

First in, first out: Word learning age and spoken word frequency as predictors of word familiarity and word naming latency

GORDON D. A. BROWN
University of Essex, Colchester, England

and

FRANCES L. WATSON
M.R.C. Applied Psychology Unit, Cambridge, England

This study is concerned with recent claims that subjective measures of word frequency are more suitable than are standard word frequency counts as indices of actual frequency of word encounter. A multiple regression study is reported, which shows that the major predictor of familiarity ratings is word learning age. Objective measures of spoken and written word frequency made independent contributions to the variance. It is concluded that rated familiarity is not an appropriate substitute for objective frequency measures. A multiple regression study of word naming latency is reported, and shows that rated word learning age is a better predictor of word naming latency than are spoken word frequency, written word frequency, rated familiarity, and other variables. Possible theoretical explanations for age-of-acquisition effects are discussed and it is concluded that early-learned words have a more complete representation in a phonological output lexicon. This conclusion is related to relevant developmental literature.

The effects of word frequency on lexical decision time have been widely studied, and models of word recognition have accorded a central role to accounts of these effects. However, frequency effects in word naming have been less thoroughly investigated. Accounts of such effects in word naming followed naturally from theories of lexical decision tasks that assigned the effect of frequency to an early stage of lexical analysis, but the trend toward postaccess accounts of frequency effects has left the locus of frequency effects in word naming less clear. For example, Balota and Chumbley (1984) showed that words that show a strong frequency effect in a lexical decision task do not show a significant frequency effect when presented singly in a semantic classification task. However, Balota and Chumbley (1985) did find a frequency effect in word naming. If lexical access involves automatic access to phonological information about the target word (e.g., Humphreys, Evett, & Taylor, 1982), then a pure

postaccess decision-making account of frequency effects in lexical decision would need to assume that the frequency effect in word naming reflects relatively peripheral processes, such as ease of articulation (see Balota & Chumbley, 1985; Theios & Muise, 1977).

If frequency effects in word naming occur at output, then we might expect that spoken word frequency, rather than written word frequency, would be the relevant factor in determining word naming latency. Tryk (1968) found that ratings of spoken word production frequency were correlated with an objective measure of written word frequency, but represented a distinct measure (see also Rubin, 1976). We explore this issue experimentally in the present paper, using multiple regression techniques to assess the relative sizes of the independent contributions to explained variance made by the spoken and written frequency variables. We also use experimental evidence to investigate the possibility that previously reported effects of word learning age (e.g., Carroll & White, 1973; Gilhooly & Logie, 1981) are redundant on spoken word frequency. The final discussion concerns possible loci of word frequency and age-of-acquisition effects during the later stages of word production, and a model is outlined to account for the findings.

The question of the best measures of word frequency to use in the design of experiments has recently received some attention (Gernsbacher, 1984; Gordon, 1985). Gernsbacher pointed to apparent inconsistencies in the literature concerning the interactions between objective word frequency measures and other word attributes, such

The experiments reported in this paper were carried out at the University of Sussex. We acknowledge the support of Grant CO8250011 awarded to Gordon D. A. Brown by the Economic and Social Research Council (U.K.). Frances L. Watson was on a Social Science Research Council studentship at the University of Aberdeen, working with Ken Gilhooly.

We thank Marcel Tatham for advice and assistance and Anne Cutler and Dennis Norris for helpful comments on earlier versions of the manuscript. Alice Healy, David Rubin, and an anonymous reviewer have contributed substantially to the quality of the final manuscript.

Please address all correspondence and reprint requests to Gordon Brown at the Department of Language and Linguistics, University of Essex, Colchester CO4 3SQ, England.

as orthographic regularity and degree of polysemy, and suggested that the inconsistent findings might reflect inadequacies in the objective written word frequency measures, particularly at the low-frequency end of the scale. Gernsbacher manipulated rated familiarity instead of objective word frequency in a series of experiments and indeed found that there were no interactions of the kind previously reported in the literature. However, there is a problem with the interpretation of Gernsbacher's own results, owing to uncertainty about the criteria subjects adopt in making familiarity ratings. Although the ratings that Gernsbacher obtained were reliable, insofar as there was a good degree of interrater agreement, there was no evidence for the assumption that the ratings represent a valid index of "experiential frequency." It is possible, for example, that factors such as word frequency and polysemy have interactive effects on the perceived familiarity of a word (see Gordon, 1985). If this were so, then it would scarcely be surprising that interactions between ambiguity and rated familiarity were not obtained in Gernsbacher's experiments. In general, as Rubin (1976) pointed out, there is more to subjective measures than simply objective frequency.

Gordon (1985) has also used a rated frequency variable. He found that there were good correlations between rated and objective measures across the entire frequency range, but relatively little correlation between them for low-frequency words. Furthermore, the subjective measure accounted for more reaction-time variance in a lexical decision task. However, it is again unclear whether the subjective ratings provide an accurate reflection of genuine experiential frequency. For example, it is not perhaps totally implausible to suggest that subjects might base their familiarity ratings on the length of time it takes them to recognize the relevant word. The fact that such a possibility cannot be excluded demonstrates the need for a greater understanding of the ways in which familiarity ratings are produced.

There is already evidence to suggest that rated measures reflect some aspects of subjects' experience with words other than actual frequency of encounter. If the subjective measure is a better index of true frequency, then we would expect objective measures to correlate more highly with subjective measures than objective measures correlate with each other (assuming that objective measures are just unreliable, rather than systematically biased in the same way as each other). In fact, Gordon (1985) found precisely the reverse of this: his subjective-objective correlations were significantly *lower* than were the correlations of the objective frequencies with each other. Furthermore, Howes (1954) found that subjective measures show better correlations with each other than with objective measures. This pattern of correlations suggests that the subjective and objective measures reflect qualitatively different dimensions of stimulus familiarity. If this is so, then the substitution of rated measures for objective indices as a variable to use in word recognition and production experiments may lead to a serious underestimation of the

role of other word attributes in lexical processing. Further suggestive evidence is provided by Galbraith and Underwood (1973), who found that abstract words with the same objective frequency index as control concrete words are accorded higher frequency ratings by subjects. Galbraith and Underwood suggested that "contextual variety" may influence the rating process. In addition, Rubin (1983) found that associative frequency predicts perceived familiarity.

Note that our claim is *not* that objective frequency counts provide an adequate index of actual frequency of word encounter. Rather, we suggest that the use of rated measures as a substitute may prove misleading, since a multitude of factors may influence subjects' ratings. The following analysis tests the hypothesis that other factors influence rated familiarity. In particular, it examines the roles of spoken and written word frequency, as well as that of word age-of-acquisition, in determining familiarity ratings.

ANALYSIS

An examination of the study by Rubin (1980) reveals that rated familiarity in his study was most highly correlated with rated pronunciability, associative frequency (the number of times the word was produced as the first associate to some other word), word age-of-acquisition, and four objective frequency measures. All of these factors might be expected to contribute to rated word familiarity, although we note that Rubin's rated pronunciability measure might itself be multiply determined. Brown (1984) found that rated familiarity was more highly correlated with spoken than with written word frequency and was most highly correlated with rated word age-of-acquisition. Rubin (1983) found that associative frequency predicted rated familiarity, although his analysis included no measure of age-of-acquisition. Accordingly, we examined the extent to which various factors contribute to variance in familiarity ratings by using rated familiarity as the dependent measure in multiple regression analysis.

Multiple regression techniques were employed here because such analysis provides a way of separating out the *independent* contributions of various correlated factors to the variance in the dependent measure. Multiple regression analyses are becoming more widely used in psycholinguistic research owing to the difficulty of obtaining sufficiently large stimulus samples to design orthogonal experiments that control for all variables suspected of being correlated with the variable under investigation (see Cutler, 1981). Both Whaley (1978) and Rubin (1980) reported major regression studies in attempts to determine the independent contributions of various word attributes to performance on different tasks. Our analysis differs from these studies in that the dependent variable in our study is rated familiarity rather than a performance measure, such as naming latency or lexical decision time.

Large regression studies inevitably supply a rich source of data and further hypotheses. However, we are primarily

interested in the extent to which factors other than word frequency independently predict familiarity ratings for words. In order to assess the independent contributions of the variables that we suppose might influence the rating process, it is of course necessary to include a number of variables that we do not necessarily expect to be related to word familiarity. The variables we used are discussed individually below.

Method

Selection of materials and variables. Gilhooly and Logie (1980) reported ratings of 1,944 words on a number of different dimensions, including a familiarity scale. The instructions used by Gilhooly and Logie to obtain the familiarity ratings were essentially the same as those employed by Gernsbacher (1984). We selected from this count every item that both had a Thorndike and Lorge (1944) word frequency value greater than 100 words per million and appeared in a separate spoken word frequency count (Brown, 1984). Words with a value of zero in the Brown count were not included, because the true probability of occurrence of such items is even more difficult to estimate than that of words with a frequency value of one or more. This selection process resulted in a sample of 437 items. The sample was further reduced, however, because we wished to have a measure of orthographic regularity in the regression equation. The best measure of this type is the positional bigram frequency count, published by Solso and Juel (1980), but this count gives values only for words up to 9 letters in length. We therefore discarded from our sample the 21 words with 10 or more letters, leaving a final sample of 416 items.

As there were no low-frequency words (less than 100 per million in Thorndike & Lorge, 1944) in this sample, there is no possibility that correlations obtained between the Gilhooly and Logie (1980) rated measures could have been artificially inflated due to subjects giving extreme values to unknown words.

Independent measures. All the variables used in this study were word attributes rather than subject attributes. Although it is of course possible that factors such as subject age or vocabulary score would affect familiarity ratings, it is unlikely that such variables would illuminate the issue under investigation here.

Two variables of potential interest were spoken word frequency and written word frequency. We expected that one or both of these variables would account for much of the variance in the familiarity ratings, although the main question concerned the amount of variance that would be attributed to the effect of other factors when the influence of the two word frequency measures had been partialled out. Note that the question of whether the spoken and written counts are genuinely different or reflect the same underlying dimension of experiential frequency is irrelevant in deciding whether other factors affect familiarity ratings. If, however, the two frequency measures *independently* contribute to the variance, then that would provide evidence that they represent genuinely distinct measures, rather than being different indices of the same word attribute.

Our chosen measure of written word frequency was the widely used frequency-per-million count from Kučera and Francis (1967). This was the most recent count available to us of reasonable size, and the measure was the one with which rated familiarity is most likely to be seen as competing as a variable to be used in experimental design. The measure of production word frequency from Brown (1984) is based on the London-Lund corpus of spontaneous conversation (Svartvik & Quirk, 1980). The corpus is based on surreptitiously recorded spontaneous conversations of native speakers of British English (full details can be found in Svartvik & Quirk, 1980).

Subjective ratings of word age-of-acquisition (AOA), imageability, and concreteness from the Gilhooly and Logie (1980) word sample were used as variables. All three variables might be expected to contribute to subjective familiarity on the basis of previous results. In particular, Brown (1984), using the same sample of words used in the present study, found that rated AOA was more highly correlated with rated familiarity than was any other variable, although, of course, the high correlation does not mean it would show an independent effect in the regression analysis. A "meaning uncertainty" ambiguity variable was also included (Gilhooly & Logie, 1980). Although indisputable effects of semantic variables on lexical decision or word naming latency are rare, it is possible that ambiguity would influence familiarity ratings.

Other measures included in the analysis were positional bigram frequency (Solso & Juel, 1980) and word length (in letters).

Analyses and Results

As the first stage in the analysis, summary statistics of the 416 items were obtained (see Table 1). The skews of three variables—positional bigram frequency and the two word frequency measures—were greater than 1.0. In this analysis, as throughout, those simple transformations were chosen that most effectively reduced the skews. The skew in the bigram frequency measure was reduced to below 1.0 by a square root transformation, and the word frequency measures underwent log transformations that successfully reduced the skews to below 1.0.

The correlations of familiarity with the independent measures are presented in Table 2. (The intercorrelations of all the variables can be obtained from Table 4.) Rated familiarity was most highly correlated with rated AOA ($r = -.58$), and next most highly correlated with spoken word frequency ($r = .50$) and then written word frequency ($r = .37$). The fact that the most highly correlated variable was neither of the frequency measures already suggests that familiarity ratings are not performed solely on the basis of experiential frequency.

A simultaneous multiple regression analysis was performed. The results of this are summarized in Table 3. The beta values represent the weights accorded to each variable in the final regression equation, and the F values reflect the significance of the variance that can be attributed to each variable independent of the effects of all

Table 1
Summary Statistics of Items Used in the Analysis and the Experiment

Variable	Mean	Range	SD
Latency (msec)*	577	150	25.4
Position*	220	435	126
Initial Phoneme*	32.9	80	31.6
Bigram Frequency	1345	9461	1029
Spoken Frequency	18.9	532	40.2
Written Frequency	132	974	132
Age-of-Acquisition	3.36	4.31	0.92
Familiarity	5.65	3.22	0.55
Imageability	4.91	4.32	1.02
Concreteness	4.63	4.83	1.33
Ambiguity	0.47	2.36	0.59
Length	5.62	6	1.67

*Not included in the analysis.

Table 2
Correlations Between Rated Familiarity and Independent Measures

Variable	Correlation
Age-of-Acquisition	-.580
Imageability	.245
Concreteness	.208
Ambiguity	-.058
Length	-.027
Bigram Frequency	.166
Spoken Frequency	.500
Written Frequency	.365

the other variables. The bigram frequency measure showed too small an effect to be entered into the final equation by the analysis package.

The final equation accounted for a significant 56% of the variance [$F(9,406) = 56.6, p < .01$].

Four variables contributed significant independent variance in the final simultaneous equation: age-of-acquisition, spoken word frequency, written word frequency, and word length. No other variables accounted for a significant proportion of the variance when the effects of all other variables had been partialled out.

Discussion

The factors we included in the analysis accounted for 56% of the dependent variance. Although there may be other factors, not included in our set of variables, that contribute to familiarity ratings, our results show clearly that familiarity ratings cannot be used as a substitute for objective word frequency measures (Rubin, 1976), since other factors, such as word learning age, contribute to the familiarity ratings (see also Rubin, 1983). Indeed, not only does AOA independently predict familiarity ratings, but in our study it was more highly correlated with the familiarity measure than were either of the frequency counts. It may be interesting to study subjective familiarity judgments in their own right, but subjects are clearly not able to perform the ratings exclusively on the basis of experiential frequency. This is not to deny, however, Gernsbacher's point that the inconsistent interactions observed between word frequency and other variables may be artifactual; objective word frequency count measures may not accurately reflect genuine experiential frequency, particularly for low-frequency words.

There are two features of the results that we wish to focus on here. The first concerns the *independent* effects

of the two different frequency measures, which indicate that the two different measures do represent psychologically distinct mechanisms. If the two frequency measures were simply reflecting the same underlying dimension, then they could not both have contributed significant independent variance, for any effect of the variable that was the less accurate measure of the assumed underlying dimension would have been redundant on the better measure, and would not have shown a significant effect when the influence of the better measure had been partialled out. Furthermore, it does not seem possible that other differences between the two counts could have given rise to the effects we observed. Given that a homogeneous population of raters was used to obtain the familiarity ratings, such differences (e.g., data and country of collection) could not have given rise to *independent* effects. Rather, differences other than modality could only have caused one of the measures to subsume all of the variance that would otherwise have been taken up by the other.

The second feature of the data that we discuss here is the fact that rated word age-of-acquisition is the major predictor of subjective familiarity judgment. Although the use of one subjective measure to predict another could be problematic, there is evidence for the validity of the AOA measure in a way that there is not for rated familiarity. Gilhooly and Gilhooly (1980) reported such results. In their first study, adults rated words taken from a standardized vocabulary test. The rank order of the vocabulary test words was based on age norms and was found to correlate highly with the adult ratings ($r = .93, p < .05$) (see also Lyons, Teer, & Rubenstein, 1978). In the second study reported by Gilhooly and Gilhooly, words that had been rated for AOA were given as a vocabulary test to groups of subjects whose ages ranged from 5 to 21 years. Objective AOA estimated from the test data was highly correlated with the AOA ratings ($r = .84, p < .05$). (Note that our correlations between familiarity and spoken and written frequency were only .50 and .37, respectively.) In both of the Gilhooly and Gilhooly (1980) studies, simultaneous multiple regression analyses indicated that rated AOA was the only independent predictor of the objective AOA indices, even when rated familiarity was included as a measure in their Experiment 2. Furthermore, the correlations of objective and subjective AOA with other word attribute variables were very similar. Note that the evidence we have obtained would prejudice the use of rated measures as an index of experiential frequency, even if the rated AOA measure is not a good index of actual AOA.

The fact that AOA, spoken word frequency, and written word frequency are all independently related to rated familiarity does not of course allow us to draw the conclusion that the same word variables would affect on-line lexical processing; rated familiarity might account for all the relevant variance in a reaction-time measure. This issue is addressed in the following experiment.

Table 3
Summary Multiple Regression Statistics for Analysis

Variable	Beta	F Ratio
Age-of-Acquisition	-.66	170.3*
Spoken Frequency	.29	37.6*
Written Frequency	.18	14.0*
Length	.11	7.5*
Concreteness	-.03	0.3
Ambiguity	-.04	1.5
Imagery	-.07	1.1

Note—Bigram frequency not entered ($F = .005$). * $p < .01$.

EXPERIMENT

Although the analysis reported above indicated that rated AOA is the main determinant of subjective familiarity ratings, Gilhooly and Logie (1981) found independent effects of AOA and familiarity in a word naming experiment. However, Gilhooly and Logie included no measure of word frequency in the spoken language, and, as we have shown above, spoken word frequency contributes independently to the familiarity variance. Thus the familiarity effect found by Gilhooly and Logie may disappear given a measure of spoken word frequency. Accordingly, one purpose of the present experiment is to examine the extent to which rated familiarity independently predicts word naming latency. In particular, we may be able to show that familiarity effects are redundant on a combination of AOA and spoken and written frequency measures.

A further hypothesis to be tested is that the effects of word learning age that have been obtained in word production tasks are redundant on spoken word frequency. Gilhooly and Watson (1981) reviewed the experimental literature and concluded that independent effects of AOA can only be found in tasks involving explicit word production. It is therefore possible that AOA in these experiments has been contributing variance that actually reflects an underlying influence of spoken word frequency.

The effect of spoken word frequency may be interesting in itself. If the locus of frequency effects in word naming tasks is a relatively peripheral output stage (Balota & Chumbley, 1985; Theios & Muise, 1977), then we predict that there will be no independent effect of written word frequency. If, on the other hand, frequency determines initial lexical access, then there should be no independent effect of production frequency. If there are two different loci to the frequency effect in word naming (see Balota & Chumbley, 1985), we might expect independent effects of both variables, unless all the frequency variance is more successfully captured by the rated familiarity measure.

Method

Subjects. Twenty-eight staff and students of the University of Sussex participated in the experiment. All had normal or corrected-to-normal vision, and all were right-handed. Subjects were paid for their participation in the experiment.

Materials and independent measures. The materials were those used in the above analysis. Rated familiarity was included as an independent measure in the present experiment. Thus the measures included written word frequency, spoken word frequency, concreteness, ambiguity, imageability, positional bigram frequency, rated AOA, word length, and rated familiarity. Rated pronounciability was not included as a variable, because such a measure might well be multiply determined and thus might subsume effects of purer variables. Presentation position was included as a variable, as was a measure designed to capture variance attributable to differing voice onset characteristics of various initial phonemes. This scale was constructed with the help of a trained phonetician. For each initial phoneme in our word sample, the length of time between offset of a vocalized consonant and energy onset (at a level detectable by

the voice key) of the subsequently uttered phoneme was measured. The resulting times covered a range of 80 msec and were used as an independent measure.

Procedure and apparatus. The experiment was run on-line using a microprocessor interfaced to a PDP-11/40 computer, and was controlled using the experimental program described by Norris (1984). Reaction times were recorded to the nearest millisecond. Each subject was seated in a darkened sound-attenuated cubicle facing a video display unit screen that was used for stimulus presentation. A microphone connected to a voice key was positioned in front of the subject. The experimenter used headphones to monitor responses for naming or pronunciation errors. Subjects were provided with written instructions for the task. The need to speak clearly and to name each word as fast as possible as soon as it appeared on the screen in front of them was emphasized.

The materials were divided into four experimental blocks. Presentation order was random. Ninety-five further items were selected from the same sources as the test items for use as practice items. A block of 87 items formed an initial practice block, and 2 practice items were presented at the beginning of each experimental block. Subjects rested between blocks of trials. There was an interval of 1,400 msec between a response and the onset of the subsequent item.

Analyses and Results

Trials that resulted in mispronunciations or accidental triggering of the voice key accounted for less than 1% of the data. These items were discarded. The data were averaged across subjects prior to analysis, since there were no subject-dependent variables, and there were sufficient averaged data to yield reliable regression results. A measure of reliability for the reaction times was calculated and indicated that they were indeed reliable across subjects (Cronbach's $\alpha = .82$). Three variables exhibited skews of greater than 1.0; skews in written and spoken frequency were reduced to 0.28 and 0.22 by a log transformation, and a skew in the bigram frequency measure was reduced to 0.76 by application of a square root transformation (see Table 1 for summary statistics).

The complete correlation matrix is presented as Table 4. Mean naming latency was most highly correlated with AOA ($r = .3$) and with rated familiarity ($r = -.28$). Thus spoken word frequency did not correlate more highly with latency than did either of these two variables, although spoken word frequency was more highly correlated with reaction time than was written word frequency.

A simultaneous multiple regression analysis was performed on the data to determine the independent contributions of each of the variables when the effects of the others were partialled out.

The multiple R for the final equation was .38; a significant proportion of the variance was accounted for [$F(11,404) = 6.04, p < .01$]. AOA displayed a significant independent effect, whereas the effects of rated familiarity, word length, and initial phoneme all just failed to reach significance at the conventional level. The effects of positional bigram frequency, concreteness, ambiguity, and the two word frequency measures were all clearly nonsignificant (all $F < 1.0$). A summary of the results is presented in Table 5.

Table 4
Correlation Matrix from Experiment

	2	3	4	5	6	7	8	9	10	11	12
1. Latency	-.11	.30	-.28	-.13	-.01	.05	.21	-.12	-.17	-.10	-.09
2. Imageability		-.65	.24	.85	.17	.10	.29	.02	-.13	-.18	.07
3. Age-of-Acquisition			-.58	-.58	.09	-.05	.51	-.13	-.12	.04	-.04
4. Familiarity				.21	-.06	.06	-.21	.17	.50	.37	-.01
5. Concreteness					-.02	.09	.30	.05	-.10	-.16	.11
6. Ambiguity						-.02	.14	.03	.11	.13	.14
7. Position							-.05	-.04	-.03	-.05	-.08
8. Length								.13	.06	.03	.04
9. Bigram Frequency									.22	.20	.00
10. Spoken Frequency										.70	.00
11. Written Frequency											-.01
12. Initial Phoneme											

Note—Decimal points have been omitted from the correlation matrix.

Discussion

We discuss the results of the experiment in terms of the contributions of word frequency and word familiarity before going on to discuss in more detail the theoretical implications of the AOA effect, which was clearly not redundant on spoken word frequency.

In our analysis the effect of rated familiarity approached significance, whereas measures of spoken and written word frequency did not approach significance. These results do suggest that in a task of this type the familiarity measure captures all the variance normally attributed to word frequency. However, because factors other than experiential frequency influence familiarity judgments, the subjective measure unfortunately cannot be used as an index of experiential frequency, since familiarity captures variance due to other factors as well.

Is it possible that effects that have been assumed in the past to reflect genuine experiential frequency are really due to word learning age? Given the uncertainty in the measures currently available, it is not clear how such a hypothesis could be disproved at present. Although the effect of rated familiarity in our experiment did not attain significance at the conventional level, other experiments have demonstrated significant independent effects of word familiarity and AOA (e.g., Gilhooly & Logie, 1981). But such results do not mean that genuine experiential frequency effects in word naming must exist, because any given effect of familiarity might not be due to the fre-

quency component of the familiarity measure. For example, independent effects of the rated familiarity measure might reflect some dimension, such as associative frequency, which is known to influence familiarity ratings (Rubin, 1983). The data are at least consistent with the possibility that pure word frequency effects do not occur in simple word naming, and that some other component of the familiarity measure, unrelated to frequency, independently predicts variance in word naming times.

GENERAL DISCUSSION

In previous sections of this paper, we have claimed that rated AOA is the best predictor of subjective rated familiarity and word naming latency. The rated measure is assumed to provide a valid index of an objective measure (Gilhooly & Gilhooly, 1980). In view of the importance of the AOA measure, we discuss in this section theoretical explanations of the AOA effects that have been observed in various lexical processing tasks.

It has been claimed that unambiguous AOA effects can be found only in tasks involving overt word naming (see Carroll & White, 1973; Gilhooly & Logie, 1981; Rubin, 1980; Whaley, 1978). Both Gilhooly and Watson (1981) and Watson (1985), on the basis of exhaustive reviews of the literature, have concluded that AOA effects occur reliably only in tasks that require overt word production. We take this as our starting point in the discussion that follows. Assuming that the source of the AOA effect is somewhere within the word production system, we consider three broad classes of explanation.

AOA-related factors. The first possible explanation is that other age-related factors might underlie the effects that have been observed. There are three main possibilities. The effects could be due to AOA per se, to word residence time, or to total lifespan frequency. By virtue of the correlations between these and other factors, certain interactions might be expected between such factors as AOA, frequency, and subject age. We take it that if residence time is the relevant variable, then AOA effects would be redundant on (subject age - AOA). If cumulative lifespan frequency is the underlying effect, then resi-

Table 5
Summary Multiple Regression Statistics for Experiment

Variable	Beta	F Ratio
Age-of-Acquisition	0.23	7.17*
Familiarity	-.013	3.35†
Length	0.10	3.02†
Initial Phoneme	-.09	3.51†
Imageability	0.15	2.31
Position	0.05	1.13
Bigram Frequency	-.04	0.7
Spoken Frequency	-.04	0.3
Concreteness	-.07	0.6
Ambiguity	0.02	0.2
Written Frequency	-.02	0.1

* $p < .01$. † $.05 < p < .10$.

dence time and AOA effects would be redundant on (subject age – AOA) \times frequency (assuming that subjects do not differ significantly in the number of words they hear/produce per unit time). Thus the ideal experiment would contain all these factors as independent measures and would examine their relative importance. Unfortunately the ideal experiment has never been carried out, owing to the lack of independence of the various measures. However, the patterns of interaction that have been obtained (Carroll & White, 1973; Gilhooly, 1984; Gilhooly & Gilhooly, 1979; Lachman, Shaffer, & Hennrikus, 1974) are not consistent with any of the AOA-related factors giving rise to apparent effects of AOA per se (Watson, 1985). It seems that the effects are probably due to the age at which a word is learned, or to some independent variable correlated with AOA.

Are some words learned early because they are easily pronounced? The second possibility we wish to consider is that language has evolved to make early-learned words easy to pronounce, and that this influences order of word learning by children (Farwell, 1976). The only certain way to distinguish the two possible directions of causality, apart from a diachronic study, would be to obtain some independent measure of physical difficulty of articulation, perhaps in terms of characteristics of the vocal musculature and so forth. Correlations between AOA and rated pronunciability cannot inform us about the direction of causality, owing to the uncertainty of the attributes that would affect the rated pronunciability variable. Although there does appear to be a fixed developmental order to acquisition of phonology (Jakobson, 1968; Menn, 1983), this cannot reflect any very peripheral difficulty, since children use a much wider range of sound in the “babbling” stage before acquiring any words at all. Smith (1973) reports *puggle* for target *puddle* at the same stage of development as *puddle* for target *puzzle*. This kind of evidence suggests that phonological development is not constrained by the structure of the articulatory apparatus, but reflects some principle of developing internal organization (Menn, 1983). Thus, we reject the possibility that purely *articulatory* difficulty might be a causal factor in determining word learning age. The suggestion that phonological difficulty (of representation) partly determines AOA also appears to be excluded by the data, for children acquire words before they are able to produce them accurately. Indeed, it seems clear that children represent more distinctions for perceptual purposes than they represent for production purposes (e.g., Morton & Smith, 1974). It seems unlikely, in conclusion, that words are learned early because they are easy to say, rather than vice versa. In the absence of further evidence on this issue, we turn to the psychologically more interesting hypothesis that early-learned words are easier to say because they are learned early.

Are some words easily pronounced because they are learned early? If AOA does determine pronunciability, there are several ways in which this might happen. Likely

hypotheses concern the quality or availability of various phonological representations.

The phonological output lexicon appears to be a possible source of AOA effects. This lexicon is phonologically organized, and, when access to it fails, word substitution errors will occur (Fay & Cutler, 1977). These errors will be phonologically related to the target. Ease of mapping onto the output lexicon cannot be affected by target-word AOA, for if the early AOA words in the phonological output lexicon were more available, then speech errors of this type would be words that were learned earlier than the target words. Fay and Cutler (1977) have shown that errors do not have a higher frequency than the targets they replace, and, given the high correlation between AOA and word frequency, we can therefore exclude the hypothesis that early-learned words have more *accessible* representations in the phonological store. However, it is possible that the *quality* of phonological information in the store differs for early acquired words. We term this the “completeness hypothesis.” Some evidence for this hypothesis is provided by Aitchison and Straf (1982) in a study of children’s malapropisms. They found that children, in their storage of word phonologies, accord relatively less weight to the maximally informative beginnings of words, suggesting that their relatively undeveloped phonological storage systems are less economical, and make less use of redundancy in word endings. If this is so, then a more complete phonological representation might be permanently available for early acquired words.

When later acquired words must be pronounced, then, more time might have to be spent in generating phonological information not directly represented in the phonological lexicon, and it is suggested here that this might be the cause of AOA effects in adults. The completeness hypothesis assumes that, for example, a common word ending will be represented in the phonological output lexicon by a single, more abstract symbol, which can be expanded via look-up procedures into a full phonological representation prior to further processing. Phonetic accommodation, for example, will take place at later stages. Words are stored in a relatively complete form by young children, and these representations remain complete throughout adult life.

There is other evidence that children have difficulty with developing phonological storage. Vihman (1981) has discussed children’s tendency to use homonymy (one word is used by the child to stand for several adult words), suggesting storage limitations. Ferguson and Farwell (1975) have suggested that children learn words as wholes, without translating them into smaller phonetic units. Menn (1983) has discussed the use of systematic simplification by children.

In conclusion, we suggest that the effects of rated AOA in tasks that require overt word naming reflect the fact that the phonological output representations are stored in a relatively complete form during the early stages of vocabulary acquisition. As vocabulary size increases,

storage limitations require more efficient strategies, and only minimal information is stored explicitly. Although this would represent a great saving in storage capacity, there is a processing cost, as the phonological information that is not directly represented has to be generated whenever required. This takes time, and this is reflected in word naming latency.

Regarding the effect of AOA on familiarity ratings, it seems entirely possible that subjects might use the quality of phonological lexical representation available to them as an index of how frequently they have encountered that word. Assuming that subjects do not have direct access to frequency information, they may perform the familiarity ratings in a number of different ways. Some variable, such as number or richness of semantic associations, might perhaps be the relevant variable (Galbraith & Underwood, 1973; Rubin, 1983), for such a dimension would be highly correlated with actual AOA. Although semantic variables are assumed not to be responsible for the AOA effect we have observed on word naming time, because such variables do not in general have strong effects on such low-level tasks, such semantic variables may well affect familiarity ratings. The present data do not enable us to draw a definite conclusion on this issue.

Several important conclusions follow from the results reported in the present paper. Gernsbacher's (1984) suggestion that familiarity ratings be used as a measure of experiential frequency was not supported. Familiarity ratings are made partly on the basis of other word characteristics, most particularly word learning age.

We have also shown that AOA is a better predictor of word naming latency than is either familiarity or objective word frequency, and that the AOA effects that have been previously reported in the literature are not redundant on spoken word frequency.

Spoken and written word frequency have independent effects on rated familiarity, and this suggests that they influence different psychological stages of processing and are not simply different indices of one underlying dimension. Further research is needed to evaluate the respective contributions of these two variables in various tasks.

Finally, it has been suggested that early-learned words are given a more complete representation in a phonological output lexicon and that, as a result of this, the pronunciations of these words can be synthesized more rapidly throughout later life.

REFERENCES

- AITCHISON, J., & STRAF, M. (1982). Lexical storage and retrieval: A developing skill? In A. Cutler (Ed.), *Slips of the tongue and language production* (pp. 197-241). Amsterdam: Mouton.
- BALOTA, D. A., & CHUMBLEY, J. I. (1984). Are lexical decisions a good measure of lexical access? The role of word frequency in the neglected decision stage. *Journal of Experimental Psychology: Human Perception & Performance*, **10**, 340-357.
- BALOTA, D. A., & CHUMBLEY, J. I. (1985). The locus of word frequency effects in the pronunciation task: Lexical access and/or production? *Journal of Memory & Language*, **24**, 89-106.
- BROWN, G. D. A. (1984). A frequency count of 190,000 words in the *London-Lund Corpus of English Conversation*. *Journal of Behavior Research Methods, Instruments, & Computers*, **16**, 502-532.
- CARROLL, J. B., & WHITE, M. N. (1973). Word frequency and age of acquisition as determiners of picture-naming latency. *Quarterly Journal of Experimental Psychology*, **25**, 85-95.
- CUTLER, A. (1981). Making up materials is a confounded nuisance, or: Will we be able to run any psycholinguistic experiments at all in 1990? *Cognition*, **10**, 65-70.
- FARWELL, C. B. (1976). *Some strategies in the early production of fricatives*. Stanford, CA: Stanford University Linguistics Department.
- FAY, D., & CUTLER, A. (1977). Malapropisms and the structure of the mental lexicon. *Linguistic Inquiry*, **8**, 505-520.
- FERGUSON, C. A., & FARWELL, C. B. (1975). Words and sounds in early language acquisition. *Language*, **51**, 431-439.
- GALBRAITH, R. C., & UNDERWOOD, B. J. (1973). Perceived frequency of abstract and concrete words. *Memory & Cognition*, **1**, 56-69.
- GERNSBACHER, M. A. (1984). Resolving 20 years of inconsistent interactions between lexical familiarity and orthography, concreteness, and polysemy. *Journal of Experimental Psychology: General*, **113**, 256-281.
- GILHOOLY, K. J. (1984). Word age-of-acquisition and residence time in lexical memory as factors in word naming. *Current Psychological Research & Reviews*, **3**, 24-31.
- GILHOOLY, K. J., & GILHOOLY, M. L. M. (1979). Age-of-acquisition effects in lexical and episodic memory tasks. *Memory & Cognition*, **7**, 214-223.
- GILHOOLY, K. J., & GILHOOLY, M. L. M. (1980). The validity of age-of-acquisition ratings. *British Journal of Psychology*, **71**, 105-110.
- GILHOOLY, K. J., & LOGIE, R. H. (1980). Age-of-acquisition, imagery, concreteness, familiarity, and ambiguity measures for 1944 words. *Behavior Research Methods & Instrumentation*, **12**, 395-427.
- GILHOOLY, K. J., & LOGIE, R. H. (1981). Word age-of-acquisition, reading latencies and auditory recognition. *Current Psychological Research*, **1**, 251-262.
- GILHOOLY, K. J., & WATSON, F. L. (1981). Word age-of-acquisition effects: A review. *Current Psychological Reviews*, **1**, 269-286.
- GORDON, B. (1985). Subjective frequency and the lexical decision latency function: Implications for mechanisms of lexical access. *Journal of Memory & Language*, **24**, 631-645.
- HOWES, D. (1954). On the interpretation of word frequency as a variable affecting speed of recognition. *Journal of Experimental Psychology*, **48**, 106-112.
- HUMPHREYS, G. W., EVETT, L. J., & TAYLOR, D. E. (1982). Automatic phonological priming in visual word recognition. *Memory & Cognition*, **10**, 576-590.
- JAKOBSON, R. (1968). *Child language aphasia and phonological universals*. The Hague: Mouton.
- KUCERA, H., & FRANCIS, W. (1967). *Computational analysis of present-day American English*. Providence, RI: Brown University Press.
- LACHMAN, R., SHAFFER, J. P., & HENNRICKUS, D. (1974). Language and cognition: Effects of stimulus codability, name-word frequency and age of acquisition on lexical reaction time. *Journal of Verbal Learning & Verbal Behavior*, **13**, 613-625.
- LYONS, A. W., TEER, P., & RUBENSTEIN, H. (1978). Age-at-acquisition and word recognition. *Journal of Psycholinguistic Research*, **7**, 179-187.
- MENN, L. (1983). Development of speaking skills. In B. Butterworth (Ed.), *Language production 2*. London: Academic Press.
- MORTON, J., & SMITH, N. V. (1974). Some ideas concerning the acquisition of phonology. In *Current problems in psycholinguistics* (pp. 161-176). Paris: Editions du CRNS.
- NORRIS, D. (1984). A computer-based programmable tachistoscope for nonprogrammers. *Behavior Research Methods, Instruments, & Computers*, **16**, 25-27.
- RUBIN, D. C. (1976). Applying psychometric methods in linguistic research: Some recent advances. *Linguistics*, **168**, 63-66.
- RUBIN, D. C. (1980). 51 properties of 125 words: A unit analysis of verbal behavior. *Journal of Verbal Learning & Verbal Behavior*, **19**, 736-755.

- RUBIN, D. C. (1983). Associative asymmetry, availability, and retrieval. *Memory & Cognition*, **11**, 83-92.
- SMITH, N. V. (1973). *The acquisition of phonology: A case study*. Cambridge: Cambridge University Press.
- SOLSO, R. L., & JUEL, C. L. (1980). Positional frequency and versatility of bigrams for two- through nine-letter English words. *Behavior Research Methods & Instrumentation*, **12**, 297-343.
- SVARTVIK, J., & QUIRK, R. (1980). *A corpus of English conversation*. Lund, Sweden: Gleerup.
- THEIOS, J., & MUISE, J. G. (1977). The word identification process in reading. In N. J. Castellan, D. B. Pisoni, & G. R. Potts (Eds.), *Cognitive theory*. Hillsdale, NJ: Erlbaum.
- THORNDIKE, E. L., & LORGE, I. (1944). *The teacher's word book of 30,000 words*. New York: Teachers College Press, Columbia University.
- TRYK, H. E. (1968). Subjective scaling of word frequency. *American Journal of Psychology*, **81**, 170-177.
- VIHMAN, M. M. (1981). Phonology and the development of the lexicon: Evidence from children's errors. *Journal of Child Language*, **8**, 239-264.
- WATSON, F. L. (1985). *Age of acquisition effects in lexical processing*. Unpublished doctoral dissertation, University of Aberdeen, Scotland.
- WHALEY, C. P. (1978). Word-nonword classification time. *Journal of Verbal Learning & Verbal Behavior*, **17**, 143-154.

(Manuscript received February 26, 1985;
revision accepted for publication October 20, 1986.)