

The Calgary semantic decision project: concrete/abstract decision data for 10,000 English words

Penny M. Pexman¹ · Alison Heard¹ · Ellen Lloyd¹ · Melvin J. Yap²

© Psychonomic Society, Inc. 2016

Abstract Psycholinguistic research has been advanced by the development of word recognition megastudies. For instance, the English Lexicon Project (Balota et al., 2007) provides researchers with access to naming and lexical-decision latencies for over 40,000 words. In the present work, we extended the megastudy approach to a task that emphasizes semantic processing. Using a concrete/abstract semantic decision (i.e., does the word refer to something concrete or abstract?), we collected decision latencies and accuracy rates for 10,000 English words. The stimuli were concrete and abstract words selected from Brysbaert, Warriner, and Kuperman's (2013) comprehensive list of concreteness ratings. In total, 321 participants provided responses to 1,000 words each. Whereas semantic effects tend to be quite modest in naming and lexical decision studies, analyses of the concrete/abstract semantic decision responses show that a substantial proportion of variance can be explained by semantic variables. The item-level and trial-level data will be useful for other researchers interested in the semantic processing of concrete and abstract words.

Keywords Semantic decision task · Concreteness · Semantic processing · Word meaning

Electronic supplementary material The online version of this article (doi:10.3758/s13428-016-0720-6) contains supplementary material, which is available to authorized users.

✉ Penny M. Pexman
pexman@ucalgary.ca

¹ Department of Psychology, University of Calgary, 2500 University Drive NW, Calgary, Alberta T2N1N4, Canada

² National University of Singapore, Singapore, Singapore

In recent years, the publication of large datasets of behavioral responses to linguistic stimuli has been an important development for language researchers. The most influential of these has been the English Lexicon Project, which provides naming and lexical decision task (LDT) latencies for over 40,000 words (Balota et al., 2007). As evidence of its impact, consider that more than 1,000 citations now reference the article that describes the English Lexicon Project database. Additional LDT datasets have since been made available through the British Lexicon Project (Keuleers, Lacey, Rastle, & Brysbaert, 2012), as well as for other languages including French (Ferrand et al., 2010), Dutch (Keuleers, Diependaele, & Brysbaert, 2010), Malay (Yap, Liow, Jalil, & Faizal, 2010), and Chinese (Sze, Rickard Liow, & Yap, 2014). Since these datasets each involve responses to thousands of items, they allow researchers to evaluate effects of different (and often correlated) psycholinguistic variables and to do so with considerable statistical power (for discussions, see Balota, Yap, Hutchison, & Cortese, 2012; Brysbaert, Stevens, Mander, & Keuleers, 2016; Keuleers & Balota, 2015).

Using these datasets, researchers have learned a great deal about the lexical characteristics that influence LDT responses and have been able to test the effects of new variables as they emerge. For instance, the effects of word frequency (typically, faster responses to more frequent words) have been compared to those of contextual diversity (the number of unique passages/documents in which a word appears; Adelman, Brown, & Quesada, 2006), with results suggesting that contextual diversity is the better predictor. Similarly, the effects of orthographic neighborhood size (Coltheart, Davelaar, Jonasson, & Besner, 1977) have been compared to those of orthographic Levenshtein distance (a measure of words' orthographic similarity; Yarkoni, Balota, & Yap, 2008), with results showing that orthographic Levenshtein distance was the better predictor of LDT performance; responses were

faster for words that were orthographically less distinct. These and other lexical characteristics explain considerable variance in LDT performance.

In contrast, much less variance in LDT performance is explained by words' semantic characteristics (see Pexman, 2012, for a review). While seminal studies have shown that semantic information does play a role (e.g., Balota, Cortese, Sergent-Marshall, Spieler, & Yap, 2004; Buchanan, Westbury, & Burgess, 2001), it is assumed that LDT responses are primarily based on orthographic familiarity (Balota, Ferraro, & Connor, 1991). For those researchers interested in semantic processing, this constraint limits the utility of the LDT datasets. In one example, Yap, Tan, Pexman, and Hargreaves (2011) examined the influence of four measures of semantic richness on lexical decision latencies: number of features (e.g., *cheese* [high] vs. *basket* [low]; McRae, Cree, Seidenberg, & McNorgan, 2005), average radius of co-occurrence (e.g., *prison* [high] vs. *tweezers* [low]; Shaoul & Westbury, 2010), contextual dispersion (e.g., *whistle* [high] vs. *parsnip* [low]; Brysbaert & New, 2009), and number of senses (e.g., *book* [high] vs. *axe* [low]; Miller, 1990). After first controlling for a number of lexical characteristics such as frequency, orthographic neighborhood size, and orthographic Levenshtein distance, they found that these four semantic variables explained only 2 % of additional variance in lexical decision latencies. As Yap et al. (2011) showed, meaning influences in LDTs tend to be quite modest. Accordingly, researchers who are interested in questions of semantic representation and processing often use other tasks—in particular, those that require more extensive consideration of word meaning by participants. In the present study, we chose a concrete/abstract semantic decision (SDT) for this purpose. In this task, words are presented one at a time and participants are asked to decide whether each presented word refers to something concrete or abstract. The purpose of the present study was to generate a large SDT dataset to facilitate future research in ways that cannot be accomplished using existing LDT datasets.

To our knowledge, the only other SDT dataset currently available is from a recent study by Taikh, Hargreaves, Yap, and Pexman (2015). The authors collected behavioral responses for 288 pictures and, separately, their corresponding word labels. In that study, the decision was living/nonliving. Taikh et al. conducted regression analyses of behavioral responses to a subset of these items (i.e., those for which a full set of lexical and semantic predictor variables were available). The analysis of SDT responses to *word* stimuli is of particular relevance to the present study. Their analysis examined living/nonliving SDT latencies to 196 words, with lexical and task-specific variables entered on the first step and semantic richness variables on the second step. The results showed that the semantic richness variables explained 11 % of variance in living/nonliving SDT responses, over and above the 24 % explained by the lexical and task variables. Conversely, a

parallel analysis for the same items using English Lexicon Project LDT latencies as the outcome variable showed different results. When LDT (rather than SDT) latencies were regressed on the same predictor variables, lexical and task variables now explained 61 % of the variance, while the semantic richness variables explained 3 %. This provides some evidence that meaning variables can play a stronger role in an SDT than in the LDT, with the caveat that the item set used by Taikh et al. was quite limited. The number of items therein is relatively small, and is limited to concrete concepts. In addition, the results of that study included several effects that seemed particular to the living/nonliving decision. Taikh et al. speculated that the living/nonliving decision encouraged participants to focus on certain aspects of meaning, such as animacy, which may have contributed to the particular pattern of semantic effects that was observed.

Indeed, there is now strong evidence that the decision chosen in a semantic task can influence the effects observed. For instance, Hino, Pexman, and Lupker (2006) compared responses to words having many unrelated meanings with responses to unambiguous words. This type of semantic ambiguity effect was inhibitory in a living/nonliving SDT and also in a human/nonhuman SDT, but was null when the decision was vegetable/nonvegetable. Similarly, Pexman, Holyk, and Monfils (2003) examined number of features effects in three different semantic decisions, and found that number-of-features effects were large and facilitatory in a bird/nonbird SDT, larger with a living/nonliving SDT, and largest with a concrete/abstract SDT. Also evidence from fMRI data suggests differences in the brain regions that are associated with living/nonliving versus concrete/abstract semantic decision-making (Hargreaves, White, Pexman, Pittman, & Goodyear, 2012).

In perhaps the most fine-grained manipulation of decision category to date, Tousignant and Pexman (2012) examined the body–object interaction effect (faster responses to words that refer to objects with which the body can easily interact—e.g., *mask* [high] vs. *ship* [low]; Siakaluk, Pexman, Aguilera, Owen, & Sears, 2008; Tillotson, Siakaluk, & Pexman, 2008) in four different versions of an SDT. All four versions of the SDT presented the same word lists and varied only in how the decision was framed to participants: as action/nonaction, action/entity, entity/action, or entity/nonentity. The body–object interaction effect was null when the decision was action/nonaction, large and facilitatory when the decision was action/entity or entity/action, and largest when the decision was entity/nonentity. The authors took these findings as evidence that semantic processing is highly context-dependent, and that participants make adjustments to semantic processing in response to the task context to optimize performance. Similarly, Jared and Seidenberg (1991) found differences in the semantic effects observed in narrow (e.g., flower/nonflower) versus broad (e.g., living/nonliving) decisions, and recommended that

researchers avoid specific categories in decision tasks. In the present study, we chose the broadest decision that we could (concrete/abstract) so that a large number of items could be presented under the same task demands. In the sections of this article that follow, we describe our data collection procedure and offer some preliminary description and analyses of the dataset. This includes comparisons of the present results to those of previous smaller-scale SDT studies, because there can sometimes be differences in the results of small-scale and megastudies (Sibley, Kello, & Seidenberg, 2009). By making this dataset available to other researchers, we hope to facilitate future studies on the semantic processing of concrete and abstract words.

Method

Participants

The participants were 321 undergraduate students at the University of Calgary who participated for partial course credit. Nine of these participants had SDT accuracy below 70 %, and so their data were removed from the final dataset and from all analyses; the data for the remaining 312 participants (225 female, 87 male) were analyzed. All subsequent descriptions correspond to this final set. Participants were asked to provide their age, and 296 did so (age range = 17–66 years; mean age = 21.75 years, $SD = 5.82$). Prior to taking part in the study, all participants completed a prescreen questionnaire on which they reported their level of English fluency. Only participants who self-reported as being “completely fluent” were eligible to participate. All participants had normal or corrected-to-normal vision. Participants were randomly assigned to one of ten versions of the study, comprising unique word lists (further description follows). The numbers and characteristics of the participants who completed the different versions of the study are presented in Table 1.

Apparatus

Words were presented via a widescreen 24-in. ASUS monitor (VG248QE), which was controlled by a Dell OptiPlex 9020 PC. The monitor has a rapid refresh rate of 144 Hz and a 1-ms response time.

Stimuli

The word stimuli were selected from Brysbaert et al.’s (2013) comprehensive list of concreteness ratings for English lemmas. This list contains the concreteness ratings of 40,000 known English lemmas, rated on a scale of 1 (*abstract*) to 5 (*concrete*). From this list we selected 18,000 words consisting of nouns, verbs, adjectives, and adverbs. These included 9,000

of the words rated as most concrete and 9,000 of the words rated as most abstract. The concreteness ratings ranged from 3.78 to 5 for concrete words, and 1.04 to 2.08 for abstract words. Slang or obscenities, one-letter words, and words with spaces or dashes were eliminated. Next, we selected 10,000 items that could be divided into ten lists of 1,000 words and matched (using the Match program; van Casteren & Davis, 2007), such that in each list the abstract and concrete words did not differ significantly on word length or frequency (measured as the log of the SUBTLEXus frequency values; Brysbaert & New, 2009). Each resulting list of 1,000 words was assigned to a different version of the experiment. The mean lexical characteristics for each of the ten versions are presented in Table 2. Thus, across versions, participants collectively gave responses to 5,000 concrete words and 5,000 abstract words. All of these words are also present in the ELP database.

Procedure

PC-compatible computers running E-Prime software (Schneider, Eschman, & Zuccolotto, 2001) were used for stimulus presentation and data collection. Participants were individually tested in our university laboratory. Before beginning the SDT, each participant first completed the Modified Edinburgh Handedness Survey. We used the results of this survey to ensure that all participants responded to concrete words during the SDT using their dominant hand and abstract words using their nondominant hand. For participants whose score on the handedness survey was zero (fully ambidextrous, $n = 2$), the hand they preferred to write with was designated as the dominant hand. Next, each participant was administered the shortened version of the North American Adult Reading Test (NAART35; Uttl, 2002) in order to assess vocabulary skill.

To align with the definitions used in the Brysbaert et al. (2013) study, participants were provided with the following onscreen instructions for the SDT:

Concrete words are defined as things or actions in reality, which you can experience directly through your senses. These words are experience-based.

Night, bridle, and lynx are examples of concrete words. Abstract words are defined as something you cannot experience directly through your senses or actions. These words are language-based, as their meaning depends on other words.

Have, limitation, and outspokenness are examples of abstract words.

The researcher verbally repeated these instructions and confirmed that the participant understood the distinction between the concrete and abstract word categories. The

Table 1 Participant characteristics and mean Calgary semantic decision task (SDT) response latencies and accuracies, by version

| Variable | Version Number | | | | | | | | | | All Versions |
|--------------------------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | |
| Total <i>n</i> | 31 | 31 | 31 | 32 | 32 | 31 | 31 | 31 | 30 | 32 | 312 |
| Left-handed <i>n</i> | 3 | 1 | 1 | 4 | 3 | 1 | 1 | 0 | 1 | 1 | 16 |
| Mean age (<i>SD</i>) | 21.38 (7.03) | 21.59 (4.63) | 20.41 (2.63) | 23.55 (11.59) | 20.47 (1.96) | 20.31 (2.44) | 22.47 (3.43) | 21.90 (5.73) | 22.90 (5.37) | 22.40 (6.06) | 21.75 (5.82) |
| Female <i>n</i> | 21 | 21 | 20 | 22 | 24 | 25 | 19 | 24 | 22 | 27 | 225 |
| Mean NAART (<i>SD</i>) | 14.87 (5.38) | 15.71 (5.91) | 15.00 (5.59) | 15.50 (7.98) | 15.66 (5.44) | 13.74 (6.09) | 14.58 (7.31) | 17.06 (5.83) | 16.30 (5.55) | 14.03 (6.34) | 15.24 (6.19) |
| Mean SDT latency (<i>SD</i>) | 1,006.31 (187.39) | 1,003.73 (230.85) | 1,020.82 (200.44) | 982.13 (139.61) | 1,004.91 (159.45) | 1,036.24 (212.27) | 995.21 (170.61) | 1,004.58 (185.15) | 995.12 (160.26) | 1,022.26 (234.34) | 1,007.13 (188.24) |
| Mean SDT accuracy (<i>SD</i>) | 88.56 (3.82) | 86.10 (5.14) | 86.27 (4.04) | 86.07 (5.05) | 87.64 (4.21) | 84.67 (5.23) | 86.68 (4.39) | 86.87 (4.04) | 87.20 (5.67) | 83.33 (6.09) | 86.33 (4.96) |
| Mean concrete latency (<i>SD</i>) | 964.78 (178.54) | 968.02 (226.41) | 964.50 (181.75) | 951.89 (133.92) | 970.62 (142.18) | 1,001.00 (210.12) | 952.35 (166.79) | 966.32 (166.17) | 953.66 (144.69) | 994.60 (233.47) | 968.86 (179.60) |
| Mean concrete accuracy (<i>SD</i>) | 88.39 (4.93) | 86.12 (6.80) | 87.76 (4.44) | 86.38 (5.62) | 87.07 (6.62) | 86.20 (5.24) | 86.42 (6.78) | 86.49 (4.96) | 87.46 (6.58) | 84.02 (8.41) | 86.62 (6.17) |
| Mean abstract latency (<i>SD</i>) | 1,048.62 (200.97) | 1,041.02 (238.50) | 1,081.58 (227.92) | 1,014.51 (151.75) | 1,040.25 (182.29) | 1,073.84 (217.65) | 1,039.55 (179.81) | 1,043.36 (207.99) | 1,038.04 (182.56) | 1,052.86 (242.88) | 1,047.28 (202.92) |
| Mean abstract accuracy (<i>SD</i>) | 88.74 (4.08) | 86.08 (5.31) | 84.78 (6.12) | 85.77 (7.08) | 88.22 (3.26) | 83.14 (7.63) | 86.92 (4.75) | 87.26 (4.76) | 86.94 (5.99) | 82.64 (8.91) | 86.04 (6.24) |

NAART = North American Adult Reading Test

researcher reminded each participant that words in the study were not restricted to nouns and that even verbs and adjectives could fall into the concrete or abstract categories.

Participants were next provided with 24 practice items before beginning the experimental trials. Stimuli were presented one at a time in the center of the screen in white lowercase letters against a black background (Courier New, font size 18). Each trial began with the presentation of a fixation screen depicting two horizontal lines positioned above and below a gap where a word would appear. Participants were asked to focus on the gap between the lines. After 500 ms, the stimulus word was presented in the gap; the horizontal lines remained on the screen. Individual stimuli remained on the screen until

the participant made a response or for a maximum of 3,000 ms. Using an external response box connected to the serial port, participants responded using their dominant hand for concrete words and their nondominant hand for abstract words. The interstimulus interval was 500 ms. A feedback screen was presented for 1,000 ms following any incorrect responses (“incorrect”) or when no response was detected (“no response detected”).

Following completion of the practice trials, the researcher invited each participant to ask any additional questions. For example, the researcher might explain the correct categorization of a given word if a participant indicated they were unsure why a response had been incorrect. The researcher then left

Table 2 Mean lengths, frequencies and concreteness ratings for concrete and abstract words by version (standard deviations in parentheses)

| Version | Length | | | Frequency | | | Concreteness | | |
|--------------|-------------|-------------|----------------|-------------|-------------|----------------|--------------|-------------|----------------|
| | Concrete | Abstract | <i>t</i> (998) | Concrete | Abstract | <i>t</i> (998) | Concrete | Abstract | <i>t</i> (998) |
| 1 | 8.24 (2.26) | 8.25 (2.27) | 0.08 | 1.59 (0.79) | 1.59 (0.79) | 0.00 | 4.27 (0.43) | 2.02 (0.28) | 98.49*** |
| 2 | 8.19 (2.24) | 8.22 (2.26) | 0.18 | 1.59 (0.77) | 1.59 (0.77) | 0.04 | 4.31 (0.42) | 2.04 (0.28) | 100.13*** |
| 3 | 8.03 (2.10) | 8.05 (2.10) | 0.13 | 1.59 (0.76) | 1.59 (0.77) | 0.02 | 4.27 (0.43) | 2.02 (0.29) | 96.48*** |
| 4 | 7.93 (2.04) | 7.94 (2.05) | 0.12 | 1.64 (0.75) | 1.64 (0.75) | 0.00 | 4.27 (0.45) | 2.03 (0.28) | 95.05*** |
| 5 | 7.89 (2.07) | 7.90 (2.08) | 0.13 | 1.59 (0.76) | 1.59 (0.76) | 0.03 | 4.30 (0.41) | 2.04 (0.29) | 100.45*** |
| 6 | 8.08 (1.97) | 8.10 (1.97) | 0.14 | 1.59 (0.72) | 1.59 (0.72) | 0.01 | 4.28 (0.43) | 2.02 (0.27) | 98.27*** |
| 7 | 7.93 (1.87) | 7.95 (1.88) | 0.13 | 1.61 (0.74) | 1.61 (0.74) | 0.02 | 4.29 (0.41) | 2.03 (0.29) | 101.01*** |
| 8 | 7.85 (1.59) | 7.88 (1.61) | 0.29 | 1.59 (0.73) | 1.59 (0.74) | 0.08 | 4.30 (0.42) | 2.03 (0.29) | 100.51*** |
| 9 | 7.68 (1.34) | 7.75 (1.36) | 0.82 | 1.60 (0.81) | 1.57 (0.84) | 0.47 | 4.28 (0.44) | 2.02 (0.28) | 95.87*** |
| 10 | 8.21 (1.95) | 8.38 (1.99) | 1.34 | 1.65 (0.79) | 1.64 (0.77) | 0.23 | 4.28 (0.43) | 2.02 (0.29) | 97.75*** |
| All versions | 8.00 (1.97) | 8.04 (1.98) | 0.99 | 1.61 (0.77) | 1.60 (0.77) | 0.31 | 4.28 (0.43) | 2.03 (0.28) | 311.13*** |

*** $p < .001$, ** $p < .01$, * $p < .05$

the participant alone in the testing room to complete the procedure independently. Throughout the experimental trials, each participant made semantic decisions for 500 concrete and 500 abstract words. Each word list was divided into four blocks consisting of 250 words. Trials were randomized within blocks, and block order was fixed. Breaks were provided between each block. The first and third breaks did not have a set time limit; participants were simply told to press a button when they were ready to continue. To manage participant fatigue, the duration of the second break was a mandatory 3 min. After the mandatory break, a warning screen (black lettering on a white background) appeared for 1,000 ms, signaling to the participant that the trials were about to resume. On average, participants took 80 min to complete the entire procedure.

Full item-level and trial-level datasets are available as supplements to this article, and descriptions of the variables in each file are provided in the [Appendix](#).

Results

Trials with incorrect responses (12.49 %) were excluded from the latency analyses. Responses faster than 250 ms (0.02 %) were likewise excluded before computing the latency means and standard deviations for each participant in each block. Note that responses slower than 3,000 ms were also automatically excluded, because all trials timed out after 3,000 ms (0.49 %). Next, latencies beyond 3 *SDs* from each participant's mean in each block were eliminated, removing a further 1.37 % of the responses.

The correlation between participant age and North American Adult Reading Test score was significant, $r(296) = .30$, $p < .001$, such that older participants tended to have higher vocabulary scores. We used partial correlations to investigate the relationship between vocabulary score and SDT performance, independent of age (Table 3). Even with age controlled, participant vocabulary scores were still correlated with the speed and accuracy of SDT responses, such that participants with higher vocabulary scores had faster and more accurate SDT responses for both concrete and abstract words.

Response latencies were standardized as *z* scores, since these minimize the influence of a participant's overall processing speed and variability (Faust, Balota, Spieler, & Ferraro, 1999). Using these scores, we ran a series of hierarchical linear regression analyses to compare the semantic effects on Calgary SDT latencies to those found in previous studies. In the first two sets of regression analyses reported next, we made direct comparisons to previous (smaller-scale) studies, so we used the same predictors as in the original studies. In the final set of regression analyses reported below, we used the semantic predictors that allowed us to include as many items

Table 3 Partial correlations between NAART scores and Calgary SDT response latency and accuracy, with participant age controlled

| Variable | 1 | 2 | 3 | 4 | 5 | 6 |
|----------------------|---------|---------|---------|---------|--------|--------|
| 1. NAART score | – | | | | | |
| 2. SDT latency | –.29*** | – | | | | |
| 3. SDT accuracy | .61*** | –.26*** | – | | | |
| 4. Concrete latency | –.31*** | .99*** | –.30*** | – | | |
| 5. Concrete accuracy | .49*** | –.22*** | .80*** | –.30*** | – | |
| 6. Abstract latency | –.26*** | .99*** | –.22*** | .94*** | –.14* | – |
| 7. Abstract accuracy | .49*** | –.20** | .79*** | –.18** | .26*** | –.22** |

NAART = North American Adult Reading Test; SDT = semantic decision task. *** $p < .001$, ** $p < .01$, * $p < .05$

as possible in the present dataset, to examine the contributions of lexical and semantic variables across the dataset.

Pexman, Hargreaves, Siakaluk, Bodner, and Pope (2008) examined SDT latencies based on the concrete/abstract decision for 514 concrete words from McRae et al.'s (2005) feature-listing norms. Using hierarchical regression, they entered log word frequency (HAL; Lund & Burgess, 1996), orthographic neighborhood size, and word length as control variables in Step 1, and three semantic richness variables in Step 2: number of semantic neighbors (Durda, & Buchanan, 2006), contextual dispersion (Zeno, Ivens, Millard, & Duvvuri, 1995), and number of features. Using the same variables, we ran the same analysis on latencies to the 192 items that are common to the Pexman et al. study and the present SDT. The results are presented in Table 4. The patterns of results for the two data sets are the same; frequency is the only significant control variable, and both contextual dispersion and number of features are significant semantic predictors, while number of semantic neighbors is not.

Similarly, SDT latencies to 202 abstract words were examined in an earlier study by Zdrazilova and Pexman (2013). The task used by Zdrazilova and Pexman was a go/no-go SDT; participants decided whether each item referred to something abstract, pressing a button to respond “yes” to abstract words and withholding a response for concrete words. Using hierarchical regression, they entered log word frequency (SUBTL; Brysbaert & New, 2009), orthographic Levenshtein distance, and age of acquisition ratings in Step 1, and six semantic richness variables in Step 2: context availability, sensory experience ratings (Juhasz, Yap, Dicke, Taylor, & Gullick, 2011), valence, arousal, number of semantic neighbors, and number of associates (Nelson, McEvoy, & Schreiber, 1998). Using the same variables, we ran the same analysis on latencies to the 125 items that are common to the Zdrazilova and Pexman study and the present SDT. The results are presented in Table 5. Despite the fact that the Zdrazilova and Pexman SDT used a go/no-go procedure and the present SDT did not, the patterns of results for the two data sets are the same;

Table 4 Hierarchical regression results for 192 concrete words from Pexman et al. (2008), and the corresponding Calgary SDT results for the same items

| Variable | Pexman et al. (2008) SDT | | | | | Calgary SDT | | | | |
|-------------------|--------------------------|------------|---------|--------|--------------|-------------|------------|---------|--------|--------------|
| | <i>B</i> | <i>SEB</i> | β | R^2 | ΔR^2 | <i>B</i> | <i>SEB</i> | β | R^2 | ΔR^2 |
| Step 1 | | | | .06* | .06* | | | | .16*** | .16*** |
| Log HAL frequency | −9.65 | 3.31 | −.24** | | | −.07 | .02 | −.36*** | | |
| ON | −0.14 | 2.07 | −.01 | | | −.00 | .01 | −.02 | | |
| Word length | −0.23 | 3.14 | −.01 | | | .01 | .01 | .06 | | |
| Step 2 | | | | .18*** | .12*** | | | | .23*** | .07** |
| NSN | −0.12 | 0.56 | −.02 | | | −.00 | .00 | −.05 | | |
| CD | −95.06 | 26.01 | −.28*** | | | −.38 | .12 | −.23** | | |
| NF | −4.27 | 1.30 | −.23** | | | −.01 | .01 | −.15* | | |

SDT = semantic decision task; HAL = hyperspace analogue to language; ON = orthographic neighborhood size; NSN = number of semantic neighbors; CD = contextual dispersion; NF = number of features. *** $p < .001$, ** $p < .01$, * $p < .05$

frequency was the only significant lexical variable, and sensory experience rating was the only significant semantic predictor. The values in Table 5 also suggest that more variance was explained overall in the Zdrzilova and Pexman dataset. This may be due to the go/no-go procedure, which Zdrzilova and Pexman speculated would have encouraged participants to focus on factors that are diagnostic of abstractness, rather than simply on the absence of concreteness.

Using a similar analytic approach, we next assessed the variance explained by lexical and semantic predictors in the present SDT for a much larger set of items, and compared to English Lexicon Project LDT latencies for the same items. We expected that lexical variables might explain more variance in LDT responses, while semantic variables might explain more variance in SDT responses. We analyzed the responses to concrete and abstract words separately, since these represent different responses in the SDT. Using hierarchical regression, we entered log contextual dispersion (Brysbaert & New, 2009), orthographic Levenshtein distance, orthographic neighborhood size, and word length as lexical variables in Step 1, and in Step 2 we entered three semantic richness variables for which we had a large number of values, to make the analysis inclusive of responses to most of the items: concreteness, average radius of co-occurrence (Shaoul & Westbury, 2010), and semantic diversity (Hoffman, Lambon Ralph, & Rogers, 2013). As is illustrated in Table 6, the lexical variables did tend to explain more of the variance in LDT than in SDT latencies. Furthermore, the semantic richness variables tended to explain more variance in the SDT than in the LDT, particularly for concrete words. We discuss the observed patterns of effects for semantic richness variables in more detail below.

Finally, we checked for practice effects across experiment blocks. Given the length of the experimental sessions, it is not surprising that participants did tend to get faster across blocks. A one-way analysis of variance on the response latencies revealed a main effect of block [$F(3, 933) = 58.94, p < .001, \eta_p^2$

$= .159$]. Participants tended to speed up from the first block ($M = 1,048.33, SD = 393.74$) to the second ($M = 1,002.10, SD = 382.97$) [$t(311) = 8.35, p < .001$], but had similar latencies for the second and third blocks ($M = 1,000.00, SD = 383.23$) [$t(311) = 0.54, p = .59$]. Participants then got faster from the third to the fourth blocks ($M = 970.28, SD = 366.70$) [$t(311) = 6.61, p < .001$]. In addition, we evaluated the possibility that participants were relying on different types of information to make their semantic decisions in the first block and the last block. We did this by running the regression analyses presented in Table 6 separately for Block 1 data and Block 4 data. The results showed, for both blocks, the same patterns of effects as in the overall analysis. As such, we assume that participants did not shift their reliance on different types of lexical or semantic information across the experimental blocks. Users who wish to control for variability in latencies across blocks can do so with the Block variable in the item-level data file, or in a more fine-grained way using the FullRunOrder variable in the trial-level data file.

Discussion

The overarching purpose of the present study was to generate a relatively large dataset of SDT responses. We chose a decision category that was sufficiently broad to allow inclusion of a large number of items but that still required meaning retrieval for each item presented. We capitalized on the existing concreteness norms generated by Brysbaert et al. (2013) to select concrete and abstract word stimuli. As we described in the introduction, the decision chosen for a SDT will necessarily shape the responses generated. Participants tend to focus on dimensions of meaning that are diagnostic of the decision (Tousignant & Pexman, 2012). Certainly, the breadth of the concrete/abstract decision would likely make it less susceptible to these effects than a more narrow decision might be (e.g.,

Table 5 Hierarchical regression results for 125 abstract words from Zdrzilova and Pexman (2013), and the corresponding Calgary SDT results for the same items

| Variable | Zdrzilova and Pexman (2013) Go/No-Go SDT | | | | | Calgary SDT | | | | |
|---------------------|--|------------|---------|--------|--------------|-------------|------------|---------|-------|--------------|
| | <i>B</i> | <i>SEB</i> | β | R^2 | ΔR^2 | <i>B</i> | <i>SEB</i> | β | R^2 | ΔR^2 |
| Step 1 | | | | .15*** | .15*** | | | | .08* | .08* |
| Log SUBTL frequency | −50.61 | 13.18 | −.47*** | | | −.14 | .06 | −.29* | | |
| OLD | −5.44 | 8.11 | −.06 | | | −.05 | .04 | −.11 | | |
| AoA | −9.15 | 9.59 | −.11 | | | .01 | .05 | .02 | | |
| Step 2 | | | | .35*** | .20*** | | | | .19** | .11* |
| CA | −12.59 | 17.61 | −.09 | | | .05 | .09 | .09 | | |
| SER | −62.30 | 13.83 | −.48*** | | | −.25 | .07 | −.41** | | |
| Valence | 3.88 | 6.38 | .05 | | | .02 | .03 | .06 | | |
| Arousal | −29.14 | 16.86 | −.16 | | | −.05 | .09 | −.05 | | |
| NSN | 0.20 | 0.15 | .11 | | | .00 | .00 | .00 | | |
| NoA | 0.35 | 1.03 | .03 | | | .00 | .01 | .01 | | |

SDT = semantic decision task; SUBTL = subtitle; OLD = orthographic Levenshtein distance; AoA = age of acquisition; CA = context availability; SER = sensory experience rating; NSN = number of semantic neighbors; NoA = number of associates. *** $p < .001$, ** $p < .01$, * $p < .05$

vegetable/nonvegetable, bird/nonbird), but the decision will, nonetheless, have influenced responses. As evidence, consider the large proportion of variance explained by the concreteness dimension in the regression analyses in Table 6; concreteness was facilitatory for concrete words and inhibitory for abstract words. Researchers wishing to control for these effects of typicality should include the concreteness dimension in analyses of this dataset. Since we chose our stimuli from Brysbaert et al.'s (2013) comprehensive concreteness ratings norms, those values are available for every item in the present dataset and we have included them in the item-wise dataset in order to make it relatively straightforward for users to perform this type of adjustment for typicality.

Previous lexical-semantic studies have tended to focus on concrete words, and have identified a number of dimensions that are important to concrete meaning, including sensorimotor dimensions like imageability and BOI (e.g., Amsel, 2011; Amsel & Cree, 2013; Amsel, Urbach, & Kutas, 2012; Cortese & Fugett, 2004; Siakaluk, et al., 2008; Yap, Pexman, Wellsby, Hargreaves, & Huff, 2012; Yap et al., 2011). Some of these dimensions are likely not relevant for abstract words; for example, body-object interaction by definition applies only to words that refer to objects or entities. Less research attention has been given to dimensions of abstract meaning, however, so there is much we do not yet understand about the semantic representation of abstract concepts. Indeed, the regression results in Table 6 show that the variables tested explained more variance for concrete words than for abstract words in the SDT. With responses to 5,000 abstract items, the present dataset offers the opportunity to examine new questions about abstract word meaning. Our results, preliminary though they are, provide some intriguing evidence that semantic effects may differ for concrete and abstract words.

In particular, the patterns of semantic diversity effects for concrete and abstract words show some interesting differences. While we found that this semantic variable was facilitatory in the LDT for both concrete and abstract words, in the SDT it was facilitatory for abstract words but inhibitory for concrete words (Table 6). Hoffman et al. (2013) devised the construct of semantic diversity and assumed that words that appear in more diverse contexts have more varied meanings. In a previous study, Hoffman and Woollams (2015) showed that semantic diversity effects vary across tasks; in the LDT, responses were faster to high-semantic-diversity words than to low-semantic-diversity words, but in a semantic relatedness judgment task the effect was reversed, with faster responses to low-semantic-diversity words than to high-semantic-diversity words. This pattern is consistent with some of the previous literature on semantic ambiguity effects, but that literature is quite mixed (e.g., Hargreaves, Pexman, Pittman, & Goodyear, 2011; Piercey & Joordens, 2000; Rodd, Gaskell, & Marslen-Wilson, 2002; Yap et al., 2011). Our results point to one potential explanation for some of the mixed results—concreteness. That is, our results suggest different effects of ambiguity in the SDT for concrete and abstract words. Using the present dataset, the reasons for these differences could be explored in future studies.

Similarly, the effects of the average radius of co-occurrence differed in this study for concrete and abstract words. The average radius of co-occurrence indexes a word's closeness or similarity to its neighbors in lexical co-occurrence space (Shaoul & Westbury, 2010). Previous studies have reported facilitatory effects of the average radius of co-occurrence in the LDT, but null effects in the SDT for both concrete (Yap et al., 2012; Yap et al., 2011) and abstract (Zdrzilova & Pexman, 2013) stimuli. In the present analyses, controlling

Table 6 Hierarchical regression results for Calgary SDT and English Lexicon Project LDT

| Variable | Calgary SDT | | | | | ELP LDT | | | | |
|--------------------------------|-------------|------------|---------|--------|--------------|----------|------------|---------|--------|--------------|
| | <i>B</i> | <i>SEB</i> | β | R^2 | ΔR^2 | <i>B</i> | <i>SEB</i> | β | R^2 | ΔR^2 |
| Concrete Words ($n = 2,856$) | | | | | | | | | | |
| Step 1 | | | | .18*** | .18*** | | | | .32*** | .32*** |
| Log SUBTL CD | -.24 | .01 | -.34*** | | | -.17 | .01 | -.34*** | | |
| OLD | -.22 | .02 | -.45*** | | | .08 | .01 | .21*** | | |
| ON | .00 | .00 | .02 | | | -.00 | .00 | -.01 | | |
| Length | .11 | .01 | .48*** | | | .03 | .01 | .16*** | | |
| Step 2 | | | | .42*** | .24*** | | | | .33*** | .01*** |
| Concreteness | -.55 | .02 | -.51*** | | | -.07 | .01 | -.09*** | | |
| ARC | .05 | .07 | .01 | | | -.03 | .05 | -.01 | | |
| SemD | .05 | .03 | .03* | | | -.11 | .02 | -.10*** | | |
| Abstract Words ($n = 3,747$) | | | | | | | | | | |
| Step 1 | | | | .11*** | .11*** | | | | .40*** | .40*** |
| Log SUBTL CD | -.17 | .01 | -.33*** | | | -.21 | .01 | -.39*** | | |
| OLD | .02 | .01 | .05 | | | .07 | .01 | .16*** | | |
| ON | .02 | .00 | .15*** | | | .00 | .00 | .01 | | |
| Length | .00 | .01 | .01 | | | .04 | .00 | .24*** | | |
| Step 2 | | | | .20*** | .09*** | | | | .41*** | .01*** |
| Concreteness | .28 | .02 | .23*** | | | -.05 | .02 | -.04** | | |
| ARC | .20 | .05 | .08*** | | | -.14 | .04 | -.05** | | |
| SemD | -.21 | .02 | -.17*** | | | -.12 | .02 | -.10*** | | |

SDT = semantic decision task; ELP = English Lexicon Project; SUBTL = subtitle; OLD = orthographic Levenshtein distance; ON = orthographic neighborhood size; ARC = average radiance of co-occurrence; SemD = semantic diversity. *** $p < .001$, ** $p < .01$, * $p < .05$

for several other lexical and semantic variables, the facilitatory effect of average radius of co-occurrence was present in the LDT only for abstract words. Furthermore, for abstract words there was an inhibitory effect of average radius of co-occurrence on SDT latencies. This may be consistent with Mirman and Magnuson's (2006) findings of trade-offs between close and distant semantic neighbors. That is, Mirman and Magnuson noted that while greater semantic neighborhood density is typically facilitatory, close neighbors can sometimes exert an inhibitory effect on semantic processing. To explain the different pattern of results that we observed across tasks, we would need to further assume that task demands interact with semantic neighborhood structure to exert different effects of average radius of co-occurrence in the LDT and SDT. These and other differences between concrete and abstract meaning could be explored in future studies utilizing this dataset along with more fine-grained measures of semantic neighborhood characteristics.

The limited set of analyses in the present study are intended merely to assess the potential of our dataset for testing the effects of semantic variables. Preliminary results from these analyses suggest that the database has promise, but certainly there are many more semantic variables we have not tested

here, as well as the promise of novel variables not yet characterized. Indeed, many researchers now assume that semantic representation is multidimensional—that is, composed of several different types of information, including both linguistic or language-based information and experiential or object-based information (e.g., Barsalou, Santos, Simmons, & Wilson, 2008; Binder & Desai, 2011; Dove, 2009; Louwerse, 2010; Vigliocco, Meteyard, Andrews, & Kousta, 2009). The Calgary SDT dataset offers researchers the opportunity to test the independent and joint effects of these variables on the processing of concrete and abstract word meanings.

For instance, it has been argued that emotion information may play a particularly important role in the representation of abstract meaning (Vigliocco et al., 2009), but the literature on emotion variables in lexical-semantic processing is quite mixed. Some studies have shown that valence has a linear effect on lexical processing, with faster LDT latencies to positive than to negative words (Estes & Adelman, 2008; Kuperman, Estes, Brysbaert, & Warriner, 2014; Larsen, Mercer, Balota, & Strube, 2008). Other studies have shown that the effect of valence is better described as an inverted U-shape, with faster LDT latencies for both positive and negative words as compared with neutral words (Kousta, Vinson, &

Vigliocco, 2009; Vinson, Ponari, & Vigliocco, 2014; Yap & Seow, 2014). Finally, other ways of measuring emotional information have been characterized, and these need to be compared to the more traditional constructs of valence and arousal (Moffat, Siakaluk, Sidhu, & Pexman, 2015; Newcombe, Campbell, Siakaluk, & Pexman, 2012). These issues could be pursued with the present dataset.

Similarly, it has been argued that contextual and situational information is particularly important to abstract meaning (Wilson-Mendenhall, Simmons, Martin, & Barsalou, 2013), but we need to better characterize this type of information, and there are ongoing efforts to do so (e.g., Moffat et al., 2015; Recchia & Jones, 2012). The Calgary SDT dataset offers researchers the unprecedented opportunity to explore each of these issues using a task for which substantial variance is explained by semantic variables. As such, use of the present dataset holds strong potential for allowing new insights and progress.

Author note This work was supported by a Discovery Grant from the Natural Sciences and Engineering Research Council (NSERC) of Canada to P.M.P. The authors thank Ian S. Hargreaves and Amanda Fernandez for assistance with stimulus selection and programming.

References

- Adelman, J. S., Brown, G. D. A., & Quesada, J. F. (2006). Contextual diversity, not word frequency, determines word naming and lexical decision times. *Psychological Science*, 17, 814–823.
- Amsel, B. D. (2011). Tracking real-time neural activation of conceptual knowledge using single-trial event-related potentials. *Neuropsychologia*, 49, 970–983. doi:10.1016/j.neuropsychologia.2011.01.003
- Amsel, B. D., & Cree, G. S. (2013). Semantic richness, concreteness, and object domain: An electrophysiological study. *Canadian Journal of Experimental Psychology*, 67, 117–129.
- Amsel, B. D., Urbach, T. P., & Kutas, M. (2012). Perceptual and motor attribute ratings for 559 object concepts. *Behavior Research Methods*, 44, 1028–1041. doi:10.3758/s13428-012-0215-z
- Balota, D. A., Cortese, M. J., Sergent-Marshall, S. D., Spieler, D. H., & Yap, M. J. (2004). Visual word recognition of single-syllable words. *Journal of Experimental Psychology: General*, 133, 283–316. doi:10.1037/0096-3445.133.2.283
- Balota, D. A., Ferraro, F. R., & Connor, L. T. (1991). On the early influence of meaning in word recognition: A review of the literature. In P. J. Schwanenflugel (Ed.), *The psychology of word meanings* (pp. 187–218). Hillsdale, NJ: Erlbaum.
- 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. doi:10.3758/BF03193014
- Balota, D. A., Yap, M. J., Hutchison, K. A., & Cortese, M. J. (2012). Megastudies: What do millions (or so) of trials tell us about lexical processing? In J. S. Adelman (Ed.), *Visual word recognition: Vol. 1. Models and methods, orthography and phonology* (pp. 90–115). Hove, UK: Psychology Press.
- Barsalou, L. W., Santos, A., Simmons, W. K., & Wilson, C. D. (2008). Language and simulation in conceptual processing. In M. De Vega, A. M. Glenberg, & A. C. Graesser (Eds.), *Symbols and embodiment: Debates on meaning and cognition* (pp. 245–283). Oxford, UK: Oxford University Press.
- Binder, J. R., & Desai, R. H. (2011). The neurobiology of semantic memory. *Trends in Cognitive Sciences*, 15, 527–536. doi:10.1016/j.tics.2011.10.001
- Brysbaert, M., & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods*, 41, 977–990. doi:10.3758/BRM.41.4.977
- Brysbaert, M., Stevens, M., Mander, P., & Keuleers, E. (2016). The impact of word prevalence on lexical decision times: Evidence from the Dutch Lexicon Project 2. *Journal of Experimental Psychology: Human Perception and Performance*, 42, 441–458. doi:10.1037/xhp0000159
- Brysbaert, M., Warriner, A. B., & Kuperman, V. (2013). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods*, 46, 904–911. doi:10.3758/s13428-013-0403-5
- Buchanan, L., Westbury, C., & Burgess, C. (2001). Characterizing semantic space: Neighborhood effects in word recognition. *Psychonomic Bulletin & Review*, 8, 531–544.
- Coltheart, M., Davelaar, E., Jonasson, J. T., & Besner, D. (1977). Access to the internal lexicon. In S. Dornic (Ed.), *Attention and performance VI* (pp. 535–555). Hillsdale, NJ: Erlbaum.
- 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
- Dove, G. O. (2009). Beyond perceptual symbols: A call for representational pluralism. *Cognition*, 110, 412–431.
- Durda, K., & Buchanan, L. (2006). WordMine2 [Online]. Available at www.wordmine2.org
- Estes, Z., & Adelman, J. S. (2008). Automatic vigilance for negative words is categorical and general. *Emotion*, 8, 453–457. doi:10.1037/a0012887
- Faust, M. E., Balota, D. A., Spieler, D. H., & Ferraro, F. R. (1999). Individual differences in information-processing rate and amount: Implications for group differences in response latency. *Psychological Bulletin*, 125, 777–799. doi:10.1037/0033-2909.125.6.777
- Ferrand, L., New, B., Brysbaert, M., Keuleers, E., Bonin, P., Méot, A., . . . Pallier, C. (2010). The French Lexicon Project: Lexical decision data for 38,840 French words and 38,840 pseudowords. *Behavior Research Methods*, 42, 488–496. doi:10.3758/BRM.42.2.488
- Hargreaves, I. S., Pexman, P. M., Pittman, D. J., & Goodyear, B. G. (2011). Tolerating ambiguity: Ambiguous words recruit the left inferior frontal gyrus in absence of a behavioral effect. *Experimental Psychology*, 58, 19–30.
- Hargreaves, I. S., White, M., Pexman, P. M., Pittman, D., & Goodyear, B. G. (2012). The question shapes the answer: The neural correlates of task differences reveal dynamic semantic processing. *Brain and Language*, 120, 73–78.
- Hino, Y., Pexman, P. M., & Lupker, S. J. (2006). Ambiguity and relatedness effects in semantic tasks: Are they due to semantic coding? *Journal of Memory and Language*, 55, 247–273.
- Hoffman, P., Lambon Ralph, M. A., & Rogers, T. T. (2013). Semantic diversity: A measure of semantic ambiguity based on variability in the contextual usage of words. *Behavior Research Methods*, 45, 718–730. doi:10.3758/s13428-012-0278-x
- Hoffman, P., & Woollams, A. M. (2015). Opposing effects of semantic diversity in lexical and semantic relatedness decisions. *Journal of Experimental Psychology: Human Perception and Performance*, 41, 385–402.
- Jared, D., & Seidenberg, M. S. (1991). Does word identification proceed from spelling to sound to meaning? *Journal of Experimental Psychology: General*, 120, 358–394.

- Juhasz, B. J., Yap, M. J., Dicke, J., Taylor, S. C., & Gullick, M. M. (2011). Tangible words are recognized faster: The grounding of meaning in sensory and perceptual systems. *Quarterly Journal of Experimental Psychology*, 64, 1683–1691. doi:10.1080/17470218.2011.605150
- Keuleers, E., & Balota, D. A. (2015). Megastudies, crowdsourcing, and large datasets in psycholinguistics: An overview of recent developments. *Quarterly Journal of Experimental Psychology*, 68, 1457–1468.
- 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), 1–15. 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. doi:10.3758/s13428-011-0118-4
- Kousta, S.-T., Vinson, D. P., & Vigliocco, G. (2009). Emotion words, regardless of polarity, have a processing advantage over neutral words. *Cognition*, 112, 473–481. doi:10.1016/j.cognition.2009.06.007
- Kuperman, V., Estes, Z., Brysbaert, M., & Warriner, A. B. (2014). Emotion and language: Valence and arousal affect word recognition. *Journal of Experimental Psychology: General*, 143, 1065–1081.
- Larsen, R. J., Mercer, K. A., Balota, D. A., & Strube, M. J. (2008). Not all negative words slow down lexical decision and naming speed: Importance of word arousal. *Emotion*, 8, 445–452. doi:10.1037/1528-3542.8.4.445
- Louwerse, M. (2010). Symbol interdependency in symbolic and embodied cognition. *Topics in Cognitive Science*, 3, 1–30.
- Lund, K., & Burgess, C. (1996). Producing high-dimensional semantic spaces from lexical co-occurrence. *Behavior Research Methods, Instruments, & Computers*, 28, 203–208. doi:10.3758/BF03204766
- McRae, K., Cree, G. S., Seidenberg, M. S., & McNorgan, C. (2005). Semantic feature production norms for a large set of living and nonliving things. *Behavior Research Methods*, 37, 547–559. doi:10.3758/BF03192726
- Miller, G. A. (1990). Word Net: An on-line lexical database. *International Journal of Lexicography*, 3, 235–312.
- Mirman, D., & Magnuson, J. S. (2006). The impact of semantic neighborhood density on semantic access. In R. Sun & N. Miyake (Eds.), *Proceedings of the 28th Annual Conference of the Cognitive Science Society* (pp. 1823–1828). Mahwah, NJ: Erlbaum.
- Moffat, M., Siakaluk, P. D., Sidhu, D. M., & Pexman, P. M. (2015). Situated conceptualization and semantic processing: Effects of emotional experience and context availability in semantic categorization and naming tasks. *Psychonomic Bulletin & Review*, 22, 408–419. doi:10.3758/s13423-014-0696-0
- Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (1998). The University of South Florida word association, rhyme, and word fragment norms [Database]. Retrieved March 9, 2011, from <http://w3.usf.edu/FreeAssociation>
- Newcombe, P. I., Campbell, C., Siakaluk, P. D., & Pexman, P. M. (2012). Effects of emotional and sensorimotor knowledge in semantic processing of concrete and abstract nouns. *Frontiers in Human Neuroscience*, 6, 275. doi:10.3389/fnhum.2012.00275
- Pexman, P. M. (2012). Meaning-level influences on visual word recognition. In J. S. Adelman (Ed.), *Visual word recognition: Vol. 2. Meaning and context, individuals, and development* (pp. 24–43). Hove, UK: Psychology Press.
- Pexman, P. M., Hargreaves, I. S., Siakaluk, P. D., Bodner, G. E., & Pope, J. (2008). There are many ways to be rich: Effects of three measures of semantic richness on visual word recognition. *Psychonomic Bulletin & Review*, 15, 161–167. doi:10.3758/PBR.15.1.161
- Pexman, P. M., Holyk, G. G., & Monfils, M.-H. (2003). Number-of-features effects and semantic processing. *Memory & Cognition*, 31, 842–855. doi:10.3758/BF03196439
- Piercey, C. D., & Joordens, S. (2000). Turning an advantage into a disadvantage: Ambiguity effects in lexical decision versus reading tasks. *Memory & Cognition*, 28, 657–666.
- Recchia, G., & Jones, M. N. (2012). The semantic richness of abstract concepts. *Frontiers in Human Neuroscience*, 6, 315. doi:10.3389/fnhum.2012.00315
- Rodd, J., Gaskell, G., & Marslen-Wilson, W. (2002). Making sense of semantic ambiguity: Semantic competition in lexical access. *Journal of Memory and Language*, 46, 245–266. doi:10.1006/jmla.2001.2810
- Schneider, W., Eschman, A., & Zuccolotto, A. (2001). *E-Prime user's guide*. Pittsburgh, PA: Psychology Software Tools.
- Shaoul, C., & Westbury, C. (2010). Exploring lexical co-occurrence space using HiDex. *Behavior Research Methods*, 42, 393–413. doi:10.3758/BRM.42.2.393
- Siakaluk, P. D., Pexman, P. M., Aguilera, L., Owen, W. J., & Sears, C. R. (2008). Evidence for the activation of sensorimotor information during visual word recognition: The body–object interaction effect. *Cognition*, 106, 433–443. doi:10.1016/j.cognition.2006.12.011
- Sibley, D. E., Kello, C. T., & Seidenberg, M. S. (2009). Error, error everywhere: A look at megastudies of word reading. In N. Taatgen & H. van Rijn (Eds.), *Proceedings of the 31st annual conference of the cognitive science society* (pp. 1036–1041). Austin, TX: Cognitive Science Society.
- 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, 263–273. doi:10.3758/s13428-013-0355-9
- Taikh, A., Hargreaves, I. S., Yap, M., & Pexman, P. M. (2015). Semantic classification of pictures and words. *Quarterly Journal of Experimental Psychology*, 68, 1502–1518.
- Tillotson, S. M., Siakaluk, P. D., & Pexman, P. M. (2008). Body–object interaction ratings for 1,618 monosyllabic nouns. *Behavior Research Methods*, 40, 1075–1078. doi:10.3758/BRM.40.4.1075
- Tousignant, C., & Pexman, P. M. (2012). Flexible recruitment of semantic richness: Context modulates body–object interaction effects in lexical–semantic processing. *Frontiers in Human Neuroscience*, 6, 53. doi:10.3389/fnhum.2012.0053
- Uttl, B. (2002). North American Adult Reading Test: Age norms, reliability, and validity. *Journal of Clinical and Experimental Neuropsychology*, 24, 1123–1137.
- van Casteren, M., & Davis, M. H. (2007). Match: A program to assist in matching the conditions of factorial experiments. *Behavior Research Methods*, 39, 973–978. doi:10.3758/BF03192992
- Vigliocco, G., Meteyard, L., Andrews, M., & Kousta, S. (2009). Toward a theory of semantic representation. *Language and Cognition*, 1, 219–247. doi:10.1515/LANGCOG.2009.011
- Vinson, D., Ponari, M., & Vigliocco, G. (2014). How does emotional content affect lexical processing? *Cognition and Emotion*, 28, 737–746.
- Wilson-Mendenhall, C. D., Simmons, W. K., Martin, A., & Barsalou, L. (2013). Contextual processing of abstract concepts reveals neural representations of non-linguistic semantic content. *Journal of Cognitive Neuroscience*, 25, 920–935.
- Yap, M. J., Pexman, P. M., Wellsby, M., Hargreaves, I. S., & Huff, M. J. (2012). An abundance of riches: Cross-task comparisons of semantic richness effects in visual word recognition. *Frontiers in Human Neuroscience*, 6, 72. doi:10.3389/fnhum.2012.00072
- 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. doi:10.3758/BRM.42.4.992
- Yap, M. J., & Seow, C. S. (2014). The influence of emotion on lexical processing: Insights from RT distributional analysis. *Psychonomic Bulletin & Review*, 21, 526–533. doi:10.3758/s13423-013-0525-x

- Yap, M. J., Tan, S. E., Pexman, P. M., & Hargreaves, I. S. (2011). Is more always better? Effects of semantic richness on lexical decision, speeded pronunciation, and semantic classification. *Psychonomic Bulletin & Review*, 18, 742–750. doi:[10.3758/s13423-011-0092-y](https://doi.org/10.3758/s13423-011-0092-y)
- Yarkoni, T., Balota, D. A., & Yap, M. J. (2008). Moving beyond Coltheart's *N*: A new measure of orthographic similarity. *Psychonomic Bulletin & Review*, 15, 971–979. doi:[10.3758/PBR.15.5.971](https://doi.org/10.3758/PBR.15.5.971)
- Zdravilova, L., & Pexman, P. M. (2013). Grasping the invisible: Semantic processing of abstract words. *Psychonomic Bulletin & Review*, 20, 1312–1318. doi:[10.3758/s13423-013-0452-x](https://doi.org/10.3758/s13423-013-0452-x)
- Zeno, S. M., Ivens, S. H., Millard, R. T., & Duvvuri, R. (Eds.). (1995). *The educator's word frequency guide*. Brewster, NJ: Touchstone.