CONTEXTUAL PROBABILITY AND WORD FREQUENCY AS DETERMINANTS OF PAUSES AND ERRORS IN SPONTANEOUS SPEECH*

GEOFFREY W. BEATTIE
University of Sheffield
AND
B. L. BUTTERWORTH
University of Cambridge

This study investigated the relationship between the contextual probability of lexical items in spontaneous speech, as measured by the Cloze procedure, and word frequency. It also attempted to determine the relative importance of the two variables in causing delay, in the form of hesitation, in the production of spontaneous speech. The analysis revealed that content words of low contextual probability tended to be more infrequent than other words, and that both contextual probability and word frequency were associated with hesitation in speech. Contextual probability had an effect on hesitation even when word frequency was held constant, but word frequency had no effect when contextual probability was controlled. Analysis of certain types of errors, also, revealed that word frequency may play an important role in the lexical selection process.

INTRODUCTION

A number of studies have demonstrated a relationship between the location of various types of hesitation in spontaneous speech and the relative uncertainty of the succeeding lexical items (Goldman-Eisler, 1958; Tannenbaum, Williams and Hillier, 1965; Cook, 1969). For example, Goldman-Eisler (1958, p. 98), employing the Shannon guessing technique to measure transitional probability (defined as "The ratio of frequency of correct guesses to the total number of guesses made",[1] found that 28 out of 34 words preceded by an unfilled pause (silence $\geqslant$ 250 msec.) were of comparatively low transitional probability ($\leqslant$ 0.10). This study has, however, been strongly criticized, firstly because the speech sample used was clearly unrepresentative of spontaneous speech generally; only "grammatically correct ... well constructed sentences ... logically consistent with the whole utterance" were considered (Goldman-Eisler, 1968, p. 36). Such a sample also necessarily excludes the large subset of between-sentence pauses (Butterworth, 1972). Goldman-Eisler (1972) herself, in fact, demonstrated that 77.9% of sentences in spontaneous speech are divided from each other by pauses longer than 500 msec. The second major criticism of the Goldman-Eisler (1958) study has centred on the statistical treatment of the data (Boomer, 1970).

* *This research was carried out while the senior author was at the Psychological Laboratory, University of Cambridge, and a member of Trinity College.*

[1] *But see Boomer (1970) for discussion of some inconsistencies in this definition.*

In a later study, Tannenbaum *et al.* (1965), using the Cloze procedure (Taylor, 1953) to measure contextual probability,[2] found that hesitations (particularly filled pauses – *ah, er, um,* etc.) preceded words of significantly lower contextual probability than words occurring in fluent contexts. Cook (1969), again using the Cloze procedure, found contextual probability to be lower for every part of speech (except pronouns and auxiliary verbs) following filled pauses. The conclusions from these studies have been that pauses "anticipate increases of information in subsequent speech" and that they involve "acts of choice" between lexical items (Goldman-Eisler, 1968, p. 48). This view appears to have gone unchallenged: "Hesitation pauses that occur within phrases or clauses seem to reflect the difficulty of choosing a word out of many alternatives" (Taylor, 1976, p. 169).

However, such a conclusion may not be warranted by the data since transitional, or contextual, probability may have been confounded with word frequency, which has itself been shown to be an important determinant of delay in word, or sentence, production in a number of experiments. Word frequency may affect measures of transitional, or contextual, probability because judges may be less familiar with more uncommon words and may, therefore, find them more difficult to guess. A number of experiments have demonstrated that word frequency does cause delay in word and sentence production. Oldfield and Wingfield (1965), in an object-naming situation, demonstrated a linear relationship between latency of response to a presented picture of an object and the log "frequency of occurrence" of the object name. Freedman and Loftus (1971) found a virtually linear relationship between word frequency of response and response latency in a situation where subjects produced a word falling into the semantic space defined by a noun category paired with either an adjective (e.g. *daffodil* in response to *flower-yellow*) or a letter (e.g. *zebra* in response to *animal-z*). Taylor (1969), in a sentence production task, found that topic word frequency had a greater effect than topic abstractness in delaying sentence production. The mean latency to begin speaking on infrequent concrete topics was 3.59 sec. compared with 2.91 sec. for frequent abstract topics. Mercer (1976), in a study of spontaneous speech, found that within hesitant phonemic clauses (Trager and Smith, 1957), there was a positive relationship between the number of words following a pause up to the next pause, or terminal juncture, and the average word frequency of the fluent sequence. Mercer thus concluded that word frequency was a primary influence on the fluency of production of spontaneous speech. It should be noted, however, that in this study Mercer did not attempt to control for transitional probability, which again may have been confounded with word frequency (a confounding in the reverse direction to that discussed earlier). Neither does Mercer appear to have considered the relationship between individual hesitations and the frequency of the immediately succeeding lexical items but rather, following Boomer (1965), the average

[2] *In the Shannon guessing technique employed by Goldman-Eisler, judges guess each successive word in a sentence, they therefore have only the preceding linguistic context available to them when guessing. In the Cloze procedure every nth word is deleted and replaced by a blank which judges attempt to guess. Therefore, both the preceding and the following linguistic contexts are available. The Cloze procedure thus yields a measure of contextual probability rather than strict transitional probability.*

frequency of words within the phonemic clause in which the hesitation occurred.

Boomer (1965), in another study of spontaneous speech, analysed the distribution of hesitations with respect to phonemic clauses and, having observed that these tended to occur towards the beginnings of such clauses, concluded (following Lounsbury, 1954) that the phonemic clause was the fundamental unit of encoding. Boomer also argued that the observed distribution could not be a function of transitional probability (as observed by Goldman-Eisler, 1958) but in so doing clearly demonstrated that transitional probability and word frequency were not kept conceptually distinct. Boomer's (1965, p. 88) argument was that "primary stress typically occurs towards the end of a phonemic clause, almost invariably the last or next to last word in the clause receives the stress. And, as Berry (1953) has shown, primary stress is negatively related to word frequency . . . Thus the high-information lexical words tend to occur towards the end of phonemic clauses." Here transitional probability and word frequency are clearly confused. In principle they need not always covary. For example, in the sentence "Too many cooks spoil the soup," *soup* would probably have a lower transitional probability than *broth* whilst being of higher frequency. It is an empirical question whether transitional probability and word frequency are systematically related in spontaneous speech. They may well be; word frequency may affect judges' guesses of words in context because judges may be completely unfamiliar with certain low frequency words or alternatively they may know the words but think of them less readily.

The aim of the present study is to investigate the relationship between hesitations in spontaneous speech and the contextual probability and word frequency of the succeeding lexical items.

PROCEDURE

Subjects

Subjects were native speakers of English, all having completed at least an undergraduate degree; their mean age was 24 years.

Material

Samples of speech from three subjects were taken from 3½ hours spontaneous dyadic discussions, in which subjects were asked to pick a proposition they agreed with, from a list of propositions, and argue for it. Samples were transduced into a graphic representation of phonation and silence by means of a signal detector linked to an Ediswan pen-oscillograph. The words uttered were matched to the phonations given by the pen-oscillograph record, and therefore, the locations of unfilled pauses (silence $\geqslant$ 200 msec., cf. Boomer, 1965) with respect to the verbal content of the utterance could be determined accurately. Filled pauses (*ah, er, um,* etc.) were identified on the speech transcripts.

Method

The Cloze procedure was used to measure contextual probability. This technique

gives results broadly comparable to the version of Shannon's sequential guessing-game technique employed by Goldman-Eisler (1958) (cf. Schiavetti and Burke, 1974; Burke and Schiavetti, 1975). The procedure consists of deleting words from a text, and employing a number of judges to guess the deleted items. In this study, five protocols of each text were prepared, each with every fifth word deleted such that no word was deleted on two protocols. Each of the five protocols was given to a set of five judges, thus each word in the text became associated with a Cloze score of 0, 1, 2, 3, 4 or 5 according to how many judges guessed it correctly. All filled hesitation, filled pauses (*ah, er, um,* etc.), repetition ("all repetitions judged non-significant semantically"), and false starts ("all incomplete or self-interrupted utterances") (Maclay and Osgood, 1959) were eliminated from the texts. The residue was represented in normal orthography and punctuation.

Judges were 25 native English-speaking undergraduates with a mean age of 20 years.

Word frequency values were taken from Kučera and Francis (1967). Only content words (Fries, 1952) were included in the analysis reported in this study. Content words immediately preceded by an unfilled or filled pause were considered hesitant; those not immediately preceded by an unfilled or filled pause were considered fluent. Content words preceded by a repetition or false start were not considered hesitant unless the repetition or false start was associated with either an unfilled or filled pause. Although there is some evidence that repetition may be used for general cognitive planning (Beattie and Bradbury, 1979), there is evidence to suggest that it is not used for short range *forward* planning, unlike unfilled and filled pauses (Tannenbaum *et al.*, 1965).

The present corpus consists of 253 content words and 53 speech locations occupied by a hesitation (either a single unfilled pause, filled pause or combination of the two).

RESULTS

The mean word frequency values (per 10^6) for hesitant and fluent lexical items with high, medium or low Cloze scores are presented in Table 1. It is apparent that contextual probability and word frequency are related; high Cloze score lexical items tend to have higher frequencies in English than medium or low Cloze score items. There is also a trend for lexical items immediately preceded by hesitations to be of lower frequency than their fluent counterparts (although this does not appear to be the case with high Cloze score lexical items).

These data were analysed firstly by comparing the word frequency of hesitant and fluent lexical items. Two word frequency criteria were adopted (10^2 per 10^6, and 10^3 per 10^6) and the numbers of lexical items with high (4, 5), medium (2, 3) or low (0, 1) Cloze scores above or below these criteria were noted. See Table 2.

The data contained in Table 2 are decomposed and analysed in the succeeding tables. It was discovered firstly that there were significantly more low Cloze score lexical items immediately preceded by a hesitation (compared with medium or high Cloze score items) than occurred in fluent contexts ($G = 4.880$, $p < 0.05$; Sokal and Rohlf, 1973). See Table 3.

TABLE 1

Mean word frequency values (per 10^6, Kučera and Francis, 1967) for hesitant and fluent lexical items with high (4, 5), medium (2, 3) or low (0, 1) Cloze scores

	Cloze Score		
Lexical Items	High	Medium	Low
Hesitant	4648.0 (n=18)	1418.3 (n=7)	343.1 (n=28)
Fluent	3372.2 (n=77)	2565.2 (n=51)	854.1 (n=72)
Both	3613.9 (n=95)	2426.8 (n=58)	711.0 (n=100)

TABLE 2

Cloze score and word frequency for hesitant and fluent lexical items

		Cloze Score		
Lexical Items	Word Frequency	High (4, 5)	Medium (2, 3)	Low (0, 1)
Hesitant	$> 10^3$ per 10^6	11	2	2
	$\leqslant 10^3$ per 10^6	7	5	26
Fluent	$> 10^3$ per 10^6	43	18	11
	$\leqslant 10^3$ per 10^6	34	33	61
Hesitant	$> 10^2$ per 10^6	17	5	13
	$\leqslant 10^2$ per 10^6	1	2	15
Fluent	$> 10^2$ per 10^6	70	41	48
	$\leqslant 10^2$ per 10^6	7	10	24

TABLE 3

Cloze scores of hesitant and fluent lexical items

Lexical Items	Cloze Score: High/Medium	Cloze Score: Low
Hesitant	25	28
Fluent	128	72
Both	153	100

$G = 4.880$ ($p < 0.05$)

TABLE 4

Cloze score and word frequency

Frequency	Cloze score: High	Cloze score: Low
$> 10^3$ per 10^6	54	13
$\leqslant 10^3$ per 10^6	41	87

$G = 41.586$ ($p < 0.001$)

This result replicates that previously reported by Goldman-Eisler (1958), Tannenbaum *et al.* (1965) and Cook (1969). However, it was also discovered that low Cloze score items were significantly more infrequent than high Cloze score items ($G = 41.586$, $p < 0.001$). See Table 4. In the case of low Cloze score lexical items 87.0% had a frequency $\leqslant 10^3$ per 10^6, compared with 43.2% of high Cloze score items. This is positive evidence that Cloze score and word frequency are related, and therefore probably confounded in earlier studies.

It was also discovered that lexical items preceded by a hestitation were significantly more infrequent than were fluent lexical items ($G = 3.994$, $p < 0.05$). See Table 5. Considering all lexical items in the present sample with frequencies $\leqslant 10^2$ per 10^6 (i.e. 1 in 10^4), 30.5% were preceded by a hesitation compared with 18.0% of words of higher

TABLE 5

Word frequency (Kučera and Francis, 1967) of hesitant and fluent lexical items

Lexical Items	Word frequency	
	$\leqslant 10^2$ per 10^6	$> 10^2$ per 10^6
Hesitant	18	35
Fluent	41	159
Both	59	194

$G = 3.994$ ($p < 0.05$)

TABLE 6

Cloze scores of hesitant and fluent lexical items, with frequencies $\leqslant 10^3$ per 10^6 (Kučera and Francis, 1967)

	Cloze Score	
	High/Medium	Low
Hesitant	12	26
Fluent	67	61

$G = 4.350$ ($p < 0.05$)

frequency.

However, when Cloze score was controlled (see Table 2) there was no significant difference in the word frequency of hesitant and fluent lexical items (considering the relative numbers of words with frequencies $\leqslant 10^2$ per 10^6 and $> 10^2$ per 10^6 in hesitant and fluent contexts). In the case of low Cloze score items only $-G = 2.632$ ($p > 0.05$), and for high Cloze score items only $-G = 0.001$ ($p >> 0.05$). For low Cloze score items, only 38.5% of infrequent words ($\leqslant 10^2$ per 10^6) were hesitant compared with 21.3% of the more frequent items (word frequency $> 10^2$ per 10^6). For high Cloze score lexical items, only 12.5% of infrequent items ($\leqslant 10^2$ per 10^6) were hesitant, compared with 19.5% of the more frequent items ($> 10^2$ per 10^6).

On the other hand, when word frequency was partially controlled, by considering only low frequency lexical items ($\leq 10^3$ per 10^6), it was discovered that words of low Cloze score were still significantly more likely to be hesitant than were words of medium, or high, Cloze score ($G = 4.350, p < 0.05$). See Table 6. Considering only the low frequency lexical items, 29.9% of words of low Cloze score were preceded by a hesitation, compared with 15.2% of words of medium or high Cloze score. This significant effect does not emerge with words of higher frequency ($G = 0.042, p >> 0.05$).

CONCLUSIONS

(1) Words of low contextual probability in spontaneous speech are significantly more likely to be hesitant than are words of higher contextual probability.

(2) Words of low frequency in spontaneous speech are significantly more likely to be hesitant than are words of higher frequency.

(3) Words of low contextual probability in spontaneous speech are significantly more infrequent than are words of higher contextual probability.

(4) When contextual probability is held constant, there is no significant difference in the word frequency of hesitant and fluent lexical items.

(5) When word frequency is held constant (low frequency lexical items only), words of low contextual probability are more likely to be hesitant than are words of higher contextual probability.

DISCUSSION

This study demonstrated that the contextual probability of lexical items in a continuous sample of spontaneous speech, as measured by the predictability of these words in context, is related to word frequency. Unpredictable, high-information, lexical items are significantly more infrequent that are the more predictable lexical items. These results make it difficult to interpret the earlier studies (Goldman-Eisler, 1958; Tannenbaum *et al.* 1965; Cook, 1969) which demonstrated a relationship between hesitations in speech and the relative unpredictability of the succeeding lexical items. A confounding of these two variables undoubtedly was present. In the present study hesitations were again observed to precede relatively unpredictable lexical items, but in addition, they preceded relatively infrequent lexical items.

However, in the present study it was possible to hold one of these variables constant and to determine whether the other variable resulted in a delay in the production of spontaneous speech. When contextual probability was held constant, there was no significant difference in the word frequency of hesitant and fluent lexical items. However, when word frequency was partially controlled (by considering only low frequency lexical

items), it was observed that unpredictable lexical items were more likely to be preceded by a hesitation than were the more predictable words. These results suggest that some hesitations indeed reflect the actual "act of choice" between lexical items which share the same or similar semantic features and fit the same linguistic contexts. Lexical items produced in spontaneous speech following such "acts of choice" would presumably have low contextual probabilities, as measured by the Cloze procedure, because there would be a number of alternative words which would be appropriate in that context, and therefore judges would find it difficult to guess the exact word produced by the original speaker. The time required for the original speaker to choose between alternative words is in addition to any time invested in accessing the less readily available infrequent lexical items (Oldfield and Wingfield, 1965; Mercer, 1976).

There is evidence from another (related) source that word frequency may play an important role in the lexical selection process. The other evidence comes from speech errors. Two common types of errors are word blends and word substitutions. Blends occur when "non-existent words are produced as the result of composites of two words with similar semantic features" (Fromkin, 1973, p. 235), e.g. "Don't shell so loud" where *shell* seems to be a blend of *shout* and *yell* (from Hockett, 1967). Word substitution occurs when a complete word unintentionally substitutes for the intended word, e.g. "He got hot under the belt (collar)" (from Fromkin, 1973, Appendix V).

According to Fromkin (1973, p. 235), blends reflect difficulties in choosing between alternative lexical items with similar semantic features: "In selecting words, it appears that he (the speaker) is matching semantic features. Where there are a number of alternative possibilities, rather than making an immediate selection, he brings them both into a buffer, storage compartment, with their phonological specifications. Either a selection occurs at this point, or the words are blended." The evidence obtained in the present study, of course, suggests that the selection, when it is successful, requires time in the form of a hesitation at such points. Moreover, analysis of the Fromkin (1973) corpus suggests that word frequency may affect this lexical selection process. Content words which blended rather than substituted were significantly closer in frequency. In a corpus of 64 word blends[3] (Fromkin, 1973, Appendix U) and 65 word substitutions (Appendix V), it was discovered that 39.1% of the words combining in blends had a frequency difference $\leqslant$ 10 per 10^6, 1 in 10^5 (cf. Kučera and Francis, 1967), compared with 20.0% of the words which completely substituted ($G = 4.808, p < 0.05$). This result suggests that in the lexical selection process alternative words sharing semantic features are more likely to be simultaneously processed up to the phonological output stage if they are of similar frequencies. Of course, the present study has demonstrated a relationship between word frequency and contextual probability. Words similar in frequency are also likely to be relatively similar in contextual probability. Thus, contextual probability probably played some role in mediating the effects described here. The present research cannot determine whether word frequency *per se*, or word frequency and contextual probability, is responsible for these effects since it is not possible to

[3] *One blend in the Fromkin corpus involved function words, and this was therefore excluded from the present analysis.*

compute Cloze scores on the speech error data obtained by Fromkin, because insufficient linguistic context is reported. Future research must determine whether word frequency or contextual probability is mainly responsible for affecting the probability of content words blending rather than substituting in spontaneous speech.

Nevertheless, this result does suggest that during certain hesitations in spontaneous speech where lexical selection occurs, speakers not only have some knowledge of the semantic features of the words being searched for (Butterworth and Beattie, 1978) but also some knowledge of the approximate frequency and/or contextual probability of the word. Word frequency and/or contextual probability thus appear to be important factors in guiding lexical selection processes. Exclusion of function words from the present study meant that pauses at syntactic junctures tended to be excluded. The focus of this study was thus on lexical items, and pauses used for lexical search. The results do not, of course, imply that higher level planning units do not operate in spontaneous speech. Indeed, there is considerable evidence for such units (Boomer, 1965; Henderson, Goldman-Eisler and Skarbek, 1966; Goldman-Eisler, 1967; Butterworth, 1975; Beattie, 1978, 1979, in press).

The results do, however, suggest that word frequency and contextual probability are a pervasive influence on the planning processes involved in spontaneous speech.

In conclusion, these results suggest that the contextual probabilities of lexical items in spontaneous speech are related to their frequency of usage in the language as a whole, but, nevertheless, contextual probability does appear to cause delay in the production of spontaneous speech (in the form of a hesitation) even when word frequency is held constant. The evidence from speech errors also suggests that both word frequency and contextual probability may play important roles in the lexical selection process.

REFERENCES

BEATTIE, G.W. (1978). Sequential temporal patterns of speech and gaze in dialogue. *Semiotica,* **23**, 29-52.

BEATTIE, G.W. (1979). Planning units in spontaneous speech: Some evidence from hesitation in speech and speaker gaze direction in conversation. *Linguistics,* **17**, 61-78.

BEATTIE, G.W. (in press). The role of language production processes in the organisation of behaviour in face-to-face interaction. In Butterworth, B. (ed.), *Language Production* (London).

BEATTIE, G.W. and BRADBURY, R.J. (1979). An experimental investigation of the modifiability of the temporal structure of spontaneous speech. *J. psycholing. Res.*, **8**, 225-48.

BERRY, J. (1953). Some statistical aspects of conversational speech. In Jackson, W. (ed.), *Communication Theory* (London).

BOOMER, D.S. (1965). Hesitation and grammatical encoding. *Language and Speech,* **8**, 148-58.

BOOMER, D.S. (1970). Review of F. Goldman-Eisler's *Psycholinguistics: Experiments in Spontaneous Speech. Lingua,* **25**, 152-64.

BURKE, J.P. and SCHIAVETTI, N. (1975). Effects of cumulative context and guessing methods on estimates of transition probability in speech. *Language and Speech,* **18**, 299-311.

BUTTERWORTH, B.L. (1972). *Semantic Analysis of the Phasing of Fluency in Spontaneous Speech.* Ph.D. Dissertation, University College, London.

BUTTERWORTH, B.L. (1975). Hesitation and semantic planning in speech. *J. psycholing. Res.*, **4**, 75-87.

BUTTERWORTH, B. and BEATTIE, G.W. (1978). Gesture and silence as indicators of planning in speech. In R.N. Campbell and P.T. Smith (eds.), *Recent Advances in the Psychology of Language: Formal and Experimental Approaches*, Vol. 4B (New York), pp. 347-60.

COOK, M. (1969). Transition probabilities and the incidence of filled pauses. *Psychonomic Science*, **16**, 191-2.

FREEDMAN, J.L. and LOFTUS, E.F. (1971). Retrieval of words from long-term memory. *J. verb. Learn. verb. Behav.*, **10**, 107-15.

FRIES, C.C. (1952). *The Structure of English* (New York).

FROMKIN, V. (1973). The non-anomalous nature of anomalous utterances. In V.A. Fromkin (ed.), *Speech Errors as Linguistic Evidence* (The Hague).

GOLDMAN-EISLER, F. (1958). Speech production and the predictability of words in context. *Quart. J. exper. Psychol.*, **10**, 96-106.

GOLDMAN-EISLER, F. (1967). Sequential temporal patterns and cognitive processes in speech. *Language and Speech*, **10**, 122-32.

GOLDMAN-EISLER, F. (1968). *Psycholinguistics: Experiments in Spontaneous Speech* (London).

GOLDMAN-EISLER, F. (1972). Pauses, clauses, sentences. *Language and Speech*, **15**, 103-13.

HENDERSON, A., GOLDMAN-EISLER, F. and SKARBEK, A. (1966). Sequential temporal patterns in spontaneous speech. *Language and Speech*, **9**, 207-16.

HOCKETT, C.F. (1967). Where the tongue slips, there slip I. Reprinted in V.A. Fromkin (ed.), (1973). *Speech Errors as Linguistic Evidence* (The Hague).

KUČERA, H. and FRANCIS, W.N. (1967). *Computational Analysis of Present-Day American English* (Providence).

LOUNSBURY, F.G. (1954). Transitional probability, linguistic structure and systems of habit – family hierarchies. In C.E. Osgood and T.A. Sebeok (eds.), *Psycholinguistics: A Survey of Theory and Research Problems* (Baltimore).

MACLAY, H. and OSGOOD, C.E. (1959). Hesitation phenomena in spontaneous English speech. *Word*, **15**, 19-44.

MERCER, N. McK. (1976). Frequency and availability in the encoding of spontaneous speech. *Language and Speech*, **19**, 129-43.

OLDFIELD, R.C. and WINGFIELD, A. (1965). Response latencies in naming objects. *Quart. J. exper. Psychol.*, **4**, 272-81.

SCHIAVETTI, N. and BURKE, N.P. (1974). Comparison of methods of estimating transition probability in speech. *Language and Speech*, **17**, 347-52.

SOKAL, R.R. and ROHLF, F.J. (1973). *Introduction to Biostatistics* (San Francisco).

TANNENBAUM, P.H., WILLIAMS, F. and HILLIER, C.S. (1965). Word predictability in the environments of hesitation. *J. verb. Learn. verb. Behav.*, **4**, 134-40.

TAYLOR, I. (1976). *Introduction to Psycholinguistics* (New York).

TAYLOR, L. (1969). Content and structure in sentence production. *J. verb. Learn. verb. Behav.*, **8**, 170-5.

TAYLOR, W.L. (1953). "Cloze" procedure: a new tool for measuring readability. *Journalism Quarterly*, **30**, 415-33.

TRAGER, G.L. and SMITH, H.L. (1957). *An Outline of English Structure* (Washington, D.C.).