518

JOURNAL OF MEMORY AND LANGUAGE 24, 717–733 (1985)

The Roles of Phoneme Frequency, Similarity, and Availability in the Experimental Elicitation of Speech Errors

ANDREA G. LEVITT

Wellesley College, Wellesley, Massachusetts, and Haskins Laboratories, New Haven, Connecticut

AND

ALICE F. HEALY

University of Colorado, Boulder

In two experiments subjects read aloud pairs of nonsense syllables rapidly presented on a display screen or repeated the same syllables presented auditorily. The error patterns in both experiments showed significant asymmetry, thus lending support to explanations of the error generation process that consider certain phonemes to be "stronger" than others. Further error analyses revealed substantial effects of phoneme frequency in the language and effects of phoneme similarity, which depended on the feature system used to index similarity. Phoneme availability (the requirement that an intruding phoneme be part of the currently presented stimulus) was also important but not essential. We argue that the experimental elicitation of errors provides critical tests of hypotheses generated by the analysis of naturally occurring speech errors.

Recent interest in speech errors has focused largely on the evidence such errors provide about levels of linguistic analysis and psychological models of the speech production process. For example, Fromkin (1971), basing her analysis on a corpus of naturally occurring speech errors, found evidence in support of the independence of various levels of linguistic analysis, including both phonemes and phonetic features. On the other hand, Garrett (1980), also basing his analysis on spontaneous-error collections, examined speech error distributions for the constraints they provide about a model of sentence production.

The development of experimental techniques for the elicitation of speech errors (see, for example, Motley & Baars, 1976) provides a new source of data, which, when used in conjunction with the evidence from naturally occurring errors, greatly facilitates the modeling of speech error generation. As Fowler (1983) points out, the experimental elicitation of speech errors permits tape recording of subjects' responses so that errors are less likely to be misheard or overlooked. Furthermore, experimental elicitation provides more thorough tests of hypotheses generated by the analysis of spontaneous error collections, especially

This research was supported in part by NICHD Grant HD01994 and BRS Grant RR05596 to Haskins Laboratories, by NSF Grant BNS 80-00263 to Yale University, and by NSF Grant BNS 80-25020 to the Institute of Cognitive Science, University of Colorado, Boulder. The experimental work was conducted while the second author was supported by a Senior Faculty Fellowship from Yale University. This work was completed while the first author was a NATO postdoctoral fellow at the Centre d'étude des processus cognitifs et du langage, Ecole des hautes études en sciences sociales, Paris, France. The authors are grateful to Mary LaRue for her help in testing subjects, transcribing tapes, and tabulating the data for Experiment 1, to Michele Noe for help in testing subjects and transcribing tapes for Experiment 2, to Harriet Magen for help in transcribing tapes and tabulating data for Experiment 2, and to Tony Hatoun for help in analyzing the results of both experiments. We also thank Bruno Repp and Stefanie Shattuck-Hufnagel for helpful comments on an earlier version of this manuscript. Requests for reprints should be sent to Andrea G. Levitt, Haskins Laboratories, 270 Crown Street, New Haven, CT 06511-6695.

0749-596X/85 $3.00

when portions of the error pattern in the naturally occurring corpus are based on relatively few examples. On the other hand, there is always the danger of introducing influences in the laboratory that do not apply in more natural settings.

Shattuck-Hufnagel and Klatt (1979) analyzed collections of naturally occurring segment substitution errors and contrasted two types of error generation explanations. In the case of the first type of explanation, it is assumed that some segments are "strong" whereas others are "weak." Strong segments might be those that occur more frequently in the language, are acquired earlier, are unmarked in phonological theory, or are easier to articulate. The precise definition of segment strength is less important than the role strong segments play. Each segment substitution error has an intended, or target, segment source for the intruding error. The explanation predicts that strong segments appear more often as intrusions, whereas weak segments appear more often as targets in segmental substitution errors. A confusion matrix of such speech errors should thus be asymmetrical. This asymmetry would reflect the pattern of strength versus weakness of the segments involved.

In the case of the second type of explanation, on the other hand, the tendency of one segment (y) to substitute for another segment (x) would be related to their degree of similarity, but substitutions of x to y and y to x would be equally frequent. A confusion matrix of speech errors, if such errors arose as predicted by this type of explanation, should thus be symmetrical.

Shattuck-Hufnagel and Klatt (1979) analyzed the confusion matrix generated by 1620 substitution errors. The matrix proved to be asymmetrical. However, further analysis revealed that the asymmetry was due almost exclusively to four consonant segments /s, š, č, t/, such that errors of the type /s/ to /š/, /s/ to /č/, and /t/ to /č/ were all more frequent, respectively, than /š/ to /s/, /č/ to /s/, and /č/ to /t/. Once this source of asymmetry was removed, the confusion matrix of segmental errors was no longer significantly asymmetrical. However, the pattern of errors for /s, š, č, t/, which contributed most to the asymmetry of the matrix, could not be accounted for by stronger segments intruding more often, since, according to Shattuck-Hufnagel and Klatt, /š/ and /č/, for example, are less frequent and acquired later than /s/ (i.e., they are weaker), yet they intruded more often.

Shattuck-Hufnagel and Klatt proposed to account for the asymmetrical pattern of their confusion matrix in terms of a palatalization mechanism. They checked their corpus for factors that might "palatalize" the pronunciation of a nonpalatal consonant (e.g., /s/ becoming /š/), but no difference was found between the source consonant environments in which palatalizing and nonpalatalizing errors occurred. When the vowel environments of the target utterances were examined, Shattuck-Hufnagel and Klatt found that a palatalizing error occurred proportionately more often before a high vowel (e.g., /i/), but that this difference was not statistically significant. However, their calculations were based on a relatively small number of observations. The effect of the following vowel might indeed be reliable given a larger number of observations.

The authors concluded that the evidence from their data suggests that errors arise during the speech production process when one of two simultaneously available segments is mis-selected for a slot in an utterance, with the two segments generally being equally likely to be mis-selected.

Notice, however, that an explanation assuming that phonemes are not equal in strength, in particular one for which a strong segment is defined as a more frequent segment in the language,[1] does not

---

[1] Motley and Baars (1975) found experimental evidence that consonant frequency in initial position affects the tendency of initial consonants in pairs of CVC nonsense words to interchange. Hence, frequency in the language seems like an appropriate initial index of phoneme strength.

receive a fair test in a corpus of naturally collected errors, because the prior probabilities of occurrence for all the segments are not equal. Imagine an explanation of the error generation process according to which segment strength is defined by segment frequency and similar segments are likely to substitute for one another. Such an explanation would predict that the rate that a frequent segment would be mispronounced given that it was intended would be lower than the rate that an infrequent segment would be mispronounced given that it was intended. So, for example, for /s/ and /š/, similar segments that might easily be confused, with /s/ as the stronger because it is more frequent, the rate of /s/ being mispronounced given that it was intended should be lower than the rate of /š/ being mispronounced. But the collection of naturally generated speech errors reflects the frequency of occurrence of phonemes in English, not just the error rates given that the phonemes are intended. Thus, since /s/ is much more frequent in the language than /š/, it will occur much more often as an intended phoneme, so that it will occur more frequently as a target than /š/, even if its rate of occurrence as a target given that it was intended is lower. Furthermore, /š/, which is likely to substitute for /s/ because it is very similar, will appear more often as an intrusion than as a target, because of the high prior probability or frequency of /s/ as an intended phoneme. Note that the asymmetry arises because of the segmental similarity of /s/ and /š/ *and* a great discrepancy in their relative frequencies of occurrence in English. An experimental elicitation of errors using these segments in source utterances provides a good way of avoiding the problem of unequal frequencies of occurrence, because in the experimental situation, the intended utterances can be assigned equal prior probabilities. If frequency contributes to segment strength and if strength is a factor in the error generation process, then /s/ should appear more often as an intrusion and /š/ more often as a target, in the controlled experimental situation.

Intuitively, /s/ and /š/ seem quite similar, but similarity between two segments has not been clearly defined in the speech error context, although several investigators (Nooteboom, 1969; MacKay, 1970; Fromkin, 1971) have discussed the role of features in the error generation process. One way of defining segment similarity might be on the basis of the number of shared features. Clearly, the choice of a particular feature system can be crucial. Given a particular feature system, segments might need to share all or almost all features and only differ on some single individual feature (e.g., anterior or high) or type of feature (e.g., features for place of articulation) for errors to occur frequently. The role of segment similarity can be assessed in two ways: (1) Does the similarity of two segments in an utterance affect the tendency of subjects to make errors on those segments and (2) given that an error has occurred, how similar is the intruding phoneme to its intended target?

Another issue is whether it is necessary for the target and intruding segments to be simultaneously available for a substitution error to occur that involves them. In a very broad view, the availability of a segment as an error source should be a function of its frequency in the language. A narrower view might define segment availability such that the source of an error need occur within a relatively constrained portion of the intended utterance. One could assess this narrower view of availability experimentally by seeing whether substitutions of y for x are more likely to occur when y is part of the stimulus.

Finally, it may be that Shattuck-Hufnagel and Klatt's observed asymmetry involving /s, š, č, t/ does reflect a palatalizing mechanism but there were insufficient observations in the environment of high vowels or palatal consonants. Again, the experimental situation permits a direct test of this hypothesis.

The basic technique for the experimental elicitation of speech errors involves what Baars (1980) calls the "completing plans framework." Essentially, the subject is given two alternative plans for the production of an utterance and is required to make a rapid response. For example, the subject might see the series of word pairs "give book, go back, get boot, bad goof" flashed rapidly on a screen. Notice that the fourth word pair, the test pair, "bad goof" involves a reversal of the initial consonant pattern found in the first three pairs, the bias pairs. After the test pair, at the sound of a buzzer, the subject would be expected to say the now-occluded final pair as quickly as possible. Under these conditions, a number of subjects will produce a speech error and may even spoonerize the test pair, reversing the initial consonants, and say "gad boof" instead.

We adapted this basic technique for the purposes of our study. Since previous work (Baars, Motley, & MacKay, 1975) has shown that there is output monitoring for the lexical status of spoonerized words (e.g., that "gad boof," which contains two nonlexical items, will occur less often as an error for "bad goof" than "darn bore," which contains two lexical items, will occur as an error for "barn door" in a similar sequence), we chose pairs of nonsense CV syllables as stimuli.[2] In pilot work, we found that subjects tended to make a greater number of errors when they were asked to pronounce both the bias and test items than when they pronounced only the test items. Hence, we required subjects to pronounce all of the items flashed before them on a screen.[3] Furthermore, pilot work indicated that when the bias pairs had a consistent vowel pattern (e.g., compare the bias series "right lean, ripe leap, ride leak" with the one given above), more errors tended to occur than when the vowel pattern was inconsistent (see also Dell, 1984). Thus, we restricted our bias pairs to those with consistent vowel patterns. We created our CV stimuli from the four consonants in Shattuck-Hufnagel and Klatt's data base that had been responsible for the initial asymmetry /s, š, č, t/, plus the additional consonant phoneme /θ/. The addition of /θ/ allowed us to test whether similarity, defined as a single feature difference, depends on a specific feature, since the consonants in the pairs /š, č/ and /θ, t/ differ on the single feature *continuant,* according to Chomsky and Halle (1968), whereas the consonants in the pair /s, θ/ differ on the single feature *strident.* The consonant /θ/ also provides another relatively infrequent, but nonpalatal phoneme to test against the infrequent palatal set /š, č/. We chose the vowels /a, i, u/ for the test set, so as to be able to assess whether vowel height, high /i, u/ versus low /a/, or vowel height and frontness, front high /i/ versus /a, u/, might be the possible source of palatalizing errors.

EXPERIMENT 1

In Experiment 1, pairs of CV nonsense stimuli were presented visually, and subjects were asked to read all presented items as rapidly as possible.

Method

Materials. Using the set of consonant phonemes /s, š, č, t, θ/, written as *s, sh, ch, t,* and *th,* respectively, and the set of vowels /a, i, u/, we constructed pairs of CV nonsense syllables. Since we eliminated pairs with matched consonants (e.g., *ta ti*) as well as those with matched vowels (e.g., *sa*

---

[2] Although none of the CV nonsense pairs represented common lexical items as visually presented, six of them did represent common lexical items as pronounced: *si* = "see"; *shi* = "she"; *ti* = "tea"; *su* = "sue"; *shu* = "shoe"; *tu* = "two."

[3] It is possible that this rapid reading procedure is influenced by articulatory interference of the type involved in tongue twisters as well as by the factors producing higher-level slips of the tongue. However, Cohen (1973) found that the pattern of speech errors induced via a rapid reading procedure was of a very similar nature to that of a naturally collected corpus.

ta), there were 20 possible consonant permutations and 6 possible vowel permutations for a total of 120 test stimuli. A set of 120 filler pairs of CV nonsense syllables was analogously constructed using another set of consonants /r, l, b, v, m/ and the same set of vowels /a, i, u/.

Design. Each of the 120 test stimuli was preceded by three identical bias pairs of nonsense syllables which were constructed analogously to the test CV pair set and in which the order of the vowels was preserved but that of the consonants was switched. For example, for the test stimulus *su ti,* the presentation order was *tu si, tu si, tu si, su ti*. In order to prevent subjects from anticipating a switch after three identical CV nonsense pairs, 30 of the test CV nonsense syllables were also presented as distractors in groups of four (e.g., *tu si, tu si, tu si, tu si*), 30 in groups of three, 30 in pairs, and 30 singly. The 120 filler CV nonsense pairs also served to divert subjects' attention from the test stimuli consonants and pattern of presentation. Thirty of the filler CV nonsense syllables were presented in groups of four, 30 in groups of three, 30 in groups of two, and 30 singly. For half the trials with the filler syllables, the last item preserved the consonant order (e.g., *ra li, ra li, ra li, ra li*) and for half the trials the last item reversed the consonant order (e.g., *ra li, ra li, ra li, la ri*). The presentation of the test stimuli, distractors, and filler sequences was in pseudorandom order with the constraint that there were four test sequences, four filler sequences, and four distractor sequences in every block of twelve sequences. There was a total of 1080 pairs of CV nonsense syllables presented to subjects.

Subjects. Thirteen men and women participated in the experiment. Four were volunteers from the Haskins Laboratories staff (who were relatively knowledgeable phonetically), and nine were Yale University undergraduates receiving course credit for their participation. [Five additional subjects (one volunteer and four students) were tested, but their data were not analyzed because they failed to read a substantial number of the syllable pairs, and it was often not possible to determine what syllable pair they were responding to when they did utter something.]

Apparatus and procedure. The pairs of CV nonsense syllables were projected under program control onto the self-refreshing screen of a Decgraphic 11 GT-40 computer terminal hooked up to a PDP 11/45 computer at the rate of two syllable pairs a second. Subjects were asked to pronounce each syllable pair aloud as accurately as possible. During this task, subjects listened to white noise presented over Grason-Stadler TDH 39-300Z headphones in order to encourage them to speak up as loudly as possible and to minimize their ability to monitor their own utterances. Subjects' responses to the stimuli were recorded via a Sony F-27S microphone onto a Sony cassette tape recorder model TC-110B for later analysis.

Subjects were told that the nonsense syllables they would see would be composed of three vowel sounds, spelled as *i, a,* and *u*. They were instructed to pronounce the letter *i* as /i/ as in the word *eat,* the letter *a* as /a/ as in the word *father,* and the letter *u* as /u/ as in the word *boot*. They were also told to pronounce the letter pair *th* as in the word *think, sh* as in *shoe,* and *ch* as in *church*. Subjects were then shown CV nonsense syllable pairs typewritten on a sheet of paper and asked to read them aloud. Their pronunciation was checked, and if they did not pronounce the letters as instructed, they were asked to do so. There were 29 CV nonsense pairs from the filler set presented first to subjects as practice with the computer apparatus.

Results

Subjects' responses to all 1080 CV stimulus pairs were transcribed by one listener and then checked by another. Across the 13 subjects, there were 185 disagreements (1.3%), which were resolved by relistening to the disputed pairs until a consensus was reached. A response was scored as an error if the pair deviated in any way from the stimulus; thus, null responses were scored as errors. The results for the 120 test stimuli are summarized in Table 1 in terms of error frequencies as a function of consonant pair and vowel pair.

As is clear from Table 1, the vowel pairs did not have consistent effects on error rates. An analysis of variance was conducted on the error data summed across vowel pairs in order to determine the significant effects due to consonant pairs. Two factors were included in the analysis, one for the 10 different combinations of consonants, and the second to assess the effect of consonant frequency on error rates, such that the first permutation of the consonant pair had the more frequent of the two consonants preceding the less frequent consonant (with frequency determined by Dewey, 1923), and the second permutation had the less frequent consonant preceding the more frequent one, as revealed in the ordering of Table 1. Both main effects were significant. The consonant pairs were significantly different from one another, $F(9,108) = 6.89, p < .0001$, and consonant pairs for which the less frequent consonant preceded the more frequent consonant had a significantly greater number of errors, $F(1,12) = 5.76, p = .0335$. The interaction of consonant pairs and frequency was not significant, $F(9,108) = 1.50$, p = .1560.

Feature analysis 1. A further analysis was performed on the same data in order to test the hypothesis that the number of feature differences between each consonant in a target pair was crucial in determining the error rate. Since there are a variety of competing feature analyses and since the choice of a single feature system could bias our results, we chose to contrast two phonetic

TABLE 1
FEATURE DIFFERENCES SEPARATING CONSONANTS IN A PAIR AND ERROR FREQUENCIES FOR TEST STIMULI IN EXPERIMENT 1 AS A FUNCTION OF CONSONANT PAIR AND VOWEL PAIR

Feature differences			Vowel pair						
C&H	B&G	Consonant pair	ai	ia	au	ua	ui	iu	Total
1	1	sh–ch	7	7	3	9	8	5	39
		ch–sh	7	5	4	8	6	7	37
1	3	t–th	3	3	2	1	3	0	12
		th–t	2	4	2	2	3	1	14
1	1	s–th	3	2	1	2	2	1	11
		th–s	3	3	5	4	3	7	25
2	1	s–sh	1	2	6	2	3	1	15
		sh–s	3	4	5	8	8	7	35
2	2	t–s	4	2	3	1	1	3	14
		s–t	2	3	1	4	0	0	10
3	1	sh–th	6	5	3	5	2	7	28
		th–sh	8	4	2	4	6	10	34
3	2	t–ch	1	2	4	3	2	2	14
		ch–t	2	3	5	2	4	2	18
3	2	s–ch	2	0	7	3	3	3	18
		ch–s	3	6	3	4	1	2	19
4	3	t–sh	3	2	4	2	1	4	16
		sh–t	3	5	0	7	4	1	20
4	2	ch–th	2	5	4	4	3	4	22
		th–ch	5	2	5	6	5	5	28
		Total	70	69	69	81	68	72	429

feature systems: the well-known system devised by Chomsky and Halle (1968), henceforth *C & H,* and another one derived from a corpus of speech errors in English and German by van den Broecke and Goldstein (1980), henceforth *B & G*. First, the consonant pairs were divided into four feature difference classes according to *C & H* (see Table 1), and errors were averaged across consonant pairs in each class. The main effect of feature difference class was not significant, $F(3,36) = 1.09$, $p = .3672$. Furthermore, the error rate did not monotonically increase or decrease with the number of feature differences, and the error rate for the consonant pair *sh-ch* differed greatly from that for *th-t,* though both consonant pairs differ on the same single feature.

Next, the consonants were divided into three feature difference classes according to *B & G* (see Table 1). With this feature set, the main effect of feature difference class was significant, $F(2,24) = 14.22$, $p = .0002$. The mean number of errors per subject for consonant pairs differing on one feature was 2.2, on two features, 1.4, and on three features, 1.2.

Substitution errors. A separate analysis was made of substitution errors, in which the correct consonant in a syllable of a test stimulus was replaced by another consonant in the stimulus set. The resulting confusion matrix is presented in Table 2.

In order to determine whether the relative frequency with which each consonant segment intrudes is the same as the frequency with which it appears as a target, we computed a χ^2 statistic comparing the two distributions and found that they were in fact significantly different from one another, $\chi^2(4) = 69.1$, $p < .01$. One striking discrepancy between the previous study by Shattuck-Hufnagel and Klatt (1980) and ours concerns the asymmetrical pattern of substitutions involving *sh* and *s*. In the earlier study, there were more replacements of *s* by *sh* than vice versa, whereas the opposite was found in the present study. This discrepancy may be attributable in part to visual factors. Perhaps, consonant segments that contain the same letters *(sh/s* and *th/t)* are particularly likely to be confused, especially in the direction of letter deletion. An analysis that eliminates such confusions, by combining the *sh* and *s* segments and the *th* and *t* segments, yields a marginally significant difference between the target and intrusion distribution, $\chi^2(2) = 4.8$, $p < .10$.

TABLE 2
SUBSTITUTION ERRORS IN EXPERIMENT 1 AS A FUNCTION OF TARGET CONSONANT AND INTRUSION CONSONANT

	Target					
Intrusion	T	S	SH	CH	TH	Total
T	—	10	6	5	27	48
S	6	—	54	6	10	76
SH	4	28	—	49	5	86
CH	6	6	26	—	18	56
TH	5	5	10	9	—	29
Total	21	49	96	69	60	295

Frequency analysis. To determine whether the incidence of errors for each target consonant phoneme is related to the log frequency of that segment in English, we computed a Pearson Product-Moment correlation coefficient relating the frequency with which each of the five consonants occurred as a target to its log frequency in English (Dewey, 1923). As expected according to the strength explanation, there was a negative correlation, although it did not reach standard levels of statistical significance $r(3) = -.696$, $p > .10$. A significant negative correlation was found when the frequency analysis of Shattuck-Hufnagel and Klatt (1979) was used instead of that of Dewey (1923), $r(3) = -.887$, $p < .05$. This new frequency analysis, henceforth the content count, was derived from the speech sample of Carterette and Jones (1974) and includes only content words, not function words or common bound morphemes.[4]

---

[4] The rank order of the consonant phonemes by the Dewey (1923) count is *t* > *s* > *sh* > *ch* > *th,* whereas by the content count it is *t* > *s* > *th* > *ch* > *sh*.

A similar analysis was conducted to compare intrusion frequency and log frequency in the language. The correlations in this case were not significant for the Dewey (1923) count, $r(3) = .284, p > .10$, nor for the content count, $r(3) = -.054, p > .10$.

In view of the high correlations for target frequency and despite the low correlations for intrusion frequency, frequency in the language in addition to visual confusions may be a source of the asymmetry in intrusions noted earlier. In order to test this hypothesis, for the 10 consonant pairs (e.g., *ch-t*), we compared how often the more frequent phoneme intruded for the less frequent phoneme (*t* for *ch*) rather than vice versa (*ch* for *t*). For one test we used the Dewey count, which yielded a significant difference, $t(9) = 2.41, p < .05$, and for a second test we used the more recent content count, which was not significant $t(9) < 1$. By both counts, the more frequent phoneme in the pair intruded more often on the average than did the less frequent phoneme, in accord with a strength explanation of speech errors.

Feature analysis 2. A second feature analysis was performed on the substitution data to see whether more substitutions of y for x occur when x and y differ by a single phonetic feature than when they differ by more. For the *C & H* features, the mean number of substitution errors involving a change of one feature was 20, of two features 24, of three features 6, and of four features 9. Clearly though one- and two-feature changes are more frequent than three- and four-feature changes, there is not a monotonic decrease in the number of substitution errors as the number of feature changes increases. Indeed, *sh-ch* and *th-t*, which differ on the same single feature according to *C & H*, show mean substitution rates of 38 and 16, respectively. Furthermore, there are complementary asymmetries in the substitution rates for these two pairs (see Table 2) such that the feature change to [+ continuant] involves fewer errors for the pair *t-th* but more errors for the pair *ch-sh*.

For the *B & G* features, the mean number of substitution errors involving a change of one feature was 23, of two features 8, and of three features 10. Although there is not a perfect monotonic decrease in the number of substitution errors as the number of feature changes increases, it is clear that the single feature substitution errors are most frequent.

Availability analysis. A further analysis was performed on the substitution errors to assess the role of segment availability. We determined the number of times a substitution error of y for x occurred in the environment of y (i.e., how often did the intrusion phoneme /t/ occur for the target phoneme /s/ when the test consonant pair was *t-s* or *s-t*). By comparing that number to the overall number of y for x substitutions, we determined the percentage of times that a substitution occurred when the error was part of the intended utterance (see Table 3). For substitution errors of y for x, y was part of the intended utterance 47.5% of the time. Since x was paired with phonemes other than y three times as often as it was paired with y, the appropriate chance percentage is 25%. Hence, segment availability in the stimulus does seem to influence error rate. However, it clearly is not necessary for the intruding phoneme to be part of the intended utterance, since the majority of the substitutions of y for x occur when y is not part of the intended utterance, defined narrowly here as the test CV nonsense syllable pair.

Furthermore, phoneme frequency seems to influence the importance of availability. When the direction of the substitution error involves a change from a relatively more frequent (strong or +) to a relatively less frequent (weak or −) phoneme (see Table 3), then it is more important that the infrequent segment be available, than when the direction of the substitution involves a change from a relatively weak to a relatively strong phoneme. Thus, by the Dewey count of phoneme frequency, when a change involves strong (+) to weak (−), the weak segment is available 58.1% of the

TABLE 3
RELATIVE FREQUENCY OF TARGET PHONEME (x) AND INTRUDING PHONEME (y) AND PERCENTAGE OF ERRORS OF THE TYPE x CHANGES TO y WHEN y WAS AVAILABLE IN THE STIMULUS IN EXPERIMENT 1

Target phoneme	Intruding phoneme	Relative freq.				Number of x to y errors		
		Dewey		Content				
x	y	x	y	x	y	y available	Total	%
sh	ch	+	−	−	+	13	26	50.0
ch	sh	−	+	+	−	24	49	49.0
t	th	+	−	+	−	2	5	40.0
th	t	−	+	−	+	5	27	18.5
s	th	+	−	+	−	3	5	60.0
th	s	−	+	−	+	5	10	50.0
s	sh	+	−	+	−	15	28	53.6
sh	s	−	+	−	+	16	54	29.6
t	s	+	−	+	−	4	6	66.7
s	t	−	+	−	+	6	10	60.0
sh	th	+	−	−	+	8	10	80.0
th	sh	−	+	+	−	3	5	60.0
t	ch	+	−	+	−	2	6	33.3
ch	t	−	+	−	+	2	5	40.0
s	ch	+	−	+	−	3	6	50.0
ch	s	−	+	−	+	3	6	50.0
t	sh	+	−	+	−	4	4	100.0
sh	t	−	+	−	+	4	6	66.7
ch	th	+	−	−	+	7	9	77.8
th	ch	−	+	+	−	11	18	61.1
				Total		140	295	47.5

time, whereas when the change involves weak (−) to strong (+), the strong segment is available only 41.6% of the time, $t(9) = 3.19$, $p < .05$. The same pattern obtains with the content count (53.8% from strong (+) to weak (−), 42.3% from weak (−) to strong (+)), although the latter set of differences is not significant, $t(9) < 1$.

On the other hand, the availability of the intruding phoneme did not vary regularly with the number of feature differences separating each consonant pair. By the *C & H* feature set, the intruding phoneme was available 42.6% of the time when there was a single feature difference between the consonants in a pair, 41.8% of the time when there were two feature differences, 55.3% of the time when there were three feature differences, and 70.3% of the time when there were four feature differences. Although this pattern suggests the possibility that it is more important that the intruding phoneme be available when consonant pairs differ by three or more features, it is not confirmed in the pattern of availability for the *B & G* features. In that case, the intruding phoneme was available 46.5% of the time when the consonants in a pair differed on a single feature, 57.6% of the time when they differed on two features, but only 35.7% of the time when the consonants differed on three features.

Discussion

The results of Experiment 1 show that the likelihood of an error occurring for a given segment in a test pair depends in part on the relative frequency in English of the individual segments in the pair. Thus, the matrix generated by the substitution errors showed significant asymmetry. There was a high negative correlation between the frequency of an error occurring for a target segment and its log frequency of occurships in speech errors. Indeed, van den Broecke and Goldstein (1980) compared a number of feature systems, along with the one they devised on the basis of English and German speech errors, and found that "feature systems designed without incorporating evidence from speech errors are all capable of showing meaningful structure in phonological speech errors as they occur" (p. 63). Nonetheless, segment similarity emerges as a significant effect in our data only when we use the *B & G* features to determine segment similarity. That the segment similarity effects in our data are best demonstrated by the *B & G* features, derived from the analysis of naturally occurring speech errors in English and German, suggests that the errors we find in our experimental situation are analogous to those occurring in collections of naturally occurring utterances.

Availability. When naturally occurring speech errors are analyzed, the assumption is often made that errors are most likely to occur when similar segments are simultaneously available. Yet the results of our experiments suggest that availability, here defined in narrow terms as a substitution of x for y when x is part of the stimulus, is important but not necessary, since the percentage of the x for y substitutions in both experiments that occur when x is part of the stimulus is substantially greater than the chance value but no greater than 50%. Indeed, the substitution errors in the corpus examined by Shattuck-Hufnagel and Klatt (1979) include 30% with no known source word. It is possible that the actual proportion of naturally occurring speech errors that have no source in the surrounding context might be higher than that estimated by Shattuck-Hufnagel and Klatt, and it might be wrong to assume in such cases that the intruding error was part of the intended utterance (see Harley, 1984, for a discussion of higher level non-plan-internal errors). Finally, we find that segment availability becomes increasingly important as the frequency of the intruded phoneme decreases and perhaps, to a lesser extent, as the featural similarity between the intruded and target phonemes decreases.

However, it is difficult to compare the relative magnitudes of the effects of phoneme frequency and availability (see Sechrest & Yeaton, 1982). Moreover, the influence of phoneme frequency on the importance of availability suggests that both effects may stem from the same activation mechanism. The frequency effect may be reflecting differences in the base activation levels of phonemes, whereas the availability effect may reflect transient increases in phoneme activation that result from being part of the intended utterance.[5]

Conclusions. The results of our two experiments provide support for an explanation of the speech error generation process in which a segment's strength is a function of its frequency of occurrence in English: Weak (or infrequent) segments tend to serve as targets whereas strong (or frequent) segments tend to serve as intrusions. The role of phoneme frequency is a consistently important one. Phoneme availability also plays a role, though perhaps more restricted than expected. Furthermore, availability may be reflecting the same activation mechanism responsible for the frequency effect. Finally, the notion that the segments that interact in speech errors are likely to be similar is best supported by our data when segment similarity is defined in terms of a feature set derived from naturally occurring speech errors.

REFERENCES

ANDERSON, V. A. (1942). *Training the speaking voice*. New York: Oxford Univ. Press.

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

BAARS, B. J., MOTLEY, M. T., & MACKAY, D. G.

---

[5] We are indebted to Marcel Just for making this point.

(1975). Output editing for lexical status in artificially elicited slips of the tongue. *Journal of Verbal Learning and Verbal Behavior,* **14**, 382–391.

Berry, M. L., & Eisenson, J. (1947). *The defective in speech.* New York: Crofts.

Borden, G. J., & Harris, K. S. (1980). *A speech science primer.* Baltimore: Williams & Wilkins.

Carterette, E. C., & Jones, M. H. (1974). *Informal speech.* Berkeley: Univ. of California Press.

Chomsky, N., & Halle, M. (1968). *The sound pattern of English.* New York: Harper & Row.

Cohen, A. (1973). Errors of speech and their implication for understanding the strategy of language users. In V. A. Fromkin (Ed.), *Speech errors as linguistic evidence* (pp. 88–92). The Hague: Mouton.

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 and Cognition,* **10**, 222–233.

Dewey, G. (1923). *Relative frequency of English speech sounds.* Cambridge, MA: Harvard Univ. Press.

Ferguson, C. A. (1978). Fricatives in child language acquisition. In V. Honsa & M. J. Hardman-de-Bautista (Eds.), *Papers on linguistics and child language* (pp. 73–115). The Hague: Mouton.

Fowler, C. A. (1983). Review of V. A. Fromkin (Ed.), *Errors in linguistic performance: Slips of the tongue, ear, pen and hand. Linguistics,* **19**, 819–840.

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

Garrett, M. F. (1980). The limits of accommodation: Arguments for independent processing levels in sentence production. In V. A. Fromkin (Ed.), *Errors in linguistic performance: Slips of the tongue, ear, pen and hand* (pp. 263–271). New York: Academic Press.

Harley, T. A. (1984). A critique of top-down independent levels models of speech production: Evidence from non-plan-internal speech errors. *Cognitive Science,* **8**, 191–219.

Ingram, D., Christensen, L., Veach, S., & Webster, B. (1980). The acquisition of word initial fricatives and affricates in English by children two and six years. In G. H. Yeni-Komshian, J. F. Kavanagh, & C. A. Ferguson (Eds.), *Child psychology* (Vol. 1, pp. 169–192). London: Academic Press.

Lester, L., & Skousen, R. (1974). The phonology of drunkenness. *Papers from the parasession on natural phonology* (pp. 233–239). Chicago: Chicago Linguistics Society.

MacKay, D. (1970). Spoonerisms: The structure of errors in the serial order of speech. *Neuropsychologia* **8**, 323–350.

Moskowitz, A. J. (1973). The acquisition of phonology and syntax. In K. J. J. Hintikka, J. M. E. Moravcsik, & P. Suppes (Eds.), *Approaches to natural language* (pp. 48–84). Boston: Reidel.

Motley, M. T., & Baars, B. J. (1975). Encoding sensitivities to phonological markedness and transitional probability: Evidence from spoonerisms. *Human Communication Research* **1**, 353–361.

Motley, M. T., & Baars, B. J. (1976). Semantic bias effects on the outcomes of verbal slips. *Cognition,* **4**, 177–187.

Nooteboom, S. G. (1969). The tongue slips into patterns. In A. G. Sciarone, A. J. van Essen, & A. A. van Raad (Eds.), Nomen Society, *Leyden studies in linguistics and phonetics* (pp. 114–132). The Hague: Mouton

Norman, D. A. (1981). Categorization of action slips. *Psychological Review,* **88**, 1–15.

Sander, E. K. (1972). When are speech sounds learned? *Journal of Speech and Hearing Disorders,* **37**, 55–63.

Sechrest, L. & Yeaton, W. H. (1982). Magnitudes of experimental effects in social science research. *Evaluation Research,* **6**, 579–600.

Shattuck-Hufnagel, S. (1982). Position of errors in tongue twisters and spontaneous speech: Evidence for two processing mechanisms? (No. 1, pp. 1–8). *Working Papers,* Speech Communications Group, Research Laboratory of Electronics, MIT.

Shattuck-Hufnagel, S., & Klatt, D. H. (1979). The limited use of distinctive features and markedness in speech production: Evidence from speech error data. *Journal of Verbal Learning and Verbal Behavior,* **18**, 41–55.

Shattuck-Hufnagel, S., & Klatt, D. H. (1980). How single phoneme error data rule out two models of error generation. In V. A. Fromkin (Ed.), *Errors in linguistic performance: Slips of the tongue, ear, pen and hand* (pp. 35–46). New York: Academic Press.

van den Broecke, M. P. R., & Goldstein, L. (1980). Consonant features in speech errors. In V. A. Fromkin (Ed.), *Errors in linguistic performance: Slips of the tongue, ear, pen, and hand* (pp. 47–65). New York: Academic Press.

Velleman, S. L. (1983). *Children's production and perception of English voiceless fricatives.* Unpublished doctoral dissertation, University of Texas at Austin.

(Received September 11, 1984)
(Revision received February 5, 1985)