

Norms of age of acquisition and concreteness for 30,000 Dutch words



Marc Brysbaert^{a,*}, Michaël Stevens^a, Simon De Deyne^b, Wouter Voorspoels^b, Gert Storms^b

^a Ghent University, Belgium

^b University of Leuven, Belgium

ARTICLE INFO

Article history:

Received 11 March 2014

Received in revised form 22 April 2014

Accepted 23 April 2014

Available online 13 May 2014

PsycINFO:

2340

2720

Keywords:

Norms

Age of acquisition

Concreteness

Word recognition

ABSTRACT

Word processing studies increasingly make use of regression analyses based on large numbers of stimuli (the so-called megastudy approach) rather than experimental designs based on small factorial designs. This requires the availability of word features for many words. Following similar studies in English, we present and validate ratings of age of acquisition and concreteness for 30,000 Dutch words. These include nearly all lemmas language researchers are likely to be interested in. The ratings are freely available for research purposes.

© 2014 Elsevier B.V. All rights reserved.

1. Introduction

Research on word recognition is rapidly changing. Authors realise that the traditional small-scale factorial experiments are not the best approach because they lack power (Keuleers, Diependaele, & Brysbaert, 2010), do not give information about the full range of variables (Kuperman, Estes, Brysbaert, & Warriner, *in press*), and are open to experimenter bias in stimulus selection (Forster, 2000; Kuperman, *in press*). A better approach is to treat word recognition studies not as experiments in which word features can be manipulated but as correlational studies in which covariations between word features and word processing performance can be assessed (Baayen, Feldman, & Schreuder, 2006; Balota, Cortese, Sergent-Marshall, Spieler, & Yap, 2004; Lewis & Vladeanu, 2006). As a result, researchers have collected word processing times for thousands of words in so-called lexicon projects. Thus far, this happened in American English (Balota et al., 2007), Dutch (Keuleers et al., 2010), Malay (Yap, Rickard Liow, Jalil, & Faizal, 2010), French (Ferrand et al., 2010), British English (Keuleers, Lacey, Rastle, & Brysbaert, 2012), and Chinese (Sze, Rickard Liow, & Yap, *in press*).

At the same time, an optimal use of the lexicon projects requires information about the word features for (ideally) the entire database. This is easy for word variables that can be calculated on the

basis of the words themselves or corpus analyses, such as word length, various measures of word frequency, and similarity to other words but requires a major investment for variables that are based on subjective ratings.¹ These are variables like age of acquisition, concreteness, imageability, familiarity, valence, and arousal. They are investigated for their own sake or must be controlled for in order not to confound the effect of the variable of interest.

The situation is rapidly improving for the English language, where age-of-acquisition ratings have been collected for 30,000 words (Kuperman, Stadthagen-Gonzalez, & Brysbaert, 2012), affective ratings for 14,000 words (Warriner, Kuperman, & Brysbaert, 2013), and concreteness ratings for 40,000 words (Brysbaert, Warriner, & Kuperman, *in press*). The main reason for this improvement is that in English, one can make use of Amazon Mechanical Turk, a service created by the company Amazon where Internet users can earn a small amount of money by doing so-called Human Intelligence Tasks. These are usually short rating or translation tasks. Because there are several tens of thousands of Mechanical Turk workers, large-scale rating studies can be done in a matter of weeks at an affordable price. In addition, if some basic controls are included, the ratings are as reliable and valid as those collected under traditional laboratory circumstances (for evidence, see the references above).

¹ An extra complication is that it is difficult to secure funding for the collection of such ratings, because research councils seem to have an aversion for research proposals that are not driven by theory falsification, even though good hypothesis testing critically depends on access to this information.

* Corresponding author at: Department of Experimental Psychology, Ghent University, Henri Dunantlaan 2, B-9000 Gent, Belgium. Tel.: +32 9 264 94 25; fax: +32 9 264 64 96. E-mail address: marc.brysbaert@ugent.be (M. Brysbaert).

The situation is different for other languages because Amazon Mechanical Turk is based in the United States and has much fewer users/workers in languages other than English or Spanish (also the payment happens in dollars via the American branch of the company). This means that Amazon Mechanical Turk is a less interesting tool to collect data for languages such as Dutch. However, Moors et al. (2013) recently proposed an alternative solution. They showed that asking a limited group of participants to rate a list of 4,300 words returns the same outcome as the traditional approach of asking a large number of participants to rate 300 words each. The costs for paying the participants are the same, but the logistics become much more feasible. Also, participants are more interested and motivated when they can earn more money (because of the larger time investment).²

Arguably, the two most important word norms based on subjective ratings are age of acquisition (AoA) and concreteness. AoA refers to the age at which a word has been acquired and explains some 5% of variance in lexical decision times after the effects of word frequency, word length, and similarity to other words have been partialled out (Kuperman et al., 2012). This is even more when a suboptimal word frequency measure is used (Brysbaert & Cortese, 2011). The impact of AoA is due to the fact that the order of acquisition is an important variable in the organisation of the mental lexicon and the semantic system (Bai, Ma, Dunlap, & Chen, 2013; Catling, Dent, Preece, & Johnston, 2013; Cortese & Schock, 2013; Cuetos, Herrera, & Ellis, 2010; Palmer & Havelka, 2010) and to the fact that AoA is an important proxy for estimating the cumulative frequency with which people have come across words in their life (Lete & Bonin, 2013).

Concreteness evaluates the degree to which a concept denoted by a word refers to a perceptible entity. It is an important variable in memory research ever since Paivio formulated his 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. The variable gained extra interest within the embodied view of cognition (Barsalou, 1999; Fischer & Zwaan, 2008; Wilson, 2002), certainly after it was established that words referring to easily perceptible entities co-activate the brain regions involved in the perception of those entities, and that action-related words co-activate the motor cortex involved in executing the actions. 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 of direct experiences; others are mostly used in text and discourse.

Concreteness is also much researched in psycholinguistics. These are a few examples of recently examined topics related to concreteness. Are there hemispheric differences in the processing of concrete and abstract words (Oliveira, Perea, Ladera, & Gamito, 2013)? Does concreteness affect bilingual and monolingual word processing (Barber, Otten, Kousta, & Vigliocco, 2013; Connell & Lynott, 2012; Gianico-Relyea & Altarriba, 2012; Kaushanskaya & Rehtzige, 2012)? Do concrete and abstract words differ in affective connotation (Ferré, Guasch, Moldovan, & Sánchez-Casas, 2012; Kousta, Vigliocco, Vinson, Andrews, & Del Campo, 2011)? Do neuropsychological patients differ in the comprehension of concrete and abstract words (Loiselle et al., 2012)?

Imageability and familiarity are less interesting variables because imageability is highly correlated with concreteness and seems to stress the visual modality too much (Connell & Lynott, 2012). The importance

of familiarity is likely to be minimal, once one has a good word frequency measure and information about AoA (Brysbaert & Cortese, 2011). Valence and arousal have recently gained interest (e.g., Kuperman et al., in press) but could not be included in the present study (see, however, Moors et al., 2013, who collected values for 4,300 words).

AoA ratings were available for a few thousand words in Dutch. Ghyselinck, De Moor, and Brysbaert (2000) collected norms for some 3,000 short words. Ghyselinck, Custers, and Brysbaert (2003) collected ratings for a further 2,300 words from much used semantic categories (such as clothes, animals, utensils, birds, etc.). Finally, Moors et al. (2013) collected ratings for 4,300 words. To our knowledge, there are no large collections of concreteness ratings, but imageability norms were collected by Van Loon-Vervoorn (1985) for 6,100 words. The correlations between concreteness and imageability reported in the literature range from 0.78 to 0.85 (Friendly, Franklin, Hoffmann, & Rubin, 1982; Gilhooly & Logie, 1980; Paivio, Yuille, & Madigan, 1968).

Below, we describe the collection of concreteness and AoA ratings for 30,000 Dutch words.

2. Method

2.1. Stimulus materials

On the basis of dictionaries and corpus analyses, we selected a list of 30,000 'interesting' words. Interesting was defined in terms of the following:

1. Words are lemmas (unless the inflected form is highly frequent; e.g., 'eyes' in addition to 'eye').
2. No proper nouns are used.
3. The words are likely to be known to the participants.
4. No long, transparent compound words are included. Dutch is a language in which compounds do not have spaces, meaning that hundreds of thousands of words can be made by combining base words. Bertram and Hyona (2003) reviewed the reasons why low frequency, long, and transparent compounds are unlikely to be represented in the mental lexicon (they are parsed into their constituent meanings on the spot).

2.2. Participants

The participants were 74 students and scientific collaborators from Ghent University who completed the AoA lists and 75 students and collaborators from the University of Leuven who completed the concreteness lists. Of the Ghent participants, 11 were male and 63 female. Their mean age was 21.8 years (range, 18–32 years). Of the Leuven participants, 21 were males and 54 females. Their average age was 25.08 years (range, 17–63 years). Ghent and Leuven are two towns in Flanders (the Dutch speaking half of Belgium) separated by 80 km. Moors et al. (2013) found no differences in the ratings of the two universities for the variables they investigated, and there are no reasons to expect this would be otherwise for the present ratings. Still, as a precaution, the Ghent students provided the AoA norms, given that all previous AoA ratings in Dutch were collected there. The Leuven students provided the concreteness norms. More students started the study, but they are not included in the analyses because they did not return their list, arguably because they lost interest after a few trials.

2.3. Methods

For the concreteness ratings, the master stimulus list was divided into five lists of 6,000 words each. Each participant got a different permutation (15 raters per list). The lists started with the same 10 calibrator words covering the entire range of values from very concrete to very abstract, based on the authors' judgment and the imageability ratings of Van Loon-Vervoorn (1985). The concreteness instructions

² Language researchers seem to have a peculiar aversion to time intensive studies. The most often mentioned reasons are fatigue effects and lack of motivation (as if people are not used to working for a few hours at a task). Other objections are practice effects and long term priming (as if participants in psychology experiments should be uncertain about the task they are doing). As it happens, there is good evidence that you get better data if the participants have some experience with the task. Certainly for the lexicon projects it is becoming clear that time intensive studies with a limited group of participants are providing less noise than short studies with a large group (Keuleers et al., 2010, 2012).

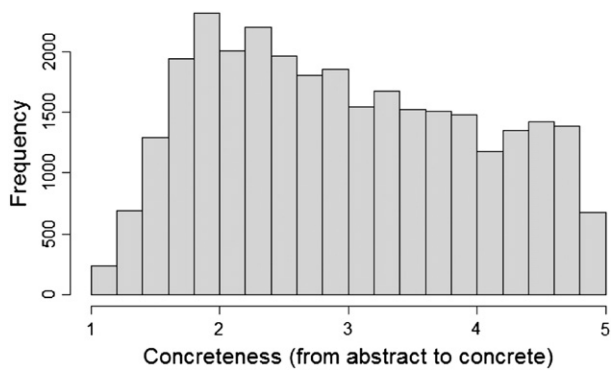


Fig. 1. Distribution of the concreteness ratings (1 = very abstract/language based, 5 = very concrete/experience based).

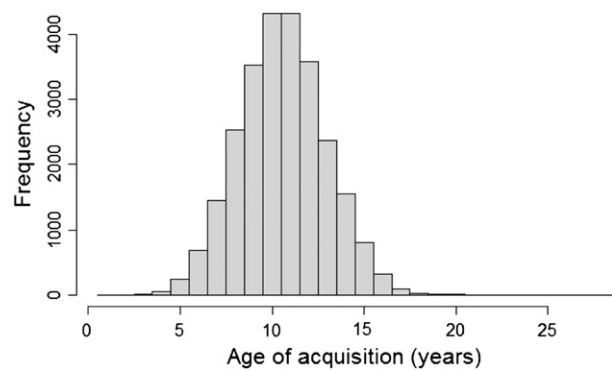


Fig. 2. Distribution of the AoA ratings collected in the present study (AoA is age in years at which the word was acquired).

were the same as used in Brysbaert et al. (in press). They can be consulted (in the original language) in Appendix A. The ratings went from 1 (very abstract/language based) to 5 (very concrete/experience based).

For AoA, list preparation was largely the same as for the concreteness ratings, except that the 4,300 stimuli of Moors et al. (2013) were not included and that the remaining words were distributed over 4 lists of 6,500 words each (18–20 raters per list). Instructions were the same as in Moors et al. (2013) (see Appendix A for the precise instructions in Dutch). Each list included the same 22 calibrator words and 62 control words covering the entire range of AoA. The calibrator words were presented at the beginning; the control words were interspersed in the list. They all came from the stimulus list of Moors et al. (2013). As for the concreteness ratings, calibrator words were included to expose participants to the full range of values early on. The control words were added to the AoA lists, to make it possible to compare the new ratings to those collected by Moors et al. (2013) in a different study, as recommended by Stadthagen-Gonzalez and Davis (2006). Participants were simply asked to enter the age (in years) at which they thought they learned the word.

Participants were given Excel files with the words in column A and asked to enter their rating in column B. They could freely decide when to complete the list over a period of 2 weeks. It took them on average 5–7 hours to finish a list, for which they received 50€. The participants were told in advance (and had to sign a sheet explaining this) that they would be paid only if their results correlated positively with those of the other participants. This was done to make sure that the students would take the task seriously and would not generate numbers at random. The students were also asked to use the letter N if they did not know the word well enough to give a rating.

3. Results and discussion

No students had to be excluded because of bad data. The intraclass correlation coefficient of reliability for the concreteness ratings was 0.92 (confidence interval 0.91–0.93; there were no noteworthy differences between the lists). To check the validity of the ratings, we correlated them with the imageability ratings collected by Van Loon-Vervoorn (1985). There was an overlap of 5,683 words between both lists. The correlation amounted to 0.76, very similar to the values reported for the English language (see the Introduction).³ Fig. 1 shows the distribution of the concreteness ratings.

³ The high correlation with imageability is in line with the observation we made in English that our concreteness ratings are the same as the existing concreteness ratings despite the fact that we stressed the fact that the meaning of abstract words is based on language rather than on experience (Brysbaert et al., in press). As noticed by a reviewer, in future it may be better to omit this part of the instructions. It does not change the participants' responses and the exclusion makes the instructions more theory-neutral.

The intraclass correlation coefficient of reliability for the AoA ratings was 0.93 (confidence interval: 0.91–0.94). To check for their congruent validity, we correlated them with the values collected by Ghyselinck et al. (2000, 2003) and Moors et al. (2013). There was an overlap of 989 words with Ghyselinck et al. (2000), which resulted in a correlation of $r = 0.88$. There was an overlap of 1,112 words with Ghyselinck et al. (2003) resulting in a correlation of $r = 0.91$. For the 84 calibrators and fillers overlapping with Moors et al. (2013), the correlation was 0.98 (remember that this was a subsample spread evenly across the entire range). Hence, as expected on the basis of previous research, the present AoA ratings are as good as the existing ones.

Fig. 2 shows the distribution of AoA values. As for the English ratings, the distribution is a normal distribution centred on the primary school age. This suggests that the ratings do not reflect the true age at which words have been acquired but rather the order in which the words have been learned (see Kuperman et al., 2012, for further discussion).

4. Availability

Two supplemental Excel files contain all the information discussed in the present study about AoA and concreteness ratings for 30,000 Dutch words. In addition, there is a third file summarising all information collected on AoA ratings in Dutch thus far. These are the AoA ratings from Ghyselinck et al. (2000, 2003), Moors et al. (2013), and the present study. While combining this information, we noticed that the ratings of Ghyselinck et al. (2000, 2003) not only correlated well with our ratings but also had the same means and standard deviations for the overlapping stimuli. This was not true for the Moors et al. (2013) ratings, which had a mean of nearly 1 year less. This was largely due to the fact that the participants from Rotterdam gave earlier estimates than those from Ghent. Hence, we limited the data to the 16 participants from Ghent in order to improve the comparability of the data sets. In line with the recommendations made by Stadthagen-Gonzalez and Davis (2006), we further added 0.4 to the Moors et al. (2013) mean ratings in order to bring the ratings to the same level as those of the other three studies (there was no need to adjust the slope). All files can be downloaded from the website of the journal or from Ghent University.

Acknowledgement

This research was made possible by the grant 'Wetenschappelijke Onderzoeksgemeenschap Taalverwerking' from the FWO Flanders.

Appendix A

A.1. Instructions for AoA ratings

Welkom bij deze beoordelingsstudie. Vul eerst en vooral de gegevens rechts van deze instructiepagina in.

Als je onderaan op het 'DE WOORDENLIJST' klikt, dan vind je de 6500 woorden terug die je dient te beoordelen.

We vragen je te beoordelen op welke leeftijd je de woorden hebt geleerd.

Hiermee bedoelen we de leeftijd waarop je het woord voor het eerst begreep toen iemand anders het in je bijzijn gebruikte, ook als je het woord zelf niet gebruikte.

Je vult hiervoor een geheel getal in.

Enkele voorbeelden:

Als je denkt dat je 'banaan' geleerd hebt toen je 3 jaar was, vul je 3 in.

Als je denkt dat je 'boekhouder' geleerd hebt toen je 11 jaar was, vul je 11 in.

Dit cijfer dien je in te vullen in de kolom naast het woord (kolom B). In kolom C verschijnt dan de betekenis van het cijfer dat jij gekozen hebt.

Als je een fout cijfer hebt ingevuld, verschijnt in kolom C een rood vlak met de boodschap 'FOUTE CODE'.

Probeer een zo juist mogelijke schatting te maken, maar denk ook niet te lang na bij elk woord.

Als je een woord niet kent, vul dan 'N' in.

A.2. Instructions for concreteness ratings

Sommige woorden verwijzen naar dingen of acties in de werkelijkheid. Je kunt ze direct ervaren via een van je vijf zintuigen of door de actie uit te voeren. We noemen deze woorden concrete woorden. Andere woorden verwijzen naar betekenissen die we niet rechtstreeks kunnen ervaren, maar die we kunnen kennen door andere woorden te gebruiken. Dit zijn abstracte woorden. Nog andere woorden vallen tussen beide extremen in, omdat we ze tot op zekere hoogte kennen via ervaring en we ook taal gebruiken om hun betekenis te vatten.

We vragen je om voor elk woord aan te geven hoe concreet de mening van het woord is. Hiervoor gebruik je een 5-puntenschaal gaande van abstract naar concreet. Een concreet woord verwijst naar iets wat bestaat in de realiteit. Je kunt het ervaren via een van je zintuigen (ruiken, proeven, voelen, horen, zien) of door een actie uit te voeren. De gemakkelijkste manier om het woord te definiëren is door het te tonen. Om het woord "zoet" uit te leggen, kun je iemand bijvoorbeeld suiker laten proeven. Om "springen" uit te leggen, kun je op en neer springen of een videoclip tonen van iemand die springt. Om de betekenis van "sofa" uit te leggen, kun je een sofa of een tekening van een sofa tonen.

Een abstract woord verwijst naar iets wat je niet rechtstreeks kunt ervaren. De betekenis ervan hangt af van taal. De gemakkelijkste manier om het woord uit te leggen is door andere woorden te gebruiken. Zo is er geen eenvoudige manier om het woord "wet" te tonen, maar je kunt het woord uitleggen door andere woorden te gebruiken.

Omdat we woorden verzamelen voor alle woorden uit een woordenboek, zul je verschillende soorten woorden krijgen, die je misschien niet meteen in dit soort studie verwacht. Probeer bij elk woord na te gaan in hoeverre de betekenis concreet (= ervaringsgebaseerd) is of abstract (= taalgebaseerd).

De kans is groot dat je een aantal woorden niet goed genoeg kent om een score te geven. Gebruik dan de letter N (of n) om aan te duiden dat je dit woord niet goed genoeg kent om te antwoorden. Ook dit is belangrijk voor ons, want het heeft geen zin om woorden te gebruiken in ons onderzoek, die niet door studenten gekend zijn.

Dus:

1 = abstract (vooral op taal gebaseerd)

2 = meer abstract dan concreet (taalcomponent is belangrijker dan ervaringscomponent)

3 = evenveel abstract als concreet

4 = meer concreet dan abstract (ervaringscomponent is belangrijker dan taalcomponent)

5 = concreet (vooral op ervaring gebaseerd)

N = dit woord ken ik niet goed genoeg om een rating te geven

(The ratings to be used remained visible at the top of the screen throughout the entire rating list).

Appendix B. Supplementary data

Supplementary data to this article can be found online at <http://dx.doi.org/10.1016/j.actpsy.2014.04.010>.

References

- Andrews, M., Vigliocco, G., & Vinson, D. (2009). Integrating experiential and distributional data to learn semantic representations. *Psychological Review*, 116, 463–498.
- Baayen, R. H., Feldman, L. B., & Schreuder, R. (2006). Morphological influences on the recognition of monosyllabic monomorphemic words. *Journal of Memory and Language*, 55, 290–313.
- Bai, L. L., Ma, T. F., Dunlap, S., & Chen, B. G. (2013). Age of acquisition affects the retrieval of grammatical category information. *Quarterly Journal of Experimental Psychology*, 66, 786–800.
- Balota, D. A., Cortese, M. J., Sergent-Marshall, S. D., Spieler, D. H., & Yap, M. J. (2004). Visual word recognition for single syllable words. *Journal of Experimental Psychology: General*, 133, 283–316.
- Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. A., Kessler, B., Loftis, B., et al. (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. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences*, 22, 577–660.
- Bertram, R., & Hyona, J. (2003). The length of a complex word modifies the role of morphological structure: Evidence from eye movements when reading short and long Finnish compounds. *Journal of Memory and Language*, 3, 615–634.
- 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.
- Brysbaert, M., Warriner, A. B., & Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods*, <http://dx.doi.org/10.3758/s13428-013-0403-5> (in press).
- Catling, J., Dent, K., Preece, E., & Johnston, R. (2013). Age-of-acquisition effects in novel picture naming: A laboratory analogue. *Quarterly Journal of Experimental Psychology*, 66, 1756–1763.
- Connell, L., & Lynott, D. (2012). Strength of perceptual experience predicts word processing performance better than concreteness or imageability. *Cognition*, 125, 452–465.
- Cortese, M. J., & Schock, J. (2013). Imageability and age of acquisition effects in disyllabic word recognition. *Quarterly Journal of Experimental Psychology*, 66, 946–972.
- Cuetos, F., Herrera, E., & Ellis, A. W. (2010). Impaired word recognition in Alzheimer's disease: The role of age of acquisition. *Neuropsychologia*, 48, 3329–3334.
- Ferrand, L., New, B., Brysbaert, M., Keuleers, E., Bonin, P., Méot, A., et al. (2010). The French Lexicon Project: Lexical decision data for 38,840 French words and 38,840 pseudowords. *Behavior Research Methods*, 42(2), 488–496.
- 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.
- 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.
- Forster, K. I. (2000). The potential for experimenter bias effects in word recognition experiments. *Memory & Cognition*, 28, 1109–1115.
- Friendly, M., Franklin, P. E., Hoffmann, D., & Rubin, D. C. (1982). The Toronto word pool: Norms for imagery, concreteness, orthographic variables, and grammatical usage for 1,080 words. *Behavior Research Methods & Instrumentation*, 14, 375–399.
- Ghyselinck, M., Custers, R., & Brysbaert, M. (2003). Age-of-acquisition ratings for 2332 Dutch words from 49 different semantic categories. *Psychologica Belgica*, 43, 181–214.
- Ghyselinck, M., De Moor, W., & Brysbaert, M. (2000). Age-of-acquisition ratings for 2,816 Dutch four- and five-letter nouns. *Psychologica Belgica*, 40, 77–98.
- Gianico-Relyea, J. L., & Altarriba, J. (2012). Word Concreteness as a Moderator of the Tip-of-the-Tongue Effect. *Psychological Record*, 62, 763–776.
- Gilhooly, K. J., & Logie, R. H. (1980). Age of acquisition, imagery, concreteness, familiarity, and ambiguity measures for 1944 words. *Behavior Research Methods & Instrumentation*, 12, 95–427.
- Kaushanskaya, M., & Rehtzige, 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, <http://dx.doi.org/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.
- 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. (2014). Virtual experiments in megastudies: A case study of language and emotion. *Quarterly Journal of Experimental Psychology* (in press).
- Kuperman, V., Estes, Z., Brysbaert, M., & Warriner, A.B. (2014). Emotion and language: Valence and arousal affect word recognition. *Journal of Experimental Psychology: General*, <http://dx.doi.org/10.1037/a0035669> (in press).
- Kuperman, V., Stadthagen-Gonzalez, H., & Brysbaert, M. (2012). Age-of-acquisition ratings for 30 thousand English words. *Behavior Research Methods*, *44*, 978–990.
- Lete, B., & Bonin, P. (2013). Does frequency trajectory influence word identification? A cross-task comparison. *Quarterly Journal of Experimental Psychology*, *66*, 973–1000.
- Lewis, M. B., & Vladeanu, M. C. (2006). What do we know about psycholinguistic effects? *Quarterly Journal of Experimental Psychology*, *59*, 977–986.
- Loiselle, M., Rouleau, I., Nguyen, D. K., Dubeau, F., Maccoir, J., Whatmough, C., et al. (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.
- Moors, A., De Houwer, J., Hermans, D., Wanmaker, S., van Schie, K., Van Harmelen, A., et al. (2013). Norms of valence, arousal, dominance, and age of acquisition for 4300 Dutch Words. *Behavior Research Methods*, *45*, 169–177.
- 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, Monograph Supplement*, *76*, 1.
- Palmer, S. D., & Havelka, J. (2010). Age of acquisition effects in vocabulary learning. *Acta Psychologica*, *135*, 310–315.
- Stadthagen-Gonzalez, H., & Davis, C. J. (2006). The Bristol norms for age of acquisition, imageability and familiarity. *Behavior Research Methods*, *68*, 598–605.
- Sze, W. P., Rickard Liow, S. J., & Yap, M. J. (2014). The Chinese Lexicon Project: A repository of lexical decision behavioral responses for 2,500 Chinese characters. *Behavior Research Methods*, *46*(1), 263–273.
- Van Loon-Vervoorn, W. A. (1985). *Voorstelbaarheidswaarden van Nederlandse woorden [Imageability values of Dutch words]*. Lisse: Swets & Zeitlinger.
- 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. (2013). Norms of valence, arousal, and dominance for 13,915 English lemmas. *Behavior Research Methods*, *45*, 1191–1207.
- Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bulletin & Review*, *9*, 625–636.
- Yap, M. J., Rickard Liow, S. J., Jalil, S. B., & Faizal, S. S. B. (2010). The Malay lexicon project: A database of lexical statistics for 9,592 words. *Behavior Research Methods*, *42*, 992–1003.