Behav Res (2018) 50:1187–1197
DOI 10.3758/s13428-017-0936-0

Imageability ratings across languages

Adrià Rofes[1] · Lilla Zakariás[2] · Klaudia Ceder[3] · Marianne Lind[4,5] · Monica Blom Johansson[3] · Vânia de Aguiar[1] · Jovana Bjekić[6] · Valantis Fyndanis[4] · Anna Gavarró[7] · Hanne Gram Simonsen[4] · Carlos Hernández Sacristán[8] · Maria Kambanaros[9] · Jelena Kuvač Kraljević[10] · Silvia Martínez-Ferreiro[11] · İlknur Mavis[12] · Carolina Méndez Orellana[13] · Ingrid Sör[3] · Ágnes Lukács[14] · Müge Tunçer[12] · Jasmina Vuksanović[6] · Amaia Munarriz Ibarrola[15] · Marie Pourquie[16] · Spyridoula Varlokosta[17] · David Howard[18]

Published online: 13 July 2017

Abstract Imageability is a psycholinguistic variable that indicates how well a word gives rise to a mental image or sensory experience. Imageability ratings are used extensively in psycholinguistic, neuropsychological, and aphasiological studies. However, little formal knowledge exists about whether and how these ratings are associated between and within languages. Fifteen imageability databases were cross-correlated using nonparametric statistics. Some of these corresponded to unpublished data collected within a European research network—the Collaboration of Aphasia Trialists (COST IS1208). All but four of the correlations were significant. The average strength of the correlations (rho = .68) and the variance explained (R^2 = 46%) were moderate. This implies that factors other than imageability may explain 54% of the results. Imageability ratings often correlate across languages. Different possibly interacting factors may explain the moderate strength and variance explained in the correlations: (1) linguistic and cultural factors; (2) intrinsic differences between the databases; (3) range effects; (4) small numbers of words in each database, equivalent words, and participants; and (5) mean age of the participants. The results suggest that imageability ratings may be used cross-linguistically. However, further understanding of the factors explaining the variance in the correlations will be needed before research and practical recommendations can be made.

Keywords Imageability · Linguistics · Cross-linguistic · Correlations

Electronic supplementary material The online version of this article (doi:10.3758/s13428-017-0936-0) contains supplementary material, which is available to authorized users.

✉ Adrià Rofes
rofesa@tcd.ie

[1] Trinity College Dublin, Dublin, Ireland
[2] University of Potsdam, Potsdam, Germany
[3] Uppsala University, Uppsala, Sweden
[4] University of Oslo, Oslo, Norway
[5] Statped, Oslo, Norway
[6] University of Belgrade, Belgrade, Serbia
[7] Universitat Autònoma de Barcelona, Barcelona, Spain
[8] Universitat de València, València, Spain
[9] Cyprus University of Technology, Limassol, Cyprus
[10] University of Zagreb, Zagreb, Croatia
[11] University of Copenhagen, Copenhagen, Denmark
[12] Anadolu University, Eskişehir, Turkey
[13] Universidad Católica de Chile, Santiago, Chile
[14] Budapest University of Technology and Economics, Budapest, Hungary
[15] University of the Basque Country, UPV/EHU, Leioa, Spain
[16] Basque Center on Cognition Brain and Language, Donostia–San Sebastián, Spain
[17] National and Kapodistrian University of Athens, Athens, Greece
[18] Newcastle University, Newcastle upon Tyne, UK

Imageability (also named *imagery*) is a psycholinguistic variable that is used to indicate how well a word gives rise to a mental image or sensory experience. Imageability ratings are typically collected through paper- or web-based questionnaires. Words such as "apple" or "house," for example, are typically rated as high in imageability, whereas words such as "fact" or "hope" are rated as low in imageability (Paivio, Yuille, & Madigan, 1968). Imageability ratings are used in empirical studies of language. Early examples are the association and analogy work of Francis Galton (1822–1911) and Carl Gustav Jung (1875–1961), or the statistical approaches of Friedrich Wilhelm Kaeding (1843–1928) and George Kingsley Zipf (1902–1950), among many others (Levelt, 2014, p. 449). Imageability ratings are also relevant in neuropsychological and aphasiological studies. Published datasets, varying in length, exist for languages such as Chinese (Ma, Golinkoff, Hirsh-Pasek, McDonough, & Tardif, 2009), English (e.g., Bird, Franklin, & Howard, 2001; Coltheart, 1981; Cortese & Fugett, 2004; Schock, Cortese, & Khanna, 2012), French (Desrochers & Thompson, 2009), Italian (Della Rosa, Catricalà, Vigliocco, & Cappa, 2010; Rofes, de Aguiar, & Miceli, 2015), Japanese (Nishimoto, Ueda, Miyawaki, Une, & Takahashi, 2012), Norwegian (Lind, Simonsen, Hansen, Holm, & Mevik, 2015; Simonsen, Lind, Hansen, Holm, & Mevik, 2013), and Swedish (Blomberg & Öberg, 2015). However, despite some of this excellent work, little is known about the associations of imageability ratings between and within languages (see Blomberg & Öberg, 2015, for a recent analysis of Swedish and English).

Psycholinguistic studies

In a seminal study, Paivio et al. (1968) found a high positive correlation between imageability and concreteness ratings. The authors stressed that these two variables are not the same, because concreteness ratings have a dichotomous nature, while imageability ratings respond to a scale. For example, the word "apple" is concrete because it refers to an object or material, whereas "fact" is not concrete, because it cannot be experienced by the senses. At the same time, "apple" is higher in imageability than "fact," but "apple" is also higher in imageability than "appliance"—even though the latter also refers to a concrete object. Paivio et al. also found that words "associated with sensory experience (usually affective in nature) but not [referring to] specific things or classes of things," such as "affection," "blessing," "ghost," "delirium," and "hierarchy," were higher in imageability than concreteness, whereas words that had an "infrequent association with [a] concrete sensory experience," such as "antitoxin," "encephalon," and "originator," were higher in concreteness than imageability. Despite these arguments, many scholars have interchangeably used the terms *imageability* and *concreteness* (e.g., McMullen & Bryden, 1987; Tyler & Moss, 1997; Tyler, Moss, Galpin, & Voice, 2002). Indeed, the two variables have a high degree of correlation and many similarities (e.g., in the Medical Research Council [MRC] psycholinguistic database [Coltheart, 1981], the two variables correlate at rho = .84).

Psycholinguistic studies have shown that words that are rated high in imageability are processed differently—typically, faster and more accurately—than low-imageability words. This phenomenon has been named the *imageability effect* and has been attributed to different factors: from word differences in age of acquisition (e.g., Carroll & White, 1973; Morrison & Ellis, 1995; Stoke, 1929) or perceptual features (e.g., Plaut & Shallice, 1993), to a separate conceptualization of high- and low-imageability words in the mental lexicon (e.g., Paivio, 2014). Imageability effects have been shown in lexical decision tasks (e.g., Cortese & Schock, 2013; Schwanenflugel, Harnishfeger, & Stowe, 1988; cf. Tyler et al., 2002), in word production paradigms (e.g., Alario et al., 2004; Strain, Patterson, & Seidenberg, 1995; cf. Bleasdale, 1987; Coltheart, Laxon, & Keating, 1988), and in word recognition memory (Cortese, Khanna, & Hacker, 2010; Cortese, McCarty, & Schock, 2015). Imageability effects have also been shown in tasks that use sentences. Holmes and Langford (1976), for example, indicated that healthy individuals recalled sentences constructed with low-imageability words (e.g., "Many factors affected the crucial choice") less accurately than sentences constructed with high-imageability words (e.g., "Many sailors deserted the sinking vessel"). Furthermore, neuroimaging studies have indicated asymmetrical engagement of the left and right perisylvian and entorhinal cortices when healthy individuals hear or read high-imageability as opposed to low-imageability words (e.g., Wise et al., 2000), and also when they perform semantic similarity judgment tasks (e.g., Sabsevitz, Medler, Seidenberg, & Binder, 2005).

Imageability ratings have also been used to control experimental conditions in multiple empirical language studies, because failing to do so may create undesired artifactual results. Naming and reading differences have been shown to disappear when items (i.e., nouns vs. verbs; function vs. content words) were matched for imageability in studies of healthy individuals (Davelaar & Besner, 1988), people with dyslexia (Allport & Funnell, 1981), and people with aphasia (e.g., Franklin, Howard, & Patterson, 1995; Hanley & Kay, 1997; Howard & Franklin, 1988). These results are in contrast with those of other studies on people with aphasia, in which, even when items are matched for imageability, differences were found in naming and sentence completion tasks (e.g., Berndt, Haendiges, Burton, & Mitchum, 2002; Kambanaros & Grohmann, 2015; Rofes, Capasso, & Miceli, 2015). In relation to this, it has been argued that even when nouns and verbs are matched for imageability, differences may still exist in the cognitive processes necessary to process nouns and verbs, because participants take significantly longer to rate

the imageability of nouns than of verbs (Chiarello, Shears, & Lund, 1999). Other work in which experimental stimuli were matched for imageability includes studies finding separate effects of imageability and grammatical class during single-word comprehension using fMRI (Bedny & Thompson-Schill, 2006); studies testing the efficacy of a linguistically motivated protocol to treat people with post-stroke aphasia (de Aguiar et al., 2015); cross-linguistic comparisons of bidialectal children speaking Greek and Cypriot Greek (Kambanaros, Grohmann, & Michaelides, 2013); and effects of context and word class on the retrieval of words in Chinese speakers with aphasia (Law, Kong, Lai, & Lai, 2015). Further discussion of different ways of matching items is beyond the scope of this article.

Neuropsychological and aphasiological studies

Imageability ratings, along with ratings for frequency, word length, regularity of spelling and grammatical category, are considered a source of evidence to identify impairments in specific levels of language processing. Other relevant sources of information include the number and type of errors. Shallice (1988) called this the "critical variable approach." This approach has helped us understand the underlying deficits that explain why a person with deep dyslexia may read "sandal" when given the word "scandal." In this example, it is assumed that reading "scandal" may also activate the word representations of "sandal" and other words, because the words are very similar at the orthographic level. The production of "sandal" will be favored over "scandal" if the person has an impairment in abstract word semantics (in which imageability plays an important role). This is because the word "sandal" has a higher imageability value than "scandal" (see, e.g., Whitworth, Webster, & Howard, 2014, p. 11).

In other studies, people with aphasia after stroke have been shown to retrieve words with high imageability more accurately than words with low imageability, because low-imageability words are typically thought to be more difficult to process at the semantic level (e.g., Luzzatti et al., 2002; Nickels & Howard, 1994). However, opposite results have been found in the same population (Warrington, 1981), as well as in people with neurodegenerative diseases (Breedin, Saffran, & Coslett, 1994).

Motivation for the present study

There is little knowledge about the associations of imageability ratings between and within languages. Imageability is a linguistic variable related to meaning. That is, it reflects the richness of the semantic representations of words (Breedin et al., 1994; Plaut & Shallice, 1993). Therefore, finding cross-linguistic correlations in imageability ratings between words that are semantically equivalent may indicate lexical/semantic similarities across languages. By *semantic equivalence*, we mean words for which a language expert and proficient speaker of both languages provides a direct translation.

Concepts such as "apple" and "house" may be thought of as easy to imagine among speakers of the same language but also among speakers of different languages. This is because they can be represented with a semantically equivalent word (e.g., "apple" vs. "mela" in Italian, or "house" vs. "kuća" in Serbian). At the same time, it could be argued that concepts that are dependent on cultural or socio-economic factors, such as "golf," "handrail," or "priest," may not have the same imageability ratings across languages. Along these lines, Blomberg and Öberg (2015) reported a strong positive correlation between English and Swedish imageability ratings. The authors argued that imageability ratings "can be reliably transferred between the two languages, although some caution should be taken, since for some individual words, some ratings might differ substantially" (p. 351).

If a positive finding for cross-linguistic similarities holds, existing imageability ratings in a widely studied language such as English might be used to norm and validate newly obtained ratings in a less studied language or be used as approximate measures for the new language of interest. This could be useful at a practical level, since many languages possess few or no databases available that yield information on imageability (or on other variables; see, e.g., Proctor & Vu, 1999). This lack of available ratings contrasts with a growing interest in empirical language studies and the need to adapt assessment materials to new languages (Fyndanis et al., 2017). Therefore, such a finding could be used as an argument to utilize existing ratings of other languages and to speed up the adaptation of test materials into less researched languages. To the best of our knowledge, specific criteria to decide whether imageability ratings can be used across languages are nonexistent. In this exploratory study, we assessed different criteria, including the numbers of semantically equivalent words between databases, the correlation values (rho), and the variances explained (R^2). Additionally, we discuss linguistic and cultural factors, intrinsic differences between databases, range effects, and the mean age of participants.

In the present study, members the Collaboration of Aphasia Trialists (CATs; COST Action IS1208) compared ratings of 13 European languages (i.e., Basque, Catalan, Croatian, English, Greek, Cypriot Greek, Hungarian, Italian, Norwegian, Serbian, Spanish, Swedish, and Turkish). These data were collected as part of a project for which we adapted The Comprehensive Aphasia Test to a range of languages spoken in Europe (Fyndanis et al., 2017). We expected to find strong positive correlations between the imageability ratings collected for different languages, provided that the words entered in the correlations were semantically equivalent.

Method

Fifteen imageability databases were considered. These corresponded to unpublished data for ten different languages—namely, Basque, Catalan, Croatian, Greek, Cypriot Greek, Hungarian, Serbian, Spanish, Swedish, and Turkish. We also included four published sets of imageability ratings: three English datasets (Bird et al., 2001; Coltheart, 1981; Cortese & Fugett, 2004), one Italian (Rofes, de Aguiar, & Miceli, 2015), and one Norwegian (Lind et al., 2015; Simonsen et al., 2013). Detailed information on the total numbers of words, informant characteristics, modality, scale used, and references for the published databases can be found in Table 1.

Some differences existed between the databases. The numbers of participants ranged between 20 and 399, and the mean ages of participants between 21 and 65 years. Eight of the 15 databases were collected using a Web-based survey, and six with a paper-based survey. The Greek database was collected with both a paper-based and a Web-based survey. Thirteen of the databases were collected using a 7-point scale, and two using a 5-point scale. Furthermore, the Hungarian database only included nouns. Also, the Norwegian database included nouns, verbs, and adjectives, but only the nouns were used in this study.

Instructions

The imageability ratings in all languages were obtained following the instructions by Paivio et al. (1968):

> The purpose of this experiment is to rate a list of words as to the ease or difficulty with which they arouse mental images. Any word which, in your estimation, arouses a mental image (i.e., a mental picture, or sound, or other sensory experience) very quickly and easily should be given a *high imagery* [imageability] rating; any word that arouses a mental image with difficulty or not at all should be given a *low imagery* [imageability] rating. Think of the words "apple" or "fact." Apple would probably arouse an image relatively easily and would be rated as high [imageability]; fact would probably do so with difficulty and would be rated as low [imageability]. (p. 4)

Semantically equivalent words

The numbers of semantically equivalent words between each of the two English databases and each of the other languages ranged between four and 467. Semantic equivalence between two words was determined as follows: The relevant words for each language were listed, and for each word a language expert (native speaker of the language) indicated a corresponding English word equivalent. For example, for the word "poma" in Catalan, the English equivalent "apple" was given.

Statistical analyses

We correlated (i.e., Spearman's rho coefficient) the semantically equivalent words between all databases based on their English translation. We excluded all correlations in which fewer than 20 words were shared between databases, since

Table 1 Number of words, participant characteristics, and modality in which the data were obtained for each language

Language	Total Words	Participants	Modality	Scale	Reference
Basque	260	43 (mean age = 42, *SD* = 17)	Web-based	7-point	
Catalan	202	32 (university undergraduates)	Web-based	7-point	
Croatian	608	27–46 (mean age = 44, *SD* = 18)	Web-based	5-point	Kuvač Kraljević & Olujić, 2017
English	2020	78 (mean age = 65; *SD* = 9)	Paper-based	7-point	Bird et al., 2001
English	9240	Various databases, not reported	Paper-based	7-point	Coltheart, 1981
English	3000	31 (university undergraduates)	Paper-based	7-point	Cortese & Fugett, 2004
Greek	76	118 (mean age = 42, *SD* = 10.2)	Paper & Web-based	7-point	
Cypriot Greek	80	40 (mean age = 39; *SD* = 14)	Paper-based	7-point	
Hungarian	207	31-37 (mean age = 44, *SD* = 12)	Web-based	7-point	
Italian	292	50 (mean age = 28, *SD* = 11)	Web-based	7-point	Rofes, de Aguiar, & Miceli, 2015
Norwegian	917	399 (mean age = 38, *SD* = 16)	Web-based	7-point	Lind et al., 2015; Simonsen et al., 2013
Serbian	82	30 (mean age = 31, *SD* = 12)	Paper-based	7-point	
Spanish	256	20 (mean age = 22, *SD* = 5)	Web-based	5-point	
Swedish	190	52 (mean age = 41, *SD* = 17)	Web-based	7-point	
Turkish	176	22-29 (mean age = 21; *SD* = 1)	Paper-based	7-point	

In the Croatian database, the values 27–46 indicate the range of participants that rated each word, since different words were rated by different numbers of participants. In the Norwegian database, not all participants rated all the words. That is, "the mean number of ratings for each word in the database [was] 23.5, with a standard deviation of 2.7. The range of ratings [was] 11–52" (Simonsen et al., 2013, p. 439). For the Hungarian and Turkish databases, each number in a pair (31–37 and 22–29, respectively) indicates the number of participants who rated each list, since two lists were used in these studies.

correlations with very few data points are vulnerable to error (see, e.g., Bonett & Wright, 2000). We calculated 105 correlations and excluded 36 because they contained fewer than 20 words in common. In total, we included 69 correlations. The correlation across the English databases of Bird et al. (2001) and Cortese and Fugett (2004) had the greatest number of semantically equivalent words (467). This was followed by the English databases of Cortese and Fugett (2004) and Coltheart (1981), and then the English database of Cortese and Fugett (2004) with the Norwegian database (296 and 251 semantically equivalent words, respectively). We also measured the amount of variation that could be explained by the relationship between each pair of databases. This is called the *variance explained* (R^2). For example, given that Basque and Catalan correlate at a rho = .74, the variance in Basque is "explained" or predicted by the Catalan database by 55%. We calculated this variance using the following formula: $rho^2 \times 100 = \%variance$ (in the example, $.74^2 \times 100 = 55\%$).

Results

A summary of the correlations for the lists of semantically equivalent words across languages can be found in Table 2. A full description of each of the correlations, including the mean imageability for each set of semantically equivalent words, number of equivalent words, rho, *p* value, and variance explained (R^2), can be found in the online supplement (Table S1).

We obtained 65 significant correlations, and four did not lead to significant results (i.e., English [Bird et al.] and Turkish; English [Bird et al.] and Catalan; English [MRC database] and Spanish; Basque and Hungarian). The strengths of the correlations ranged from low (rho = .31 for Norwegian and Turkish) to high (rho = .92 for Catalan and Turkish) and had a moderate median value (rho = .68). The variances explained ranged from 9% (for Norwegian and Turkish) to 85% (for Catalan and Turkish) and had a median value of 46%. A matrix scatterplot representing the variability in numbers of semantically equivalent words and R^2s across datasets can be found in Fig. 1.

Discussion

Imageability ratings have been collected, studied, and used to control experimental conditions in numerous psycholinguistic, neuropsychological, and aphasiological studies (e.g., Hanley & Kay, 1997; Kambanaros et al., 2013; Nickels & Howard, 1994; Paivio et al., 1968; Wise et al., 2000). Excellent work has been put forward to understand the consistency of such ratings within and between languages (i.e., Bird et al., 2001; Blomberg & Öberg, 2015; Cortese & Fugett, 2004; Simonsen et al., 2013). Yet, to the best of our knowledge, no studies had considered this issue across multiple languages and using the same instructions to collect the imageability ratings (Paivio et al., 1968). In this study, members of CATs addressed this issue across 13 European languages. Imageability ratings often correlated across languages. The median strength across correlations was moderate (rho = .68), and the variance explained reached 46%. This implies that at least 54% of the variation in this dataset was due to factors other than imageability.

The finding significant of correlations across databases can be explained by the fact that imageability is a linguistic variable that reflects the richness of the semantic representation of a word (Breedin et al., 1994; Plaut & Shallice, 1993), and such representations should be relatively similar within and between languages (e.g., Bird et al., 2001; Blomberg & Öberg, 2015; Cortese & Fugett, 2004; Cortese et al., 2012). Even though associations across languages possibly exist, our present results should be interpreted cautiously, since the moderate strength of the correlations and the variances explained indicate that there is no clear one-to-one correspondence between imageability ratings across languages. We discuss below some of the possible explanations for this remaining unexplained variance, including linguistic and cultural factors; intrinsic differences between the databases; range effects; the numbers of words, equivalent words, and participants; and the mean ages of participants.

Linguistic and cultural factors

The fact that words are semantically equivalent does not necessarily imply that these words also share similar ratings for other psycholinguistic variables such as frequency of usage, age of acquisition, word length, regularity of spelling, and so forth. Blomberg and Öberg (2015) found that the English word "sorrow" is higher in imageability but lower in frequency than its Swedish semantic equivalent "sorg"; the English word "anger" is less imageable than the Swedish semantic equivalent "ilska"; and the English word "position" is lower in age of acquisition than the Swedish semantic equivalent "position." Another factor that possibly could have affected the results—albeit to a lesser extent, given that we only compared ratings from European speakers—is that ratings for imageability (and other variables) may also depend on the cultural setting. Simonsen et al. (2013) pointed out that "Most Norwegian children have to swallow a spoonful of cod liver oil every day at least through the winter months. It is fair to assume that the Norwegian word for cod liver oil, *tran*, has a high imageability compared to languages spoken in countries where this is not the custom" (p. 436).

Matching words for a series of linguistic variables was not possible in the present study. This is because some of the

Table 2 Numbers of semantically equivalent words and variances explained (R^2) for the correlations in each language

	Basque	Catalan	Croatian	Eng_Bird	Eng_Cortese	Eng_MRC1	Greek	Cypriot_Greek	Hungarian	Italian	Norwegian	Serbian	Spanish	Swedish
Basque														
Catalan	35*;													
	55%													
Croatian	110*;	46*;												
	42%	56%												
Eng_ Bird	56*;	20;	100*;											
	38%	15%	49%											
Eng_ Cortese	152*;	64*;	162*;	467*;										
	48%	37%	67%	46%										
Eng_ MRC1	61*;	23*;	91*;	105*;	296*;									
	31%	77%	62%	47%	76%									
Greek	26*;	6;	23*;	12;	18;	17;								
	28%	NA	61%	NA	NA	NA								
Cypriot _ Greek	20*;	6;	25*;	23*;	22*;	13;	64*;							
	43%	NA	67%	30%	50%	NA	76%							
Hungarian	26;	11;	35*;	24;	73*;	28*;	10;	11;						
	8%	NA	45%	NA	9%	49%	NA	NA						
Italian	44*;	9;	46*;	111*;	139*;	33*;	6;	7;	20*;					
	16%	NA	37%	74%	45%	76%	NA	NA	69%					
Norwegian	145*;	44*;	145*;	133*;	251*;	135*;	27*;	31*;	66*;	70*;				
	24%	27%	35%	38%	42%	25%	38%	28%	66%	19%				
Serbian	28*;	5;	61*;	15;	31;	15;	7;	7;	4;	11;	34*;			
	20%	NA	64%	NA	25%	NA	NA	NA	NA	NA	40%			
Spanish	37*;	10;	29*;	45*;	102*;	27;	4;	4;	14;	13;	39*;	7;		
	58%	NA	52%	35%	32%	12%	NA	NA	NA	NA	35%	NA		
Swedish	31*;	18;	43*;	27*;	75*;	28*;	12;	12;	12;	19;	64*;	12;	9;	
	13%	NA	47%	53%	55%	48%	NA	NA	NA	NA	60%	NA	NA	
Turkish	66*;	28*;	83*;	31;	65*;	45*;	13;	15;	19;	30*;	84*;	19;	20*;	24*;
	28%	85%	47%	9%	38%	43%	NA	NA	NA	41%	9%	NA	67%	48%

Eng_Bird = data from Bird et al. (2001); Eng_Cortese = data from Cortese and Fugett (2004); Eng_MRCI1 = English data from Coltheart (1981). We report results only for correlations with ≥20 semantically equivalent words. Significant correlations are marked with an asterisk (*).

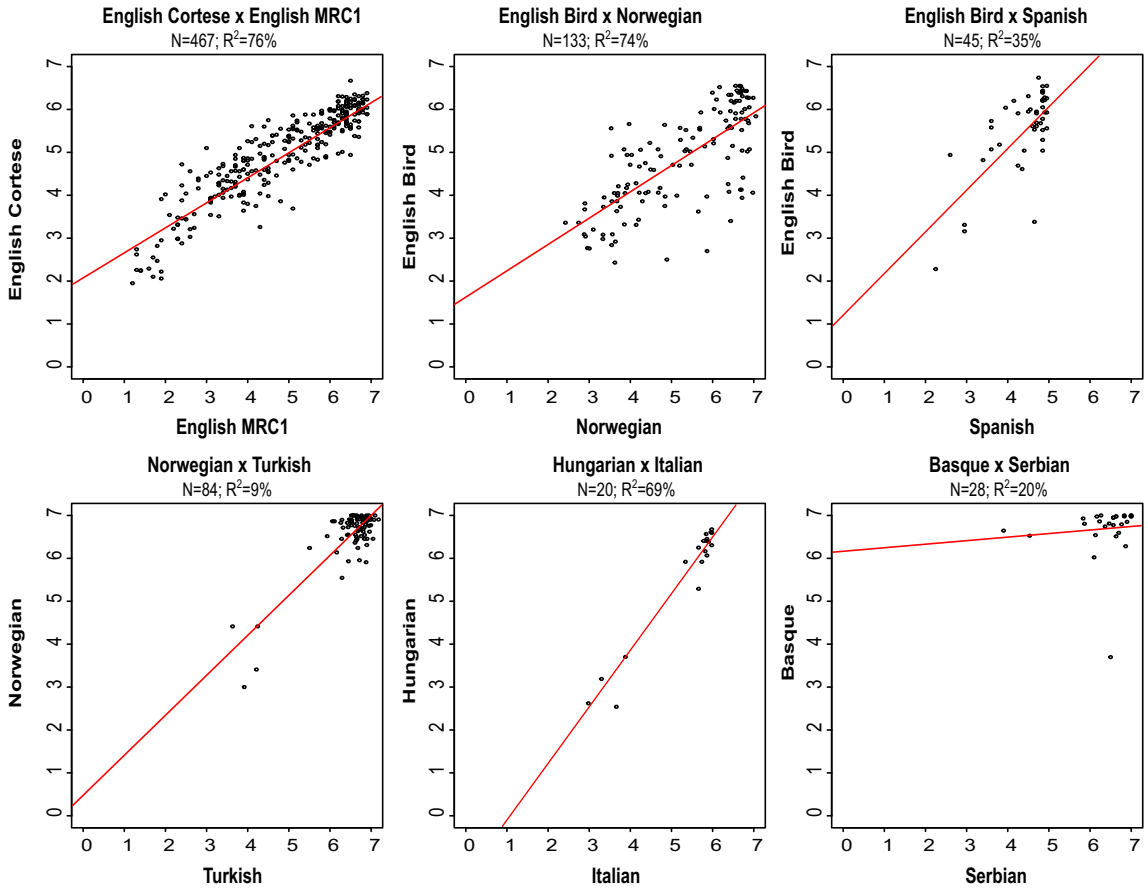

Fig. 1 Matrix scatterplot representing the variability across datasets. From top left to bottom right: English (Cortese & Fuggett, 2004) and English (MRC) as an example of a correlation within the same language, with a large number of semantically equivalent words and high variance explained; English (Bird et al., 2001) and Norwegian, as an example with a large number of semantically equivalent words, high variance explained, and data collected by asking participants to explicitly differentiate word categories (nouns vs. verbs); English (Bird et al., 2001) and Spanish, as an example with an average number of semantically equivalent words, moderate variance explained, and data collection in English by explicitly differentiating word category (nouns vs. verbs), and in Spanish according to the language differentiates word categories in the word form; Norwegian and Turkish, for an average number of equivalent words and low variance explained; Hungarian and Italian, for a low number of equivalent words and a relatively high variance explained; and Basque and Serbian, for a low number of equivalent words and low variance explained

languages did not have norms for all these linguistic variables and, when they did, some of the databases did not include ratings for all of the words. Matching all words for these psycholinguistic variables would have minimized the number of words entered in the study and, hence, also reducing the overall power of the statistical analyses. Having said that, we performed a secondary analysis on the age-of-acquisition ratings for languages that in this study contained ≥100 semantically equivalent words for imageability. The analyses comprised materials from Basque (Duñabeitia et al., 2017), two English databases (Bird et al., 2001; Cortese & Khanna, 2008; Schock et al., 2012), Norwegian (Lind et al., 2015; Simonsen et al., 2013), Italian (Rofes, de Aguiar, & Miceli, 2015), and Spanish (Alonso, Fernández, & Díez, 2015). The results indicated that, also for age of acquisition, the median strength across correlations was moderate (rho = .53), and the variance explained reached 28%. This implies that at least 72% of the variation in these datasets was due to factors other than age of acquisition. Further details can be found in the supplementary materials (Table S2).

Intrinsic differences between databases

Despite the fact that all imageability ratings in all of the databases were collected following the instructions by Paivio et al. (1968), many of the databases did not take into account rating differences that may have appeared due to the fact that participants might not have known the grammatical category of each word they were rating. For example, if a participant is presented with the English word "brief," she may not know whether it is a noun or a verb, unless it is read as "a brief" or "to brief"—in fact, "brief" could also be an adjective, as in "a brief history." The same holds for Norwegian, in which a participant having to rate the word "føde" may consider it either a verb, "give birth," or a noun, "food," unless the infinitive marker "å" is used, as in "å føde" (to give birth).

Disambiguating cases of homonymy is relevant because nouns and verbs have different imageability and other psycholinguistic values (e.g., Howard & Franklin, 1988; Whitworth et al., 2014). This aspect is particularly relevant for some languages, such as English and Norwegian. In fact, such variation was taken into account in the English database of Bird et al. (2001) and also in the Norwegian database (Lind et al., 2015; Simonsen et al., 2013). This specific aspect may not be as relevant for other languages in which the difference between nouns and verbs is marked morphologically and orthographically. Catalan and Spanish infinitives, for example, are marked with -(a/e/i)r, as in "cantar" (to sing). Also, Turkish infinitives are marked with -m(e/a)k, as in "bakmak" (to look) and "almak" (to take). A priori, having controlled for this factor makes the English database of Bird et al. and the Norwegian database different from the other databases. In this study, we noticed no special patterns regarding homonymy. English correlated with almost all databases except for Greek and Serbian, which are languages in which homonymy is not an issue, since very few noun and verb homonyms exist. Additionally, Greek and Serbian correlated with Basque and with Norwegian. Basque is also a language with very few noun and verb homonyms, and the Norwegian database was controlled for homonymy (Lind et al., 2015; Simonsen et al., 2013).

Range effects

When finding semantically equivalent words between languages, the numbers of words entered in the correlation may have clustered around specific parts of the distribution (e.g., Poulton, 1975). In our study, this implied that we could be correlating subsets in which the majority of words had been rated as high in imageability. This is a reasonable explanation for some of the nonsignificant correlations, because many of the databases were collected as part of another project that aimed at adapting a language battery that includes highly picturable items (e.g., for object-naming tasks) into multiple European languages (Fyndanis et al., 2017). If this was the only contributing factor, however, it would be hard to explain why a database such as that of Croatian, as opposed to other databases that were collected as part of that project, would significantly correlate with all of the databases. Indeed, to avoid range effects for those databases that were collected anew, we instructed each language team to include 20 to 100 items that were expected to produce low imageability ratings, based on the items in the database of Bird et al. (i.e., 2 to 3 points out of 7 in imageability). Also, those databases that had already been collected contained larger number of words and, therefore, included a wider range of imageability scores. Finally, range effects may not be accounted for by the fact that the Croatian and Spanish databases used a 5-point scale, as opposed to the rest of databases, which used a 7-point scale. This is because the two scales produce similar results; for example, they share the same mean score when rescaled (Dawes, 2012).

Numbers of words, equivalent words, and participants

The relatively small number of words that some of the databases contained (cf. 9,240 words in the MRC Psycholinguistic database vs. 202 words in the Catalan database) diminished the potential number of equivalent words between databases. We minimized the effects of this factor by only considering those correlations between languages in which we found at least 20 equivalent words in common. This resulted in the exclusion of 36 out of 105 correlations. Also, there was potentially unexplained variability in the imageability ratings for each word due to the varying numbers of participants in each survey. Again, in those databases that were collected anew, we tried to minimize this factor by including at least 20 individuals in each survey (e.g., Basque = 43 participants; Greek = 118 participants; Spanish = 20 participants; Swedish = 52 participants; Turkish = 51 participants).

Mean ages of participants

The mean age of participants was higher in the study by Bird et al. (2001) than in many of the other studies. If we take the mean age of participants as a factor, we see that the Catalan and Turkish databases were rated by people around 20–25 years of age (some undergraduate students, others not), whereas the Basque, Hungarian, Greek, and Swedish databases were rated by populations with a mean value of 40–45 years of age. The latter value could be thought of as closer to the 65 years of age in the database by Bird et al. (2001). This could explain some of the differences in significance testing—for example, the fact that the English database of Bird et al. did not correlate with the Catalan and Turkish databases. These results would be in line with an effect of age found in the Norwegian imageability study—from age 30 and upward, the imageability ratings increased systematically and significantly with participant age, with the largest difference found between 40 and 50 years (Simonsen et al., 2013). In the same vein, Bird et al. indicated that specific word ratings for some variables, such as age of acquisition, may have differed depending on the age of the participants.

Additionally, it could be argued that older individuals may have richer semantic representations, due to experience, since vocabulary scores increase with age (Diaz, Johnson, Burke, & Madden, 2014). If this holds, Catalan would have obtained lower imageability scores than Hungarian, Greek, and Swedish. However, on average, Hungarian, Greek, and Swedish were approximately 2 points lower in imageability than Catalan, despite the fact that older participants generally

provide higher imageability scores than younger participants. Additionally, the Turkish scores were very similar to those from Hungarian, Greek, and Swedish (range = 5.55–6.43; see the supplementary materials). This is also in contrast to older individuals providing higher imageability scores than younger individuals. Despite these results, it is worth noting that none of our databases included ratings from people of age 70 or older. Obtaining the ratings of people of age 70 or older may be relevant, since a decline in vocabulary scores has been reported in these individuals (Alwin & McCammon, 2001), and such a decline may be related to differences in imageability.

Future directions

A future study may consider a smaller number of languages and words matched for a series of variables (frequency, age of acquisition, linguistic typology, and cultural factors) using the same methods. The study could assess whether or not the strength and variance explained in the correlations is higher when these variables are considered, as opposed to not considered, for word selection. To the best possible extent, such a study might also avoid including words that are obviously strongly dependent on cultural factors. It could also be interesting to study how bilingual and multilingual speakers conceptualize the imageability of specific words, and also to look at speakers with different levels of literacy/education. Given our present results, we would expect these speakers to rate words with the same meaning inconsistently, regardless of the language, although some differences could emerge relative to literacy/education.

Conclusion

The high number of significant correlations between databases indicates that imageability ratings are, to a large extent, similar across languages. We have argued in favor of similarities in imageability between databases and discussed different reasons for the moderate strength between the correlations and the low variance explained. All these reasons possibly interact in our dataset. In sum, these are exciting results from a practical perspective, since they suggest that imageability ratings from one language may be used in another language. However, more accurate results may be obtained when collecting scores for each individual language.

Author note The Collaboration of Aphasia Trialists (CATs research network) is funded by the European Cooperation in Science and Technology (COST, Action IS1208). For more information, please visit www.aphasiatrials.org. This project was partially supported by the Global Brain Health Institute (A.R.). The Basque team (A.M.I. and M.P.) was partially supported by the Basque Government (Grant No. IT983-16-GIC 15/129) and MINECO/FEDER (FFI2015-68589-C2-1-P). The work by the authors from Norway (M.L. and H.G.S.) was partly supported by the Research Council of Norway through its Centres of Excellence funding scheme (223265). The Croatian study (J.K.K.) was supported by the Croatian Science Foundation and the project "Adult Language Processing" (ALP, Grant HRZZ-2421-UIP-11-2013). The Catalan study (A.G.) was supported by project FFI2014-56968-C4-1-P. The Serbian study (J.B. and J.V.) was supported by the Ministry of Education Science and Technological development grant (#IO175012). The Turkish study (I.M. and M.T.) was supported by Anadolu University, Scientific Research Project (BAP) Grant 1509S632. The Spanish study (S.M.-F.) was partly supported by PROGRAM (University of Copenhagen Excellence Programme for Interdisciplinary Research) and projects from the Ministerio de Economía y Competitividad (FFI2015-68589-C2-1-P and FFI2014-61888-EXP) The Greek group thanks Sophia Apostolopoulou and Michaela Nerantzini for their contribution to data collection.

References

Alario, F. X., Ferrand, L., Laganaro, M., New, B., Frauenfelder, U. H., & Segui, J. (2004). Predictors of picture naming speed. *Behavior Research Methods, Instruments, & Computers, 36,* 140–155. doi:10.3758/BF03195559

Allport, D. A., & Funnell, E. (1981). Components of the mental lexicon. *Philosophical Transactions of the Royal Society B, 295,* 397–410.

Alonso, M. A., Fernández, A., & Díez, E. (2015). Subjective age-of-acquisition norms for 7,039 Spanish words. *Behavior Research Methods, 47,* 268–274. doi:10.3758/s13428-014-0454-2

Alwin, D. F., & McCammon, R. J. (2001). Aging, cohorts, and verbal ability. *Journals of Gerontology, 56B,* S151–S161.

Bedny, M., & Thompson-Schill, S. L. (2006). Neuroanatomically separable effects of imageability and grammatical class during single-word comprehension. *Brain and Language, 98,* 127–139.

Berndt, R. S., Haendiges, A. N., Burton, M. W., & Mitchum, C. C. (2002). Grammatical class and imageability in aphasic word production: Their effects are independent. *Journal of Neurolinguistics, 15,* 353–371.

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. doi:10.3758/BF03195349

Bleasdale, F. A. (1987). Concreteness-dependent associative priming: Separate lexical organization for concrete and abstract words. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13,* 582–594. doi:10.1037/0278-7393.13.4.582

Blomberg, F., & Öberg, C. (2015). Swedish and English word ratings of imageability, familiarity and age of acquisition are highly correlated. *Nordic Journal of Linguistics, 38,* 351–364.

Bonett, D. G., & Wright, T. A. (2000). Sample size requirements for estimating Pearson, Kendall and Spearman correlations. *Psychometrika, 65,* 23–28. doi:10.1007/BF02294183

Breedin, S. D., Saffran, E. M., & Coslett, H. B. (1994). Reversal of the concreteness effect in a patient with semantic dementia. *Cognitive Neuropsychology, 11,* 617–660. doi:10.1080/02643299408251987

Carroll, J. B., & White, M. N. (1973). Word frequency and age of acquisition as determiners of picture-naming latency. *Quarterly Journal of Experimental Psychology, 25,* 85–95. doi:10.1080/14640747308400325

Chiarello, C., Shears, C., & Lund, K. (1999). Imageability and distributional typicality measures of nouns and verbs in contemporary English. *Behavior Research Methods, Instruments, & Computers, 31,* 603–637. doi:10.3758/BF03200739

Coltheart, M. (1981). The MRC psycholinguistic database. *Quarterly Journal of Experimental Psychology, 33,* 497–505. doi:10.1080/14640748108400805

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. doi:10.1111/j.2044-8295.1988.tb02270.x

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

Cortese, M. J., & Khanna, M. M. (2008). Age of acquisition ratings for 3,000 monosyllabic words. *Behavior Research Methods, 40,* 791–794. doi:10.3758/BRM.40.3.791

Cortese, M. J., Khanna, M. M., & Hacker, S. (2010). Recognition memory for 2,578 monosyllabic words. *Memory, 18,* 595–609.

Cortese, M. J., McCarty, D. P., & Schock, J. (2015). A mega recognition memory study of 2897 disyllabic words. *Quarterly Journal of Experimental Psychology, 68,* 1489–1501. doi:10.1080/17470218.2014.945096

Cortese, M. J., & Schock, J. (2013). Imageability and age of acquisition effects in disyllabic word recognition. *Quarterly Journal of Experimental Psychology, 66,* 946–972. doi:10.1080/17470218.2012.722660

Davelaar, E., & Besner, D. (1988). Word identification: Imageability, semantics, and the content-functor distinction. *Quarterly Journal of Experimental Psychology, 40,* 789–799.

Dawes, J. G. (2012). Do data characteristics change according to the number of scale points used? An experiment using 5 point, 7 point and 10 point scales. *International Journal of Market Research, 50,* 61–77.

de Aguiar, V., Bastiaanse, R., Capasso, R., Gandolfi, M., Smania, N., Rossi, G., & Miceli, G. (2015). Can tDCS enhance item-specific effects and generalization after linguistically motivated aphasia therapy for verbs? *Frontiers in Behavioral Neuroscience, 9,* 190.

Della Rosa, P. A., Catricalà, E., Vigliocco, G., & Cappa, S. F. (2010). Beyond the abstract–concrete dichotomy: Mode of acquisition, concreteness, imageability, familiarity, age of acquisition, context availability, and abstractness norms for a set of 417 Italian words. *Behavior Research Methods, 42,* 1042–1048. doi:10.3758/BRM.42.4.1042

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

Diaz, M. T., Johnson, M. A., Burke, D. M., & Madden, D. J. (2014). Age-related differences in the neural bases of phonological and semantic processes. *Journal of Cognitive Neuroscience, 26,* 2798–2811.

Duñabeitia, J. A., Casaponsa, A., Dimitropoulou, M., Martí, A., Larraza, S., & Carreiras, M. (2017). BaSp: A Basque–Spanish database of translation equivalents. Manuscript in preparation.

Franklin, S., Howard, D., & Patterson, K. (1995). Abstract word anomia. *Cognitive Neuropsychology, 12,* 549–566.

Fyndanis, V., Lind, M., Varlokosta, S., Kambanaros, M., Soroli, E., Ceder, K., . . . Howard, D. (2017). Cross-linguistic adaptations of The Comprehensive Aphasia Test: Challenges and solutions. *Clinical Linguistics and Phonetics*. Advance online publication. doi:10.1080/02699206.2017.1310299

Hanley, R. J., & Kay, J. (1997). An effect of imageability on the production of phonological errors in auditory repetition. *Cognitive Neuropsychology, 14,* 1065–1084.

Holmes, V. M., & Langford, J. (1976). Comprehension and recall of abstract and concrete sentences. *Journal of Verbal Learning and Verbal Behavior, 15,* 559–566. doi:10.1016/0022-5371(76)90050-5

Howard, D., & Franklin, S. (1988). *Missing the meaning? A cognitive neuropsychological study of the processing of words by an aphasic patient*. Cambridge: MIT Press.

Kambanaros, M., & Grohmann, K. K. (2015). Grammatical class effects across impaired child and adult populations. *Frontiers in Psychology, 6*(1670), 1–17. doi:10.3389/fpsyg.2015.01670

Kambanaros, M., Grohmann, K. K., & Michaelides, M. (2013). Lexical retrieval for nouns and verbs in typically developing bilectal children. *First Language, 33,* 182–199.

Kuvač Kraljević, J., & Olujić, M. (2017). *Croatian Lexical Database*. Manuscript submitted for publication.

Law, S. P., Kong, A. P. H., Lai, L. W. S., & Lai, C. (2015). Effects of context and word class on lexical retrieval in Chinese speakers with anomic aphasia. *Aphasiology, 29,* 81–100.

Levelt, W. (2014). *A history of psycholinguistics: The pre-Chomskyan era*. Oxford: Oxford University Press.

Lind, M., Simonsen, H. G., Hansen, P., Holm, E., & Mevik, B.-H. (2015). Norwegian words: A lexical database for clinicians and researchers. *Clinical Linguistics and Phonetics, 29,* 276–290.

Luzzatti, C., Raggi, R., Zonca, G., Pistarini, C., Contardi, A., & Pinna, G. D. (2002). Verb–noun double dissociation in aphasic lexical impairments: The role of word frequency and imageability. *Brain and Language, 81,* 432–444.

Ma, W., Golinkoff, R. M., Hirsh-Pasek, K., McDonough, C., & Tardif, T. (2009). Imageability predicts the age of acquisition of verbs in Chinese children. *Journal of Child Language, 36,* 405–423.

McMullen, P. A., & Bryden, M. P. (1987). The effects of word imageability and frequency on hemispheric asymmetry in lexical decisions. *Brain and Language, 31,* 11–25.

Morrison, C. M., & Ellis, A. W. (1995). Roles of word frequency and age of acquisition in word naming and lexical decision. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 116–133. doi:10.1037/0278-7393.21.1.116

Nickels, L., & Howard, D. (1994). A frequent occurrence? Factors affecting the production of semantic errors in aphasic naming. *Cognitive Neuropsychology, 11,* 289–320.

Nishimoto, T., Ueda, T., Miyawaki, K., Une, Y., & Takahashi, M. (2012). The role of imagery-related properties in picture naming: A newly standardized set of 360 pictures for Japanese. *Behavior Research Methods, 44,* 934–945. doi:10.3758/s13428-011-0176-7

Paivio, A. (2014). Intelligence, dual coding theory, and the brain. *Intelligence, 47,* 141–158.

Paivio, A., Yuille, J. C., & Madigan, S. A. (1968). Concreteness, imagery, and meaningfulness values for 925 nouns. *Journal of Experimental Psychology, 76*(1, Pt. 2), 1–25. doi:10.1037/h0025327

Plaut, D. C., & Shallice, T. (1993). Preservative and semantic influences on visual object naming errors in optic aphasia: A connectionist account. *Journal of Cognitive Neuroscience, 5,* 89–117.

Poulton, E. C. (1975). Range effects in experiments on people. *American Journal of Psychology, 88,* 3–32.

Proctor, R. W., & Vu, K.-P. L. (1999). Index of norms and ratings published in the Psychonomic Society journals. *Behavior Research Methods, Instruments, & Computers, 31,* 659–667. doi:10.3758/BF03200742

Rofes, A., Capasso, R., & Miceli, G. (2015). Verb production tasks in the measurement of communicative abilities in aphasia. *Journal of Clinical and Experimental Neuropsychology, 37,* 483–502. doi:10.1080/13803395.2015.1025709

Rofes, A., de Aguiar, V., & Miceli, G. (2015). A minimal standardization setting for language mapping tests: An Italian example. *Neurological Sciences, 36,* 1113–1119.

Sabsevitz, D. S., Medler, D. A., Seidenberg, M., & Binder, J. R. (2005). Modulation of the semantic system by word imageability. *NeuroImage, 27,* 188–200. doi:10.1016/j.neuroimage.2005.04.012

Schock, J., Cortese, M. J., & Khanna, M. M. (2012). Imageability estimates for 3,000 disyllabic words. *Behavior Research Methods, 44,* 374–379. doi:10.3758/s13428-011-0162-0

Schwanenflugel, P. J., Harnishfeger, K. K., & Stowe, R. W. (1988). Context availability and lexical decisions for abstract and concrete

words. *Journal of Memory and Language, 27,* 499–520. doi:10.1016/0749-596X(88)90022-8

Simonsen, H. G., Lind, M., Hansen, P., Holm, E., & Mevik, B.-H. (2013). Imageability of Norwegian nouns, verbs and adjectives in a cross-linguistic perspective. *Clinical Linguistics & Phonetics, 27,* 435–446. doi:10.3109/02699206.2012.752527

Stoke, S. M. (1929). Memory for onomatopes. *Pedagogical Seminary and Journal of Genetic Psychology, 36,* 594–596.

Strain, E., Patterson, K., & Seidenberg, M. S. (1995). Semantic effects in single-word naming. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 1140–1154. doi:10.1037/0278-7393.21.5.1140

Tyler, L. K., & Moss, H. E. (1997). Imageability and category-specificity. *Cognitive Neuropsychology, 14,* 293–318.

Tyler, L. K., Moss, H. E., Galpin, A., & Voice, J. K. (2002). Activating meaning in time: The role of imageability and form-class. *Language and Cognitive Processes, 17,* 471–502. doi:10.1080/01690960143000290

Warrington, E. K. (1981). Concrete word dyslexia. *British Journal of Psychology, 72,* 175–196.

Whitworth, A., Webster, J., & Howard, D. (2014). *A cognitive neuropsychological approach to assessment and intervention in aphasia: A clinician's guide*. Hove: Psychology Press.