*Cognition*, 10 (1981) 65–70



# Making up materials is a confounded nuisance, or: Will we be able to run any psycholinguistic experiments at all in 1990?

ANNE CUTLER*

*University of Sussex*

Research in cognitive psychology tends to be paradigm-driven at the best of times, and the seventies haven't even been the best of times. The most judicious attempts to break the mould can be self-defeating; see Lockhart (1978), for example, bewailing the fact that the 'levels of processing' approach, devised by Craik and Lockhart (1972) as an attempt to inject more real-life relevance into memory research, was enthusiastically taken up by the field and developed into a self-perpetuating paradigm.

Psycholinguistics exemplifies the general predicament. Its history over the past decade chronicles as much as anything else the continual discovery of new confounds. In order to facilitate this exercise, psycholinguists now conscientiously publish their materials in full. The more materials are published, the more confounds can be and are discovered. (Publish and perish.)

In the following pages I will illustrate, by way of a few judicious examples, what this means for the ordinary designer of psycholinguistic experiments; and of course, since I too wish to make an immortal contribution to the psycholinguistic literature, I may not refrain from pointing to a few confounds myself.

## Example 1: What happened to the ambiguity effect in phoneme-monitoring

In the early seventies there was held to be an 'ambiguity effect' in phoneme-monitoring; when the word preceding the target-bearing word was ambiguous, reaction times to detect the target were slower than when the preceding word was unambiguous (Foss 1970), and this was true even when prior context made it quite clear which meaning of the ambiguous word was intended (Foss and Jenkins 1973; Cutler and Foss 1974). The effect was

*Reprint requests should be sent to Anne Cutler, Laboratory of Experimental Psychology, University of Sussex, Brighton BNI 9QG, England.

explained as an increase in processing load due to the necessity to choose between the alternative meanings of the ambiguous word. But then along came Mehler, Segui and Carey (1978), and Newman and Dell (1978), who pointed out that the unambiguous control words in these experiments had more often than not been longer than the ambiguous words. Perhaps this added length had allowed the subjects a little extra processing time in the unambiguous condition. Indeed, when length was controlled, the ambiguity effect disappeared; in fact by making ambiguous words longer than the unambiguous controls, it was possible to produce phoneme-monitoring reaction times which were faster following ambiguous than unambiguous words. Newman and Dell pointed out yet another confounding factor: the ambiguous words often began with sounds which were phonologically similar to the target sound beginning the following word, whereas the unambiguous control words rarely did; judicious control of this factor also removed any indication of an ambiguity effect.

The lesson to be learnt from this episode was, of course: take more care in constructing phoneme-monitoring experiments! No conclusions were drawn about the processing of ambiguous words. As it turned out, the hypothesis behind the early phoneme-monitoring work on lexical ambiguity was not entirely ill-conceived; it really does appear to be the case that occurrence of an ambiguous word in a sentence results in all possible meanings of the word being momentarily activated, irrespective of disambiguating context (Swinney 1979). It is not the case, however, that this produces processing difficulty measurable *via* phoneme-monitoring response time.

### Example 2: Timing, frequency and intensity are very important[1]

As listeners process an utterance, they pay close attention to the prosodic (timing, frequency, intensity) variations; in fact, they will follow prosodic continuity at the expense of semantic continuity (Darwin 1975). Phoneme-monitoring response times are faster to targets on stressed words than to targets on unstressed words (Shields, McHugh and Martin 1974; Cutler and Foss 1977). Moreover, listeners can use the prosodic contour to direct their attention to the most highly stressed parts of an utterance, leading to faster monitoring times (Cutler 1976; Cutler and Darwin 1981). Obviously sentence prosody is a most important factor in sentence comprehension, and

[1] As the actress said to the bishop.

ought to be taken into judicious consideration in designing and interpreting sentence processing experiments—particularly monitoring experiments. But is it? Well, usually not.[2]

## Example 3: How to calculate word frequency

All psycholinguists learn at their supervisor's knee the importance of frequency of occurrence as a predictor of word recognition time; everybody controls for frequency. And so we should, since Whaley (1978) has shown that it may well be the single strongest influence on lexical decision reaction time.

Theoretical explanations for the frequency effect in word recognition appeal to the structure and access of the mental lexicon Forster 1976; Morton 1978. Among the things we know about the mental lexicon is that words regularly inflected, e.g., for tense and number, do not appear to be represented independently of their uninflected form (Gibson and Guinet 1971; Murrell and Morton 1974; Stanners, Neiser, Hernon and Hall 1979). Therefore the frequency with which a particular lexical representation (say, *pick*) is accessed ought to be better approximated by the summed frequency of the uninflected with the regularly inflected forms (i.e., *pick* + *picked* + *picks* + *picking*—151 in Kučera and Francis [1967])—than by the surface frequency of *pick* alone (55)). Sure enough, the combined frequency produces stronger frequency effects than the surface frequency (Rosenberg, Coyle and Porter 1966; Taft 1979). It follows, then, that one does better to match experimental materials on this combined frequency measure than on surface frequency. Some judicious experimenters do this (e.g. Bradley 1978); most don't.[3,4]

Alas for us all, this is not even the whole story. Judicious matching on combined frequency, though more difficult than matching on surface

[2] Consider for instance a recent word-monitoring experiment in which reaction times to the same target word were compared in three types of context: normal sentences, semantically anomalous but grammatically acceptable sentences, and scrambled strings of words. One of the important results of this experiment was that reaction times were faster in the normal sentences than in the abnormal, and faster in the strings which were only semantically anomalous than in those which were syntactically anomalous as well. The speaker who recorded the experimental sentences was not aware which words were the targets. One might expect more prosodic cues to the location of content words (all of the target words were content words) in the syntactically regular conditions than in the scrambled condition, and since sentence stress is semantically determined, more prosodic cues to sentence stress in the semantically normal condition than in the semantically anomalous condition. Thus these experimental materials may have confounded prosodic cues with syntactic and semantic cues.

Notes 3 and 4. *(Please see overleaf)*

frequency, can be done. However, it turns out that we ought to match on surface frequency *as well.* Taft (1979) found that combined frequency correctly predicted reaction time differences when surface frequency was controlled, but surface frequency also correctly predicted reaction time differences when combined frequency was controlled!

## 1981

Consider a judicious psycholinguist constructing materials for an experiment comparing nouns, verbs and adjectives. Ideally he would like to create matched triples of an unambiguous noun with an unambiguous verb and an unambiguous adjective. They should be matched, as we have seen, on both surface and combined frequency. Naturally they should be matched on length. At this point it is already clear to the experimenter that the task is probably impossible[5]; and he has not even begun to consider further variables on which they might be matched, such as association, age of acquisition, autobiographical memory, categorizability, concreteness, digram frequency, imagery, goodness, letter frequency, number of meanings, orthographic regularity, meaningfulness, emotionality and recognition threshold Rubin 1980; Whaley 1978; Jastrzembski 1981.

No easier task confronts the psycholinguist designing a phoneme-monitoring experiment in which the experimental sentences contain, say, words of high morphological complexity while the control sentences contain morphologically simple words. On the basis of several models which describe specifically the process of monitoring for phonemes (Cutler and Norris 1979; Foss and Blank 1980), the experimenter can design the materials so that the

[3] A recent lexical decision experiment on prefixes, for instance, investigated words which could occur either alone or with prefixes. Words like *pending*, which have lower frequency than their prefixed relatives (e.g., *impending*) were compared with words like *bark*, which have higher frequency than prefixed forms (e.g., *embark*). Each word was matched on length and surface frequency with a non-prefixable control word; *pending* with *picking*, for instance, and *bark* with *bull.* It was predicted that there would be no reaction time difference between *bark* words and their controls, but *pending* words would be responded to slower than their controls because the higher frequency prefixed form would interfere; this pattern of results was indeed found. But when one looks at combined frequency the matching turns out to be imbalanced. Thus while *pending* and *picking* each have a surface frequency of 14, the combined frequency measures are respectively 14 (i.e., no other form of *pend* occurs) and 151. In fact of the 20 *pending* words, 16 were less frequent than their controls on the combined frequency measure, whereas the *bark* words did not differ significantly from their controls in this respect (11 more frequent, 8 less frequent, one equal).

[4] The Thorndike-Lorge (1944) word count sums frequencies across regular inflections; unfortunately it is also 23 years more out of date than the Kučera-Francis (1967) count.

[5] Even with the invaluable assistance of Coltheart (1981).

monitoring response reflects the lexical characteristics of either the target-bearing word or the preceding word, according to choice. But the experimenter still has to match materials. The sentence prosody must be equivalent across experimental and control sentences, for example, the phonological similarity of target and initial sound of preceding word must be controlled. And most of the word recognition factors described above will be relevant.

This is not nearly the end of the story. We have not yet begun, for instance, to assess the whole new range of possible confounds opened up by Marslen-Wilson's observation (Marslen-Wilson and Welsh 1978) that words differ markedly in the position of their recognition point, i.e., the point at which, counting from left to right, they become unique from all other words of the language. The prospects are gloomy. If it goes on this way in the eighties, psycholinguists will literally be lost for words. Perhaps it is time for us to take matters—or rather, materials—into our own hands. Judicious choice of language in all our writings, for instance, combined with judicious extension of our fields of publication to other literary domains, could eventually allow us to exercise (judicious) control over the ratings assigned to words in future frequency counts. By way of a beginning, this essay represents a modest attempt to upgrade the frequency rating of the word *judicious*.

## References


Bradley, D. C. (1978) *Computational Distinction of Vocabulary Type*. Ph. D. Dissertation, Cambridge, Mass., MIT.

Coltheart, M. (1981) *Psycholinguistic Database User Manual*. London, Medical Research Council

Craik, F. I. M. and Lockhart, R. S. (1972) Levels of processing: A framework for memory research. *J. verb. Learn. verb. Behav., 11*, 671–684.

Cutler, A. (1976) Phoneme-monitoring reaction time as a function of preceding intonation contour. *Percep. Psychophys., 20*, 50–60.

Cutler, A. and Darwin, C. J. (1981) Phoneme-monitoring reaction time and preceding prosody: Effects of stop closure duration and of fundamental frequency. *Percep. Pshychophys., 29*, 217–224.

Cutler, A. and Foss, D. J. (1974) Comprehension of ambiguous sentences: The locus of context effects. Paper presented to the Midwestern Psychological Association, Chicago.

Cutler, A. and Foss, D. J. (1977) On the role of sentence stress in sentence processing. *Lang. Sp., 20*, 1–10.

Cutler, A. and Norris, D. G. (1979) Monitoring sentence comprehension. In W. E. Cooper and E. C. T. Walker (eds.), *Sentence Processing: Psycholinguistic Studies Presented to Merrill Garrett*. Hillsdale, N J, Erlbaum.

Darwin, C. J. (1975) On the dynamic use of prosody in speech perception. In A. Cohen and S. Nooteboom (eds.), *Structure and Process in Speech Perception*. Heidelberg, Springer.

Forster, K. I. (1976) Accessing the mental lexicon. In R. J. Wales and E. C. T. Walker (eds.), *New Approaches to Language Mechanisms*. Amsterdam, North-Holland.

Foss, D. J. (1970) Some effects of ambiguity upon sentence comprehension. *J. verb. Learn. verb. Behav., 9,* 699–706.
Foss, D. J. and Blank, M. (1980) Identifying the speech codes. *Cog. Psychol., 12,* 1–31.
Foss, D. J. and Jenkins, C. M. (1973) Some effects of context on the comprehension of ambiguous sentences. *J. verb. Learn. verb. Behav., 12,* 577–589.
Gibson, E. J. and Guinet, L. (1971) Perception of inflections in brief visual presentations of words. *J. verb. Learn. verb. Behav., 10,* 182–189.
Jastrzembski, J. E. (1981) Multiple meanings, number of related meanings, frequency of occurrence, and the lexicon. *Cog. Psychol., 13,* 278–305.
Kučera, H. and Francis, W. N. (1967) *Computational Analysis of Present-Day American English.* Providence, R I, Brown University Press.
Lockhart, R. S. (1978) Method and content in the study of human memory. In J. P. Sutcliffe (ed.), *Conceptual Analysis and Method in Psychology: Essays in Honour of W. M. O'Neil.* Sydney, Sydney University Press.
Marslen-Wilson, W. and Tyler, L. K. (1980) The temporal structure of spoken language understanding. *Cog., 8,* 1–71.
Marslen-Wilson, W. and Welsh, A. (1978) Processing interactions and lexical access during word recognition in continuous speech. *Cog. Psychol., 10,* 29–63.
Mehler, J., Segui, J. and Carey, P. W. (1978) Tails of words: Monitoring ambiguity. *J. verb. Learn. verb. Behav., 17,* 29–35.
Morton, J. (1978) Word recognition. In J. Morton and J. C. Marshall (eds.), *Psycholinguistics Series 2: Structures and Processes.* London, Elek Books.
Murrell, G. A. and Morton, J. (1974) Word recognition and morphemic structure. *J. exper. Psychol., 102,* 963–968.
Newman, J. E. and Dell, G. S. (1978) The phonological nature of phoneme-monitoring: A critique of some ambiguity studies. *J. verb. Learn. verb. Behav., 17,* 359–374.
Rosenberg, S., Coyle, P. J. and Porter, W. L. (1966) Recall of adverbs as a function of the frequency of their adjective roots. *J. verb. Learn. verb. Behav., 5,* 75–76.
Rubin, D. C. (1980) 51 properties of 125 words: A unit analysis of verbal behavior. *J. verb. Learn. verb. Behav., 19,* 736–755.
Shields, J. L., McHugh, A. and Martin, J. G. (1974) Reaction time to phoneme targets as a function of rhythmic cues in continuous speech. *J. exper. Psychol., 102,* 250–255.
Stanners, R. F., Neiser, J. J., Hernon, W. P. and Hall, R. (1979) Memory representation for morphologically related words. *J. verb. Learn. verb. Behav., 18,* 399–412.
Swinney, D. A. (1979) Lexical access during sentence comprehension; (Re)consideration of context effects. *J. verb. Learn. verb. Behav., 18,* 545–569.
Taft, M. (1979) Recognition of affixed words and the word frequency effect. *Mem. Cog., 7,* 263–272.
Taft, M. and Forster, K. I. (1975) Lexical storage and retrieval of prefixed words. *J. verb. Learn. verb. Behav., 14,* 638–647.
Thorndike, E. L. and Lorge, I. (1944) *The Teacher's Word Book of 30,000 Words.* New York, Columbia University.
Whaley, C. P. (1978) Word-nonword classification time. *J. verb. Learn. verb. Behav., 17,* 143–154.