Subjective frequency estimates
for 2,938 monosyllabic words

DAVID A. BALOTA, MAURA PILOTTI, and MICHAEL J. CORTESE
Washington University, St. Louis, Missouri

Subjective frequency estimates for large sample of monosyllabic English words were collected from 574 young adults (undergraduate students) and from a separate group of 1,590 adults of varying ages and educational backgrounds. Estimates from the latter group were collected via the internet. In addition, 90 healthy older adults provided estimates for a random sample of 480 of these words. All groups rated words with respect to the estimated frequency of encounters of each word on a 7-point scale, ranging from never encountered to encountered several times a day. The young and older groups also rated each word with respect to the frequency of encounters in different perceptual domains (e.g., reading, hearing, writing, or speaking). The results of regression analyses indicated that objective log frequency and meaningfulness accounted for most of the variance in subjective frequency estimates, whereas neighborhood size accounted for the least amount of variance in the ratings. The predictive power of log frequency and meaningfulness were dependent on the level of subjective frequency estimates. Meaningfulness was a better predictor of subjective frequency for uncommon words, whereas log frequency was a better predictor of subjective frequency for common words. Our discussion focuses on the utility of subjective frequency estimates compared with other estimates of familiarity. The raw subjective frequency data for all words are available at http://www.artsci.wustl.edu/~dbalota/labpub.html.

One might argue that no single variable has been studied more in psycholinguistics and memory research than word frequency. Information about the frequency of occurrence of words in the language has been gathered from either objective frequency counts or subjective familiarity ratings. Objective frequency counts typically include extensive collections of written samples of word usage. For example, the Brown corpus consists of 1,014,000 words of text (Kučera & Francis, 1967), whereas the Celex corpus (Dutch Center for Lexical Information, 1995) includes approximately 17,900,000 items—a large portion of which is from written sources (16,600,000). These counts presumably provide an objective estimate of the average frequency of exposure (and hence an estimate of familiarity) to words in a given language.

A number of researchers have noted some problems with objective frequency counts. For example, the counts are primarily taken from written materials. Thus, it is possible that the sampling is quite different from those of estimates of spoken, heard, or written frequency. As Gernsbacher (1984) pointed out, based on normative data, the

words *boxer*, *icing*, and *joker* have the same frequency count as *loire*, *gnome*, and *assay* [all being 1 per million according to the Kučera & Francis (1967) norms]. Moreover, the words *hype* and *babe* are not in either the Kučera and Francis or the much larger and more recent Celex (Dutch Center for Lexical Information, 1995) norms.

Because subjective ratings of familiarity are unlikely to be affected by the sampling biases that plague objective ratings, Gernsbacher and others (e.g., Gilhooly & Logie, 1980) have argued that subjective familiarity ratings might serve as a better reflection of the relative frequency of exposure to a word. These ratings are typically untimed estimates of the familiarity of individual words. For example, Gernsbacher presented the following instructions to her subjects:

> On the following pages is a list of words. Some of these words will be words that you know and use very often; that is, you will be VERY FAMILIAR with these words. Others will be words that you might not have ever seen before and surely do not recognize. Thus, you are VERY UNFAMILIAR with these words. Of course, there will also be words that can fall within those two extremes. Some words you might recognize, but not use or hear very often; you are familiar with them. Others you might just slightly recognize, perhaps you have seen them a few times and thus are SLIGHTLY FAMILIAR to you. And conversely, there will be words that you barely recognize, SLIGHTLY UNFAMILIAR, or words that you can recognize but perhaps have only heard or read once or twice, UNFAMILIAR.

However, just as in the case of objective frequency estimates, there is also some concern about rated familiarity estimates. Although Gernsbacher's (1984) instructions

attempted to provide considerable direction to the participants, it is possible that the participants not only rated the printed string for familiarity, but may also have included other variables such as the availability of a clear and vivid meaning for the word or possibly the familiarity of sublexical spelling to sound correspondence. The critical concern is that additional sources of information are available to the participant in an untimed rating of familiarity.

Because of this potential difficulty with subjective familiarity estimates, one must be very careful in using familiarity (as typically measured) as a way to estimate frequency of the stimulus item. For example, as Balota, Ferraro, and Conner (1991) point out, it is quite possible for familiarity ratings to mask effects of other variables. Imagine an experiment in which stimuli produce an effect of meaningfulness, with words that have many meanings producing shorter response latencies compared with words that have few meanings. A careful researcher may demonstrate that this effect is eliminated when familiarity is controlled (see, e.g., Gernsbacher, 1984). However, because meaningfulness may be accessed by participants when they make their untimed estimates of familiarity, one might eliminate the effect of meaningfulness via controlling for familiarity. In fact, Toglia and Battig (1978) found that their familiarity ratings were related to four different measures of semantic attributes of selected words. Among these, a measure of meaningfulness (an index of a word's semantic associations with other words) was the best predictor of their subjective familiarity scores ($r = .82$).

In addition to semantic variables, the orthographic/ phonological patterns present in common words tend to be more frequent overall (Coltheart, Davelaar, Jonasson, & Besner, 1977; Landauer & Streeter, 1973) and may also contribute to a person's sense of familiarity. Therefore, it is possible that participants will, in part, rely on how frequently they have encountered the sublexical components of an orthographic/phonological pattern (a word) in making their untimed familiarity ratings. This is particularly important because, as in the example of meaningfulness presented above, if participants allow orthographic similarity to other words to influence their familiarity ratings, then, by controlling stimuli for familiarity, they might inadvertently eliminate the effect of spelling to sound correspondence. Interestingly, Peereman, Content, and Bonin (1998) recently demonstrated that some aspects of sound to spelling relations, as reflected by feedback consistency measures, may in fact be eliminated in lexical decision performance when standard familiarity ratings are controlled. This is precisely why one needs to provide evidence concerning the type of information that contributes to familiarity ratings.

Subjective frequency estimates may also be age dependent. For instance, it is possible that the subjective ratings collected from young adults (usually college students) may be different from the ratings observed for older adults. This question becomes particularly important in light of the usually superior performance of older adults in vocabulary tests (see O'Dowd, 1984) and their greater reliance on word frequency (as measured by objective frequency counts) in psycholinguistic tasks such as naming and lexical decision (see Balota & Ferraro, 1996; Spieler & Balota, 2000). Unfortunately, to our knowledge, studies have yet to examine the influence of age on subjective frequency of occurrence estimates.

In light of these unresolved issues, we collected subjective frequency estimates for a large corpus of monosyllabic words. The corpus was based on the stimulus materials that have been used to develop computational models of word identification (see Plaut, McClelland, Seidenberg, & Patterson, 1996; Seidenberg & McClelland, 1989) and to examine word identification processes in psycholinguistic experiments (see Balota & Spieler, 1999; Spieler & Balota, 1997). We attempted to avoid the potential ambiguity with "familiarity" ratings, by explicitly requesting that participants provide information about how frequently they have encountered a word. In this way, we hoped to minimize the potential influence of additional sources of information that may come into play with familiarity ratings (e.g., semantic and/or orthographic/ phonological). Thus, we had the participants rate words on a relatively neutral frequency-of-exposure 7-point scale (1 = never, 2 = once a year, 3 = once a month, 4 = once a week, 5 = once every two days, 6 = once a day, 7 = several times a day). In addition to these overall ratings, we also asked the participants to rate each word on the basis of the frequency with which they had seen, heard, written, and spoken that word. These source-specific ratings served as a measure for assessing whether the participants could discriminate among different modalities of presentation. Finally, in order to examine whether subjective ratings vary substantially across different participant groups, we administered the frequency rating task to three groups: young adults (college students), healthy older adults (age range: 65–85), and a large sample of adults of mixed ages and backgrounds, who participated in the experiment by accessing a web page specifically designed for this task. For technical reasons, the latter sample did not rate the words as a function of modality of presentation.

METHOD

Participants

Three groups of participants took part in the present study. First, 574 undergraduates were recruited from Washington University (mean age = 19.5 years, range 17–29 years). Second, 90 healthy older adults (mean age = 73.4 years, range 61–80 years) were recruited from the Aging and Development subject pool at Washington University. The older adults performed better ($p < .001$) than the young adults on the Shipley vocabulary test (Shipley, 1940). Third, 1,590 individuals provided ratings via a web page specifically designed for this task (mean age = 40.4 years, range 14–84 years). Table 1 shows the percentage of individuals at each education level for the web-based sample and for the older adult group. As shown in Table 1, the web-based individuals reported being more highly educated than the healthy older adults. (Although we did not collect this information for the young adults, the vast majority of these individuals would fit the category *some college education but without a degree*.) All participants reported English as their first language.

Table 1
Percentage of Individuals at Each Education Level

Education Level	Healthy Older Adults	Web Subjects
No high school	4.4%	.4%
Some high school	46.7%	2.1%
High school graduate	10.0%	13.4%
Some college	32.2%	38.5%
College graduate	6.7%	24.3%
Some graduate school	0.0%	6.9%
Completed graduate degree	0.0%	11.1%
Not listed	0.0%	2.9%
Declined to answer	0.0%	0.4%

Stimulus Materials and Procedure

The stimuli were 2,938 monosyllabic words. These words involved a wide range of frequencies: from 0 to 69,971 counts per million, according to the Kučera and Francis' (1967) objective frequency norms, and from 1 to 62,218 per million, according to the Celex (Dutch Center for Lexical Information, 1995) norms. The words ranged from two to eight letters in length.

The stimuli for the Washington University undergraduates were printed in random order and assembled into 17 sets of 160 words and 2 sets of 109 words. Each of the 160-word sets was divided into 5 subsets of 32 words and the remaining sets were divided into 4 subsets that varied in length from 27 to 28 words. To minimize order effects, a Latin square design was used to order the subsets of each set. This procedure produced a total of 85 booklets of 160 words and 8 booklets of 109 words. Each of the 85 booklets of 160 words was given to 6 participants, and each of the eight booklets of 109 words was given to 8 participants. Therefore, each word in the stimulus set was rated by either 30 or 32 participants, and the order of the sets of items was counterbalanced across participants.

The stimuli for the web-based participant pool included a slightly smaller number of items (2,824 words). The words were randomly organized and assembled into 28 sets of 100 and one set of 24. Each set of 100 words was then divided into independent subsets of 10 words. Each word was rated by 40–69 participants, and each participant rated 100 words, with the only exception being that one list included only 24 words. Harris Interactive implemented the web-based program and collected the data.

All participants were given the following instructions:

> Words differ in how commonly or frequently they have been encountered. Some words are encountered very frequently, whereas other words are encountered infrequently. The purpose of this study is to rate a list of 160 words with respect to frequency. We believe that your ratings will be important to future studies involving word recognition. You should base your ratings according to the following 7-point scale: 1 = never, 2 = once a year, 3 = once a month, 4 = once a week, 5 = every two days, 6 = once a day, 7 = several times a day.

In addition, the healthy young and older adults also rated each word on four additional dimensions (i.e., read, heard, wrote, or said). These individuals were given the additional following instruction: "Because words can be encountered in many different ways, we ask you to provide five different ratings for each word." The young participants reported their answers on scan forms, whereas the older adults were also given the option of reporting their answers directly on the test sheets. The young participants were tested in groups of 7–40, and older participants performed the task at home. Older adults rated a subset of 480 words randomly selected from the larger corpus.

RESULTS

Data for all the individual words are available at http://www.artsci.wustl.edu/~dbalota/labpub.html.

The mean ratings and *SD*s for each of the questions for the young group and the overall estimate from the web-based group are displayed in Table 2. As shown, there are large differences for the young group in the reported frequency of encounters for the various perceptual domains of reading, hearing, writing, and saying. Interestingly, the participants appeared to match the likelihood of hearing a word to the overall likelihood of encountering a word (3.87 vs. 3.88), even though the words were presented visually. This is somewhat surprising because one might expect the total number of times encountered to be the sum of the reading, hearing, writing, and saying estimates, but this was clearly not the case. In addition, the participants, on average, rated reading and saying the word higher than writing it. Individuals rated the frequency of encounters in writing as less likely than all other domains (all ps < .001). Finally, one should note that the web-based sample had relatively higher ratings for overall frequency of encountering words [$t(958) = 3.23$, $p < .001$], possibly due to the participants' relatively high self-reported education levels (see Table 1). It is also possible that the presence of lower ratings for the written, said, and read estimates for the young adults lowered their perceived overall encounters compared with the web-based group, who did not make these ratings. Although reliably higher frequency of encounter estimates were given by the web-based group, as discussed below, there was considerable similarity across the two samples.

Table 3 presents the mean ratings for each of the questions for the young adults, the older adults, and the web-based sample for the same set of 480 words. It is noteworthy that the older adults' ratings for the written domain were lower than those of young adults [$t(958) = 6.33$, $p < .001$]. The opposite was true for the read dimension [$t(958) = 5.57$, $p < .001$], indicating that there was an age difference in perception of encounters in the different modalities. Although the difference in the written domain could be attributed to the fact that the younger adults were college undergraduates and therefore were more likely to have many writing assignments, this does not account for the opposite effect for the reading domain. Interestingly, there were no group differences either in the remaining source-specific ratings or in the overall subjective frequency measures (ts < 1.5). These findings indicate that, although there were age–group differences in the extent to which participants reported experiencing words in some perceptual domains, the participants were able to integrate these experiences into an

Table 2
Mean Ratings and Standard Deviations
for Young Adults and the Web-Based Sample

Domain	Number of Items	Mean Rating	SD
Young adults			
Read	2,938	3.41	1.06
Heard	2,938	3.87	1.24
Written	2,938	2.78	1.06
Said	2,938	3.59	1.32
Encountered	2,938	3.88	1.26
Web-based sample			
Encountered	2,824	4.16	1.29

Table 3
Mean Ratings and Standard Deviations for Young Adults,
Older Adults, and the Web-Based Sample
for the Same Set of Words

Domain	Number of Items	Mean Rating	SD
Young Adults			
Read	480	3.55	1.03
Heard	480	3.99	1.19
Written	480	2.86	1.01
Said	480	3.74	1.27
Encountered	480	4.01	1.22
Older Adults			
Read	480	3.91	0.98
Heard	480	4.09	1.01
Written	480	2.50	0.73
Said	480	3.73	1.09
Encountered	480	4.05	1.06
Web-Based Sample			
Encountered	480	4.27	1.26

overall dimension of word frequency, which remained largely invariant across the two age groups. The overall ratings of familiarity produced by the web-based adult group were reliably higher than those yielded by the young adults [$t(958) = 3.23$, $p < .001$] or the older adults [$t(958) = 2.91, p < .004$]. Although the web-based group produced a slightly higher overall rating, it should be noted that the mean frequency of encounter estimates for each word were highly correlated between groups (young and old, $r = .92$, young and web-based, $r = .95$, and old and web-based, $r = .93$).

We will now attempt to provide some insight into the types of information the participants used to make their frequency of encounter estimates. First, we consider the predictive power of log frequency taken from both the Kučera and Francis (1967) and the Celex (Dutch Center for Lexical Information, 1995) norms. The computation of log frequency was based on both the Kučera and Francis and the Celex corpora in order to obtain an estimate of the contribution of objective word frequency to subjective frequency of encounter ratings. As noted earlier, although the Kučera and Francis norms are still the most common set of norms used in memory and psycholinguistic research, they are relatively old and are based on a much smaller sample than the Celex norms. Therefore, one may argue that the incidence of words in the Kučera and Francis corpus may not be a good representation of a participant's linguistic experience (see Burgess & Livesay, 1998). Second, we consider the predictive power of a widely cited set of subjective familiarity ratings taken from the Toglia and Battig (1978) norms. Toglia and Battig's familiarity ratings were used to address whether standard familiarity ratings might produce a different pattern of results than the current frequency of exposure ratings. It is noteworthy that in Toglia and Battig's instructions, participants were told to regard the word *person* as highly familiar and *amorphous* as unfamiliar, whereas our instructions did not provide any example of word familiar-

ity. Because Toglia and Battig's examples of word familiarity differ greatly in meaning, it is possible that these examples may have encouraged participants to rely more heavily on the semantic attributes of the stimulus material than did our neutral instructions. Third, we consider an orthographic similarity index (Coltheart's orthographic N, which is the number of words or "neighbors" of the same length that can be formed by changing only one letter—e.g., the word *bent* has *bend*, *cent*, *dent*, *lent*, *belt*, as neighbors). This measure reflects the extent to which subjective frequency estimates could be driven by the frequency of orthographic neighbors—a factor that has been central to models of word recognition (see Andrews, 1997, for a review). Finally, we included a semantic estimate (i.e., *meaningfulness*, taken from the Toglia & Battig, 1978, norms) in order to determine whether the meaning of the stimulus word would influence the estimates of encounters.

The intercorrelations among these factors and the overall estimates of frequency of encounters are displayed in Table 4. The top panel displays the correlations among the factors based on all the items that each group rated (i.e., 2,938 for the young adults, 2,824 for the web-based sample, and 480 for the older adults). The bottom panel displays the correlations among the factors for the same 480 words that all three groups rated.

There are six points to note with regard to Table 4. (1) The pattern of correlations is quite consistent between the top and bottom panels, suggesting that these results are stable across different sample sizes. (2) The present subjective frequency estimates and the standard familiarity ratings of Toglia and Battig (1978) were related, but clearly were not identical (accounting for 42%–58% of the variance). This suggests that the present subjective frequency of encounter estimates appear to tap into a different type of information than do Toglia and Battig's more standard familiarity ratings. (3) The correlations between the present subjective frequency estimates and the objective estimates taken from Kučera and Francis (1967) and Celex (Dutch Center for Lexical Information, 1995) are consistently larger (ranging between .78 and .83) than the correlations between Toglia and Battig's subjective familiarity ratings and these same objective frequency estimates (ranging between .53 to .62). This suggests that the present norms may tap into a more accurate representation of frequency of occurrence. (4) The present frequency of encounter estimates were more strongly related to the estimates of log frequency than to either meaningfulness or neighborhood size. In contrast, Toglia and Battig's familiarity ratings were equally related to meaningfulness (.60) and log frequency (.58 and .62 for Kučera & Francis and Celex, respectively). (5) The ratings of the mixed adult group, albeit larger in magnitude than those of the young and older adult groups, appeared to rely on the same sources of information, as reflected by the similar pattern of correlations. (6) Although the Celex norms produced slightly larger correlations with the present frequency of encounter estimates than did the

Table 4
Intercorrelations Among the Selected Factors for
Young Adults, Older Adults, and Web-Based Groups

	SubF-Young	SubF-Old	SubF-Web	OLF-KF	OLF-Cel	NSize	M-T&B
			All Items per Group				
OLF-KF	0.78*	0.79*	0.78*				
OLF-Cel	0.81*	0.83*	0.81*	0.95*			
NSize	0.14*	0.18*	0.15*	0.16*	0.17*		
M-T&B	0.29*	0.16*	0.28*	0.09*	0.14*	−0.01	
Fam-T&B	0.76*	0.65*	0.75*	0.58*	0.62*	0.04	0.60*
			Same 480 Words Across Groups				
OLF-KF	0.79*	0.79*	0.78*				
OLF-Cel	0.83*	0.83*	0.82*	0.95*			
NSize	0.16*	0.18*	0.18*	0.24*	0.20*		
M-T&B	0.17*	0.16*	0.13	0.09	0.06	0.04	
Fam-T&B	0.70*	0.65*	0.65*	0.53*	0.59*	0.11	0.50*

Note—*$p < .05$. SubF = Subjective Frequency estimates from current study. OLF-KF = Objective log frequency based on Kučera and Francis (1967). OLF-Cel = Objective log frequency based on Celex (Dutch Center for Lexical Information, 1995). NSize = Neighborhood size. M-T&B = Meaningfulness based on Toglia and Battig (1978). Fam-T&B = Familiarity ratings based on Toglia and Battig (1978).

Kučera and Francis norms, these two sets of norms were strongly correlated (.95).

To clarify the predictive power of each of the selected factors, we conducted a set of multiple regression analyses with the overall frequency of encounter estimates as the dependent measure and Toglia and Battig's (1978) familiarity measure, log frequency (Kučera & Francis, 1967, corpora), meaningfulness, and neighborhood size as predictor variables [note that the same analyses conducted with the log frequency measure of the Celex (Dutch Center for Lexical Information, 1995) corpus yielded identical results]. Table 5 includes all the data available to each group, whereas Table 6 displays the results from the young and the web-based sample for the same set of items that was included in the older adult sample.[1] In addition to these analyses, the bottoms of Tables 5 and 6 display the predictive power of each of the factors on the Toglia and Battig familiarity ratings.

As shown in these tables, objective frequency estimates from Kučera and Francis (1967) were better predictors of the present subjective frequency of encounter ratings than was meaningfulness, and the predictive power of neighborhood size was negligible across all the participant groups. Interestingly, as is shown at the bottom of Table 5, both log frequency and meaningfulness were reliable predictors of Toglia and Battig's (1978) familiarity ratings, whereas neighborhood size remained a negligible factor.

It is possible that the relatively negligible contribution of neighborhood size to subjective frequency estimates could be attributed to the overall familiarity of words in a participant's mental lexicon. As discussed earlier, participants may be prone to rely on orthographic similarity to other words stored in the lexicon when the target word is relatively less well known. In order to examine this issue, we collapsed across the frequency of encounter estimates for both the young and the mixed adult groups and computed the mean subjective frequency estimates

for each of the 2,824 items. (Because of the small number of subjective frequency estimates in the older adult group and the small number of low familiarity words in Toglia & Battig's, 1978, set, we did not include these data in these analyses.) We then divided the data into three sets of words on the basis of their mean frequency of en-

Table 5
Simultaneous Regression Analyses for the Subjective
Frequency Estimates and Toglia and Battig (1978)
Familiarity Estimates as a Function of Subject Group

Predictors	Beta	t value	$p <$	Semipartial
	Young Adults			
Log frequency (Kučera & Francis, 1967)	**.79**	**48.36**	**.001**	**.81**
Neighborhood size	−.01	−.56	n.s.	−.02
Meaningfulness Toglia & Battig (1978)	**.19**	**11.67**	**.001**	.32
$R^2 = .68$	$F(3,1219) = 875.90, p < .001$			
	Older Adults			
Log frequency (Kučera & Francis, 1967)	**.78**	**17.86**	**.001**	**.78**
Neighborhood size	.03	−59	n.s.	.04
Meaningfulness (Toglia & Battig, 1978)	**.12**	**2.78**	**.006**	**.19**
$R^2 = .64$	$F(3,204) = 118.64, p < .001$			
	Web-Based Sample			
Log frequency (Kučera & Francis, 1967)	**.77**	**44.62**	**.001**	**.79**
Neighborhood size	.02	.88	n.s.	.03
Meaningfulness (Toglia & Battig, 1978)	**.19**	**11.39**	**.001**	**.31**
$R^2 = .65$	$F(3,1205) = 756.26, p < .001$			
	Toglia and Battig's Sample			
Log frequency (Kučera & Francis, 1967)	**.54**	**30.37**	**.001**	**.66**
Neighborhood size	−.02	−1.23	n.s.	−.04
Meaningfulness (Toglia & Battig, 1978)	**.54**	**30.61**	**.001**	**.66**
$R^2 = .63$	$F(3,1219) = 683.69, p < .001$			

Table 6
Simultaneous Regression Analyses for Subjective Frequency Estimates and Toglia and Battig's (1978) Familiarity Estimates as a Function of Subject Group

Predictors	Beta	t value	p<	Semipartial
	Young Adults			
Log frequency (Kučera & Francis, 1967)	.82	19.91	.001	.81
Neighborhood size	−.01	−.17	n.s.	−.01
Meaningfulness (Toglia & Battig, 1978)	.12	3.15	.002	.13
$R^2 = .68$	$F(3,204) = 144.38, p < .001$			
	Web-Based Sample			
Log frequency (Kučera & Francis, 1967)	.80	18.95	.001	.80
Neighborhood size	.02	.53	n.s.	.04
Meaningfulness (Toglia & Battig, 1978)	.09	2.21	.029	.15
$R^2 = .66$	$F(3,204) = 131.75, p < .001$			
	Toglia and Battig's Sample			
Log frequency (Kučera & Francis, 1967)	.53	10.42	.001	.59
Neighborhood size	−.03	−.52	n.s.	−.04
Meaningfulness (Toglia & Battig, 1978)	.48	9.83	.001	.57
$R^2 = .51$	$F(3,204) = 70.58, p < .001$			

counter estimates: low subjective frequency words (range: 1.25–3.36; 942 items), medium subjective frequency words (range: 3.37–4.55; 934 items), and high subjective frequency words (range: 4.56–7.00; 948 items). If reliance on neighborhood size is subjective-frequency dependent, neighborhood size should yield a larger contribution to subjective frequency estimates of low subjective frequency words compared with high subjective frequency words.

Table 7 displays the results of the regression analyses conducted separately on high, medium, and low subjective-frequency words. In these analyses, meaningfulness was the best predictor for the ratings of low subjective-frequency words. In contrast, objective log frequency was the best predictor for the ratings of high subjective-frequency words. Interestingly, neighborhood size provided a modest contribution to these ratings, but only for medium and high subjective-frequency words (note that the effect of neighborhood size approached significance only for high subjective frequency words). Hence, in contrast to our predictions, the participants appeared to rely more on the semantic attributes of the stimulus words when the words were less familiar and more on log frequency and orthographic similarity when the words were more familiar. This pattern might be due to the meanings of frequent words' being readily available to all the participants, and hence does not account for any of the variability in the subjective frequency rating estimates. On the other hand, the participants are likely to be differentially familiar with the meanings of less common words. As a result, the extent to which participants know the meanings of words

becomes a more influential factor in judging the subjective frequency of occurrence for low-frequency words.[2]

DISCUSSION

The present results can be summarized in four main points. (1) The present subjective frequency of encounter ratings appear to be a more accurate reflection of actual exposure to stimulus words compared with Toglia and Battig's (1978) familiarity ratings. This observation was supported by the larger correlation between objective log frequency estimates and subjective frequency estimates in the present sample compared with Toglia and Battig's familiarity norms. In addition, this observation was supported by the evidence that Toglia and Battig's familiarity ratings were more highly related to a semantic variable (i.e., meaningfulness of the word) compared with the present subjective frequency estimates. (2) Subjective frequency estimates from different participant groups obtained with the present sample appear to rely on very similar sources of information. This observation was supported by the relatively similar ratings across healthy young, older, and the web-based subjects, as reflected by the high correlations across these participant groups. Moreover, the consistency across groups in the predictive power of objective log frequency, orthographic N, and meaningfulness indicated that the participants used the same information in making their ratings. (3) Meaningfulness appears to be a more reliable predictor of the present subjective frequency ratings for relatively rare words than do log frequency and neighborhood size, whereas log frequency appears to be a better predictor

Table 7
Simultaneous Regression Analyses With the Three Predictor Variables as a Function of Subjective Frequency Estimates (High, Medium, and Low)

Predictors	Beta	t value	p<	Semipartial
	Low Familiarity			
Log frequency (Kučera & Francis, 1967)	.336	8.583	.000	.394
Neighborhood size	.019	.505	n.s.	.025
Meaningfulness (Toglia & Battig, 1978)	.483	12.212	.000	.521
$R^2 = .44$	$F(3,401) = 103.07, p < .001$			
	Medium Familiarity			
Log frequency (Kučera & Francis, 1967)	.318	6.616	.001	.315
Neighborhood size	−.047	−1.017	n.s.	−.051
Meaningfulness (Toglia & Battig, 1978)	.139	2.908	.004	.144
$R^2 = .14$	$F(3,397) = 21.71, p < .001$			
	High Familiarity			
Log frequency (Kučera & Francis, 1967)	.619	14.523	.000	.588
Neighborhood size	.07	1.846	.066	.092
Meaningfulness (Toglia & Battig, 1978)	−.055	−1.293	n.s.	−.065
$R^2 = .43$	$F(3,399) = 100.38, p < .001$			

than meaningfulness for relatively common words. (4) The orthographic similarity to other words stored in the lexicon was only a modest predictor of subjective frequency of encounter estimates, and this predictive power was limited to medium and high subjective-frequency words.

The strong relation between the present subjective frequency ratings and the objective frequency estimates suggests that participants do have access to relatively pure information about frequency of occurrence. However, it should also be noted that these ratings rely on different sources of information, depending on the subjective frequency value of the words in the sample. As noted, subjective frequency estimates were more strongly related to meaningfulness for low subjective-frequency words, whereas, for medium and high subjective-frequency words, objective frequency and orthographic similarity to other words played a larger role. The frequency-dependent nature of the influence of these variables is not surprising, because the meanings of common words should be highly familiar to participants, making this variable a relatively insensitive measure for common words. The relatively larger influence of meaningfulness for low-frequency words is also consistent with a recent observation by Paap, Johansen, Chun, and Vonnahme (2000, Experiment 5) suggesting that college participants may not recognize very low-frequency words when presented out of context. As Paap et al. noted, for such very rare words, it is possible that semantic characteristics may play more of a role. Although there was an influence of meaningfulness for rare words in the present subjective frequency of occurrence estimates, the important point for the present discussion is that such an influence is considerably smaller than the influence of meaningfulness in the Toglia and Battig (1978) subjective familiarity norms (see Table 5). Hence, we believe that the present subjective frequency of encounter estimates are more successful at focusing individuals on the targeted information (i.e., frequency of exposure to a stimulus word).

Given the present evidence suggesting the important role of instructions for familiarity ratings, it is important to note that there has been considerable variability across past studies in the type of instructions that investigators have used for developing familiarity norms. For example, in another large scale familiarity rating study, Nusbaum, Pisoni, and Davis (1984) asked participants to rate over 19,000 words on a 7-point scale ranging from 7, *You recognize a word and are confident that you know the meaning of the word* to 1, *You have never seen the word before*. A rating of 6 in the Nusbaum et al. norms indicated *You think you know the meaning of the word but are not certain that the meaning you know is correct*. Of course, this set of norms directly emphasized the meaning of the word, and most words were given very high ratings. In order to directly address the predictive power of the Nusbaum et al. norms compared with the present norms, we partialled out the Nusbaum et al. norms and found that the present norms still predicted a highly reliable 61% of the variance of Kučera and Francis (1967)

and 67% of the variance in the Celex (Dutch Center for Lexical Information, 1995) norms (both $ps < .0001$). In contrast, the Nusbaum et al. norms did not predict a reliable amount of variance in the objective norms after the present subjective frequency estimate was partialled out (both $ps > .15$). Of course, if one is primarily interested in whether participants know the meaning of the words, the Nusbaum et al. norms are probably a good proxy, but there is some question as to whether these are useful as measures of a frequency-based estimate of familiarity.

In contrast to estimates of knowing the meaning of the stimulus word, Amano, Kondo, and Kakehi (1995), Gernsbacher (1984), Toglia and Battig (1978) had participants rate items on a numeric scale ranging from familiar to unfamiliar. However, as supported by the present analyses, it is possible that participants can use multiple sources of information to make familiarity ratings. Thus, one needs to exercise caution in partialling out this variable to examine the "pure" influence of other variables in standard word recognition studies. We would argue that relatively neutral estimates concerning the subjective frequency, as used in the present study, are a better proxy.

It is also noteworthy that the participants did vary their subjective frequency estimates across the different domains of reading, writing, hearing, and speaking. Specifically, participants consistently rated hearing words more frequently than writing words. In addition, older adults rated the frequency of encountering words via print as higher than the younger adults, whereas the opposite was true for writing words. Although it is possible that these may reflect perceived rather than real differences in frequency of exposure, it is noteworthy that all estimates were highly intercorrelated (all $rs > .90$). The differences in frequency of exposure to different forms of lexical representations have intriguing implications regarding modality specific representations and exemplar based models of lexical processing (e.g., Goldinger, 1998). This issue has been recently explored by Gaygen and Luce (1998), who found evidence for both modality specific and modality independent influences of subjective frequency estimates on a smaller sample of words (252 words, also see Amano et al., 1995; Garber & Pisoni, 1991; Pisoni & Garber, 1990).

Although there was a strong correlation between objective log frequency estimates and the present subjective frequency estimates, the correlation was not perfect. Thus, one might ask the question whether the present subjective frequency estimates account for more of the variance in a speeded lexical processing task than do objective frequency estimates. In order to address this possibility, we investigated the relationship between objective log frequency and subjective frequency of encounter estimates for the mega-study of lexical decision performance (including the 2,938 single syllable words used in the present sample), reported by Balota, Cortese, and Pilotti (1999). The results indicated that the present subjective frequency estimates accounted for 21% and 18% of the variance after controlling for Kučera and Francis (1967)

and Celex (Dutch Center for Lexical Information, 1995) norms, respectively (both ps < .0001). In contrast, Kučera and Francis and Celex accounted for less than 1% of the variance after partialing out the present subjective frequency norms. Thus, subjective frequency of encounter estimates appear to be better at predicting the ease of lexical processing than do large-scale objective frequency of exposure norms.

CONCLUSION

The present study provides subjective frequency of exposure estimates for the vast majority of all single syllable words in the English language. We have argued that participants are more likely to rely on semantic variables when asked to make standard judgments of familiarity compared with the present frequency of exposure estimates. The important inference for studies of lexical processing is that one needs to exercise caution in partialing out the potential confound of familiarity when addressing the unique influence of a targeted variable on lexical processing. Without such caution, it is possible that the investigator will throw out the baby with the bathwater (i.e., inadvertently partial out the targeted variable, which may covary with standard estimates of familiarity). Based on the relatively strong relationship between the present subjective frequency estimates and objective frequency estimates, and the predictive power of the present subjective frequency estimates for lexical decision performance, we believe that if one wishes to partial out subjective untimed ratings, one should use subjective frequency estimates instead of familiarity ratings.

REFERENCES

AMANO, S., KONDO, T., & KAKEHI, K. (1995). Modality dependency of familiarity ratings of Japanese words. *Perception & Psychophysics*, **57**, 598-603.

ANDREWS, S. (1997). The effect of orthographic similarity on lexical retrieval: Resolving neighborhood conflicts. *Psychonomic Bulletin & Review*, **4**, 439-461.

BALOTA, D. A., CORTESE, M. J., & PILOTTI, M. (1999, November). *Item-level analyses of lexical decision performance: Results from a megastudy.* Poster presented at the annual meeting of the Psychonomic Society, Los Angeles.

BALOTA, D. A., & FERRARO, F. R. (1996). Lexical, sublexical, and implicit memory processes in healthy older adults and in individuals with dementia of the Alzheimer type. *Neuropsychology*, **10**, 82-95.

BALOTA, D. A., FERRARO, F. R., & CONNOR, L. T. (1991). On the early influence of meaning in word recognition: A review of the literature. In P. J. Schwanenflugel (Ed.), *The psychology of word meaning* (pp. 187-222). Hillsdale, NJ: Erlbaum.

BALOTA, D. A., & SPIELER, D. H. (1999). Word frequency, repetition, and lexicality effects in word recognition tasks: Beyond measures of central tendency. *Journal of Experimental Psychology: General*, **128**, 32-55.

BURGESS, C., & LIVESAY, K. (1998). The effects of corpus size in predicting reaction time in a basic word recognition task: Moving on from Kučera and Francis. *Behavior Research Methods, Instruments, & Computers*, **30**, 272-277.

COLTHEART, M., DAVELAAR, E., JONASSON, J. T., & BESNER, D. (1977). Access to the internal lexicon. In S. Dormic (Ed.), *Attention and performance VI* (pp. 535-555). Hillsdale, NJ: Erlbaum.

DUTCH CENTER FOR LEXICAL INFORMATION (1995). *The Celex lexical database.* Nijmegen: Dutch Center for Lexical Information.

GARBER, E. E., & PISONI, D. B. (1991). *Lexical memory in visual and auditory modalities: A second report.* (Research on Speech Perception, Prog. Rep. No. 17, pp. 213-227). Bloomington: Indiana University, Speech Research Laboratory.

GAYGEN, D. E., & LUCE, P. A. (1998). Effects of modality on subjective frequency estimates and processing of spoken and printed words. *Perception & Psychophysics*, **60**, 465-483.

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., & LOGIE, R. H. (1980). Age-of-acquisition, imagery, concreteness, familiarity, and ambiguity measures for 1,944 words. *Behavior Research Methods & Instrumentation*, **12**, 395-427.

GOLDINGER, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review*, **105**, 251-279.

KUČERA, M., & FRANCIS, W. N. (1967). *Computational analysis of present-day American English.* Providence, RI: Brown University Press.

LANDAUER, T. K., & STREETER, L. A. (1973). Structural differences between common and rare words: Failure of equivalence assumptions for theories of word recognition. *Journal of Verbal Learning & Verbal Behavior*, **12**, 119-131.

NUSBAUM, H. C., PISONI, D. B., & DAVIS, C. K. (1984). *Sizing up the Hoosier mental lexicon: Measuring the familiarity of 20,000 words* (Research in Speech Perception, Prog. Rep. No. 10). Bloomington: Indiana University, Speech Research Laboratory.

O'DOWD, S. C. (1984). Does vocabulary decline qualitatively in old age? *Educational Gerontology*, **10**, 357-368.

PAAP, K. R., JOHANSEN, L. S., CHUN, E., & VONNAHME, P. (2000). Neighborhood frequency does affect performance in the Reicher task: Encoding or decision? *Journal of Experimental Psychology: Human Perception & Performance*, **26**, 1691-1720.

PEEREMAN, R., CONTENT, A., & BONIN, P. (1998). Is perception a two-way street? The case of feedback consistency in visual word recognition. *Journal of Memory & Language*, **39**, 151-174.

PISONI, D. B., & GARBER, E. E. (1990). Lexical memory in visual and auditory modalities: The case for a common mental lexicon. In *Proceedings of the 1990 International Conference on Spoken Language Processing* (pp. 401-404). Kobe: Acoustical Society of Japan.

PLAUT, D. C., MCCLELLAND, J. L., SEIDENBERG, M. S., & PATTERSON, K. (1996). Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychological Review*, **103**, 56-115.

SEIDENBERG, M. S., & MCCLELLAND, J. L. (1989). A distributed developmental model of word recognition and naming. *Psychological Review*, **103**, 56-115.

SHIPLEY, W. C. (1940). A self-administering scale for measuring intellectual impairment and deterioration. *Journal of Psychology*, **9**, 371-377.

SPIELER, D. H., & BALOTA, D. A. (1997). Bringing computational models of word naming down to the item level. *Psychological Science*, **8**, 411-416.

SPIELER, D. H., & BALOTA, D. A. (2000). Factors influencing word naming in younger and older adults. *Psychology & Aging*, **15**, 225-231.

STONE, G. O., VANHOY, M., & VAN ORDEN, G. C. (1997). Perception is a two-way street: Feedforward and feedback phonology in visual word recognition. *Journal of Memory & Language*, **36**, 337-359.

TOGLIA, M. P., & BATTIG, W. F. (1978). *Handbook of semantic word norms.* Hillsdale, NJ: Erlbaum.

NOTES

1. We also conducted regression analyses on the items rated by the young adult group, the mixed adult group, and Toglia and Battig's (1978) subjects (1,236 items) so we could examine the contribution of log frequency, neighborhood size, and meaningfulness on a larger data set. The results closely replicated those reported in Tables 5 and 6. Young adult group: R^2 = .68; $F(3,1205)$ = 857.11, p < .001. Log frequency: Beta = .79; t = 47.87, p < .001, partial R^2 = .81. Neighborhood size: Beta =

$-.01; t = -.44$, n.s.; partial $R^2 = -.01$. Meaningfulness: Beta $= .19; t = 11.51$, $p < .001$; partial $R^2 = .32$. Mixed adult group: $R^2 = .65$; $F(3,1205) = 756.26$, $p < .001$. Log frequency: Beta $= .77$; $t = 44.62$, $p < .001$; partial $R^2 = .79$. Neighborhood size: Beta $= .02$; $t = .88$, n.s.; partial $R^2 = .03$. Meaningfulness: Beta $= .19$; $t = 11.39$, $p < .001$; partial $R^2 = .31$. Toglia and Battig's group: $R^2 = .62$; $F(3,1205) = 663.84$, $p < .001$. Log frequency: Beta $= .54$; $t = 30.03$, $p < .001$; partial $R^2 = .65$. Neighborhood size: Beta $= -.02$; $t = -1.02$, n.s.; partial $R^2 = -.03$. Meaningfulness: Beta $= .54$; $t = 30.18$, $p < .001$; partial $R^2 = .66$.

2. It is worth noting that we also examined other orthographic similarity indices that have been identified as predictors of word identification such as the frequency of the neighbors (see Andrews, 1997) and feedback consistency (the consistency of the relations between phonology and orthography; see Peereman et al., 1998; Stone, Vanhoy, & Van Orden, 1997). None of these factors was found to yield a contribution to frequency ratings that was higher than the modest effect produced by neighborhood size, suggesting that structural similarity factors indeed play a rather marginal role in subjective frequency ratings.