Psychonomic Bulletin & Review
1996, 3 (4), 395-421

Language production: Methods and methodologies

KATHRYN BOCK
University of Illinois at Urbana-Champaign, Urbana, Illinois

Methodological problems have been a longstanding barrier to the systematic exploration of issues in language production. Recently, however, production research has broadened beyond traditional observational approaches to include a diverse set of experimental paradigms. This review surveys the observational and experimental methods that are used to study production, the questions to which the methods have been directed, and the theoretical assumptions that the methods embody. Although tailored to the investigation of language production, most of the methods are closely related to others that are widely employed in cognitive research. The common denominator of these procedures is verbal responding. Because the processing complexities of verbal responses are sometimes overlooked in research on memory, perception, attention, and language comprehension, the methodological assumptions of production research have implications for other experimental procedures that are used to elicit spoken words or sentences.

The commonplace standard of skill in a language is the ability to speak it fluently. Most people have this skill in their native tongue. It comes as no surprise, then, that language production is one of the three core topics in psycholinguistics, along with language comprehension and language acquisition. Yet a common preface to textbook discussions of language production is a notification to the reader that the discussion will be brief, speculative, or both. The reason usually offered is that, in contrast to research on language comprehension, research on language production is scarce. Garnham wrote that "it is easier to study language understanding than language production, and comprehension has therefore been more widely investigated" (1985, p. 205). D. W. Carroll sounded the same theme, saying that "far more is presently known about receiving language than producing it" (1994, p. 190).

The paucity of production research is typically attributed to the problems of achieving the ideals of experimental control and measurement. Production is "an intrinsically more difficult subject to study than language comprehension" (D. W. Carroll, 1994, p. 190), because it is "extremely difficult to perform experiments dealing with production processes" (Foss & Hakes, 1978, p. 171). It is hard to control the input to language production processes in the way that the input to language comprehension can be controlled and, in the face of the diversity of the output, even harder to develop a defensible set of response measurements. Two decades ago, the consequence was that "practically anything that one can say about speech production must be considered speculative even by the standards current in psycholinguistics" (Fodor, Bever, & Garrett, 1974, p. 434).

When one extends one's sights beyond the traditional experiment, the quantity and diversity of information available about production is overwhelming. Speech fills the air and, as Levelt (1989) observed, different facets of these abundant data are the province of disciplines that run the gamut from artificial intelligence through articulatory phonetics to psychoanalysis, rhetoric, and sociolinguistics. Analyses of spoken language from all of these disciplines offer valuable clues about production processes to anyone who is willing to look. But the business of psycholinguistics is to turn these clues into empirically testable hypotheses about the mechanisms that convert thought into speech. Presently, research on language production is undergoing a rapid transformation from an observational enterprise to one with a set of experimental paradigms and modeling techniques for examining different kinds of questions.

Behind much of the emerging research on production is a framework that is sketched in Figure 1. It includes three processing components. The first component creates a nonverbal *message*, which represents what the speaker intends to communicate. The second component, *grammatical encoding*, encompasses the selection of semantically appropriate words (by locating lexical entries—technically, *lemmas*—in the mental lexicon) and the assignment of the lemmas to roles in a syntactic structure. The third component, *phonological encoding*, is responsible for spelling out the sound forms of the words (technically, their *lexemes*) and the prosodic properties of the utterance as a whole (the utterance's "musical" features, including qualities corresponding to tempo, rhythm, pitch, and timbre). The output systems guide the actual production of the ut-

I would like to thank G. S. Dell, G. L. Murphy, J. Stemberger, G. Vigliocco, and an anonymous reviewer for their comments on earlier drafts of this review. Preparation of the paper was supported in part by grants from the National Institutes of Health (R01 HD21011) and the National Science Foundation (SBR 94-11627). Correspondence should be addressed to K. Bock, Beckman Institute, University of Illinois, Urbana, IL 61801 (e-mail: kbock@s.psych.uiuc.edu).

Copyright 1996 Psychonomic Society, Inc.

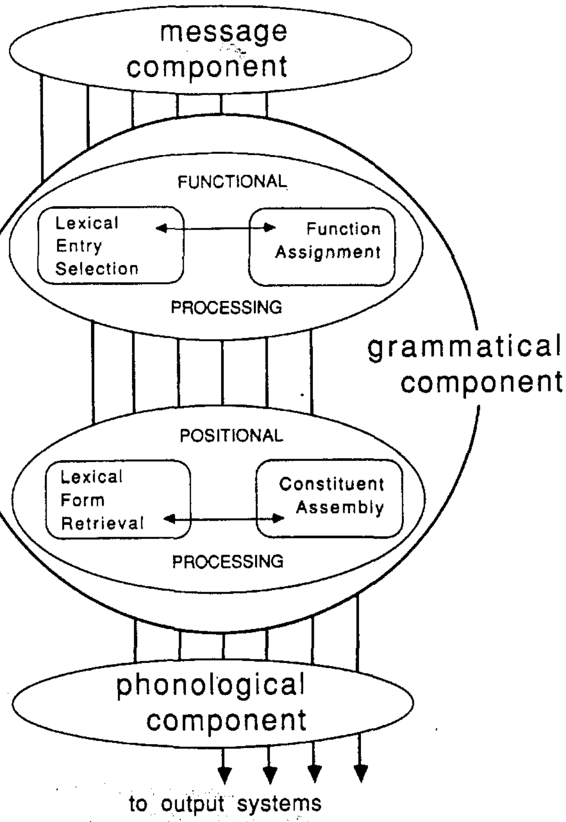

Figure 1. The components of language production. The links among the components are staggered from left to right to represent the flow of information over time. From "Sentence Production: From Mind to Mouth" (p. 185), by J. K. Bock, in J. L. Miller and P. D. Eimas (Eds.), *Handbook of Perception and Cognition: Speech, Language, and Communication*, 1995, San Diego: Academic Press. Copyright 1995 by Academic Press. Reprinted with permission.

terance. Levelt (1989) reviewed and analyzed much of the literature relevant to processing in these components. More recent perspectives can be found in Bock and Levelt (1994) and Bock (1995).

The implications of research on language production extend well beyond the narrow confines of psycholinguistics. Because spoken responses are widely employed in psychological research on memory, attention, and perception, an understanding of the cognitive precursors of speech is important to many experimental paradigms. However, the contribution of verbal output systems to the performance of experimental tasks is sometimes overlooked. The implicit rationale for this oversight might be called the *mind-in-the-mouth assumption*. The assumption is that the properties of a stimulus's mental representation are transparently reflected in the verbal response to the stimulus. This assumption in turn motivates an erroneous supposition that is pervasive in the literature: What one says, how one says it, and how long it takes to say it are unsullied reflections of input processing and interpretation. One result is that perceptual and cognitive explanations of task performance are couched almost exclusively in terms of transformations and interpretations of the input; only rarely are the contributions of a complex series of production processes acknowledged. This oversight goes hand in hand with the scantiness of theoretical and empirical attention to language production mechanisms.

In what follows I first sketch some of the manifestations of the mind-in-the-mouth assumption, in order to highlight the general difficulty of disentangling the mechanisms of production from those of encoding and understanding information. This sets the stage for a methodologically oriented survey of research on language production. The survey is accompanied by sketches of the questions to which production research has been directed, along with analyses of the processing assumptions that underlie the methods that have been employed. The goal is to provide a wide-ranging synopsis for those who are unfamiliar with the methods and results of production research and, for those who know the methods and results well, to make explicit the major assumptions and problems of the methods that are in current use.

The Mind-in-the-Mouth Assumption: Verbal Responses in Cognitive Research

In experimental psychology at large, language production is far more familiar as a convenient modality of response than as a topic of theoretical interest. Verbal responses are widely used for assessing the consequences of immediate perception in traditional perceptual tasks and for measuring the capacity of immediate memory in traditional short-term memory tasks. Verbal response plays a similar part in assessing higher level visual processing and categorization. In research on word recognition in reading and listening, the production (naming) of words often serves as an index of successful recognition. In the study of language comprehension, a long line of research relies on the (re)production of sentences and entire discourses. In each case, the properties of the verbal responses themselves and the processes that produce them may receive little attention beyond the contributions of a globally characterized response stage or a box in an information processing flowchart that is labeled "output."

It is easy to overlook the intricacies of deploying a verbal response, and psychology has a long tradition of doing so. In 1909 E. B. Titchener delivered a lecture at the University of Illinois in which he claimed to be able to "read off what I have to say from a memory manuscript" (1909, p. 8), a statement that captures a commonplace view of speaking. Cognitive psychologists have become sensitive to the fact that processing a stimulus requires sensory, perceptual, and memory processes on an order of complexity that is unexpected from normal experience. However, beyond their motor components, the processes that create responses are often conceived in a fashion stark enough to satisfy the strictest behaviorist. The implicit assumption is that responses transparently reveal the cognitive consequences of recognition and comprehension—the mind in the mouth. Some of the consequences of this assumption are examined below in the context of studies of picture naming (including verbal responding in the Stroop task), categorization, word recognition, and sentence comprehension.

Picture naming and Stroop naming. The picture-naming task requires subjects to respond to pictures with an appropriate word, typically a noun denoting the picture's basic-level category. This naming episode is sometimes preceded or accompanied by a prime or distractor, in the form of a word or another picture. The Stroop task, in its most familiar variants, requires naming the ink color in which a word is printed or naming a picture upon which a written word is superimposed. In the color-word version, the word denotes a color that is congruent or incongruent with the ink color. For example, the word *blue* printed in blue ink would be a congruent stimulus (since the correct naming response fits both the ink color and the word itself), whereas the word *red* printed in blue ink is incongruent. This task, like picture naming, demands processing within the language production system to support word naming. (Comprehensive reviews of research on the naming of pictures and Stroop stimuli can be found in Glaser, 1992, and MacLeod, 1991.)

Because picture naming and Stroop naming are widely used, they offer a representative glimpse of the part that hypotheses about language production play in theorizing about cognitive performance. It is a bit part. This is not because investigators are unaware that responding is critical in these tasks. Particularly in the Stroop task and variants of it, investigators have long recognized that much of the interference that is observed may arise during responding. Despite this, few efforts have been made to analyze verbal response processes into their cognitive components. Instead, much research assumes an undifferentiated "response stage," which may be interpreted as comprising little more than articulation.

Consider Figure 1 again. In terms of the model shown there, naming demands that a stimulus first be represented in a way that allows it to make contact with the lexicon. That is, it requires the mental representation of a referent and a meaning—a "message"—even when all that is required is naming an object or a color. In the case of color naming, the message may be a categorization of the stimulus (e.g., a categorization of the color perceived under normal conditions in light at a wavelength of 475 millimicrons; perhaps "color of the clear daytime sky"). The message must locate an appropriate lexical entry (a lemma), which yields information about how the word that it represents is normally used in utterances, whether as a noun, a verb, or other part of speech; if as a noun whether it is count or mass; if count whether it is singular or plural, and so on. So, an entry for a word used to denote a color like that of the clear daytime sky should give access to usage information along the lines of [NOUN [mass]]. Locating a lexical entry establishes that a fitting single word exists in the speaker's mental lexicon for conveying a message. The ease of selecting the entry for production may depend on such factors as the frequency with which it is accessed (that is, lemma frequency; see Dell, 1990, Experiment 3 and discussion) and typicality or codability (R. Brown, 1976; Lachman, Shaffer, & Hennrikus, 1974). Once the entry is selected, the sounds of the word must be retrieved and assembled (e.g., [b][l][u]). This process is also frequency sensitive (Jescheniak & Levelt, 1994) and sensitive to the abstract phonological structure of the word (Dell, 1986). Only after selection and retrieval can articulation proceed.

Experimental tasks make different demands on these processes. For example, the traditional Stroop color-word task calls on response processes that must retrieve the name of the color of the ink in which the distractor is printed. This requires locating the lexical entry that is appropriate for conveying the ink color (e.g., "color of the clear daytime sky: NOUN [mass]") and retrieving its phonological form ([b][l][u]). In addition, the phonological form of the distractor word (the pronunciation of the written word; e.g., [r][ɛ][d]) is directly activated from the input and set into competition with the phonological form [b][l][u] retrieved via the lexical entry.

In an interactive-activation model like that of Dell (1986), activation spreading up from the distractor's word form to its associated lexical entry creates one source of interference in the Stroop task, allowing the distractor to vie with the entry for the ink color for encoding. Further interference can arise during assembly of the word form, particularly when word forms corresponding to the activated entries share phonological features (Dell, 1984).[1] Eliminating the verbal response can reduce phonological interference, but competition among lexical entries may remain if lexical selection is needed (for instance, to mediate a button-press response; Keele, 1972). Only if lexical selection and production are bypassed entirely should it be possible to eliminate Stroop-like effects (McClain, 1983; see MacLeod, 1991, for discussion). All of these are response processes, but they go well behind articulation to the cognitive mechanisms of language production that make articulation possible.

An analogous but more straightforward argument applies to picture naming. Glaser (1992) took some steps toward integrating existing chronometric research on picture naming and primed picture naming with questions about language production. However, outside of the production literature itself, there is little acknowledgment of the complexity of the production processes that subserve picture naming and Stroop naming.

Categorization. Closely related to the issues in picture naming are questions about categorization. Here as well, it is common to call on tasks that require verbal responses and to treat the responses as direct revelations of mental attention, thought, or opinion. A study by Higgins, Bargh, and Lombardi (1985) provides an interesting example. Higgins et al. presented ambiguous personality descriptions like the following, and asked students to categorize the person described: "Other than business engagements, his contacts with people are rather limited. He feels he doesn't really need to rely on anyone." This individual might be called "aloof" or "independent." What Higgins et al. discovered was that the tendency to use one or the other of these terms varied as a function of earlier experience with the words themselves. Prior to making the personality categorization, the students had been exposed to one or the other of the terms, either *aloof* or *independent*. Because the students were more likely to use the primed term in their categorization, the researchers concluded that the students' interpretation of personality characteristics changed as a consequence of exposure to different words.

Notice that the verbal response offered the main evidence for a difference in the underlying evaluations of the person described. This verbal response, either "aloof" or "independent," was primed. Speakers are likely to use (and overuse) recently encountered words when other words would do just as well (Kubovy, 1977); many of the phenomena of implicit memory depend on just this. For example, patterns of word association change depending on prior exposure to alternative associates, even when the initial exposure occurred under anesthesia (Kihlstrom, Schacter, Cork, Hurt, & Behr, 1990). Likewise, speakers occasionally replace one word with another from the same semantic or pragmatic domain (Hotopf, 1980): Everyone is acquainted with such errors as saying *left* when *right* was quite consciously intended, or addressing a child by the name of a sibling—or the name of the family dog. For these reasons, it is a long and uncertain step to the conclusion that variations in response patterns directly reflect differences in underlying evaluations. Conceivably, the students in the Higgins et al. (1985) experiment intended to convey none of the affective nuances implied by *aloof* versus *independent*, but simply produced an accessible word that was broadly compatible with the description. Similarly, instances of so-called unconscious plagiarism in single-word production tasks (A. S. Brown & Murphy, 1989) may reflect only transient patterns of activation in the lexicon.

Word recognition. At least since Morton (1969), a typical view of the process of reading aloud is that the reader speaks a word that is made available by the recognition system. Basic findings in both of the field's paradigmatic tasks, lexical decision (word–nonword decisions) and naming (word pronunciation), are interpreted primarily in terms of hypothesized properties of recognition. A controversy over the role of production processes in creating frequency effects in the naming task (Balota & Chumbley, 1985, 1990; Monsell, 1990, 1991; Monsell, Doyle, & Haggard, 1989; Paap, McDonald, Schvaneveldt, & Noel, 1987; Savage, Bradley, & Forster, 1990) exposes how limited the role accorded to production can be: By "production," little more may be meant than the issuing of motor commands to the articulators (see Monsell, 1987, for a more detailed conceptualization).

A similar perspective can be found in the first groundbreaking efforts to use neuroimaging techniques for localizing the cognitive processes engaged in word recognition. Petersen, Fox, Posner, Mintun, and Raichle (1989) used positron emission tomography (PET) in an attempt to localize the processes involved in the comprehension of words. In one of their manipulations, they addressed the semantic processing of visually presented words such as *hammer*, asking subjects to generate and produce words for actions typically carried out with the object denoted (e.g., *pound*). To isolate the semantic processing of *hammer*, they subtracted the cortical activation patterns produced when the subject saw the word *hammer* and said it aloud from the patterns detected in the verb-generation conditions. The reasoning was that the difference between the two sets of patterns represented semantic processing, with visual processing and articulation stripped away. Overlooked in this comparison was the nature of the processing that is required to support the voluntary retrieval and phonological assembly of a spontaneously generated word (*pound*). More recent research places questions about retrieval at center stage (e.g., Warburton, Wise, Price, Weiller, Hadar, Ramsay, & Frackowiak, 1996), but still overlooks the cognitive architecture of normal language production in interpreting the results.

There have been a few notable efforts to disentangle the recognition-dependent characteristics of visual-word processing from characteristics that are due to higher level production components of word-naming tasks (e.g., Forster & Davis, 1991). Still, without detailed analyses of production processes and production data, it is easy for even well-intentioned efforts to come up short. For instance, Mozer (1983) examined the phenomenon of letter migration in word recognition, in which a letter from one word is illusorily seen as belonging to another. Subjects in Mozer's study viewed a briefly presented word pair such as *line–lace*, and were then cued to report one of the two words aloud (e.g., "line"). Mozer found an increased incidence of errors such as saying "lice" instead of "line." These errors were attributed to the migration of a letter from the nontarget word (the *c* from *lace*) into the target.

Because the erroneous response is also a type of speech error in which the /s/ ending from *lace* is anticipated as the ending of *line*, Mozer (1983) conducted a second experiment to explicitly assess the contribution of speech errors to task performance. However, the only responses that were judged to be speech errors were exchanges such as "lice–lane." Since errors such as "lice–lace" were about 10 times more frequent than exchanges in the data, Mozer argued against a production explanation. Yet in spontaneous speech-error corpora, anticipations are about 10 times more frequent than exchanges (Nooteboom, 1973).

The mind-in-the-mouth perspective is also discernible in studies of spoken word recognition. When spoken or written reproduction of an auditorily perceived word is taken as the measure of successful identification (Goldinger, Luce, Pisoni, & Marcario, 1992; Marslen-Wilson, 1973; Radeau, Morais, & Dewier, 1989; Slowiaczek & Hamburger, 1992), the processes of production come into play. However, the task analysis and explanatory framework are likely to center exclusively on recognition mechanisms. For example, Marslen-Wilson's (1973) classic work on speech shadowing employed speakers who shadowed sentences aloud, engaging the entire production system. However, the point of the work was the speed of word recognition.

Again, it is difficult to reconcile this narrow focus with the known facts of word production. Consider the role of frequency, which is the most widely studied variable in word recognition. Word frequency has a similar relation to response times in picture naming, a production task, and in lexical decision, a recognition task. In picture naming, the time to produce the target word is an inverse logarithmic function of the word's frequency (Jescheniak & Levelt, 1994; Oldfield & Wingfield, 1965), and in lexical decision, the time to make a word–nonword judgment is likewise an inverse logarithmic function of the word's frequency (Forster, 1990). These functions may be roughly similar, as Figure 2 suggests. The figure shows a scatterplot of the picture-naming times for the words used by Oldfield and Wingfield (1965) and the lexical decision times for the words from Forster and Chambers's (1973) plotted against the logarithm of word frequency (Francis & Kučera, 1982). What makes this convergence worthy of note is that in the former case the words are the responses, whereas in the latter they are the stimuli.

Is there any convergence in the operations of recognizing and producing words that might begin to account for this coincidence? The frequency effects in picture naming cannot be ascribed to processing the picture, either the encoding of the visual array or the identification of the depicted object (Jescheniak & Levelt, 1994; Wingfield, 1968), and the effects are too large to be explained in terms of articulatory effects alone (like those reported by Balota & Chumbley, 1985; see Jescheniak & Levelt, 1994). This leaves the processes called lexical entry selection and lexical form retrieval in Figure 1.

To separate their respective contributions to production, one can elicit high- and low-frequency words that have the same phonological forms (e.g., *him* vs. *hymn*) and compare them to frequency-matched nonhomophones. If frequency effects reside in lexical entries, homophones and nonhomophones should behave similarly in frequency-sensitive tasks, but if frequency effects reside at the level of phonological forms, low-frequency homophones should behave more like high-frequency nonhomophones. Using different variations on this logic, Jescheniak and Levelt (1994) and Griffin (1995) found that the most robust frequency effects in picture naming seem to arise within the processes that retrieve the phonological representation of a to-be-spoken word form (i.e., in lexical form retrieval; see also Dell, 1990).

Given these results, one interpretation of the convergence in the frequency functions is that some of what is attributed to frequency in word recognition may be associated with the processes of retrieving and assembling a phonological word form, whether for pronunciation or for lexical identification. The word-recognition results of McCann, Besner, and Davelaar (1988) point in the same direction, although the authors characterized their phono-

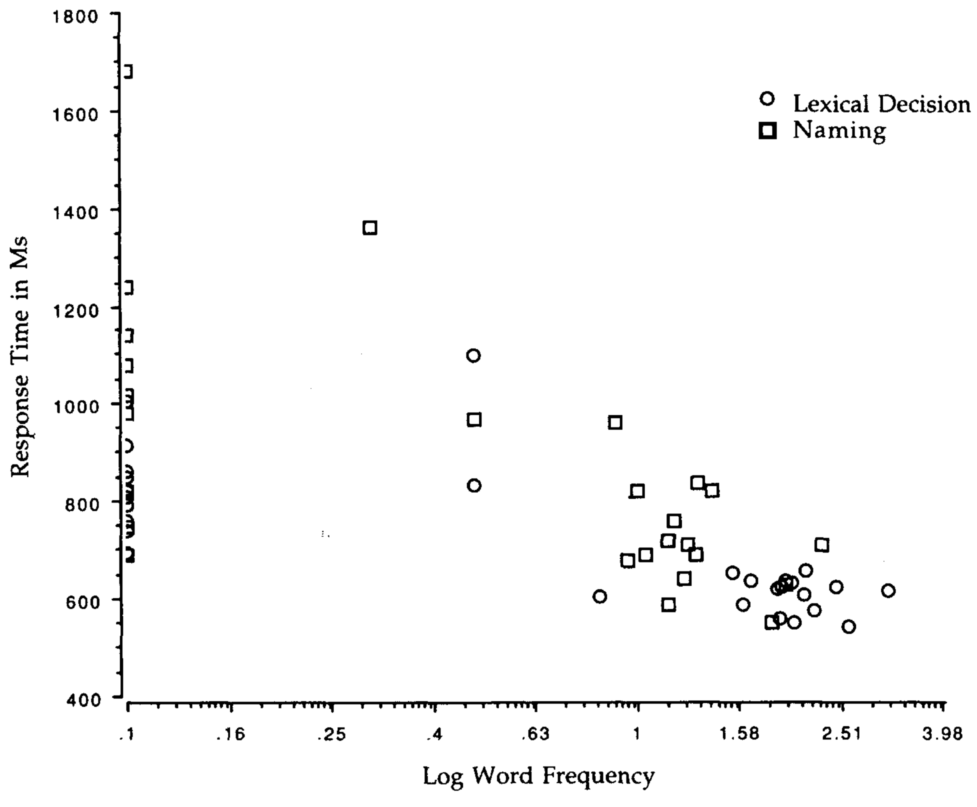

Figure 2. Picture naming latencies from Oldfield and Wingfield (1965) and lexical decision times from Forster and Chambers (1973) plotted against the logarithm of word frequency (Francis & Kučera, 1982).

logical frequency effects as the product of bias within a decision stage for lexical identification.

It is also possible to explain variations in frequency effects in different word-recognition tasks in terms of differential reliance on components of the lexicon that are shared by recognition and production processes (Monsell, 1987). Consider Dell's (1990) modeling demonstration that lexical-entry-based frequency differences can explain variations in the incidence of phonological speech errors, which arise during lexical form retrieval. An analogous account may help to explain the typical disparity in the magnitude of the word-frequency effect in lexical decision (where frequency effects are relatively large) and naming (where frequency effects are relatively small). Specifically, when the pronunciation of a visually or auditorily presented word can be accomplished by bypassing the lexical entry and directly activating phonological segments, the only frequency effects observable will be those due to feedback rebounding to the segments from the indirectly activated lexical entry. Lexical decision is more likely to reflect any differences in frequency that are encoded at the lexical entry, because it is more likely to call on the lexical entry in the first place. Consistent with this account, considerable work in the word-recognition literature shows that frequency effects in naming and lexical decision become more similar when naming requires accessing the lexical entry (for instance, when the task is to pronounce real words but not nonwords; Forster, 1990; Paap et al., 1987).

This last argument is a simple extension of models of reading according to which words are recognized either as whole visual patterns that associate directly to lexical entries or as a result of mappings from spelling patterns to sound patterns (e.g., the dual-route theory of visual word recognition; Coltheart, Curtis, Atkins, & Haller, 1993). What is gained from the extension is an integrated and elaborated view of lexical processing from recognition through production. Such a view is amenable to powerful experimental tests, coupling word-recognition tasks to word-production tasks for convergent validation.

Language comprehension and language memory. In the study of language comprehension and memory, there are even more formidable obstacles to inferring the processes of interest directly from the characteristics of verbal responses. Sentence recall offers an example. Recall requires comprehension, and so recall measures are sometimes used to assess variations in language understanding. However, as observed earlier, recall is also heavily influenced by the processes of language production (see Bock & Brewer, 1974; Levelt & Kempen, 1975). Unrecognized, this contribution can create misinterpretations. For example, studies of sentence recall beginning in the 1960s showed that active sentences are more likely to be recalled than passives (Mehler, 1963). A common interpretation of this result is that actives are easier to understand. The data are inadequate to support this conclusion, however. For a host of reasons (see E. V. Clark & H. H. Clark, 1978, for some of them), people tend to *produce* actives rather than passives, and one consequence is that many sentences that are presented as passives are recalled as actives. When this is taken into account, it appears that passives and actives can be understood equally well (James, Thompson, & Baldwin, 1973). What differs is how they are reconstructed—reproduced—in the production phase that recall requires.

The same conclusion applies to immediate recall. The immediate recall of verbal materials is traditionally regarded as the emptying of a short-term store. However, the process may be better likened to one of producing a response by assembling highly activated linguistic elements, using the mechanisms of production to do so. This introduces subtle changes into the recalled material (Lombardi & Potter, 1992; Potter & Lombardi, 1990).

Strikingly, the amount of information that *can* be veridically reproduced from immediate memory is itself tied to the lower level processes of production: The duration of traditionally defined short-term memory traces is highly correlated with rate of articulation (Baddeley & Hitch, 1974; Baddeley, Thomson, & Buchanan, 1975; Schweickert & Boruff, 1986). Generally, people can immediately recall about as many items as they can pronounce in 2 sec. This interacts with differences among languages in the normal sizes of the lexical units employed in short-term memory tasks and with developmental differences in speech rate (Cheung & Kemper, 1993; Ellis & Hennelly, 1980; Hoosain, 1982; Hulme, Thomson, Muir, & Lawrence, 1984; Kynette, Kemper, Norman, & Cheung, 1990; Stigler, Lee, & Stevenson, 1986; Zhang & Simon, 1985). The appearance is that standard ways of assessing both long- and short-term memory are tightly interwoven with the processes of normal language production.

A different concern about covert production effects arises in connection with the norming tasks that are sometimes employed to create materials for experiments on language comprehension. One common norming technique relies on the elicitation of completions of sentence fragments. High- versus low-frequency completions then serve as stimuli in experimental conditions designed to assess the impact of highly plausible versus less plausible materials on such processes as structural ambiguity resolution. For example, *the song* in *The singer remembered the song* ... is likely to be taken as the direct object of *remembered*, whereas in *The singer thought the song* ..., *the song* is more likely to be taken as the subject of an upcoming clause, as in *The singer thought the song was unsingable*. The use of production norms for creating highly plausible materials could not be faulted if plausibility alone were the result. However, production norming insures that the materials will be not merely plausible, but also highly predictable. The concern then becomes that effects attributed to normal processes of ambiguity resolution in comprehension are in fact highly atypical consequences of expectations that make use of the machinery of language production, effects that are customarily regarded as response biases.

Summary. Studies of picture recognition, Stroop interference, linguistic categorization, visual and auditory word recognition, and language comprehension often tap components of the language production system as well as input systems. In order to arrive at models of recognition, comprehension, and production tasks that can be integrated into an overarching theory of normal language performance, it is essential to consider what the relevant components of processing may be. In principle, language tasks that call on recognition, comprehension, and production offer the possibility of converging on an interpretation of the knowledge that supports language processing. In practice, the achievement of this goal requires a more analytic consideration of experimental tasks than the tasks sometimes receive in studies of human information processing.

One step in this analysis is to understand the methods and the methodology of current research in language production. By *methods*, I mean the experimental techniques themselves. By *methodology*, I mean to denote the logic of the experimental tasks, in terms of their processing analyses and assumptions within the broad framework shared by virtually all current language production models. The exposition of these methods and their rationale is the purpose of the survey that follows. The emphasis is not on reviewing the substantive results of production research (for that see Bock, 1987a, 1995; Bock & Levelt, 1994; Levelt, 1989), but on examining some of the new purposes to which some familiar experimental procedures have been applied. Throughout, the focus is on studies of grammatical and phonological encoding, because it is these topics that intersect most often and most directly with other areas of cognitive research.

THE METHODS OF LANGUAGE PRODUCTION RESEARCH

The survey begins with standard observational methods and the issues that they address. Consideration of these methods paves the way for an analysis of emerging experimental techniques and the assumptions behind them. In both cases, the review emphasizes research on normal adult speech, with occasional tangents into areas that diverge from that standard.

Observational Methods

One side benefit of the barriers to applying experimental logic to language production was the emergence of a sizable contemporary literature that includes elegant and insightful analyses of spontaneous speech and its problems (see Fromkin, 1971; Garrett, 1975; and Goldman-Eisler, 1968, for some of the beginnings of this work). Until quite recently, in fact, almost all research on language production involved the observation of everyday talk. This research has served to identify basic problems and establish widely shared assumptions, blazing a trail through rough terrain. Because many issues from this literature have found their way into current experimental work, the research offers a useful introduction to some basic questions.

Many of these observational studies investigate departures from fluent speech, including hesitations or outright errors. Less often, statistical analyses of large speech samples have been carried out to determine distributional features. The latter approach has gained in popularity with the increased availability and power of computational methods for corpus research. This section begins with a brief overview of the resources that are presently available for corpus-based studies of spoken language, and then turns to the procedures that are commonly used in observational research on speech errors and other dysfluencies.

Distributional analyses of spontaneous speech. A valuable but underused option in observations of language production is to examine the distribution of linguistic features of all kinds in large samples of normal speech. A major barrier has been the scarcity of publicly available, transcribed-speech corpora. For many years, the only alternative to do-it-yourself recording and transcription was the London-Lund corpus (Svartvik & Quirk, 1980), which contains prosodic transcriptions of tape-recorded British English speech. Frequency counts for spoken words are also lacking (see Howes, 1966, for an exception), but this is changing as machine-readable corpora appear. These corpora include the CHILDES database (MacWhinney & Snow, 1990), which offers samples of adult speech to children as well as child speech, and is available on CD-ROM.[2] Other sources of transcribed speech, such as Bartlett's (1937) *Familiar Quotations* and the White House transcripts (Gold, 1974), can sometimes be used effectively for special purposes, given adequate precautions against their significant and idiosyncratic limitations.

An interesting analysis of distributional results can be found in Deese (1984). Deese examined the transcriptions of 25 h of tape-recorded speech from American speakers, looking for occurrences of various sentence types, verb forms, prosodic markers, and speech errors, among other things. Other examples of corpus-based work include that of Clark and Fox Tree (cited by Smith & Clark, 1993), who used the London-Lund corpus to locate distinguishing features of the uses of "um" and "uh." Kelly (1986, 1988; Kelly & Bock, 1988) relied on Bartlett (1937) and the White House transcripts to explore the prosodic contexts of alternative lexical stress patterns and the phonological and semantic factors that are associated with alternative word orders.

A more focused but still naturalistic approach to the analysis of spontaneous speech is possible when the corpus is restricted to speech recorded under controlled conditions. Goldman-Eisler (1968) pioneered this approach. More recently, Berman and Slobin (1994) used the technique to examine developmental and cross-language variations in the linguistic devices used in telling a single story about a boy, his dog, and a missing frog. All of the speakers first viewed the story in pictures alone, and then narrated it in their own words. By correlating elements of the events and their sequence with the linguistic devices used to convey them, Berman and Slobin were able to draw inferences about the relative ease with which speakers of different languages encode different semantic functions.

There is one publicly available spoken corpus of conversational speech elicited in a constrained setting. It contains both speech and transcripts from an extensive set of dialogues between Scots English speakers, and is available on CD-ROM.[3]

Dysfluency in spontaneous speech. Many of the existing distributional studies are concerned in one way or another with dysfluencies. *Dysfluencies* are generally distinguished from speech errors as interruptions to the stream of speech. They come in several varieties. The usual taxonomy follows Maclay and Osgood (1959) and includes hesitations (silent pauses), filled pauses (such as *hmm*, *umm*, *er*, and *uh*), false starts, and repetitions (of sounds, words, or whole phrases). Closely related in purpose to many studies of dysfluency are studies that deal not with dysfluencies per se, but with the durations of various grammatical junctures that naturally punctuate speech (e.g., Ford, 1982; Holmes, 1988) and with the occurrence of gesture and other nonverbal behaviors that accompany speech (e.g., Beattie, 1983, chap. 4; McNeill, 1992).

Dysfluencies have diverse causes. Garrett (1982) observed that hesitations (filled and silent pauses) may reflect transient increases in processing load, normal advance planning and retrieval for an upcoming structural unit, or delay created by the momentary inaccessibility of a needed piece of information. Garrett dubbed these three culprits "don't bother me, I'm busy," "wait till the boat's loaded," and "it's in the mail," respectively. So, the fact that speakers often pause after an initial function word (such as *the*) and before the first content word in a clause or sentence (Boomer, 1965) can be explained in at least three ways: (1) as a tendency to reconsider what one is about to say before one says it ("don't bother me, I'm busy"); (2) as time spent in planning the details of the next constituent ("wait till the boat's loaded"); or (3) as a delay in the retrieval of the next word to be spoken ("it's in the mail").

Of these possibilities, "loading the boat" has received the most concerted attention. The goal in this work is to isolate the locus and scope of forward preparation in speech (Butterworth, 1980), examining how far ahead speakers plan and where they tend to hesitate while doing the cognitive work needed for an upcoming stretch of speech. In these investigations, researchers have explored the distributions and durations of dysfluencies to assess whether single words or larger units function as minimal planning ranges (Boomer, 1965; H. H. Clark & E. V. Clark, 1977, chap. 7; Fodor et al., 1974, chap. 7; Ford & Holmes, 1978; Lounsbury, 1965; for reviews see Bock & Cutting, 1992; Garrett, 1982). Similar techniques have been applied to questions about global planning in the production of monologue (Beattie, 1983; Deese, 1984, p. 104). Other investigators (Blackmer & Mitton, 1991; Levelt, 1983; Nooteboom, 1980) have examined interruptions in speech with an eye toward "backward" attention, or how speakers detect and correct errors already made.

A major methodological issue in some of this work relates to the classification of observed dysfluencies into one or another category. Lying behind this issue are questions about whether differences in the surface forms of dysfluencies reflect different production problems (of the three sorts noted by Garrett, 1982) or different kinds of preparation (e.g., conceptual, lexical, or syntactic preparation). For many dysfluencies, there are reasons to suspect that they do not have equivalent sources. Maclay and Osgood (1959) found few sizable correlations among the four types of dysfluencies they examined, either within or across speakers. Deese (1980) reported that pauses (both filled and silent) were more likely in planned than in unplanned speech, whereas false starts, corrections, and repetitions were more likely in unplanned speech. Levelt's (1983) analysis of repairs revealed that different kinds of production problems elicited different hesitation markers. Smith and Clark (1993) showed that even the choice between "um" and "uh" can have a functional underpinning in the speaker's tacit evaluation of the likelihood of retrieving the answer to an information question (although hearers do not seem to capitalize on the distinction; Brennan & Williams, 1995). The natural trend thus seems to be toward different distributions for different types of dysfluency.

The importance of accurate classification has spawned a perennial controversy over the temporal and distributional criteria for identifying silent pauses as dysfluencies (Butterworth, 1980). Silence fills a sizable portion of spontaneous speech, on average between 40% and 50% of speaking time (Goldman-Eisler, 1968), but its occurrence need not indicate uncertainty or mental difficulty on the part of the speaker. Some pauses mark linguistic boundaries, such as clause boundaries. Other pauses may be used for purely stylistic effect. Yet others may be designed to aid the hearer's interpretation. And some silences are prosodically controlled (Ferreira, 1993; Grosjean, Grosjean, & Lane, 1979; A. S. Meyer, 1994) or purely articulatory (such as those that accompany the closure in a stop consonant). So, when is silence part of the stream of fluent speech, and when is it dysfluent? For identifying linguistically conditioned gaps in the speech stream, some investigators specify a ceiling of 200–250 msec (Butterworth, 1980), but for other silent pauses, there is no simple convention for differentiating alternative contributors to pause durations. The quest for a defensible taxonomy occupies an enormous literature (see Drommel, 1980, and other papers in Dechert & Raupach, 1980).

Silent pauses can be and usually are identified mechanically from acoustic records, but most research on dysfluency relies at some point on the judgments of listeners to detect that a dysfluency has occurred. The reliability and validity of these judgments has been examined in several studies. Maclay and Osgood (1959) reported an average 95% agreement between two judges scoring four types of dysfluencies, and Schachter, Christenfeld, Ravina, and Bilous (1991) reported reliabilities of .98 and .99 for the detection of filled pauses. J. G. Martin and Strange (1968, Experiment 4) examined filled- and silent-pause detection by naive judges across a range of different listening conditions. The naive judges detected 95% of the pauses marked by more experienced judges. For silent pauses, J. G. Martin (1970) compared the judgments of two naive listeners to each other and to spectrographic records. The agreement between the judges was 92%, and between the judges and the spectrograph, 90%. The durations of pauses can also be subjectively evaluated. Subjective judgments of duration correlated .85 with the measured length of pauses in a study reported by Deese (1980).

Though listeners are good at detecting genuine dysfluency, they falter in an important way: J. G. Martin (1970) found that a slowing of speech was likely to be perceived as a pause, even when no silence occurred. This points indirectly to a major limitation on the use of pauses as tools for exploring the cognitive effort associated with production. Since speakers may slow their speech in response to underlying disruptions, without overt hesitation, a complete picture must include speech-rate variations. Although rate differences are rarely assessed in the traditional hesitation literature, studies of speech timing standardly include both types of variables (Cooper & Paccia-Cooper, 1980; Ferreira, 1993; Huitema, 1993; Meyer, 1994). Cooper and Paccia-Cooper (1980, chap. 7) provided evidence that speakers may, in fact, slow down while preparing upcoming material.

Table 1
Examples of Speech Errors

Type	Example: Intended utterance	Example: Error	Unit Involved*
Sound errors			
Contextual errors			
Exchange	York library	lork yibrary	Phoneme
	snow flurries	flow snurries	Consonant cluster
	clear blue	glear plue	Feature
Anticipation	reading list	leading list	Phoneme
	couch is comfortable	comf is . . .	Syllable or rime
Perseveration	beef noodle	beef needle	Phoneme
Anticipatory addition	eerie stamp	steerie stamp	Consonant cluster
Perseveratory addition	blue bug	blue blug	Phoneme
Shift	black boxes	back bloxes	Phoneme
Deletion†	same state	same sate	Phoneme
Noncontextual errors			
Substitution	department	jepartment	Phoneme
Addition	winning	winnding	Phoneme
Deletion	tremendously	tremenly	Syllable
Morpheme errors			
Contextual errors			
Exchange	self-destruct instruction	self-instruct de . . .	Prefix
	thinly sliced	slicely thinned	Stem
Anticipation	my car towed	my tow towed	Stem
Perseveration	explain . . . rule insertion	rule exsertion	Prefix
Shift	gets it	get its	Inflectional suffix
Addition	dollars deductible	dedollars deductible	Prefix
	some weeks	somes weeks	Inflectional suffix
Noncontextual errors			
Substitution	conclusion	concludement	Derivational suffix
Addition	to strain it	to strained it	Inflectional suffix
Deletion	he relaxes	he relax	Inflectional suffix
Word errors			
Contextual errors			
Exchange	writing a letter to my mother	writing a mother to my letter	Noun
Anticipation	sun is in the sky	sky is in the sky	Noun
Perseveration	class will be about discussing the test	discussing the class	Noun
Addition	these flowers are purple	these purple flowers are purple	Adjective
Shift	something to tell you all	something all to tell you	Quantifier
Noncontextual errors			
Substitution	pass the pepper	pass the salt	Noun
	Liszt's second Hungarian rhapsody	second Hungarian restaurant	Noun
Blend	athlete/player	athler	Noun
	Taxi/cab	tab	Noun
Addition	the only thing I can do	the only one thing	Quantifier
Deletion	I just wanted to ask that	I just wanted to that	Verb

Note—From "A Spreading Activation Theory of Retrieval in Sentence Production," by G. S. Dell, 1986, *Psychological Review*, 93, p. 285. Copyright 1986 by the American Psychological Association. Reprinted with permission. *The units involved in the examples for a given type do not exhaust the set of possible units for that type. In general, a given type can occur with many different units. †The deletion category appears both under the heading of sound misorderings and sound noncontextual errors. This is because some deletions, such as in the *same state* example, seem to involve a contextual influence, whereas other deletions do not.

Errors in spontaneous speech. Speech errors come in myriad forms. In this diversity, however, an orderly taxonomy has been identified (Table 1). There is a relatively small number of error types (including exchanges, additions, shifts, deletions, anticipations, perseverations, substitutions, blends) that may involve elements from the identifiable linguistic context (contextual errors) or from outside that context (noncontextual errors). Within each error type, virtually every linguistic element may be represented (ranging from phonetic features through phonemes to words, phrases, and clauses). The examples in Table 1 offer a small sampling of the linguistic elements that can be found in each of the error categories.

The collection and analysis of such errors is the foundation and starting point for much of the current research on production. An investigator needs little more than a good ear for language, a pen, and some paper (the repositories of error collections, my own included, are laced with table napkins). Armed with these materials, one can examine an enormous range of questions. The two most influential uses of error analyses have been in testing linguistic claims about language structure (Fromkin, 1971) and psychological claims about the architecture of the language production system (Garrett, 1975).

The latter questions are the relevant ones for this review. They have been most fruitfully addressed in terms of what stays right in an utterance when other things go wrong. For example, consider the class of errors like "You just count wheels on a light" (when "lights on a wheel" was intended; Stemberger, 1985b). Notice that the plural inflection *-s* stayed put, even when the word stem *light* moved. This so-called stranding of inflections is a very regular feature of errors: Stemberger (1985b) reported that it occurred in 89% of those errors in his corpus in which it could have occurred. The upshot is that the production system seems to deal with word stems and inflections somewhat independently, and puts inflections into position more reliably.

The implicit theory behind the use of such data is that errors are most likely to occur at the joints of an articulated system, points where pieces must be joined together. Like garden paths in language comprehension, errors in language production and, importantly, the constraints on errors in production reveal the interchanges of the routes that the processes follow.

Two examples will illustrate the logic of the analysis. First, consider the error "He called her yesterday," produced instead of "She called him yesterday." This is not an exchange of pronouns: The error is not "Him called she yesterday." The implication is that the error occurred before the cases of the pronouns were assigned, before pronoun form was determined. This is generally true of pronoun exchanges, and it implies that there may be a point at which the subjects and objects of sentences have been identified when the words to express them have not yet been retrieved. Second, consider "it certainly run outs fast" (Garrett, 1980). In this case, the speaker says "run out/s/" when "run/z/ out" was intended. This error, like the pronoun error, is not a simple movement of a sound from one position to another. The suffix *-s* is pronounced in the way that is appropriate to its new phonetic environment. Because this appears to be true for all errors of this type, the pronunciation is evidently established after the point at which suffixation occurs. These kinds of constraints on errors thus illuminate the partitions of the production mechanism and the nature of the elements that it manipulates.

Despite its manifest value, there are drawbacks to the error-observation enterprise. Although none of the drawbacks are irremediable, the need for remedies must be kept in mind by anyone contemplating the enterprise of error collection. The chief pitfalls are (1) the relative scarcity of errors, (2) the ambiguity of categorization, and (3) the potential for bias in the collection of errors.

The scarcity of errors. Scarcity comes as something of a surprise given the widespread impression that speech is full of mistakes. The source of this impression is not outright error, however, but dysfluency. Blackmer and Mitton (1991) reported that callers on a radio call-in program interrupted themselves more than once every 5 sec, but only 3% of the interruptions were to correct speech errors. The remainder were dysfluencies. Obvious syntactic errors occur less than 5 times in every 1,000 sentences (Deese, 1984; Heeschen, 1993); lexical selection errors occur less than once in every 1,000 words (Bock & Levelt, 1994); and phonological encoding errors occur less than 4 times in every 10,000 words (Deese, 1984; Garnham, Shillcock, Brown, Mill, & Cutler, 1982). To surmount the problem of scarcity without sacrificing naturalistic observation, investigators typically collect errors over many years. Some investigators have pooled observations collected by groups of students and the like, but this introduces other difficulties (see Dell & Reich, 1981).

Ambiguities in error classification. Ambiguity in classification arises because deviant patterns can almost always be explained in more than one way (Cutler, 1988). For instance, the error in "start the boat on the motor" could be a simple exchange of the words boat and motor, or a misassignment of two complete noun phrases, "the boat" and "the motor." Ambiguity at this level is a particular problem in English, which lacks the markers that specify grammatical functions in languages such as German, Russian, and other richly inflected languages. However, the classification problem pervades errors of all kinds. To overcome it, investigators call on multiple converging criteria whenever possible. Garrett's (1975, 1980, 1988) error analyses are paradigm examples of this strategy at work.

Bias. The danger of bias is that the collection can display properties of perception, personality, or theoretical proclivity that may be erroneously attributed to language production. To counter this, most practiced observers apply safeguards against the many sources of bias that can contaminate a collection that is intended for scientific use. In what follows, I will sketch some of the known biases and the measures used to minimize them.

Attentional and perceptual biases arise because errors must be heard to be counted as errors. Listeners may hear errors that did not happen or fail to hear errors that did. Since the latter problem can occur when listeners are more interested in the content of speech than in its form, as is typically true, errors that are less likely to affect meaning may be less noticeable (Tent & Clark, 1980). Even with full attention, errors at some points in the stream of speech may be more readily detected than similar errors at other points, and errors involving some sound features are more readily perceived than errors involving others (Cole, Jakimik, & Cooper, 1978; Cole & Perfetti, 1980). Since adult native speakers are demonstrably bad at hearing sounds that are outside the segmental inventory of their language (Werker & Lalonde, 1988), it is perhaps not surprising that so few "phonetically impossible noises" turn up among reported speech errors (Wells, 1951). All of these problems are exacerbated by the distracting conditions of normal error collection. Cutler (1982) summarized some of the evidence pertinent to the biases created by differences in detectability, and Ferber (1991) showed how unreliable error detection can be.

The simple act of recording an error can also introduce biases. Some errors may be easier to remember than others, even for the brief period required to write them down. Critical elements of the error context may not be accurately or reliably noted. Errors involving subtle features of sound are hard to transcribe as well as hard to hear. The kinds of notation systems that observers know and use, from common orthographies (e.g., the Roman alphabet) to technical systems (e.g., the International Phonetic Alphabet), may themselves be the source of a kind of bias, imposing discrete categories on articulatory patterns that are not physically discrete (Mowrey & MacKay, 1990).

Biases attributable to theoretical proclivities are nicely illustrated in Ellis's (1980) analysis of Freud's corpus of speech errors. The commonest kinds of speech errors (those involving single sounds) are underrepresented in the list of errors appended to *The Psychopathology of Everyday Life*, whereas errors involving whole words are overrepresented. For instance, a sound-deletion error like "tubbled" (where the /m/ from the intended word *tumbled* was deleted) would be less likely to be included in the corpus than the word-deletion error in a host's parting admonition: "Next time, don't stay so long... I mean, don't stay away so long," although errors of the former type are more frequent in speech. Since whole-word errors are easier game for psychodynamic interpretation, one suspects that Freud paid undue attention to just those errors for which he had a ready explanation.

An unavoidable source of bias in error collections stems from the distributional characteristics of the language itself. Some kinds of errors will be observed more frequently than other kinds simply because the relevant linguistic or conversational environments themselves differ in frequency. An instructive example comes from a debate about whether sound errors tend to create real words. The majority of sound errors, 55% to 60%, result in nonwords (Garrett, 1988). This nonword majority conceals a trend toward real-word errors. Consider the errors "peg" (for *beg*) and "ped" (for *bed*). In both cases, the error is a consequence of the production of /p/ in place of /b/. However, in the first case the error creates another word (*peg*) and in the second it does not (*ped*). In order to assess whether /p/ is more likely to replace /b/ when it creates a real word than when it does not, the actual incidence of word-creating errors when /p/ replaces /b/ must be compared to the number of cases in which it would be *possible* for a substitution of /p/ for /b/ to yield a real word: How often can such a sound error create another English word purely by accident? This chance probability can be relatively low, because the majority of words in a language have very few first-degree relatives (words that differ from them by only one phoneme). In Dell and Reich (1981), the chance probabilities of word outcomes ranged between 18% and 45%. A greater-than-chance tendency for sound errors to create real words is therefore fully compatible with a preponderance of nonword outcomes.

The difficulty of circumventing all of these biases is one argument for experimental approaches. Still, the ecological value of naturalistic observation makes it worthwhile to seek ways of minimizing bias. Several types of controls have been used. Most investigators set out with the goal of recording all errors, without regard to the value or interest of the error. To increase the likelihood of catching sound slips and other low-level errors, some investigators monitor speech for errors only during limited time periods (Harley, 1990a; Stemberger, 1990). To minimize both detection and recording problems, investigators sometimes cull errors from transcriptions of tape-recorded speech (Boomer & Laver, 1968; Garnham et al., 1982; Wijnen, 1992). To guard against distributional biases, the chance probabilities of error outcomes can be calculated for some kinds of errors. Two methods for doing this are currently in use. One method derives the chance probability of an error from the words in the error corpus itself (see Dell & Reich, 1981), and another is based on estimates of the frequency of occurrence of particular units in the language (e.g., Stemberger & Treiman, 1986).

There have been a few assessments of the reliability of naturalistic observation. Comparisons of errors observed "on the fly" with errors tabulated from transcribed, tape-recorded speech have revealed somewhat different distributions, especially for less salient elements of speech (Wijnen, 1992). MacKay (1980) noted informally that subjects listening to taped errors under optimal conditions often disagreed with each other (and even with themselves over successive occasions). Ferber (1991) reported performance that can only be described as abysmal: Trained listeners failed to detect the majority of slips that occurred in a recorded discussion, failed to record the slips correctly on nearly half of the occasions when they did detect them, and for the most part, failed to record the same slips.

Such are the rigors of field research. Fortunately, however, the perilousness of the enterprise does not nullify the reliability of the major results that have emerged from careful studies. The best testimony to this can be found in Stemberger (1992). Stemberger reviewed the convergences and divergences of observational and experimental findings in the speech error literature, primarily the literature on phonological and lexical errors, and found good agreement in almost all cases.

The speech errors of children present special challenges and special opportunities. Children's errors have been called upon to illuminate both the development of the language production system (Aitchison & Straf, 1982; Gerken, 1991; Stemberger, 1989; Wijnen, 1992) and the child's acquisition of linguistic knowledge (E. V. Clark, 1995; Pinker, 1995). The former issue is one of linguistic performance: How do children develop the retrieval and assembly skills required to fluently formulate utterances? The latter, however, is an issue of linguistic competence: What do children know about language, and when do they know it? The first issue is more germane to theories of adult language production, whereas the second has been a central concern of linguistic theory.

The dilemma in these enterprises is that any error, taken by itself, could be explained as a mere mistake by a speaker who knows better (a performance error) or as a knowledge gap in a speaker who does *not* know better (a competence deficit). Since children are exceptionally vulnerable to the joint effects of production problems and competence gaps, accounts of their errors must take care to distinguish haphazard mistakes from systematic developmental deviations. To achieve this, most investigators identify errors relative to norms established by a child's current usage, but this demands intensive, ongoing observation. And even when this groundwork is carefully laid, it can remain unclear whether an error reflects something about the child's production system or something about the child's knowledge.

A study by Bowerman (1978) illustrates the problem. Bowerman interpreted a set of children's verb errors (e.g., "you put the pink one to me" instead of "you gave the pink one to me") as reflecting the emergence of integrated knowledge about verb semantics; that is, as the product of a change in the child's still-incomplete knowledge of language. Bowerman's argument hinged on the fact that the children began to make these errors only after a relatively long period of correct usage. They therefore knew the meanings of the words. The errors, she suggested, arose only when meaning similarities came to be implicitly recognized, perhaps by integration into a new knowledge structure. However, Bowerman acknowledged an alternative: The integrated knowledge structures were there all along, and the errors emerged from changes in the way the structures were addressed during actual production; that is, they were the product of changes in the child's performance system. An increase in the breadth of the search space, an increase in the rate of spread of activation through a lexical network, and an increase in speech rate could all create such errors. So, the error "you put the pink one to me" is a straightforward example of a word substitution, in which an intended word (*gave*) is replaced by another word that is loosely related in meaning (*put*).

At the other developmental extreme, errors in aphasic speech can offer clues about the mechanisms of production from patterns of dissolution (Saffran, Schwartz, & Marin, 1980). The relationship between pathological and normal deviations from fluent speech has sometimes been examined from this standpoint (Butterworth & Howard, 1987; Garrett, 1992; Harley, 1990b; Lapointe & Dell, 1989; Stemberger, 1984, 1985a). In some ways, the analysis of aphasic errors is even more vexed than analyses of children's errors. The neurophysiological substrates of the system may be unstable to unknown degrees; the characteristics of the patient's premorbid language may be uncertain; the status of the language knowledge that normally supports production may be indeterminate; and the target of the erroneous utterance may not be deducible, exacerbating the ambiguity that afflicts even the classification of "normal" errors.

In the face of such problems, it is noteworthy that there are broad similarities in error distributions for normal and aphasic speakers, with some studies reporting only quantitative differences in susceptibility to error. As interesting as this may be, it reveals little more than normal speech error analysis about how production mechanisms are organized for fluent speech. More detailed studies have, however, uncovered disparities in both the kinds and quantities of error produced by aphasics and normal speakers. In one instance, Schwartz, Saffran, Bloch, and Dell (1994) found the same kinds of deviations from normal error patterns in aphasic speech and in child speech, raising the possibility of a unified account of acquisition and dissolution that also illuminates specific mechanisms within the normal production system.

Aphasic speech errors serve to inform theories of aphasia as well as theories of normal production. There is increasing interest among aphasiologists in using accounts of normal production as a backdrop for explanations of aphasic speech deficits (see, among others, Blanken, Dittmann, Haas, & Wallesch, 1987; Buckingham, 1986; Byng & Black, 1989; Heeschen, 1993; Kohn, Lorch, & Pearson, 1989; Kolk, 1987; Pate, Saffran, & Martin, 1987; Saffran, Berndt, & Schwartz, 1989; Schwartz, 1987; Zingeser & Berndt, 1988). Although methodological issues in the study of aphasic speech are beyond the scope of this article, Berndt (1991) has provided an excellent review of research on aphasic production.

Experimental Methods

In the psycholinguistic study of language production, naturalistic observation has played a substantially larger role than experimentation. The reasons can be found in the many challenges that confront efforts to impose experimental control on speaking. In principle, speech is a domain in which experimental logic is sometimes seen to contravene the normally free rein of thought and expression (Chomsky, 1986, pp. 222–223), and so concerns about ecological validity run unusually deep. In practice, hard problems arise in crafting viable experimental methods.

Some of the practical problems are traceable to the difficulty of characterizing nonlinguistic messages, which are the input to the production process (Foss & Hakes, 1978; for review see Fodor et al., 1974, chap. 7). The worries range from the philosophical (e.g., the assumption that an "input" exists has been criticized as vicious regress) to the empirical (lacking a characterization of the input makes it hard to identify, manipulate, and control theoretically interesting independent variables). The product of production, speech itself, creates additional problems. Experiments require the selection of specific measurable responses, and in production, there is often no easy way to get specific responses. As Fodor et al. put it,

> Getting the data we need (e.g., on the relative production complexity of sentences of different syntactic types) would involve somehow getting the subject spontaneously to produce sentences of an antecedently specified message or form, and there is no very satisfactory procedure for doing this. (1974, p. 397)

If the desired response is stipulated to speakers in advance, the investigator risks distorting or circumventing the underlying processes. If the response is not stipulated, the problem of what I will call *exuberant responding* arises. The problem is that something other than the target is very often produced, with the consequence that many of the utterances must be set aside because of their uncertain bearing on the questions of interest. In short, there are dilemmas about gaining control of both the input and the output.

In the face of such difficulties, why bother? Obviously, because there is no more powerful way to test explanatory hypotheses about patterns of language use. The word "pattern" is important: The contemporary study of production deals less with what people say than how they say it. It emphasizes the linguistic forms of speech rather than their content. This makes the enterprise simultaneously more tractable and less Orwellian than it might otherwise promise to be. From a scientific standpoint, the experimental method circumvents many of the uncertainties that attend the observation of speech errors (A. S. Meyer, 1992). It severely curtails the problem of bias, isolates the phenomena of interest, increases the opportunities for systematic observation, and makes the phenomena available for further scrutiny.

For these things to be feasible, however, production hypotheses must be formulated and tested in ways that sidestep the input and output dilemmas. Recent advances in production research are largely attributable to the emergence of solutions to these problems.

On the input side, the problem of characterizing the message has been temporarily bypassed with techniques that are fashioned to keep the eliciting stimulus the same across experimental manipulations. This solution is aimed at creating message constancy, encouraging or requiring speakers to express the same idea. By design, the critical manipulations do not change the message itself, but they do change the state of the linguistic knowledge that is used or the system through which the to-be-produced information passes. This class of techniques is related to standard priming paradigms in cognitive psychology, and has been fruitful in several production domains. A simple illustration comes from experiments by Bock (1986a, 1987b). In these studies, the speakers repeated a specified word (a prime) and then saw and described a picture of an event (Figure 3). The priming word (*search*) was not appropriate for describing the event, but was phonologically or semantically related to a target word that was very likely to appear in the event description (*church*). The event descriptions comprised the responses of interest, with theoretical attention centering on the effect of priming on the order in which words were produced.

On the output side, two types of solutions to the problem of exuberant responding are in general use. One, *normative elicitation*, employs materials that naturally tend to elicit a desired response (e.g., pictures of events or objects that have uniform descriptions, questions that elicit specific answers, definitions that clearly point to a specific word, or foreign words that have known translations). Nonconforming responses are set aside as beyond the scope of the hypotheses under test. The second method, *specified elicitation*, tends to be used primarily when the processes of interest do not require message formulation or grammatical encoding, and involve phonological or articulatory mechanisms only. In these cases, the responses are often stipulated (for example, a written version of the desired response might be presented to the subject prior to a production trial), and deviations are treated as errors. The assumption is that the same phonological and articulatory mechanisms are exercised regardless of whether the desired response is spontaneous or specified in advance.

There are caveats, of course, that reflect imperfections in these solutions. First, the assumption of message constancy may not always be met. For example, in Bock's priming experiments, the priming word could have affected the processing of the pictured event and the formation of the message that described it, not the lexical and syntactic processes that were of primary interest.[4] More generally, whenever comprehension or interpretation of the eliciting stimuli are required prior to a production task, it can be difficult to ensure that production processes have been isolated unless appropriate controls are instituted. This is the *comprehension contamination* problem. Second, the solutions to response exuberance can severely limit the generality of the conclusions. With normative elicitation, as the number of nonconforming responses grows large, the statistical power of the experiment diminishes. With specified elicitation, there is always a danger that the processes that support the production of a designated response diverge in essential ways from the normal mechanisms of production.

With this as background, I will survey two general classes of methods that have been used in experimental research on production. The first employs priming and interference techniques in eliciting both speech errors and normal speech. The second relies on the direct manipulation of messages and the pragmatic contexts in which the messages are produced.

Manipulating the pathways: Error elicitation. Befitting the origins of research on production, contemporary experimental work has often involved the elicitation

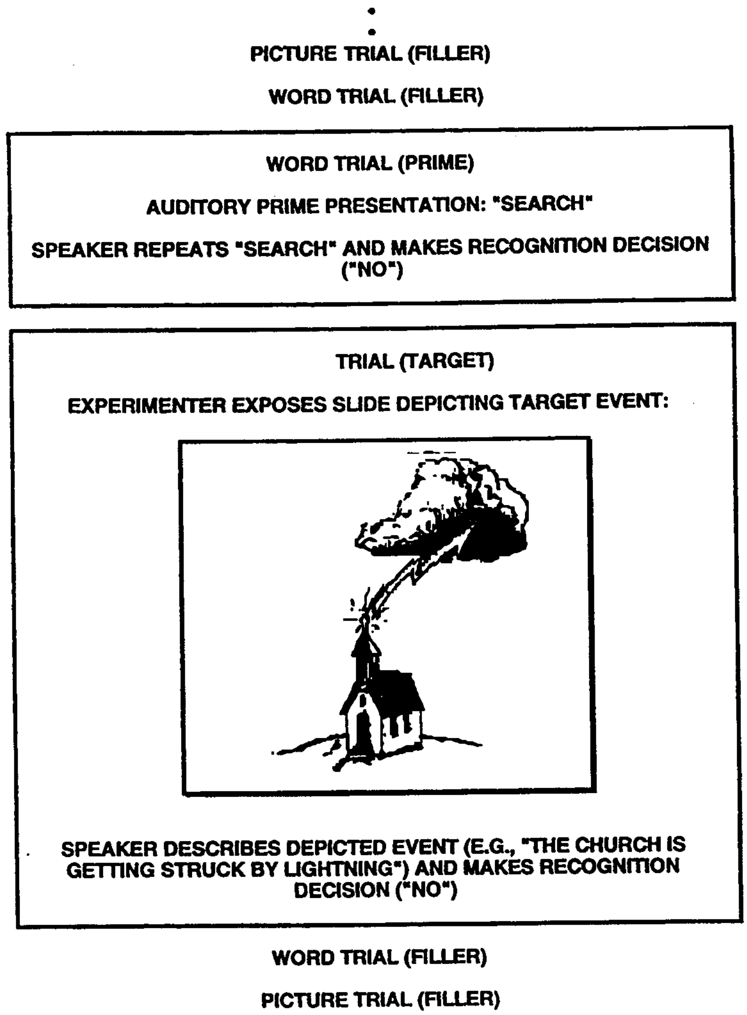

Figure 3. Sequence of events on a word-primed sentence production trial. The word prime was related to another word in each of two target sentences (e.g., *Lightning is striking the church; The church is getting struck by lightning*). The primes were phonologically related to the target sentence's subject or object (*frightening*, *search*) or semantically related to them (*thunder*, *worship*). The target sentences were produced spontaneously in describing the picture. The dependent measure was the percentage of target sentences that were produced with the primed word before the unprimed word. From "Syntactic Presistence in Language Production," by J. K. Bock, 1986, *Cognitive Psychology*, *18*, p. 361. Copyright 1986 by Academic Press. Adapted with permission.

of speech errors. Many of the experiments in this tradition use variants of a paradigm that was developed by Baars, Motley, and MacKay (1975). This is the SLIP paradigm (for spoonerisms of laboratory-induced predisposition; Motley, 1980). In it, subjects silently read successive pairs of words such as

bash door
bean deck
bell dark
darn bore

within a longer list. The subjects are cued to repeat some of the pairs aloud immediately after reading them. The pair "darn bore" is an example. Although it would normally be a simple matter to produce these words correctly, subjects make mistakes. On approximately 6% of the trials, instead of saying "darn bore," they say "barn door." Evidently, the preceding word pairs prime a specific pattern of successive word onsets (/b/–/d/) that perseverates into the cued trial.

Baars (1992) called this a competing-plans technique, and the label aptly describes a general strategy for error elicitation.[5] The scheme is to induce a set to produce an utterance with particular characteristics, and then to elicit an utterance whose features conflict with those characteristics. In the face of the conflict, the utterance tends to go awry in a predictable way. The SLIP paradigm has been used to assess the contributions to speech errors of phoneme frequency (Levitt & Healy, 1985), speech rate (Dell, 1986), lexical bias (Dell, 1986; Baars et al., 1975), phoneme repetition (Dell, 1984), and word class (Dell, 1990), and to diagnose the structure of consonant clusters (Stemberger & Treiman, 1986).

The success of competing-plans techniques can be attributed to their ability to induce conflicts similar to those that may normally give rise to errors (Baars, 1980a, 1980b; Dell, 1986). The techniques accomplish this through a time-honored form of experimental artifice. The spoken materials and the tasks themselves may deviate, sometimes widely, from the circumstances of normal speaking. Silent reading and written-word pronunciation, as they are used in the SLIP paradigm, bear little resemblance to spontaneous speech. Nonetheless, SLIP and related methods yield error patterns comparable to those that have been observed naturalistically (Stemberger, 1992). Clearly, the validity of SLIP results does not rest on simulating the everyday environments of errors, but on creating the processing circumstances that lead to error. Silent reading often evokes sounds, and pronouncing written words sets word-production processes in train. The paradigm is thus a caricature of speech, distilling and magnifying the essential conditions for error.

Competing-plans methods are especially useful for increasing the *rate* at which errors occur. Other, simpler techniques simulate real error environments in order to increase the number of errors available for observation. These techniques have been applied to the elicitation of phonological, inflectional, lexical, and grammatical errors.

To elicit phonological errors, speakers have been asked to produce conventional or novel tongue-twisters (Butterworth & Whittaker, 1980; Shattuck-Hufnagel, 1992). Shattuck-Hufnagel used novel tongue-twisters like "It's the peal of the tone from the pan on the tool" to examine the role of syllable position and stress in the encoding of sound segments. Sevald and Dell (1994) developed a complementary version of this method in which speakers were asked to repeat strings like "pick tuck puck tick" as rapidly as possible. Rather than errors per se, the dependent measure was the number of correct repetitions of the string in a specified time period.

To observe inflectional errors, subjects have been asked to produce an inflected form of a presented word (e.g., the past tense of a verb or the plural of a noun; Bybee & Slobin, 1982; Lapointe & Dell, 1989; MacKay, 1976; Stemberger & MacWhinney, 1986a, 1986b). The frequency of inflectional errors in a critical environment is compared to the frequency of the same errors in a control environment. For example, Stemberger and MacWhinney (1986a) asked speakers to produce the past-tense forms of verb stems like *guide* and *grab*. In performing this task, the speakers were significantly more likely to omit the past-tense inflection on words whose stems end in the same sound that forms the past tense (such as *guided*) than on words which do not (such as *grabbed*).

To induce lexical errors, N. Martin, Weisberg, and Saffran (1989) had subjects describe arrays of pictured objects. The names of the objects were semantically related, phonologically related, both semantically and phonologically related, or unrelated. This elicited semantic word substitutions like "house" for *church*, phonological substitutions like "skull" for *squirrel*, and mixed errors like "carrot" for *cabbage*, where the error was both semantically and phonologically similar to the target word.

Finally, there is a related method for eliciting a class of grammatical mistakes, errors in subject–verb number agreement (Bock & Cutting, 1992; Bock & Eberhard, 1993; Bock & Miller, 1991). In these experiments, speakers are given phrases such as *The key to the cabinets*. The speakers' task is to create a full sentence using the prescribed phrase as the sentence subject. In the critical conditions, the grammatical number (singular or plural) of the first and second noun in the phrases differs, creating a number conflict. Because the speakers have to supply verbs in order to create full sentences, it is possible to observe errors in number agreement on the verbs (e.g., "The key to the cabinets *were* lost"). The frequency of such verb errors is compared to the incidence of the same verb error in a minimally contrasting environment (e.g., after *The key to the cabinet*). Similar materials and techniques can be used to explore pronoun–antecedent number agreement (Bock, 1995). Unsurprisingly, number errors are more likely after number conflicts. More interesting are the possibilities for using variations in error patterns to assess the kinds of information and the kinds of processing demands that are normally in play during the creation of long-distance dependencies in speech.

All of these procedures aim to create conditions that mimic normal production in essential respects. Some evidence of their success can be found in the distributions of elicited errors. These distributions appear to be similar to the distributions of the same errors in normal speech, and similar errors are absent from control environments (e.g., Bock & Miller, 1991; Stemberger & MacWhinney, 1986b). This makes it possible to infer something about the factors that lead to mistakes in speech.

There is an important continuity between the competing-plans methods and the simple elicitation methods. Simple elicitation relies on materials that naturally contain or create a kind of competition. Tongue twisters require the repetition of closely related phonological features (as in the familiar "she sells sea shells" and the novel "chef's sooty shoe soles"). Inflectional errors occur when speakers have to inflect verb stems that end in sounds that mimic real inflections. Lexical substitutions occur between semantically and phonologically related words. Agreement errors occur when speakers have to implement agreement with a complex subject noun phrase that contains a number conflict. As noted earlier, all of these conflicts are to production what ambiguity is to comprehension: They require selections that serve to reveal the choice points within the processing system (Baars, 1992).

Most of the experimental error-elicitation methods rest on the assumption that multiple sources of information are normally in play during production, with errors arising as a consequence of interference among competing elements. There are two theoretical frameworks that offer contrasting accounts for the performance of these tasks. One framework embodies the idea that the processes of production are staged so that information flows in one direction only; in terms of Figure 1, from messages through grammatical and phonological encoding to articulation (Garrett, 1988; see also Levelt, 1989). The theory assumes that similar elements at each level of processing may compete with one another for inclusion in a speech plan. Because the information flow is unidirectional, the elements that are in play at any given level are, for the most part, those that are consistent with the processes completed at the next higher level.

A different framework postulates interactions between lower and higher levels of processing, so that events at lower levels may influence the top-down flow of information (Dell, 1986). This model likewise incorporates mechanisms of underlying competition among alternative ways of realizing speech plans. However, in the cases of word and sound errors, the competition is regulated not only by events at higher levels but also by those at lower levels of processing, simulated in terms of feedback of activity within a connectionist lexical network.

Both of these frameworks focus on explanations for speech errors, but they also offer accounts of processing in normal production within the same architectures that give rise to error. There is thus a close connection between the explanations of error and theories of normal performance. This in turn offers a bridge from experimentally elicited speech errors to accounts of error-free speech.

Manipulating the pathways: Priming normal speech. The competing-plans rationale has likewise been applied to the elicitation of normal, error-free speech, but with a different goal. Whereas pathway manipulation in speech-error elicitation is designed to create an error, in normal elicitation it is designed to change the probability or the ease of producing a specific correct utterance. In some instances the goal (or result) is to disrupt production, whereas in others it is to facilitate production. In the latter case, in place of a competing plan, it might be more appropriate to speak of a cooperating plan.

In what follows, it will become apparent that the conditions that lead to disruption and facilitation are not well understood: So-called interference techniques do not always disrupt performance, and so-called priming techniques do not always facilitate performance. This situation mirrors one that has developed in applications of similar techniques to questions of word recognition (for reviews see Glaser, 1992, and MacLeod, 1991). Since there is as yet no clear reason why one or the other outcome is obtained, I will draw a rough-and-ready distinction between interference and priming paradigms in terms of the temporal relations between stimulus presentation and response generation. In interference, or *concurrent-stimulation* paradigms, the conditions of presentation are such that the processing of the interfering stimulus (the distractor) and the generation of the production target overlap in time. In priming, or *successive-stimulation* paradigms, a priming stimulus (the prime) is presented and its processing is typically completed prior to the elicitation of the production target.

Concurrent-stimulation paradigms have been used with both specified-response elicitation and normative-response elicitation. In one specified-response paradigm, subjects are given two responses (e.g., the words *dog* and *dot*) and instructed to prepare to say one of them ("dot"). Most of the time they are cued to produce the prepared response ("dot"), but on occasional trials, they are cued to produce the alternative ("dog"). Phonological relationships between the two words tend to yield slower response times than when the words are unrelated (Dell & O'Seaghdha, 1992; see also D. E. Meyer & Gordon, 1985; Yaniv, Meyer, Gordon, Huff, & Sevald, 1990). A related technique has been developed for the investigation of higher level production processes (Cutting & Bock, in press).

More commonly, a distractor is injected into the time course of an elicited response. In a typical application, picture–word interference, the primary task is timed picture naming: Speakers simply name a pictured object as rapidly as possible. On some or all of the naming trials, an auditory word is played or a visual word is superimposed on the picture. These distractors may be phonologically related, semantically related, or unrelated to the naming target, and timed to occur simultaneously with picture presentation or shortly before or after (A. S. Meyer & Schriefers, 1991; Roelofs, 1992, 1993; Schriefers, 1992, 1993; Schriefers & Meyer, 1990; Schriefers, Meyer, & Levelt, 1990; see Glaser, 1992, for a review of the literature on picture–word interference). The time taken to name the depicted object is the measure of interest.

In the first use of this concurrent-stimulation paradigm to explore issues of normal word production, Schriefers et al. (1990) presented auditory distractor words 150 msec prior to picture presentation, simultaneous with picture presentation, or 150 msec after picture presentation. The distractors were phonologically related, semantically related, or unrelated to the to-be-named picture. The major result was that early in the time course of naming, semantic distractors slowed picture naming more than unrelated or phonological distractors, whereas later in the time course, phonological distractors facilitated naming more than unrelated or semantic distractors. To help to ensure that the task tapped word production rather than picture processing, a separate control experiment paired the semantic distractors with the pictures in the same way, but used a nonproduction task. In the control task, the subjects received a recognition test on the pictures. During the recognition test, the distractors were presented with the pictures exactly as they had been in the production experiment. The distractors on this occasion produced no effect on recognition times, from which Schriefers et al. inferred that the interference did not arise during the perception or interpretation of the pictures.

This experiment illustrates the elegant logic of the concurrent-stimulation paradigm. The goal is a surgical strike: The distractor is targeted to tap into a single ongoing process, and the nature, magnitude, and timing of its effect are used to infer something about the process and its temporal limits. A nonproduction control is used to help rule out certain forms of comprehension contamination, leaving production processes as a likely culprit in the observed effects.

The chief drawbacks of the technique are associated with uncertainties about the consequences of comprehending the distractor. For an experimental effect to occur, the subject must hear or read the distractor and recover mental representations of its sound and meaning. This gives the distractor its own processing path and time course. The distractor's comprehension path is assumed to intersect the picture's production path at some point, perhaps in the semantic or phonological representation of the picture name. The exact locus and nature of the intersection are unknown, and depend on the details of the relationship between comprehension and production. So, the distractor's time course may run opposite to the timing of the targeted production processes. A distractor's sound may be recovered earlier than its meaning, whereas the target's sound may be recovered somewhat *later* than its meaning. The implication is that a distractor designed to tap phonological processing could have its intended effect sooner than a distractor designed to tap semantic processing, yet the process it is intended to affect comes relatively late in production processing.

The usual safeguard against these problems is to use unrelated distractors to establish a baseline, and to make the comparisons of interest against them. Time courses will nonetheless vary for different words, so that unrelated distractors may run faster or slower than related distractors. Since concurrent-stimulation effects can change depending on very small variations in the relative timing of input and output processes (see A. S. Meyer & Schriefers, 1991, Experiment 3 for an instructive example), the selection of unrelated distractors requires great care. A second safeguard is an explicit model that attempts to take account of each source of variation within the task in predicting experimental results (Roelofs, 1992).

Closely linked to the concurrent-stimulation paradigm is a procedure developed by Jones (1989; Jones & Langford, 1987) for investigating a common frustration, a problem that speakers sometimes characterize as having a word "on the tip of the tongue." Following Brown and McNeill (1966), Jones induced occasional tip-of-the-tongue (TOT) states by presenting the dictionary definitions of rare words and asking subjects to produce the words that corresponded to them. On the heels of the definition, Jones presented a distractor word that was semantically related, phonologically related, or unrelated to the target, to examine whether the injection of the distractor into the retrieval process would affect the likelihood of a TOT report. Although Jones found more TOTs after phonological distractors than after semantic or unrelated distractors, subsequent studies that have used the paradigm suggest that this may have been due to differences among the TOT targets that were assigned to each condition. When targets were equated for their normal susceptibility to TOT states, neither semantically nor phonologically related distractors induced more TOTs than unrelated distractors, and phonological distractors actually tended to enhance target retrieval (A. S. Meyer & Bock, 1992; Perfect & Hanley, 1992).

In language-production research, concurrent-stimulation paradigms are conceptually very similar to successive-stimulation (priming) paradigms. From a descriptive standpoint, the major difference between the two is in the subject's response to the distractor versus the response to the prime. In concurrent stimulation, the distractor is just that: Only the target is actually produced. This procedural difference has a theoretical consequence. With concurrent-stimulation tasks, it is the intersection of the target and distractor paths (presumably in a linguistic representation) that is held responsible for any effects. In successive-task paradigms, however, because the prime and target are both produced, one after the other, representations and production processes are tapped together. If primes and targets share just one critical feature that is absent from controls, experimental effects may be attributed to the shared feature or to any production processes that depend on that feature.

A disadvantage of the full-response requirement of the successive-stimulation paradigm is that it cannot easily be used to isolate the time course of production processes, since the priming response runs to completion before the target response begins. However, this same feature sometimes confers an advantage over concurrent stimulation. When a production process is the focus of interest, and not a specific representation, priming makes it possible to observe the consequences for that process of "prior exercise."

An illustration comes from structural priming (Bock, 1986b, 1989; Bock & Loebell, 1990; Bock et al., 1992). In structural priming tasks, as illustrated in Figure 4, speakers hear and then reproduce priming sentences that instantiate a specific syntactic construction (an active sentence in the figure). They then view an event and describe it in a single sentence. The event is structured in a way that permits a description in either the target construction (the primed construction, an active sentence) or an alternative construction (e.g., a passive, "The boy was awakened by the alarm clock"), and the measure is the proportion of target constructions that are employed relative to the alternative. Other subjects describe the same event after a prime of the opposite kind (e.g., "The boat was carried by five people"). In these circumstances, the primed construction is produced more often than the alternative. This appears to be attributable to the procedures that create the structural configurations of sentences, and not to shared function words, to shared intonational features, or to shared thematic roles (Bock, 1989; Bock & Loebell, 1990).[6]

The successive-stimulation paradigm has also been used to examine the processes involved in word production. Some experiments by Wheeldon and Monsell (1994) nicely illustrate word-production priming, along with one of its apparent paradoxes. In Wheeldon and Mon-

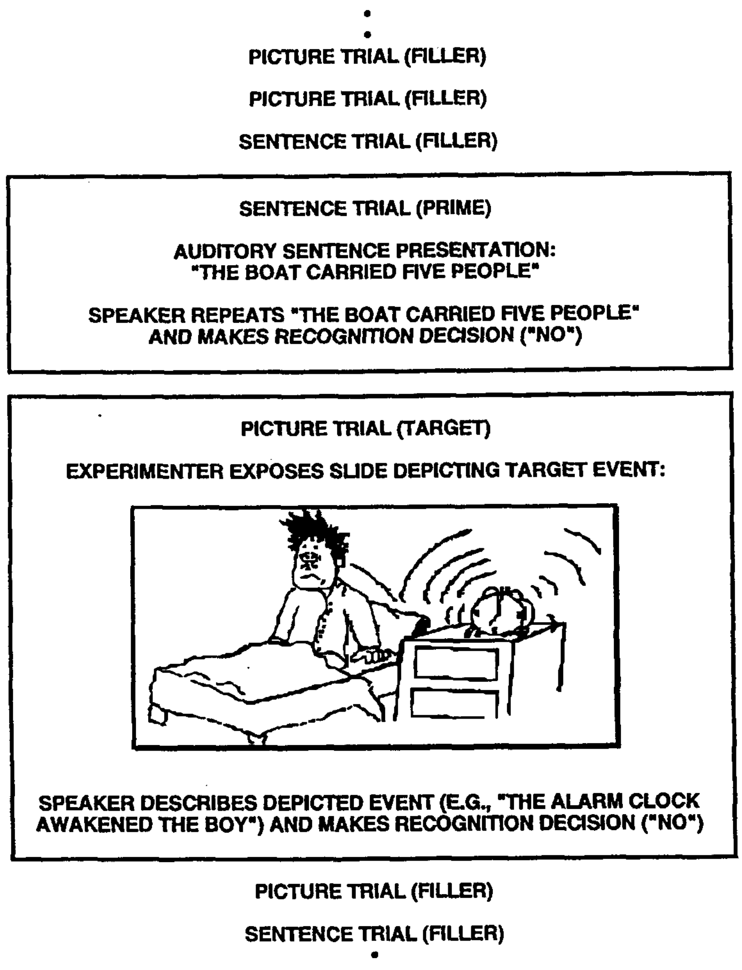

Figure 4. A structural priming trial. From "From Conceptual Roles to Structural Relations: Bridging the Syntactic Cleft," by J. K. Bock, H. Loebell, and R. Morey, 1992, *Psychological Review*, *99*, p. 157. Copyright 1992 by the American Psychological Association. Adapted with permission.

sell's experiments, the subjects received dictionary definitions with instructions to produce the defined words. These words served as the primes, and were sometimes related and sometimes unrelated to the targets. The targets were elicited two trials later, when the subjects received a picture to name. Wheeldon and Monsell measured the time required for subjects to name the targets when they were related or unrelated to the preceding prime.

The paradox is that related primes actually slowed production of the target responses, relative to unrelated primes. One of the unsolved puzzles in language processing research is why superficially similar manipulations sometimes help and sometimes hurt target performance (see Dell & O'Seaghdha, 1992, and Wheeldon & Monsell, 1994, for discussion). In another successive-stimulation paradigm, one in which single-word primes preceded a picture description (Figure 2), Bock (1986a) found that semantically related primes facilitated target production relative to unrelated primes. Using exactly the same task, Bock (1987b) showed that phonologically related primes inhibited target production.[7] In the concurrent-stimulation paradigm, facilitation has been demonstrated with phonological distractors, and inhibition with semantic distractors (Schriefers et al., 1990). Also in the concurrent-stimulation paradigm, interference has been observed when related distractors were members of the target–response set and facilitation when they were not (Roelofs, 1993). It is as yet unclear whether these seesaw patterns should be attributed to variations in expectancies and response monitoring (Roediger, Neely, & Blaxton, 1983), to something more akin to activation versus inhibition of closely related representations (Carr & Dagenbach, 1990; Dagenbach, Carr, & Barnhardt, 1990), to cooperation versus competition among linguistic alternatives (Peterson, Dell, & O'Seaghdha, 1989; Wheeldon & Monsell, 1994), or to something else entirely.

All of these hypotheses about sources of facilitation and interference bear on the mechanisms of word production. In producing words, the primary processes involve the retrieval or activation of stored information, and can be modeled along lines similar to memory retrieval. Different considerations apply to structural priming in whole-sentence production (Bock, 1990). Phrase and sentence production involve the assembly of information into ofttimes novel configurations. To explain priming-dependent changes in the probability of alternative structural configurations, one may have to go beyond the kinds of theories that apply to the revival of stored representations to theories that can account for processes like procedural learning. Such learning may require the abstraction of tacit patterns for the assembly of sequences of actions (e.g., Nissen & Bullemer, 1987), and apply to the assembly of syllables (Sevald, Dell, & Cole, 1995), novel words, and phrases, as well as sentences.

In all of the experiments described so far, the emphasis has been on characteristics of the production responses themselves. There are other studies in which characteristics of production mechanisms are inferred from responses to nonproduction probes. These probes and the responses to them may intercept an ongoing production episode (as in concurrent-stimulation paradigms) or follow a production episode (as in successive-stimulation paradigms). A task of the former type was used by Levelt, Schriefers, Vorberg, Meyer, Pechmann, and Havinga (1991). In it, subjects performed a lexical decision on an auditory word (e.g., "chair") that occasionally interrupted the subjects' primary task, naming a picture (e.g., a picture of a desk). The interruption came at times that were designed to intercept the naming response prior to its articulation, at different points in its development. Lexical decision latencies were the measure of interest. Figure 5 plots these latencies for lexical decision targets that were semantically related to the target or unrelated to it, at different points after the exposure of the to-be-named picture. The data suggest selective early interference from the meaning of the production target on the lexical-decision response.[8]

A task in which probes followed production episodes was developed by Yaniv and Meyer (1987) to explore tip-of-the-tongue (TOT) states. In their experiments, subjects attempted to name words to definitions (e.g., "large bright colored handkerchief; brightly colored square of silk material with red or yellow spots, usually worn round the neck"), sometimes experiencing TOT states in the process. They then performed lexical decisions on series of words that included the targets from the name-to-definition trials (e.g., "bandanna"). Figure 6 gives the lexical decision times for words that were unretrieved on the naming trials, and had or had not elicited strong TOTs at that time. The figure shows that the stronger the subjective certainty that an unretrieved word was a known one, figuratively at the very tip of the tongue, *the faster* the subsequent lexical decision.

In the procedures used by both Levelt et al. (1991) and Yaniv and Meyer (1987), certain characteristics of production processes may be inferred from responses to probes that tap features of an ongoing or previous production episode. The selectivity of conflicts and conver-

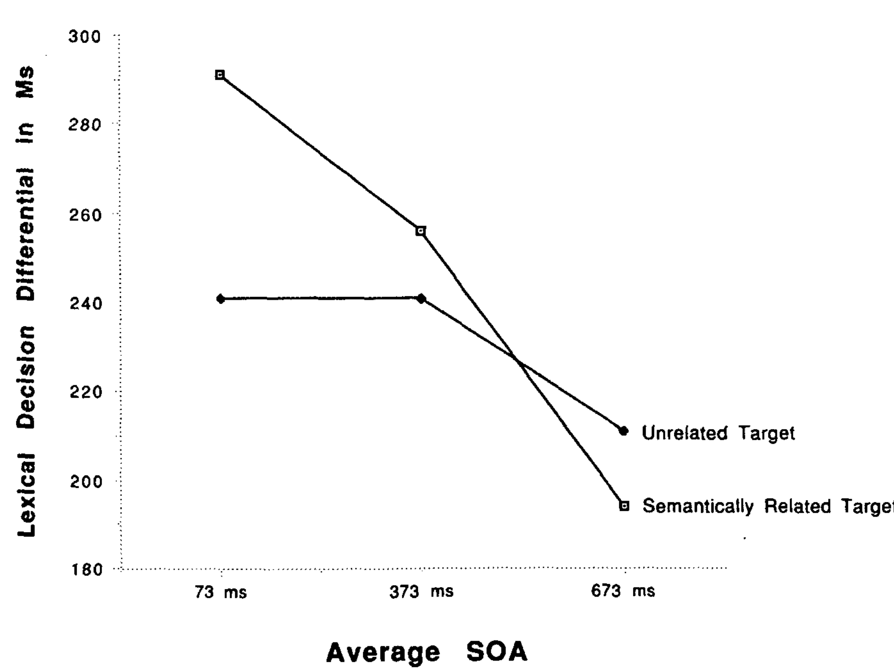

Figure 5. Results from two probe conditions in Levelt et al. (1991, Experiment 3). Differential scores reflect latencies on auditory-lexical-decision probe trials minus base lexical decision latencies (established in a control experiment). Probes were presented at one of three intervals after exposure of a to-be-named picture. Stimulus onset asynchronies (SOAs) were calibrated separately for individual subjects; the SOAs shown are averages in the short, medium, and long SOA conditions.

gences between probes and production targets offers yet another window on how information is represented and retrieved for speaking.

Manipulating the message. The most direct method for exploring production is also one of the oldest. It involves creating minimal contrasts in the message the speaker intends to convey, where "message" refers to the product of the message component represented at the top of Figure 1. At issue, then, are the contents and the representational format of a speaker's preverbal communicative intentions.

A typical experimental manipulation in this genre varies the elements of enacted scenes or events. The technique was exploited by Charles Osgood, who called it Simply Describing (Osgood, 1971; also see Bates & Devescovi, 1989; Sridhar, 1988). In his experiment, Osgood created vignettes with simple objects, placing a ball on a plate, rolling one ball so that it struck another, taking poker chips from a tabletop, and so on. The subjects were instructed merely to say what they saw, as though describing the event to a hypothetical 6-year-old child "outside the door." The sequence of events was arranged so as to create those contrasts believed to be critical for the use of certain words or grammatical devices. For example, to elicit a switch from the indefinite article (*a*) to the definite (*the*), the first vignette showed Osgood holding a black ball, followed by a second vignette in which the black ball was lying on a table. Subjects predictably said "a ball" to refer to the ball in the first vignette, but "the ball" to refer to it in the second, just as I did in the immediately preceding sentence.

Closely related are methods that elicit descriptions of displays, line drawings, or pictures in which the depicted elements themselves vary along some dimension that is potentially relevant to message formulation (e.g., Carlson-Radvansky & Irwin, 1993, Experiment 4; Flores d'Arcais, 1987; Johnson-Laird, 1968). The bare bones of this technique are evident in a simple experiment by H. H. Clark and Chase (1974, Experiment 1). They merely asked speakers to describe visual arrays such as the one on the left or the right below:

*	o
o	*

Clark and Chase discovered a strong tendency to describe these arrays in top-to-bottom fashion, as "The star is above the circle" in the array on the left, or "The circle is above the star" in the array on the right.

The simplicity of this method belies its treacherous theoretical underpinnings. Because there is little agreement on such basic issues as whether "the language of thought"—the hypothetical format in which messages are represented—is made up of features that are com-

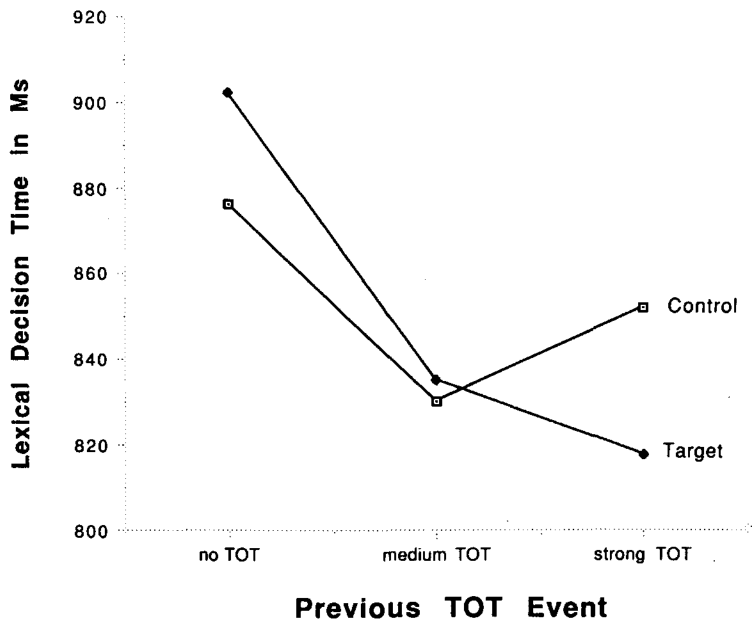

Figure 6. Lexical decision latencies to target words that had elicited a strong tip-of-the-tongue (TOT) experience, no TOT experience, or something in between during a retrieval attempt on an earlier naming-to-definition trial. Controls are the same words in the absence of the previous retrieval attempt. From "Activation and Metacognition of Inaccessible Stored Information: Potential Bases for Incubation Effects in Problem Solving," by I. Yaniv and D. E. Meyer, 1987, *Journal of Experimental Psychology: Learning, Memory, & Cognition, 13*, p. 194. Copyright 1987 by the American Psychological Association. Adapted with permission.

bined into word meanings (McNamara & Miller, 1989) or of elements that correspond rather directly to word meanings (Fodor, 1975; Fodor, Garrett, Walker, & Parkes, 1980), the seemingly straightforward idea of a "minimal contrast" can be difficult to implement in a theoretically revealing way. An alternative is to elicit descriptions of elements that contrast on known dimensions, but this runs the risk of triviality: It is seldom of interest that English speakers describe blue objects as blue and red objects as red.

In place of semantic contrasts, then, most investigators who use this method or extensions of it have manipulated so-called pragmatic contrasts, creating differences in perspective and information structure (see Levelt, 1989, chap. 3). Although the theoretical grounding of these contrasts is in many cases at least as slippery as the grounding of semantic contrasts, there is one advantage: It is possible to keep the event structure or the propositional content of a message the same while changing the speaker's perspective, the discourse context, or the communicative situation. The usual goal is to shift the kind or amount of attention given to different message components, changing which ones are in some sense more important (Flores d'Arcais, 1975; Forrest, 1993; Tannenbaum & Williams, 1968), add new information (MacWhinney & Bates, 1978; Nooteboom & Terken, 1982), or establish links to old information (Delis & Slater, 1977; MacWhinney & Bates, 1978; Nooteboom & Terken, 1982).

In most applications, the syntactic consequences of these pragmatic manipulations are examined, and in many instances, the specific focus is on what serves as the sentence subject. In one study, Forrest (1993) manipulated speakers' attention to alternative elements of simple scenes by cuing the locations in which one of the components would appear. The scenes contained just two elements (e.g., a heart and a star), and the speakers' task was to describe where one of them was with respect to the other (above, below, left, right). Forrest found that the cued element tended to serve as the subject, and that utterances were initiated more rapidly when it did. Earlier work in this tradition is reviewed in Bates and MacWhinney (1982), Bock (1982), and Levelt (1989).

Simply Describing is limited to messages that can be efficiently and unambiguously depicted or enacted, but the range of ideas that language expresses extends much further. To broaden the possibilities of message manipulation, a recall-based method is sometimes used. I will call this Simply Remembering. Simply Remembering requires subjects to remember sentences or passages for later recall (Bock, 1977; Bock & Brewer, 1974; Bock & Irwin, 1980; Bock & Warren, 1985; Kelly, Bock, & Keil, 1986; McDonald, Bock, & Kelly, 1993). Although recollection may seem to be a risky way to study production, the fact that recall is so rarely reproductive and so often reconstructive gives considerable warrant to the method. The argument rests on two familiar but still important points. The first is that people have a great deal of difficulty recalling the precise wording of what they hear, even after brief intervals (Potter & Lombardi, 1990). The second is that people have very little difficulty remembering the gist of what they hear (Jarvella, 1971). This offers a means for implanting messages that are more abstract or more elaborate than those likely to be elicited from pictures. Elicited recall of these messages resembles normal speech in so many ways that it has seemed reasonable to suggest that what is called "reconstruction" in memory for linguistic materials is in many respects equivalent to the process of language production (Bock, 1982; Levelt & Kempen, 1975; Lombardi & Potter, 1992).[9] However, the controlled comparisons of Simply Describing and Simply Remembering that would help to establish this point remain to be done.

Simply Describing and Simply Remembering have both been enhanced with manipulations of context (e.g., Prentice, 1967; Turner & Rommetveit, 1968). One of the most effective contextual manipulations is questioning. The experimenter asks a question in order to elicit a picture description (Bates & Devescovi, 1989; Carroll, 1958) or to elicit verbal recall of a previously learned piece of information (Bock, 1977). For example, someone who is viewing a scene in which a truck hits a car could be asked, "What happened to the car?" or, "What did the truck do?" Similarly, someone who has heard a story about a psychologist treating a neurotic poodle could later be asked, "What happened to the poodle?" or, "What did the psychologist do?" In each of these instances, the first question ("what happened to?") tends to be answered with a passive construction and the second ("what did it do?") with an active, although the events or memories queried are objectively the same. The goal, once more, is to keep the content of the message the same while changing the speaker's perspective on the content. In this way, message manipulation begins to converge on the logic of pathway manipulation.

This convergence can create a methodological pitfall whose theoretical consequences are worked out in detail in Bock (1982). Because changes in perspective are often prompted by features of the *linguistic* environment—sounds, words, and sentence structures—contextual manipulations may simultaneously affect both nonlinguistic messages and linguistic representations within the production system. The consequence is a confounding of message manipulation and the sort of pathway manipulation that creates lexical priming or interference effects. For example, a simple question like, "What did Billy eat?" has two undisputed effects: It topicalizes a specific referent (the person who is named Billy) and makes accessible the representation of his name (the word "Billy" and its component sounds). Both of these factors can affect the production of an answer. People position topicalized entities differently than they position nontopics (Bock, 1977) and they position primed words differently than they position unprimed words (Bock, 1986a, 1987b; Bock & Irwin, 1980; Perfetti & Goldman, 1975). It is difficult to disentangle these effects, yet the way in which they interact has important implications for explaining how language production is organized.

In order to assess the contributions of brute accessibility to production, investigators have elicited utterances that contain words whose accessibility varies naturally. This has been done in Simply Describing (Levelt & Maassen, 1981) and in Simply Remembering (Bock & Warren, 1985; Kelly et al., 1986; McDonald et al., 1993). The general conclusion from these and other studies is that the effects of lexical accessibility on formulation are small but not negligible (see Levelt, 1989, pp. 276-282), making it necessary to control for variations in accessibility in experiments whose goal is the elucidation of message effects on formulation processes.

Summary. In experimental studies of language production, the focus is on the organization of the language-manipulating components of production processes. Most methods start from the assumption of an initial nonlinguistic representation that is elaborated lexically, syntactically, and phonologically in the course of speaking. A common methodological aim is to keep the propositional content of the message constant across variations in processing. This makes it possible to create and trace perturbations in the elaborative mechanisms themselves, rather than examining the effects of variations in message features. The advantage to this approach is that it has opened a way to explore basic psycholinguistic performance questions about the flow and interplay of different kinds of linguistic information.

The disadvantages deserve prominent notice, however. The major one is that this approach sidesteps some exceptionally thorny issues about the relationships between conceptual representations and production processes. A consequence is that production theories have so far failed to address fundamental concerns about the relationships between thought and language. Some of these concerns have to do with the mapping between conceptual and grammatical categories (Bock et al., 1992; Slobin, 1996), and others with the effects of grammatical categories on conceptual ones (Lucy, 1992a, 1992b). Even so, when viewed as an item from the agenda for future research, the difficulties inherent in tackling these concerns promise to be lessened by current advances in our understanding of the grammatical and phonological processes of language production, as well as by advances in our ability to assess these processes of production.

CONCLUSIONS

The contemporary study of language production draws on both observational and experimental methods that jointly define areas of theoretical interest. This helps to give the investigation of production ecological grounding that is increasingly coupled with empirical rigor.

There is a methodogical preoccupation in the production literature with discriminating the mechanisms of production from the mechanisms of comprehension. Despite the intricacy of the relationships between production and comprehension, this separation of mechanisms promises to be an important contribution of production research. From the inception of cognitive studies of perception, language comprehension, language acquisition, memory, and farther-flung areas of research, verbal responses have provided much of the data of interest. Despite this, psychology has largely overlooked the contributions of verbal output systems. The maturation of empirical research on language production and the detailing of theories about how production works should help to correct this imbalance.

REFERENCES

Aitchison, J., & Straf, M. (1982). Lexical storage and retrieval: A developing skill? In A. Cutler (Ed.), *Slips of the tongue and language production* (pp. 197-241). Berlin: Mouton.

Baars, B. J. (1980a). The competing plans hypothesis: An heuristic viewpoint on the causes of errors in speech. In H. W. Dechert & M. Raupach (Eds.), *Temporal variables in speech: Studies in honor of Frieda Goldman-Eisler* (pp. 39-49). The Hague: Mouton.

Baars, B. J. (1980b). On eliciting predictable speech errors in the laboratory. In V. A. Fromkin (Ed.), *Errors in linguistic performance: Slips of the tongue, ear, pen, and hand* (pp. 307-318). New York: Academic Press.

Baars, B. J. (1992). A dozen competing-plans techniques for inducing predictable slips in speech and action. In B. J. Baars (Ed.), *Experimental slips and human error: Exploring the architecture of volition* (pp. 129-150). New York: Plenum.

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

Baddeley, A. D., & Hitch, G. J. (1974). Working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 8, pp. 47-90). New York: Academic Press.

Baddeley, A. D., Thomson, N., & Buchanan, M. (1975). Word length and the structure of short-term memory. *Journal of Verbal Learning & Verbal Behavior*, **14**, 575-589.

Balota, D. A., & Chumbley, J. I. (1985). The locus of word-frequency effects in the pronunciation task: Lexical access and/or production? *Journal of Memory & Language*, **24**, 89-106.

Balota, D. A., & Chumbley, J. I. (1990). Where are the effects of frequency in visual word recognition tasks? Right where we said they were! Comment on Monsell, Doyle, and Haggard (1989). *Journal of Experimental Psychology: General*, **119**, 231-237.

Bartlett, J. (1937). *Familiar quotations* (11th ed.). Boston: Little, Brown.

Bates, E., & Devescovi, A. (1989). Crosslinguistic studies of sentence production. In B. MacWhinney & E. Bates (Eds.), *The crosslinguistic study of sentence processing* (pp. 225-253). Cambridge: Cambridge University Press.

Bates, E., & MacWhinney, B. (1982). Functionalist approaches to grammar. In E. Wanner & L. R. Gleitman (Eds.), *Language acquisition: The state of the art* (pp. 173-218). Cambridge: Cambridge University Press.

Beattie, G. (1983). *Talk: An analysis of speech and non-verbal behaviour in conversation*. Milton Keynes, U.K.: Open University Press.

Berman, R. A., & Slobin, D. I. (1994). *Relating events in narrative: A crosslinguistic developmental study*. Hillsdale, NJ: Erlbaum.

Berndt, R. S. (1991). Sentence processing in aphasia. In M. T. Sarno (Ed.), *Acquired aphasia* (pp. 223-270). San Diego: Academic Press.

Blackmer, E. R., & Mitton, J. L. (1991). Theories of monitoring and the timing of repairs in spontaneous speech. *Cognition*, **39**, 173-194.

Blanken, G., Dittmann, J., Haas, J.-C., & Wallesch, C.-W. (1987). Spontaneous speech in senile dementia and aphasia: Implications for a neurolinguistic model of language production. *Cognition*, **27**, 247-274.

Bock, J. K. (1977). The effect of a pragmatic presupposition on syntactic structure in question answering. *Journal of Verbal Learning & Verbal Behavior*, **16**, 723-734.

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

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

Bock, J. K. (1986b). Syntactic persistence in language production. *Cognitive Psychology*, **18**, 355-387.

Bock, J. K. (1987a). Coordinating words and syntax in speech plans. In A. Ellis (Ed.), *Progress in the psychology of language* (pp. 337-390). London: Erlbaum.

Bock, J. K. (1987b). An effect of the accessibility of word forms on sentence structures. *Journal of Memory & Language*, **26**, 119-137.

Bock, J. K. (1989). Closed-class immanence in sentence production. *Cognition*, **31**, 163-186.

Bock, J. K. (1990). Structure in language: Creating form in talk. *American Psychologist*, **45**, 1221-1236.

Bock, J. K. (1995). Sentence production: From mind to mouth. In J. L. Miller & P. D. Eimas (Eds.), *Handbook of perception and cognition: Speech, language, and communication* (Vol. 11, pp. 181-216). San Diego: Academic Press.

Bock, J. K., & Brewer, W. F. (1974). Reconstructive recall in sentences with alternative surface structures. *Journal of Experimental Psychology*, **103**, 837-843.

Bock, J. K., & Cutting, J. C. (1992). Regulating mental energy: Performance units in language production. *Journal of Memory & Language*, **31**, 99-127.

Bock, J. K., & Eberhard, K. M. (1993). Meaning, sound, and syntax in English number agreement. *Language & Cognitive Processes*, **8**, 57-99.

Bock, J. K., & Irwin, D. E. (1980). Syntactic effects of information availability in sentence production. *Journal of Verbal Learning & Verbal Behavior*, **19**, 467-484.

Bock, J. K., & Levelt, W. J. M. (1994). Language production: Grammatical encoding. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 945-984). San Diego: Academic Press.

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

Bock, J. K., Loebell, H., & Morey, R. (1992). From conceptual roles to structural relations: Bridging the syntactic cleft. *Psychological Review*, **99**, 150-171.

Bock, J. K., & Miller, C. A. (1991). Broken agreement. *Cognitive Psychology*, **23**, 45-93.

Bock, J. K., & Warren, R. K. (1985). Conceptual accessibility and syntactic structure in sentence formulation. *Cognition*, **21**, 47-67.

Boomer, D. S. (1965). Hesitation and grammatical encoding. *Language & Speech*, **8**, 148-158.

Boomer, D. S., & Laver, J. D. M. (1968). Slips of the tongue. *British Journal of Disorders of Communication*, **3**, 2-12.

Bowerman, M. (1978). Systematizing semantic knowledge: Changes over time in the child's organization of word meaning. *Child Development*, **49**, 977-987.

Brennan, S. E., & Williams, M. (1995). The feeling of another's knowing: Prosody and filled pauses as cues to listeners about the metacognitive states of speakers. *Journal of Memory & Language*, **34**, 383-398.

Brown, A. S., & Murphy, D. R. (1989). Cryptomnesia: Delineating inadvertent plagiarism. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **15**, 432-442.

Brown, R. (1976). Reference—In memorial tribute to Eric Lenneberg. *Cognition*, **4**, 125-153.

Brown, R., & McNeill, D. (1966). The "tip of the tongue" phenomenon. *Journal of Verbal Learning & Verbal Behavior*, **5**, 325-337.

Buckingham, H. W. (1986). The scan-copier mechanism and the positional level of language production: Evidence from phonemic paraphasia. *Cognitive Science*, **10**, 195-217.

Butterworth, B. (1980). Evidence from pauses in speech. In B. Butterworth (Ed.), *Language production* (Vol. 1, pp. 155-176). London: Academic Press.

Butterworth, B., & Howard, D. (1987). Paragrammatisms. *Cognition*, **26**, 1-37.

Butterworth, B., & Whittaker, S. (1980). Peggy Babcock's relatives. In G. E. Stelmach & J. Requin (Eds.), *Tutorials in motor behavior* (pp. 647-656). Amsterdam: North-Holland.

Bybee, J. L., & Slobin, D. I. (1982). Rules and schemas in the development and use of the English past tense. *Language*, **58**, 265-289.

Byng, S., & Black, M. (1989). Some aspects of sentence production in aphasia. *Aphasiology*, **3**, 241-263.

Carlson-Radvansky, L. A., & Irwin, D. E. (1993). Frames of reference in vision and language: Where is above? *Cognition*, **46**, 223-244.

Carr, T. H., & Dagenbach, D. (1990). Semantic priming and repetition priming from masked words: Evidence for a center-surround attentional mechanism in perceptual recognition. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **16**, 341-350.

Carroll, D. W. (1994). *Psychology of language* (2nd ed.). Pacific Grove, CA: Brooks/Cole.

Carroll, J. B. (1958). Process and content in psycholinguistics. In *Current trends in the description and analysis of behavior* (pp. 175-200). Pittsburgh, PA: University of Pittsburgh Press.

Cheung, H., & Kemper, S. (1993). Recall and articulation of English and Chinese words by Chinese–English bilinguals. *Memory & Cognition*, **21**, 666-670.

Chomsky, N. (1986). *Knowledge of language: Its nature, origin, and use*. New York: Praeger.

Clark, E. V. (1995). Language acquisition: The lexicon and syntax. In J. L. Miller & P. D. Eimas (Eds.), *Handbook of perception and cognition: Speech, language, and communication* (Vol. 11, pp. 303-337). San Diego: Academic Press.

Clark, E. V., & Clark, H. H. (1978). Universals, relativity, and language processing. In J. H. Greenberg (Ed.) *Universals of human language: Method and theory* (pp. 225-277). Stanford, CA: Stanford University Press.

Clark, H. H., & Bly, B. (1995). Pragmatics and discourse. In J. L. Miller & P. D. Eimas (Eds.), *Handbook of perception and cognition: Speech, language, and communication* (Vol. 11, pp. 371-410). San Diego: Academic Press.

Clark, H. H., & Chase, W. G. (1974). Perceptual coding strategies in the formation and verification of descriptions. *Memory & Cognition*, **2**, 101-111.

Clark, H. H., & Clark, E. V. (1977). Psychology and language: An introduction to psycholinguistics. New York: Harcourt Brace Jovanovich.

Cole, R. A., Jakimik, J., & Cooper, W. E. (1978). Perceptibility of phonetic features in fluent speech. *Journal of the Acoustical Society of America*, **64**, 44-56.

Cole, R. A., & Perfetti, C. A. (1980). Listening for mispronunciations in a children's story: The use of context by children and adults. *Journal of Verbal Learning & Verbal Behavior*, **19**, 297-315.

Coltheart, M., Curtis, B., Atkins, P., & Haller, M. (1993). Models of reading aloud: Dual-route and parallel-distributed-processing approaches. *Psychological Review*, **100**, 589-608.

Cooper, W. E., & Paccia-Cooper, J. (1980). *Syntax and speech*. Cambridge, MA: Harvard University Press.

Cutler, A. (1982). The reliability of speech error data. In A. Cutler (Ed.), *Slips of the tongue and language production* (pp. 7-28). Berlin: Mouton.

Cutler, A. (1988). The perfect speech error. In L. M. Hyman & C. N. Li (Eds.), *Language, speech and mind: Studies in honour of Victoria A. Fromkin* (pp. 209-223). London: Routledge.

Cutting, J. C., & Bock, J. K. (in press). That's the way the cookie bounces: A study of experimentally elicited idiom blends. *Memory & Cognition*.

Dagenbach, D., Carr, T. H., & Barnhardt, T. M. (1990). Inhibitory semantic priming of lexical decisions due to failure to retrieve weakly activated codes. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **16**, 328-340.

Dechert, H. W., & Raupach, M. (Eds.) (1980). *Temporal variables in speech*. The Hague: Mouton.

Deese, J. (1980). Pauses, prosody, and the demands of production in language. In H. W. Dechert & M. Raupach (Ed.), *Temporal variables in speech* (pp. 69-84). The Hague: Mouton.

Deese, J. (1984). *Thought into speech: The psychology of a language*. Englewood Cliffs, NJ: Prentice-Hall.

Delis, D., & Slater, A. S. (1977). Toward a functional theory of reduction transformations. *Cognition*, **5**, 119-132.

Dell, G. S. (1984). Representation of serial order in speech: Evidence from the repeated phoneme effect in speech errors. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **10**, 222-233.

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

DELL, G. S. (1990). Effects of frequency and vocabulary type on phonological speech errors. *Language & Cognitive Processes*, **5**, 313-349.

DELL, G. S., & O'SEAGHDHA, P. G. (1992). Stages of lexical access in language production. *Cognition*, **42**, 287-314.

DELL, G. S., & REICH, P. A. (1981). Stages in sentence production: An analysis of speech error data. *Journal of Verbal Learning & Verbal Behavior*, **20**, 611-629.

DROMMEL, R. H. (1980). Towards a subcategorization of speech pauses. In H. W. Dechert & M. Raupach (Eds.), *Temporal variables in speech: Studies in honour of Frieda Goldman-Eisler* (pp. 227-238). The Hague: Mouton.

ELLIS, A. (1980). On the Freudian theory of speech errors. In V. A. Fromkin (Ed.), *Errors in linguistic performance: Slips of the tongue, ear, pen, and hand* (pp. 123-131). New York: Academic Press.

ELLIS, N. C., & HENNELLY, R. A. (1980). A bilingual word-length effect: Implications for intelligence testing and the relative ease of mental calculation in Welsh and English. *British Journal of Psychology*, **71**, 43-51.

FERBER, R. (1991). Slip of the tongue or slip of the ear? On the perception and transcription of naturalistic slips of the tongue. *Journal of Psycholinguistic Research*, **20**, 105-122.

FERREIRA, F. (1993). The creation of prosody during sentence production. *Psychological Review*, **100**, 233-253.

FLORES D'ARCAIS, G. B. (1975). Some perceptual determinants of sentence construction. In G. B. Flores d'Arcais (Ed.), *Studies in perception: Festschrift for Fabio Metelli* (pp. 344-373). Milan, Italy: Martello-Giunti.

FLORES D'ARCAIS, G. B. (1987). Perceptual factors and word order in event descriptions. In G. Kempen (Ed.), *Natural language generation* (pp. 441-451). Dordrecht: Martinus Nijhoff.

FODOR, J. A. (1975). *The language of thought*. New York: Thomas Y. Crowell.

FODOR, J. A., BEVER, T. G., & GARRETT, M. F. (1974). *The psychology of language*. New York: McGraw-Hill.

FODOR, J. A., GARRETT, M. F., WALKER, E. C. T., & PARKES, C. H. (1980). Against definitions. *Cognition*, **8**, 263-367.

FORD, M. (1982). Sentence planning units: Implications for the speaker's representation of meaningful relations underlying sentences. In J. Bresnan (Ed.), *The mental representation of grammatical relations* (pp. 797-827). Cambridge, MA: MIT Press.

FORD, M., & HOLMES, V. M. (1978). Planning units and syntax in sentence production. *Cognition*, **6**, 35-53.

FORREST, L. B. (1993). *Syntactic subject and focus of attention*. Unpublished master's thesis, University of Oregon, Portland.

FORSTER, K. I. (1990). Lexical processing. In D. N. Osherson & H. Lasnik (Eds.), *An invitation to cognitive science: Language* (Vol. 1, pp. 95-131). Cambridge, MA: MIT Press.

FORSTER, K. I., & CHAMBERS, S. M. (1973). Lexical access and naming time. *Journal of Verbal Learning & Verbal Behavior*, **12**, 627-635.

FORSTER, K. I., & DAVIS, C. (1991). The density constraint on form-priming in the naming task: Interference effects from a masked prime. *Journal of Memory & Language*, **30**, 1-25.

FOSS, D. J., & HAKES, D. T. (1978). *Psycholinguistics: An introduction to the psychology of language*. Englewood Cliffs, NJ: Prentice-Hall.

FRANCIS, W. N., & KUČERA, H. (1982). *Frequency analysis of English usage: Lexicon and grammar*. Boston: Houghton Mifflin.

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

GARNHAM, A. (1985). *Psycholinguistics: Central topics*. London: Methuen.

GARNHAM, A., SHILLCOCK, R. C., BROWN, G. D. A., MILL, A. I. D., & CUTLER, A. (1982). Slips of the tongue in the London-Lund corpus of spontaneous conversation. In A. Cutler (Ed.), *Slips of the tongue and language production* (pp. 251-263). Berlin: Mouton.

GARRETT, M. F. (1975). The analysis of sentence production. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 9, pp. 133-177). New York: Academic Press.

GARRETT, M. F. (1980). Levels of processing in sentence production. In B. Butterworth (Ed.), *Language production* (pp. 177-220). London: Academic Press.

GARRETT, M. F. (1982). Production of speech: Observations from normal and pathological language use. In A. Ellis (Ed.), *Normality and pathology in cognitive functions* (pp. 19-76). London: Academic Press.

GARRETT, M. F. (1988). Processes in language production. In F. J. Newmeyer (Ed.), *Linguistics: The Cambridge survey, III. Language: Psychological and biological aspects* (pp. 69-96). Cambridge: Cambridge University Press.

GARRETT, M. [F.] (1992). Disorders of lexical selection. *Cognition*, **42**, 143-180.

GERKEN, L. (1991). The metrical basis for children's subjectless sentences. *Journal of Memory & Language*, **30**, 431-451.

GLASER, W. R. (1992). Picture naming. *Cognition*, **42**, 61-105.

GOLD, G. (ED.) (1974). *The White House transcripts*. New York: Bantam.

GOLDINGER, S. D., LUCE, P. A., PISONI, D. B., & MARCARIO, J. K. (1992). Form-based priming in spoken word recognition: The roles of competition and bias. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **18**, 1211-1238.

GOLDMAN-EISLER, F. (1968). *Psycholinguistics: Experiments in spontaneous speech*. London: Academic Press.

GRIFFIN, Z. M. (1995). *The interaction of semantic contextual constraint and word frequency in word production*. Unpublished master's thesis, University of Illinois, Urbana-Champaign.

GROSJEAN, F., GROSJEAN, L., & LANE, H. (1979). The patterns of silence: Performance structures in sentence production. *Cognitive Psychology*, **11**, 58-81.

HARLEY, T. A. (1990a). Environmental contamination of normal speech. *Applied Psycholinguistics*, **11**, 45-72.

HARLEY, T. A. (1990b). Paragrammatisms: Syntactic disturbance or breakdown of control? *Cognition*, **34**, 85-91.

HEESCHEN, C. (1993). Morphosyntactic characteristics of spoken language. In G. Blanken, J. Dittman, H. Grim, J. Marshall, & C. Wallesch (Eds.), *Linguistic disorders and pathologies*. Berlin: Walter de Gruyter.

HIGGINS, E. T., BARGH, J. A., & LOMBARDI, W. (1985). Nature of priming effects on categorization. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **11**, 59-69.

HOLMES, V. M. (1988). Hesitations and sentence planning. *Language & Cognitive Processes*, **3**, 323-361.

HOOSAIN, R. (1982). Correlation between pronunciation speed and digit span size. *Perceptual & Motor Skills*, **55**, 1128.

HOTOPF, W. H. N. (1980). Semantic similarity as a factor in whole-word slips of the tongue. In V. A. Fromkin (Ed.), *Errors in linguistic performance: Slips of the tongue, ear, pen, and hand* (pp. 97-109). New York: Academic Press.

HOWES, D. (1966). A word count of spoken English. *Journal of Verbal Learning & Verbal Behavior*, **5**, 572-604.

HUITEMA, J. S. (1993). *Planning referential expressions in speech production*. Unpublished doctoral dissertation, University of Massachusetts, Amherst.

HULME, C., THOMSON, N., MUIR, C., & LAWRENCE, A. (1984). Speech rate and the development of short-term memory span. *Journal of Experimental Child Psychology*, **38**, 241-253.

JAMES, C. T., THOMPSON, J. G., & BALDWIN, J. M. (1973). The reconstructive process in sentence memory. *Journal of Verbal Learning & Verbal Behavior*, **12**, 51-63.

JARVELLA, R. J. (1971). Syntactic processing of connected speech. *Journal of Verbal Learning & Verbal Behavior*, **10**, 409-416.

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, & Cognition*, **20**, 824-843.

JOHNSON, M. K., & RAYE, C. L. (1981). Reality monitoring. *Psychological Review*, **88**, 67-85.

JOHNSON-LAIRD, P. N. (1968). The choice of the passive voice in a communicative task. *British Journal of Psychology*, **59**, 7-15.

JONES, G. V. (1989). Back to Woodworth: Role of interlopers in the tip-of-the-tongue phenomenon. *Memory & Cognition*, **17**, 69-76.

JONES, G. V., & LANGFORD, S. (1987). Phonological blocking in the tip of the tongue state. *Cognition*, **26**, 115-122.

KEELE, S. W. (1972). Attention demands of memory retrieval. *Journal of Experimental Psychology*, **93**, 245-248.

KELLY, M. H. (1986). *On the selection of linguistic options*. Unpublished doctoral dissertation, Cornell University.

Kelly, M. H. (1988). Rhythmic alternation and lexical stress differences in English. *Cognition*, **30**, 107-137.

Kelly, M. H., & Bock, J. K. (1988). Stress in time. *Journal of Experimental Psychology: Human Perception & Performance*, **14**, 389-403.

Kelly, M. H., Bock, J. K., & Keil, F. C. (1986). Prototypicality in a linguistic context: Effects on sentence structure. *Journal of Memory & Language*, **25**, 59-74.

Kihlstrom, J. F., Schacter, D. L., Cork, R. C., Hurt, C. A., & Behr, S. E. (1990). Implicit and explicit memory following surgical anesthesia. *Psychological Science*, **1**, 303-306.

Kohn, S. E., Lorch, M. P., & Pearson, D. M. (1989). Verb finding in aphasia. *Cortex*, **25**, 57-69.

Kolk, H. (1987). A theory of grammatical impairment in aphasia. In G. Kempen (Ed.), *Natural language generation* (pp. 377-391). Dordrecht: Martinus Nijhoff.

Kubovy, M. (1977). Response availability and the apparent spontaneity of numerical choices. *Journal of Experimental Psychology: Human Perception & Performance*, **3**, 359-364.

Kynette, D., Kemper, S., Norman, S., & Cheung, H. (1990). Adults' word recall and word repetition. *Experimental Aging Research*, **16**, 117-121.

Lachman, R., Shaffer, J. P., & Hennrikus, D. (1974). Language and cognition: Effects of stimulus codability, name-word frequency, and age of acquisition on lexical reaction time. *Journal of Verbal Learning & Verbal Behavior*, **13**, 613-625.

Lapointe, S. G., & Dell, G. S. (1989). A synthesis of some recent work in sentence production. In G. N. Carlson & M. K. Tanenhaus (Eds.), *Linguistic structure in language processing* (pp. 107-156). Dordrecht: Kluwer.

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., & Kempen, G. (1975). Semantic and syntactic aspects of remembering sentences: A review of some recent continental research. In A. Kennedy & A. Wilkes (Eds.), *Studies in long term memory* (pp. 201-216). New York: Wiley.

Levelt, W. J. M., & Maassen, B. (1981). Lexical search and order of mention in sentence production. In W. Klein & W. Levelt (Eds.), *Crossing the boundaries in linguistics* (pp. 221-252). Dordrecht: Reidel.

Levelt, W. J. M., Schriefers, H., Vorberg, D., Meyer, A. S., Pechmann, T., & Havinga, J. (1991). The time course of lexical access in speech production: A study of picture naming. *Psychological Review*, **98**, 122-142.

Levitt, A. G., & Healy, A. F. (1985). The roles of phoneme frequency, similarity, and availability in the experimental elicitation of speech errors. *Journal of Memory & Language*, **24**, 717-733.

Lombardi, L., & Potter, M. C. (1992). The regeneration of syntax in short term memory. *Journal of Memory & Language*, **31**, 713-733.

Lounsbury, F. G. (1965). Transitional probability, linguistic structure, and systems of habit-family hierarchies. In C. E. Osgood & T. A. Sebeok (Eds.), *Psycholinguistics: A survey of theory and research problems* (pp. 93-101). Bloomington: Indiana University Press.

Lucy, J. A. (1992a). *Grammatical categories and cognition: A case study of the linguistic relativity hypothesis.* Cambridge: Cambridge University Press.

Lucy, J. A. (1992b). *Language diversity and thought: A reformulation of the linguistic relativity hypothesis.* Cambridge: Cambridge University Press.

MacKay, D. G. (1972). The structure of words and syllables: Evidence from errors in speech. *Cognitive Psychology*, **3**, 210-227.

MacKay, D. G. (1976). On the retrieval and lexical structure of verbs. *Journal of Verbal Learning & Verbal Behavior*, **15**, 169-182.

MacKay, D. G. (1980). Speech errors: Retrospect and prospect. In V. A. Fromkin (Ed.), *Errors in linguistic performance: Slips of the tongue, ear, pen, and hand* (pp. 319-332). New York: Academic Press.

Maclay, H., & Osgood, C. E. (1959). Hesitation phenomena in spontaneous English speech. *Word*, **15**, 19-44.

MacLeod, C. M. (1991). Half a century of research on the Stroop effect: An integrative review. *Psychological Bulletin*, **109**, 163-203.

MacWhinney, B., & Bates, E. (1978). Sentential devices for conveying givenness and newness: A cross-cultural developmental study. *Journal of Verbal Learning & Verbal Behavior*, **17**, 539-558.

MacWhinney, B., & Snow, C. (1990). The Child Language Data Exchange System: An update. *Journal of Child Language*, **17**, 457-472.

Marslen-Wilson, W. (1973). Linguistic structure and speech shadowing at very short latencies. *Nature*, **244**, 522-523.

Martin, J. G. (1970). On judging pauses in spontaneous speech. *Journal of Verbal Learning & Verbal Behavior*, **9**, 75-78.

Martin, J. G., & Strange, W. (1968). The perception of hesitation in spontaneous speech. *Perception & Psychophysics*, **3**, 427-438.

Martin, N., Weisberg, R. W., & Saffran, E. M. (1989). Variables influencing the occurrence of naming errors: Implications for models of lexical retrieval. *Journal of Memory & Language*, **28**, 462-485.

McCann, R. S., Besner, D., & Davelaar, E. (1988). Word recognition and identification: Do word-frequency effects reflect lexical access? *Journal of Experimental Psychology: Human Perception & Performance*, **14**, 693-706.

McClain, L. (1983). Effects of response type and set size on Stroop color–word performance. *Perceptual & Motor Skills*, **56**, 735-743.

McDonald, J. L., Bock, J. K., & Kelly, M. H. (1993). Word and world order: Semantic, phonological, and metrical determinants of serial position. *Cognitive Psychology*, **25**, 188-230.

McNamara, T. P., & Miller, D. L. (1989). Attributes of theories of meaning. *Psychological Bulletin*, **106**, 355-376.

McNeill, D. (1992). *Hand and mind: What gestures reveal about thought.* Chicago: University of Chicago Press.

Mehler, J. (1963). Some effects of grammatical transformations on the recall of English sentences. *Journal of Verbal Learning & Verbal Behavior*, **2**, 346-351.

Meyer, A. S. (1992). Investigation of phonological encoding through speech error analyses: Achievements, limitations, and alternatives. *Cognition*, **42**, 181-211.

Meyer, A. S. (1994). Timing in sentence production. *Journal of Memory & Language*, **33**, 471-492.

Meyer, A. S., & Bock, J. K. (1992). The tip-of-the-tongue phenomenon: Blocking or partial activation? *Memory & Cognition*, **20**, 715-726.

Meyer, A. S., & Schriefers, H. (1991). Phonological facilitation in picture–word interference experiments: Effects of stimulus onset asynchrony and types of interfering stimuli. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **17**, 1146-1160.

Meyer, D. E., & Gordon, P. C. (1985). Speech production: Motor programming of phonetic features. *Journal of Memory & Language*, **24**, 3-26.

Monsell, S. (1987). On the relation between lexical input and output pathways for speech. In A. Allport, D. G. MacKay, W. Prinz, & E. Scheerer (Eds.), *Language perception and production: Relationships between listening, speaking, and writing* (pp. 273-311). London: Academic Press.

Monsell, S. (1990). Frequency effects in lexical tasks: Reply to Balota and Chumbley. *Journal of Experimental Psychology: General*, **119**, 335-339.

Monsell, S. (1991). The nature and locus of word frequency effects in reading. In D. Besner & G. W. Humphreys (Eds.), *Basic processes in reading: Visual word recognition* (pp. 148-197). Hillsdale, NJ: Erlbaum.

Monsell, S., Doyle, M. C., & Haggard, P. N. (1989). Effects of frequency on visual word recognition tasks: Where are they? *Journal of Experimental Psychology: General*, **118**, 43-71.

Morton, J. (1969). Interaction of information in word recognition. *Psychological Review*, **76**, 165-178.

Motley, M. T. (1980). Verification of "Freudian slips" and semantic prearticulatory editing via laboratory-induced spoonerisms. In V. A. Fromkin (Ed.), *Errors in linguistic performance: Slips of the tongue, ear, pen, and hand* (pp. 133-147). New York: Academic Press.

Mowrey, R. A., & MacKay, I. R. A. (1990). Phonological primitives: Electromyographic speech error evidence. *Journal of the Acoustical Society of America*, **88**, 1299-1312.

Mozer, M. C. (1983). Letter migration in word perception. *Journal of Experimental Psychology: Human Perception & Performance*, **9**, 531-546.

Nissen, M. J., & Bullemer, P. (1987). Attentional requirements of learning: Evidence from performance measures. *Cognitive Psychology*, **19**, 1-32.

Nooteboom, S. G. (1973). The tongue slips into patterns. In V. A. Fromkin (Ed.), *Speech errors as linguistic evidence* (pp. 144-156). The Hague: Mouton.

Nooteboom, S. G. (1980). Speaking and unspeaking: Detection and correction of phonological and lexical errors in spontaneous speech. In V. A. Fromkin (Ed.), *Errors in linguistic performance: Slips of the tongue, ear, pen, and hand* (pp. 87-95). New York: Academic Press.

Nooteboom, S.G., & Terken, J. M. B. (1982). What makes speakers omit pitch accents? An experiment. *Phonetica*, **39**, 317-336.

Oldfield, R. C., & Wingfield, A. (1965). Response latencies in naming objects. *Quarterly Journal of Psychology*, **17**, 273-281.

Osgood, C. E. (1971). Where do sentences come from? In D. D. Steinberg & L. A. Jakobovits (Eds.), *Semantics: An interdisciplinary reader in philosophy, linguistics and psychology* (pp. 497-529). Cambridge: Cambridge University Press.

Paap, K. R., McDonald, J. E., Schvaneveldt, R. W., & Noel, R. W. (1987). Frequency and pronounceability in visually presented naming and lexical decision tasks. In M. Coltheart (Ed.), *Attention and performance XII: The psychology of reading* (pp. 222-243). Hillsdale, NJ: Erlbaum.

Pate, D. S., Saffran, E. M., & Martun, N. (1987). Specifying the nature of the production impairment in a conduction aphasic: A case study. *Language & Cognitive Processes*, **2**, 43-84.

Perfect, T. J., & Hanley, J. R. (1992). The tip-of-the-tongue phenomenon: Do experimenter-presented interlopers have any effect? *Cognition*, **45**, 55-75.

Perfetti, C. A., & Goldman, S. R. (1975). Discourse functions of thematization and topicalization. *Journal of Psycholinguistic Research*, **4**, 257-271.

Petersen, S. E., Fox, P. T., Posner, M. I., Mintun, M., & Raichle, M. E. (1988). Positron emission tomographic studies of the cortical anatomy of single-word processing. *Nature*, **331**, 585-589.

Peterson, R. R., Dell, G. S., & O'Seaghdha, P. (1989). A connectionist model of form-related priming effects. In *Proceedings of the Cognitive Science Society* (pp. 166-203). Hillsdale, NJ: Erlbaum.

Pinker, S. (1995). Why the child holded the baby rabbits: A case study in language acquisition. In L. R. Gleitman & M. Liberman (Eds.), *An invitation to cognitive science: Language* (Vol. 1, pp. 107-133). Cambridge, MA: MIT Press.

Potter, M. C., & Lombardi, L. (1990). Regeneration in the short-term recall of sentences. *Journal of Memory & Language*, **29**, 633-654.

Prentice, J. L. (1967). Effects of cuing actor versus cuing object on word order in sentence production. *Psychonomic Science*, **8**, 163-164.

Radeau, M., Morais, J., & Dewier, A. (1989). Phonological priming in spoken word recognition: Task effects. *Memory & Cognition*, **17**, 525-535.

Roediger, H. L. III, Neely, J. H., & Blaxton, T. A. (1983). Inhibition from related primes in semantic memory retrieval: A reappraisal of Brown's (1979) paradigm. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **9**, 478-485.

Roelofs, A. (1992). A spreading activation theory of lemma retrieval in speaking. *Cognition*, **42**, 107-142.

Roelofs, A. (1993). Testing a non-decompositional theory of lemma retrieval in speaking: Retrieval of verbs. *Cognition*, **47**, 59-87.

Saffran, E. M., Berndt, R. S., & Schwartz, M. F. (1989). The quantitative analysis of agrammatic production: Procedure and data. *Brain & Language*, **37**, 440-479.

Saffran, E. M., Schwartz, M. F., & Marin, O. S. M. (1980). Evidence from aphasia: Isolating the components of a production model. In B. Butterworth (Ed.), *Language production* (Vol. 1, pp. 221-241). London: Academic Press.

Savage, G. R., Bradley, D. C., & Forster, K. I. (1990). Word frequency and the pronunciation task: The contribution of articulatory fluency. *Language & Cognitive Processes*, **5**, 203-236.

Schachter, S., Christenfeld, N., Ravina, B., & Bilous, F. (1991). Speech disfluency and the structure of knowledge. *Journal of Personality & Social Psychology*, **60**, 362-367.

Schriefers, H. (1992). Lexical access in the production of noun phrases. *Cognition*, **45**, 33-54.

Schriefers, H. (1993). Syntactic processes in the production of noun phrases. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **19**, 841-850.

Schriefers, H., & Meyer, A. S. (1990). Experimental note: Cross-modal, visual–auditory picture–word interference. *Bulletin of the Psychonomic Society*, **28**, 418-420.

Schriefers, H., Meyer, A. S., & Levelt, W. J. M. (1990). Exploring the time course of lexical access in language production: Picture–word interference studies. *Journal of Memory & Language*, **29**, 86-102.

Schwartz, M. F. (1987). Patterns of speech production deficit within and across aphasia syndromes: Application of a psycholinguistic model. In M. Coltheart, G. Sartori, & R. Job (Eds.), *The cognitive neuropsychology of language* (pp. 163-199). Hillsdale, NJ: Erlbaum.

Schwartz, M. F., Saffran, E. M., Bloch, D. E., & Dell, G. S. (1994). Disordered speech production in aphasic and normal speakers. *Brain & Language*, **47**, 52-88.

Schweickert, R., & Boruff, B. (1986). Short-term memory capacity: Magic number or magic spell? *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **12**, 419-425.

Sevald, C. A., & Dell, G. S. (1994). The sequential cuing effect in speech production. *Cognition*, **53**, 91-127.

Sevald, C. A., Dell, G. S., & Cole, J. S. (1995). Syllable structure in speech production: Are syllables chunks or schemas? *Journal of Memory & Language*, **34**, 807-820.

Shattuck-Hufnagel, S. (1992). The role of word structure in segmental serial ordering. *Cognition*, **42**, 213-259.

Slobin, D. I. (1996). From "thought and language" to "thinking for speaking." In J. Gumperz & S. C. Levinson (Eds.), *Rethinking linguistic relativity*. Cambridge: Cambridge University Press.

Slowiaczek, L. M., & Hamburger, M. (1992). Prelexical facilitation and lexical interference in auditory word recognition. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **18**, 1239-1250.

Smith, V. L., & Clark, H. H. (1993). On the course of answering questions. *Journal of Memory & Language*, **32**, 25-38.

Sridhar, S. N. (1988). *Cognition and sentence production: A cross-linguistic study*. New York: Springer-Verlag.

Stemberger, J. P. (1984). Structural errors in normal and agrammatic speech. *Cognitive Neuropsychology*, **1**, 281-313.

Stemberger, J. P. (1985a). Bound morpheme loss errors in normal and agrammatic speech: One mechanism or two? *Brain & Language*, **25**, 246-256.

Stemberger, J. P. (1985b). An interactive activation model of language production. In A. Ellis (Ed.), *Progress in the psychology of language* (pp. 143-186). London: Erlbaum.

Stemberger, J. P. (1989). Speech errors in early child language production. *Journal of Memory & Language*, **28**, 164-188.

Stemberger, J. P. (1990). Wordshape errors in language production. *Cognition*, **35**, 123-157.

Stemberger, J. P. (1992). The reliability and replicability of naturalistic speech error data: A comparison with experimentally induced errors. In B. J. Baars (Ed.), *Experimental slips and human error: Exploring the architecture of volition* (pp. 195-215). New York: Plenum.

Stemberger, J. P., & MacWhinney, B. (1986a). Form-oriented inflectional errors in language processing. *Cognitive Psychology*, **18**, 329-354.

Stemberger, J. P., & MacWhinney, B. (1986b). Frequency and the lexical storage of regularly inflected forms. *Memory & Cognition*, **14**, 17-26.

Stemberger, J. P., & Treiman, R. (1986). The internal structure of word-initial consonant clusters. *Journal of Memory & Language*, **25**, 163-180.

Stigler, J. W., Lee, S.-Y., & Stevenson, H. W. (1986). Digit memory in Chinese and English: Evidence for a temporally limited store. *Cognition*, **23**, 1-20.

Svartvik, J., & Quirk, R. (Eds.) (1980). *A corpus of English conversation*. Lund, Sweden: Gleerup.

Tannenbaum, P. H., & Williams, F. (1968). Generation of active and passive sentences as a function of subject or object focus. *Journal of Verbal Learning & Verbal Behavior*, **7**, 246-250.

Tent, J., & Clark, J. E. (1980). An experimental investigation into the perception of slips of the tongue. *Journal of Phonetics*, **8**, 317-325.

Titchener, E. B. (1909). *Lectures on the experimental psychology of the thought-processes*. New York: Macmillan.

Turner, E. A., & Rommetveit, R. (1968). Focus of attention in recall of active and passive sentences. *Journal of Verbal Learning & Verbal Behavior*, **7**, 543-548.

Warburton, E., Wise, R. J. S., Price, C. J., Weiller, C., Hadar, U., Ramsay, S., & Frackowiak, R. S. J. (1996). Noun and verb retrieval by normal subjects studied with PET. *Brain*, **119**, 159-179.

Wells, R. (1951). Predicting slips of the tongue. *The Yale Scientific Magazine*, **26**, 9-30.

Werker, J. F., & Lalonde, C. E. (1988). Cross-language speech perception: Initial capabilities and developmental change. *Developmental Psychology*, **24**, 672-683.

Wheeldon, L. R., & Monsell, S. (1994). Inhibition of spoken word production by priming a semantic competitor. *Journal of Memory & Language*, **33**, 332-356.

Wijnen, F. (1992). Incidental word and sound errors in young speakers. *Journal of Memory & Language*, **31**, 734-755.

Wingfield, A. (1968). Effects of frequency on identification and naming of objects. *American Journal of Psychology*, **81**, 226-234.

Yaniv, I., & Meyer, D. E. (1987). Activation and metacognition of inaccessible stored information: Potential bases for incubation effects in problem solving. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **13**, 187-205.

Yaniv, I., Meyer, D. E., Gordon, P. C., Huff, C. A., & Sevald, C. A. (1990). Vowel similarity, connectionist models, and syllable structure in motor programming of speech. *Journal of Memory & Language*, **29**, 1-26.

Zhang, G., & Simon, H. A. (1985). STM capacity for Chinese words and idioms: Chunking and the acoustical loop hypothesis. *Memory & Cognition*, **13**, 193-201.

Zingeser, L. B., & Berndt, R. S. (1988). Grammatical class and context effects in a case of pure anomia: Implications for models of language production. *Cognitive Neuropsychology*, **5**, 473-516.

NOTES

1. Roelofs (1992) has offered an explicit analysis of Stroop interference within a noninteractive production framework. Like Dell (1986), Roelofs has provided an account of production processes that link a name to a visual stimulus, and he has used the framework to account for a range of Stroop effects.

2. Requests for copies of the CD-ROM may be sent to B. MacWhinney, Department of Psychology, Carnegie Mellon University, Pittsburgh, PA 15213.

3. The dialogues were elicited in a laboratory setting and constrained with respect to conversational topic. Information about the corpus is available in the United States from E. Hodas, Linguistic Data Consortium, 441 Williams Hall, University of Pennsylvania, Philadephia, PA 19104-6305 (e-mail: ehodas@unagi.cis.upenn.edu).

4. The results of the experiments indicated that picture processing and message formation were unlikely to be responsible for the observed effects, but the logical possibility of input effects remains.

5. The term *plan* is somewhat misleading. The relevant sense of the word is not a conscious, strategic plot, but a tacit assembly or structural representation of linguistic information.

6. The processes that create these structural configurations are unknown, but the configurations themselves can be described in terms of a hierarchically organized phrase structure. So, the target sentence "Two children are showing the picture to their teacher" corresponds to the phrase structure [Noun Phrase [Verb Phrase [Noun Phrase][Prepositional Phrase]]]. This target sentence is primed as readily by sentences with a corresponding structure but different function words (e.g., [The secretary [made [a cake] [for her boss]]]) as by sentences with a corresponding structure and some of the same function words (e.g., [The secretary [took [a cake [to her boss]]]]; Bock, 1989). Likewise, the target sentence is primed as well by sentences with a corresponding structure but different thematic roles ([The wealthy woman [drove [her Mercedes [to the church]]]]) as by sentences with a corresponding structure and corresponding thematic roles ([The wealthy woman [gave [her Mercedes [to the church]]]]; Bock & Loebell, 1990). If the structure of the priming sentence differs from that of the target, priming is no longer observed (so [Susan [brought [a book [to Stella]]]] is an effective prime for the structure [Noun Phrase [Verb Phrase [Noun Phrase][Prepositional Phrase]]], but [Susan [brought [a book to study]]] is not; Bock & Loebell, 1990).

7. In these experiments, the primary measure of facilitation and inhibition was production order (whether a primed target preceded or followed an unprimed target).

8. Only two of the several conditions in the experiment are included, for purposes of illustration.

9. It might be argued that there is a recognition component to verbal recall that is absent from production, but even that is debatable. Rememberers monitor their recall against some episodic representation of experience (Johnson & Raye, 1981): They check to see if they recognize it as what happened. Since the memory representation of a previous verbal experience is more likely to be a representation of meaning than of form, the basis of this kind of monitoring in recall is the remembered meaning. But speakers likewise monitor, and in much the same way: They keep track of their speech to ensure that what they say is what they mean (Levelt, 1983, 1989).

(Manuscript received May 15, 1995;
revision accepted for publication May 2, 1996.)