Proceedings of the Twenty-Fourth International Florida Artificial Intelligence Research Society Conference

Simulating Human Ratings on Word Concreteness

Shi Feng, Zhiqiang Cai, Scott Crossley, and Danielle S. McNamara

Institute for Intelligent Systems
University of Memphis
Memphis, TN, 38152
sfeng, zcai@ memphis.edu, eslscc@langate.gsu.edu, dsmcnamr@ memphis.edu

Abstract

Psychological measures of concreteness of words are generally estimated by having humans provide ratings of words on a concreteness scale. Due to the limits of this technique, concreteness ratings in current word databases (e.g., MRC) are incomplete due to the limited size of the word samples. In this study, we use available linguistic databases to formulate a computational model to simulate human ratings on word concreteness. The computational model includes *Lexical Type*, *Latent Semantic Analysis Dimensions*, *Hypernymy Levels*, *Word Frequency* and *Word Length*. Our results indicate that the model accounts for 64% variance of human ratings.

Introduction

A single word in the human language has many complex dimensions such as semantics, parts of speech, lexical type, imagability, concreteness, familiarity, etc. It is important to know the dimensions of words in languages so that we can develop a better theoretical understanding of language and also to build tools that simulate human intelligence and performance. One important dimension of words is their level of *concreteness*. Concrete words such as house, poodle, and tiger evoke mental images quickly and easily in contrast to less concrete words such as *causality*, *evolution* and *mortal*. Words with higher concreteness are easier to imagine, comprehend, and memorize (e.g., Paivio, 1991).

Because of the importance of word concreteness to comprehension, processing, and memory, word concreteness also plays an important role in the fields of text and discourse and computational linguistics. Indeed, word concreteness is among the most important indices provided by Coh-Metrix (e.g., Graesser, McNamara, Louwerse, & Cai, 2004; McNamara & Graesser, in press).

In several studies, word concreteness values have played important roles in distinguishing between types of texts and parts of texts. For example, McCarthy, Renner, Duncan, Duran, Lightman, and McNamara (2008) included word concreteness as one of the indices to identify topic sentencehood. Crossley and McNamara (2009) included word concreteness to assess lexical differences in writings by first and second English language speakers. Graesser, Jeon, Cai, and McNamara (2008) used word concreteness in genre classification. Most recently, word concreteness has emerged in one of eight aspects of language in text that characterizes text difficulty (Duran, Bellissens, Taylor, & McNamara, 2007; McNamara & Graesser, in press)

However, word concreteness is not an attribute that a computer can directly compute. One means of assessing the characteristics of words is by having humans rate them on the dimensions of interest. Humans are proficient in categorizing words into linguistic dimensions, but it is impractical to have humans rating tens of thousands of words that we would need for psycholinguistic research. As a consequence, any particular corpus of words will comprise a limited number of words. Then, either the word concreteness values are not available for certain words, or averages over corpora of words can be misleading because the average contains an indeterminate number of missing values. Thus, as powerful as concreteness values have been, they have been imperfect.

Our goal in this study is to develop a computational model to predict word concreteness to overcome this problem. Our approach is to use information about words available from other sources to build a computational algorithm that will predict word concreteness, even when word concreteness human ratings are not available.

Word Databases

MRC. One widely used source for concreteness ratings is the MRC Psycholinguistic database (Paivio, Yuille, & Madigan, 1968; Toglia & Battig, 1978; Gilhooly & Logie, 1980). MRC is an online service that provides a resource for public research purposes. MRC contains 150,837 words and provides information on 26 different linguistic properties. It differs from other machine usable dictionaries in that it not only includes syntactic information but also psychological data for the entries. However, it provides concreteness ratings for only 8,228 words. Therefore, the following databases are used in this study to develop an algorithm to predict word concreteness.

WordNet. A second word database resource for psycholinguistic research is WordNet. WordNet is an online lexical database that includes English nouns, verbs, adjectives, and adverbs which are organized into sets of synonyms, each representing a lexicalized concept, and the semantic relations linking these synonym sets (Miller, Beckwith, Fellbaum, Gross, & Miller, 1990). WordNet contains more than 166,000 word form and sense pairs, and incorporates a variety of semantic relations that can be defined between word forms and word senses.

WordNet categorizes words into 45 lexical types, such as *food*, *plant*, *act*, *feel*, and *communication*. Because a word may have multiple senses, a word can be of multiple lexical types. In addition, these different lexical types have differences in concreteness. For example, words of *food* type tend to be more concrete than words of *feeling* type because one is a concrete object, whereas the other is an abstract human affect.

WordNet also provides estimates of word hypernymy levels. Hypernymy relations estimate the semantic links between words taxonomically. For example, *animal* is semantically related to *dog*, but animal is superordinate taxonomically to dog. As such, animal is more abstract, whereas dog is more concrete. Hence, words that are lower in hypernymy values (superordinate terms) also tend to be more abstract (Crossley, Salsbury, & McNamara, 2009; Graesser, McNamara, Louwerse, & Cai, 2004; Graesser, Jeon, Cai, & McNamara, 2008).

Latent Semantic Analysis. Latent Semantic Analysis (LSA) is a computational algorithm that uses a statistical method to yield a representation comprised of hundreds of dimensions (≈ 300) that can be used to perform tasks related to human assessments of language. LSA is not a complete model of language but it has been highly successful in predicting or simulating a number of language-related tasks (Landauer, McNamara, Dennis, & Kintsch, 2007). For example, it shares an 85%-90% overlap with expert human readers in assessing word sorting evaluations, word synonymy judgments, vocabulary learning (Landauer & Dumais, 1997), and word relatedness judgments (Landauer, Foltz, & Laham, 1998). LSA has been effectively used in assessing word similarities and solving other language problems (Landauer et al., 1998). In LSA, each word is represented by an N-dimensional vector. We hypothesized that some of the dimensions would provide information about word concreteness. Given a large corpus, an LSA space would thus provide dimensional attributes for as many terms as there were in the corpus.

CELEX. Word frequency is another attribute that can be obtained for large set of words. The CELEX database from the Dutch Centre for Lexical Information contains word frequency count for more than 160,000 words (Baayen, Piepenbrock, & Gulikers, 1996).

The Current Study

In this study, we compute word attributes from WordNet, LSA and CELEX and use these attributes to simulate human ratings in the MRC database. Our overarching goal is to optimize Coh-Metrix estimations of word concreteness, as well as other computational linguistic algorithms. Our goal is to develop an algorithm that simulates human ratings of word concreteness using a variety of freely available, automated lexical features. If successful, such an approach will allow us to estimate the human concreteness values of words that have not been judged by human raters. As such, we can potentially solve limitations resulting from relying on using word databases for human ratings.

Method

Human Rating of Concreteness

For this study, we examined the human ratings of concreteness provided in the MRC Psycholinguistic database. These concreteness values are based on the works of Paivio, Yuille, and Madigan (1968), Toglia and Battig (1978), and Gilhooly and Logie (1980), who used human subjects to rate large collections of words for psychological properties. Specifically, participants in these studies were asked to score the concreteness of words based on a numerical scale (from 1 to 7). A word that refers to an object, material, or person generally received a higher concreteness score than an abstract word (Toglia & Battig, 1978). Although the MRC database has 150,837 word entries, only a subset of these words were rated for concreteness.

For this study, we selected 3521 unique nouns from the database that had concreteness ratings. For the 3521 nouns, the minimum concreteness rating is 195 and the maximum is 670. The mean concreteness rating is 459.28 and the standard deviation is 116.63. Ten percent of the nouns have a concreteness rating below 291 and 10% of the nouns have a concreteness rating above 600. The lexical features examined were lexical types, LSA dimensions, word hypernymy, word frequency, and word length (i.e., number of letters). These are each discussed in greater detail in the following sections.

Lexical Types. WordNet contains data on about 81,000 nouns that are sub-classified into 26 different lexical types. These lexical types are presented in Table 1 (ordered from least to most concrete), as well as the mean concreteness for each lexical type computed by averaging the concreteness ratings for each of the selected 3521 nouns that are assigned to that lexical type.

Table 1: Mean Concreteness Values for Each Lexical Type in WordNet

Lexical Type	N	Mean	SD	Examples
feeling	156	323.62	84.35	hate, fear
motive	13	362.46	103.61	obsession,
relation	66	371.83	85.45	Causality
cognition	497	372.26	100.58	Algebra
state	452	383.85	100.48	measles, sneeze
attribute	397	387.03	117.57	rigidity, agility
process	62	394.39	92.64	evolution, vapor
act	847	400.51	99.22	tennis, battle
time	139	404.47	94.73	Daybreak
event	295	408.65	93.47	surf, faint
tops	43	409.02	100.93	animals, mortal
communication	701	416.35	106.84	movie, medal
phenomena	120	440.36	102.51	typhoon, sleet
possession	163	446.85	104.21	coin, money
group	338	460.94	95.17	bunch, corps
location	212	466.10	101.11	lair, exterior
quantity	179	468.17	106.70	quart, volt
person	657	489.70	96.57	clown, boy
shape	111	507.13	86.24	Rectangle
object	198	517.39	87.22	pond, beach
artifact	1160	526.38	85.61	Necklace
substance	244	534.82	79.80	Firewood
body	191	540.17	82.42	Forearm
animal	277	558.91	73.44	robin, ox
food	265	560.13	76.27	breakfast, lunch
plant	182	560.49	78.33	larch, tulip

For each lexical type, we created a proportion index. The proportion index of a given word *w* to a given lexical type *T* is defined as the proportion of the number of senses of *w* belonging to the lexical type *T* divided by the total number of senses of *w*. For instance, the word "line" belongs to 10 lexical types (*artifact, communication, cognition, group, location, shape, act, phenomena, possession,* and *quantity*). Each of these lexical types contains a variety of senses of the word "line" (i.e., 8 senses for the lexical type *artifact*, 5 for *communication*, 4 for *group*, 3 for *location*, 2 for *shape*, etc.). Thus, the proportion score for the word *line* in the lexical type *artifact* would be .276 (the 8 senses contained in the lexical type *artifact divided by* the 29 senses).

LSA Dimension Attributes. We also examined LSA dimensions as potential predictors of human scores of lexical concreteness. We limited our analysis to the first 156 dimensions (of the potential ≈300 dimension reported in an LSA space). The LSA space we use in this study was generated from the Touchstone Applied Science Associates (TASA) corpus. The TASA corpus contains about 37,000 documents comprising 90,000 words. The LSA space for this corpus is a 300-dimensional vector representation of all the words in the corpus. The 300-dimensional vectors were generated by compressing the weighted co-occurrence word-document matrix using singular value decomposition technique. These 300 dimensions correspond to the largest singular values of the word-document matrix. We computed LSA dimensions scores because we hypothesized that they store important word information and we predicted they could explain a certain amount of variance in human judgments of word concreteness.

Hypernymy Level. We examined links between concreteness scores and hypernymy scores following the hypothesis that word specificity would correlate with word concreteness (Crossley, Salsbury, & McNamara, 2009). Hypernymic relations are hierarchical associations between hypernyms (superordinate words) and hyponyms (subordinate words). A hypernym is defined as a word that is more general than a related word (*animal* compared to *dog*) and a hyponym is more specific than a related word (*dog* as compared to *animal*). Hypernymic relations in WordNet form a tree structure. Each sense of a word is mapped to a certain node on the tree. For example, the sense of the noun "line" as "the trace of a moving point" has 5 specific hypernyms ("line" => "shape" => "attribute" => "abstraction" => "abstract entity" => "entity"). When computing the hypernymy score, the word "line" would receive a score of 5 (for this sense). For each sense of the word, a score would be computed. For example, the sense of "line" as "a formation of people or things one behind another" has the following hypernym chain: "line"=> "formation" => "arrangement" => "group" => "abstraction" => "abstract entity" => "entity". This sense would receive a hypernymy score of 6. For the final score for the word, we computed an average hypernymy score for all the senses contained within the word. For some senses of a noun, the hypernyms may form a tree of multiple branches, instead of a chain. In that case, we simply count the hypernymy level of every branch and take the average over the branches as the hypernymy score for the given sense. We hypothesize that words that receive higher scores (and thus are more specific) would be more concrete than words that receive lower scores.

Polysemy. We also investigated links between word polysemy and word concreteness. Polysemous words are words that have more than one sense. For instance, the word "list" has 30 senses and is thus highly polysemous.

The word "class" has 8 senses. The word "apple" has two senses, and is thus not highly polysemous. Highly polysemous words are generally more frequent (Zipf, 1945). Additionally, highly polysemous words also exhibit higher degrees of ambiguity (Davies & Widdowson, 1974). We computed polysemy scores by calculating the number of senses for each word contained in WordNet. We hypothesized that words that contained more senses (and were potentially more ambiguous) would be less concrete.

Word Frequency. We computed frequency scores for the nouns in the MRC Database using the CELEX Database (Baayen, Piepenbrock, & Gulikers, 1996). The CELEX Database provides word frequencies for more than 160,000 word forms. To compute the CELEX frequency score, we calculated the logarithm of the CELEX frequency count for each selected noun. We hypothesized that more frequent words would be more concrete.

Word Length. We computed the length of a word as the number of letters in the word. Word length is a strong proxy for word frequency, with the advantage that it is available for all words rather than only a subset of words. Typically, more frequent (shorter) words tend to be more concrete and thus we expected word length to be negatively correlated with concreteness values.

Statistical Analysis

We separated the 3521 nouns from the MRC database into *training* and *test* sets based on a 67/33 split. The training set comprised 2348 words and their related concreteness scores. The test set comprised 1173 words and their related concreteness scores. We then calculated correlations to examine what lexical features of the words in the training set correlated with the human scores of concreteness. A step-wise regression analysis on the training set was used to examine which lexical variables were most predictive of human scores of word concreteness. Lastly, we used the model reported in the regression analysis on the held-back test set to examine how well the model predicted the variance in concreteness scores on words for which it had not been trained.

Results

Pearson Correlations Training Set

Pearson correlations between concreteness and 186 variables were conducted; these variables included the 26 lexical types, the 156 LSA dimensions, the hypernym scores, number of letters in the word, the logarithm of the word frequency and the polysemy scores. 125 insignificant variables were excluded from the regression training set described below. We selected 61 variables that are significantly correlated to concreteness ratings ($p<0.05$), These 61 variables include 21 lexical types, hypernym, number of letters, logarithm of word frequency and 37 LSA dimensions. The excluded non-significant lexical types from the correlation analysis were *phenomenon*, *quantity*, *location*, *group*, and *possession*.

The selected LSA dimensions were: 1-4, 6-9, 11,13-15, 17, 21-22, 24-25, 27, 30-33, 38, 49, 54, 61, 63, 68, 73, 80, 102, 110, 112, 117, 129, 135, and 139. Interestingly, our results suggest that the information about concreteness is mainly stored in earlier dimensions.

Multiple Regression Training Set

A stepwise regression analysis was conducted with the 61 variables that demonstrated small effect sizes or greater. These 61 variables were regressed onto the human concreteness scores for the 2348 words in the training set. The variables were checked for multicollinearity. Coefficients were checked for variance inflation factors (VIF) values. All VIF values were at about 1 which is equivalent of tolerance levels well beyond the .2 threshold. This model indicates that the model data did not suffer from multicollinearity (Field, 2005).

The linear regression using the 61 variables yielded a significant model, $F(39, 2353) = 108.850$, $p < .001$, $r = .802$, $r^2 = .643$. Thirty-nine variables were significant predictors in the regression. Twenty-two variables were not significant predictors. *R, r^2, ß, B,* and Standard Error information for each of the included variables is presented in Table 2. The results from the linear regression demonstrate that the combination of the 39 variables accounts for 64% of the variance in the human evaluations of word concreteness found in the MRC Psycholinguistic database.

Test Set Model

To further support the results from the multiple regression conducted on the training set, we used the *ß* weights and the constant from the training set multiple regression analysis to estimate how the model would function on an independent data set (the 1173 words along with their human ratings of concreteness held back in the test set). The model produced an estimated concreteness value for each word in the test set. We then conducted a Pearson Correlation between the estimated concreteness score and the concreteness score. This correlation along with its r^2 is indicative of the strength of the model on an independent data set. The model for the test set yielded $r = .821$, $r^2 = .674$. The results from the test set model demonstrate that the combination of the 39 variables accounted for 67% of the variance in the human scores of concreteness for 1173 words comprising the test set. A Breush-Pagan test was conducted to check for homoscedasticity. The χ^2 reported a p-value of .229, suggesting that the unstandardized residuals are normally distributed.

Table 2: Linear Regression Model to Predict Human Concreteness Ratings

Variable	r	r²	ß	B	SE
Constant				338.723	13.183
artifact	0.403	0.162	.274	100.201	7.528
Num of letters	0.482	0.232	-.087	-4.460	0.710
food	0.538	0.289	.156	133.960	12.253
animal	0.584	0.341	.136	108.841	12.241
person	0.619	0.383	.098	41.250	7.777
substance	0.649	0.421	.158	131.885	12.096
plant	0.676	0.457	.113	107.407	13.271
body	0.695	0.483	.093	111.313	16.171
log word freq.	0.708	0.501	.228	14.863	0.998
cognition	0.722	0.521	-.205	-150.144	11.120
object	0.733	0.538	.070	82.252	16.139
attribute	0.743	0.552	-.184	-130.687	10.569
feeling	0.751	0.565	-.151	-187.385	16.499
LSA 4	0.759	0.575	-.121	-1341.564	148.352
hypernym	0.766	0.587	.139	10.205	1.125
act	0.770	0.593	-.168	-91.019	9.298
LSA 1	0.774	0.599	-.135	-1927.644	225.602
shape	0.776	0.603	.039	83.742	27.368
LSA 15	0.778	0.606	-.048	-598.881	162.930
LSA 2	0.780	0.609	.062	752.001	161.152
state	0.782	0.612	-.100	-78.269	11.486
comm.	0.784	0.615	-.121	-63.674	8.766
time	0.787	0.620	-.080	-86.793	14.897
LSA 38	0.789	0.622	-.049	-568.807	147.478
LSA 139	0.790	0.624	-.042	-459.569	136.095
LSA 6	0.791	0.626	.059	649.575	141.753
LSA 54	0.793	0.628	.049	585.203	151.174
relation	0.794	0.630	-.042	-116.906	35.124
event	0.795	0.632	-.047	-49.636	14.321
LSA 32	0.796	0.633	.034	341.268	127.861
LSA 17	0.797	0.635	-.039	-475.259	155.894
LSA 9	0.798	0.636	.050	580.544	153.242
LSA 31	0.799	0.638	.050	587.464	155.397
LSA 3	0.799	0.639	-.035	-428.699	161.792
LSA 22	0.800	0.640	-.037	-420.308	147.305
motive	0.801	0.641	-.032	-201.791	77.775
LSA 21	0.801	0.642	.031	362.914	148.854
LSA 135	0.802	0.643	.028	296.141	132.293
LSA 30	0.802	0.643	-.026	-289.682	142.576

Discussion

The results of this study indicate that it is possible to formulate a model that predicts human ratings of noun concreteness using automated lexical indices. Our model consists of 39 attributes that include 19 lexical type attributes, 17 LSA dimension attributes, 1 hypernymy level attribute, 1 word frequency and 1 word length attribute. This model predicts 64% of the variance in human ratings of word concreteness

Our strongest predictors of human ratings of word concreteness were lexical types. Our analysis indicated that words with higher concreteness ratings were more likely to be categorized as *artifacts*, *foods*, *animals*, *people*, *substances*, *plants*, or *body parts*. Less concrete words were more likely to be categorized as related to *cognition*, *action*, *shapes*, *communication*, *relations*, *states*, *events*, *time*, or *motives*.

Our next strongest predictors of human judgments of word concreteness were the number of letters per word and word frequency. Word length was negatively correlated to human ratings on concreteness, as expected ($r = -.324$). Word frequency also had a positive correlation with word concreteness ($r = .058$).

One of our goals was to examine the role of LSA dimensions in predicting word concreteness. The results of this study indicate that a small part of word concreteness information is stored in some of the lower dimensions within the LSA space. Few studies have drawn upon information contained within the particular LSA dimensions to predict human performance. Hence, this study provides innovative evidence that the LSA dimensions provide individual contributions to simulating human cognition.

Another consideration regarded the relationship of hypernymy to word specificity and exploiting that relationship to predict concreteness. However, the results indicated that word hypernymy level explains only a small variance of concreteness. Because this result may be biased from the way in which hypernymy was assessed, in future research we intend to examine different methods of calculating hypernymy (i.e., other than counting the number of branches). We assumed that a word at the bottom of the branch would be more concrete than a word at the top of the branch. However, we might also take into account the number of levels of each branch. For example, if a hypernym branch has 6 levels, and the position of the target word is 4, than the word would be 4/6 of a relationship to concreteness. Thus, the ratio of the position of a word and the number of levels from a hypernym branch may also be predictive of concreteness. Future studies will explore this and other possible uses of hypernymy in predicting word concreteness.

We were also interested in those indices that were not predictive of human judgments of concreteness. For instance, we hypothesized that more concrete words would be less ambiguous (i.e., less polysemous). This was not the case. Within the last decade, there has been an increasing development of available digital databases and computational algorithms from which to explore the language use, language meaning, and language processing, as well as automated applications that make use of computational information about language. Here, we demonstrate how to augment some of those databases automatically, in this case for estimations of word concreteness. One can suggest, however, using such an approach to develop estimations of any number of features of language. Indeed, it may be possible to use information about a limited number of words to predict characteristics of an infinite set of words. We do not suggest, of course, that our model can necessarily replace human ratings, but our model proves to be a useful tool in assessing the relationships between the given linguistic characteristics of words and human judgments of those words. Our future research will continue to examine such characteristics.

Acknowledgments

This research was supported in part by the Institute for Education Sciences (IES R305A080589). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the IES.

References

Baayen, H., Piepenbrock, R., & Gulikers, L. (1996, Jan 15). *LDC catalog*. Retrieved from http://www.ldc.upenn.edu/Catalog/CatalogEntry.jsp?catalogId=LDC96L14.

Cohen, J (1992). A power primer. *Psychological Bulletin, 112*, 155–159

Crossley, S.A. & McNamara, D.S. (2009). Computationally assessing lexical differences in L1 and L2 writing. *Journal of Second Language Writing, 18,* 119-135.

Crossley, S.A., Salsbury, T., & McNamara, D.S. (2009). Measuring second language lexical growth using hypernymic relationships. *Language Learning, 59*, 307-334.

Davies, A., & Widdowson, H. (1974). Reading and writing. In J. Allen & S. Corder (Eds.), *Techniques in applied linguistics* (pp. 154-201). London: Oxford University Press.

Duran, N., Bellissens, C., Taylor, R., & McNamara, D.S. (2007). Qualifying text difficulty with automated indices of cohesion and semantics. In D. S. McNamara & G. Trafton (Eds.), *Proceedings of the 29th Annual Meeting of the Cognitive Science Society* (pp. 233–238). Austin, TX: Cognitive Science Society.

Fields, A. (2005). *Discovering statistics using SPSS*. London: Sage Publications.

Gilhooly, K. L., & Logie, R. H. (1980) Age of Acquisition, imagery, concreteness, familiarity and ambiguity measures for 1944 words. *Behavioral Research Methods and Instrumentation* 12, 395-427.

Graesser, A.C., McNamara, D.S., Louwerse, M., & Cai, Z. (2004). Coh-Metrix: Analysis of text on cohesion and language. *Behavior Research Methods, Instruments, & Computers, 36*, 193-202.

Graesser, A.C., Jeon, M., Cai, Z., & McNamara, D.S. (2008). Automatic analyses of language, discourse, and situation models. In W. van Peer & J. Auracher (Eds.), *New Beginnings for Study of Literature*. New Castle: Cambridge Scholars Publications.

McNamara, D.S., & Graesser, A.C. (in press). Coh-Metrix: An automated tool for theoretical and applied natural language processing. In P.M. McCarthy & C. Boonthum (Eds.), *Applied natural language processing and content analysis: Identification, investigation, and resolution.* Hershey, PA: IGI Global.

Hu, X., Cai, Z., Wiemer-Hastings, P., Graesser, A. C., & McNamara, D. S. (2007). Strengths, limitations, and extensions of LSA. In T. K. Landauer, D. S. McNamara, S. Dennis, & W. Kintsch (Eds.), *The handbook of latent semantic analysis* (pp. 401-426). Mahwah, NJ:Erlbaum.

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problems: The latent semantic analysis theory of the acquisition, induction, and representation of knowledge. *Psychological Review, 104(2)*, 211-240.

Landauer, T. K., Foltz, P. W., & Laham, D. (1998). Introduction to latent semantic analysis. Discourse Processes, 25, 259-284.

Landauer, T. K., McNamara, D. S., Dennis, S., & Kintsch, W. (Ed.). (2007). *Handbook of latent semantic analysis*. Mahwah, NJ: Lawrence Erlbaum Associates Publishers.

McCarthy, P.M., Renner, A.M., Duncan, M.G., Duconduct, N.D., Lightman, E.J., & McNamara, D.S. (2008). Identifying topic sentencehood. *Behavior Research and Methods, 40,* 647-664.

Miller, G. A., Beckwith, R., Fellbaum, C. D., Gross, D., & Miller, K. (1990). WordNet: An online lexical database. *International Journal of Lexicograph*, 3(4), 235-244.

Paivio, A. (1991). Dual coding theory: Retrospect and current status. *Canadian Journal of Psychology*, *45*, 255–287.

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

Toglia, M. P., & Battig, W. F. (1978). Handbook of semantic word norms. Somerset, NJ: John Wiley & Sons, Inc.