
Michael Studdert-Kennedy*

How can we study the process of transforming thought into speech? Can we find in the temporal structure of speech overt, measurable indications of the speaker's underlying cognitive activity? Do pauses serve a communicative function in guiding the listener's segmentation of an utterance? Are there pausal patterns common to different languages or to speech in its many genres—reading, public speaking, telling a story, answering difficult questions, idly conversing, and so on? What light may be thrown by unusual patterns of pausing on the disordered origins of aphasic speech or on the difficulties of second language learners? These and related questions are the topics of this book.

The book comprises the revised versions of 34 presentations given at a workshop on "Pausological Implications of Speech Production," held in Kassel, West Germany in June, 1978. The workshop, jointly organized by psycholinguists at the Gesamthochschule Kassel and at St. Louis University, Missouri, was attended by 37 linguists and psychologists from Canada, France, the Netherlands, Norway, the United Kingdom, the United States, and West Germany. For publication the participants wisely agreed to abandon the pretentious neologism, pausological, in favor of a more accessible word, temporal, although, as I have indicated, the papers do not deal with timing as such, that is, with the origins and mechanisms of temporal order in speech. Rather the "temporal variables" of the title are simply the frequency, duration, and location of pauses in the speech flow from which underlying nontemporal processes (that take time to occur) may be inferred.

Having said this, I should add that many of the papers fall quite outside this rubric. In fact, the papers are extraordinarily heterogeneous, in both topic and quality. Most of them are short (6-10 pages), so that reading the book straight through makes for a bumpy ride, as we jounce from grand, speculative discussion of language and the brain (Karl Pribram) to an opaque pass at extending Thom's theory of catastrophes into the dynamics of verbal planning (Wolfgang Wildgen) to formal proposals for taxonomies of speech

*Also Phonetica, in press.
+Also Queens College and Graduate Center, City University of New York.
Acknowledgment. Preparation of this work was supported in part by NICHD Grant HD-07994 to Haskins Laboratories.

pauses and their role in grammar (Thomas Ballmer, Raimund Drommeli) to the problems of the pause extraction in automatic speech recognition (Jens-Peter Köster, Hede Helfrich). Nonetheless, some coherence does emerge from the editors' grouping of the papers into the sections of the original workshop: general, syntactic, and structural, conversational, prosodic, and cross-linguistic aspects, and a final discussion.

The diversity of approaches evidently reflects some uncertainty among the participants as to what the object of study actually is. The St. Louis contingent (Daniel O'Connell and Sabine Kowal) seems to believe that the uncertainty might be resolved, if only the "field" could be granted a theoretical framework. Ballmer (p. 211) makes a valiant attempt to launch the needed theory with a taxonomy of pause types. He proposes a tripartite classification in terms of airflow intensity, controllability (unintentional vs. intentional) and the potential utility of pauses to speaker and hearer, listing under this last division some twenty-six types—and warning us that any particular pause may be classified under more than one type! The difficulty with such schemes, as Wallace Chafe points out in the final discussion (p. 327), is that interpretive (or functional) taxonomies invite disagreement. In François Grosjean's words: "There are maybe 40 or 50 different variables that can create a silence in speech. A silence may mark the end of a sentence, you can use it to breathe, you can use it to hesitate, there may be ten or fifteen different things happening during that silence" (p. 328). If this is so, there is more than enough room for disagreement on what the operative variables are. Nor are purely objective definitions of pause likely to be of greater use. For example, pause frequency and length may vary with speaker, social situation, speech rate, and a host of other contextual variables, many, if not all, of which are purely inferential. The prospect of filling the theoretical void in the face of this complexity and uncertainty is dim.

What seems to be needed is simplification: careful descriptive and experimental study with clearly defined variables. O'Connell and Kowal in their introductory "Prospectus for a science of pausology" evidently think the time for this is past: "If we are ever to transcend the trivialization which has beset modern psychology...we must find a way of engaging multilogic reality" (p. 9). For this we will find no better way than that of, say, James Joyce. In the meantime, there is science, and this calls for reliable data, systematically collected under well-controlled conditions.

An exemplary instance of an experimental approach is the work of Grosjean and his colleagues on the relations between syntactic structure and the distribution of pauses between words in a sentence. Grosjean reviews preliminary studies of spontaneous speech in interviews, showing that pauses tend to fall at major and minor constituent breaks. Later studies of oral reading showed that variations in syntactic complexity (measured, in one study, by subjects' parsings of the sentences) could account for as much as 56% of the variance in pause duration. Looking for other sources of variance, Grosjean and his colleagues noted that speakers tend to disregard syntactic breaks at certain points, so as to divide constituents into word groups of more or less equal length. They therefore worked up an elegant model to predict the distribution of pauses between words from a weighted index of syntactic complexity and constituent length. In a test of the model, they were able to
account for 72% of the total variance in pause duration. Andrew Butcher (p. 90) reports a study of pauses in the reading of a German story in which the Grosjean model accounted for 86% of the variance.

Yet the matter is not simple. If a pause can be displaced from a syntactic break, it is evidently not a necessary consequence of the speaker's syntactic organization. Moreover, an inconsistent relation between pausing and syntax throws the communicative value of pauses as syntactic markers for the listener into question. Geoffrey Beattie (p. 131) addresses this issue in a study of spontaneous speech designed to assess whether pauses serve an encoding function for the speaker or a communicative function for the listener. He combined analysis of a speaker's speech into hesitant phases (high pause/phonation ratio) and fluent phases (low pause/phonation ratio) with an analysis of the speaker's gaze toward or away from his interlocutor. Beattie found that gaze aversion was very much more likely during hesitant phases than during fluent phases, and was significantly more probable at juncture pauses in a hesitant than in a fluent phase. If we assume that gaze aversion facilitates the self-absorption necessary for clausal planning, we may conclude that pauses, particularly during hesitant phases, may indeed reflect the encoding process. Beattie suggests, further, that "...juncture pauses in fluent phases, accompanied by speaker gaze at the listener, are presumably used to segment the speech for the decoder" (p. 139). However, this attempt to rescue a communicative function for juncture pauses by assigning them a dual function, depending on whether the speaker gazes at or away from the listener, strikes me as unduly tortuous.

The issue comes up again in a lucid and energetic paper by James Deese (p. 69), illustrating, among other things, the complexity of prosodic syntax markers in fluent speech. Deese reports selected analyses of substantial bodies of formal speech recorded at public hearings and committee meetings, at graduate seminars and in radio discussions. He analyzes pause structure in terms of short range grammatical relations within sentences and of long range relations in the structure of discourse. In the short range grammatical analysis, he makes several telling (if not always new) observations: (1) sentence boundaries are frequently (24% in one sample of 1043 randomly selected boundaries) marked neither by a rising or falling intonation contour nor by a break in acoustic energy (i.e., a pause); (2) where sentence boundaries are not marked by intonation or pause, they may often be marked by increased syllable rate on both sides of the boundary; (3) in tests with words excised from context listeners are most accurate in detecting a boundary when it is marked both by intonation contour and by a pause longer than 50 msec; (4) listeners judge a given pause as longer if it occurs at a clause break than if it occurs within a clause.

The burden of these observations is that the prosodic devices by which syntactic structure may be marked in fluent speech are far from simple. Moreover, the fact that listeners' judgments of pause length may be determined by the syntactic structure, rather than the reverse, suggests that other prosodic variables may be marking the syntax and may even be determining the pause structure.

Alan Henderson (p. 198) reports an ingenious study that speaks to this last point. Starting from the well-known click studies in which reaction time
is elevated for a click placed in a syntactically marked, but prosodically unmarked, clause break, he asked whether he might not find a similar increase in reaction time for a tone placed in a syntactically unmarked, but prosodically marked, break. He measured English listeners' reaction times to a tone placed at the end of a word in each of six Czechoslovakian sentences (none of the listeners knew, or recognized the language as, Czech). The sentences were manipulated so that the tone followed either an intonation fall and a pause, a fall alone, a pause alone, or neither. Reaction times to the tone were significantly longer in the three conditions where it followed a fall than in the other conditions. From this Henderson concludes that an intonation fall is a more salient cue to segmentation than a pause. Indeed, he turns the tables completely by suggesting that "...a break in signal energy is perceived as it is because of its context rather than being a cue to the structuring of the context" (p. 205). Certainly, as Henderson also sensibly suggests, a child (or an adult) learning a language is likely to find intonation a more reliable guide to syntactic structure than pauses—for which, the participants in this workshop unanimously agreed, the determinants are many and various.

If intonation is the principal cue to syntactic segmentation, might not the correlation of pause structure with syntax simply reflect a role of intonation in determining the location and, perhaps, length of pauses? Yet it cannot be the sole determinant, since the correlation between pausing and syntax would then be as high as between intonation and syntax. What then of rhythm and rate? Here the evidence is suggestive, though certainly not conclusive. Anne Cutler (p. 183) describes errors of syllable omission in spontaneous speech that have the effect of equalizing the number of syllables per foot, and thus making the speaker's output more isochronous. Interestingly, this may be just the effect of speakers' tendencies to bisect constituents, observed by Grosjean. Ballmer (p. 216) also remarks that pauses may serve to maintain the rhythmic pattern of an utterance. Finally, as far as rate is concerned, Grosjean (pp. 92-93) reports that pauses (both breathing and nonbreathing) tend to disappear, first from minor, then from major constituent breaks as rate is increased, until, at the highest rates (391 words per minute, in the study reported) only breathing pauses at some sentence boundaries remain.

What all this comes down to, then, is that pauses in fluent speech that seem to reflect the speaker's planning of syntactic structure, may be epiphenomenal consequences of other prosodic variables. As Butcher remarks: "...it would seem...neither feasible nor desirable to investigate pausing separately from certain other dependent variables, such as intonation, rhythm, and tempo" (p. 86). Butcher goes on to conjecture that: "...rather than all prosodic variation, including pausing, being determined by the syntactic structure, pausing is determined by intonation pattern, which in turn is normally coterminous with the syntactic pattern" (p. 90). If this proves to be so, we may conclude that syntax-marking pauses have little or no direct communicative function.

Let us consider now pauses to which we might be less inclined to assign intended communicative value: unfilled and filled pauses (that is, pauses containing hesitation sounds: uh...er, and the like) in which a speaker is quite evidently at a loss for a discourse plan, that is, for what to say. The central difficulty in studying the cognitive activity that underlies these
hesitations is that, under normal circumstances, the investigator has even less idea of what speakers have in mind than the speakers themselves. One solution to the difficulty is to provide the speaker with a sort of open-ended script, a general discourse plan that the investigator knows, but that the speaker has to formulate. Thus, Wolfgang Klein (p. 159) induced much lengthy hesitation by asking people for route directions in a city. He could then compare the alternate routes, false starts, backtracks, and roadblocks in the speakers' cognitive map, inferred from their utterances, with the clear "discourse plan" laid out in an actual map of the city.

Chafe (p. 169) offered his subjects a richer opportunity for self-revelation by asking them to tell what had happened in a 7-minute color movie (with sound effects, but no dialog) they had just seen. To introduce his analysis of the resulting spontaneous narratives, Chafe quotes William James on the stream of consciousness: "Like a bird's life, it seems to be made of an alternation of flights and perchings. The rhythm of language expresses this, where every thought is expressed in a sentence, and every sentence closed by a period" (James, 1890, p. 243). Chafe applies the metaphor to describe how someone tells a story, talking in spurts of a few seconds at a time, darting from one "focus of consciousness" to another. Foci, expressed in phrases or clauses with a rising pitch contour and a brief following pause, form "clusters" (or sentences) that end with a falling contour and a somewhat longer pause. Examining the content of foci within a cluster, we see how the speaker flits from point to point, capturing different aspects of a scene, or grouping a run of small events into a single purposive action. Long hesitations between clusters often reflect "time-consuming mental processing," as the speaker switches to a new time, place, actor, event, or scene.

Chafe argues that such "hesitation-ridden speech" should not be regarded as disfluent, even if technically ungrammatical, but rather "...should actually be highly valued as an accurate expression of a speaker's thoughts" (p. 180); he expects his mode of analysis to become "...an important and necessary aspect of hesitation research" (p. 180). Perhaps he is right, but I am not sure where it all leads. What he offers seems to be little more than a traditional explication du texte, extended from works of literature to the "creative act" (p. 170) of commonplace speech production.

Indeed, Chafe's chapter, like many others in this book, inadvertently draws attention to the contrast between pauses and errors as sources of inference about the cognitive processes of a speaker. Bernard Baars remarks in the general discussion, "...slips of the tongue are revealing in a way that pauses are not. Slips say something, and if you want to make inferences regarding deeper levels of control in speaking, you have more information to go on (p. 336)." In fact, the form of errors has already served to constrain our models of language processing, and their study is by no means exhausted. This point is well illustrated by two papers in the general introductory section of the book.

The first paper, by John Laver (p. 21), reports an experiment designed to induce errors by requiring subjects to speak pairs of vowels in a /pVp/ frame at increasingly fast rates. The hypothesis was that rapid, successive execution of vowel pairs drawing on relatively distinct neuromuscular systems (e.g., front-back, high-low) might invite competition between the two systems,
leading to errors such as diphthongal glides, while different degrees of activation of roughly the same neuromuscular system (as in tense-lax front, or tense-lax back, vowel pairs) would preclude competition and so elicit few errors. This is precisely what was found. The experiment is modest and the report preliminary, but, as Laver points out, the principle of neuromuscular compatibility, illustrated in the pattern of errors, might be fruitfully applied in diverse areas of phonetic study, from the derivation of natural phonological classes to language acquisition and second language learning.

The second paper, by E. Keith Brown (p. 28), introduces the (for me) novel notion of "grammatical incoherence." An instance is the utterance of a young girl, stroking a moulding cat and holding up a hair: "How long do you suppose a life of a fur has?", spoken without hesitations and with apparent confidence that she had produced an intelligible utterance—as indeed she had. (For, as Brown remarks, listeners are far more tolerant of grammatical incoherence than of word distortion and such incoherence seldom impedes communication.) Brown uses this example to distinguish between two types of "blend," reflected in such incoherences. In a "cognitive blend" two related, but different cognitive structures with different surface realizations (fur, hair) compete, and the wrong one wins. Such errors may tell us about the organization of a speaker's lexicon and the processes of selection from it. In a "process blend," by contrast, "a single cognitive structure...may be realized by a number of surface forms and the resultant utterance is a blend of the processes that lead to these different forms" (p. 35). Thus, equivalent forms (e.g., How long a life does a hair have? How long a life has a hair? How long is the life of a hair?) may blend to produce "How long a life of a hair has?". Such errors may tell us about the processes of selecting from equivalence classes of syntactic forms. Of course, if a speaker avoids errors by pausing long enough to choose the right word or turn of phrase, we learn nothing: we detect his quandary, perhaps, but not its content. Brown's is an original and illuminating paper.

The final section of the book deals with cross-linguistic aspects. Here, it would seem, there might be an opportunity to dissociate general cognitive constraints due to syntax, tendencies toward stress or syllable timing and, perhaps, characteristic rates of speech. Thus, Grosjean, in a brief, but useful review paper (p. 307), reports that while pause time ratios in the spontaneous English and French of interviews are almost identical, they are arrived at in different ways: pauses are fewer, but longer in French, more frequent, but shorter in English. The constant ratios perhaps reflect breathing demands, common to all spoken languages; but the more frequent pauses of English reflect a tendency (syntactically governed, Grosjean implies, though it is not clear how) for speakers to insert pauses inside verb phrases, as they do not in French. On the other hand, a tendency, reported by Marc Faure (p. 287) for pauses in German to be most frequent before pronouns (as they are not in French) simply reflects a tendency, common to English, French, and German, to pause before the first or second word in a subordinate clause, of which the pronouns, in German, must be placed first.

Indeed, one may doubt the worth of including pause instruction in second language courses, recommended by Robert DiPietro (p. 320), for several reasons. First, the differences across the admittedly few languages that have been studied do not appear to be great. Alain Deschamps (p. 255) does report
that French students tend to carry French patterns over into their English; but the most general effect among second language speakers, reported by both Deschamps and Manfred Raupach (p. 263) is, not surprisingly, an increase in the frequency (not the length) of pauses within sentences. Raupach reports, further, that many individuals have idiosyncratic pause patterns in their first language that they are likely to transfer into a second language. Finally, my overall impression, gathered from many papers in this book, is that pauses—other than those introduced for deliberate rhetorical effect—are largely automatic consequences of cognitive and physiological processes over which speakers have little control.

The last point emerges with particular cogency from studies, reviewed by Grosjean, comparing the pause structures of an oral language (English) with those of a manual-facial language (American Sign Language (ASL)). Freed from the demands of breathing, a sign language can reduce the amount of time spent in pausing; the pause time ratio for ASL is, in fact, less than half that of English. On the other hand, since a sign takes longer to form than a word, the overall rate of signs per minute is less than a third of the rate of English words per minute. Yet the proposition rates in the two languages are almost identical. The paradox is resolved by noting that, while the phonological and syntactic structures of a spoken language are largely due to sequential organization over time, a highly inflected signed language, such as ASL, can make extensive use of simultaneous manual, bodily, and facial gesture, distributed in space. Quite different means are thus used in the two languages to maintain what may be a natural rate of information flow common to all languages.

Despite these differences, the durations of ASL signs seem to be influenced by many of the factors that influence word duration, such as semantic novelty and position within a phrase. Moreover, the reduced pause time ratio of ASL is accomplished by shorter, not fewer pauses, so that its pause pattern can be quite similar to that of a spoken language. In fact the distribution of pauses between signs in "recited" sentences, like the distribution of pauses between words, reflects both constituent structure and the length of constituents: the model of Grosjean and his colleagues, discussed earlier, accounted for 72% of the variance in a study of ASL, as it had in a study of speech. Of course, the communicative function of pauses, no less than their possible determination by other prosodic variables, such as rhythm and rate, are even less understood for ASL than for spoken languages. Nonetheless, cross-modal comparison between signed and spoken languages promises to isolate universal cognitive and motoric constraints on language production.

In conclusion, we can be confident that universities will not now rush to establish Departments of "Pausology." On the contrary, the message of this interesting, if uneven, book is that the study of pauses in the speech flow will be advanced not by isolation, but by integration into other areas of phonetic and general psycholinguistic study.

REFERENCES