

The reliability of speech error data*

ANNE CUTLER

1. Introduction

Collecting speech errors is enjoyable. For instance, it can give the collector the feeling of doing some useful work while on holiday, at a dinner party, or watching a television interview. And speech error collections are valuable: in the last decade research based on slips of the tongue has provided one of the major components of a long-overdue upsurge in interest in speech production processes, which have otherwise been accorded much less research attention than the processes of comprehension.

The problems associated with speech error research are well known to all in the field. Listening for errors tends to distract the listener's attention from the content of speech, analogously to the way that monitoring for sentence-internal targets reduces the amount of content understood (Johnson, 1980), and to the common experience that we take in little of a text's content when we are proof-reading it. Conversely, of course, attention to content reduces both proof-reading efficiency (Smith and Groat, 1979) as well as the speed with which sentence-internal targets are detected (Green, 1977), and without doubt it also reduces the percentage of slips of the tongue detected and recorded by the error collector. Thus no collector claims to have recorded all slips occurring in a given period of time or a given number of utterances.

If selective attention were the only problem, it could eventually be overcome by a combination of high-fidelity recording and painstaking, multiply-checked transcription. But there is the further problem that some kinds of errors are simply harder to hear than others. Every existing collection of speech errors confounds occurrence of particular types of error with detectability. In fact, it is possible that the detectability problem is so serious that even the most careful transcription of speech will be likely to miss some slips.

Another source of bias in speech error collections is the distributional

characteristics of language. Thus it is not of theoretical interest that errors are reported more often in words of one word class than in words of another if the first class also occurs more often than the second in normal speech. Error collectors are now taking linguistic distributional patterns into account (e.g. Dell and Reich, 1981). In this paper, however, I shall concentrate on the detectability problem. In the following sections I outline the various types of argument which can be made on the basis of speech error evidence, and then summarise evidence which pertains to the question of relative detectability of errors. It will be seen that the detectability problem is by no means an insuperable bar to speech error research; it is possible to identify the confounding factors and control for them when error data are interpreted. Moreover, certain types of speech error argument are completely safe to make, as they are not subject to detectability confounding at all.

2. Types of speech error argument

Speech error data have been used in support of linguistic and psycholinguistic arguments in three basic ways. The first distinct type of argument simply interprets the characteristics of errors which have been reported — ‘some errors are like this, therefore...’. The second type is concerned with the relative frequency of occurrence of particular types of error — ‘more errors are like this than like that, therefore...’. The third type is based on kinds of error which don’t occur — ‘no errors are like this, therefore...’. This three-way classification does not actually reflect a difference in kind between the three categories. For instance, a claim that particular errors don’t occur amounts to a claim that such errors have a zero frequency of occurrence — that is to say, the third type of argument is a special case of the second type. Furthermore, all interpretations of speech error data in terms of rule-governed processes constitute an argument against the ultimate null hypothesis that speech errors occur randomly; ‘Some Errors’ arguments are thus also a kind of ‘More Errors’ argument inasmuch as they claim that random rubbish occurs with insignificant frequency. Thus the categories are by no means logically exclusive. But the three-way division does correspond to a difference in emphasis, as will be seen from the examples of each type of argument given below.

2.1 'Some Errors' arguments

A typical 'Some Errors' argument, i.e. one concerned with the characterisation of occurring errors, is the argument that movement errors exhibit morphological accommodation to their environment (e.g. Garrett, 1976). In (1), for example, two words have exchanged places so that each bears the inflection originally intended for the other, yet the particular form each inflection takes is appropriate for the erroneously inflected word rather than for the originally intended word.

- (1) We had to use to wear hats.
Target: 'We used to have to ...'.

Such errors are used to justify suggestions about relative order of processes in sentence production.

Other 'Some Errors' arguments include the description of blends: in (2), the speaker reported confusing *ton* with *load* to produce the erroneous *toad*:

- (2) We all jumped on him like a toad of bricks.

From this one can argue that at certain points in the production of an utterance more than one plan can be simultaneously entertained — e.g. in selection of words (Garrett, 1980; Butterworth, this volume), or of syntactic structure (Fay, this volume). Similarly, the characteristics of word substitution errors can be invoked to support hypotheses about the word selection process; Fay and Cutler (1977) collected all the word substitution errors they could find, eliminated those that could be explained as resulting from error processes already described in the literature (e.g. semantic errors such as substitution of opposites or members of the same word field; single phoneme movement or substitution errors; blends) and found that the corpus remaining seemed to form a homogeneous class showing considerable similarity of form between error and target. The characteristics of this class of error allowed conclusions to be drawn about the organisation of the mental lexicon used in speech production.

'Environmental contamination' occurs when an unintended word finds its way into the utterance because someone else has just spoken it, or because the speaker happens to be looking at its referent, as in (3), —

- (3) Where would we be without your ribbons?
Target: '.... without your rulers'. Speaker was looking into a drawer containing typewriter ribbons.

Such errors can be cited as evidence that the processing systems of speech

production are not entirely independent of other processing systems having demands on our attention (e.g. Garrett, 1980), and this again constitutes a 'Some Errors' argument.

Finally, all arguments about 'psychological reality' are of the 'Some Errors' type. Thus the fact that single phonemes participate in errors as separate units has been seen as evidence that utterances are at some level of the production process represented as strings of phonemes; similar arguments have been made about sound features, morphemes, words and syntactic constituents (e.g. Fromkin, 1971).

2.2 'More Errors' arguments

Arguments based on the relative frequency of errors with particular characteristics are quite common in the error literature. For instance, it is a truism that error collections contain more anticipations — particularly of single sounds — than perseverations and transpositions. At least one error researcher has suggested (Nootheboom, 1969) that this indicates that 'the speaker's attention is normally directed towards the future', although the same writer also points out (Nootheboom, 1980; see also Meringer and Mayer, 1895 and many other authors for the same observation) that if a transposition is detected and corrected by the speaker when only the first erroneous segment has been uttered, it will be indistinguishable from an anticipation. This phenomenon may well have artificially inflated the frequency of anticipations in speech error collections.

Similarly, it has often been observed that stressed words are disproportionately represented in speech errors (Boomer and Laver, 1968; MacKay, 1969; Nootheboom, 1969). On the basis of this it has been claimed (Boomer and Laver, 1968; MacKay, 1969) that stressed words are not only more prominent in the acoustic form of the utterance, but also in its pre-output mental representation. Other arguments involving prosody — both errors of prosody and the prosody of errors — have also been of the 'More Errors' type. For example, errors of word stress significantly more often result in stress falling on a syllable which bears stress in a morphological relative of the intended word, as in (4):

- (4) For linguists, for linguists to judge...

in which the interference is presumably from *linguistic*; this has been used as the basis for an argument that words derived from a single base are not stored entirely independently in the mental lexicon (Cutler, 1980a). The prosodic characteristics of syllable omission errors — namely that they result significantly more often than not in errors which are more rhythmic

than the target would have been — has also been cited as evidence of the importance of rhythmicity in the generation of utterances in English (Cutler 1980b).

There have also been a number of 'More Errors' arguments interpreting single phoneme error data (e.g. Shattuck-Hufnagel and Klatt, 1979, 1980; van den Broecke and Goldstein, 1980); from the frequency with which particular phonemes substitute for one another, arguments can be constructed about the representation of utterances in pre-output store, or about the psychological justification for particular phonological descriptions.

2.3 'No Errors' arguments

The most well-known 'No Errors' argument in the speech error literature is embodied in Rulon Wells' First Law of speech errors: 'A slip of the tongue is practically always a phonetically possible noise' (Wells, 1951). Boomer and Laver (1968) and Fromkin (1971), among others, have also noted that errors which produce sequences disallowed by the phonological rules of the language in question are almost completely absent from their collections. (Not ENTIRELY absent: even Wells noted that a few exceptions had been recorded. An example noted by the present writer is recorded by Crompton (this volume).) Typically this has been interpreted as evidence that speech production is internally monitored and 'impossible' output filtered out before it is ever actually produced — although Hockett (1967) does suggest that hearers may treat phonologically deviant utterances as non-deviant (i.e. refuse to believe the unprecedented evidence of their ears).

'No Errors' arguments can sometimes be born of 'More Errors' arguments. Thus the observation that lexical stress errors arise from confusion between morphological relatives with different stressed syllables has led to the prediction that particular errors will not occur: that *administrative* is possible, and *administration*, but not *administrative*, for example; or that a stress error on *window*, which has no morphological relatives, will not occur (Cutler and Isard, 1980). Similarly, Garrett's observations that open class (lexical) words frequently exchange places in the utterance, often 'stranding' their inflections behind them, whereas in his data the inflections themselves do not exchange places leaving the lexical items in the intended position, has led him to postulate a model of speech production in which lexical items and bound morphemes have fundamentally different status and in which only lexical items CAN swap places; by implication, this amounts to a prediction (Garrett, 1980) that

exchanges between inflections never WILL occur. That is, Garrett's model accounts for errors like (5), but predicts that (6) is impossible:

- (5) Take the freezes out of the steaker.
Target: 'Take the steaks out of the freezer'. From Fromkin (1973).
- (6) Take the steaker out of the freezes.

3. Potential confounding factors

This section will review factors which could influence the detectability of particular kinds of speech errors and hence the content of error collections. The four headings under which the evidence is grouped represent relatively separate lines of research rather than truly independent sources of confounding factors.

3.1 Slips of the ear

Hearing errors are attested much less often than speech errors because it is necessary for the hearer to admit to having made the slip; but they are nevertheless not uncommon in everyday life. Examples (7)–(11) are typical:

- (7) On the eve of the motor show she'll officially open tomorrow....
Perceived: 'On the eve of the motor show Sheila Fishley open...'
- (8) Because they can answer inferential questions
Perceived: 'Because they can answer in French....'
- (9) Do you know about reflexes?
Perceived: 'Do you know about Reith lectures?'
- (10) It's about time Robert May was here.
Perceived: 'It's about time to drop my brassiere'.
- (11) If you think you have any clips of the type shown...
Perceived: 'If you think you have an eclipse...'

Garnes and Bond (1975; 1980), who analysed hundreds of hearing errors collected in the course of ordinary conversations, reported that more often than not (a) the stress pattern of the utterance is correctly perceived; (b) the vowel in a stressed syllable is correctly perceived; and (c) the error does not cross a phrase boundary. The misperceived segments are more often consonants than vowels, and are usually unstressed syllables, particularly in the middle of a word (Browman, 1980); changes in the rhythmic pattern usually involve only the mislocation of a single

unstressed syllable. Thus in (10) above the primary stress in the utterance fell on the final word; *about* and *Robert* were also stressed; *time* and *May* were unreduced; *it's* and *was* were reduced. In the misperceived sequence the stressed vowels in *Robert* and *here* have been preserved (US pronunciation of *brassiere* applies [bræzi:r]), and the sequence of stressed and unreduced syllables has likewise been preserved. Consonants, however, have been misconstrued, omitted or added in the erroneous reconstruction of the utterance, and although the number of syllables has been preserved, one reduced syllable has migrated.

Not all hearing errors are as complicated as (10); (9), for instance, consists only in the misperception of two fricatives ([f] and [s], heard as [θ] and [ç] respectively), in (8) a reduced vowel has been overlooked, and in (11) the unstressed but unreduced vowel [ɛ] has been misperceived as reduced, precipitating a misplacement of the word boundary. Sometimes the error consists entirely in misplacement of the word boundary, as in (7). Purely syntactic misperceptions, in which the error consists solely in assigning the wrong syntactic structure to the utterance (e.g. (12)–(14) below) also occur, although these are rarely reported; errors in which misparsings are precipitated when a word is mistaken for its homonym, as in (15) and (16), are reported somewhat more often.

- (12) You never actually see a forklift truck, let alone person.
Perceiver attempted to access a compound noun 'forklift person', as if a second occurrence of 'forklift' has been deleted.
- (13) This result was recently replicated by someone at the University of Minnesota in children.
Perceiver assigned NP status to 'the University of Minnesota in children' (cf. 'the University of California in Berkeley').
- (14) Mr Milne came to Rothsay to impress upon this pretty leftwing gathering...
Perceiver understood 'pretty' as adjective rather than adverb.
- (15) Stretching would initiate a change.
Perceiver understood 'stretching' as verb rather than noun, and 'would' — which was contrastively stressed — as 'wood'.
- (16) One thing that Mark's formulation did...
Perceiver parsed 'One thing that marks [Formulation Did]_{NP}'.

It is repeatedly stressed by those who describe hearing errors that this evidence shows speech perception to be an active interpretative process, rather than simply passive reception of the incoming signal. Listeners strive to make the best sense they can of the speaker's message, reconstructing sounds, words and syntactic structure from the incomplete information they have received. Such reconstruction can also be de-

monstrated experimentally. Warren (1970), for example, replaced single sounds in an utterance by a brief burst of white noise, and found that his listeners reported hearing a cough-like sound occurring SIMULTANEOUSLY WITH the speech, rather than INSTEAD OF a portion of it. Listeners are obviously very efficient at constructing a meaningful message from heard speech, even in defiance of the acoustic information.

What is the relevance of this evidence for speech error collectors? Firstly, it must be noted that there is clear evidence that speech errors can precipitate hearing errors. Some of the misparsing examples above, for instance, seem to have been prompted by slightly unusual or ambiguous prosody, or marginally deviant syntax, chosen by the speaker — contrastive stress on the auxiliary in (15), deaccenting of *forklift* in (12) together with omission of the indefinite article before *person*, failure to provide intonational marking of the phrase boundary on *Minnesota* in (13), equal stress on *pretty* and *leftwing* in (14). In the following examples, however, an actual mispronunciation has caused an unintended word to be perceived:

- (17) ... and for nurses' memory for a feel — film!
The error consisted in uttering the long vowel [i] instead of its short counterpart; the speaker stopped and corrected before the end of the word. Of the dozen or so people who heard the error, one heard the speaker to say 'field'.
- (18) It's the same right-wrap apperation...
Target was 'operation', error a sound perseveration from 'wrap', and speaker stopped and corrected just before the end of the word, but not before one of his dozen or so hearers had retrieved 'apparatus' and another had retrieved 'apparition'.
- (19) Such a representation is entirely [otijos].
The speaker produced a deviant, but probably intended, pronunciation of the word 'otiose'. Of a hundred or so listeners, one reported hearing 'odious'.

In each of these cases, the perceived word was inappropriate to the context, and the hearing error was therefore noticed and reported. Moreover, two of these three errors involved vowels, and all of them concerned either the first or second phoneme of a word; as we saw above, vowels, and the beginnings of words, are likely to be correctly perceived.

Suppose, however, that the hearer had reconstructed a word appropriate to the context, possibly even the word that the speaker had originally intended. Unless the speaker spontaneously corrected his error, there would be no way for the hearer to know an error had been committed. It is impossible for error collectors to know how often this happens, but it is

not inconceivable that it happens with sufficient frequency to bias the content of speech error — and hearing error — collections. 'More Errors' arguments, where relative frequency of error types is of crucial importance, should therefore be constructed with extreme care. In particular, they should be avoided if there is specific reason to believe that the characteristics of hearing errors could have confounded the data.

The claim, mentioned above, that speech errors more often involve stressed than unstressed syllables is one case in point. This claim looks particularly impressive when it is compared with the relative frequency of occurrence of stressed and unstressed syllables in speech — there are many more unstressed syllables than stressed. Yet hearing errors, as we have seen, are very much more likely to occur on unstressed syllables. There is good reason to believe that even with relative frequency taken into account speech errors on unstressed syllables may be under-represented in the available data. It would therefore seem highly unwise for any error collector to attach a great deal of theoretical significance to the relative frequency with which stressed and unstressed syllables participate in slips of the tongue, without taking statistical account of the differences in relative detectability of sounds in stressed and unstressed syllables as revealed in hearing error studies.

3.2 Shadowing and mispronunciation detection

Several recent experimental studies have presented listeners with speech already containing errors, and have asked them to repeat this speech back as fast as possible (shadow the text), or else to make a response as soon as they hear an error. Typically, the researchers have found that mispronunciations of single sounds are very often missed; this is particularly true if the mispronunciation differs from the intended sound on only a single feature (e.g. /k/ for /t/), and if the mispronunciation is near the end rather than the beginning of the word (Cole, 1973; Marslen-Wilson and Welsh, 1978). Moreover, the more contextually predictable the distorted word, the more quickly are distortions detected (Cole and Jakimik, 1978). (Note that hearing errors are reported to be common on proper nouns (Celce-Murcia, 1980); these are often unfamiliar or unpredictable in context.) Whereas in the mispronunciation detection tasks, the effect of a distortion being overlooked is that it fails to elicit a response from the subject, in the shadowing experiments undetected distortions are restored in the shadowers' output to the form appropriate to the context (Marslen-Wilson, 1975; Marslen-Wilson and Welsh, 1978). Lackner (1980) has looked at higher-level distortions — syntactic errors such as wrong tense, number, or

word-class markings, and semantic errors produced by substitutions of entire words — and has found that at rapid presentation rates these errors, like the phonemic errors, are very often overlooked and corrected.

The relevance of these data to the study of speech errors was brought out most clearly in a study by Cohen (1980) in which the errors inserted in the text presented to listeners for shadowing were constructed in such a way that they mimicked actually reported speech errors. Cohen compared the restoration rates for anticipations versus perseverations, for instance, and found that perseverations were significantly more often overlooked and restored than were anticipations. Similarly, consonant errors were restored more often than vowel errors, and errors in unstressed syllables were restored more often than errors in stressed syllables. These latter two results are, of course, directly in line with the hearing error evidence cited in section 3.1.

Thus these experimental findings, like the hearing error data, suggest factors which might affect the relative detectability of particular types of speech error. They indicate, for instance, that single-phoneme errors are more likely to be overlooked if the error segment differs from the intended segment by only one feature. Lackner's work suggests that at very fast speech rates even gross syntactic and semantic errors might be overlooked. (This suggests that speech rate might have more of an effect on error DETECTION than it has on error COMMISSION; attempts to demonstrate that error rates rise with rate of speech, going back to Meringer (1908: 122), have all, to my knowledge, met with failure.) Finally, Cohen's finding that anticipations are detected more often than perseverations indicates that the higher frequency of anticipations in error collections, besides being confounded with the possibility of incomplete exchanges, could also be an artefact of the relatively greater detectability of anticipations. Again, it would seem imprudent to rest any major theoretical claims on this pattern of occurrence.

3.3 Perceptual confusions

There is a considerable literature on perceptual confusions: the likelihood with which sounds are confused with one another. Typically, such studies involve the presentation of isolated syllables, with or without noise masking (Peterson and Barney, 1952; Miller and Nicely, 1955; Wang and Bilger, 1973); the listeners simply report what they have heard.

For vowels, one finding relevant to error research is that identification is affected by dialect (Peterson and Barney, 1952); analogously, misunderstandings involving vowels are more common when speaker and

listener use different dialects (Celce-Murcia, 1980). In general, Peterson and Barney found high front vowels to be most accurately identified, low back vowels least accurately, with confusions tending to be between adjacent positions in the vowel space.

For consonants, a consistent finding is that some features of sounds are less likely to be mistaken than others. Whether or not a consonant is nasal is highly likely to be perceived correctly, and similarly, whether or not it is voiced. The place of articulation is more likely to be mistaken, as is whether or not the consonant is a fricative. Thus /b/ is more likely to be perceived as /d/, /g/ or /v/ (which differ from it on place of articulation or frication only) than as /m/ or /p/ (which involve a change of nasality and voice respectively).

These findings are relevant to any attempt to interpret the relative frequency of confusions between sounds in speech errors (e.g. Shattuck-Hufnagel and Klatt, 1979, 1980; van den Broecke and Goldstein, 1980). The mispronunciation detection evidence indicates that sound errors involving a change in only one feature are more likely to be overlooked; the perceptual confusion evidence adds to this the suggestion that such changes are particularly likely to be overlooked if the altered feature is one that is relatively easily confused, e.g. place of articulation. Moreover, there is another dimension of this problem which might affect the detectability of certain sound errors, and that is the size of the response set. For instance, given that nasality value is highly likely to be perceived correctly, there is a greater likelihood of the hearer mistakenly hitting on the intended target for a mispronounced nasal consonant than for a mispronounced non-nasal consonant simply because there are many more non-nasal than nasal consonants (in English and in other languages). That is to say, there seems to be a greater likelihood for an /m/ mispronounced as /n/ to be misperceived as /m/ than for a /b/ mispronounced as /d/ to be perceived as /b/. This would show up in speech error collections simply as a greater likelihood for errors to occur in non-nasal consonants (once frequency of occurrence had been controlled for).

Perceptual confusion data also show evidence of response bias, in that some sounds are more likely to be reported than others. Goldstein (1980) has demonstrated that response bias for consonants correlates with lexical frequency (i.e. the number of words containing the sound in question, as opposed to absolute frequency of occurrence) and with phonological naturalness (as measured by the probability with which a particular sound occurs across languages). Goldstein points out that the same response bias does not appear to be at work in speech errors; although the perceptual confusion experiments show an asymmetry of report as a result of bias (e.g. in Wang and Bilger's data /b/ is more likely to be reported as /p/ than

vice versa), no such asymmetries show up in the speech error data — for any pair of sounds, one is as likely to be substituted for the other as vice versa (Shattuck-Hufnagel and Klatt, 1980). The reality may of course be horrendously complex, with the reported symmetry being in fact a function of a bias-determined asymmetry IN THE OPPOSITE DIRECTION FROM THE ASYMMETRY DEMONSTRATED IN PERCEPTION (e.g. in fact /b/ occurs as an error for /p/ more often than vice versa, but because there is a response bias towards reporting /p/, some substitutions of /b/ for /p/ are misperceived as /p/, bringing the total of /b/ for /p/ substitutions down to the /p/ for /b/ substitution level); however there are at present no independent reasons for believing this to be so.

Indeed there are even some indications that the phoneme error data may be relatively uncontaminated by hearing error confounding, since the most frequently reported sound substitutions are those between sounds which are most like one another — exactly the substitutions which the evidence above suggests are most difficult to detect. Moreover, ‘manner-of-articulation and voicing features are significantly more likely to be preserved than a place-of-articulation feature, and the greatest number of single-feature errors involve a change in place of articulation’ (Shattuck-Hufnagel and Klatt, 1979). Place-of-articulation errors are of course the ones which the perceptual confusion findings (as well as the hearing error and mispronunciation detection findings) suggest should be most often overlooked. Nevertheless, studies of phoneme errors should always take account of the possible effect of perceptual confusions.

Finally, it should be noted that the effects of language and dialect-based expectations and of phonological naturalness on perceptual confusions must cast some doubt on the generality of the (‘No Errors’) argument that speech errors virtually never violate the phonotactic constraints of the language in which they are perpetrated. Phonologically deviant errors would on the present evidence be likely to be perceived as segments more probable in the language, so that it is quite possible that such errors occur at least a little more often than they have been reported; certainly it seems that the phonologically sophisticated (e.g. Ladefoged, reported in Fromkin, 1973: 25) are somewhat more likely to detect phonological violations. Butterworth and Whittaker (1980) report that impermissible consonant clusters can quite easily be elicited in a tongue twister task. The strict adherence to phonological constraints would appear therefore to be another argument on which no grandiose theoretical edifice should be erected, unless it is firmly underpinned by comparisons with measures of the relative detectability of phonologically permissible and impermissible slips.

3.4 Relative salience of beginnings and ends of words

There is abundant evidence that the initial portions of words are of crucial importance to word identification. For instance, when subjects are presented with letter strings which can be made into existing words merely by exchanging two adjacent letters, and are required to identify the distorted words, identification is hardest if the letters which have been exchanged are the first two (and easiest if the letters occur in the middle of the word; final letters are of intermediate difficulty) (Bruner and O'Dowd, 1958). Similarly, words can be guessed most speedily and more reliably from their initial fragments (Broerse and Zwaan, 1966; Nootboom, 1981). If word fragments are presented as cues in a recall task, the initial letters comprise the most effective fragment, the final letters the second most effective, and medial letters the least effective (Horowitz *et al.*, 1968); Horowitz *et al.* (1969) demonstrated that this result is independent of word frequency. The initial positions of a compound word are more important to its recognition than the other portions (Taft and Forster, 1976).

People who have a word on the 'tip of their tongue' very often have intuitions about the word's beginning or end, and these intuitions are right more often than guesses about the middle of the word (Brown and McNeill, 1966; Browman, 1978). Brown and McNeill hypothesise that memory storage of words assigns greater weight to the two ends of the words than to the middle, and probably particular weight to the initial portions.

Marslen-Wilson (1978; Marslen-Wilson and Welsh, 1978) has constructed a model of word recognition in which the initial portions of words can be the ONLY necessary cues for recognition. Marslen-Wilson calls the point at which a word becomes unique from other words of the language — scanning the word from left to right — its RECOGNITION POINT, and suggests that in auditory speech perception a word can be recognised at the latest at its recognition point. This model fails to account for why word-final segments should be more salient than word-medial segments, so that it probably over-simplifies the actual word recognition process; however, it can be shown that lexical decision latencies are affected by how far into the word the recognition point occurs (Marslen-Wilson, 1978) — compare *dwindle*, with recognition point on the third segment, with *intestate*, which only parts company with *intestine* on the final segment. Interestingly, the recognition point can be shown to be of relevance in the construction of neologisms (which of course usually consist in the addition of novel endings to stems); although speakers generally prefer to preserve the base word intact in a neologism (Cutler, 1980c), an exception to this rule can be made as long as the segmental

values and relative syllable salience of the base word are maintained up to the base word's recognition point (Cutler, 1981).

In speech errors, there is further evidence that the beginnings of words are particularly important. Form-related word substitution errors resemble their intended targets very strongly in the initial segments (Fay and Cutler, 1977), at least in adult errors (Aitchison and Straf, this volume); although there are similarities at other points of the word (Hurford, 1981), these similarities are significantly weaker than those in the initial portions (Cutler and Fay, 1982). These findings have also been interpreted as evidence that words are stored in the mental lexicon in left-to-right order, and by implication that word recognition proceeds left to right.

The clear implication for speech error collectors is that even a small distortion of initial segments is quite likely to be noticed and reported, whereas changes in later parts of the word — especially in the middle — are much more likely to go undetected. (Recall that hearing errors are also more common in the middles of words.) In fact it has often been noted that error collections contain many more examples of sound errors in initial position than in final position (Cohen, 1966; Garrett, 1980; Goldstein, 1980; van den Broecke and Goldstein, 1980). Once again, this finding may be an artefact of a hearing error pattern determined by differing psychological salience of parts of words, and once again, therefore, it must be accounted a finding on which it would be hazardous to base important theoretical claims without taking the potentially confounding factor into statistical account.

4. Levels of explanation in speech error analysis

A common confusion in the speech error literature arises from a failure to distinguish between the CAUSE of an error's occurrence and the MECHANISM by which it occurs. The two are logically distinct. For example, although the mechanism by which errors of lexical stress arise is as a result of confusion between morphologically related words, a lexical stress error may be more likely to occur if its occurrence will make the utterance easier to produce than it would otherwise have been, e.g. more rhythmical (Cutler, 1980b).

An early example of confusion between cause and mechanism is provided by the dispute in the early years of this century between Sigmund Freud and Rudolf Meringer. From Meringer's collection of speech errors, *Versprechen und Verlesen* (1895), Freud had borrowed a number of examples which he used as illustrations of his arguments in *The Psychopathology of Everyday Life*, first published in 1901. Moreover, he

suggested some tentative explanations which were characteristically 'Freudian' — for instance (20), which Meringer categorised as a syllable perseveration, was explained by Freud in terms of underlying ill-feeling of the speaker towards his boss:

- (20) Ich fordere Sie auf, auf das Wohl unseres Chefs aufzustossen.
(Target: anzustossen. 'I call on you to belch to the health of our chief' instead of '...to drink a toast to the health of our chief'.)

Meringer objected vigorously to this use of his examples, and criticised Freud's explanations on the grounds that the handful of errors which Freud had borrowed from *Versprechen und Verlesen*, in fact all the errors described in *Psychopathology*, obeyed the same rules as the thousands of other errors in Meringer's collection and could therefore be most parsimoniously described in terms of the categorisation (anticipation, perseveration, substitution, exchange, blend etc.) which Meringer had set up (see Meringer, 1923 for these arguments). Freud, on the other hand, felt that Meringer's explanations were simply vastly less interesting than his own. Both of them made the mistake of assuming that their respective explanations were on the same level. In fact, there is no logical reason why the occurrence of an error via one or another mechanism (anticipation, perseveration, etc.), or alternatively the failure of an error to be detected and corrected by internal monitoring systems prior to output, should not be rendered more likely by the fact that the error form is associated with secret desires or thoughts.

This is by no means to say that Freud's explanations of speech errors are necessarily correct. Meringer (1923) also made a number of cogent commonsense criticisms of Freud's theories, and he was neither the first nor the last to do so. The point to be made here is simply that Freud's examples cannot be dismissed by pointing out that they conform to otherwise postulated speech error mechanisms, because Freud's explanations are at a different level from those involved in the postulation of mechanism.

Others have attempted to explain away Freud's examples by classifying them in recognised error categories (Ellis, 1980) or as other known linguistic phenomena (Timpanaro, 1976; see the discussion of this work by Butterworth, this volume). There is, of course, no pressing need for error researchers to explain away Freud's data, since the logical independence of cause and mechanism explanations also means that Freud's speculations about cause have no relevance whatsoever to hypotheses about mechanism; they can simply be ignored, as indeed they largely have been.

The independence of cause and mechanism is tacitly accepted whenever an error researcher points out that error rates increase when speakers are

tired or intoxicated; it is never suggested that fatigue and drunkenness are alternative mechanisms by which errors can occur, merely that such states can precipitate, or cause, the more frequent operation of the existing mechanisms. It is often noticed by error collectors that once one error has been made, other errors seem to follow; while this effect may be solely due to heightened sensitivity in the hearers, a causal explanation could also be constructed by postulating a relaxation of the pre-output monitoring devices, thus letting more errors through.

Similarly, whereas form-related word substitution errors are hypothesised to arise by a totally different mechanism from semantically related word substitution errors (Fay and Cutler, 1977), there are a large number of cases in which form-related substitutions seem to resemble their intended targets in meaning as well (see Aitchison and Straf, this volume). A causal explanation for this phenomenon might be that routine semantic activation of words associated with the intended target occasionally results in a phonologically near neighbour of the target being activated, and that if a neighbour is activated, it is more likely than unactivated neighbours to be chosen by mistake; or alternatively, that such errors might arise independently as semantic and form-related substitutions, and that any error which has two sources is more likely to get through pre-output filters than an error with only one source.

This kind of causal explanation — that an error is more likely if it has more than one source — is in many ways not comparable to Freud's claim that an error is probable if it expresses unconscious mental states (for instance it is testable, which Freud's claims are not). But it is like Freud's suggestions in that it is logically independent of the hypothesised mechanisms involved in error phenomena. Causal explanations can, moreover, differ across languages. For instance, Cutler (1980b) suggested a causal explanation for syllable omission errors which invoked a tendency towards underlying rhythmicity in English utterances. Such an explanation would obviously not hold for languages — such as French — in which there is no tendency towards stress-timed rhythm; nonetheless, the mechanism by which syllable omission errors arise would seem to be the same in French as in English, since errors in both languages conform strictly to constraints of syllable structure (see Crompton, this volume, for a discussion of these constraints). This has led to the suggestion (Cutler 1980d) that whereas *CAUSES* of errors might differ across languages, across individuals, and across occasions, error *MECHANISMS* ought to be both speaker- and language-universal.

5. Conclusion

Not all speech errors are equally detectable; therefore all collections of speech errors assembled from everyday language behaviour are liable to be confounded by the problem of detectability. No collection as yet exists which reliably records every error in a large body of speech. The compilation by Garnham *et al.* (this volume) is a small step in this direction. The tape-recorded corpus of Boomer and Laver (1968), often cited as an example of a complete corpus, was in fact small ('more than a hundred' slips) and compiled 'over a period of several years'; like a similar (already larger) collection of recorded excerpts being amassed by the present author, it was presumably as subject to the detectability problem as any written accumulation.

Evidence from hearing error research, and from experimentation on perceptual confusions, detection of mispronunciations, and relative salience of parts of words, indicates that there are a number of factors which will act to make some slips more detectable than others. There are a few studies which have addressed this question directly. Tent and Clark (1980) studied the detection of phonetic-level and higher-level errors in isolated sentences presented under mild noise masking, and found that phonetic errors were overlooked far more often than were errors involving larger units (syllables, words); among the higher-level errors, anticipations and transpositions were nearly always detected (98% and 97% respectively), while perseverations and blends were detected less often but still in more than 75% of cases. In the phonetic errors, anticipations were detected most often (28%), transpositions nearly as often (26%) and perseverations least often (13%).

The greater detection rate for higher-level errors parallels Cohen's (1966) observation that errors which resulted in an obviously deviant meaning in the particular context were most likely to be detected. Greater detectability by the hearer, however, does not imply also greater detectability for the speaker who perpetrated the slip; Nootboom (1980) looked at self-corrections by speakers and found that phonetic errors were actually corrected slightly more often than word-level errors. Within error types, however, the relative detectability is the same for both speakers and listeners: anticipations (at both sound and word level) were corrected far more often than other errors. Whether or not the speaker spontaneously corrects an error is presumably also a factor influencing whether or not listeners perceive that an error has been committed.

The three possible types of speech error argument differ in the degree to which they are susceptible to the detectability problem. 'Some Errors' arguments, in their strongest form, require only one instance of a

particular error type to make their case. In the present volume, the paper by Meara and Ellis offers a good example of such an argument. The papers by Bierwisch and by Fay are also based on 'Some Errors' arguments. This kind of argument is without question the theoretically safest kind to make on the basis of error data collected from everyday speech.

'More Errors' arguments (of which this volume offers a typical instance in the paper by Aitchison and Straf) do not have to be avoided; but those who construct them must remember that detectability confounds are likely, and must be wary of assigning too great a theoretical importance to any pattern of distribution which independent evidence suggests may be influenced by differential detectability. As the evidence cited in section 3 showed, however, a number of 'More Errors' arguments in the speech error literature are based on error distributions which are quite unlike the distributions which would be predicted simply on the basis of the differential detectability of those errors. The proponents of these arguments therefore have cause for considerable confidence in their findings. Where differential detectability would predict the same result as the distribution of reported errors shows, though, interpretations of error patterns are less immune to criticism; but all is still not lost. What is required, as section 3 suggested, is a numerical estimate of the difference in detectability based on the available evidence, and a comparison of this difference with the difference observed in the error distribution. For instance, Tent and Clark's study found that 28% of phonetic anticipations were detected, but only 13% of phonetic perseverations. Thus for every 28 anticipations that were reported, 72 were missed, while for every 13 perseverations reported, 87 were missed. A comparative study of the relative frequency of anticipations and perseverations could adjust the reported frequencies accordingly, and would thereby control for the differential degree of detectability as operative in Tent and Clark's study.

'No Errors' arguments (with which in the present volume only Crompton briefly flirts) are obviously the least safe, since only one decisive counter-example is required to destroy them; but those who make 'No Errors' arguments have presumably always known this.

Finally, it should be pointed out that speech error research is not limited exclusively to data collected from everyday speech. As Cutler and Fay (1978) and Fowler (this volume) have pointed out, some of the problems inherent in naturalistic collection methods can be overcome by combining this methodology with recently developed laboratory techniques for the elicitation of errors. The reliability problems which confront speech error collectors are by no means insurmountable; speech error research — as the

present volume attests — is in fine health and making a valuable contribution to linguistic and psycholinguistic knowledge.

Centre for Research on
Perception and Cognition
University of Sussex
Falmer
Brighton BN1 9QG
England

Note

- * Brian Butterworth, Steve Isard, John Laver, Dennis Norris and Richard Shillcock read and commented on an earlier version of this paper. They are responsible for a vast number of improvements, but not for the many deficiencies which remain. Financial support was provided by the Science Research Council.

References

- Aitchison, J. and Straf, M. (1981). Lexical storage and retrieval: a developing skill? *Linguistics* 19-7/8, 751-795.
- Bierwisch, M. (1981). Linguistics and language error. (Translated and enlarged version of Bierwisch, 1970). *Linguistics* 19-7/8, 583-626.
- Boomer, D. S. and Laver, J. D. M. (1968). Slips of the tongue. *British Journal of Disorders of Communication* 3, 2-12.
- van den Broecke, M. P. R. and 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*. N. Y., London: Academic Press.
- Broerse, A. C. and Zwaan, E. J. (1966). The information value of initial letters in the identification of words. *Journal of Verbal Learning and Verbal Behavior* 5, 441-446.
- Browman, C. P. (1978). Tip of the tongue and slip of the ear: Implications for language processing. *UCLA Working Papers in Phonetics* 42.
- (1980). Perceptual processing: Evidence from slips of the ear. In V. A. Fromkin (ed.), *Errors in Linguistic Performance: Slips of the Tongue, Ear, Pen, and Hand*. N. Y., London: Academic Press.
- Brown, R. and McNeill, D. (1966). The 'tip-of-the-tongue' phenomenon. *Journal of Verbal Learning and Verbal Behavior* 5, 325-337.
- Bruner, J. S. and O'Dowd, D. (1958). A note on the informativeness of words. *Language and Speech* 1, 98-101.
- Butterworth, B. (1981). Speech errors: old data in search of new theories. *Linguistics* 19-7/8, 627-662.
- Butterworth, B. and Whittaker, S. (1980). Peggy Babcock's relatives. In G. E. Stelmach and J. Requin (eds), *Tutorials in Motor Behaviour*. Amsterdam: North-Holland.
- Celce-Murcia, M. (1980). On Meringer's corpus of 'slips of the ear'. In V. A. Fromkin (ed.), *Errors in Linguistic Performance: Slips of the Tongue, Ear, Pen, and Hand*. N. Y., London: Academic Press.

- Cohen, A. (1966). Errors of speech and their implication for understanding the strategy of language users. *Zeitschrift für Phonetik* 21, 177–181.
- (1980). Correcting of speech errors in a shadowing task. In V. A. Fromkin (ed.), *Errors in Linguistic Performance: Slips of the Tongue, Ear, Pen, and Hand*. N. Y., London: Academic Press.
- Cole, R. A. (1973). Listening for mispronunciations: a measure of what we hear during speech. *Perception and Psychophysics* 11, 153–156.
- Cole, R. A. and Jakimik, J. (1978). Understanding speech: how words are heard. In G. Underwood (ed.), *Strategies of Information Processing*. London: Academic Press.
- Crompton, A. (1981). Syllables and segments in speech production. *Linguistics* 19–7/8, 663–716.
- Cutler, A. (1980a). Errors of stress and intonation. In V. A. Fromkin (ed.), *Errors in Linguistic Performance: Slips of the Tongue, Ear, Pen and Hand*. N. Y., London: Academic Press.
- (1980b). Syllable omission errors and isochrony. In H. W. Dechert and M. Raupach (eds), *Temporal Variables in Speech*. The Hague: Mouton.
- (1980c). Productivity in word formation. *Papers from the Sixteenth Regional Meeting, Chicago Linguistic Society*, 45–51.
- (1980d). La leçon des lapsus. *La Recherche* 11, 686–692.
- (1981). Degrees of transparency in word formation. *Canadian Journal of Linguistics* 26, 73–77.
- Cutler, A. and Fay, D. A. (1978). Introduction. In re-issue of R. Meringer and K. Mayer *Versprechen und Verlesen* (1895), ix–xi. Amsterdam, John Benjamins.
- (1982). One mental lexicon, phonologically arranged: comments on Hurford's comments. *Linguistic Inquiry* 13 107–113.
- Cutler, A. and Isard, S. D. (1980). The production of prosody. In B. Butterworth (ed.), *Language Production*. London: Academic Press.
- Dell, G. S. and Reich, P. A. (1981). Stages in speech production: an analysis of speech error data. *Journal of Verbal Learning and Verbal Behavior* 20 611–629.
- Ellis, A. W. (1980). On the Freudian theory of speech errors. In V. A. Fromkin (ed.), *Errors in Linguistic Performance: Slips of the Tongue, Ear, Pen, and Hand*. N. Y., London: Academic Press.
- Fay, D. A. (1981). Substitutions and splices: a study of sentence blends. *Linguistics* 19–7/8, 717–749.
- Fay, D. A. and Cutler, A. (1977). Malapropisms and the structure of the mental lexicon. *Linguistic Inquiry* 8, 505–520.
- Fowler, C. (1981). Review of V. A. Fromkin (ed.), *Errors in Linguistic Performance: Slips of the Tongue, Ear, Pen and Hand*. *Linguistics* 19–7/8, 00–000.
- Freud, S. (1901). Zur Psychopathologie des Alltagslebens (Vergessen, Versprechen, Vergreifen), nebst Bemerkungen über eine Wurzel des Aberglaubens. *Monatschrift für Psychiatrie und Neurologie* 10, 1–13.
- Fromkin, V. A. (1971). The non-anomalous nature of anomalous utterances. *Language* 47, 27–52.
- (ed.) (1973). *Speech Errors as Linguistic Evidence*. The Hague: Mouton.
- Garnes, S. and Bond, Z. S. (1975). Slips of the ear: Errors in perception of casual speech. *Proceedings of the Eleventh Regional Meeting, Chicago Linguistic Society*, 214–225.
- (1980). A slip of the ear: A snip of the ear? A slip of the year? In V. A. Fromkin (ed.), *Errors in Linguistic Performance: Slips of the Tongue, Ear, Pen, and Hand*. N. Y., London: Academic Press.
- Garnham, A., Shillcock, R. C., Brown, G. D. A., Mill, A. I. D. and Cutler, A. (1981). Slips of

- the tongue in the London-Lund corpus of spontaneous conversation. *Linguistics* 19-7/8, 805-817.
- Garrett, M. F. (1976). Syntactic processes in sentence production. In R. J. Wales and E. C. T. Walker (eds.), *New Approaches to Language Mechanisms*. Amsterdam: North-Holland.
- (1980). Levels of processing in sentence production. In B. Butterworth (ed.), *Language Production*. London: Academic Press.
- Goldstein, L. (1980). Bias and asymmetry in speech perception. In V. A. Fromkin (ed.), *Errors in Linguistic Performance: Slips of the Tongue, Ear, Pen, and Hand*. N. Y., London: Academic Press.
- Green, D. W. (1977). The intermediate processing of sentences. *Quarterly Journal of Experimental Psychology* 29, 135-146.
- Hockett, C. F. (1967). Where the tongue slips, there slip I. *To Honor Roman Jakobson*. The Hague: Mouton.
- Horowitz, L. M., Chilian, P. C. and Dunnigan, K. P. (1969). Word fragments and their re-integrative powers. *Journal of Experimental Psychology* 80, 392-394.
- Horowitz, L. M., White, M. A. and Atwood, D. W. (1968). Word fragments as aids to recall: the organisation of a word. *Journal of Experimental Psychology* 76, 219-226.
- Hurford, J. (1981). Malapropisms, left-to-right listing and lexicalism. *Linguistic Inquiry* 12, 419-423.
- Johnson, N. F. (1980). Part-whole relationships in word processing: psycholinguistics in the eyeball. Paper presented to the Midwestern Psychological Association, St. Louis.
- Lackner, J. R. (1980). Speech production: Correction of semantic and grammatical errors during speech shadowing. In V. A. Fromkin (ed.), *Errors in Linguistic Performance: Slips of the Tongue, Ear, Pen, and Hand*. N. Y., London: Academic Press.
- MacKay, D. G. (1969). Forward and backward masking in motor systems. *Kybernetik* 6, 57-64.
- Marslen-Wilson, W. D. (1975). Sentence perception as an interactive parallel process. *Science* 189, 226-228.
- (1978). Sequential decision processes during spoken word recognition. Paper presented to the Psychonomic Society, San Antonio.
- Marslen-Wilson, W. D. and Welsh, A. (1978). Processing interactions and lexical access during word recognition in continuous speech. *Cognitive Psychology* 10, 29-63.
- Meara, P. and Ellis, A. W. (1981). The psychological reality of deep and surface phonological representations: evidence from speech errors in Welsh. *Linguistics* 19-7/8, 797-804.
- Meringer, R. (1908). *Aus dem Leben der Sprache: Versprechen, Kindersprache, Nachahmungstrieb*. Berlin: Behr's Verlag.
- (1923). Die täglichen Fehler im Sprechen, Lesen und Handeln. *Wörter und Sachen* 8, 122-140.
- Meringer, R. and Mayer, K. (1895). *Versprechen und Verlesen: Eine Psychologisch-Linguistische Studie*. Stuttgart: Göschen.
- Miller, G. A. and Nicely, P. E. (1955). An analysis of perceptual confusions among some English consonants. *Journal of the Acoustical Society of America* 27, 338-352.
- Nooteboom, S. G. (1969). The tongue slips into patterns. *Leyden Studies in Linguistics and Phonetics*. The Hague: Mouton.
- (1980). Speaking and unspeaking: Detection and correction of phonological and lexical errors in spontaneous speech. In V. A. Fromkin (ed.), *Errors in Linguistic Performance: Slips of the Tongue, Ear, Pen, and Hand*. N. Y., London: Academic Press.
- (1981). Lexical retrieval from fragments of spoken words: beginnings versus endings. *Journal of Phonetics* 9, 407-424.

- Peterson, G. E. and Barney, H. L. (1952). Control methods used in a study of the vowels. *Journal of the Acoustical Society of America* 24, 175–184.
- Shattuck-Hufnagel, S. and Klatt, D. (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.
- (1980). How single phoneme error data rule out two models of error generation. In V. A. Fromkin (ed.), *Errors in Linguistics Performance: Slips of the Tongue, Ear, Pen, and Hand*. N. Y., London: Academic Press.
- Smith, P. T. and Groat, A. (1979). Spelling patterns, letter cancellation and the processing of text. In P. A. Kolars, M. Wrolstad and H. Bouma (eds), *Processing of Visible Language*. New York: Plenum.
- Taft, M. and Forster, K. I. (1976). Lexical storage and retrieval of polymorphic and polysyllabic words. *Journal of Verbal Learning and Verbal Behavior* 15, 607–620.
- Tent, J. and Clark, J. E. (1980). An experimental investigation into the perception of slips of the tongue. *Journal of Phonetics* 8, 317–325.
- Timpanaro, S. (1976). *The Freudian Slip*. London: New Left Press.
- Wang, M. D. and Bilger, R. C. (1973). Consonant confusions in noise: a study of perceptual features. *Journal of the Acoustical Society of America* 54, 1248–1266.
- Warren, R. M. (1970). Perceptual restoration of missing speech sounds. *Science* 167, 392–393.
- Wells, R. (1951). Predicting slips of the tongue. *Yale Scientific Magazine* 26, 9–30.