Behav Res
DOI 10.3758/s13428-014-0466-y

Norms and reading times for acronyms in French

Patrick Bonin · Alain Méot · Séverine Millotte · Aurélia Bugaiska

© Psychonomic Society, Inc. 2014

Abstract We collected subjective frequency, age-of-acquisition, and imageability norms for 319 acronyms from French adults. Objective printed frequency, bigram frequency, and lengths in letters, phonemes, and syllables, as well as orthographic neighbors, were computed. The time taken to read acronyms aloud was also recorded. Correlational analyses indicated that the relations between the psycholinguistic variables were similar to those usually found for common words (e.g., highly imageable acronyms were more frequent and learned earlier in life than less imageable acronyms), but were generally weaker in the former than in the latter. Linear mixed-model analyses performed on the reading latencies revealed that the main determinants were the voicing feature of initial phonemes, the type of pronunciation of the acronyms (ambiguous vs. unambiguous, typical vs. atypical characteristics), length (number of letters and number of syllables), together with bigram frequency, printed frequency, and imageability. Both objective frequency and imageability interacted reliably with the ambiguous typical and ambiguous atypical properties. Accuracy was predicted by the number of letters and by imageability factors: More errors occurred on longer than on shorter acronyms, and also more errors on less imageable than on more imageable acronyms. The theoretical and methodological implications of the findings for the understanding of acronym reading are discussed. The entire set of norms and the acronym reading times (and accuracy scores), together with the acronym definitions, are provided as supplemental materials.

Keywords Acronyms · Psycholinguistic norms · Word reading times

Electronic supplementary material The online version of this article (doi:10.3758/s13428-014-0466-y) contains supplementary material, which is available to authorized users.

P. Bonin
Institut universitaire de France, Dijon, France

P. Bonin · S. Millotte · A. Bugaiska
LEAD-CNRS, University of Bourgogne, Dijon, France

A. Méot
LAPSCO-CNRS, University Blaise Pascal, Clermont-Ferrand, France

P. Bonin (✉)
LEAD-CNRS (UMR 5022), University of Bourgogne, Pôle AAFE, Esplanade Erasme, BP 26513, 21065 Dijon Cedex, France
e-mail: Patrick.Bonin@u-bourgogne.fr

Theories of word recognition mainly relate to common words (e.g., nouns, verbs, adjectives, etc.). However, any theory of word recognition must account for how adults read all types of words, pseudowords or meaningful strings of letters such as acronyms (e.g., BBC, NASA, FBI). Acronyms are generally understood to refer to all abbreviations, whether pronounceable or not, that are formed as a combination of the initial letters of a term or sentence (Izura & Playfoot, 2012), and this general meaning is what we used in the present study. Strictly speaking, however, abbreviations can be divided into those that are pronounced as words using a set of grapheme-to-phoneme rules (e.g., NASA)—and the term *acronyms* was initially limited to this kind of abbreviation (Brysbaert, Speybroeck, & Vanderelst, 2009)—and initialisms, which are pronounced as a sequence of letter names (e.g., FBI). Acronyms are fully printed in capital letters, and one difficulty when encountering a new acronym is whether it should be read just like any other word (e.g., NASA) or by saying each letter aloud (e.g., FBI). Acronyms are nowadays widely used in books, newspapers, advertising, and so on. However, the question of exactly how acronyms are processed is still poorly understood by researchers working in the field of word recognition (Slattery, Schotter, Berry, & Rayner, 2011). Recently, Izura and Playfoot (2012) collected norms for about 140 acronyms in English on several psycholinguistic variables (e.g., subjective frequency, imageability, and age of acquisition [AoA]). They also collected acronym reading times. Their findings, which will be discussed below, are interesting and important, since they have the potential to further constrain

Published online: 25 April 2014

the building of theoretical views of word recognition. As we shall review, the way in which acronyms are read by adults cannot be easily accounted for by influential models of word recognition. However, this situation may be due to the fact that only a few studies have considered acronym recognition, as compared to the vast literature that exists on common word reading. The present study therefore helps fill this gap.

In recent years, researchers have become increasingly interested in the collection of norms for long lists of words. To illustrate, ratings of AoA have been provided for 3,000 disyllabic words in English (Schock, Cortese, Khanna, & Toppi, 2012), for 1,493 words in French (Ferrand et al., 2008), and even for 30,000 English words (Kuperman, Stadthagen-Gonzalez, & Brysbaert, 2012). Similar data have been collected for imageability (e.g., Bonin, Méot, Ferrand, & Roux, 2011; Cortese & Fugett, 2004). AoA and imageability are two important variables that have caused considerable debate in the word recognition domain (e.g., Cortese & Schock, 2012; Mermillod, Bonin, Méot, Ferrand, & Paindavoine, 2012; Zevin & Seidenberg, 2002, 2004). Reaction times (RTs) and accuracy scores have also been recorded for a huge number of words in visual word recognition (i.e., megastudies), and more precisely, in the tasks of lexical decision and word reading aloud (e.g., Balota et al., 2007; Ferrand et al., 2011; Ferrand et al., 2010). One of the advantages of having RT and accuracy scores for a large number of word trials is the fact that this makes possible assessing the effect of various potentially important variables on word recognition performance. Likewise, the availability of such databases simplifies the task of assessing the importance of new variables. For instance, Juhasz, Yap, Dicke, Taylor, and Gullick (2011) recently found that sensory experience ratings (SER) of words (i.e., the degree to which a word evokes a sensory experience) predicted a significant amount of variance in lexical decision times in two megastudies of lexical processing (i.e., the English Lexicon Project [Balota et al., 2007] and the British Lexicon Project database [Keuleers, Lacey, Rastle, & Brysbaert, 2012]), even when a large number of other important psycholinguistic variables were taken into account. Juhasz and Yap (2013) also recently collected other norms for sensory experience for a set of 5,000 words in English, and again they found that sensory experience ratings had a reliable influence on word recognition performance. Thus, the use of long lists of words involving several psycholinguistic norms has proven to be a fruitful strategy when performing multiple-regression analyses of both RTs and accuracy in lexical decision and/or word reading, since this makes it possible to identify, among a large set of potential predictors, those that are the most relevant when building models of word recognition (see Balota, Yap, Hutchison, & Cortese, 2013, for a review of megastudies of word recognition). One of the advantages of the multiple-regression approach over the factorial approach lies in the potential of the former to investigate the weight of several (and potentially more) variables at the same time. Moreover, having continuous variables avoids the problem of using a categorical boundary for a variable that is nonlinearly scaled (e.g., word frequency), since this may either magnify or diminish the influence of the variable. Importantly, the use of long lists of words reduces potential selection bias from researchers when designing factorial experiments (Forster, 2000), since it has been shown that experimenters possess knowledge about variables other than those being explicitly studied that may inadvertently influence lexical performance. There is no doubt that megastudies have provided very useful data and have allowed researchers to achieve a better understanding of the representations and processes underpinning lexical decision and word reading of *common words* (Balota et al., 2013). However, one issue that still deserves attention is how acronyms are recognized by adults. Our study is directly related to that of Izura and Playfoot (2012), who collected norms and reading times in English for 146 acronyms. We collected norms and reading times for a set of 319 acronyms in French. Our study can be thought of as an extension of Izura and Playfoot's study to the French language, and we believe it will prove to be very useful, in that it will allow researchers to select materials for designing experiments on acronym processing in French. Moreover, our findings will also help shed light on the issue of whether similar or different processes are involved in acronym and common-word processing. Izura and Playfoot have reported important findings in English that constrain models of word recognition. However, before any strong implications can be drawn, it remains to be seen whether their findings can be replicated and extended to the French language, and whether our study would reveal any novel findings relative to theirs.

How are acronyms processed in adults? Empirical facts and theoretical views

One issue associated with acronym processing is to determine whether acronyms are represented in long-term memory, and are therefore part of the mental lexicon, in the same way as common words. The findings obtained by Laszlo and Federmeier (2007a) with the Reicher–Wheeler task suggest that this is the case. These authors used words (DUCT), pseudowords (DAWK), acronyms (HDTV), and illegal letter strings (GHTS) and compared letter identification in response to these stimuli. Importantly, they found a significant difference between the percentages of identification of acronyms (77 %) and illegal letter strings (73 %), and they argued that this suggested an "acronym superiority effect." Thus, one way to determine whether acronyms behave just like other (common) words has been to look for major effects that are typically found with common words. Consequently, Laszlo and Federmeier (2007b) used event-related potentials to examine whether the repetition-priming effect typically observed for words could also be found in acronym processing. They found that the N400 component for words was reduced the second time that a word was presented, as compared to the first presentation, whereas no such difference was observed for illegal letter strings. Interestingly, although the repetition effect was also found for acronyms, it was only observed on those known by the participants. Using the associative-priming paradigm, Brysbaert et al. (2009) showed that acronym primes (e.g., ISBN) yielded associative-priming effects in lexical decisions on words (e.g., BOOKS) that were similar to the priming effects observed with common words used as primes, and that the acronyms did so independently of whether they were presented in uppercase letters or in unfamiliar formats (e.g., isbn). The similarity of the effects in word and acronym processing supports the hypothesis that familiar acronyms are stored in the mental lexicon in the same way as familiar common words (see also van Elk, van Schie, & Bekkering, 2010, for spatial association activation—i.e., "left" vs. "right" with political party acronyms).

Turning to theoretical views of acronym recognition, Slattery et al. (2011) have claimed that "no present models of word recognition seriously address the issue of abbreviations" (p. 1024). One important issue, therefore, will be to determine whether the current theoretical views of word recognition are able to account for the processing of acronyms. Therefore, the question of how these models are able to account for empirical facts relating to acronym processing is a very important and challenging issue. There are two influential views of word recognition: the dual-route view (Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001) and the single-route view (Seidenberg & McClelland, 1989). It is not our intention to describe them in full here. Instead, we shall simply present their core assumptions.

The dual-route view posits that two pathways are available when reading: a lexical pathway that permits the retrieval of whole-word representations, and a nonlexical pathway that computes a pronunciation for any kind of letter string using a rule-based mechanism that uses grapheme–phoneme correspondences. It is assumed that the two pathways are functionally independent and that they are both involved in the processing of common words. The single-route view assumes that no whole-word representations are stored in memory, but that different links exist between different types of units: orthographic, phonological, and semantic units. Activation spreads continuously between these units to achieve a stable pattern of activation that corresponds to the recognition of an input string (e.g., a word). According to dual-route models, such as the DRC model (Coltheart et al., 2001), legal acronyms (i.e., wordlike strings; e.g., NASA) can be processed by both the lexical and nonlexical routes, whereas illegal acronyms (i.e., initialisms, strictly speaking) cannot be processed by the nonlexical route, since this would generate an unpronounceable output. Thus, the lexical route would be involved in the processing of this type of acronym, but only if they are relatively familiar and stored in memory. Slattery et al. (2011) recorded the eye movements of adults reading sentences comprising either legal or illegal acronyms (e.g., NCAA) using the boundary change paradigm, in order to assess parafoveal and foveal processing. Previews of the acronyms were either identical (e.g., NASA or NCAA), orthographically legal (NUSO), or illegal (NRSB). The acronyms were presented in capital letters, in "lowercase sentences," or in "capital letter sentences." The findings suggested that readers modulated their processing on the basis of low-level visual cues in parafoveal vision, and more particularly, that they exhibited a bias toward processing capitalized letter strings as illegal acronyms in parafoveal vision when the remainder of the sentence was printed in lowercase letters (i.e., they prepared to read the acronyms using the different letter names comprising them).

As we mentioned above, one fruitful strategy for constraining the building of models of visual word recognition has been to collect RT data for long lists of words and to conduct multiple-regression analyses with a potentially important set of predictors. In a study of monosyllabic words, Balota, Cortese, Sergent-Marshall, Spieler, and Yap (2004) found that a myriad of variables (e.g., length, objective frequency, subjective frequency, initial phoneme characteristics, and others) influenced participants' naming and lexical decision responses. Cortese and Khanna (2007) examined the influence of 22 variables on lexical decision and word-naming times for 2,342 words. The variables were the same as those employed by Balota et al. (2004) and were entered in the same order that the previous researchers had used, except that both the AoA and imageability variables were also included in the regression model. It appears that the most important variables to play a role in word recognition included lexical variables such as word frequency or word length, sublexical variables, and lexico-semantic (e.g., AoA) or semantic (e.g., imageability) variables . The authors found that rated AoA predicted word-naming and lexical decision times over and above other variables, and that the imageability variable accounted for unique variance only in lexical decision once AoA was controlled.

Among these variables, word frequency is certainly the most powerful variable affecting word recognition (Murray & Forster, 2004). Measures of word frequency are estimates of how often a particular person has come across that word. Some discussion has focused on how to measure this frequency (e.g., Brysbaert & Cortese, 2011; Brysbaert & New, 2009). As we will discuss below, this aspect is particularly important with regard to acronyms. Turning to the AoA variable, there is no consensus as to the locus of this variable in lexical-processing tasks. In effect, rated AoA has been claimed to act at the phonological level in word reading (Gerhand &

Barry, 1998, 1999), at the semantic level in object categorization (Johnston & Barry, 2005), as well as at the links between semantic and lexical codes (Ellis & Lambon Ralph, 2000; see Johnston & Barry, 2006, for further discussion). More importantly, serious doubts have been raised as to the status of rated AoA measures (Zevin & Seidenberg, 2002). Brysbaert and Cortese (2011) have shown that AoA accounted for less of the explained variance in word recognition times when better objective word frequency measures were used. Finally, turning to imageability ratings, several pieces of evidence can be used to support the idea that imageability is a semantic variable, as was explained by Cortese and Schock (2012). First of all, imageability effects are greater in lexical decision, in which semantic information is considered to be more important than in reading aloud, because the emphasis in the latter task is more on the computation of phonological and articulatory codes. Second, Evans, Lambon Ralph, and Woollams (2012) reported that imageability and semantic-priming effects increase together in lexical decision as nonword foils become more similar to real words. This indicates that when discriminability is low, semantic information becomes more helpful. Finally, this variable is related to the concepts referred to by the words, and not to the characteristics of the words referring to the concepts.

The strategy employed in the so-called megastudies to examine various properties that influence the processing of words was recently used by Izura and Playfoot (2012) to study the reading aloud of acronyms. Subjective norms were collected for acronyms (AoA and imageability), along with several other characteristics, such as each acronym's initial sound, number of letters, printed and rated word frequencies, and bigram and trigram frequencies. The time taken to read aloud the different acronyms was also recorded. The analyses performed on reading times revealed an influence of these variables. Moreover, imageability interacted with acronym typicality, indicating that its influence was stronger for typically pronounced acronyms (i.e., those that were named by spelling aloud each of their letters, such as DVD) than for atypically pronounced acronyms (i.e., acronyms named following the spelling-to-sound correspondences of the language, as in the DOS acronym). Thus, as had been found in other studies with common words, the reading times were influenced by lexical and sublexical factors. According to Izura and Playfoot, the findings indicated that acronym naming is a complex process affected by more variables than is generally considered.

In sum, given the small number of studies that have collected norms and RTs for acronyms, we decided to pursue the work initiated by Izura and Playfoot (2012) for a series of 319 French acronyms. As we said earlier, this kind of study is necessary in order to gain a better understanding of the mechanisms and the representations that underlie the reading of such expressions. We therefore collected norms for 319 French acronyms (Study 1) and examined the influences of the norms and other psycholinguistic variables on acronym reading times (Study 2).

Study 1: Norms for acronyms

Method

Participants

A total of 83 participants were involved in the collection of the acronym norms. All were volunteers and had normal or corrected-to-normal vision. Two groups of participants, who were students of either language therapy school or medicine, took part in the subjective frequency rating task (more participants took part in this rating task than in the other rating tasks because, as is explained in the Stimuli section, we initially started with more acronyms than the final set of 319). The first group comprised 19 adults (17 females, two males) whose mean age was 24.4 years (range 20–40), and the second group comprised 20 adults (15 females, five males) whose mean age was 22.2 years (range 19–29). Twenty-two psychology students provided the AoA ratings (19 female, three males; mean age 20.18 years old, range 17–28), whereas another 22 psychology students (20 female, two males; mean age 18.9 years old, range 18–23), all from the University of Bourgogne (Dijon, France), evaluated the words for imageability.

Stimuli

Acronyms were selected from the Internet and from several acronym dictionaries. We retained acronyms that we thought might reasonably be considered to be widely known (note that acronyms based on English words but that are well-known because they are heard in movies, such as FBI, were also retained). Thus, acronyms that were too technical were not considered for inclusion. When choosing the acronyms, we made sure to include both initialisms and acronyms pronounced as words. In addition, we also asked 39 participants taken from different surrounding areas and having different occupations to provide all of the acronyms that came to mind by writing them down on a sheet of paper. They were given 5 min to do this task. We did not ask the participants to tell us what the component words were for each acronym provided, since in most cases the participants knew the acronyms without thinking about what they stood for. All acronyms generated by the participants were retained. All of these sources combined gave us a list of 414 acronyms.

These acronyms were then rated for subjective frequency. Given that rating 414 acronyms for subjective frequency takes some time, we decided to divide the list into two sublists and to ask two independent groups of participants (20 and 19 adults, respectively) to rate one of these sublists apiece. A 5-

point scale was used to rate the frequency with which the participants thought they read, heard, or produced each acronym, with 1 = *never heard, read, or produced* and 5 = *heard, read, or produced very often*. Any acronym that was given a rating of 1 by at least half of the participants was not retained for further ratings. Consequently, the final list consisted of 319 acronyms. Among these acronyms, none contained numbers (e.g., H_2O), 38 were foreign acronyms (e.g., USA, WWF), eight were letter abbreviations of a word (e.g., TV for *television*), and 74 comprised only nouns (e.g., EEG). Finally, 113 were made up only of consonants, 132 included at least one vowel, and 74 included two or more vowels.

Procedure

We prepared booklets to collect the different ratings of age of acquisition, subjective frequency, and imageability.

AoA ratings For each acronym, the participants had to rate the age at which they thought they had learned each acronym (in its written or oral form). The five values of the scale corresponded to 3-year age bands, with 0–3 at one extreme and 12+ at the other. The values were then converted to numerical values, with 1 = *learned between 0 and 3 years* and 5 = *learned at age 12 or after.*

Subjective frequency ratings As we described above, subjective frequency ratings were collected using a 5-point scale with 1 = *never heard, read, or produced* and 5 = *heard, read, or produced very often*.

Imageability ratings A 5-point scale was used to rate the degree to which participants estimated how easy it was to form a mental image based on the referent of each acronym, with 1 = *very difficult or impossible* and 5 = *very easy* (i.e., for the acronym " NASA," participants were asked to come up with an image based on "NASA"—e.g., an astronaut—and not with an image "OF NASA").

Objective variables Following the procedure described in Izura and Playfoot (2012), the printed frequencies of the acronyms were obtained from the Internet by running the Google search engine. We used the number of hits returned by the Google search engine for French only as the index of printed frequency. The values provided in the Appendix of the supplementary materials are log transformations of the numbers of hits returned for each acronym. In order to compute the number of orthographic neighbors of each acronym (i.e., the number of words that differed by one letter from the target acronym while preserving the identity and position of the rest of the letters in the acronym), together with its bigram frequency, we used the Lexique toolbox, which is freely available on the Internet (www.lexique.org/toolbox/toolbox.pub/). Finally, the number of letters and the numbers of phonemes and syllables were computed by hand for each acronym.

Reliability of the imageability, AoA, and subjective frequency norms

The norms corresponding to each acronym are available in the supplemental material as an excel file. The definitions of all the acronyms are also provided.

The correlations between the scores ($N = 319$) corresponding to the means of even and odd participants and intraclass correlation coefficients [random effects of both participants and items—ICC(2, *k*) in Shrout and Fleiss's, 1979, terminology] are shown in Table 1. The reliability indexes are high, thus indicating that the participants' estimations were generally consistent.

Table 2 reports descriptive statistics for the different psycholinguistic variables. The first aspect of note is that the AoA scores were negatively skewed and located mostly at the top of the scale. To a lesser extent, the same was also found for the subjective frequency variable. By contrast, the distribution of imageability values was nearly uniform. In total, 194 (61 %) items began with a voiced initial sound.

We employed the criteria used by Izura and Playfoot (2012) to categorize acronym pronunciations. Consequently, acronyms read by spelling aloud each of the letter names (e.g., FBI) were classified as *typically pronounced* acronyms. As was explained by Playfoot, Izura, and Tree (2013), regular acronyms obey a single rule, which is that they are to be pronounced by naming each letter aloud, and irregular acronyms are therefore the rest. The acronyms read aloud following the spelling-to-sound correspondences of the language (e.g., NASA) were classified as *atypically pronounced acronyms*.

Acronyms made up of a combination of consonants and vowels and that look like words introduce an ambiguity at the time of reading (e.g., someone who has never seen and heard an acronym such as CABG [i.e., coronary artery bypass graft] will find it very difficult to know how to pronounce it). This ambiguity applies to both typically pronounced (e.g., HIV) and atypically pronounced (e.g., NATO) acronyms. The acronyms that consist entirely of consonants or vowels (e.g., CNN,

Table 1 Correlation (*r*) between even and odd participants, along with intraclass correlation coefficient (ICC) indexes

	r(even, odd)	ICC
Imageability	.86	.82
Subjective frequency	.89	
AoA	.88	.92

Since two different lists of words were used for collecting subjective frequency ratings, no ICC is given for this variable. AoA = age of acquisition

Table 2 Summary of descriptive statistics for the French acronyms

	Min	Max	Mean	*SD*	Q1	Med.	Q3	Skew
AoA	1.67	5	4.19	0.66	3.86	4.39	4.71	–1.23
Imageability	1.37	5	3.27	0.97	2.42	3.33	4.10	–0.09
Number of letters	2	7	3.24	0.95	3.00	3.00	4.00	1.37
Number of phonemes	2	14	5.02	1.44	4.00	5.00	6.00	1.05
Number of syllables	1	7	2.68	0.85	2.00	3.00	3.00	0.05
N	0	4	0.94	1.02	0.00	1.00	2.00	0.69
Subjective frequency	1.16	5	3.93	0.92	3.11	4.21	4.74	–0.64
Printed frequency*	3.22	8.79	6.54	0.97	5.95	6.46	7.25	–0.08
Bigram frequency*	2.00	4.35	3.63	0.41	3.42	3.67	3.92	–1.03

AoA = age of acquisition, *N* = number of orthographic neighbors. * Log-transformed

AOA), which are typically pronounced by naming each letter aloud, were categorized as unambiguous. The combination of these features—that is, pronunciation typicality and pronunciation ambiguity—yielded three different types of acronym pronunciations: (1) ambiguous and typical (e.g., HIV), (2) ambiguous and atypical (e.g., ROM), and (3) unambiguous and typical (DVD). Applying these criteria to our set of acronyms yielded the following distribution: 127 (40 %) of the acronyms had ambiguous and typical pronunciations, 72 (23 %) had ambiguous and atypical pronunciations, and 114 (36 %) had unambiguous and typical pronunciations. Six items could not be classified according to these criteria. These are indicated in bold in the supplemental materials.

The correlations between the collected norms are presented in Table 3. The first aspect of note is that the subjective estimations were highly correlated. Consequently, highly imageable acronyms were estimated to be more frequently encountered and learned earlier in life. It is worth noting that the pattern of correlations of the different variables with subjective frequency is also frequently found with words (e.g., Tsaparina, Bonin, & Méot, 2011, in Russian). Importantly, this pattern was also reported by Izura and Playfoot's (2012) acronym study. However, a stronger relation was observed here between subjective frequency and AoA. This can be partly explained by the different AoA scales used in the two studies. Since acronyms are acquired relatively late in life, the AoA scale used in French led to a relatively large negative skew (it should be noted that no departure from normality was observed in English), thus showing that a high number of acronyms are estimated to be acquired late in life (the scale values corresponding to late age bands were often chosen). At the same time, a negative skew was also observed for subjective frequency ratings in French. It was therefore not surprising that the correlation between AoA and subjective frequency was higher in French than in English. Finally, it is worthy of note that the differences in variabilities (which were generally higher in the English data), and some range restriction properties (since acronyms were estimated to be acquired relatively late and to be frequently encountered, the lower sections of the

Table 3 Correlations among the psycholinguistic variables and item-level word reading times

	Number of Letters	Number of Syllables	Number of Phonemes	*N*	Imageability	Subjective Frequency	Printed Frequency	AoA	Bigram Frequency
Mean reading times	.49***	.13*	.27***	–.17**	–.43***	–.43***	–.37***	.35***	.02
Number of letters		.16**	.40***	–.36***	–.13*	–.08	–.63***	.33***	.21***
Number of syllables			.80***	–.27***	–.08	–.08	–.33***	.10	–.03
Number of phonemes				–.51***	–.07	–.07	–.48***	.12*	–.10
N					.01	–.02	.44***	–.12*	.20***
Imageability						.81***	.21***	–.68***	–.10
Subjective frequency							.15**	–.55***	–.06
Printed frequency								–.39***	.02
AoA									.16**

N = Number of orthographic neighbors, AoA = age of acquisition. Correlations were computed using the 301 items selected for subsequent mixed-model analyses. * $p < .05$. ** $p < .01$. *** $p < .001$

scales were not often chosen) might also account for the difference observed between the two studies.

We also found that acronyms with a higher objective frequency value tended to generate mental images more easily, to be rated as learned earlier in life and to be subjectively encountered more frequently than less frequent objective ones. The underlying correlations, however, were relatively low, which was also the case in the Izura and Playfoot (2012) study. The correlations between objective frequency and subjective frequency and AoA were also lower than those generally found with common words (e.g., Desrochers & Thompson, 2009; Ferrand et al., 2008). A final property worth noting is that the more-frequent acronyms were shorter and had more word neighbors.

Izura and Playfoot (2012) pointed out that some of the correlations that they found were very different from what is typically found with words. For instance, they found a negative correlation between number of letters and number of syllables. However, unlike these authors, we found that acronyms consisting of many letters also had more syllables. Again, unlike the authors above, we found a positive relationship between the number of letters and the number of phonemes, and furthermore, we did not find that the number of letters was positively correlated with objective frequency or with the number of orthographic neighbors. In fact, precisely the opposite correlations were found. Given these differences between the two studies, we decided to take a closer look at the English data obtained by Izura and Playfoot and provided in their supplementary materials. We therefore reanalyzed their data and discovered reporting errors in the Izura and Playfoot article at the level of the Number of Letters factor. Indeed, all of the signs reported for the correlations between number of letters and all of the other norms were inverted. As a result, the discrepancies between the two studies that we noted above before the reanalysis of the data were generally not actually as high as we thought. Nevertheless, it remains true that the correlations observed on acronyms were generally lower than those observed for words. However, acronyms may be more distinctive than they seem, and the boundaries between the different types of acronyms may be more rigid than initially thought. As a result, many characteristics of acronyms, and more particularly those relating to length (in terms of letters, syllables, or phonemes) or bigram frequency probably differ depending on the type of acronym.[1] For example, although the acronym DVD (which has an unambiguous typical pronunciation and is three syllables long) and the acronym ROM (which has an ambiguous atypical pronunciation and is one syllable long) have the same number of letters, they have different numbers of syllables. The analyses performed with all three types of acronyms included in the study could potentially be different from those performed within each category/type of acronym. In order to get a clearer picture of the differences that might exist between the three types of acronyms, we ran a number of additional analyses on the differences between the means. These are summarized in Table 4.

The first aspect of note is that, as we anticipated, differences were observed essentially on the length measures. In particular, the ambiguous atypical acronyms had more letters, but fewer syllables and phonemes, than the other two categories. More surprisingly, the number of phonemes was also lower for ambiguous typical acronyms than for unambiguous typical acronyms.

Other noticeable differences between the three types of acronyms concerned objective frequency and the number of orthographic neighbors. Objective frequency was lower for the ambiguous atypical acronyms, and the number of neighbors was higher for the unambiguous typical acronyms than for the other two categories. Finally, subjective frequency and imageability did not exhibit reliable differences, whereas AoA values were higher for the ambiguous atypical acronyms. Despite the lower number of acronyms used in the English study, nearly the same pattern of results was found. The differences between French and English concerned (1) the number of orthographic neighbors, for which considerable differences emerged between the ambiguous typical category and the other two groups, and (2) AoA and objective frequency, for which no reliable differences were apparent between the three types of acronyms. Concerning these two aspects, the means suggest that they could be partly explained by the lower number of acronyms used by Izura and Playfoot (2012).

Turning now to the correlations between the norms within each category of acronyms, Table 5 shows that these sometimes differed dramatically from the (global) correlations reported in Table 3. This was particularly true for the correlations between the number of letters and the other length measures, which were clearly higher in each category of acronyms than the overall correlation. The differences in length between the three categories thus, to some extent, make the global correlations unsuitable for the study of their relations with other independent variables (IVs). The same property was observed for orthographic neighborhood. One somewhat reassuring aspect, however, was that both signs and reliability were, except in a very few cases, the same, both within the individual categories of acronyms and when they were taken as a whole.

[1] We thank an anonymous reviewer for pointing this out to us.

Table 4 Pairwise comparisons for the three types of acronyms (unambiguous typical, ambiguous typical, and ambiguous atypical, given in that order) in the French and English studies

	French				English			
	p values	Pairwise comparisons			p values	Pairwise comparisons		
Number of letters	< .001	2.8	3.0	4.3	< .001	3.3	3.2	4.3
Number of syllables	< .001	2.9	3.0	1.8	< .001	3.3	3.1	2.4
Number of phonemes	< .001	5.8	4.8	4.3	< .001	6.5	5.1	4.2
N	< .001	0.5	1.4	0.8	< .001	1.4	4.1	0.8
Imageability	> .1	3.4	3.2	3.3	> .10	5.1	5.0	5.5
Subjective frequency	> .1	4.0	3.9	4.0	> .1	2.8	2.8	2.5
Printed frequency (log10)	< .001	6.6	6.7	6.2	> .1	7.7	7.9	7.5
AoA	< .001	4.0	4.2	4.4	> .10	15.2	14.4	14
Bigram frequency (log10)	< .001	3.5	3.7	3.7	< .001	3.0	3.8	4.0

*N*s = 114, 125, and 62 for the unambiguous typical, ambiguous typical, and ambiguous atypical acronyms, respectively. Solid lines indicate $p < .001$, long-dashed lines $p < .01$, and short-dashed lines $p < .05$. For each norm, a Bonferroni correction was applied.

Other noticeable differences between the categories were as follows. First of all, the correlations between the number of orthographic neighbors and the different length measures (as well as objective printed frequency) were generally lower (in absolute values) for the ambiguous atypical acronyms than for the two other categories. Second, the same property was observed for the relations between the number of syllables and the number of letters, the number of phonemes, and objective frequency. Finally, the correlation between subjective frequency and bigram frequency was negative and reliable for ambiguous typical acronyms, but not significant in the other two groups.

In the Izura and Playfoot (2012) study, the number of ambiguous atypical acronyms was quite small ($N = 13$), with the result that the correlations computed for this group of acronyms had only a low level of reliability. We do not therefore report the same detailed analyses. However, it should be noted that the same properties were generally observed. For instance, the global correlation between number of letters and number of syllables (with no a priori transformation) in their study was .33, whereas the within-category correlations (between the numbers of letters and of syllables) were 1.0, .72, and .06 for the unambiguous, ambiguous typical, and ambiguous atypical acronyms, respectively.

Study 2: Naming times for acronyms

Method

Participants

A group of 23 psychology students (21 females, two males; mean age 21 years old, range 19–24) from the University of Bourgogne (Dijon, France) took part in the acronym reading task. The task was performed individually. The participants were all volunteers and had normal or corrected-to-normal vision.

Table 5 Global correlations and within-category acronym correlations

	2	3	4	5	6	7	8	Bigram Frequency (log)
1. Nb of letters	.16** [.85, .99, .61] ***, ***, ***	.40*** [.85, .89 .87]	–.36*** [–.73, –.53, –.33] **, ***, ns	–.13* [–.14, –.15, –.18] ns, ns, ns	–.08 [–.08, –.13, –.10] ns, ns, ns	–.63*** [–.73, –.70, –.59] ns, ns, ns	.33*** [.31, .30, .37] ns, ns, ns	.21*** [.05, .16, .39] ns, *, ns
2. Nb of syllables		.80*** [.98, .90, .71] ***, ***, ***	–.27*** [–.66, –.53, –.09] ns, ***, **	–.08 [–.11, –.16, –.00] ns, ns, ns	–.08 [–.07, –.15, –.04] ns, ns, ns	–.33*** [–.65, –.70, –.32] ns, **, ***	.10 [.25, .31, .01] ns, ns, *	–.03 [–.11, .18, .14] .05, ns, ns
3. Nb of phonemes			–.51*** [–.63, –.52, –.22] ns, **, *	–.07 [–.13, –.07, –.14] ns, ns, ns	–.07 [–.12, –.06, –.13] ns, ns, ns	–.48*** [–.62, –.67, –.46] ns, ns, *	.12* [.25, .22, .27] ns, ns, ns	–.10 [–.12, .10, .30] ns, .01, ns
4. Nb of orthographic neighbors				.01 [.10, .03, .03] ns × 3	–.02 [.02, –.04, .11] ns, ns, ns	.44*** [.62, .47, .28] ns, **, ns	–.12* [–.30, –.21, –.10] ns, ns, ns	.20*** [–.08, .26, .24] **, *, ns
5. Imageability					.81*** [.84, .80, .78] ns, ns, ns	.21*** [.22, .16, .31] ns, ns, ns	–.68*** [–.69, –.72, –.60] ns, ns, ns	–.10 [.01, –.24, .03] ns, ns, ns
6. Subjective frequency						.15** [.15, .12, .23] ns, ns, ns	–.55*** [–.58, –.54, –.51] ns, ns, ns	–.06 [.11, –.30, .13] **, ns, **
7. Objective frequency (log)							–.39*** [–.39, –.39, –.41] ns, ns, ns	.02 [.13, .02, –.12] ns, ns, ns
8. AoA								.16** [.04, .16, .26] ns, ns, ns

Nb = number. For each pair of norms, the global correlation is shown in the first line, and the within-category correlations are given in the second line in brackets. Given the numbers of acronyms in each group (*N*s = 114, 125, and 62), the correlations are reliably different from 0 at $p < .01$ if the absolute values of the observed correlations are above .24, .23, and .32, respectively. Pairwise comparisons are presented in the third line, in the following order: unambiguous typical versus ambiguous typical, unambiguous typical versus ambiguous atypical, and ambiguous typical versus ambiguous atypical. $^{*}p < .05$. $^{**}p < .01$. $^{***}p < .001$. ns, $p > .05$

Stimuli

The list of 319 acronyms normed in the previous study was used.

Procedure

We used a Macintosh computer running PsyScope software (Cohen, MacWhinney, Flatt, & Provost, 1993) to run the acronym-reading task. The computer controlled the presentation of the acronyms and recorded the RTs. After a ready signal (a plus sign) presented for 2,000 ms in the center of the screen, an acronym was presented. The participants had to say the acronym aloud as quickly as possible. A voice key connected to the button-box and the computer (and running via PsyScope) was used to record the latencies (in milliseconds). The interval between trials was set to 1,500 ms. Participants took short breaks (of 1 min) every 100 trials. The entire session lasted about 30 min.

Results

Scoring of the RT data

In all, 535 trials (7.3 %) were discarded because of technical problems (309, 4.2 %), incorrect pronunciations (145, 2 %), too-long hesitations (58, 0.8 %), or letter inversions (23, 0.3 %). We also eliminated one item (CLIS) because it elicited an error rate of over 50 %. Trials with naming times lower than 300 ms (three trials) were also excluded from the analyses. Six acronyms (listed in the supplemental materials) could not be classified according to the a priori criteria of typicality and ambiguity. In addition, 11 acronyms that were pronounced in an unexpected way (e.g., as a word when the expected pronunciation was letter by letter) by more than 50 % of the participants were set apart. The analyses reported below were therefore performed on 301 acronyms. The reading times (and accuracy scores) corresponding to each acronym are available in the supplemental materials.

Correlations and linear mixed-model analyses

The correlations between naming times and the psycholinguistic variables are presented in Table 3. As can be seen from this table, they were of the same sign, but generally somewhat higher, than those reported in the English study of Izura and Playfoot (2012). Naming times were longer for longer acronyms (with length defined in terms of the number of letters or of phonemes) and for acronyms that were estimated to be acquired later in life. Shorter naming times were associated with more imageable or more frequent (according to either rated or objective frequency norms) acronyms.

In order to compare our results with those of Izura and Playfoot (2012) in English, we started by running the same analysis on the log-transformed naming times. Participants were treated as a random factor on the basis of which adjustments of the intercept were performed using the mixed model procedure of SPSS 20. The characteristics of the items reported in Table 1 were treated as fixed effects. Two different analyses were conducted, with number of syllables and number of phonemes taken as measures of phonological length. Dummy variables were used to code the voicing feature of the initial phoneme of the acronym (reference category = voiced) and the print-to-pronunciation patterns of the three types of acronym (reference category = unambiguous typical acronyms).

Since Izura and Playfoot (2012) found that certain interaction terms between AoA, imageability, and frequency and the type of acronym were reliable, we also took these terms into account in the equation. In some ways, the Typicality factor used here for acronyms is analogous to the Regularity factor used in word-reading studies. Since the latter factor has been found to interact with semantic (i.e., imageability) and lexical (i.e., word frequency) variables in word-reading studies (e.g., Cortese & Simpson, 2000; Cortese, Simpson, & Woolsey, 1997), it was important to determine whether such interactions would also hold true in the case of acronym reading.

A quadratic term for the number of letters was also included, since a type of relation between this term and RTs had been observed in previous word recognition studies (e.g., New, Ferrand, Pallier, & Brysbaert, 2006). Finally, for the purposes of comparison with the English study only, a quadratic term for imageability was also included in the equation.[2]

In order to report coefficients comparable to the standardized coefficients provided in multiple linear regressions, all the continuous IVs were transformed into *z* scores before the computation of the quadratic terms and the interaction terms. The results for the linear mixed-model analysis (LMM), with the number of syllables taken as an acronym's phonological length, are reported in Table 6. (It is important to note that the results were nearly the same when the number of phonemes was introduced as an acronym's phonological length in the equation.)

Several findings were similar to those reported by Izura and Playfoot (2012) but, as we shall describe below, we also observed several discrepancies between our findings and theirs. As far as the common findings are concerned, we found that (1) ambiguous atypical expressions were named more slowly than typical ones; (2) reliable main effects emerged of number of letters (positive effect), of bigram frequency (negative effect), and of Google printed frequency (negative effect); and (3) reliable positive interactions could be seen between the two dummy variables coding ambiguity and printed Google frequency.

Turning now to the findings that differed between the Izura and Playfoot (2012) study and our own, the following pattern was found. First of all, in the present study, we found that acronyms with higher subjective frequencies were named faster than those with lower frequency acronyms, whereas the opposite was true in the English study. Second, we found that acronyms with a larger number of syllables took longer to read than those with a lower number, whereas this effect had not been reliable in the English study. Third, both AoA and the quadratic term of number of letters were of the same sign in both studies, but they were reliable only in the Izura and Playfoot study. Fourth, the effect of the number of orthographic neighbors was nearly zero in the present study, whereas it was positive and reliable in their study. Fifth, the interaction term between imageability and the dummy variable coding ambiguous atypical items was negative and reliable here, whereas it was not reliable in English (the opposite was observed for the second interaction term including imageability). Finally, none of the interaction terms including AoA and ambiguity reached significance here, whereas one of them (the AoA × Ambiguous Typical interaction term) was reliable in the Izura and Playfoot study.

In order to account for the differences in the findings of the English and French studies, we first had to determine whether the error mentioned above concerning the sign of the correlations between the number of letters and the other IVs in the Izura and Playfoot (2012) English study was also present in their LMM analyses. Although it was not possible to completely rule out this possibility, because the English RTs are not available in their supplemental materials, the overall agreement between the signs of the estimations of the partial coefficients of the length measures suggests that this was not the case. We therefore went on to consider other factors that might account for the discrepancies between the findings in the two languages. One of these was potential problems linked to collinearity. Indeed, a large number of IVs were included in the analyses, and some of them were more highly correlated in the present study than in the Izura and Playfoot study. To explore this possibility, we first identified the variables with

[2] The same results were found when the quadratic term of imageability was removed from the equation.

Table 6 Linear mixed-model analyses on word-reading latencies

	All Variables (R^2 = .365)		Subjective Frequency Excluded (R^2 = .353)		Izura and Playfoot (R^2 = .25)
	Coefficients	*t* and *p*	Coefficients	*t* and *p*	*t* and *p*
Step 2					
Ambiguous typical	.043	$1.73^{\dagger}$	.043	$1.73^{\dagger}$	–1.33
Ambiguous atypical	.378	12.25^{*}	.378	12.25^{*}	5.43^{**}
Step 3					
Intercept	–.139	–1.40	–.168	–1.70	
Ambiguous typical	.038	1.31	.033	1.12	–1.33
Ambiguous atypical	.331	4.53^{**}	.356	4.87^{**}	5.43^{**}
Voicing	.209	9.20^{**}	.214	9.42^{**}	5.69^{**}
Number of letters	.176	5.15^{**}	.168	4.93^{**}	4.13^{**}
Number of letters (quad.)	–.013	–1.61	–.012	–1.49	$–3.50^{**}$
Number of orthographic neighbors	–.005	–0.35	.004	0.28	2.49^{*}
Imageability (Imag)	–.049	–1.48	–.139	$–6.29^{**}$	–1.59
Imageability (quad.)	–.007	–0.58	.017	1.36	0.62
Subjective frequency (SF)	–.116	$–3.65^{**}$			2.34^{*}
Google printed frequency (PF)	–.094	$–4.01^{**}$	–.090	$–3.81^{**}$	$–3.32^{**}$
Age of acquisition (AoA)	–.036	$–1.65^{\dagger}$	–.027	–1.12	$–2.17^{*}$
Bigram frequency	–.071	$–5.89^{**}$	–.069	$–5.80^{**}$	$–5.02^{**}$
Number of syllables	.095	3.96^{**}	.108	4.50^{**}	0.34
AoA × Ambiguous typical	.014	0.39	–.001	–0.03	4.25^{**}
AoA × Ambiguous atypical	.045	0.95	.051	1.07	0.85
SF × Ambiguous typical	–.110	$–2.65^{**}$			$–1.66^{\dagger}$
SF × Ambiguous atypical	.084	1.62			1.61
PF × Ambiguous typical	.154	5.56^{**}	.147	5.28^{**}	2.84^{**}
PF × Ambiguous atypical	.131	4.09^{**}	.128	3.98^{**}	2.21^{*}
Imag × Ambiguous typical	–.022	–0.46	–.115	$–3.51^{**}$	3.18^{**}
Imag × Ambiguous atypical	–.280	$–5.09^{**}$	–.205	$–5.25^{**}$	–1.08

quad. = quadratic term. $^{*}p < .05$. $^{**}p < .01$. $^{***}p < .001$. $^{\dagger}.05 < p < .1$

a particularly high variance inflation factor (VIF) index (and thus also a high R^2 with the other IVs). We then investigated the stability of their effects when other predictor variables that were strongly related to them were removed. Difficulties due to collinearity are known to be limited to variables having strong relations. Studying the stability of the effects when some of them are excluded from the equation is one of several potential approaches when attempting to determine whether these interrelations give rise to real stability problems (e.g., Berry & Feldman, 1985).

The first variable considered was the number of letters, for which the VIF was 10. As far as this variable is concerned, the removal of the Number of Syllables factor led the VIF to decrease to nearly 5. A positive reliable length effect ($p < .001$) was still found, but the quadratic term associated with the letters number was now negative and reliable at $p < .01$. The elimination of the quadratic term with the syllable length entered in the equation led to a VIF of 6.5, with strictly the same pattern of reliable effects for number of letters and number of syllables that we reported in the analysis above. Given this pattern of results, and because the estimated partial coefficients of the number of letters (and its square) and of the number of syllables were not very different, whether or not all of these variables were entered in the equation, we decided that the collinearity noise was sufficiently weak to keep all of the terms in the equation. The second set of variables that we considered included imageability and subjective frequency, for which the VIFs were 10 and 9, respectively. It is worth mentioning that the bivariate correlation between these two factors was already found to be strong. The VIF interaction terms including them were also important (between 4.2 and 8.3). When all of the terms including one of the two variables were eliminated from the equation, the VIF was greatly reduced (the greatest VIF for all of these terms was 4.5). The absolute values of the main-effect estimation and the *t* value of the other variable became larger, while at the same time, both of the interaction terms between this IV and the variables coding ambiguity were negative and significant.

Finally, in the analysis excluding effects based on imageability, the interaction term between the IV coding ambiguous atypical expressions and AoA became significant ($p < .01$). The other reliable and nonreliable effects were the same (and of the same signs) as in the analysis including all of the terms.

Given that the strong relations found between imageability and subjective frequency with the other IVs led to difficulties in establishing stability within the estimation process, we decided to exclude one of the two variables from the analyses. Since imageability is frequently thought to be a genuine index of semantic code involvement (Evans et al., 2012), and because subjective frequency was not found to supplement objective frequency (the latter variable gave rise to the same kinds of effects that have been observed when studying isolated words), in the remainder of the article we will mainly discuss the analyses that were performed with all terms associated with subjective frequency excluded (the fourth and fifth columns of Table 6).

However, even when we limited ourselves to these analyses, discrepancies with the Izura and Playfoot study (2012) were still observed. These are outlined below. We found reliable effects of the Ambiguity, Voicing, and Number of Letters factors, as well as of imageability, objective frequency of the acronyms, bigram frequency, and number of syllables. Turning to the interaction effects, we found that imageability interacted reliably with both the (ambiguous) typical and (ambiguous) atypical properties (Fig. 1). The same interactions were also reliable with the Objective Frequency factor (Fig. 2).

As far as main effects were concerned, unlike in the English study, where imageability was not reliable, we found that this variable had a facilitatory effect on reading latencies in the present study. In the same way, the number of syllables had a positive reliable effect in French, whereas this effect was not reliable in the Izura and Playfoot (2012) study. Maybe the difference between the two languages is due to the fact that French is a syllable-timed language with fairly clear syllable boundaries, whereas English is a stress-timed language with substantial ambisyllabicity (Levelt, Roelofs, & Meyer, 1999).

The AoA factor was not reliable in the present study, whereas it was significant in the English study, albeit in the opposite direction from the expected one (i.e., shorter RTs for later-acquired acronyms than for early-acquired ones). The main effect of orthographic neighborhood was also not reliable, whereas it was significant in English, with longer naming times being observed for acronyms having a higher number of neighbors than for those having fewer.

In our study, the quadratic term of the number of letters was not reliable, whereas it was negative and reliable in the Izura and Playfoot (2012) study.

Concerning interaction effects with the IVs coding the ambiguity and typicality of the acronyms, in the present study, the effect of imageability was significantly negative for the unambiguous acronyms (reference category) and was significantly more negative for the two ambiguous categories than for the unambiguous acronyms. This suggests that the facilitatory effect of imageability was greater for ambiguous than for unambiguous acronyms (see Fig. 1). It is important to

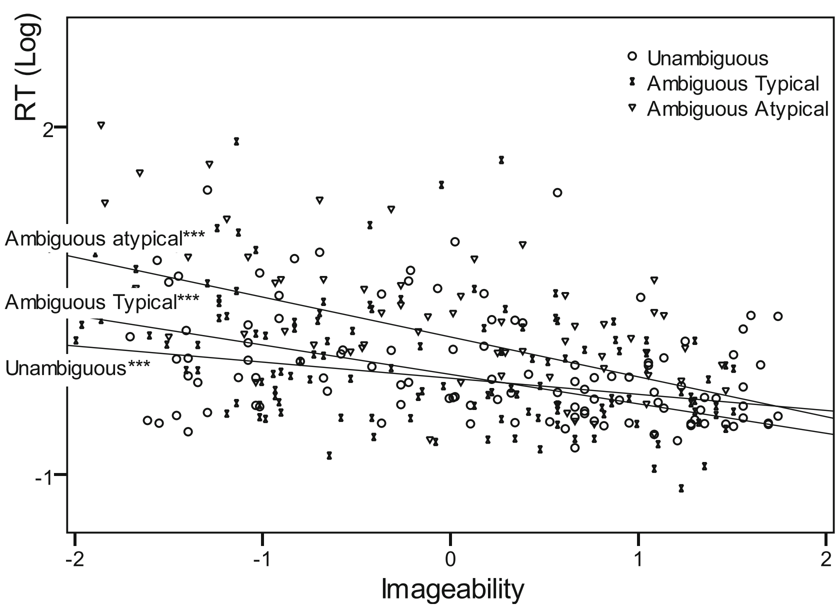

Fig. 1 Interaction between imageability and types of acronyms (regular/unambiguous, ambiguous typical, and ambiguous atypical). The figure presents bivariate plots between the by-item means of the *z* scores of reaction times (RTs, as logarithms) and imageability. Pronunciations categories are shown by different symbols. Lines represent the relation found for each category in the multilevel analysis excluding subjective frequency. Intercepts are computed for voiced acronyms, and values of all other continuous independent variables (IVs) are set at their means. Slopes are the same for all sets of other IVs. ***Slope for the category was significantly different from zero ($p < .001$)

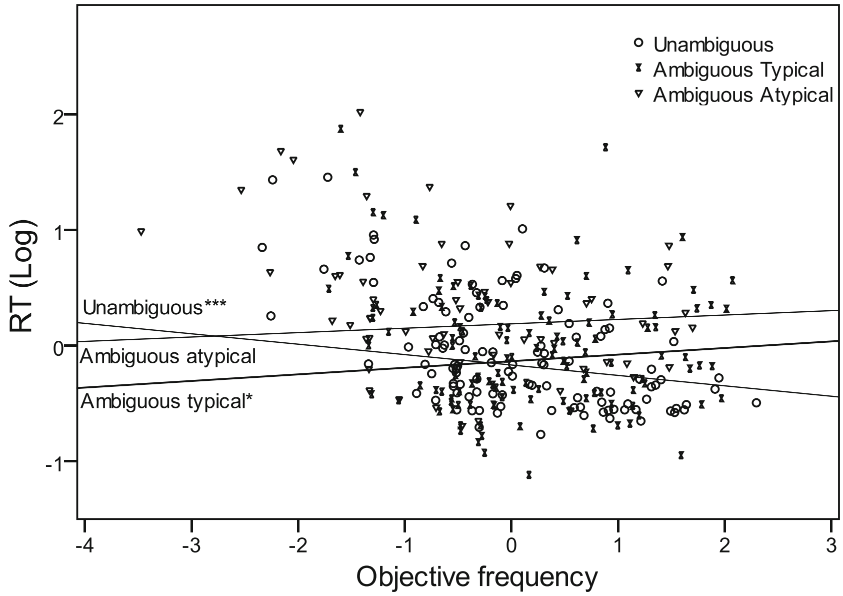

Fig. 2 Interaction between printed frequency and types of acronyms (regular/unambiguous, ambiguous typical, and ambiguous atypical). The figure presents bivariate plots between the by-item means of the *z* scores of reaction times (RTs, as logarithms) and objective frequency (also logs). See Fig. 1 caption for further information. Slopes for the categories followed by asterisks are significantly different from zero: $^{*}p < .05$, $^{***}p < .001$

mention that the same pattern was observed when subjective frequency, instead of imageability, was used. The facilitatory effect of subjective frequency was reliable for unambiguous acronyms, and reliably greater for the other two categories of acronyms.

In English, even though a negative partial coefficient of imageability was found (i.e., a facilitation effect) for the unambiguous acronyms, the effect was not reliable. The effect of imageability was reliably higher (i.e., less facilitatory) for the ambiguous typical acronyms, and no reliable differences were apparent between ambiguous atypical and unambiguous acronyms. Even though no tests were reported for the two ambiguous typical categories, the results suggest that, when other variables were controlled for, no reliable effects of imageability emerged.

Both the AoA term and the interaction terms including these IVs were unreliable in the present study, which suggests that AoA had no noticeable effect on RTs. In English, the AoA effect for the unambiguous category was negative and significant (i.e., a facilitator) and significantly lower than that for the ambiguous typical category.

Finally, the effects of printed frequency were roughly the same in the two studies—that is to say, the frequency effect was significantly negative (i.e., facilitatory) for unambiguous acronyms, and was significantly more so than for the two other types of pronunciations. Changing the reference category in the LMM showed that the frequency effect was not reliable for ambiguous atypical acronyms, whereas it was significantly positive for the ambiguous typical ones (see Fig. 2).

The differences between the English and French studies could be due, in part, to the smaller set of acronyms ($N = 146$) used in the English study, and more particularly, to the smaller number of those coded as ambiguous atypical ($N = 13$).

Like Izura and Playfoot (2012), we performed a multilevel (or mixed) logistic regression with accuracy as the dependent variable. We used the same IVs that we had used in the analysis of RTs. Participant was introduced as a random factor. However, since the 301 expressions for which reading times were analyzed yielded very few nontechnical errors—incorrect pronunciations (132, 1.9 %), too-long hesitations (56, 0.8 %), and letter inversions (20, 0.3 %)—and given the strength of the relations that were found between the IVs, the predictors were not entered simultaneously in the equation, and we instead conducted a stepwise procedure.

The first variable entered in the equation was the number of letters [$b = .098$, $t(6634) = 3.41$, $p < .001$]. The percentage of errors was higher with acronyms having a larger number of letters. In a second step, imageability reached significance [$b = -.066$; $t(6633) = -2.29$, $p < .05$], and the effect of the number of letters was still reliable [$b = .089$, $t(6633) = 3.08$, $p < .01$]: More-imageable acronyms yielded fewer errors than did less-imageable ones. It is worthy of note that the same pattern was observed when subjective frequency was used instead of imageability [$b = -.061$, $t(6633) = -2.13$, $p < .05$, and $b = .093$, $t(6633) = 3.22$, $p < .01$]. No other IVs reached significance after these two steps.

General discussion

A fruitful strategy for gaining a better understanding of how words are read is to establish RTs for long lists of words and to conduct multiple-regression analyses with potentially important variables (e.g., Cortese & Khanna, 2007). Recently, Izura and Playfoot (2012) adopted such a strategy to investigate the reading aloud of acronyms in English. This study is interesting, but limited in scope for the simple reason that it is the only research that has been made available thus far. We therefore considered it necessary to extend this work to a new (and somewhat larger) set of acronyms in French. In the following discussion, we first briefly summarize the main findings about the correlations among the collected variables for acronyms. We then discuss the main findings concerning the determinants of acronym reading.

Summary of correlations among the collected variables for acronyms

We found that acronyms that were easy to form a mental image of were more frequent and were learned earlier in life than those that less easily give rise to a mental image. This kind of relationship has often been reported in studies in which words have been used as stimuli (e.g., Balota et al., 2004; Ferrand et al., 2008). However, the correlations on acronyms found in both our study and that of Izura and Playfoot (2012) were generally lower than those observed for words.[3] These differences could be partially explained by range restrictions on some of the scales. We also found that frequently encountered acronyms were generally shorter and had more neighbors. Finally, when we analyzed the correlations for each type of acronym separately, we found certain differences, as compared to those observed for the whole set of acronyms, which were essentially observed on the length measures.

Determinants of acronym reading: Theoretical and methodological implications

In word naming, it has been found that initial phoneme characteristics are important predictors, since they account for a large amount of the variance (Balota et al., 2004; Bonin, Barry, Méot, & Chalard, 2004; Cortese & Khanna, 2007; but see also Rastle & Davis, 2002, for a more far-reaching discussion of the difficulties related to measuring onset latencies in word reading). Thus, as has been frequently reported for word naming, we found that the “voiced” feature of initial phonemes plays a significant role in reading acronyms.

Types of acronyms Concerning the different types of acronyms, one important finding is the observation that acronyms with ambiguous atypical pronunciations incur a processing cost during reading. Izura and Playfoot (2012) had already noted that ambiguous atypical acronyms were read significantly more slowly than unambiguous acronyms, and that naming times were nearly identical for ambiguous typical and unambiguous acronyms, with both being pronounced by saying each letter aloud. Similar results were observed in the present study.

Following Izura and Playfoot (2012), we propose that acronyms are processed by means of a specific mechanism; that is, each letter of the acronym is read aloud, and this is the default processing for acronyms. As a result, when an acronym looks like a word (and is therefore ambiguous) and, moreover, is not read typically (i.e., is read like a word, and not letter by letter), the default processing mode must be inhibited, and this delays the initialization of reading. Neither the dual-route view (Coltheart et al., 2001) nor the single-route view (Seidenberg & McClelland, 1989) of word reading includes a specific mechanism dedicated to the reading of acronyms, especially those that are unfamiliar, unambiguous, and regular (e.g., LSD). These types of acronyms cannot be processed by the nonlexical route, since this would generate an unpronounceable output. Within the dual-route view, only acronyms that are wordlike strings (e.g., NASA) could be processed by both the lexical and the nonlexical routes.

Orthographic variables (bigram frequency, orthographic neighbors, and length) As in English, acronyms with frequent bigrams were read faster than acronyms with less-frequent bigrams. This observation is not surprising, given that most acronyms consist of only a few letters, and that more weight is therefore given to orthographic features when processing them, especially for those that are unambiguous and typically pronounced.

Unlike in French, in English acronyms with more neighbors took longer to name than those with fewer neighbors. Neighborhood density effects for words have been reported in different tasks (e.g., word naming and lexical decision), but they act differently depending on the task (Mathey, 2001).

Turning to the length variables, we found reliable effects of both number of letters and number of syllables on acronym reading times, with the result that longer acronyms took longer to read than shorter ones. For words, New et al. (2006) found that lexical decision times decreased for short words (two to four letters), were stable for medium word lengths (five to eight letters), and increased sharply for words having nine or more letters. In their multiple regression study of multisyllabic words, Yap and Balota (2009) found a positive effect of number of letters. Finally, we also found a syllable length effect on acronym reading times. This finding is consistent with Yap and Balota’s (2009) observation that the number of syllables is positively correlated with both lexical decision and

[3] It is worthy of note, however, that the correlations are dependent on the types of words used in the corpora (e.g., monosyllabic vs. dissyllabic, or nouns only vs. nouns plus other grammatical categories).

pronunciation latencies (see also Ferrand & New, 2003). A syllable length effect in acronym reading is consistent with the idea that the syllable is one of the multiple codes that mediates lexical access and articulatory processes.

Lexical frequency and AoA An examination of the reading latencies revealed a main effect of printed word frequency, which is undoubtedly the most important variable in word recognition (Murray & Forster, 2004). We also recorded the subjective frequency of the acronyms. The correlation between the two frequency measures was reliable (and positive), but not high. This contrasts, for example, with a French study of 1,493 common words (Ferrand et al., 2008), in which the correlations between the subjective and several objective frequency measures were all high (>.70). The difference in the sizes of the correlations between acronyms and common words could be partly attributable to the way that subjective frequency is measured. Since acronyms are generally judged to be frequently encountered, the values of the scale are not all used to the same degree, and this restriction could therefore influence the values of the correlations. However, another explanation relates to the possibility that subjective frequency norms also capture other dimensions of words, beyond the frequency of encounter. Indeed, subjective frequency correlates strongly with imageability. Since imageability is assumed to be a genuine semantic variable (Evans et al., 2012), this strongly suggests that subjective frequency ratings are influenced by certain semantic features of acronyms. It is worth stressing that we found a strong positive correlation between subjective frequency and imageability, whereas in some studies conducted with common words (e.g., Ferrand et al., 2008), the correlation between the two variables has been found to be reliable but *negative*, with the result that subjectively more frequent common words tend to be less imageable. Such observations are generally due to the types of words that are included in the samples. For instance, in Bonin et al. (2011), the correlation was .158 when computed on the basis of 1,300 nouns, whereas it was –.416 when 496 verbs were included in the sample, and –.397 with 76 adverbs included in the sample.

The reliable effect of printed frequency on reading times (but not on accuracy) is in line with previous studies that have suggested that familiar acronyms are stored in the mental lexicon (Laszlo & Federmeier, 2007a). Since in the dual-route view, objective frequency effects are taken to signal the involvement of the lexical route, the observation of this effect on reading times for acronyms suggests that this view can account for the reading of familiar acronyms. However, our findings indicate that the influence of printed frequency varied as a function of the type of acronym. For example, the higher the printed frequency of the unambiguous typical acronyms, the more quickly their reading was initiated. This is similar to the relationship found in common word reading. However, and in particular, ambiguous typical acronyms with high printed frequencies were associated with a processing cost. This finding was also reported by Izura and Playfoot (2012), and suggests that acronyms consisting of a combination of consonants and vowels (e.g., HIV) introduce an ambiguity at the time of reading because these acronyms are pronounced by saying each letter aloud. Given this ambiguity, it is possible to conjecture that the reading system tends to read them as words, and that this tendency is stronger in the case of frequently seen acronyms. However, the "read-like-a-word" procedure has to be inhibited if this type of acronym is to be read correctly, and this inhibitory process takes some time. It must be stressed that word-reading models do not predict this type of reversed word frequency effect. The Word Frequency × Consistency/Regularity interaction has often been investigated in common word-reading studies (e.g., Cortese & Simpson, 2000), and it has been found that irregular words take longer to produce than regular words, in particular in the case of items of low printed frequency.

The AoA factor was not reliable here, whereas it was reliable in the English study, but the direction of this effect was opposite to that normally observed in lexical processing tasks (Johnston & Barry, 2006). The divergence between the findings in the two studies is difficult to account for. In the present study, the distribution of AoA ratings was not the same as is found for words (e.g., Ferrand et al., 2008), since most of the acronyms were rated as being late-acquired, with a restricted range of variation in the scores. This pattern of distribution of AoA ratings could account for the lack of a reliable effect of this variable. However, there is still some debate concerning what exactly this variable measures. Moreover, at a theoretical level, it has been proposed that age-limited learning effects should not be observed (Zevin & Seidenberg, 2002) because word reading in alphabetic languages involves quasi-regular relationships. These effects should be limited to domains in which arbitrary links are involved, such as object naming (see Mermillod et al., 2012).

Semantic variable (imageability) Unlike in the English study, in which imageability was not reliable, this variable had a facilitatory effect on reading latencies. The finding that imageability, a variable assumed to index semantic code activation (Evans et al., 2012), plays a role in reading acronyms aloud is consistent with previous findings that have shown an associative-priming effect in the processing of acronyms (Brysbaert et al., 2009). Importantly, imageability interacted reliably with acronym type. More precisely, the pattern of findings suggests that the influence of semantics is stronger for acronyms that are not read as words but that resemble words. The question of how exactly imageability influences the reading aloud of different types of words has given rise to some debate in the literature (e.g., Monaghan & Ellis, 2002; Strain, Patterson, & Seidenberg, 2002), and no consensus has as yet been reached.

Conclusion

Acronym reading is a complex process that is affected by more variables than is generally thought (Izura & Playfoot, 2012). We have identified several important determinants of acronym reading, most of which are common to word reading. However, we have also identified specific findings that necessarily imply changes to models of reading if our ultimate goal is to build a general view of reading, rather than to simply explain the reading of isolated common words. Finally, we hope that the norms on acronyms available for the French language (provided as supplemental materials) will be useful to researchers who want to investigate the processing of these items, which are becoming more and more numerous in written language.

Author note This work was supported by a grant from the Institut universitaire de France, given to the first author. The authors thank Mélanie Provost and Séverine Juget for their help in the collection of the data, and Gregory Francis, Cristina Izura, and one anonymous reviewer for their constructive comments on a previous version of the article.

References

Balota, D. A., Cortese, M. J., Sergent-Marshall, S., Spieler, D. H., & Yap, M. J. (2004). Visual word recognition of single-syllable words. *Journal of Experimental Psychology: General, 133,* 283–316. doi:10.1037/0096-3445.133.2.283

Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. A., Kessler, B., Loftus, B., & Treiman, R. (2007). The english lexicon project. *Behavior Research Methods, 39,* 445–459. doi:10.3758/BF03193014

Balota, D. A., Yap, M. J., Hutchison, K. A., & Cortese, M. J. (2013). Megastudies: What do millions (or so) of trials tell us about lexical processing? In J. S. Adelman (Ed.), *Visual word recognition* (Models and methods, orthography and phonology, Vol. 1, pp. 90–115). New York, NY: Psychology Press.

Berry, W. D., & Feldman, S. (1985). *Multiple regression in practice*. Beverly Hills, CA: Sage.

Bonin, P., Barry, C., Méot, A., & Chalard, M. (2004). The influence of age of acquisition in word reading and other tasks: A never ending story? *Journal of Memory and Language, 50,* 456–476. doi:10.1016/j.jml.2004.02.001

Bonin, P., Méot, A., Ferrand, L., & Roux, S. (2011). L'imageabilité: Normes et relation avec d'autres variables psycholinguistiques [Imageability: Norms and relationship with other psycholinguistic variables]. *L'Année Psychologique, 111,* 329–359.

Brysbaert, M., & Cortese, M. J. (2011). Do the effects of subjective frequency and age of acquisition survive better word frequency norms? *Quarterly Journal of Experimental Psychology, 64,* 545–559. doi:10.1080/17470218.2010.503374

Brysbaert, M., & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods, 41,* 977–990. doi:10.3758/BRM.41.4.977

Brysbaert, M., Speybroeck, S., & Vanderelst, D. (2009). Is there room for the BBC in the mental lexicon? On the recognition of acronyms. *Quarterly Journal of Experimental Psychology, 62,* 1832–1842. doi:10.1080/17470210802585471

Cohen, J., MacWhinney, B., Flatt, M., & Provost, J. (1993). PsyScope: An interactive graphic system for designing and controlling experiments in the psychology laboratory using Macintosh computers. *Behavior Research Methods, Instruments, & Computers, 25,* 257–271. doi:10.3758/BF03204507

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review, 108,* 204–256. doi:10.1037/0033-295X.108.1.204

Cortese, M. J., & Fugett, A. (2004). Imageability ratings for 3,000 monosyllabic words. *Behavior Methods and Research, Instrumentation, & Computers, 36,* 384–387. doi:10.3758/BF03195585

Cortese, M. J., & Khanna, M. M. (2007). Age of acquisition predicts naming and lexical-decision performance above and beyond 22 other predictor variables: An analysis of 2,342 words. *Quarterly Journal of Experimental Psychology, 60,* 1072–1082. doi:10.1080/17470210701315467

Cortese, M. J., & Schock, J. (2012). Imageability and age of acquisition effects in disyllabic word recognition. *Quarterly Journal of Experimental Psychology, 66,* 946–972. doi:10.1080/17470218.2012.722660

Cortese, M. J., & Simpson, G. B. (2000). Regularity effects in word naming: What are they? *Memory & Cognition, 28,* 1269–1276. doi:10.3758/BF03211827

Cortese, M. J., Simpson, G. B., & Woolsey, S. (1997). The effects of association and imageability on phonological mapping. *Psychonomic Bulletin & Review, 4,* 226–231. doi:10.3758/BF03209397

Desrochers, A., & Thompson, G. L. (2009). Subjective frequency and imageability ratings for 3,600 French nouns. *Behavior Research Methods, 41,* 546–557. doi:10.3758/BRM.41.2.546

Ellis, A. W., & Lambon Ralph, M. A. (2000). Age of acquisition effects in adult lexical processing reflect loss of plasticity in maturing systems: Insights from connectionist networks. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26,* 1103–1123. doi:10.1037/0278-7393.26.5.1103

Evans, G. A. L., Lambon Ralph, M. A., & Woollams, A. M. (2012). What's in a word? A parametric study of semantic influences on visual word recognition. *Psychonomic Bulletin & Review, 19,* 325–331. doi:10.3758/s13423-011-0213-7

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. doi:10.3758/BRM.40.4.1049

Ferrand, L., Brysbaert, M., Keuleers, E., New, B., Bonin, P., Méot, A., & Pallier, C. (2011). Comparing word processing times in naming, lexical decision, and progressive demasking: Evidence from Chronolex. *Frontiers in Psychology, 2,* 306. doi:10.3389/fpsyg.2011.00306

Ferrand, L., & New, B. (2003). Syllabic length effects in visual word recognition and naming. *Acta Psychologica, 113,* 167–183. doi:10.1016/S0001-6918(03)00031-3

Ferrand, L., New, B., Brysbaert, M., Keuleers, E., Bonin, P., Méot, A., & Pallier, C. (2010). The French Lexicon Project: Lexical decision data for 38,840 French words and 38,840 pseudowords. *Behavior Research Methods, 42,* 488–496. doi:10.3758/BRM.42.2.488

Forster, K. I. (2000). The potential for experimenter bias effects in word recognition experiments. *Memory & Cognition, 28,* 1109–1115. doi:10.3758/BF03211812

Gerhand, S., & Barry, C. (1998). Word frequency effects in oral reading are not merely age-of-acquisition effects in disguise. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24,* 267–283. doi:10.1037/0278-7393.24.2.267

Gerhand, S., & Barry, C. (1999). Age-of-acquisition and frequency effects in speeded word naming. *Cognition, 73,* B27–B36.

Izura, C., & Playfoot, D. (2012). A normative study of acronyms and acronym naming. *Behavior Research Methods, 44,* 862–889. doi:10.3758/s13428-011-0175-8

Johnston, R. A., & Barry, C. (2005). Age of acquisition effects in the semantic processing of pictures. *Memory & Cognition, 33,* 905–912. doi:10.3758/BF03193084

Johnston, R. A., & Barry, C. (2006). Age of acquisition and lexical processing. *Visual Cognition, 13,* 789–845. doi:10.1080/13506280544000066

Juhasz, B. J., & Yap, M. J. (2013). Sensory experience ratings for over 5,000 mono- and disyllabic words. *Behavior Research Methods, 45,* 160–168. doi:10.3758/s13428-012-0242-9

Juhasz, B. J., Yap, M. J., Dicke, J., Taylor, S. C., & Gullick, M. M. (2011). Tangible words are recognized faster: The grounding of meaning in sensory and perceptual systems. *Quarterly Journal of Experimental Psychology, 64,* 1683–1691. doi:10.1080/17470218.2011.605150

Keuleers, E., Lacey, P., Rastle, K., & Brysbaert, M. (2012). The British Lexicon Project: Lexical decision data for 28,730 monosyllabic and disyllabic English words. *Behavior Research Methods, 44,* 287–304. doi:10.3758/s13428-011-0118-4

Kuperman, V., Stadthagen-Gonzalez, H., & Brysbaert, M. (2012). Age-of-acquisition ratings for 30,000 English words. *Behavior Research Methods, 44,* 978–990. doi:10.3758/s13428-012-0210-4

Laszlo, S., & Federmeier, K. D. (2007a). The acronym superiority effect. *Psychonomic Bulletin & Review, 14,* 1158–1163.

Laszlo, S., & Federmeier, K. D. (2007b). Better the DVL you know: Acronyms reveal the contribution of familiarity to single-word reading. *Psychological Science, 18,* 122–126. doi:10.1111/j.1467-9280.2007.01859.x

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences, 22,* 1–38. disc. 38–75.

Mathey, S. (2001). L'influence du voisinage orthographique lors de la reconnaissance des mots écrits [The influence of orthographic neighborhood on visual word recognition]. *Canadian Review of Experimental Psychology, 55,* 1–23.

Mermillod, M., Bonin, P., Méot, A., Ferrand, L., & Paindavoine, M. (2012). Computational evidence that frequency trajectory theory does not oppose but emerges from age-of-acquisition theory. *Cognitive Science, 36,* 1499–1531. doi:10.1111/j.1551-6709.2012.01266.x

Monaghan, J., & Ellis, A. W. (2002). What exactly interacts with spelling–sound consistency in word naming? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28,* 183–206. doi:10.1037/0278-7393.28.1.183

Murray, W. S., & Forster, K. I. (2004). Serial mechanisms in lexical access: The rank hypothesis. *Psychological Review, 111,* 721–756. doi:10.1037/0033-295X.111.3.721

New, B., Ferrand, L., Pallier, C., & Brysbaert, M. (2006). Re-examining word length effects in visual word recognition: New evidence from the English Lexicon Project. *Psychonomic Bulletin & Review, 13,* 45–52. doi:10.3758/BF03193811

Playfoot, D., Izura, C., & Tree, J. (2013). Are acronyms really irregular? Preserved acronym reading in a case of semantic dementia. *Neuropsychologia, 51,* 1673–1683.

Rastle, K., & Davis, M. H. (2002). On the complexities of measuring naming. *Journal of Experimental Psychology: Human Perception and Performance, 28,* 307–314. doi:10.1037/0096-1523.28.2.307

Schock, J., Cortese, M. J., Khanna, M. M., & Toppi, S. (2012). Age of acquisition estimates for 3,000 disyllabic words. *Behavior Research Methods, 44,* 971–977. doi:10.3758/s13428-012-0209-x

Seidenberg, M. S., & McClelland, J. L. (1989). A distributed developmental model of word recognition and naming. *Psychological Review, 96,* 523–568. doi:10.1037/0033-295X.96.4.523

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin, 86,* 420–428.

Slattery, T. J., Schotter, E. R., Berry, R. W., & Rayner, K. (2011). Parafoveal and foveal processing of abbreviations during reading: Making a case for case. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37,* 1022–1031.

Strain, E., Patterson, K., & Seidenberg, M. S. (2002). Theories of word naming interact with spelling–sound consistency. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28,* 207–214. doi:10.1037/0278-7393.28.1.207

Tsaparina, D., Bonin, P., & Méot, A. (2011). Russian norms for name agreement, image agreement for the colorized version of the Snodgrass and Vanderwart pictures and age of acquisition, conceptual familiarity, and imageability scores for modal object names. *Behavior Research Methods, 43,* 1085–1099. doi:10.3758/s13428-011-0121-9

van Elk, M., van Schie, H. T., & Bekkering, H. (2010). From left to right: Processing acronyms referring to names of political parties activates spatial associations. *Quarterly Journal of Experimental Psychology, 63,* 2202–2219. doi:10.1080/17470218.2010.495160

Yap, M. J., & Balota, D. A. (2009). Visual word recognition of multisyllabic words. *Journal of Memory and Language, 60,* 502–529. doi:10.1016/j.jml.2009.02.001

Zevin, J. D., & Seidenberg, M. S. (2002). Age of acquisition effects in word reading and other tasks. *Journal of Memory and Language, 47,* 1–29. doi:10.1006/jmla.2001.2834

Zevin, J. D., & Seidenberg, M. S. (2004). Age of acquisition effects in reading aloud: Tests of cumulative frequency and frequency trajectory. *Memory & Cognition, 32,* 31–38. doi:10.3758/BF03195818