they'll still be so willing to volunteer but I'll [send them an email and ask]
B: [tell them about the free lunch]
A: I'll tell them about the free lunch

Finally, inter-turn intervals tend to cluster around a value that varies between 0 and 200ms (across languages; Stivers *et al.*, 2009), with both long gaps and long overlaps being comparatively rare. This further suggests close links between production and comprehension, and has prompted the suggestion that comprehenders might be able to anticipate turn ends (de Ruiter, Mitterer, & Enfield, 2006; Magyari & de Ruiter, 2012).

Summary

In this chapter, we have described several models that posit explicit links between production and comprehension. It is generally agreed that information about concepts, lemmas, and syntactic frames is shared between production and comprehension processes. But opinions diverge on the degree of sharing of phonological information, and most theorists assume that phonetic representations are separate (despite some dissent). Both the language production and the language comprehension literatures have internal debates about the directionality of processes, but such debates have not been framed as debates about the sharing of processes between production and comprehension until quite recently.

In Figure 7.1a, we assumed that any process that maps from "higher" linguistic levels (semantics) to "lower" linguistic levels (phonology) should be named a production process (left-to-right arrows), and that every process that maps in the inverse direction (right-to-left arrows) should be named a comprehension process. Based on this definition, we identified three accounts that assume comprehension processes take place during acts of production, in the form of self-monitoring: Levelt's (1989) Perceptual Loop theory (solid lines in Figure 7.1b), neuro-computational models of speech motor control (such as Hickok, 2012), and Pickering and Garrod's (2013) integrated account of language production and comprehension (dotted lines in Figure 7.1b). Such accounts differ in terms of the nature and speed of the loops they assume exist between production and comprehension.

We presented two accounts that assume that production processes take place during acts of comprehension to support prediction: Pickering and Garrod's integrated account, and the P-chain framework. In addition, we briefly described a number of frameworks that posit slower-cycling loops between production and comprehension; that is, loops that are mediated by long-term experience with language and that can explain the development of linguistic preferences (dashed lines in Figure 7.1b). Overall, many researchers assume some degree of sharing of processes, but the range of views on this issue is far wider than on the issue of shared representations. Finally, we noted that the assumption of links between language production and language comprehension is also motivated by language use in dialogue. We believe that more explicit theorizing on the relations between production and comprehension processes, in both monologue and dialogue, would benefit the field.

Acknowledgments

We would like to thank Franklin Chang and an anonymous reviewer for useful discussion.

NOTES

1 Note that the description provided by Pickering and Garrod (2013), and summarized here, appears to imply separate production and comprehension representations at all levels. In practice, the theory is consistent with the idea that the same representations are accessed during production as well as during comprehension. The two sets of representations they assume correspond to the output stage of production and comprehension processes respectively, but such processes may have access to the same pool of representations.

2 In the example, we focus on single word retrieval. However, Pickering and Garrod (2013) have also discussed this process in relation to constituent ordering (p. 339).

3 Note that within this single cognitive architecture, the Dual path model incorporates separate weights for the word-to-syntax ("comprehension") and syntax-to-word ("production") directions, but both of these representations are hypothesized to be used in both comprehension and production tasks.

REFERENCES

Adank, P. (2012). The neural bases of difficult speech comprehension and speech production: Two Activation Likelihood Estimation (ALE) meta-analyses. *Brain and Language*, 122(1), 42–54. doi: 10.1016/j.bandl.2012.04.014

Alario, F.X., Segui, J., & Ferrand, L. (2000). Semantic and associative priming in picture naming. *The Quarterly Journal of Experimental Psychology*, 53A(3), 741–764. doi: 10.1080/027249800410535

Altmann, G. T., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, 73(3), 247–264. doi: 10.1016/S0010-0277(99)00059-1

Arnold, J. E., Kahn, J. M., & Pancani, G. C. (2012). Audience design affects acoustic reduction via production facilitation.

Psychonomic Bulletin & Review, 19(3), 505–512. doi: 10.3758/s13423-012-0233-y

Aylett, M., & Turk, A. (2004). The smooth signal redundancy hypothesis: A functional explanation for relationships between redundancy prosodic prominence, and duration in spontaneous speech. *Language and Speech*, 47(1), 31–56. doi: 10.1177/00238309040470010201

Bard, E. G., Anderson, A. H., Sotillo, C., Aylett, M., Doherty-Sneddon, G., & Newlands, A. (2000). Controlling the intelligibility of referring expressions in dialogue. *Journal of Memory and Language*, 42(1), 1–22. doi: 10.1006/jmla.1999.2667

Barr, D. J., & Keysar, B. (2002). Anchoring comprehension in linguistic precedents. *Journal of Memory and Language*, 46(2), 391–418. doi: 10.1006/jmla.2001.2815

Bates, E. (1993). Comprehension and production in early language development: Comments on Savage-Rumbaugh et al. *Monographs of the Society for Research in Child Development, Serial No. 233*, 58*(*3–4*)*, 222–242. doi: 10.1111/j.1540-5834.1993.tb00403.x

Blackmer, E. R., & Mitton, J. L. (1991). Theories of monitoring and the timing of repairs in spontaneous speech. *Cognition*, 39(3), 173–194. doi: 10.1016/0010-0277(91)90052-6

Bock, K., Dell, G. S., Chang, F., & Onishi, K. H. (2007). Persistent structural priming from language comprehension to language production. *Cognition*, 104(3), 437–458. doi: 10.1016/j.cognition.2006.07.003

Bock, K. (1996). Language production: Methods and methodologies. *Psychonomic Bulletin & Review*, 3(4), 395–421. doi: 10.3758/BF03214545

Bock, K., Dell, G. S., Chang, F., & Onishi, K. H. (2007). Persistent structural priming from language comprehension to language production. *Cognition*, 104(3), 437–458. doi: 10.1016/j.cognition.2006.07.003

Bock, K., & Miller, C. A. (1991). Broken agreement. *Cognitive Psychology*, 23(1), 45–93. doi: 10.1016/0010-0285(91)90003-7

Bock, K., & Warren, R. K. (1985). Conceptual accessibility and syntactic structure in sentence formulation. *Cognition*, 21(1), 47–67. doi: 10.1016/0010-0277(85)90023-X

Branigan, H. P., Pickering, M. J., & Cleland, A. A. (2000). Syntactic co-ordination in dialogue. *Cognition*, 75(2), B13–B25. doi: 10.1016/S0010-0277(99)00081-5

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

Brown-Schmidt, S. (2009). Partner-specific interpretation of maintained referential precedents during interactive dialog. *Journal of Memory and Language*, 61(2), 171–190. doi: 10.1016/j.jml.2009.04.003

Chang, E. F., Niziolek, C. A., Knight, R. T., Nagarajan, S. S., & Houde, J. F. (2013). Human cortical sensorimotor network underlying feedback control of vocal pitch. *Proceedings of the National Academy of Sciences*, 110(7), 2653–2658. doi: 10.1073/pnas.1216827110

Chang, F. (2009). Learning to order words: A connectionist model of heavy NP shift and accessibility effects in Japanese and English. *Journal of Memory and Language*, 61*(*3*)*, 374–397. doi: 10.1016/j.jml.2009.07.006

Chang, F., Baumann, M., Pappert, S., & Fitz, H. (2015). Do lemmas speak German? A verb position effect in German structural priming, *Cognitive Science*, 39*(*5*)*, 1113–1130. doi: 10.1111/cogs.12184

Chang, F., Dell, G. S., & Bock, K. (2006). Becoming syntactic. *Psychological Review*, 113(2), 234–272. doi: 10.1037/0033-295X.113.2.234

Clark, H. H. (1996). *Using language*. Cambridge University Press. Cambridge, U.K.

Clark, H. H., & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition*, 22(1), 1–39. doi: 10.1017/cbo9780511620539

Damian, M. F., & Martin, R. C. (1999). Semantic and phonological codes interact in single word production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 25(2), 345–361. doi: 10.1037/0278-7393.25.2.345

D'Ausilio, A., Bufalari, I., Salmas, P., & Fadiga, L. (2012). The role of the motor system in discriminating normal and degraded speech sounds. *Cortex*, 48(7), 882–887. doi: 10.1016/j.cortex.2011.05.017

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., & Chang, F. (2014). The P-chain: relating sentence production and its disorders to comprehension and acquisition. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369(1634), 20120394. doi: 10.1098/rstb.2012.0394

De Ruiter, J. P., Mitterer, H., & Enfield, N. J. (2006). Projecting the end of a speaker's turn: A cognitive cornerstone of conversation. *Language*, 82*(3)*, 515–535. doi: 10.1353/lan.2006.0130

Elman, J. L. (1990). Finding structure in time. *Cognitive Science*, 14(2), 179–211. doi: 10.1207/s15516709cog1402_1

Evans, B. G., & Iverson, P. (2007). Plasticity in vowel perception and production: A study of accent change in young adults. *The Journal of the Acoustical Society of America*, 121(6), 3814–3826. doi: 10.1121/1.2722209

Fadiga, L., Craighero, L., Buccino, G., & Rizzolatti, G. (2002). Speech listening specifically modulates the excitability of tongue muscles: a TMS study. *European Journal of Neuroscience*, 15(2), 399–402. doi: 10.1046/j.0953-816x.2001.01874.x

Federmeier, K. D. (2007). Thinking ahead: The role and roots of prediction in language comprehension. *Psychophysiology*, 44(4), 491–505. doi: 10.1111/j.1469-8986.2007.00531.x

Federmeier, K. D., Kutas, M., & Schul, R. (2010). Age-related and individual differences in the use of prediction during language comprehension. *Brain and Language*, 115(3), 149–161. doi: 10.1016/j.bandl.2010.07.006

Federmeier, K. D., Wlotko, E. W., De Ochoa-Dewald, E., & Kutas, M. (2007). Multiple effects of sentential constraint on word processing. *Brain Research*, 1146, 75–84. doi: 10.1016/j.brainres.2006.06.101

Fowler, C. A., Brown, J. M., Sabadini, L., & Weihing, J. (2003). Rapid access to speech gestures in perception: Evidence from choice and simple response time tasks. *Journal of Memory and Language*, 49(3), 396–413. doi: 10.1016/S0749-596X(03)00072-X

Galantucci, B., Fowler, C. A., & Goldstein, L. (2009). Perceptuomotor compatibility effects in speech. *Attention, Perception, & Psychophysics*, 71(5), 1138–1149. doi: 10.3758/APP.71.5.1138

Galantucci, B., Fowler, C. A., & Turvey, M. T. (2006). The motor theory of speech perception reviewed. *Psychonomic Bulletin & Review*, 13(3), 361–377. doi: 10.3758/BF03193857

Galati, A., & Brennan, S. E. (2010). Attenuating information in spoken communication: For the speaker, or for the addressee? *Journal of Memory and Language*, 62(1), 35–51.

Gambi, C., Cop, U., & Pickering, M. J. (2015). How do speakers coordinate? Evidence for prediction in a joint word-replacement task. *Cortex*, 68, 111–128, doi: 10.1016/j.cortex.2014.09.009

Garrod, S., & Anderson, A. (1987). Saying what you mean in dialogue: A study in conceptual and semantic co-ordination. *Cognition*, 27(2), 181–218. doi: 10.1016/0010-0277(87)90018-7

Garrod, S., & Pickering, M. J. (2004). Why is conversation so easy? *Trends in Cognitive Sciences*, 8(1), 8–11. doi: 10.1016/j.tics.2003.10.016

Gennari, S. P., Mirković, J., & MacDonald, M. C. (2012). Animacy and competition in relative clause production: A cross-linguistic investigation. *Cognitive*

Psychology, 65(2), 141–176. doi: 10.1016/j.cogpsych.2012.03.002

Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review*, 105(2), 251–279. doi: 10.1037/0033-295X.105.2.251

Guenther, F. H., & Vladusich, T. (2012). A neural theory of speech acquisition and production. *Journal of Neurolinguistics*, 25(5), 408–422. doi: 10.1016/j.jneuroling.2009.08.006

Hartsuiker, R. J., & Kolk, H. H. (2001). Error monitoring in speech production: A computational test of the perceptual loop theory. *Cognitive Psychology*, 42(2), 113–157. doi: 10.1006/cogp.2000.0744

Hartsuiker, R. J., Catchpole, C. M., de Jong, N. H., & Pickering, M. J. (2008). Concurrent processing of words and their replacements during speech. *Cognition*, 108(3), 601–607. doi: 10.1016/j.cognition.2008.04.005

Hickok, G. (2012). Computational neuroanatomy of speech production. *Nature Reviews Neuroscience*, 13(2), 135–145. doi: 10.1038/nrn3158

Hickok, G. (2014). Towards an integrated psycholinguistic, neurolinguistic, sensorimotor framework for speech production. *Language, Cognition and Neuroscience*, 29(1), 52–59. doi: 10.1080/01690965.2013.852907

Hintz, F., Meyer, A.S., & Huettig, F. (2014). The influence of verb-specific featural restrictions, word associations, and production-based mechanisms on language-mediated anticipatory eye-movements. *Proceedings of the 27th Annual CUNY conference on human sentence processing*. Ohio State University. Columbus, OH.

Horton, W. S., & Keysar, B. (1996). When do speakers take into account common ground?. *Cognition*, 59(1), 91–117. doi: 10.1016/0010-0277(96)81418-1

Houde, J. F., & Jordan, M. I. (1998). Sensorimotor adaptation in speech production. *Science*, 279(5354), 1213–1216. doi: 10.1126/science.279.5354.1213

Houde, J. F., Nagarajan, S. S., Sekihara, K., & Merzenich, M. M. (2002). Modulation of the auditory cortex during speech: an MEG study. *Journal of Cognitive Neuroscience*, 14(8), 1125–1138. doi: 10.1162/089892902760807140

Indefrey, P., & Levelt, W. J. (2004). The spatial and temporal signatures of word production components. *Cognition*, 92(1), 101–144. doi: 10.1016/j.cognition.2002.06.001

Ito, M. (2008). Control of mental activities by internal models in the cerebellum. *Nature Reviews Neuroscience*, 9(4), 304–313. doi: 10.1038/nrn2332

Jarick, M., & Jones, J. A. (2008). Observation of static gestures influences speech production. *Experimental Brain Research*, 189(2), 221–228. doi: 10.1007/s00221-008-1416-7

Jaeger, F.T. (2010). Redundancy and reduction: Speakers manage syntactic information density. *Cognitive Psychology*, 61(1), 23–62. doi: 10.1016/j.cogpsych.2010.02.002

Jaeger, T. F., Furth, K., & Hilliard, C. (2012). Incremental phonological encoding during unscripted sentence production. *Frontiers in Psychology*, 3, doi: 10.3389/fpsyg.2012.00481

Jaeger, T. F., & Snider, N. E. (2013). Alignment as a consequence of expectation adaptation: Syntactic priming is affected by the prime's prediction error given both prior and recent experience. *Cognition*, 127(1), 57–83. doi: 10.1016/j.cognition.2012.10.013

Jones, J. A., & Munhall, K. G. (2002). The role of auditory feedback during phonation: studies of Mandarin tone production. *Journal of Phonetics*, 30(3), 303–320. doi: 10.1006/jpho.2001.0160

Kempen, G. (2000). Could grammatical encoding and grammatical decoding be subserved by the same processing module? *Behavioral and Brain Sciences*, 23(1), 38–39. doi: 10.1017/S0140525X00402396

Kempen, G. (2014). Prolegomena to a neurocomputational architecture for human grammatical encoding and decoding. *Neuroinformatics*, 12(1), 111–142. doi: 10.1007/s12021-013-9191-4

Kempen, G., Olsthoorn, N., & Sprenger, S. (2012). Grammatical workspace sharing during language production and language comprehension: Evidence from grammatical multitasking. *Language and Cognitive Processes*, 27(3), 345–380. doi: 10.1080/01690965.2010.544583

Kerzel, D., & Bekkering, H. (2000). Motor activation from visible speech: evidence from stimulus response compatibility. *Journal of Experimental Psychology: Human Perception and Performance*, 26(2), 634–647. doi: 10.1037/0096-1523.26.2.634

Kim, M., & Thompson, C. K. (2000). Patterns of comprehension and production of nouns and verbs in agrammatism: Implications for lexical organization. *Brain and Language*, 74(1), 1–25. doi: 10.1006/brln.2000.2315

Kraljic, T., Brennan, S. E., & Samuel, A. G. (2008). Accommodating variation: Dialects, idiolects, and speech processing. *Cognition*, 107(1), 54–81. doi: 10.1016/j.cognition.2007.07.013

Kurtić, E., Brown, G. J., & Wells, B. (2013). Resources for turn competition in overlapping talk. *Speech Communication*, 55(5), 721–743. doi: 10.1016/j.specom.2012.10.002

Lerner, G. H. (1991). On the syntax of sentences-in-progress. *Language in Society*, 20(03), 441–458. doi: 10.1017/S0047404500016572

Lesage, E., Morgan, B. E., Olson, A. C., Meyer, A. S., & Miall, R. C. (2012). Cerebellar rTMS disrupts predictive language processing. *Current Biology*, 22(18), R794–R795. doi: 10.1016/j.cub.2012.07.006

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

Levelt, W. J., & Kelter, S. (1982). Surface form and memory in question answering. *Cognitive Psychology*, 14(1), 78–106. doi: 10.1016/0010-0285(82)90005-6

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

Levy, R., & Jaeger, T. F. (2007). Speakers optimize information density through syntactic reduction. In B. Schöllkopf, J. Platt, and T. Hoffman (Eds.). *Advances in Neural Information Processing Systems*, Vol. 19 (pp. 849–856), Cambridge, MA: MIT Press.

Liberman, A. M., & Whalen, D. H. (2000). On the relation of speech to language. *Trends in Cognitive Sciences*, 4(5), 187–196. doi: 10.1016/S1364-6613(00)01471-6

Lindblom, B. (1990). Explaining phonetic variation: A sketch of the H&H theory. In W.J. Hardcastle and A. Marchal (Eds.). *Speech Production and Speech Modelling* (pp. 403–439). Springer Netherlands.

MacDonald, M. C. (2013). How language production shapes language form and comprehension. *Frontiers in Psychology*, 4, doi:10.3389/fpsyg.2013.00226

MacKay, D. G. (1987). *The organization of perception and action: A theory for language and other cognitive skills*. New York: Springer-Verlag.

MacKay, D. G., Abrams, L., & Pedroza, M. J. (1999). Aging on the input versus output side: Theoretical implications of age-linked asymmetries between detecting versus retrieving orthographic information. *Psychology and Aging*, 14(1), 3–17. doi: 10.1037/0882-7974.14.1.3

Magyari, L., & De Ruiter, J. P. (2012). Prediction of turn-ends based on anticipation of upcoming words. *Frontiers in Psychology*, 3, doi:10.3389/fpsyg.2012.00376

Mani, N., & Huettig, F. (2012). Prediction during language processing is a piece of cake—But only for skilled producers.

Journal of Experimental Psychology: Human Perception and Performance, 38(4), 843–847. doi: 10.1037/a0029284

Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word-recognition. *Cognition*, 25(1), 71–102. doi: 10.1016/0010-0277(87)90005-9

Menenti, L., Gierhan, S. M., Segaert, K., & Hagoort, P. (2011). Shared language overlap and segregation of the neuronal infrastructure for speaking and listening revealed by functional MRI. *Psychological Science*, 22(9), 1173–1182. doi: 10.1177/0956797611418347

Mitterer, H., & Ernestus, M. (2008). The link between speech perception and production is phonological and abstract: Evidence from the shadowing task. *Cognition*, 109(1), 168–173. doi: 10.1016/j.cognition.2008.08.002

Mitterer, H., & Müsseler, J. (2013). Regional accent variation in the shadowing task: Evidence for a loose perception–action coupling in speech. *Attention, Perception, & Psychophysics*, 75(3), 557–575. doi: 10.3758/s13414-012-0407-8

Niziolek, C. A., Nagarajan, S. S., & Houde, J. F. (2013). What does motor efference copy represent? Evidence from speech production. *The Journal of Neuroscience*, 33(41), 16110–16116. doi: 10.1523/JNEUROSCI.2137-13.2013

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), 1–33. doi: 10.1016/j.cogpsych.2011.05.001

Oomen, C. C., & Postma, A. (2002). Limitations in processing resources and speech monitoring. *Language and Cognitive Processes*, 17(2), 163–184. doi: 10.1080/01690960143000010

Pardo, J. S. (2006). On phonetic convergence during conversational interaction. *The Journal of the Acoustical Society of America*, 119(4), 2382–2393. doi: 10.1121/1.2178720

Pearlmutter, N. J., Garnsey, S. M., & Bock, K. (1999). Agreement processes in sentence comprehension. *Journal of Memory and Language*, 41(3), 427–456. doi: 10.1006/jmla.1999.2653

Pickering, M. J., & Ferreira, V. S. (2008). Structural priming: a critical review. *Psychological Bulletin*, 134(3), 427–459. doi: 10.1037/0033-2909.134.3.427

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, 27(02), 169–190. doi: 10.1017/S0140525X04000056

Pickering, M. J., & Garrod, S. (2007). Do people use language production to make predictions during comprehension?. *Trends in Cognitive Sciences*, 11(3), 105–110. doi: 10.1016/j.tics.2006.12.002

Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences*, 36(04), 329–347. doi: 10.1017/S0140525X12001495

Pickering, M. J., & Garrod, S. (2014). Self-, other-, and joint monitoring using forward models. *Frontiers in Human Neuroscience*, 8, doi: 10.3389/fnhum.2014.00132

Plaut, D. C., & Kello, C. T. (1999). The emergence of phonology from the interplay of speech comprehension and production: A distributed connectionist approach. In B. MacWhinney (Ed.). *The Emergence of Language* (pp. 381–415). Mahwah, NJ: Taylor and Francis.

Postma, A. (2000). Detection of errors during speech production: A review of speech monitoring models. *Cognition*, 77(2), 97–132. doi: 10.1016/S0010-0277(00)00090-1

Pulvermüller, F., & Fadiga, L. (2010). Active perception: sensorimotor circuits as a cortical basis for language. *Nature Reviews Neuroscience*, 11(5), 351–360. doi: 10.1038/nrn2811

Pulvermüller, F., Huss, M., Kherif, F., del Prado Martin, F. M., Hauk, O., & Shtyrov, Y. (2006). Motor cortex maps articulatory features of speech sounds. *Proceedings of the National Academy of Sciences*, 103(20), 7865–7870. doi: 10.1073/pnas.0509989103

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

Scott, S. K., McGettigan, C., & Eisner, F. (2009). A little more conversation, a little less action—candidate roles for the motor cortex in speech perception. *Nature Reviews Neuroscience*, 10(4), 295–302. doi: 10.1038/nrn2603

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(7), 1662–1670. doi: 10.1093/cercor/bhr249

Skipper, J. I., Nusbaum, H. C., & Small, S. L. (2005). Listening to talking faces: motor cortical activation during speech perception. *Neuroimage*, 25(1), 76–89. doi: 10.1016/j.neuroimage.2004.11.006

Skipper, J. I., van Wassenhove, V., Nusbaum, H. C., & Small, S. L. (2007). Hearing lips and seeing voices: how cortical areas supporting speech production mediate audiovisual speech perception. *Cerebral Cortex*, 17(10), 2387–2399. doi: 10.1093/cercor/bhl147

Stivers, T., Enfield, N. J., Brown, P., Englert, C., Hayashi, M., Heinemann, T., … & Levinson, S. C. (2009). Universals and cultural variation in turn-taking in conversation. *Proceedings of the National Academy of Sciences*, 106(26), 10587–10592. doi: 10.1073/pnas.0903616106

Sumner, M., & Samuel, A. G. (2009). The effect of experience on the perception and representation of dialect variants. *Journal of Memory and Language*, 60(4), 487–501. doi: 10.1016/j.jml.2009.01.001

Tourville, J. A., & Guenther, F. H. (2011). The DIVA model: A neural theory of speech acquisition and production. *Language and Cognitive Processes*, 26(7), 952–981. doi: 10.1080/01690960903498424

Tourville, J. A., Reilly, K. J., & Guenther, F. H. (2008). Neural mechanisms underlying auditory feedback control of speech. *Neuroimage*, 39(3), 1429–1443. doi: 10.1016/j.neuroimage.2007.09.054

Twomey, K., Chang, F., & Ambridge, B. (2014). Do as I say, not as I do: A lexical distributional account of English locative verb class acquisition. *Cognitive Psychology*, 73, 41–71. doi: 10.1016/j.cogpsych.2014.05.001

Wilson, S. M., Saygin, A. P., Sereno, M. I., & Iacoboni, M. (2004). Listening to speech activates motor areas involved in speech production. *Nature Neuroscience*, 7(7), 701–702. doi: 10.1038/nn1263

Wolpert, D. M. (1997). Computational approaches to motor control. *Trends in Cognitive Sciences*, 1(6), 209–216. doi: 10.1016/S1364-6613(97)01070-X

Wolpert, D. M., Miall, R. C., & Kawato, M. (1998). Internal models in the cerebellum. *Trends in Cognitive Sciences*, 2(9), 338–347. doi: 10.1016/S1364-6613(98)01221-2

Ylinen, S., Nora, A., Leminen, A., Hakala, T., Huotilainen, M., Shtyrov, Y., & Mäkelä, J. P. (2014). Two distinct auditory-motor circuits for monitoring speech production as revealed by content-specific suppression of auditory cortex. *Cerebral Cortex*, 25(6), 1576–1586. doi: 10.1093/cercor/bht351

Yuen, I., Davis, M. H., Brysbaert, M., & Rastle, K. (2010). Activation of articulatory information in speech perception. *Proceedings of the National Academy of Sciences*, 107(2), 592–597. doi: 10.1073/pnas.0904774107

Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological bulletin*, 123(2), 162–185. doi: 10.1037/0033-2909.123.2.162