

Concreteness ratings for 40 thousand generally known English word lemmas

Marc Brysbaert · Amy Beth Warriner ·
Victor Kuperman

Published online: 19 October 2013
© Psychonomic Society, Inc. 2013

Abstract Concreteness ratings are presented for 37,058 English words and 2,896 two-word expressions (such as *zebra crossing* and *zoom in*), obtained from over 4,000 participants by means of a norming study using Internet crowdsourcing for data collection. Although the instructions stressed that the assessment of word concreteness would be based on experiences involving all senses and motor responses, a comparison with the existing concreteness norms indicates that participants, as before, largely focused on visual and haptic experiences. The reported data set is a subset of a comprehensive list of English lemmas and contains all lemmas known by at least 85 % of the raters. It can be used in future research as a reference list of generally known English lemmas.

Keywords Concreteness · Ratings · Crowdsourcing · Word recognition

Concreteness evaluates the degree to which the concept denoted by a word refers to a perceptible entity. The variable came to the foreground in Paivio's dual-coding theory (Paivio, 1971, 2013). According to this theory, concrete words are easier to remember than abstract words, because they activate perceptual memory codes in addition to verbal codes. Schwanenflugel, Harnishfeger, and Stowe (1988) presented an alternative context availability theory, according to which

concrete words are easier to process because they are related to strongly supporting memory contexts, whereas abstract words are not, as can be demonstrated by asking people how easy it is to think of a context in which the word can be used.

The importance of concreteness for psycholinguistic and memory research is hard to overestimate. A search through the most recent literature gives the following, nonexhaustive list of topics related to concreteness. Are there hemispheric differences in the processing of concrete and abstract words (Oliveira, Perea, Ladera, & Gamito, 2013)? What are the effects of word concreteness in working memory (Mate, Allen, & Baqués, 2012; Nishiyama, 2013)? How are concrete and abstract concepts stored in and retrieved from long-term memory (Hanley, Hunt, Steed, & Jackman, 2013; Kousta, Vigliocco, Vinson, Andrews, & Del Campo, 2011; Paivio, 2013)? Does concreteness affect bilingual and monolingual word processing (Barber, Otten, Kousta, & Vigliocco, 2013; Connell & Lynott, 2012; Gianico-Relyea & Altarriba, 2012; Kaushanskaya & Rechtzigel, 2012)? Do concrete and abstract words differ in affective connotation (Ferre, Guasch, Moldovan, Sánchez-Casas, 2012; Kousta et al., 2011)? Do neuropsychological patients differ in the comprehension of concrete and abstract words (Loiselle et al., 2012)?

Concreteness gained extra interest within the embodied view of cognition (Barsalou, 1999; Fischer & Zwaan, 2008; Wilson, 2002)—in particular, when neuroscience established that words referring to easily perceptible entities coactivate the brain regions involved in the perception of those entities. Similar findings were reported for action-related words, which coactivate the motor cortex involved in executing the actions (Hauk, Johnsrude, & Pulvermüller, 2004). On the basis of these findings, Vigliocco, Vinson, Lewis, and Garrett (2004; see also Andrews, Vigliocco, & Vinson, 2009) presented a semantic theory, according to which the meaning of concepts depends on experiential and language-based connotations to different degrees. Some words are mainly learned on the basis

Electronic supplementary material The online version of this article (doi:10.3758/s13428-013-0403-5) contains supplementary material, which is available to authorized users.

M. Brysbaert (✉)
Department of Experimental Psychology, Ghent University, Henri
Dunantlaan 2, 9000 Gent, Belgium
e-mail: marc.brysbaert@ugent.be

A. B. Warriner · V. Kuperman
McMaster University, Hamilton, Canada

of direct experiences; others are mostly used in text and discourse. To make the theory testable, Della Rosa, Catricala, Vigliocco, and Cappa (2010) collected ratings of mode of acquisition, in which participants were asked to indicate to what extent the meaning of a word had been acquired through experience or through language. Unfortunately, to our knowledge these (Italian) norms have not yet been used to predict performance in word-processing tasks.

A final reason why concreteness has been a popular variable in psychological research is the availability of norms for a large number of words. Ratings were collected by Spreen and Schulz (1966), Paivio (both in Paivio, Yuille, & Madigan, 1968, and in unpublished data) and made available in the MRC database (Coltheart, 1981) for 4,292 words. The same database provides imageability ratings (closely related to the concreteness ratings) for 8,900 words. Throughout the years, authors have collected additional concreteness or imageability norms for specific subsets of words (e.g., Altarriba, Bauer, & Benvenuto, 1999; Schock, Cortese, & Khanna, 2012; Stadthagen-Gonzalez & Davis, 2006), which could be combined with the MRC ratings.

Impressive though the existing data sets are, developments in the past years have rendered them suboptimal. First, even 9,000 words is a limited number when viewed in the light of recently collected megastudies. For instance, the English Lexicon Project (Balota et al., 2007) contains processing times for more than 40,000 words, and the British Lexicon Project (Keuleers, Lacey, Rastle, & Brysbaert, 2012) has data for more than 28,000 monosyllabic and disyllabic words. This means that concreteness ratings are available only for limited subsets of available behavioral data sets.

A second limitation of the existing concreteness ratings is that they tend to focus too much on visual perception (Connell & Lynott, 2012; Lynott & Connell, 2009, *in press*) at the expense of the other senses and at the expense of action-related experiences. Lynott and Connell (2009) asked participant to what extent adjectives were experienced “by touch,” “by hearing,” “by seeing,” “by smelling,” and “by tasting” (five different questions). Connell and Lynott observed that these perceptual strength ratings were correlated only with concreteness ratings for vision, touch, and, to a lesser extent, smell. They were not correlated for taste and were even negatively correlated for auditory experiences. Similarly, none of the concreteness ratings collected so far includes the instruction that the actions one performs are experience based as well (and hence, concrete).

To remedy the existing limitations, we decided to collect new ratings for a large number of stimuli. This also allowed us to address another enduring problem in word recognition research—namely, the absence of a standard word list to refer to. Individual researchers use different word lists for rating studies and word recognition megastudies, mostly based on

existing word frequency lists. A problem with some of these lists is that they contain many entries that, depending on the purposes of one’s study, could qualify as noise. For instance, a study by Kloumann, Danforth, Harris, Bliss, and Dodds (2012) reported affective valence ratings for the 10,000 most frequent entries attested in four corpora. Their list included items that are unlikely to produce informative affective ratings, such as spelling variants (bday, b-day, and birthday), words with special characters (#music, #tcot), foreign words not borrowed into English (cf. the Dutch words “hij” [he] and “zijn” [to be]), alphanumeric strings (a3 and #p2), and names of people, cities, and countries. The list also included inflected word forms, which is a useful design option only if one expects inflected forms to differ in rating from lemmas (e.g., runs vs. run). When we compared Kloumann et al.’s list to a large list of English lemmas (see below), only half of the stimuli overlapped (see also Warriner, Kuperman, & Brysbaert, *in press*). This is a serious loss of investment, which is likely to further increase for less frequent entries (where the signal-to-noise ratio is even smaller).

To tackle the problem head on, we collected concreteness ratings for a list of 63,039 English lemmas one of us (M.B.) has been assembling over the years. This list does not contain proper names or inflected forms. The latter are more difficult to define in English than would be assumed at first sight, because many inflected verb forms are homonymous (and derivationally related) to uninflected adjectives (appalling) or nouns (playing). The simplest criterion to disambiguate such cases is to verify whether the word is used more often as an adjective/noun than as a verb form. This has become possible since we collected part-of-speech-dependent word frequency measures for American English (Brysbaert, New, & Keuleers, 2012). Similarly, some nouns are used more frequently in plural form than in singular form (e.g., *eyes*) or have different meanings in singular and plural (*glasses*, *aliens*). For these words, both forms were included in the list. Finally, the list for the first time also includes frequently encountered two-word spaced compound nouns (eye drops, insect repellent, lawn mower) and phrasal verbs (give away, give in, give up). The latter were based on unpublished analyses of the SUBTLEX-US corpus (Brysbaert & New, 2009). By presenting the full list, we were able to see which words are known to the majority of English speakers independently of word frequency. One way often used to select words for megastudies is to limit the words to those with frequencies larger than one occurrence per million words (e.g., Ferrand et al., 2010; Keuleers, Diependaele, & Brysbaert, 2010). This is a reasonable criterion but may exclude generally known words with low frequencies, which arguably are the most interesting to study the limitations of the existing word frequency measures.

In summary, we ran a new concreteness rating study (1) to obtain concreteness ratings for a much larger sample of

English words, (2) to obtain ratings based on all types of experiences, and (3) to define a reference list of English lemmas for future studies.

Method

Materials

The stimuli consisted of a list of 60,099 English words and 2,940 two-word expressions. The list was built on the basis of the SUBTLEX-US corpus (Brysbaert & New, 2009), supplemented with words from the English Lexicon Project (Balota et al., 2007), the British Lexicon Project (Keuleers et al., 2012; if necessary, spellings were Americanized), the corpus of contemporary American English (Davies, 2009), words used in various rating studies and shop catalogs, and words encountered throughout general reading. Although it is unavoidable that the list missed a few widely known words, care was taken to include as many entries as we could find.¹

Data collection

The stimuli were distributed over 210 lists of 300 words. Each list additionally included 10 calibrator words and 29 control words. The calibrator words represented the entire concreteness range (based on the MRC ratings) to introduce the participants to the variety of stimuli they could encounter. These words were placed in the beginning of each list. They were *shirt, infinity, gas, grasshopper, marriage, kick, polite, whistle, theory, and sugar*. Care was taken to include words referring to nonvisual senses and actions. The control words were from the entire concreteness range as well, used to detect noncompliance with the instructions (see below). Like the calibrator words, the same set of control words were used in all lists, to make sure that we used fixed criteria throughout. Control words were scattered randomly throughout the lists.

As in our previous studies (Kuperman, Stadthagen-Gonzalez, & Brysbaert, 2012; Warriner et al., *in press*) participants were recruited via Amazon Mechanical Turk's crowdsourcing Web site. Responders were restricted to those who self-identified as current residents of the U.S. The completion of a single list by a given participant is referred to as an assignment, given that participants were allowed to rate more than one list.

The following instructions were used:

Some words refer to things or actions in reality, which you can experience directly through one of the five senses. We call

these words concrete words. Other words refer to meanings that cannot be experienced directly but which we know because the meanings can be defined by other words. These are abstract words. Still other words fall in-between the two extremes, because we can experience them to some extent and in addition we rely on language to understand them. We want you to indicate how concrete the meaning of each word is for you by using a 5-point rating scale going from abstract to concrete.

A concrete word comes with a higher rating and refers to something that exists in reality; you can have immediate experience of it through your senses (smelling, tasting, touching, hearing, seeing) and the actions you do. The easiest way to explain a word is by pointing to it or by demonstrating it (e.g. To explain 'sweet' you could have someone eat sugar; To explain 'jump' you could simply jump up and down or show people a movie clip about someone jumping up and down; To explain 'couch', you could point to a couch or show a picture of a couch).

An abstract word comes with a lower rating and refers to something you cannot experience directly through your senses or actions. Its meaning depends on language. The easiest way to explain it is by using other words (e.g. There is no simple way to demonstrate 'justice'; but we can explain the meaning of the word by using other words that capture parts of its meaning).

Because we are collecting values for all the words in a dictionary (over 60 thousand in total), you will see that there are various types of words, even single letters. Always think of how concrete (experience based) the meaning of the word is to you. In all likelihood, you will encounter several words you do not know well enough to give a useful rating. This is informative to us too, as in our research we only want to use words known to people. We may also include one or two fake words which cannot be known by you. Please indicate when you don't know a word by using the letter N (or n).

So, we ask you to use a 5-point rating scale going from abstract to concrete and to use the letter N when you do not know the word well enough to give an answer.

Abstract (language based)			Concrete (experience based)	
1	2	3	4	5
N = I do not know this word well enough to give a rating.				

In the instructions, we stressed that we made a distinction between experience-based meaning acquisition and language-based meaning acquisition (cf. Della Rosa et al., 2010) and that experiences must not be limited to the visual modality. We used a 5-point rating scale based on Laming's (2004) observation that 5 is the maximum number of categories humans can distinguish consistently. When people are asked to make finer distinctions,

¹ Indeed, M.B. would appreciate receiving suggestions of missing words that should have been included.

they start using the labels inconsistently to such an extent that no extra information is obtained and the further scale precision is illusionary. In addition, we did not want to overtax the participants' working memory, because they had to keep in mind to use the N alternative in case they did not know the word well enough to give a valid rating.

On average, assignments were completed in approximately 14 min. Participants received 75 U.S. cents per completed assignment. After reading a consent form and the instructions, participants were asked to indicate their age, gender, first language(s), country/state resided in most between birth and age 7, and educational level. Subsequently, they were reminded of the scale anchors and presented with a scrollable page in which all words in the list were shown to the left of an answer box. Once finished, participants clicked 'Submit' to complete the study.

We aimed at 30 respondents per list. However, missing values due to subsequent exclusion criteria resulted in some words having less than 20 valid ratings. Several of the lists were reposted until the vast majority of words reached at least 25 observations per word. Data collection began January 25, 2013 and was completed by mid April 2013.

Results

Data trimming

Altogether 2,385,204 ratings were collected. Around 4 % of the data were removed due to missing responses, lack of variability in responses (i.e., providing the same rating for all words in the list), or the completion of fewer than 100 ratings per assignment. Further cleaning involved lists for which the correlation with the MRC ratings of the control words was between $-.5$ and $.2$. (The ones with correlations below $-.5$ were assumed to come from participants who misunderstood the instructions and used the opposite ordering; these scores were converted. This was the case for 149 assignments or 2.5 % of the total number.) Nonnative English speakers were also removed. Finally, assignments for which the correlation across the entire list was less than $.1$ with the average of the other raters were removed. Of the remaining data, 1,676,763 were numeric ratings, and 319,885 were "word not known" responses. These data came from 4,237 workers completing 6,076 assignments. There were more valid data from female participants (57 %) than from male participants. Of the participants, 1,542 (36 %) were of the typical student age (17–25 years old) and 40 (1 %) were older than 65 years. The remainder came from the ages in between these two groups. The distribution across educational levels is shown in Table 1.

Table 1 Distribution of the educational levels of the valid respondents

Some high school	46
High school graduate	355
Some college–no degree	1,438
Associates degree	442
Bachelors degree	1,389
Masters degree	422
Doctorate	112
Educational level not specified	33

Final list

Because ratings are only useful for well known words, we used a cutoff score of 85 % known. In practice, this meant that not more than 4 participants out of the average of 25 raters indicated that they did not know the word well enough to rate it. This left us with a list of 37,058 words and 2,896 two-word expressions (i.e., a total of 39,954 stimuli).

Validation

The simplest way to validate our concreteness ratings is to correlate them with the concreteness ratings provided in the MRC database (Coltheart, 1981). There were 3,935 overlapping words (the nonoverlapping words were mostly words not known to a substantial percentage of participants in our study, inflected forms, and words differing in spelling between British and American English). The correlation between both measures was $r=.919$, which is surprising given that our instructions emphasized—to a larger extent than the MRC instructions—the importance of action-related experiences. Also, when we look at the stimuli with the largest residuals between MRC and our ratings (Table 2), we see that they are more understandable as the outcome of different interpretations of ambiguous words than as differences between perception and action.

To further understand the essence of our ratings, we correlated them with the perceptual strength ratings collected by Lynott and Connell (2009, *in press*; downloaded on May 1, 2013 from <http://personalpages.manchester.ac.uk/staff/louise.connell/lab/norms.html>). As was indicated in the Introduction, these authors asked their participants to indicate how strongly they had experienced the stimuli with their auditory, gustatory, haptic, olfactory, and visual senses. Lynott and Connell also calculated the maximum perceptual strength of a stimulus, defined as the maximum value of the previous five ratings. Of the 1,001 words for which ratings were available, 615 had concreteness ratings in the MRC database and in our database. The correlation between the two concreteness ratings was very similar to that of the complete database ($r=.898$, $N=615$). Table 3 shows the

Table 2 Differences between the MRC ratings and the present ratings of concreteness: The 20 words with the largest negative and positive residuals

Words much lower in our ratings			Words much higher in our ratings		
WORD	MRC	OUR	WRD	MRC	OUR
concern	509	1.70	site	408	4.56
general	408	1.62	on	262	3.25
originator	491	2.52	stop	308	3.68
outsider	468	2.33	grate	432	4.82
patient	487	2.50	flow	311	3.72
chic	454	2.26	lighter	400	4.53
master	498	2.63	himself	285	3.50
conspirator	464	2.37	pour	356	4.14
dreamer	442	2.19	devil	274	3.41
gig	525	2.89	sear	292	3.59
ally	485	2.61	their	257	3.34
gloom	399	1.86	precipitate	350	4.19
mortal	406	1.96	facility	279	3.58
religion	375	1.71	dozen	396	4.66
equality	342	1.41	month	345	4.20
buffer	509	2.89	can	365	4.55
connoisseur	483	2.70	drop	320	4.21
evaluate	388	1.85	logos	299	4.41
forelock	565	3.28	tush	287	4.45
earl	500	2.85	concert	252	4.35

correlations with the perceptual strength ratings. Again, it is clear that our concreteness ratings provide very much the same information as the MRC concreteness ratings, despite the differences in instructions. In particular, both concreteness ratings correlate best with haptic and visual strength and show a negative correlation with auditory strength.

Discussion

Recent technological advances have made it possible to collect valid word ratings at a much faster pace than in the past. In

Table 3 Correlations between concreteness ratings and the perceptual strength ratings collected by Connell and Lynott (2009, *in press*)

	Concreteness_Our	Concreteness_MRC
Auditory strength	-.259**	-.234**
Gustatory strength	.023	.054
Haptic strength	.410**	.364**
Olfactory strength	.187**	.243**
Visual strength	.449**	.399**
Maximal strength	.495**	.440**

** $p < .001$

particular, the availability of Amazon Mechanical Turk (AMT) and the kindness of Internet surfers in providing good scientific data at an affordable price have made it possible to collect ratings for tens of thousands of words, rather than hundreds of words. In the present article, we discuss the collection of concreteness ratings for about 40,000 generally known English lemmas.

The high correlation between our ratings and those included in the MRC database ($r = .92$) attests to both the reliability and the validity of our ratings (for similar findings with AMT vs. lab-collected ratings, see also Kuperman et al., 2012; Warriner et al., *in press*). At the same time, the high correlation shows that the extra instructions we gave for the inclusion of nonvisual and action-related experiences did not seem to have much impact. Gustatory strength was not taken into account and auditory strength even correlated negatively, because words such as *deafening* and *noisy* got low concreteness ratings (1.41 and 1.69, respectively) but high auditory strength ratings (5.00 and 4.95). Apparently, raters cannot take into account several senses at the same time (Connell & Lynott, 2012).

The fact that our concreteness ratings are very similar to the existing norms (albeit for a much larger and more systematically collected stimulus sample) means that other criticisms recently raised against the ratings apply to our data set as well.² One concern, for instance, is that concreteness and abstractness may be not the two extremes of a quantitative continuum (reflecting the degree of sensory involvement, the degree to which words meanings are experience based, or the degree of contextual availability), but two qualitatively different characteristics. One argument for this view is that the distribution of concreteness ratings is bimodal, with separate peaks for concrete and abstract words, whereas ratings on a single, quantitative dimension usually are unimodal, with the majority of observations in the middle (Della Rosa et al., 2010; Ghio, Vaghi, & Tettamanti, 2013). As Fig. 1 shows, the bimodality of the distribution is true even for the large data set we collected, although it seems to be less extreme than reported by Della Rosa et al. Other arguments for qualitative differences between abstract and concrete concepts are that they can be affected differently by brain injury and that their representations may be organized in different ways (Crutch & Warrington, 2005; Duñabeitia, Avilés, Afonso, Scheepers, & Carreiras, 2009).

A further criticism raised against concreteness ratings is that concrete and abstract may not be basic level categories but superordinate categories (or maybe even ad hoc categories; Barsalou, 1983), which encompass psychologically more important subclasses, such as *fruits*, *vegetables*, *animals*, and *furniture* for concrete concepts and *mental-state-related*, *emotion-related*, and *mathematics-related notions* for

² We thank an anonymous reviewer for pointing us to this literature.

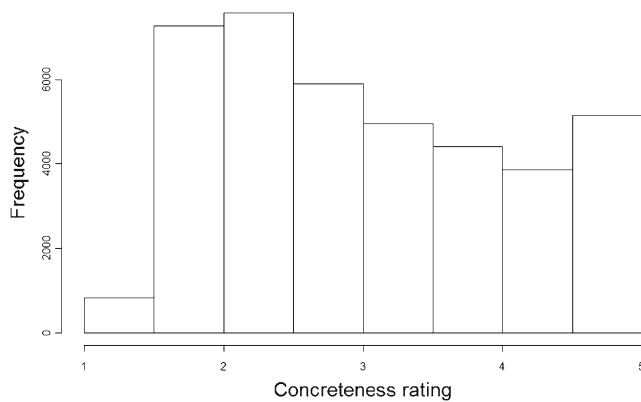


Fig. 1 Distribution of the concreteness ratings ($N=39,954$): 1=*very abstract* (language-based), 5=*very concrete* (experience-based)

abstract concepts (Ghio et al., 2013). If true, this criticism implies that not much information can be gained from concreteness information and that more fine-grained information is needed about the basic level categories (also Mahon & Caramazza, 2011).

The above criticisms perfectly illustrate that each study involves choices and, therefore, is limited in scope. What we won on the one hand (information about a variable for the entire set of interesting English lemmas) has been achieved at the expense of information richness on the other hand. This can be contrasted with the approach taken by Della Rosa et al. (2010), Ghio et al. (2013), Rubin (1980), and Clark and Paivio (2004), among others, who collected information about a multitude of word features, so that the correspondences between the measures could be determined. This, however, was achieved at the expense of the number of items for which information could be collected.

It is clear that our study cannot address all questions raised about concreteness norms. However, it provides researchers with values of an existing, much researched variable for an exhaustive word sample. More focused research is needed to further delineate the uses and limitations of the variable. For instance, it can be wondered how the low concreteness rating of *myth* (2.17) relates to the high perceptual strength rating of the same word (4.06, coming from auditory strength)³ and what the best value is for *atom*, given that the concreteness rating (3.34) is much higher than the perceptual strength rating (1.37). Similarly, it may be asked what the much lower concreteness rating of *loving* (1.73) than of *sailing* (4.17) means, given that many more participants are likely to have experienced the former than the latter (remember that we defined concrete as “experience-based” and abstract as “language-based”). These examples remind us that collecting a lot of information about a variable does not by itself make the variable more “real.” It only allows us to study the variable in more detail.

³ Arguably because people *hear* about myths.

Next to concreteness information, the research described in this article provides us with a reference list of English lemmas for future word recognition research. To achieve this, we presented a rather exhaustive list of lemmas to our participants, so that we made no a priori selection. On the basis of our findings, we can conclude that such a big list contains about one third of words not known to enough native speakers to warrant further inclusion in rating studies (to be fair to our participants, many of these stimuli referred to little known animals and plants). For future research, it seems more efficient to focus on the 40,000 generally known words than to continue including words that will have to be discarded afterward. At the same time, our research shows that some of the well-known words have low frequencies, as measured nowadays. These obviously include all two-word expressions (which are absent in most word frequency lists), but also compound words that were concatenated in our list because this is how they were used in the study we took them from (such as *birdbath* and *birdseed* from ELP) but that, in normal text, are usually written separately. Further well-known words with low frequencies are derivations of familiar words (such as *bloodlessness*, *borrowable*, and *brutalization*) and, more intriguingly, some words referring to familiar objects (such as *canola*, *lollypop*, *nectarine*, *nightshirt*, *thimble*, *wineglass*, and *bandanna*). By focusing on these stimuli, we can better understand the limitations of current-day word frequency measures. An interesting conceptual framework in this respect may be found in the papers of Vigliocco and colleagues (Andrews et al., 2009; Kousta et al., 2011; Vigliocco et al., 2004). Apparently, some words are well known to us because we daily experience the objects they refer to, but we rarely communicate about them, making them rather obscure in language corpora. Our database for the first time allows us to zoom in on these stimuli.

Availability

The data discussed in the present article are available in an Excel file, provided as [supplementary materials](#). The file contains eight columns:

1. The word
2. Whether it is a single word or a two-word expression
3. The mean concreteness rating
4. The standard deviation of the concreteness ratings
5. The number of persons indicating they did not know the word
6. The total number of persons who rated the word
7. Percentage participants who knew the word
8. The SUBTLEX-US frequency count (on a total of 51 million; Brysbaert & New, 2009)

References

- Altarriba, J., Bauer, L. M., & Benvenuto, C. (1999). Concreteness, context availability, and image ability ratings and word associations for abstract, concrete, and emotion words. *Behavior Research Methods*, 31(4), 578–602.
- Andrews, M., Vigliocco, G., & Vinson, D. (2009). Integrating experiential and distributional data to learn semantic representations. *Psychological Review*, 116, 463–498.
- Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. A., Kessler, B., Loftis, B., ... Treiman, R. (2007). The English Lexicon Project. *Behavior Research Methods*, 39, 445–459.
- Barber, H. A., Otten, L. J., Kousta, S. T., & Vigliocco, G. (2013). Concreteness in word processing: ERP and behavioral effects in a lexical decision task. *Brain and Language*, 125(1), 47–53.
- Barsalou, L. W. (1983). Ad hoc categories. *Memory & Cognition*, 11, 211–227.
- Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences*, 22, 577–660.
- Brysbaert, M., & New, B. (2009). Moving beyond Kucera 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(4), 977–90.
- Brysbaert, M., New, B., & Keuleers, E. (2012). Adding Part-of-Speech information to the SUBTLEX-US word frequencies. *Behavior Research Methods*, 44, 991–997.
- Clark, J. M., & Paivio, A. (2004). Extensions of the Paivio, Yuille, and Madigan (1968) norms. *Behavior Research Methods, Instruments & Computers*, 36(3), 371–383.
- Coltheart, M. (1981). The MRC psycholinguistic database. *The Quarterly Journal of Experimental Psychology*, 33, 497–505.
- Connell, L., & Lynott, D. (2012). Strength of perceptual experience predicts word processing performance better than concreteness or imageability. *Cognition*, 125, 452–465.
- Crutch, S. J., & Warrington, E. K. (2005). Abstract and concrete concepts have structurally different representational frameworks. *Brain*, 128(3), 615–627.
- Davies, M. (2009). The 385+ million word corpus of contemporary American English: Design, architecture, and linguistic insights. *International Journal of Corpus Linguistics*, 14(2), 159–190.
- 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.
- Duñabeitia, J. A., Avilés, A., Afonso, O., Scheepers, C., & Carreiras, M. (2009). Qualitative differences in the representation of abstract versus concrete words: evidence from the visual-world paradigm. *Cognition*, 110(2), 284–292.
- Ferré, P., Guasch, M., Moldovan, C., & Sánchez-Casas, R. (2012). Affective norms for 380 Spanish words belonging to three different semantic categories. *Behavior Research Methods*, 44, 395–403.
- 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(2), 488–496.
- Fischer, M. H., & Zwaan, R. A. (2008). Embodied language: A review of the role of the motor system in language comprehension. *The Quarterly Journal of Experimental Psychology*, 61, 825–850.
- Gianico-Relyea, J. L., & Altarriba, J. (2012). Word Concreteness as a Moderator of the Tip-of-the-Tongue Effect. *Psychological Record*, 62, 763–776.
- Ghio, M., Vaghi, M. M. S., & Tettamanti, M. (2013). Fine-Grained Semantic Categorization across the Abstract and Concrete Domains. *PLoS ONE*, 8(6), e67090. doi:10.1371/journal.pone.0067090
- Hanley, J. R., Hunt, R. P., Steed, D. A., & Jackman, S. (2013). Concreteness and word production. *Memory & Cognition*, 41, 365–377.
- Hauk, O., Johnsrude, I., & Pulvermüller, F. (2004). Somatotopic representation of action words in human motor and premotor cortex. *Neuron*, 41, 301–307.
- Kaushanskaya, M., & Rechtzigel, K. (2012). Concreteness effects in bilingual and monolingual word learning. *Psychonomic Bulletin & Review*, 19, 935–941.
- Keuleers, E., Diependaele, K., & Brysbaert, M. (2010). Practice effects in large-scale visual word recognition studies: A lexical decision study on 14,000 Dutch mono- and disyllabic words and nonwords. *Frontiers in Psychology*, 1, 174. doi:10.3389/fpsyg.2010.00174
- 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.
- Kloumann, I. M., Danforth, C. M., Harris, K. D., Bliss, C. A., & Dodds, P. S. (2012). Positivity of the English language. *PloS one*, 7(1), e29484. doi:10.1371/journal.pone.0029484
- Kousta, S. T., Vigliocco, G., Vinson, D. P., Andrews, M., & Del Campo, E. (2011). The representation of abstract words: Why emotion matters. *Journal of Experimental Psychology: General*, 140, 14–34.
- Kuperman, V., Stadthagen-Gonzalez, H., & Brysbaert, M. (2012). Age-of-acquisition ratings for 30 thousand English words. *Behavior Research Methods*, 44, 978–990.
- Laming, D. (2004). *Human judgement: The eye of the beholder*. London: Thompson Learning.
- Loiselle, M., Rouleau, I., Nguyen, D. K., Dubeau, F., Macoir, J., Whatmough, C., & Joubert, S. (2012). Comprehension of concrete and abstract words in patients with selective anterior temporal lobe resection and in patients with selective amygdalo-hippocampectomy. *Neuropsychologia*, 50, 630–639.
- Lynott, D., & Connell, L. (2009). Modality exclusivity norms for 423 object properties. *Behavior Research Methods*, 41, 558–564.
- Lynott, D., & Connell, L. (in press). Modality exclusivity norms for 400 nouns: The relationship between perceptual experience and surface word form. *Behavior Research Methods*.
- Mahon, B. Z., & Caramazza, A. (2011). What drives the organization of object knowledge in the brain? *Trends in Cognitive Sciences*, 15, 97–103.
- Mate, J., Allen, R. J., & Baqués, J. (2012). What you say matters: Exploring visual–verbal interactions in visual working memory. *The Quarterly Journal of Experimental Psychology*, 65, 395–400.
- Nishiyama, R. (2013). Dissociative contributions of semantic and lexical-phonological information to immediate recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 39, 642–648.
- Oliveira, J., Perea, M. V., Ladera, V., & Gamito, P. (2013). The roles of word concreteness and cognitive load on interhemispheric processes of recognition. *Laterality*, 18, 203–215.
- Paivio, A. (1971). *Imagery and verbal processes*. New York: Holt, Rinehart, and Winston.
- Paivio, A. (2013). Dual Coding Theory, Word Abstractness, and Emotion: A Critical Review of Kousta et al. (2011). *Journal of Experimental Psychology: General*, 142, 282–287.
- Paivio, A., Yuille, J. C., & Madigan, S. A. (1968). Concreteness, imagery, and meaningfulness values for 925 nouns. *Journal of experimental psychology*, 76, 1–25.
- Rubin, D. C. (1980). 51 properties of 125 words: A unit analysis of verbal behavior. *Journal of Verbal Learning and Verbal Behavior*, 19, 736–755.
- Schock, J., Cortese, M. J., & Khanna, M. M. (2012). Imageability estimates for 3,000 disyllabic words. *Behavior Research Methods*, 44(2), 374–9.

- 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.
- Spreen, O., & Schulz, R. W. (1966). Parameters of abstraction, meaningfulness, and pronunciability for 329 nouns. *Journal of Verbal Learning and Verbal Behavior*, 5, 459–468.
- Stadthagen-Gonzalez, H., & Davis, C. J. (2006). The Bristol norms for age of acquisition, imageability, and familiarity. *Behavior Research Methods*, 38(4), 598–605.
- Vigliocco, G., Vinson, D. P., Lewis, W., & Garrett, M. F. (2004). Representing the meanings of object and action words: The featural and unitary semantic space hypothesis. *Cognitive Psychology*, 48, 422–488.
- Warriner, A. B., Kuperman, V., & Brysbaert, M. (in press). Norms of valence, arousal, and dominance for 13,915 English lemmas. *Behavior Research Methods*.
- Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bulletin & Review*, 9, 625–636.