Remarks and Replies

One Mental Lexicon, Phonologically Arranged: Comments on Hurford's Comments

Anne Cutler, David A. Fay

In our 1977 article, "Malapropisms and the Structure of the Mental Lexicon", we concluded that the word substitution error data which we had described "provided evidence for the existence of but one mental lexicon, and for the organization of the entries in the lexicon according to phonological properties" (p. 517). Hurford (1981) offers a critique of the 1977 article; however, his quarrel is not with these conclusions, but with our further suggestions about the nature of the phonological organization within the lexicon. We proposed that words were simply arranged in terms of phonological similarity on a left-to-right basis, so that any word's nearest neighbor is the word which sounds most like it left-to-right. Our basis for this suggestion was the fact that malapropisms (word substitution errors which are unrelated to the intended word in meaning, but very like it in sound) resemble their target words most strongly in the initial portions; we hypothesized that a malapropism arises when the language production device mistakenly selects, instead of the intended word, its nearest neighbor in the lexicon.

Hurford offers a convincing demonstration that the malapropism and target pairs tend to be more alike than would be predicted by chance at other points in the word than in the initial portions.[1] We are in no doubt that the regularities to which he points

Preparation of this remark was supported (for the first author) by a grant from the Science Research Council. Thanks to Dennis Norris for useful discussions on lexical models.

[1] Hurford reports his analysis in the form of a 2-by-2 table in which the rows represent the malapropism and pseudo-error corpus, respectively, the columns the number of matched and mismatched segments after the departure point. He reports the statistical significance of this distribution in terms of the χ^2 statistic, using the formula $(A-D)^2/A+D$, where A and D are two cells in a 2-by-2 matrix. This test, logically identical to the Cochran Q test, is known as the McNemar test for the significance of changes (McNemar (1969, 261); Siegel (1956, 63)) and is intended for use in comparing a binary distribution (pass–fail; match–mismatch) under two conditions. The two cells on which the χ^2 formula is based are those which represent different values in the two conditions, i.e. those which were matched under one condition, mismatched under the other. Thus, the matrix which Hurford should have reported was one in which the rows and columns reported the match/mismatch distribution in the malapropisms and pseudo-errors, respectively; the matrix which he did report is

Linguistic Inquiry, Volume 13, Number 1, Winter 1982
0024-3892/82/010107–07 $02.50/0

are indeed characteristic of malapropisms. The conclusions which Hurford draws from his observations are:

(a) Malapropisms and targets are alike not only in their initial portions but at both extremities.

(b) It is not the case that words are arranged in the mental lexicon by left-to-right phonemic structure.

(c) Words with derivational suffixes are listed in the lexicon intact rather than as stems to which affixes are attached by rule in speech production.

In these comments we would like to draw somewhat different implications from Hurford's further analysis of our data.

First, inspection of the corpus reveals that Hurford's "both extremities" hypothesis is, like the hypothesis of "left side similarity only", an underestimate of the phonological correspondences between malapropisms and targets. The malapropisms resemble their targets as much in the middle as at the end (e.g. *deserved* for *diverged, miraculous* for *spectacular*). To assess the implications of this, it is necessary to make quite clear how Hurford's reanalysis of our data differs from our original presentation. We established

---

one on which it makes no sense to calculate McNemar's test, since the entries in the cells, representing the matches and mismatches for the two sets of errors *to the same target segments*, are not independent. Moreover, the two cells which he chose to enter into the calculation of χ^2 were malapropism matches and pseudo-error mismatches; it is hard to imagine what a comparison of these two values could be expected to show. For the record, we have carried out the analysis which Hurford should have done (we are grateful to Hurford for making his pseudo-error corpus available to us). We report this revised analysis in table (i).

Table (i)

	Number of segments	
	Malapropisms	
	Match	Mismatch
Pseudo-errors Match	106	31
Pseudo-errors Mismatch	137	251

With correction for continuity, $\chi^2(1) = [(137-31)-1]^2/137+31 = 65.63$, $p < .001$. Similarly, Hurford's second table—the same analysis based only on those examples in which neither malapropism nor target was suffixed (Hurford reports finding 61 such pairs, but we found only 60)—should be replaced by table (ii).

Table (ii)

	Number of segments	
	Malapropisms	
	Match	Mismatch
Pseudo-errors Match	9	12
Pseudo-errors Mismatch	36	79

Here, $\chi^2(1) = 11.02$, $p < .001$. These values should be substituted for those reported by Hurford. It will be noted that the total number of segments that we have tallied differs slightly from Hurford's total. This could result from the fact that in our analysis affricates were counted as one segment rather than two, or it could be that in some cases Hurford chose an inappropriate transcription, perhaps one which did not allow for the fact that the majority of the errors we reported were committed by speakers of American English.

the obvious phonological similarity between malapropism and target by example alone. Later in the article we compared the malapropism–target pairs with a set of semantically related word substitution errors and their targets. For each pair, we located the point at which (counting from the beginning of the word) the target and error departed from identity, and compared the distinctive feature marking of the segments at this departure point. Thus, for *divergence* versus *diversion* the segments at the departure point are [ǰ] and [ž], which differ on only one feature. The modal number of features on which the segments at the departure point differed was much lower for the malapropisms than for the semantic errors; on the basis of this, we argued that phonological similarity in the mental lexicon was defined by distinctive feature correspondence (assuming left-to-right arrangement of words by phonological similarity). As indirect evidence in favor of the left-to-right hypothesis, we noted that (a) the vowels in the first syllables of malapropisms and targets were alike significantly more often than the vowels in the stressed syllables; (b) a distinctive feature analysis on the departure point counting from the end of the word revealed no difference between the set of malapropisms and the set of semantic errors.

It is clear that the very nature of our analysis prevented us from taking account of the similarities between error and target which lay *between* the points of departure from identity counting from the left and from the right, however obvious these similarities might have been (the vowel in *bleed/breathe*, for example, or the second and third syllables of *miraculous/spectacular*).

Hurford's analysis was of a different kind. He constructed a pseudo-error corpus in which the features most characteristic of malapropisms—correspondence to the target in grammatical category, initial phonological segments, number of syllables, and stress pattern—were held constant.[2] Thus, the pseudo-error for *breathe* was *bet* (real error *bleed*); the pseudo-error for *spectacular* was *irrational* (real error *miraculous*). Hurford then simply tallied the number of segments on which the targets corresponded with the malapropisms and the pseudo-errors, respectively, to the right of the point of departure from identity. Note that this does not mean that additional similarities between malapropism and target which are revealed by this analysis will of necessity occur at the rightward extremity of the word, since if the departure point occurs early—as it does in *bleed/breathe* and *miraculous/spectacular*—the pseudo-error analysis will embrace the end, the middle, and possibly all or part of the beginning of the word. In fact, it is clear that Hurford's analysis covers more than the rightward extremity from the very fact that for the 156 target words in the Fay and Cutler corpus there are a total of 272 segments occurring *before* the departure point, but approximately twice as many after it. Thus, on average Hurford's analysis embraces nearly two-thirds of the target word.

The importance of this point is that Hurford's argument in favor of the unanalyzed

[2] In constructing his corpus in this manner, Hurford concedes that malapropisms resemble their targets strongly in the initial portions.

Table 1

Number of corresponding segments

	Malapropisms	
	Match	Mismatch
Match	12	3
Pseudo-errors		
Mismatch	32	82

listing of derived words rests upon the right-extremity claim. His argument is as follows:

1. If malapropisms and their targets resemble each other at each extremity and if derived words are listed in the lexicon in their base form (i.e. without suffixes), then right-extremity resemblances could be due merely to coincidence of suffix.

2. Right-extremity resemblances are not due merely to coincidence of suffix because nonderived targets also show significantly more post-departure point correspondences with the malapropisms than with the pseudo-errors.

3. Therefore, suffixed words are not listed in the lexicon in their base form.

Clearly this is not a valid argument as it stands. To make it work, it would be necessary to show (a) that malapropisms resemble their targets *at each extremity and not in the middle*, and (b) that derived malapropism–target pairs are not alike in the middle (i.e. at the end of their base). To assess the validity of these hypotheses, we examined 53 malapropism–target pairs which had been excluded from Hurford's nonderived subset of the data. In each pair, both malapropism and target bore derivational suffixes; examples are *photogenic* for *photographic, participate* for *precipitate*. We carried out the same analysis as Hurford had performed, using Hurford's pseudo-error corpus as comparison.[3] That is to say, we counted the number of segments in the target to the right of the departure point which were matched by segments in the malapropism and the pseudo-error, respectively; but in this case our analysis did not include the derivational suffix. Thus, for *photogenic/photographic/photostatic* (malapropism/target/pseudo-error) the sections to be compared (excluding the beginning up to and including the departure point, and excluding the suffix) were [ɛn], [ræf], and [tæt], respectively; for *participate/precipitate/perambulate* the sections were [rtɪsɪp], [əsɪpɪt], and [ræmbjul], respectively. The result of this analysis is given in table 1. This difference is statistically significant; $\chi^2(1) = 22.4$, $p < .001$. Thus, it can be seen that once again the malapropisms have significantly more segments in common with the target words after the departure

[3] This is not to say that we have no reservations about Hurford's pseudo-error corpus. A number of his examples do not in fact match the target on the dimensions which he claims to have controlled, and many are of such low frequency of occurrence (e.g. *Mozarab, luff, hopsacking, jennet, amyloid*) that it is doubtful that they form part of the English vocabulary represented in the mental lexicons of speakers such as those from whom our error examples are drawn.

point than do Hurford's pseudo-errors. In fact, the malapropism–target correspondences in the base of the 53 suffixed words analyzed here are rather more striking than the correspondences in the 60 nonsuffixed words mentioned in footnote 1; it could thus be suggested that this fact provides evidence that suffixed words are indeed listed in the lexicon in their base form, the additional correspondences due to suffix overlap being purely coincidental. However, we do not wish to take a position on the question of storage of derived words here, merely to demonstrate that Hurford's argument about the storage of derived words is invalid. For future discussion of this issue, though, we draw the reader's attention to the following points:

1. There is considerable independent evidence, both experimental and from speech errors, that speakers have internalized knowledge about the morphological structure of words. (See for example Stanners et al. (1979); Cutler (1980); Fay (in press).)

2. Some recent evidence suggests that affixed words may be listed in the mental lexicon both in their base form and in their full form (a relevant experiment is Stanners et al. (1979); see Cutler (in press) for a review).

Finally, let us consider whether Hurford's data are, as he claims, evidence against the left-to-right phonological organization of entries in the mental lexicon. In fact, it is quite possible to construct a hypothetical model in which left-to-right lexical arrangement is preserved, but in which malapropisms could nevertheless exhibit similarity to their targets at all points. Consider a system of the following sort: at the point at which the production device terminates its semantically determined search for a word, it finds, not the actual lexical entry for the word, but rather its address. This address is an n-place expression in which each place can have m possible values (where m equals, perhaps, the number of phonemes in English), specifying a location in a list of words arranged strictly by left-to-right phonological order. The address is not arbitrary, but is instead a direct function of the phonological structure of the word. Malapropisms arise when the production device makes a mistake in reading off the address and proceeds to an erroneous location. This system could account for the characteristics of malapropisms in the following ways:

1. The system is extremely efficient. It rarely makes mistakes, and when it does err, it generally errs on only one place—getting the whole address mixed up is something that happens only in cases of extreme pathology. (Random word substitutions, in which the error has no semantic, phonological, or contextual relation to the target, hardly ever happen in normal speakers, but do occur in certain cases of language disturbance.)

2. The address is read from left to right, and for some reason, perhaps as a result of increasing memory load, the probability of error increases the further one goes into the word; that is, the beginning is highly likely to be right, and if an error occurs it is more likely to occur towards the end of the word. (Although malapropisms and targets are alike all through the word, they are significantly *more* alike in the first half than in the second half of the word. The reader is invited to check this by constructing a

phonological representation of the malapropisms and their targets (listed in Fay and Cutler (1977)) and dividing each representation arbitrarily at the midpoint. The number of matched segments in comparison with mismatched segments before the midpoint is significantly greater than after the midpoint—$p < .001$.)

3. Although errors are most likely to occur towards the end of words, it can happen that an earlier part of the address is misread. In this case, since the numerical value of the address is not arbitrary, but directly reflects the word's phonological structure, the error will be a word which sounds like the target in all respects but the beginning (*spell* for *tell*) or the middle (*onion* for *oven*).

4. Again, since the value of an address is not arbitrary, many possible addresses will correspond to unfilled slots in the list. When the production device errs on one place of the target address, it may well proceed to an address which does not correspond to an existing word. The procedure in this instance would be to pick the nearest word to the erroneous address. This could result in errors which are different from the target at two points (*miraculous* for *spectacular*).

Of course, this model is in many ways ad hoc; the point of presenting it here is that it embodies left-to-right listing and yet is compatible with Hurford's results. The fact that such a model can be constructed implies that nothing in Hurford's results would force one to abandon the left-to-right hypothesis.

Thus, we have shown in these comments that none of Hurford's three conclusions follows from his reanalysis of the malapropism data. However, although the implications of his data are not as he suggests, Hurford has provided a valuable demonstration that the malapropism–target pairs are in fact more alike than Fay and Cutler originally claimed. In doing this, he has strengthened our original conclusion that the mental lexicon has a primary organization in terms of phonological properties; and, as we originally argued, this phonologically organized lexicon is the one and only lexicon available to the language comprehension and production devices.

References

Cutler, A. (1980) "Errors of Stress and Intonation," in V. A. Fromkin, ed., *Errors in Linguistic Performance: Slips of the Tongue, Ear, Pen, and Hand*, Academic Press, New York.

Cutler, A. (in press) "Lexical Complexity and Sentence Processing," in G. B. Flores d'Arcais and R. J. Jarvella, eds., *The Process of Language Understanding*, Wiley, Chichester, Sussex.

Fay, D. A. (in press) "The Mental Representation of Prefixed Words: Evidence from Prefix Errors in Spontaneous Speech," *Journal of Verbal Learning and Verbal Behavior*.

Fay, D. A. and A. Cutler (1977) "Malapropisms and the Structure of the Mental Lexicon," *Linguistic Inquiry* 8, 505–520.

Hurford, J. R. (1981) "Malapropisms, Left-to-Right Listing, and Lexicalism," *Linguistic Inquiry* 12, 419–423.

McNemar, Q. (1969) *Psychological Statistics*, fourth edition, Wiley, New York.

Siegel, S. (1956) *Nonparametric Statistics*, McGraw-Hill, New York.

Stanners, R. F., J. J. Neiser, W. P. Hernon, and R. Hall (1979) "Memory Representation for Morphologically Related Words," *Journal of Verbal Learning and Verbal Behavior* 18, 399–412.

(Cutler)
Laboratory of Experimental Psychology
University of Sussex
Brighton, BN1 9QG, U.K.

(Fay)
Bell Laboratories
Naperville, Illinois 60566

On the Difference between *Eating* and *Eating Something:* Activities versus Accomplishments

Anita Mittwoch

1. Classical TG derived (1) from (2) by a deletion transformation (Katz and Postal (1964, 81ff.)).

(1) John ate.
(2) John ate something.

Bresnan (1978) treats (1) as syntactically intransitive but functionally transitive and captures the semantic relationship between (1) and (2) by a lexical mapping rule:

(3) *eat*: V, [____ NP], NP_1 EAT NP_2
[____], $\exists y$, NP_1 EAT y

Fodor and Fodor (1980) point out that *eat* and *eat something* are not equivalent in combination with other quantifiers or in opaque contexts (the same is true for other verbs that behave like *eat*):

(4) a. Everybody ate something.
b. Everybody ate.
(5) a. Bill believes that John ate something.
b. Bill believes that John ate.

In (4a) and (5a) *something* can have narrow or wide scope but in (4b) and (5b) the implicit pronoun can have only the narrow scope reading. Fodor and Fodor accordingly propose that the inference from (1) to (2) should be captured neither on the syntactic nor on the functional level, on both of which *eat* without an object would be intransitive, and that

I wish to thank John Lyons and an anonymous reader for comments on an earlier version of this remark, and my colleagues Mark Stein and Joseph Taglicht for many fruitful suggestions and for painstaking criticism.

Linguistic Inquiry, Volume 13, Number 1, Winter 1982
0024-3892/82/010113–10 $02.50/0