

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/24274573>

Subjective frequency and imageability ratings for 3,600 French nouns

Article in Behavior Research Methods · May 2009

DOI: 10.3758/BRM.41.2.546 · Source: PubMed

CITATIONS

46

READS

192

2 authors:



Alain Desrochers

University of Ottawa

79 PUBLICATIONS 1,364 CITATIONS

[SEE PROFILE](#)



Glenn L Thompson

University of Ottawa

12 PUBLICATIONS 129 CITATIONS

[SEE PROFILE](#)

Some of the authors of this publication are also working on these related projects:



Assessment of French reading skills [View project](#)

Subjective frequency and imageability ratings for 3,600 French nouns

ALAIN DESROCHERS AND GLENN L. THOMPSON
University of Ottawa, Ottawa, Ontario, Canada

Subjective frequency and imageability estimates for a sample of 3,600 French nouns were collected from two independent groups of 72 young adults each. Both groups received standard instructions and provided their ratings on a 7-point scale. The timing, sequencing, presentation of lexical stimuli, and recording of responses were controlled by a computer. All estimates of internal consistency and test-retest reliability ($\geq .98$) confirm the high level of precision and reliability of the ratings. Correlations with ratings drawn from similar studies were found to be positive and significant for subjective frequency ($r \geq .85$) and for imageability ($r \geq .69$). Subjective frequency was positively and significantly correlated with objective frequency estimates drawn from 10 different sources ($r \geq .42$). Subjective frequency and imageability were significantly correlated ($r = .26$), a relationship that was driven primarily by a sudden drop in imageability ratings for words with a subjective frequency rating below 2.5. The methodological implications of these findings are discussed. The ratings can be downloaded as supplemental materials from brm.psychonomic-journals.org/content/supplemental.

Words are both perceptual and symbolic objects. As perceptual objects, they can have visual, auditory, and tactile forms, and these forms are extended in space or time. Word forms are typically distinct (except for homonyms), but they may share components with other words (e.g., strokes, letters, phonemes, syllables, morphemes). As symbolic objects, they may refer to concrete or abstract concepts and they may share semantic features with other words. In sum, words are multidimensional stimuli, and this characteristic makes their use in rigorous scientific investigation particularly challenging. Meeting this challenge (e.g., minimizing the risk of confounding lexical properties or of misattributing their effects) requires the cataloguing of a vast array of lexical attributes.

The method employed in documenting lexical attributes often depends on the nature of the variable under consideration. For example, some word properties can be drawn from an analysis of their entries in a lexical database: their structural components (e.g., their graphemes, phonemes, syllables, and morphemes) or their relationships to other words (e.g., orthographic or phonological neighborhood, uniqueness point, homonymy, and synonymy). Other word properties can be assessed only by having human judges provide an estimate on a rating scale. Two such properties are word subjective frequency and imageability. The primary goal of the present research is to investigate the characteristics of word subjective frequency and imageability, and provide ratings for a pool of 3,600 French nouns. Before we describe the method used to obtain these ratings, we briefly discuss the nature of each of these word attributes and their relevance to psychological research.

Subjective Frequency

Words are known to vary considerably in their frequency of use in speech and print. *Objective word frequency* is typically estimated by summing the occurrences of each particular word token in large sets of written text or text transcribed from speech. Usually, it reflects the frequency with which a sample of professional writers or speakers uses these words. *Subjective word frequency*, by contrast, is ordinarily measured by having human judges rate a large set of words on a 7-point Likert-like scale, where 1 is assigned to words that are perceived as being *rare*, and 7 to words that are perceived as *commonly used*. These ratings reflect the average frequency with which ordinary language users think they have heard or read a set of words. The correlation between objective word frequency and subjective word frequency estimates is substantial, but not perfect. In recent studies, in both English and French, the Pearson product-moment correlation between these two variables was found to range from .46 to .83 (e.g., Balota, Pilotto, & Cortese, 2001; Benjafield & Muckenheim, 1989; Bonin et al., 2003; Desrochers & Bergeron, 2000; Gonthier, Desrochers, Thompson, & Landry, in press; Stadthagen-Gonzalez & Davis, 2006; Toglia & Battig, 1978).

Beyond measurement error, many factors may account for a less-than-perfect correlation between objective and subjective word frequency (see Baayen, 2001; Beauchemin, Martel, & Théoret, 1992). For instance, objective frequency estimates may be drawn from texts that vary widely in genres (e.g., novels, poetry, technical writing, newspapers, books for children). These samples of text are typically produced

A. Desrochers, alain.desrochers@uottawa.ca

by professional and skilled writers whose own vocabularies often exceed those of their readers. Such samples tend toward an overestimation of the actual frequency of rare words and an underestimation for common words. Dated texts may yield an overestimation for words that are no longer in common usage, as well as an underestimation for words recently introduced into the language (e.g., recent concepts or artifacts). Some words occur only in highly specific contexts (e.g., the vocabulary of horse racing), which may or may not be represented in the text corpora used to obtain the word frequency estimates. This phenomenon is known as *underdispersion*, and it primarily affects low-frequency words (Baayen, 2001). To address these limitations, at least in part, subjective frequency has been used to confirm the assignment of words to low- or high-frequency classes by objective frequency (e.g., Monsell, Doyle, & Haggard, 1989; see also Thompson & Desrochers, *in press*) or as an independent variable in its own right (e.g., Gernsbacher, 1984). Because subjective frequency ratings are potentially codetermined by other lexical characteristics, such as age of acquisition, conceptual familiarity, meaningfulness, and imageability (see Baayen, 2005; Balota et al., 2001; Barry, Hirsh, Johnston, & Williams, 2001; Bonin, Chalard, Méot, & Fayol, 2001; Ferrand et al., 2008; Thompson & Desrochers, *in press*), it is not simply a matter of replacing objective frequency with subjective frequency, but rather is a matter of utilizing both types of ratings in a cross-validation strategy in order to compensate for their respective weaknesses in estimating the underlying “frequency” construct.

Word frequency is important in psychological research because its influence on language and memory processes is pronounced and robust. For instance, responses to frequently used words are typically faster and more accurate than those to rare words, whether these responses involve visual or auditory recognition, oral reading, written transcription, lexical categorization, or semantic evocation (for reviews, see Monsell, 1991; Norris, 2006). Word frequency may also moderate the effect of other variables, such as orthographic regularity and imageability (e.g., de Groot, 1989; James, 1975; Lupker, Brown, & Colombo, 1997; Strain & Herdman, 1999). Indeed, that these effects are observed primarily with low-frequency words (i.e., words for which objective frequency estimates are least reliable) highlights the importance of using both objective and subjective frequency ratings as criteria in experimental-item selection. The consistent application of this methodological strategy requires the collection of frequency ratings for a more extensive set of lexical stimuli than is currently available (i.e., only 4,972 French words, by our count).

Imageability

Word referents vary in the extent to which they can evoke a sensory experience. The concept of imageability refers to the estimated ease with which human judges conjure a mental image to a particular word. A high-imageability word (e.g., *house*, *apple*, *truck*) can evoke a mental image readily and spontaneously, whereas a low-imageability word (e.g., *neutrality*, *agreement*, *help*) does so with difficulty, if at all, and information retrieval from memory is laborious. Imageability is typically assessed

by having human judges rate a large set of words on a 7-point Likert-like scale, where 1 is assigned to words that are hardest to image and 7 is assigned to words that are easiest to image. Imageability can be construed as a special case of a more general semantic contrast that distinguishes concrete from abstract words. Ratings instructions typically refer to words’ capacity to arouse a sensory experience, such as a mental picture or sound. The ease and/or speed with which this sensory experience occurs are presented as the primary rating criteria. For concreteness, participants are asked to rate the extent to which words refer to concrete objects, persons, places, or things that can be seen, heard, felt, smelled, or tasted. In this case, no reference is made to the ease or speed of the sensory experience (see Paivio, Yuille, & Madigan, 1968; Toglia & Battig, 1978). Despite these differences, imageability and concreteness ratings are found to be significantly correlated ($r \geq .48$; Benjafield & Muckenheim, 1989; Bird, Franklin, & Howard, 2001; Bonin et al., 2003; Toglia & Battig, 1978).

Like word frequency, imageability influences people’s reactions to words. Highly imageable words can yield faster and more accurate responses than low-imageability words in tasks that involve lexical or semantic decisions or oral reading (Baluch & Besner, 2001; Coltheart, Laxon, & Keating, 1988; Cortese, Simpson, & Woolsey, 1997; de Groot, 1989; Glass & Meany, 1978; Kroll & Merves, 1986; Raman & Baluch, 2001; Richardson, 1980; Strain & Herdman, 1999; Strain, Patterson, & Seidenberg, 1995, 2002; van Hell & de Groot, 1998; Zevin & Balota, 2000). It is noteworthy that the effect of imageability in lexical decision can be suppressed when words have a high frequency of use (e.g., *James*, 1975), and it can be considerably reduced in oral reading when grapheme-to-phoneme correspondences are consistent (e.g., Strain & Herdman, 1999). Studies of human memory have so far yielded the richest and most consistent evidence for the importance of imageability in language processing. Overall, highly imageable words have been found to be easier to encode and/or retrieve than low-imageability words in free recall, cued recall, and recognition tasks (e.g., Coltheart & Winograd, 1986; Denis, 1979; Ferlazzo, Conte, & Gentilomo, 1993; Paivio, 1971; Walker & Hulme, 1999). In sum, there is substantial evidence to suggest that it is important to control for the effect of imageability in a variety of experimental contexts. The implementation of this control is limited by the fact that imageability estimates currently exist for only a relatively small pool of words in the French lexicon (5,068 words, by our count).

Shared Variance Between Subjective Frequency and Imageability Ratings

Ratings of subjective frequency and imageability are dissociable in principle. A highly imageable word can have a high or a low frequency of use (e.g., *house* vs. *obelisk*). Similarly, a frequently used word can be very easy or very hard to image (e.g., *house* vs. *need*). In the practice of subjective rating estimation, however, there is a dependency relationship between imageability and word frequency: Human judges can differentiate degrees

of imageability only when a word is known to them, and the probability of knowing a word varies as a function of its frequency of use in the language. This dependency accounts for the fact that subjective frequency and imageability are often found to be significantly correlated in normative studies (e.g., Benjafield & Muckenheim, 1989; Bonin et al., 2003; Desrochers & Bergeron, 2000; Ferrand et al., 2008; Forget, 2005; Gonther et al., in press; Stadthagen-Gonzalez & Davis, 2006; Toglia & Battig, 1978). In a detailed analysis, Gonther et al. showed that as subjective frequency values get larger, the between-word variability in imageability ratings increases, and so does the residual variance about the regression line relating these two lexical attributes. The fact that the residual variance in the linear model is heteroscedastic indicates that subjective frequency predicts imageability adequately only at lower levels of frequency. The pattern of this relationship suggests that subjective frequency sets an upper bound for imageability, at least for low-frequency words. We shall return to the methodological implications of this dependency after we have attested to its presence in the data collected in the present study.

The theoretical interpretation of word frequency and imageability effects on language and memory processes has been in dispute for many years (for reviews, see Kosslyn, Thompson, & Ganis, 2006; Monsell, 1991; Norris, 2006; Paivio, 1986; Sabsevitz, Medler, Seidenberg, & Binder, 2005). Such controversy is fueled in part by the multidimensional nature of lexical stimuli and the multicollinearity of their attributes. There is indeed ample evidence that word frequency and imageability are correlated with other lexical attributes (e.g., age of acquisition, conceptual familiarity, morphological family size, orthographic neighborhood, word length; for demonstrations, see Baayen, 2005; Ferrand et al., 2008; Köhler, 1986; Stadthagen-Gonzalez & Davis, 2006). It is therefore important that these variables be documented in detail for a wide range of words, within and across grammatical categories, to optimize item selection in research applications. Such detailed analysis can be expected to help identify confounded variables and their interactions, thereby enhancing the precision of experimentally controlled and manipulated lexical variables in the design of experiments. The present study was specifically intended to provide, for a large sample of French nouns (about half of which were previously undocumented), reliable ratings of subjective frequency and imageability for research applications in experimental psychology and related disciplines. The quality of the data was enhanced through the implementation of a computerized data-collection procedure, which allowed various forms of stimulus randomization and data screening based on response latency, and a research design that permitted the estimation of several reliability indices.

METHOD

Participants

A total of 144 young adults took part in the present study. They were randomly assigned to each part of the study; 72 (45 female) provided subjective frequency ratings, and 72 (47 female) provided imageability ratings. They were recruited from introductory psy-

chology classes at the University of Ottawa in Canada. Their ages ranged from 18 to 28 years ($M = 20.23$, $SD = 2.41$, for the subjective frequency group; $M = 20.12$, $SD = 2.27$, for the imageability group). All participants were native speakers of French and had normal or corrected-to-normal vision. None reported a history of reading problems. Data were collected over four experimental sessions, for which the participants each received \$40.

Materials and Equipment

The participants rated a total of 1,008 French nouns in each of the four experimental sessions. These items included a general practice list of 20 items and four experimental blocks of 247 items. Each experimental block comprised 10 filler items, followed by 237 experimental items. Of these experimental items, 225 were shown only once, 6 were repeated items drawn from other experimental blocks of the current session, and 6 were repeated items drawn from experimental blocks of other sessions. A total of 4,032 items were rated on both subjective frequency and imageability by independent groups of participants, with 96 items repeated within session and 96 repeated between sessions. After excluding the practice items and the repeated items, 3,600 items remained in the final set of ratings.

The item selection process was constrained by several methodological considerations. Practice items were drawn from existing sources of subjective frequencies and imageability estimates and were selected so as to encourage participants to make use of the entire range of the 7-point rating scale. Experimental items were selected so as to (1) extend the pool of French nouns for which subjective frequency estimates (1,772 previously undocumented words were included here, increasing the existing coverage of the lexicon by 36%) and imageability estimates (1,721 previously undocumented words were included here, increasing coverage of the lexicon by 34%) were already available, (2) share a sufficient number of items with existing sources of normative data (i.e., a minimum of 50 words per source whenever possible) in order to assess concurrent validity, and (3) increase the probability of obtaining symmetrical distributions for both variables across the entire range of the 7-point scale. The basic statistical characteristics of the experimental items are presented in Table 1.

The items were shown at the center of a video monitor, in black 20-point Arial characters on a white background. The timing, sequencing, presentation of stimuli, recording of responses, and response latencies were controlled by a PC-compatible computer and E-Prime software Version 1.0 (Schneider, Eschman, & Zuccolotto, 2002a, 2002b).

Procedure

The data-collection procedure is an extension of that developed by Cortese and Fugett (2004). Participants were tested individually in a quiet, dimly lit room over four sessions of approximately 60 min each. The sessions were separated by 7 to 10 days. Upon arriving for a testing session, participants first received or were reminded of the rating instructions (see the Appendix). Participants responded to a set of 20 practice trials and then a set of 988 experimental trials, divided into four blocks. Opportunities for breaks were provided at the end of each block of experimental trials. The order of all possible pairs of practice and experimental item lists was counterbalanced over the four sessions of the study, yielding 24 distinct orders. Over the sam-

Table 1
Basic Statistical Characteristics of the Subjective Frequency and Imageability Ratings

	Subjective Frequency	Imageability
Minimum rating	1.07	1.09
Maximum rating	6.46	6.99
Mean rating	3.58	4.16
SD	1.11	1.48
Skewness	.066	.31
Kurtosis	-.70	-1.15

ple of participants, each pair of practice and experimental item lists was presented an equal number of times in the first, second, third, and fourth sessions. Within each block of trials, the items were presented in a different random order for each participant. This randomization strategy was facilitated by the choice of a computerized data-collection procedure, and it extended prior paper-and-pencil implementations (e.g., Desrochers & Bergeron, 2000, employed 3 different possible item orders). This methodological control is justified by the observation of item-order effects in visual word recognition (Perea & Carreiras, 2003; Taylor & Lupker, 2001); explicit judgments on item attributes may be similarly sensitive to sequential effects.

Each trial involved the following sequence of events. A fixation point, a “+” sign, was shown for 500 msec at the center of the screen. A word then appeared in the same location as the fixation point and remained in view until a response was made. Participants entered their ratings via the numeric keypad on the right side of the keyboard. The response was followed by a 900-msec interstimulus interval. No feedback was provided on any aspect of the response. Although response latency was measured on every trial, both participants and experimenter were blind to this aspect of the study. Moreover, participants were not given any time constraints, nor were they instructed to speed up their responses. We assumed that the large number of items to be rated in each series would provide the necessary incentive for the participants to proceed swiftly and maintain a quick and steady pace. The response latencies provided additional information for the detection of outliers.

The instructions given to the participants were adapted from those used by Denis (1975) and Paivio et al. (1968), and they are provided in the Appendix. With respect to subjective frequency, previous work has demonstrated that it is preferable to emphasize frequency of occurrence rather than “familiarity” in the instructions, because this reduces the extent to which such ratings are influenced by other lexical characteristics, such as meaningfulness (see Balota et al., 2001). Accordingly, the subjective frequency instructions used here emphasize the frequency with which words are encountered rather than familiarity, but we elected not to adopt the full set of verbal anchors proposed by Balota et al. for the collection of subjective frequency ratings.

Although attaching the numerical values of the 7-point scale to explicit frequencies (e.g., 1 = *never*, 2 = *once a year*, etc.) may provide advantages (e.g., resistance to contextual effects), the choice of verbal labels is arbitrary and requires that frequency intervals of differing magnitudes be applied to otherwise equidistant interval-scale units (i.e., the difference between *never* and *once a year* is not equal to the difference between 1 and 2). Although it is true that Balota et al.’s (2001) norms are especially strongly correlated with objective frequency (about .10 more strongly than other similar norms in English), this may be attributable not to improved measurement of subjective frequency (i.e., improved validity) but, rather, to an artifact of measurement associated with their choice of scale. One possibility is that the use of explicit verbal anchors in Balota et al.’s study produced a positively skewed distribution much like that observed with objective frequency estimates (although they did not report the distributional characteristics of their norms), which could explain the particularly strong correlation that they obtained with objective frequency, because similarly skewed variables tend to be more highly correlated than variables that differ markedly in distribution shape, all other things being equal (Cohen, Cohen, West, & Aiken, 2003). Until the influence of the different anchoring options on subjective ratings is better understood and an evidence-based consensus is reached, we consider it preferable to maintain comparability with prior studies (e.g., Denis, 1975; Paivio et al., 1968).

RESULTS AND DISCUSSION

Prior to the statistical analysis and subsequent computation of the norms, the data were screened for outliers within each item such that ratings falling outside a ± 3.5 standard deviation cutoff were discarded. In addition, the

response latencies allowed the imposition of a lower-bound criterion below which responses could be considered invalid (e.g., accidental keypresses). Visual inspection of the reaction time distribution suggested that responses with latencies shorter than 300 msec were drawn from a distinct distribution; thus, responses that were slower than this cutoff were discarded from the database. No upper-bound criterion was adopted because speed of responding was not explicitly emphasized in the instructions. On average, the participants in the imageability and subjective frequency samples lost 0.54% ($SD = 1.14$) and 0.63% ($SD = 1.35$) of their data, respectively.

The presence of an inverted-U quadratic function relating item means to item variances in rating data may be taken as an empirical indicator that participants are indeed using the entire range of the scale, whereas the absence of such a function suggests a failure to make use of the full scale of measurement (Forget, 2005). We therefore proceeded to fit a quadratic function to the relationship between the item means and item variances for subjective frequency and imageability. For subjective frequency, a simple linear function explained only 9% of the variance in item variances [$F(1,3597) = 361.17, p < .001$, constant = 1.31, beta = 0.14]. In contrast, a model incorporating an additional quadratic component was also significant, but explained nearly half of the total variance [$r^2 = .46, F(2,3597) = 1,536.98, p < .001$], with the following parameters: constant = -1.20, first beta = 1.69, second beta = -0.22. Similarly, for imageability, the linear component of the function relating item means and item variances was significant [$F(1,3596) = 133.82, p < .001$], but explained only 3.6% of the variance in item variances (constant = 11.68, beta = -0.90). In contrast, the addition of a quadratic component added a large amount of explained variance [$r^2 = .36, F(2,3595) = 1,028.06, p < .001$], with the following parameters: constant = -23.32, first beta = 17.24, second beta = -2.08. The results of this analysis and a visual inspection of the associated scatterplots confirmed the presence of variance compression at both ends of the rating scale, supporting the idea that participants were actually using the full range of the scale. It is noteworthy that the standardized residuals resulting from this analysis provide more valid estimates of the relative disagreement among participants about an item’s rating than do item standard deviations; the standardized residuals may be viewed as an estimate of dispersion that is corrected for the location of the item on the 7-point scale, where lower residuals denote higher agreement.

The database, including the mean frequency and imageability ratings, the standard deviations, and the standardized residuals, is accessible online: brm.psychonomic-journals.org/content/supplemental. We now report and discuss the main statistical characteristics of the ratings, the reliability estimates, and the relationships among these ratings and estimates drawn from other studies.

Basic Statistical Characteristics

The items were selected so as to obtain as broad a dispersion of the ratings as possible on the 7-point scale. The skewness estimates (see Table 1) confirm that the rating

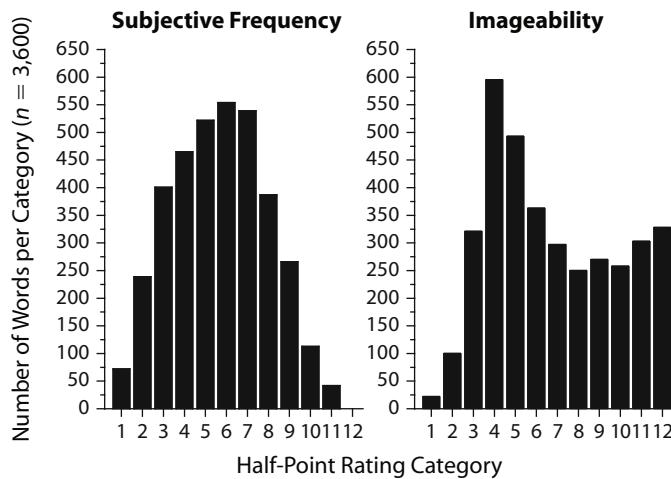


Figure 1. Distribution of mean subjective frequency and mean imageability ratings by half-point category.

distributions are fairly symmetrical for subjective frequency and slightly skewed to the right for imageability. Both kurtosis estimates are negative, reflecting a distribution that is platycurtic relative to the normal distribution. Figure 1 presents the number of words in each half-point category of each rating scale (i.e., 1 = 1.0–1.49; 2 = 1.5–1.99 . . . 11 = 6.0–6.49; 12 = 6.5–7.0). These graphs confirm the broad dispersion of the ratings even though their distribution is closer to a rectangular one for imageability than for subjective frequency. The density of words in the tails of the distribution is sparse for subjective frequency, whereas a relatively large number of words populate the high end of the scale for imageability.

Rating Reliability

This study was designed so that several reliability estimates could be calculated. Table 2 reports the standard error of measurement, two estimates of internal consistency of the ratings, and within-session and between-sessions test-retest reliability estimates. These estimates confirm that the standard error of the measurement is small for both subjective frequency and imageability. All estimates of internal consistency and test-retest reliability ($\geq .98$) attest to the accuracy and the stability of the subjective frequency and imageability ratings. The stability of the ratings over time is remarkable when one considers that the interval between testing sessions was at least a week.

Rating reliability can also be investigated by using participants as a unit of analysis. Estimates of reliability can be computed for each participant and these estimates averaged to get a reliability estimate for the group as a whole. The group means and standard deviations for these indices are reported in Table 3. Using the analogy of classical test development, one can compute a participant-total correlation that serves to identify participants whose ratings differ markedly from the rest of the group. This procedure is equivalent to computing item-total correlations to verify the suitability of test items. The participants in the present study produced participant-sample correlations ranging

from .45 to .83 for subjective frequency (excluding one outlier at .21) and .41 to .85 for imageability. It is noteworthy that all but one exceptional case met the generally accepted criterion of $r \geq .40$ for an item-total correlation.

It is also possible to express the discrepancy between participants and the average group ratings through the analysis of difference scores (i.e., the average difference over all items between a participant's rating and that of the rest of the group). The results reported in Table 3 indicate that, on average, most participants provide ratings that are within 1 SD of the group mean. For subjective frequency, only 9 participants had a mean absolute z score greater than 1, and no participant exceeded 1.5. Similarly, for imageability, only 10 participants had a mean absolute z score greater than 1, and no participant exceeded 1.15. Thus, the results indicate that none of the participants diverged systematically from the rest of the group (e.g., by misinterpreting the rating scale).

The stability of participants' responses across time was estimated for the items repeated within session and those repeated across sessions via test-retest correlations and within-subjects standard errors (i.e., typical error; see Hop-

Table 2
Reliability Estimates of the Mean Subjective Frequency and Imageability Ratings

Item Set and Indicator	Subjective Frequency	Imageability
Nonrepeated Items ($n = 3,600$)		
Standard error of measurement	.14	.18
Cronbach's alpha coefficient	.98	.98
Guttman split-half (lower bound)	.98	.99
Items Repeated Within Sessions ($n = 96$)		
Cronbach's alpha coefficient	.98	.99
Test-retest correlation (Pearson)	.99	.99
Items Repeated Across Sessions ($n = 96$)		
Cronbach's alpha coefficient	.98	.99
Test-retest correlation (Pearson)	.99	.99

Note—Cronbach's alpha coefficients for repeated items were nearly identical in the first and second presentations.

Table 3
Average Reliability of Subjective Frequency and Imageability Ratings, Computed Within Participants

Type of Variability	Subjective Frequency	Imageability
Between-Subjects		
Mean average absolute z score	.80 (.23)	.80 (.16)
Participant-sample correlation	.69 (.11)	.71 (.10)
Within-Subjects Within Session ($n = 96$)		
Mean typical error	.77 (.26)	.88 (.26)
Mean test-retest correlation	.80 (.15)	.79 (.13)
Regression coefficient analysis of practice effect	-.06, $t(71) = 3.55$	-.06, $t(71) = 2.59$
Within-Subjects Across Session ($n = 96$)		
Mean typical error	.95 (.27)	1.02 (.24)
Mean test-retest correlation	.59 (.19)	.73 (.14)
Regression coefficient analysis of practice effect	-.15, $t(71) = 4.17$	-.11, $t(71) = 2.80$

Note.—The average absolute z score reflects the average difference between a participant's ratings and the group averages, where difference is expressed as the absolute value of z scores. The participant-sample correlation reflects the linear relationship between a subject's ratings and the total score of the group on each item (group excludes the participant in question). Values within parentheses denote standard deviations, except in the case of reported significance tests, in which case the values denote N .

kins, 2000). For the test-retest analysis, we elected to use intraclass correlation coefficients because of the relatively small sample size of repeated responses for each participant (96 items). The means and standard deviations reported in Table 3 attest to the global stability of the ratings. A closer inspection of the variability in the sample revealed that individual test-retest correlations for subjective frequency ranged from .44 to .94 within session and from .12 to .85 across sessions. For the imageability task, the within-session test-retest correlations ranged from .44 to .97, and the across-sessions correlations, from .28 to .96. The broad range of observed values is somewhat misleading when one considers that the mean correlation is quite high and that all but a few participants achieved individual reliability estimates greater than .50. To determine whether the rank order of participants on test-retest reliability remained stable over the two testing intervals (within or across sessions), the correlation estimates for the within- and across-sessions reliability were themselves correlated, yielding values of .87 and .77 for subjective frequency and imageability, respectively. The results of this analysis indicate that some participants are more reliable than others regardless of the time interval between test and retest occasions.

The second index of stability reported in Table 3, the typical error, indicates that when participants respond to the same items on two occasions, their responses vary by about 1 unit, on average, on the 7-point scale. A general guideline, then, would be that items must differ by roughly 2 units on the 7-point scale before they exceed a 95% confidence interval boundary at this level of analysis. This can be viewed as an empirically based rule of thumb for stimulus selection. Note that this estimate of stability reflects within-subjects variability rather than the stability of the rating provided by the group as a whole, which is estimated by the standard errors reported in Table 2. When the standard error is considered, items that differ by about half a point might be considered significantly different. The significant practice effects reported in Table 3 revealed a systematic increase

in mean rating from Time 1 to Time 2 regardless of the testing time interval or the dependant variable considered, although the magnitude of the effects is larger in the between-sessions comparisons, which suggests that in the latter cases some type of memory consolidation might have been at work.

Relationship to Similar Normative Studies

Subjective frequency and imageability instructions across extant studies are generally similar to those used by Denis (1975), Paivio et al. (1968), and Toglia and Battig (1978), sometimes with minor modifications to serve particular purposes. Furthermore, subjective ratings are typically mapped on a 7-point Likert-like scale. These regularities permit us to compare the ratings obtained in the present study with those reported in similar studies on the items they have in common. Because no set of words is common to all normative studies, no firm conclusion can be drawn on rating similarity. Nevertheless, correlation coefficients yield a useful approximation of ratings similarity across studies.

Table 4
Correlations With Similar Ratings and Number of Common Words on Which the Estimate Is Based

Sources of Similar Ratings	Number of Words in Common	Present Study	
		Subjective Frequency	Imageability
Desrochers & Bergeron (2000)	796	.95	.96
Bonin et al. (2003)	586	.85	.69
Flieller & Tournois (1994)	555	.86	.76
Hogenraad & Orianne (1981)	513	—	.89
Ferrand et al. (2008)	424	.84	—
Gonthier et al. (in press)	348	.91	.96
Forget (2005)	279	.94	.97
Hammoud (1984)	239	—	.96
Robin (2006)	148	—	.97
Denis (1975)	100	.87	.93

Note.—All correlation coefficients are statistically significant, $p < .001$.

Table 5
**Correlations of the Logarithm (Base 10) of Objective Frequency
With Subjective Frequency and Imageability**

Sources of Objective Frequency Estimates	Number of Words in Common	Present Study	
		Subjective Frequency	Imageability
New et al. (2004)—Lexique 2	3,505	.67*	.03†
New & Pallier (2005)—Lexique 3	3,494	.56*	.18*
Imbs (1971)	3,481	.60*	-.02
Lété et al. (2004)—MANULEX	3,127	.52*	.42*
Baudot (1992)	3,127	.66*	-.06†
Lambert & Chesnet (2001)—NOVLEX	2,296	.42*	.31*
Verlinde & Selva (2001)	2,260	.60*	-.15*
Engwall (1984)	693	.67*	.12*
Vikis-Freibergs (1974)	631	.46*	.33*
Juilland et al. (1970)	478	.58*	-.18*
Lyne (1985)	346	.16†	-.04

* $p < .001$. † $p < .05$.

Table 4 presents the correlations between the present ratings and ratings drawn from shared words in nine similar studies. Some of these studies were carried out in Canada (Desrochers & Bergeron, 2000; Forget, 2005; Gonthier et al., in press; Hammoud, 1984), and others in France (Bonin et al., 2003; Flieller & Tournois, 1994; Robin, 2006) or Belgium (Hogenraad & Orianne, 1981). The reported coefficients confirm that the subjective frequency and imageability ratings in the present study are highly and positively correlated with those found in other studies (most $r_s \geq .85$). Each of these studies employed a paper-and-pencil survey method for data collection. The high correlations between the norms reported here and those reported previously suggest that the methodological advantages of a computerized procedure (e.g., item-level randomization) may actually be subtle. The lowest correlation coefficient was found with Bonin et al.'s (2003) imageability ratings. This result is not entirely surprising, since the highest correlation involving Bonin et al.'s (2003) norms and the other normative studies was .71 (i.e., with Flieller & Tournois, 1994). Thus, the relatively modest correlation between our imageability norms and those of Bonin et al. (2003) is actually consistent with that observed in other studies. Overall, the level of agreement between the norms reported here and those reported previously, representing samples of participants from a wide range of geographical areas, is quite high for both subjective frequency and imageability.

It is unclear what aspect of Bonin et al.'s (2003) imageability norms might be deflating correlations with other studies. One possibility is that Bonin et al.'s (2003) almost exclusive reliance on words with highly concrete referents created a context that slightly distorted the way in which participants responded in that study. A quick search for words without picturable referents among the low-imageability words in Bonin et al.'s (2003) study yielded only the word *saint*. In contrast, the present study has many abstract words, such as *boutade*, *recel*, *lemme*, and *ignomnie*, in the low-imageability end of the spectrum. The other studies considered here included similarly abstract words.

A closer look at Bonin et al.'s (2003) imageability norms sheds some light on the reasons why their correlation with

other studies is lower than expected. For instance, the participants in their study appear not to have made use of the full lower range of the rating scale (minimum value of 2.52 after rescaling to reflect the 7-point scale, vs. the minimum of 1.18 in the present study), suggesting that the ratings may have been slightly compressed toward the higher end of the scale. If this were indeed the case, a lack of compression in the item variances should be evident with low-imageability items in the Bonin et al. norms. To test this hypothesis, these norms were submitted to a curve-fitting procedure analogous to that performed on the present set of norms. This analysis revealed that a linear function explains most of the variance relating mean imageability ratings and their respective item variances [$r^2 = .70$, $F(1,863) = 2,056.84$, $p < .001$, constant = 12.90, beta = -2.65]. This function indicates that item variance decreased with increasing mean imageability. The addition of a quadratic component to this model explained only an additional 7.7% of the variance [$F(2,862) = 1,469.77$, $p < .001$, constant = -0.80, first beta = 4.93, second beta = -1.01]. In sum, the pattern of results observed with the Bonin et al. data can be contrasted with that reported for the present data, which yielded a predominantly quadratic (inverted-U) function relating item means and item variances (see also Forget, 2005). Whatever the explanation (e.g., the use of only items with concrete referents), there is reason to believe that the participants in the Bonin et al. imageability study were not making full use of the rating scale, which may account for the relatively modest correlations with previously published French-language norms and those reported in this article. This observation highlights the possible role of subtle methodological decisions in determining how participants make use of the rating scale, such as the use of a stimulus set that is broadly distributed on the dimension being measured and instructions to participants that emphasize the importance of using the full range of the rating scale.

Relationship Between Subjective Frequency, Objective Frequency, and Imageability

Table 5 reports the Pearson product-moment correlation between the subjective frequency ratings obtained in

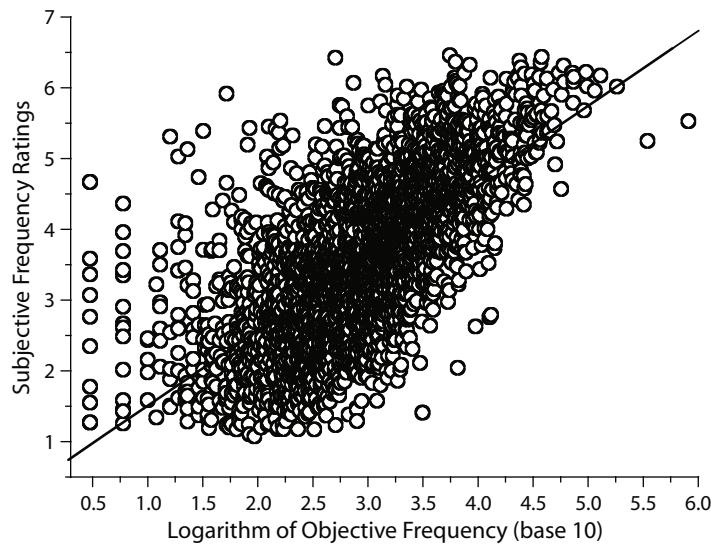


Figure 2. Relationship between subjective frequency and objective frequency.

the present study and the logarithm (base 10) of objective frequencies drawn from 11 distinct studies. The corpora that were used to compute these objective frequency estimates are quite heterogeneous. Some comprise sets of text that were sampled from a variety of genres (Baudot, 1992; Imbs, 1971; Juillard, Brodin, & Davidovitch, 1970; New & Pallier, 2005; New, Pallier, Brysbaert, & Ferrand, 2004), whereas others contain sets of text drawn from specific genres (e.g., novels, Engwall, 1984; newspapers, Verlinde & Selva, 2001; schoolbooks, Lété, Sprenger-Charolles, & Colé, 2004; youth literature, Lambert & Chesnet, 2001; business correspondence, Lyne, 1985). In contrast, the estimates provided by Vikis-Freibergs (1974) are based on free-association responses to the Kent-Rosanoff stimuli. It is worth noting that, irrespective of the types of text or methods of investigation used, the subjective frequency ratings reported in the present study are positively and significantly correlated with all estimates of objective frequency under consideration.

Figure 2 provides a scatterplot of the relationship between the subjective frequency ratings and the log (base 10) of the objective frequency estimates taken from the most widely used source of French word frequency data in language-processing research at the moment (Lexique 2; New et al., 2004). A visual inspection of this graph confirms that most data points are located near the linear regression line. However, the accuracy of the equation relating the two variables breaks down at the low end of the log scale. This observation was verified formally by splitting objective frequency into quintiles and then testing for differences in residual variance among the interval-based categories. This analysis confirmed that the regression equation relating objective frequency and subjective frequency decreases in precision as objective frequency decreases [$F(4,3499) = 26.81, p < .001, MS_e = 1.17, \eta^2 = .03$]. Tukey LSD post hoc tests indicated that words of the lowest frequency are the least precise ($M = 1.00,$

$SD = 1.76$), differing significantly from the second ($M = .74, SD = .93$), third ($M = .72, SD = .89$), fourth ($M = .54, SD = .81$), and fifth ($M = .45, SD = .66$) quintiles. Overall, the general trend was for increasing precision with increasing objective frequency, but the increase from the second to the third quintile fell short of significance, as did that from the fourth to the fifth. The observed pattern is consistent with the observation, noted in the introduction, that objective frequency estimates themselves are less precise at the lower end of the frequency distribution. This result does not support the claim, advanced by Stadthagen-Gonzalez and Davis (2006), that subjective frequency is simply redundant with objective frequency: The former may do much better at measuring the relative frequency of infrequently occurring words (for a similar observation, see Gernsbacher, 1984). Subjective frequency remains an invaluable tool for cross-validating objective frequency estimates in the low-frequency range, whether frequency is being explicitly manipulated or simply controlled in the design of experiments.

As noted in the introduction, the image-evoking capacity of words was expected to decrease as word frequency decreased. This relationship is confirmed by a positive and significant correlation ($r = .26, p < .001$), between the subjective frequency and imageability ratings. This pattern is present in several other data sets (e.g., Benjafield & Muckenheim, 1989, $r = .64$; Bonin et al., 2003, $r = .46$; Desrochers & Bergeron, 2000, $r = .24$; Forget, 2005, $r = .39$; Gonthier et al., in press, $r = .64$) but is nil in others (e.g., Denis, 1975; Flieller & Tournois, 1994). Figure 3 provides a closer look at the relationship between the two sets of ratings collected in the present study. It can be seen that the dispersion of the scatterplot is approximately square above the subjective frequency value of 2.5, which suggests a null or near-null linear relationship. In contrast, for words of lower subjective frequency, the relationship is clearly much stronger, as evidenced by

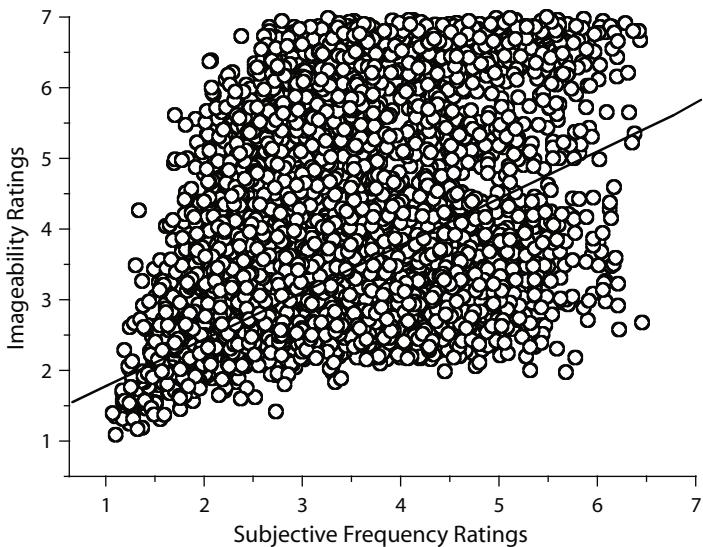


Figure 3. Relationship between imageability and subjective frequency.

the absence of items with both low subjective frequency and high imageability, which suggests that mean subjective frequency rating sets an upper bound for imageability. In practical terms, Figure 3 clearly indicates that the risk of confounding subjective frequency and imageability is particularly acute at the low end of the frequency scale and that the magnitude of the linear relationship observed between the two variables in a given study depends on the relative weight that low-frequency items were given in the item-selection process.

The relationship between imageability and objective frequency (see Table 5) closely depends on the source of frequency data. The correlations between these two variables range from $-.18$ to $.42$. The reasons for this variability are not well understood, but they may involve text-sampling decisions. For example, it could be that the correlations tend to be positive when word tokens are drawn primarily from texts intended for children (e.g., Lambert & Chesnet, 2001; Lété et al., 2004) but negative when sampled from texts intended for adults (e.g., Baudot, 1992; Verlinde & Selva, 2001).

CONCLUSIONS

This study was designed to provide ratings of word subjective frequency and imageability for a large number of French nouns. These items were selected so as to maximize the dispersion of estimates over the 7-point scale used to obtain the ratings, and their distributions were indeed found to be platycurtic. The ratings were also found to satisfy several reliability criteria and to be highly correlated with ratings drawn from similar studies. There is thus compelling evidence that these ratings meet strict psychometric standards.

The large number of items used in this study also permitted us to address specific methodological issues. A serious one concerns the correlation between subjective frequency and imageability. A closer analysis of this re-

lationship revealed that imageability varies relatively independently from subjective frequency when words are perceived as having an intermediate or high frequency of use. However, when words have a low frequency of use, their imagery-evoking value is consistently perceived as being low, even when their referents actually are concrete. This relationship highlights the risk of confounding the two variables in stimulus selection when low-frequency words are considered. It also emphasizes the merits of thoroughly analyzing the psychometric characteristics of lexical variables and paying special attention to the assumption of homoscedasticity (i.e., an even dispersion of residuals about the regression line) when interpreting statistics that are intended to summarize their linear relationship (i.e., correlation coefficients).

AUTHOR NOTE

This research was funded in part by a grant from the Canadian Language and Literacy Research (to Jean Saint-Aubin, Raymond Klein, and A.D.) and a grant from the Social Sciences and Humanities Research Council of Canada (to A.D. and Stanislaw Szpakowicz). We thank Robert Hogenraad and Serge Verlinde for making their data available to us in digital format. We also thank the members of the Cognitive Psychology of Language Laboratory of the University of Ottawa for their assistance with the data collection, as well as Jean Saint-Aubin and Patrick Bonin for their feedback and useful suggestions on an earlier version of this article. All correspondence should be addressed to A. Desrochers, School of Psychology, University of Ottawa, Ottawa, ON, K1N 6N5 Canada (e-mail: alain.desrochers@uottawa.ca).

REFERENCES

- BAAYEN, R. H. (2001). *Word frequency distributions*. London: Kluwer.
- BAAYEN, R. H. (2005). Data mining at the intersection of psychology and linguistics. In A. Cutler (Ed.), *Twenty-first century psycholinguistics: Four cornerstones* (pp. 69-83). Mahwah, NJ: Erlbaum.
- BALOTA, D. A., PILOTTI, M., & CORTESE, M. J. (2001). Subjective frequency estimates for 2,938 monosyllabic words. *Memory & Cognition*, **29**, 639-647.
- BALUCH, B., & BESNER, D. (2001). Basic processes in reading: Semantics affects speeded naming of high-frequency words in an alphabetic script. *Canadian Journal of Experimental Psychology*, **55**, 63-69.

- BARRY, C., HIRSH, K. W., JOHNSTON, R. A., & WILLIAMS, C. L. (2001). Age of acquisition, word frequency, and the locus of repetition priming of picture naming. *Journal of Memory & Language*, **44**, 350-375.
- BAUDOT, J. (1992). *Fréquence d'utilisation des mots en français écrit contemporain*. Montréal: Presses de l'Université de Montréal.
- BEAUCHEMIN, N., MARTEL, P., & THÉORET, M. (1992). *Dictionnaire de fréquence des mots du français parlé au Québec*. New York: Peter Lang.
- BENJAFIKA, J., & MUCKENHEIM, R. (1989). Dates of entry and measures of imagery, concreteness, goodness, and familiarity for 1,046 words sampled from the Oxford English Dictionary. *Behavior Research Methods, Instruments, & Computers*, **21**, 31-52.
- BIRD, H., FRANKLIN, S., & HOWARD, D. (2001). Age of acquisition and imageability ratings for a large set of words, including verbs and function words. *Behavior Research Methods, Instruments, & Computers*, **33**, 73-79.
- BONIN, P., CHALARD, M., MéOT, A., & FAYOL, M. (2001). Age-of-acquisition and word frequency in the lexical decision task: Further evidence from the French language. *Current Psychology of Cognition*, **20**, 401-443.
- BONIN, P., MéOT, A., AUBERT, L., MALARDIER, N., NIEDENTHAL, P., & CAPPELL-TOCZEK, M.-C. (2003). Normes de concréétude, de valeur d'imagerie, de fréquence subjective et de valence émotionnelle pour 866 mots. *Année Psychologique*, **104**, 655-694.
- COHEN, J., COHEN, P., WEST, S. G., & AIKEN, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). Mahwah, NJ: Erlbaum.
- COLTHEART, V., LAXON, V. J., & KEATING, C. (1988). Effects of word imageability and age of acquisition on children's reading. *British Journal of Psychology*, **79**, 1-12.
- COLTHEART, V., & WINOGRAD, E. (1986). Word imagery but not age of acquisition affects episodic memory. *Memory & Cognition*, **14**, 174-179.
- CORTESE, M. J., & FUGGETT, A. (2004). Imageability ratings for 3,000 monosyllabic words. *Behavior Research Methods, Instruments, & Computers*, **36**, 384-387.
- CORTESE, M. J., SIMPSON, G. B., & WOOLSEY, S. (1997). Effects of association and imageability on phonological mapping. *Psychonomic Bulletin & Review*, **4**, 226-231.
- DE GROOT, A. M. B. (1989). Representational aspects of word imageability and word frequency as assessed through word association. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **15**, 842-845.
- DENIS, M. (1975). *Représentation imagée et activité de mémorisation*. Paris: Centre National de la Recherche Scientifique.
- DENIS, M. (1979). *Les images mentales*. Paris: Presses Universitaires de France.
- DESROCHERS, A., & BERGERON, M. (2000). Valeurs de fréquence subjective et d'imagerie pour un échantillon de 1,916 substantifs de la langue française. *Canadian Journal of Experimental Psychology*, **56**, 274-325.
- ENGWALL, G. (1984). *Vocabulaire du roman français (1962-1968): Dictionnaire des fréquences*. Stockholm, Sweden: Almqvist & Wiksell.
- FERLAZZO, F., CONTE, S., & GENTILOMO, A. (1993). Event-related potentials and recognition memory: The effects of word imagery value. *International Journal of Psychophysiology*, **15**, 115-122.
- FERRAND, L., BONIN, P., MéOT, A., AUGUSTINOVA, M., NEW, B., PALLIER, C., & BRYSBAERT, M. (2008). Age of acquisition and subjective frequency estimates for all generally known monosyllabic French words and their relation with other psycholinguistic variables. *Behavior Research Methods*, **40**, 1049-1054.
- FLIELLER, A., & TOURNOIS, J. (1994). Imagery value, subjective and objective frequency, date of entry into the language, and degree of polysemy in a sample of 998 French words. *International Journal of Psychology*, **29**, 471-509.
- FORGET, H. (2005). Valeurs d'imagerie et de fréquence subjective de 354 mots du vocabulaire sexuel de la langue française. *Canadian Journal of Behavioral Science*, **37**, 49-69.
- 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.
- GLASS, A., & MEANY, P. (1978). Evidence for two kinds of low-typical instances in a categorization task. *Memory & Cognition*, **6**, 622-628.
- GONTHIER, I., DESROCHERS, A., THOMPSON, G. L., & LANDRY, D. (in press). Normes d'imagerie et de fréquence subjective pour 1,760 mots monosyllabiques de la langue française. *Canadian Journal of Experimental Psychology*.
- HAMMOUD, R. (1984). *Utilisation de l'image mentale et du champ d'associations dans l'enseignement du vocabulaire arabe à des débutants adultes francophones*. Québec: Centre international de recherches sur le bilinguisme de l'Université Laval.
- HOGENRAAD, R., & ORIANNE, E. (1981). Valences d'imagerie de 1130 mots de la langue française parlée. *Psychologica Belgica*, **21**, 21-30.
- HOPKINS, H. W. (2000). Measures of reliability in sports medicine and science. *Sports Medicine*, **30**, 1-15.
- IMBS, P. (ED.) (1971). *Études statistiques sur le vocabulaire français: Dictionnaire des fréquences. I. Tables alphabétiques*. Paris: Didier.
- JAMES, C. T. (1975). The role of semantic information in lexical decisions. *Journal of Experimental Psychology: Human Perception & Performance*, **1**, 130-136.
- JUILLAND, A., BRODIN, B., & DAVIDOVITCH, C. (1970). *Frequency dictionary of French words*. The Hague: Mouton.
- KÖHLER, R. (1986). *Zur linguistischen Synergetik: Struktur und Dynamik der Lexik*. Bochum: Brockmeyer.
- KOSSLYN, S. M., THOMPSON, W. L., & GANIS, G. (2006). *The case for mental imagery*. New York: Oxford University Press.
- KROLL, J. F., & MERVES, J. S. (1986). Lexical access for concrete and abstract words. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **12**, 92-107.
- LAMBERT, E., & CHESNET, D. (2001). NOVLEX: Une base de données lexicales pour les élèves de primaire. *L'Année Psychologique*, **101**, 277-288.
- LÉTÉ, B., SPRENGER-CHAROLLES, L., & COLÉ, P. (2004). MANULEX: A grade-level lexical database from French elementary school readers. *Behavior Research Methods, Instruments, & Computers*, **36**, 156-166.
- LUPKER, S. J., BROWN, P., & COLOMBO, L. (1997). Strategic control in a naming task: Changing route or changing deadlines? *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **23**, 570-590.
- LYNE, A. A. (1985). *The vocabulary of French business correspondence: Word frequencies, collocations and problems of lexicometric method*. Geneva: Slatkine-Champion.
- MONSELL, S. (1991). The nature and locus of word frequency effects in reading. In D. Besner & G. W. Humphreys (Eds.), *Basic processes in reading: Visual word recognition* (pp. 148-177). Hillsdale, NJ: Erlbaum.
- MONSELL, S., DOYLE, M. C., & HAGGARD, P. N. (1989). Effects of frequency on visual word recognition tasks: Where are they? *Journal of Experimental Psychology: General*, **118**, 43-71.
- NEW, B., & PALLIER, C. (2005). *Manuel de Lexique 3*. Unpublished document, Laboratoire de psychologie expérimentale, Université René-Descartes, Boulogne-Billancourt, France.
- NEW, B., PALLIER, C., BRYSBAERT, M., & FERRAND, L. (2004). Lexique 2: A new French lexical database. *Behavior Research Methods, Instruments, & Computers*, **36**, 516-524.
- NORRIS, D. (2006). The Bayesian Reader: Explaining word recognition as an optimal Bayesian decision process. *Psychological Review*, **113**, 327-357.
- PAIVIO, A. (1971). *Imagery and verbal processes*. New York: Holt, Rinehart & Winston.
- PAIVIO, A. (1986). *Mental representations: A dual coding approach*. New York: Oxford University Press.
- PAIVIO, A., YUILLE, J. C., & MADIGAN, S. A. (1968). Concreteness, imagery and meaningfulness values for 925 nouns. *Journal of Experimental Psychology*, **76**, 1-25.
- PEREA, M., & CARREIRAS, M. (2003). Sequential effects in the lexical decision task: The role of the item frequency of the previous trial. *Quarterly Journal of Experimental Psychology*, **56A**, 385-401.
- RAMAN, I., & BALUCH, B. (2001). Semantic effects as a function of reading skills in word naming of a transparent orthography. *Reading & Writing: An Interdisciplinary Journal*, **14**, 599-614.
- RICHARDSON, J. T. (1980). Concreteness, imagery, and semantic categorization. *Journal of Mental Imagery*, **4**, 51-58.
- ROBIN, F. (2006). Normes françaises pour des dessins, des noms concrets et des noms abstraits. *Canadian Journal of Experimental Psychology*, **60**, 221-236.

- SABSEVITZ, D. S., MEDLER, D. A., SEIDENBERG, M., & BINDER, J. R. (2005). Modulation of the semantic system by word imageability. *NeuroImage*, **27**, 221-236.
- SCHNEIDER, W., ESCHMAN, A., & ZUCCOLOTTO, A. (2002a). *E-Prime reference guide*. Pittsburgh: Psychology Software Tools.
- SCHNEIDER, W., ESCHMAN, A., & ZUCCOLOTTO, A. (2002b). *E-Prime user's guide*. Pittsburgh: Psychology Software Tools.
- STADTHAGEN-GONZALEZ, H., & DAVIS, C. J. (2006). The Bristol norms for age of acquisition, imageability, and familiarity. *Behavior Research Methods*, **38**, 598-605.
- STRAIN, E., & HERDMAN, C. M. (1999). Imageability effects in word naming: An individual differences analysis. *Canadian Journal of Experimental Psychology*, **53**, 347-359.
- STRAIN, E., PATTERSON, K. E., & SEIDENBERG, M. S. (1995). Semantic effects in single word naming. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **21**, 1140-1154.
- STRAIN, E., PATTERSON, K. E., & SEIDENBERG, M. S. (2002). Theories of word naming interact with spelling-sound consistency. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **28**, 207-214.
- TAYLOR, T. E., & LUPKER, S. J. (2001). Sequential effects in naming: A time-criterion account. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **27**, 117-138.
- THOMPSON, G. L., & DESROCHERS, A. (in press). Corroborating bias indicators: Global and local agreement among objective and subjective estimates of printed word frequency. *Behavior Research Methods*.
- TOGLIA, M. P., & BATTIG, W. F. (1978). *Handbook of semantic word norms*. Hillsdale, NJ: Erlbaum.
- VAN HELL, J. G., & DE GROOT, A. M. D. (1998). Disentangling context availability and concreteness in lexical decision and word translation. *Quarterly Journal of Experimental Psychology*, **51A**, 41-63.
- VERLINDE, S., & SELVA, T. (2001). Nomenclature de dictionnaire et analyse de corpus. *Cahiers de Lexicologie*, **79**, 113-139.
- VIKIS-FREIBERGS, V. (1974). *Fréquence d'usage des mots au Québec: Étude psycholinguistique d'un échantillon de la région montréalaise*. Montréal: Presses de l'Université de Montréal.
- WALKER, I., & HULME, C. (1999). Concrete words are easier to recall than abstract words: Evidence for a semantic contribution to short-term serial recall. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **25**, 1256-1271.
- ZEVIN, J. D., & BALOTA, D. A. (2000). Priming and attentional control of lexical and sublexical pathways during naming. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **26**, 121-135.

SUPPLEMENTAL MATERIALS

The subjective frequency and imageability ratings from this article may be downloaded from brm.psychonomic-journals.org/content/supplemental.

APPENDIX

Verbatim Instructions for the Subjective Frequency Ratings

Dans le langage que nous entendons quotidiennement, les mots de la langue française apparaissent avec des fréquences plus ou moins élevées. Certains mots sont très fréquemment utilisés; d'autres le sont de façon très rare ou même exceptionnelle. Le but de cette recherche est d'évaluer un grand nombre de mots quant à la fréquence—ou bien la rareté—de leur apparition dans le langage. Les mots que vous estimerez être le plus fréquemment utilisés obtiendront une cote de fréquence élevée; ceux qui ne sont utilisés que de manière rare ou exceptionnelle obtiendront une cote de fréquence faible. Prenons l'exemple du mot « heure » et celui du mot « heurtoir ». Le mot « heure » est un mot que nous entendons très souvent prononcer de nombreuses fois dans la même journée (p. ex. « Quelle heure est-il? », « C'est l'heure de dîner », etc.). Pour cette raison, il obtiendra une cote de fréquence élevée. En revanche, le mot « heurtoir » est un mot que nous n'entendons prononcer que très rarement autour de nous. C'est pourquoi il recevra une cote de fréquence faible. Attention: Il ne s'agira pas d'estimer la fréquence ou la familiarité des objets désignés par les mots, mais d'estimer la fréquence d'utilisation des mots eux-mêmes dans le langage.

Les mots sur lesquels nous vous demanderons de porter un jugement vous seront présentés un par un au centre de l'écran d'ordinateur. Nous vous demandons d'inscrire votre évaluation de la fréquence en utilisant le clavier numérique de l'ordinateur. Exprimez votre jugement en utilisant une échelle en sept points, où la valeur 1 marque l'extrémité de l'échelle correspondant à la plus faible fréquence et où 7 marque l'extrémité de l'échelle correspondant à la fréquence la plus élevée. Les mots qui vous paraissent les plus fréquents dans le langage obtiendront la cote 7; par contre, les mots qui vous paraissent pratiquement inutilisés obtiendront la cote 1. Les mots de fréquence intermédiaire obtiendront évidemment une cote intermédiaire appropriée entre les deux extrêmes. Rappelez-vous que vous pouvez utiliser toute l'étendue de l'échelle, de la valeur 1 jusqu'à la valeur 7. Ne vous souciez pas de savoir avec quelle fréquence vous utilisez un chiffre en particulier, du moment qu'il s'agit de votre impression réelle. Il n'y a pas de bonnes ou de mauvaises réponses dans ces évaluations. Assurez-vous toutefois que chacune de vos évaluations reflète bien votre pensée. Essayez d'être précis dans vos évaluations, mais sans vous attarder exagérément sur chaque mot.

Certains mots vous seront présentés deux fois. Lorsqu'ils vous seront présentés, il s'agit simplement de les évaluer à nouveau. Avez-vous des questions sur ce qu'il faut faire? [Questions are answered by the experimenter.] Maintenant, je vais vous faire faire quelques essais d'entraînement pour vous permettre de vous familiariser avec la tâche et avec l'appareillage. [The participant completes the practice trials.] Maintenant, je vais vous laisser travailler seul(e). Le reste de la séance est divisé en quatre listes. Vous pourrez prendre une pause entre les listes pour vous reposer. À la fin de la pause, vous n'aurez qu'à appuyer sur une touche du clavier numérique pour continuer.

APPENDIX (Continued)**Verbatim Instructions for the Imageability Ratings**

Les noms communs de la langue française varient dans leur capacité à susciter la formation d'images mentales représentant des objets ou des évènements. Certains mots évoquent une expérience sensorielle, visuelle ou auditive, très rapidement ou très spontanément, tandis que d'autres n'évoquent cette expérience qu'avec difficulté, c'est-à-dire avec un certain délai, ou même pas du tout. Le but de cette recherche est d'évaluer un grand nombre de mots quant à la facilité—ou bien la difficulté—with laquelle ils donnent lieu à l'apparition d'images mentales. Les mots qui, pour vous, provoquent l'apparition d'une image mentale très rapidement et très facilement obtiendront une cote élevée de valeur d'imagerie; les mots qui provoquent l'apparition de cette image avec difficulté ou encore ne provoquent l'apparition d'aucune image obtiendront une cote faible de valeur d'imagerie.

Pensez, par exemple, au mot « fraise » et au mot « connaissance ». Le mot « fraise » va certainement évoquer pour vous une image très facilement. Pour cette raison, il obtiendra une cote élevée de valeur d'imagerie. En revanche, le mot « connaissance » n'évoquera sans doute une image qu'avec difficulté. C'est pourquoi il recevra une cote faible de valeur d'imagerie. Attention: Il peut arriver que les mots vous fassent penser à d'autres mots qui leur sont associés (p. ex. « couteau » et « fourchette »), mais il est très important que votre évaluation porte spécifiquement sur le mot qui vous est présenté.

Les mots sur lesquels nous vous demandons de porter un jugement seront présentés un par un au centre de l'écran d'ordinateur. Nous vous demandons d'inscrire votre évaluation en utilisant le clavier numérique de l'ordinateur. Exprimez votre jugement en utilisant une échelle en 7 points, où la valeur 1 marque l'extrémité de l'échelle correspondant à la plus faible valeur d'imagerie et où 7 marque l'extrémité de l'échelle correspondant à la valeur d'imagerie la plus élevée. Les mots qui, pour vous, évoquent une image mentale de la façon la plus immédiate et la plus forte obtiendront la cote 7; par contre, les mots qui évoquent une image avec le plus de difficulté ou même n'en évoquent aucune obtiendront la cote 1. Les mots de difficulté intermédiaire obtiendront évidemment une cote intermédiaire appropriée entre les deux extrêmes. Rappelez-vous que vous pouvez utiliser toute l'étendue de l'échelle, de la valeur 1 jusqu'à la valeur 7. Ne vous souciez pas de savoir avec quelle fréquence vous utilisez un chiffre en particulier, du moment qu'il s'agit de votre impression réelle. Il n'y a pas de bonnes ou de mauvaises réponses dans ces évaluations. Assurez-vous toutefois que chacune de vos évaluations reflète bien votre pensée. Essayez d'être précis dans vos évaluations, mais sans vous attarder exagérément sur chaque mot.

Certains mots vous seront présentés deux fois. Lorsqu'ils vous seront présentés, il s'agit simplement de les évaluer à nouveau. Avez-vous des questions sur ce qu'il faut faire? [Questions are answered by the experimenter.] Maintenant, je vais vous faire faire quelques essais d'entraînement pour vous permettre de vous familiariser avec la tâche et avec l'appareillage. [The participant completes the practice trials.] Maintenant, je vais vous laisser travailler seul(e). Le reste de la séance est divisé en quatre listes. Vous pourrez prendre une pause entre les listes pour vous reposer. À la fin de la pause, vous n'aurez qu'à appuyer sur une touche du clavier numérique pour continuer.

(Manuscript received August 25, 2008;
revision accepted for publication November 17, 2008.)