CHAPTER

18

Successful Speaking: Cognitive Mechanisms of Adaptation in Language Production

Gary S. Dell and Cassandra L. Jacobs

Beckman Institute, University of Illinois, Urbana, IL, USA

The language production system works. If a person is older than the age of 4, has no major brain pathology, and has been exposed to linguistic input that accords with their perceptual and motor abilities, then they will have developed a production system that transmits what they want to say. It works when the goal is only to say "hi," and when the speaker attempts to communicate a complicated novel thought that takes several sentences to convey.

Successful linguistic communication is achieved by a division of labor between the speaker and the listener (Ferreira, 2008). Both the production and comprehension systems have to do their job. The speaker has to say something apt and understandable, and the listener must do the rest, which can include compensating for any of the speaker's errors or other infelicities.

In this chapter, we focus on how the production system keeps up its end so that the listener is not overly burdened. Our central claim is that the production system benefits from a number of what we call *speaker tuning mechanisms*. Speaker tuning mechanisms are properties of the system that adapt it to current circumstances and to circumstances that are generally more likely. These include *implicit learning* mechanisms that create long-term adaptive changes in the production system, and a variety of short-term adaptive devices, including *error monitoring*, *availability-based retrieval*, *information-density sensitivity*, and, finally, *audience design*. Although we characterize these mechanisms in cognitive rather than neural terms, we include some pointers to relevant neurobiological data and mechanisms. In the following, we describe the production system generally and then focus on the long-term and then short-term speaker tuning mechanisms.

18.1 LANGUAGE PRODUCTION

The production system turns thoughts into sequences of words, which can be spoken aloud, inwardly spoken, or written down. Traditionally (Levelt, 1989), the production process consists of determining the semantic content of one's utterance (*conceptualization*), translating that content into linguistic form (*formulation*), and *articulation*, as illustrated in Figure 18.1. Here, we focus on the second of these stages, which describes how intended meaning, sometimes called the *message*, is turned into an ordered set of words that are specified for their phonological content. That is, the formulation stage describes how CHASE (CAT1, RAT1, past) becomes /ðə.kæt′.čest′.ðə.ræt′/. (The "1" in "CAT1" represents a particular definite CAT). Much of the psycholinguistic and neuroscience research on formulation has concerned three subprocesses: (i) *lexical access*, the retrieval of appropriate words; (ii) *grammatical encoding*, the specification of the order and grammatical forms of those words; and (iii) *phonological encoding*, determining the pronunciation of the sequence of words. These are discussed in turn.

18.1.1 Lexical Access

Most experimental, clinical, and theoretical research on production has concerned lexical access and focuses on the production of single-word utterances. When given a picture that has been identified as the concept CAT, how does the speaker retrieve the word "cat"? Lexical access has been characterized as a two-step process (Garrett, 1975). First, the concept is mapped

 DOI: http://dx.doi.org/10.1016/B978-0-12-407794-2.00018-3

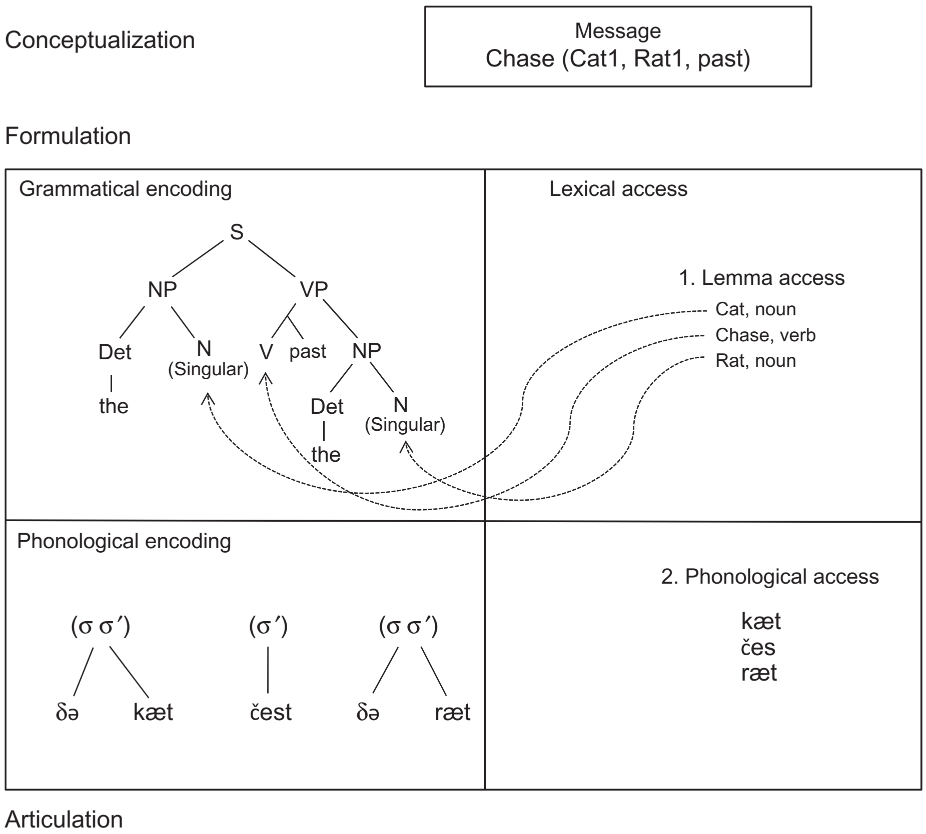

FIGURE 18.1 Components of the language production system.

onto an abstract lexical representation, variously called the *lemma*, the *L-level* representation, or simply the *word node*. This abstraction identifies the grammatical properties of the word such as its syntactic category (e.g., noun) and other grammatically relevant features (e.g., number, grammatical gender). Importantly, this level does not specify anything about pronunciation. That comes in the second step, where the word's phonological form, most often viewed as a sequence of phonemes, is retrieved. Intuitive support for the two-step notion comes from speech errors (Fromkin, 1971). Slips can profitably be divided into those that might have arisen during the first step (e.g., semantic errors such as "dog" for "cat") and those that could have happened in the second step (e.g., "cap" for "cat"). Furthermore, the tip-of-the tongue state ("I know that word! It's on the tip of my tongue") can be characterized as getting stuck between the steps.

Much of the research on lexical access has concerned just how separate the two steps are. For example, the *modular discrete-step* view states that the first step must be completed before the second step can begin (Levelt, Roelofs, & Meyer, 1999). Alternatively, one could allow for *cascading*, which blurs the boundaries between the steps by allowing for phonological properties of potential word candidates to be retrieved before a single lexical item has been settled on in the first step. Or, one could allow for *interaction*, which blurs the steps even further by allowing for relevant representations at each step to influence one another through the interactive spread of activation (see Dell, Nozari, & Oppenheim, 2014 for a recent review of the evidence for interaction between the steps, and see Dell, Schwartz, Nozari, Faseyitan, & Coslett, 2013 and Ueno, Saito, Rogers, & Lambon Ralph, 2011 for proposals regarding the neural correlates of the steps).

18.1.2 Grammatical Encoding

Although most production research concerns single-word utterances, the hallmark of production is the ability to construct multiword utterances, particularly those that the speaker has never said or even heard before. For example, William Blake famously used the phrase, "fearful symmetry" to characterize the tiger. And it is not just the poets who are linguistically inventive. Since Chomsky (1959) emphasized the creativity of language, it is a psycholinguistic cliché that most of what speakers say is novel. Regardless of whether this claim is strictly true, there is no doubt that theories must explain the production of novel utterances. The usual explanation is that the production system uses syntactic-sequential abstractions that specify how word categories can combine to express structured messages. For Blake's phrase, the relevant abstractions would dictate that, in English, adjectives (*fearful*) precede nouns (*symmetry*). Production models of grammatical encoding (Bock, 1982; Chang, Dell, & Bock, 2006; Kempen & Hoenkamp, 1987) differ considerably, but all recognize the distinction between categorically specified abstractions and lexical items. Typically, the abstractions are characterized as *frames*, that is, structures that specify the sequence and phrasal membership of syntactically categorized word-sized slots (Dell, 1986; Garrett, 1975). So, there might be a noun-phrase frame with slots for a singular indefinite determiner, an adjective, and a singular count noun. And this frame may occupy a larger slot in a clausal frame, and so on. Because of the separation between words and their slots, the system has the means to encode new phrases (e.g., "a poetic tiger") by putting known words into known frames in new combinations. Evidence for such a system comes from dissociations in aphasia between individuals with lexical retrieval deficits and those with deficits in syntactic-sequential processes (e.g., see Gordon & Dell, 2003, for review), from functional imaging data that identify different brain areas for word-retrieval and word-combination mechanisms (e.g., Hagoort, 2013), and from structural priming studies, which are reviewed later.

18.1.3 Phonological Encoding

The retrieval of the phonological form of a word results in a sequence of phonological segments: k æ t. The segments are then put together with the segments of surrounding words, and the resulting sequence must be characterized in terms of its syllables and how those syllables are stressed (Levelt et al., 1999). These processes must respect the phonological properties of the language being spoken, including how segments combine to make syllables (*phonotactic* knowledge), how syllables are organized into higher-level prosodic structures, and how timing, pitch, and intensity vary as a function of those structures. Ultimately, this organized phonological structure guides the articulatory process. The phonological encoding process has been studied by assessing the response time to produce words and syllables (e.g., Cholin, Dell, & Levelt, 2011; Meyer, 1991), by examining phonological speech errors (Warker & Dell, 2006), by measuring the articulatory and acoustic details of utterances (e.g., Goldrick & Blumstein, 2006; Goldstein, Pouplier, Chen, Saltzman, & Byrd, 2007; Lam & Watson, 2010), and more recently by event-related brain potentials and other imaging techniques (e.g., Qu, Damian, & Kazanina, 2012).

18.2 LONG-TERM SPEAKER TUNING: IMPLICIT LEARNING

The production system does its job because it has learned to do so, and the basis for that learning is experience in comprehending and speaking (Chang et al., 2006). Learning, however, is not just something that children do. The typical adult speaker says approximately 16,000 words per day (Mehl, Vazire, Ramirez-Esparza, Slatcher, & Pennebaker, 2007) and hears and reads many more. This experience adapts the production system so that it is able to make effective choices in the particular circumstances that it finds itself. We refer to this continual process of adaptation as *implicit learning*. We claim that this adaptation is a kind of learning because the changes induced are not short-lived, and that the learning is implicit because it is an automatic consequence of linguistic experience that occurs without any intention to learn or awareness of what has been learned. In the remainder of this section, we review implicit-learning research in each of the three production subprocesses mentioned previously. For lexical access, we consider mechanisms of lexical repetition priming and frequency effects, and the possibility of phrasal frequency effects. For grammatical encoding, we discuss the hypothesis that structural priming in production is a form of implicit learning. And, for phonological encoding, we review studies that find implicit learning of novel phonotactic patterns.

18.2.1 Implicit Learning of Words and Phrases

The production system adapts to make the words that it is most likely to use easier to retrieve and articulate. In particular, words that we have said recently are easier to say than words that we have not said recently (*repetition priming*; e.g., Mitchell & Brown, 1988). We are also, in general, faster and more accurate at producing words that we have more experience saying, that is, frequent words (Caramazza, Costa, Miozzo, & Bi, 2001; Jescheniak & Levelt, 1994). Both repetition priming and frequency effects are thought to arise from and can be explained by implicit learning, which optimizes the production system for situations that are more likely to happen.

One manifestation of implicit learning in word production is *cumulative semantic interference* (e.g., Howard, Nickels, Coltheart, & Cole-Virtue, 2006). When we have to name a picture of the same thing twice (e.g., *crow*), we benefit from repetition priming. But, if instead of repeating the picture's name we next have to name something that is similar in meaning, but not the same word (e.g., *finch*), then we produce this word more slowly and have a greater chance of error (e.g., Schnur, Schwartz, Brecher, & Hodgson, 2006). This negative effect is semantic interference. Oppenheim, Dell, and Schwartz (2010) investigated the "dark" (semantic interference) and "light" (repetition priming) sides of word production using a computational model, aptly called the "dark-side" model. In the model, each experience with a word tunes the production system by prioritizing words that are recently used and, importantly, deprioritizing their competitors, that is, semantically similar words. This tuning consists of the strengthening of connections to words when they are used, but weakening of connections to these words' competitors. As a result, when a word is repeated it becomes relatively more active in the lexical network, effectively by leeching activation from similar words. In this way, repeating the word *crow* becomes easier, whereas naming different, but semantically similar, words in a sequence (e.g., *crow*, *finch*, *gull*) becomes increasingly difficult. This effect shows that the production system is adaptive, because words that are used and will likely be used again become easier to say, whereas words that could potentially interfere with those words are rendered less accessible and, hence, less disruptive.

We described how lexical access in production involves two steps, retrieval of the abstract lexical item and then retrieval of the item's phonological form. We also noted that semantic errors such as "dog" for "cat" can occur at the first step, but phonological errors such as "cap" or "dat" for "cat" occur during the second step. We also said that recently spoken or high-frequency words (e.g., "cat" as opposed to "feline") are less vulnerable to error because an implicit learning process enhances their retrieval. But does the greater ease associated with common or repeated words apply to both steps or just one of them? Jescheniak and Levelt (1994) proposed that frequency effects in word retrieval are felt largely during the second step. Others (e.g., Knobel, Finkbeiner, & Caramazza, 2008) claim that both steps benefit when the target word is frequent, because implicit learning should have an effect throughout the retrieval process. Kittredge, Dell, Verkuilen, and Schwartz (2008) addressed this question by looking at how target-word frequency affects semantic and phonological errors during picture naming. They presented aphasic participants with pictures to name that varied, among other factors, in their word frequency. They found, as expected, that the odds of saying the right word increased with the frequency of the target, demonstrating that common words are "protected" by their frequency. This protective power was found to prevent both semantic and phonological errors, suggesting that both steps of lexical retrieval benefit from frequency and, more generally that the production system keeps track of likely events at all levels.

The production system also seems to keep track of and adapts to the degree to which words combine. Janssen and Barber (2012) explored this by looking at whether the frequency of the combination of two words (e.g., *red car* or *red hammer*) predicted how easily that phrase was generated. In particular, they presented participants with pictures of colored objects and had them name the object and its color with an appropriate phrase. They found that frequent phrases had faster naming latencies than would be predicted just by the frequency of the first or second word. This suggests that the production system tunes itself to probable events beyond the word level by keeping track of word combinations as well.

18.2.2 Structural Priming

One of the classic findings in psycholinguistics is structural priming, also known as syntactic priming, or structural repetition. Structural priming is the tendency for speakers to reuse recently experienced structures. Bock (1986a) gave experimental participants pictures that can be described with either of the two kinds of dative structures (double objects, "The woman handed the boy the paint brush," versus prepositional datives, "The woman handed the paint brush to the boy"). Participants described these pictures after saying an unrelated prime sentence that used either a double-object or prepositional dative structure. Priming was seen in the tendency for speakers to use the structure of the prime when describing the picture. Similar effects were seen for other structural alternations such as active transitive sentences ("Lightning is striking the church") versus passives ("The church is struck by lightning"). The important aspect of this priming is that it appears to be the persistence of an abstract syntactically characterized structure (e.g., the frame: *Noun_phrase Auxillary_verb Main_verb Prepositional_phrase* for a full passive), and not the lexical content of the utterance, its meaning, or its intonational properties (Bock & Loebell, 1990). As such, structural priming provides evidence for a production process that uses structural abstractions during grammatical encoding.

Bock and Griffin (2000) claimed that structural priming is not just a temporary change to the system, but instead it is a form of implicit learning, akin to the connection weight changes that characterize learning in connectionist models. They provided evidence for this claim by showing that the effect of a prime persists undiminished over at least 10 unrelated sentences (several minutes). If priming were due to temporary activation of a structure, then the prime's influence would rapidly decay. The evidence that the learning is implicit is that it occurs in brain-damaged speakers who have no explicit memory of the prime sentence (Ferreira, Bock, Wilson, & Cohen, 2008).

Chang et al. (2006) created a computational model that reflected the idea that structural priming is implicit learning. They trained a connectionist model to simulate a child experiencing sentences one word at a time. The model was also given a representation of the intended meaning of some of the sentences that it experienced, with this meaning presumably having been inferred by the child from context. The model learned by "listening" to each sentence and trying to predict each word. When the actual next word was heard, the model then compared its prediction to that word, thus generating a prediction error signal. This error signal was the impetus for the model to change its connection weights so that its future predictions were more accurate. By using prediction error, the model learned the linguistic patterns in the language (e.g., syntactic structures) and how those patterns mapped onto meaning (e.g., Elman, 1993). After this learning, the model was able to produce because "prediction is production" (Dell & Chang, 2014); generating the next word from a representation of previous words and intended meaning is, computationally, a production process. When given a representation of intended meaning, the model's sequence of word predictions constituted the production of a sentence. The key aspect of this model, for our purposes, is that it accounted for structural priming through learning. Even after the model attained "adult" status, it continued to learn. When a prime sentence was experienced, the model's connection weights were changed ever so slightly to favor the subsequent production of sentences with the same structure. Experiencing, for example, a double-object dative inclined the model to produce that structure later. Because the priming was based on weight change, it is a form of learning, thus accounting for Bock and Griffin's finding that structural priming is undiminished over time. Also, the evidence that the implicit learning that characterizes structural priming is based on prediction error comes from demonstrations that less common, and hence more surprising, prime structures lead to more priming than common ones (e.g., Jaeger & Snider, 2013).

18.2.3 Phonotactic Learning

Young children implicitly learn the phonotactic patterns of their language through experience. Such patterns include knowledge about where certain consonants can occur in the syllables in their language; for example, in English, /h/ only occurs at the beginning of a syllable (the *onset*) and /ng/ occurs only at the end (the *coda*). Because of their phonotactic knowledge, English speakers can readily produce the phonotactically legal nonword "heng," but not the illegal "ngeh." Evidence that the production system actively uses this knowledge comes from the *phonotactic regularity effect* on speech errors: slips tend to be phonotactically legal. One might mistakenly produce "nun" as "nung," a phonotactically legal nonword, but not as "ngun" (Wells, 1951).

Warker and Dell (2006) and Dell, Reed, Adams, and Meyer (2000) created an experimental analogue to the phonotactic regularity effect in which participants recited four-syllable tongue twisters such as "hes feng kem neg" at a fast pace. Unbeknownst to the participants, the syllables followed artificial phonotactic patterns that were present only in the experimental materials. For example, a participant's syllables might follow the pattern: *During the experiment, /f/ can only be a syllable onset and /s/ can only be a syllable coda* (as in the example four-syllable sequence above). Participants would recite several hundred of these sequences in each of four experimental sessions on consecutive days. Because of the fast speech rate, slips were reasonably common. Most often, these involved movements of consonants from one place to another, such as "hes feng kem neg" being spoken as "**f**es feng kem neg," in which /f/ moved to the first syllable. The crucial feature of the study was whether these slips respected the phonotactics of the experienced syllables. As expected, slips of /h/ and /ng/ respected English phonotactics; /h/ is always moving to an onset position and /ng/ is always moving to a coda position. The crucial finding, though, was that slips of the artificially restricted consonants (/f/ and /s/ in our example) also respected the local phonotactics of the experiment. Notice in the example that /f/ slips to an onset position, that is, the slip is "legal" with regard to the experimental phonotactic patterns. And this was not just a small statistical tendency; 98% of the slips of experimentally restricted consonants were "legal" in this respect, whereas consonants that were not experimentally restricted the way that /f/ and /s/ were often slipped from onset to coda or vice versa (Dell et al., 2000).

Finding that slips respected the experimental distributions of consonants suggests that participants implicitly learned these distributions, and this learning affected their slips. But is this effect truly one of learning, as opposed to some very temporary priming of preexisting knowledge (e.g., priming of a rule that /f/ can be an onset in English)? Evidence that true learning is occurring comes from exposing participants to more complex "second-order" constraints such as: *if the vowel is /ae/, then /f/ must be an onset and /s/ must be a coda, but if the vowel is /I/, then /s/ must be an onset and*

/f/ must be a coda. Warker and Dell (2006) found that participants' slips did *not* follow this vowel-dependent second-order constraint on the first day of a 4-day experiment. On the second and subsequent days, though, the slips did obey the constraint (e.g., more than 90% of slips were legal). This suggests that the effect requires *consolidation*, a period of time (possibly involving sleep; Warker, 2013) in which the results of the experience are registered in a relatively permanent way in the brain. After consolidation, the effects appear to remain at least for 1 week (Warker, 2013). Because the effect requires consolidation and is persistent in time, it appears to be a form of learning. Thus, phonotactic-like knowledge and its expression in speech production errors can be tuned by an implicit learning process.

18.3 SHORT-TERM SPEAKER TUNING

Implicit learning is not the only mechanism that allows the production system to fluently generate appropriate, grammatically correct utterances that listeners can easily interpret. There are several adaptive phenomena in production that involve immediate processing, rather than long-term learning. These short-term tuning mechanisms include error monitoring, availability-based production choices, sensitivity to information density, and audience design.

18.3.1 Error Monitoring

Speakers help their listeners by avoiding making speech errors or, when an error occurs, by attempting to correct it. Catching slips before they happen or fixing them after they do requires that speakers do error monitoring. Studies of monitoring suggest that we detect at least half of our overt slips after they happen, and that we detect and block potential errors before they can occur (Baars, Motley, & MacKay, 1975; Levelt, 1983). Evidence that errors can be detected before they are spoken comes from the existence of very rapid detections of overt errors. Levelt (1983) gave the example of "v—horizontal." The speaker started to say "vertical," but quickly stopped and replaced it with the correct "horizontal." The fact that speech was stopped right away (within 100 msec of the onset of the erroneous /v/) demonstrates that the error was almost certainly detected before articulation began. How is this possible? There are two theories of error detection. One is that speakers detect errors by comprehending their own speech and noting if there is a mismatch with what they intended (Hartsuiker & Kolk, 2001; Levelt, 1983). This view—the *perceptual loop theory*—allows for the comprehension of internal speech before it is produced to explain the fact that errors can be detected before articulation. The alternative is that error detection occurs within the production system itself. An example of this is the *conflict detection theory* of Nozari, Dell, and Schwartz (2011), which proposes that the production system can assess the extent to which its decisions are conflicted and assumes, when conflict is high, that an error is likely. For example, suppose that during word access, the word CAT was selected during the first lexical-access step, but DOG was also nearly as activated as CAT. That can be taken as a sign that there was a possible error during that step. Similarly, if a particular speech sound, for example, /d/, is selected while another, /k/, is almost as active, again that can be a signal that there may have been a mis-selection, this time during the second access step. Nozari et al. used a computational model to demonstrate that the association between high conflict and error likelihood is a strong one, but also that the association no longer holds when the production system is functioning very poorly. Thus, for some aphasic individuals, conflict would not be an effective predictor of error and such individuals would be expected to have trouble detecting their own errors.

To test the conflict detection theory of monitoring and the competing perceptual loop theory, Nozari et al. (2011) examined how successful aphasic individuals were at detecting their own errors in a picture-naming task. The perceptual loop account predicts that good error detection should be associated with good comprehension because detection is performed by the comprehension system in that theory. In contrast, the conflict detection theory expects good detection to be associated with production rather than comprehension skill. The results supported the conflict detection account. The aphasic patients with higher rates of error detection had relatively good production skills, and comprehension ability was unrelated to error detection rate. Furthermore, Nozari and colleagues showed that patients who were relatively better at the first step of lexical access, but poor at the second step, could detect their first-step errors (e.g., semantic errors) but not their second-step errors (phonological errors). The complement was true as well—doing better on the second step in production implied better detection of second-step errors in particular. These results show that dissociations in production abilities for the lexical-access steps are mirrored in differential abilities to detect errors at the two steps, exactly as expected by the conflict detection theory.

Do the results of Nozari et al. (2011) mean that we do not detect errors by comprehending our own speech? No. These results only point to another possible mechanism for error detection, particularly a mechanism that can detect errors before they happen. It seems likely that many overtly spoken slips are detected simply by hearing them, as proposed in the perceptual loop theory. In support of this claim, Lackner and Tuller (1979) found that using noise to mask a speaker's speech diminished the speaker's ability to detect their overt phonological errors, demonstrating that perception of the auditory signal plays a role in detection. It is therefore likely that multiple mechanisms contribute to the monitoring process. For example, speakers appear to guard their speech against slips that create taboo words (Motley, Camden, & Baars, 1982; see also Severens, Kühn, Hartsuiker, & Brass, 2012 for an fMRI study of frontal brain regions involved in taboo-word monitoring).

18.3.2 Availability-Based Production

The adage, "think before you speak" advises speakers to fully plan their utterances before saying them. The fact that the adage exists suggests that speakers do not routinely do this. Instead, language production involves some degree of *incrementality* (Kempen & Hoenkamp, 1987): utterances are often constructed and spoken in a piecemeal fashion, with the result that one might start talking before having planned the entire sentence. Because production can be incremental, the retrievability of the various parts of the utterance can influence its structure. For example, when attempting to produce the message illustrated in Figure 18.1, suppose that we are able to retrieve "rat," but have not yet retrieved "cat." We can start the utterance as "The rat..." and then, because English allows for a passive structure, can continue with "was chased by the cat" as we eventually retrieve the other words. Thus, the production system may opportunistically take advantage of the words that are retrieved first and may start with those words. This is the essence of availability-based production. What is retrieved first tends to be said first. More generally, what is available tends to be spoken as soon as it can. Although this strategy occasionally results in false starts, it makes for an efficient production system (Bock, 1982).

Bock (1986b) provided support for availability-based production by asking speakers to describe pictures such as one in which lightning is striking a church. This can be described with either an active ("Lightning is striking a church") or a passive ("The church is struck by lightning") structure. Earlier in this chapter, we showed how this structural choice can be influenced by structural priming. It turns out that this choice is also sensitive to the relative availability of the words "lightning" and "church." Bock found that participants who had recently experienced the word "thunder," which presumably makes "lightning" more available, were more likely to describe the picture with the active form, making the primed word come out earlier. Similarly, priming "church" made the passive more likely.

One can also see availability-based production at work in choices about optional words. A sentence such as "The coach knew **that** you missed practice" can be produced with no "that," without changing the meaning of the sentence. So, what determines whether you include the "that"? One possibility is that speakers engage in *audience design*: when faced with a production choice, they choose what will make the sentence easier for their listener to understand. Notice that if the "that" is missing, then the sentence has a temporary ambiguity when "you" is heard. The "you" can be either the direct object of "knew" or the subject of a new embedded clause. Including the "that" removes the ambiguity. Ferreira (2008) and Ferreira and Dell (2000) suggested an alternative explanation for when "that" is present in these sentences. It has to do with the availability of the material after "that." If "you" has already been retrieved and is ready to go at the point in the sentence after "The coach knew...," then the speaker is more likely to omit "that." But if the speaker is not quite ready with "you," then including "that" is a convenient way to pause and buy time. As described in the subsequent section on information-density sensitivity, there is evidence that speakers do attempt to stretch time out at certain points in a sentence, and this can be thought of as an example of this. Here the issue is whether speakers produce "that" to disambiguate the utterance for their listeners, or because their production systems naturally produce whatever is available.[1] If the "you" is immediately available after "The coach knew," then the sentence can grammatically continue without the "that." Ferreira and Dell tested these ideas by comparing the production of four kinds of sentences:

I knew (that) ***I*** *missed practice.* (embedded pronoun is repeated and unambiguously nominative)
You knew (that) ***you*** *missed practice.* (embedded pronoun is repeated and ambiguous)
I knew (that) ***you*** *missed practice.* (embedded pronoun is not repeated and ambiguous)
You knew (that) ***I*** *missed practice.* (embedded pronoun is not repeated and unambiguously nominative)

The sentences were presented and then recalled in situations in which the participants could not remember whether there had been a "that" in the sentence

[1]Of course, the grammar does not always allow you to eliminate "that" as a complementizer; "the girl that saw the boy is here" must have it in Standard American English.

and, hence, they tended to use their natural inclinations about whether to include "that." The key variable was the percentage of recalled sentences with "that." The hypothesis that speakers include "that" to help their listeners predicts that the two sentences with the ambiguous embedded "you" will include more instances of "that" than the unambiguous conditions that have the clearly nominative pronoun, "I," as the embedded subject. The availability hypothesis predicts that because of repetition priming, the embedded pronoun (**I** or **you**) will be more available if it had just been said as the subject of the main clause. Because their embedded pronouns should be quite available, the two conditions with repeated pronouns are expected to have fewer instances of "that." Across several experiments, there was no tendency for more "that"s in the ambiguous sentences, but repeating the pronoun caused the percentage of "that"s to decrease by approximately 9%. The results clearly supported the availability hypothesis, providing another demonstration that the production system's decisions are opportunistically guided by what is easily retrieved. In the next section, we approach the question of the production of optional words like "that" from another angle.

18.3.3 Information-Density Sensitivity

One way that people may alter production in the short-term is by monitoring for and adjusting the probabilistic characteristics of what they are about to say. Taking a cue from information theory (Shannon, 1948), it has been proposed that speakers control the *rate of information* conveyed in their utterances so that there are as few as possible points in which the rate is extremely high or extremely low. Recall that, on a formal level, words or structures that are *less* likely contain *more* information and that, in the reverse case, *redundant* or predictable items are associated with less information. The idea is that keeping the information rate constant at a level that listeners can handle maximizes the effective transmission to the listener. Too fast a rate leads to loss of transmission, and too slow a rate wastes time. The hypothesized information constancy in production is termed the *smooth signal redundancy hypothesis* or, alternatively, *uniform information density* (UID). This tendency can be assumed to apply at all levels of language production, including lexical choice (Mahowald, Fedorenko, Piantadosi, & Gibson, 2013), syntactic structure (Jaeger, 2006, 2010), and phonetic and phonological output (Aylett & Turk, 2004).

Lexical, syntactic, phonological, and pragmatic predictability and given-ness, as constrained by the discourse or experiment, strongly influence the durations of individual words as would be expected from the UID. For example, the word *nine* in "A stitch in time saves *nine*" is shorter than in the phrase "I'd like *nine*" (Lieberman, 1963) because the *nine* in the first example is highly predicted by the previous words. Speakers moderate duration and other prosodic cues in response to these linguistic factors, as has been demonstrated experimentally and in the wild, and this effect is robust even when a large number of other factors are taken into account (Jurafsky, Bell, Gregory, & Raymond, 2001).

Aylett and Turk (2004) examined the relationship between reduction and redundancy, or the contribution of statistical predictability to the short-term manifestation of phonetic output. They modeled the durations of syllables as a function of the degree to which they were predicted by the preceding information and the predictability of the word itself in discourse as well as its frequency of occurrence. They found evidence that individuals regulate the distribution of information in the signal (modulating various prosodic cues like duration, volume, and pitch) in that these cues represent a tradeoff between predictability and acoustic prominence. So, when a word is less predictable, it will carry more information in the sense that it is unexpected, but speakers take this into account by providing additional cues as to the identity of an upcoming word or syllable, such as articulating the word more loudly.

Jaeger (2006, 2010) identified analogous behavior in syntactic flexibility as a function of information density. Using evidence from the optional *that* structure that we introduced in the previous section, he demonstrated that the choice of whether to include a "that" provides the language production system with a means of redistributing information so that information density is more uniform across the utterance. For example, in sentences such as "My boss confirmed/thinks (that) we were absolutely crazy," speakers were more likely to include "that" when the presence of a complement clause (e.g., "...we were absolutely crazy") is unexpected given the main verb (e.g., "confirmed"). This is because *confirm* is a verb that most often takes a noun as its argument (e.g., "...confirmed the result"), and so the presence of a complement clause is less probable and therefore more surprising. In contrast, the verb *think* often takes a complement clause and, because that is more expected, "that" was less likely to be included in the utterance. In general, including a "that" when the complement clause is unexpected makes upcoming linguistic material less surprising for the listener, because "that" very commonly signals for a complement clause. The resulting overall structure is much more even in its syntactic surprisal. In this way, the speaker's choices about optional *that* are sensitive to the goal of minimizing peaks and valleys in the information conveyed during the incremental production of the sentence.

These choices presumably translate into less effort for the listener. It is also possible that producing sentences with more UID directly aids the fluency of the production process, because what may be highly surprising to a listener may be relatively more difficult for speakers to create.

When we say that speakers "monitor" and "adjust" information density, this implies active online control. But control of information rate is not necessarily the result of an active short-term adaptation. Instead, the mechanisms that achieve good information rates may be learned as speakers gain experience about what production choices lead to effective comprehension (Jaeger & Ferreira, 2013). For example, speakers may consistently include "that" in their complement clauses introduced by main verbs such as "confirm" because they have learned that failure to do so leads to misunderstanding. With this view, there is no active control of information density; only the retention of successful speaking habits.

18.3.4 Audience Design

The language production system adapts to one's partner, not only by avoiding high information rates but also by considering the partner's specific needs and abilities; a speaker uses syntax, words, and phonology that the partner will likely understand. As we have outlined throughout this chapter, the production system can adapt to internal moment-by-moment demands, and it can change itself in the long-term as a function of experience. In this final section, we consider how the production system goes beyond what is easy for *it* to do, and instead considers what might best help the other person understand. This consideration is known as *audience design*. We discuss two examples of such design. First, we consider the way individuals use words that result in more effective communication on a cooperative task via a process called *entrainment* (Brennan & Hanna, 2009; Clark & Wilkes-Gibbs, 1986), and then how talkers can change their own pronunciation of words to facilitate understanding in *phonetic convergence* (Pardo, 2006).

Entrainment, or the convergence on a single term between two talkers in a conversation, is a necessary part of communication. It is estimated that approximately 50% of discourse entities are mentioned multiple times in a conversation or text (Recasens, de Marneffe, & Potts, 2013). Given this degree of repetition, it would be useful if speakers could agree on labels for those entities. If one party to a conversation referred to a particular plant as a "bush" and the other called it a "tree," then confusion is likely. Agreement on terms through entrainment removes the confusion. In experimental settings, entrainment has been examined by looking at how participants in a cooperative task describe an object and how that description changes as the participants continue to interact. In Clark and Wilkes-Gibbs (1986), participants had to cooperate to sort a set of abstract visual shapes (made up of "tangrams") often resembling people or animals. Over the course of several turns, both partners came to use similar, eventually convergent terms or short phrases to describe the items. Early on, a speaker might recognize that the other person does not understand their initial description (e.g., "The next one is the rabbit." "huh?"), requiring that the speaker elaborate (e.g., "That's asleep, you know, it looks like it's got ears and a head pointing down."). As the experiment continues, the objects' labels become increasingly shorter and the listener's errors in interpreting what is said become rare, suggesting that talkers have optimized label length and form for communicative efficiency. The description of such a figure can go from the very complex on the first exchange ("looks like a person who's ice skating, except they're sticking two arms out in front") to shorter, multiphrasal ("the person ice skating, with two arms") to finally a single noun phrase ("the ice skater"). Thus, the entrainment process adapts the production systems of both participants in such a manner that communication success, rather than production ease, is the goal.

Phonetic convergence is a phenomenon where individuals adopt the phonetic and phonological representations of the other talker during a conversation. A person may adopt features of another's accent, such as the famous US southern "pin-pen" merger or a northern cities vowel shift (e.g., "dawg" becomes "dahg"), or even more subtle features such as differences in voice-onset time. Pardo (2006) demonstrated such convergence experimentally. Participants completed a map task where one partner's map (the receiver's) needs to be drawn to look like the other's (the giver's). Like the tangram task used by Clark and Wilkes-Gibbs, the map task requires cooperative communication. There were many places on the map with standard names provided to the talkers (e.g., *abandoned monastery*, *wheat field*, etc.). Phonetic convergence of the speech of the talker pairs was assessed by naïve participants who were asked to judge the degree of similarity of the pairs' pronunciations for these place names as they did the task. Not only did all conversation partners show some degree of convergence, but also these effects arose after very little interaction time—many partners showed convergent phonetics as early as before the halfway point in their dialogue, with convergence persisting into the second half as well. This convergence demonstrates that individuals engage in audience design by adopting the phonetic features of their conversation partner during a cooperative task. Because it is presumably easier for each speaker to use his or her own accent, phonetic convergence counts as another example in which the adaptation suits the goal of communication, rather than the immediate ease of the production systems of the individual speakers.

18.4 CONCLUSION

Language production is, in one sense, difficult. The speaker has to decide on something worth saying, choose words (out of a vocabulary of 40,000), appropriate syntax, morphology, and prosody, and ultimately has to articulate at the rate of two to three words per second. In another sense, production is easy. We think it takes little effort. Particularly when we are talking about familiar topics, we can at the same time walk, drive, or even play the piano (Becic et al., 2010). The seeming paradox that something so difficult is yet so easy is resolved when we consider the mechanisms presented in this chapter. The production system is continually being tuned by the extraordinary amount of experience we have. We say 16,000 words per day and hear and read a lot more. The implicit learning that results from this input effectively trains the system and tunes it well to its current circumstances. But implicit learning is not the whole story. The production system also makes use of a variety of moment-by-moment mechanisms to compensate for and prevent errors, to promote fluency, and to make the job of the listener easier.

Acknowledgments

Preparation of this chapter was supported by NIH DC000191 and by an NSF fellowship to Cassandra Jacobs.

References

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*, 31–56.

Baars, B. J., Motley, M. T., & MacKay, D. G. (1975). Output editing for lexical status from artificially elicited slips of the tongue. *Journal of Verbal Learning and Verbal Behavior*, *14*, 382–391.

Becic, E., Dell, G. S., Bock, K., Garnsey, S. M., Kubose, T., & Kramer, A. F. (2010). Driving impairs talking. *Psychonomic Bulletin & Review*, *17*, 15–21.

Bock, J. K. (1982). Towards a cognitive psychology of syntax: Information processing contributions to sentence formulation. *Psychological Review*, *89*, 1–47.

Bock, K. (1986a). Syntactic persistence in language production. *Cognitive Psychology*, *18*, 355–387.

Bock, K. (1986b). Meaning, sound, and syntax: Lexical priming in sentence production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *12*, 575–586.

Bock, K., & Griffin, Z. M. (2000). The persistence of structural priming: Transient activation or implicit learning? *Journal of Experimental Psychology: General*, *129*, 177–192.

Bock, K., & Loebell, H. (1990). Framing sentences. *Cognition*, *35*, 1–39.

Brennan, S. E., & Hanna, J. E. (2009). Partner-specific adaptation in dialog. *Topics in Cognitive Science*, *1*, 274–291.

Caramazza, A., Costa, A., Miozzo, M., & Bi, Y. (2001). The specific-word frequency effect: Implications for the representation of homophones in speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*, 1430–1450.

Chang, F., Dell, G. S., & Bock, K. (2006). Becoming syntactic. *Psychological Review*, *113*, 234–272.

Cholin, J., Dell, G. S., & Levelt, W. J. M. (2011). Planning and articulation in incremental word production: Syllable-frequency effects in English. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*, 109–122.

Chomsky, N. (1959). A review of Skinner's Verbal Behavior. *Language*, *35*, 26–58.

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

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

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*, *369*, 20120394.

Dell, G. S., Nozari, N., & Oppenheim, G. M. (2014). Word production: Behavioral and computational considerations. In M. Goldrick, V. S. Ferreira, & M. Miozzo (Eds.), *The Oxford handbook of language production*. Oxford, UK: Oxford University Press.

Dell, G. S., Reed, K. D., Adams, D. R., & Meyer, A. S. (2000). Speech errors, phonotactic constraints, and implicit learning. A study of the role of experience in language production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*, 1355–1367.

Dell, G. S., Schwartz, M. F., Nozari, N., Faseyitan, O., & Coslett, H. B. (2013). Voxel-based lesion-parameter mapping: Identifying the neural correlates of a computational model of word production. *Cognition*, *128*, 380–396.

Elman, J. L. (1993). Learning and development in neural networks: The importance of starting small. *Cognition*, *48*, 71–99.

Ferreira, V. S. (2008). Ambiguity, availability, and a division of labor for communicative success. *Psychology of Learning and Motivation*, *49*, 209–246.

Ferreira, V. S., Bock, K., Wilson, M. P., & Cohen, N. J. (2008). Memory for syntax despite amnesia. *Psychological Science*, *19*, 940–946.

Ferreira, V. S., & Dell, G. S. (2000). The effect of ambiguity and lexical availability on syntactic and lexical production. *Cognitive Psychology*, *40*, 296–340.

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

Garrett, M. F. (1975). The analysis of sentence production. *Psychology of Learning and Motivation*, *9*, 133–177.

Goldrick, M., & Blumstein, S. (2006). Cascading activation from phonological planning to articulatory processes: Evidence from tongue twisters. *Language and Cognitive Processes*, *21*, 649–683.

Goldstein, L., Pouplier, M., Chen, L., Saltzman, E., & Byrd, D. (2007). Dynamic action units slip in speech production errors. *Cognition*, *103*, 386–412.

Gordon, J. K., & Dell, G. S. (2003). Learning to divide the labor: An account of deficits in light and heavy verb production. *Cognitive Science*, *27*, 1–40.

Hagoort, P. (2013). MUC (Memory, Unification, Control) and beyond. *Frontiers in Psychology*, *4*, 416.

Hartsuiker, R. J., & Kolk, H. H. J. (2001). Error monitoring in speech production: A computational test of the perceptual loop theory. *Cognitive Psychology*, *42*, 113−157.

Howard, D., Nickels, L., Coltheart, M., & Cole-Virtue, J. (2006). Cumulative semantic inhibition in picture naming: Experimental and computational studies. *Cognition*, *100*, 464−482.

Jaeger, T. F. (2006). *Redundancy and syntactic reduction in spontaneous speech*. Ph.D. dissertation. Stanford University.

Jaeger, T. F. (2010). Redundancy and reduction: Speakers manage syntactic information density. *Cognitive Psychology*, *61*, 23−62.

Jaeger, T. F., & Ferreira, V. S. (2013). Seeking predictions from a predictive framework. *Behavioral and Brain Sciences*, *36*, 359−360.

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*, 57−83.

Janssen, N., & Barber, H. A. (2012). Phrase frequency effects in language production. *PLoS ONE*, *7*, e33202.

Jescheniak, J. D., & Levelt, W. J. M. (1994). Word frequency effects in speech production: Retrieval of syntactic information and of phonological form. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*, 824−843.

Jurafsky, D., Bell, A., Gregory, M., & Raymond, W. D. (2001). Probabilistic relations between words: Evidence from reduction in lexical production. *Typological Studies in Language*, *45*, 229−254.

Kempen, G., & Hoenkamp, E. (1987). An incremental procedural grammar for sentence formulation. *Cognitive Science*, *11*, 201−258.

Kittredge, A. K., Dell, G. S., Verkuilen, J., & Schwartz, M. F. (2008). Where is the effect of frequency in word production? Insights from aphasic picture-naming errors. *Cognitive Neuropsychology*, *25*, 463−492.

Knobel, M., Finkbeiner, M., & Caramazza, A. (2008). The many places of frequency: Evidence for a novel locus of the frequency effect in word production. *Cognitive Neuropsychology*, *25*, 256−286.

Lackner, J. R., & Tuller, B. H. (1979). Role of efference monitoring in the detection of self-produced speech errors. In W. E. Cooper, & E. C. T. Walker (Eds.), *Sentence processing: Psycholinguistic studies presented to Merrill Garrett*. Hillsdale, NJ: Erlbaum.

Lam, T. Q., & Watson, D. G. (2010). Repetition is easy: Why repeated referents have reduced prominence. *Memory & Cognition*, *38*, 1137−1146.

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

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

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

Lieberman, P. (1963). Some effects of semantic and grammatical context on the production and perception of speech. *Language and Speech*, *6*, 172−187.

Mahowald, K., Fedorenko, E., Piantadosi, S. T., & Gibson, E. (2013). Info/information theory: Speakers choose shorter words in predictive contexts. *Cognition*, *126*, 313−318.

Mehl, M. R., Vazire, S., Ramirez-Esparza, N., Slatcher, R. B., & Pennebaker, J. W. (2007). Are women really more talkative than men? *Science*, *317*, 82.

Meyer, A. S. (1991). The time course of phonological encoding in language production: Phonological encoding inside a syllable. *Journal of Memory and Language*, *30*, 69−89.

Mitchell, D. B., & Brown, A. S. (1988). Persistent repetition priming in picture naming and its dissociation from recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*, 213−222.

Motley, M. T., Camden, C. T., & Baars, B. J. (1982). Covert formulation and editing of anomalies in speech production: Evidence from experimentally elicited slips of the tongue. *Journal of Verbal Learning and Verbal Behavior*, *21*, 578−594.

Nozari, N., Dell, G. S., & Schwartz, M. F. (2011). Is comprehension necessary for error detection? A conflict-based account of monitoring in speech production. *Cognitive Psychology*, *63*, 1−33.

Oppenheim, G. M., Dell, G. S., & Schwartz, M. F. (2010). The dark side of incremental learning: A model of cumulative semantic interference during lexical access in speech production. *Cognition*, *114*, 227−252.

Pardo, J. S. (2006). On phonetic convergence during conversational interaction. *The Journal of the Acoustical Society of America*, *119*, 2382−2393.

Qu, Q., Damian, M. F., & Kazanina, N. (2012). Sound-sized segments are significant for Mandarin speakers. *Proceedings of the National Academy of Sciences*, *109*, 14265−14270.

Recasens, M., de Marneffe, M. C., & Potts, C. (2013). The life and death of discourse entities: Identifying singleton mentions. In *Proceedings of NAACL-HLT*, (pp. 627−633). Atlanta, GA.

Schnur, T. T., Schwartz, M. F., Brecher, A., & Hodgson, C. (2006). Semantic inference during blocked-cyclic naming: Evidence from aphasia. *Journal of Memory and Language*, *54*, 199−227.

Severens, E., Kühn, S., Hartsuiker, R. J., & Brass, M. (2012). Functional mechanisms involved in the internal inhibition of taboo words. *Social, Cognitive, and Affective Neuroscience*, *7*, 431−435.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, *27*, 623−656.

Ueno, T., Saito, S., Rogers, T. T., & Lambon Ralph, M. A. (2011). Lichtheim 2: synthesizing aphasia and the neural basis of language in a neurocomputational model of the dual dorsal-ventral language pathways. *Neuron*, *72*, 385−396.

Warker, J. A. (2013). Investigating the retention and time course for phonotactic constraint learning from production experience. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*, 96−109.

Warker, J. A., & Dell, G. S. (2006). Speech errors reflect newly learned phonotactic constraints. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*, 387−398.

Wells, R. (1951). Predicting slips of the tongue. *Yale Scientific Magazine*, *3*, 9−30.