Language Production means talking, but not merely that. When people talk, it is usually because they have something to say. Psycholinguistic research on language production concerns itself with the cognitive processes that convert nonverbal communicative intentions into verbal actions. These processes must translate perceptions or thoughts into sounds, using the patterns and elements of a code that constitutes the grammar of a language. For theories of language production, the goal is to explain how the mind uses this code when converting messages into spontaneous speech in ongoing time. This requires an explanation of the action system that puts language knowledge to use.

The action system for language production has a COGNITIVE ARCHITECTURE along the lines shown in figure 1. Imagine a speaker who wishes to draw a listener's attention to a rabbit browsing in the garden. The process begins with a communicative intention, a message, that stands at the interface between thought and language. Little is known about the content or structure of messages, but they are assumed to include at least conceptual categorizations (in figure 1, tacit rabbit-knowledge) and the information needed for making distinctions such as tense, number, aspect, and speaker's perspective. Less certain is whether messages habitually include different kinds of information as a function of the language being spoken, along the lines proposed in the Sapir-Whorf hypothesis (see LINGUISTIC RELATIVITY HYPOTHESIS; Slobin 1996).

Of primary interest to contemporary theories of production are the processing components dubbed grammatical and phonological in figure 1 (following Levelt 1989). These are the processes immediately responsible for recruiting the linguistic information to create the utterances that convey messages. Grammatical encoding refers to the cognitive mechanisms for retrieving, ordering, and adjusting words for their grammatical environments, and phonological encoding refers to the mechanisms for retrieving, ordering, and adjusting sounds for their phonological environments.

The motivation for separating these components comes from several lines of evidence for a division between word-combining and sound-combining processes. Speech errors suggest that there are two basic sorts of elements that are manipulated by the processes of production, roughly corresponding to words and sounds (Dell 1995). So-called tip-of-the-tongue states (the familiar frustration of being unable to retrieve a word that one is certain one knows) can carry word-specific grammatical information, in the absence of sound information (Miozzo and Carraza 1997; Vigliocco, Antonini, and Garrett 1997). Electrophysiological evidence also suggests that grammatical information about words is accessible about 40 ms before information about sounds (van Turennout, Hagoort, and Brown 1998). Finally, the arbitrariness of the linguistic mapping from meaning to sound creates a computational problem that can only be solved by a mediating mechanism (Dell et al. 1997). These and other observations argue that there are distinct grammatical and phonological encoding mechanisms.

Grammatical encoding Adult speakers of English know between 30,000 and 80,000 words. The average for highschool graduates has been estimated at 45,000 words. These words can be arranged in any of an infinite number of ways that conform to the grammar of English. The ramifications of this can begin to be appreciated in the number of English sentences with 20 or fewer words, which is about 10^20. Using these resources, speakers must construct utterances to convey specific messages. They normally do so in under two seconds, although disruptions are common enough that average speakers spend about half of their speaking time in not speaking—hemming, hawing, and pausing between three and twelve times per minute (Goldman-Eisler 1968). These
disfluencies reflect problems in retrieving a suitable set of words and arranging them into a suitable syntactic structure. What is suitable, of course, is not merely a matter of grammaticality (though it is also that) but of adequacy for conveying a particular message to particular listeners in particular places and times.

Lexical selection and retrieval are integral to grammatical processing because of the kinds of information that words carry about their structural and positional requirements. In everyday language use, words are rarely produced in isolation. Instead, they occupy places within strings of words, with their places determined in part by their grammatical categories (e.g., in most English declarative sentences, at least one noun will precede a verb) and their morphology determined in part by their positions with respect to other words (e.g., a present-tense English verb accompanying a singular subject will be inflected differently than the same verb accompanying a plural subject: *A razor cuts*, whereas *Scissors cut*). Thus, speakers must recover information about grammatical class and morphology.

Lexical selection involves locating a lexical entry (technically, a *lemma*) that adequately conveys some portion of a message, ensuring that there exists a word in one's mental lexicon that will do the job. A rough analogy is looking for a word in a reverse dictionary, which is organized semantically rather than alphabetically. If the desired meaning is listed in the dictionary with a single word that expresses the sense, there is an appropriate word to be had in the language; if not, the search fails. The mental lexicon is presumably accessible in a comparable fashion, permitting speakers to determine whether they know a word that conveys the meaning they intend. Most English speakers, for
example, will find at least one lemma for their concept of a member of the family Oryctolagus cuniculus.

Locating a lemma yields basic information about how a word combines with other words. This corresponds to information about grammatical class (noun, verb, adjective, etc.) and other grammatical features that control a word's combinatorial privileges and requirements (e.g., nouns must be specified as mass or count, and if count, as singular or plural; verbs must be specified as intransitive or transitive, and if transitive, as simple or ditransitive, etc.; cf. LEXICON). The lemma for an instance of Oryctolagus cuniculus, for example, is a noun, count, and singular. These features in turn affect the determination of syntactic functions such as subject phrases, predicate phrases, and their arrangement (cf. SYNTAX).

Once a lemma is found, the word's morphology (technically, its lexeme) may have to be adjusted to its syntactic environment. In connected speech, this will encompass inflectional processes (e.g., making a verb singular or plural). Lexical retrieval processes yield an abstract specification of the morphological structure of the selected word. So, retrieving the lexeme for the count noun that denotes a member of the family Oryctolagus cuniculus should yield a word stem for rabbit.

The role of active morphological processes in language use is currently disputed in some quarters. The issue is whether regularly inflected words are stored and retrieved from memory in the same way as uninflected words (Seidenberg 1997) or require a separable set of specifically inflectional operations (Marslen-Wilson and Tyler 1997). Although this debate has been confined primarily to research on word recognition, logically comparable issues arise regarding language production. In production, however, it may be harder to account for the available data with passive retrieval mechanisms (see Bock, Nicol, and Cutting, forthcoming).

Phonological encoding Words can be comfortably articulated at a rate of four per second, calling on more muscle fibers than may be required for any other mechanical performance of the human body (Fink 1986). Yet errors in the production of speech sounds are rare, only occurring once in every 5,000 words or so (Deese 1984). Controlling the activity requires some specification of phonological segments (/l/, /æ/, /r/, etc.), syllabification, and metrical structure. The broad outlines of phonological encoding are sketched similarly in current theories (Dell et al. 1997; Levelt, Roelofs, and Meyer, forthcoming). Counter to the intuition that words are stored as wholes, sound segments are actually assembled into word forms during the encoding process. Consonantal and vocalic segments must be selected and assigned to positions within syllabic frames. Additionally, the syllables and sounds must be integrated into the stream of speech: In a sequence of words such as rabbit in, the /t/ in rabbit will be produced differently than it is in rabbit by. One implication is that a description of the sound segments in an isolated word is insufficient to explain the word's form in connected speech.

Where theories of word production diverge is in their view of the relationship between phonological encoding and the higher level processes of lexical selection and retrieval. The discrete-stage view (Levelt, Roelofs, and Meyer, forthcoming) argues that each step of the retrieval process is completed before the next is begun (discrete processing), and without feedback to higher level processes from lower levels (strict feedforward processing). In contrast, interactive views (Dell et al. 1997) embrace the possibilities of partial information from one stage affecting the next (cascaded processing) and of information from lower levels affecting higher levels of processing (feedback).

What is at stake in this theoretical battle, in part, is the role that an explanation for speech errors should play in the account of normal speech. Speech errors are a traditional foundation for the study of language production (Dell 1986; Fromkin 1973; Garrett 1988), and the properties of errors have long been viewed as informative about the production of error-free speech. Consider the word exchange by a speaker who intended to say minds of the speakers and instead uttered "speakers of the minds." Or the sound exchange that yielded "accipitral activity" instead of occipital activity. Such errors point to the embodiment of rule-like constraints in the arrangement process. When words exchange, they exchange almost exclusively with other words from the same syntactic class (noun, verb, adjective, and so on). When sounds exchange, they exchange almost exclusively with other sounds from the same phonological class (consonant or vowel). The net effect is that erroneous utterances are almost always grammatical, albeit often nonsensical. In the spirit of exceptions proving the rule, this has been taken to mean that speech errors represent small, local, and most importantly, principled departures from normal retrieval or assembly operations.

Models of word production that incorporate discrete stages have been less successful in accounting for the distribution and the features of speech errors than interactive views, in part because of a difference in explanatory emphasis. Levelt, Roelofs, and Meyer, forthcoming) elaborate a discrete-stage approach that is designed primarily to account for experimental data about the time course of word production, and not for errors, for the simple reason that errors are departures from the norm. What the production system does best, and remarkably well under the circumstances, is retrieve words and sounds appropriate for speakers' communicative intentions. Within the approach endorsed by Levelt, Roelofs, and Meyer, errors are a product of aberrations from the basic operating principles of the production system and are correspondingly rare events. By this argument, errors demand a separate theory.

Despite these differences, the leading theories of language production are in agreement that knowledge about words comes in pieces and that the pieces are not recovered all at once. In normal speaking, the semantic, syntactic, and phonological properties of words are called upon in quick succession, not simultaneously. Thus, what normally feels like a simple, unitary act of finding and saying a word is actually a complicated but fast assembly of separate, interlocking features. More broadly, speaking cannot be likened to the recitation of lines, as E. B. Titchener once did in describing it as "reading from a memory manuscript" (1909). It involves active, ongoing construction of utterances from rudimentary linguistic parts.
Communicative processes All the processes of language production serve communication, but certain activities tailor messages and utterances to the needs of particular listeners at particular places and times. The tailoring requirements are legion. They range from such patent demands as choosing language (English? Spanish?) and gauging loudness (whisper? shout?) to the need to infer what the listener is likely to be thinking or capable of readily recollecting. These are aspects of PRAGMATICS (cf. PSYCHOLINGUISTICS). A common shortcoming of speakers is presuming too much, failing to anticipate the myriad misconstruals to which any given utterance or expression is vulnerable. The source of this presumptuousness is transparent: Speakers know what they intend. For them, there is no ambiguity in the message.

The direct apprehension of the message sets speakers apart from their listeners, for whom ambiguity is rife. In this crucial respect, language production has little in common with language comprehension. In other respects, however, successful communication demands that production and comprehension share certain competencies. They must somehow draw on the same linguistic knowledge, because we speak as well as understand our native languages.

The cognitive processing systems responsible for comprehension and production may nonetheless be distinct. Research on language disorders suggests a degree of independence between them, because people with disorders of production can display near-normal comprehension abilities, and vice versa (Caramazza 1997). At a minimum, the flow of information must differ, leading from meaning to sound in production and from sound to meaning in comprehension.

This simple truth about information flow camouflage the deep questions that are at stake in current debates about the isolability and independence of the several cognitive and linguistic components of production. The questions are a piece of the overarching debate about modularity, to do with whether language and its parts are in essence the same as other forms of cognition and more broadly, whether all types of knowledge are in essence the same in acquisition and use. Accordingly, answers to the important questions about language production bear on our understanding of fundamental relationships between LANGUAGE AND THOUGHT, between free will and free speech, and between natural and artificial intelligence.

See also CONCEPTS; CONNECTIONIST APPROACHES TO LANGUAGE; MEANING; NATURAL LANGUAGE GENERATION; SENTENCE PROCESSING

—Kathryn Bock

References


Bock, J. K., J. Nicol, and J. C. Cutting. (Forthcoming). The ties that bind: Creating number agreement in speech. Journal of Memory and Language.


