

# Automatic Expansion of the MRC Psycholinguistic Database

## Imageability Ratings

Ting Liu<sup>1</sup>, Kit Cho<sup>1</sup>, George Aaron Broadwell<sup>1</sup>, Samira Shaikh<sup>1</sup>, Tomek Strzalkowski<sup>1,2</sup>, John Lien<sup>1</sup>, Sarah Taylor<sup>3</sup>, Laurie Feldman<sup>1</sup>, Boris Yamrom<sup>1</sup>, Nick Webb<sup>1</sup>, Umit Boz<sup>1</sup>, Ignacio Cases<sup>1</sup>, Ching-sheng Lin<sup>1</sup>

1 University at Albany, State University of New York 2 Institute of Computer Science, Polish Academy of Sciences 3 Sarah M. Taylor Consulting, LLC

E-mail: tliu@albany.edu, kitwcho@gmail.com

### Abstract

Recent studies in metaphor extraction across several languages (Broadwell et al., 2013; Strzalkowski et al., 2013) have shown that word imageability ratings are highly correlated with the presence of metaphors in text. Information about imageability of words can be obtained from the MRC Psycholinguistic Database (MRCPD) for English words and *Léxico Informatizado del Español Programa (LEXESP)* for Spanish words, which is a collection of human ratings obtained in a series of controlled surveys. Unfortunately, word imageability ratings were collected for only a limited number of words: 9,240 words in English, 6,233 in Spanish; and are unavailable at all in the other two languages studied: Russian and Farsi. The present study describes an automated method for expanding the MRCPD by conferring imageability ratings over the synonyms and hyponyms of existing MRCPD words, as identified in Wordnet. The result is an expanded MRCPD+ database with imageability scores for more than 100,000 words. The appropriateness of this expansion process is assessed by examining the structural coherence of the expanded set and by validating the expanded lexicon against human judgment. Finally, the performance of the metaphor extraction system is shown to improve significantly with the expanded database. This paper describes the process for English MRCPD+ and the resulting lexical resource. The process is analogous for other languages.

**Keywords:** MRC Psycholinguistic Database (MRCPD), wordnet, metaphor, imageability

## 1. Introduction

Scientists engaged in research using lexical stimuli must take into account a multitude of variables associated with words (e.g. familiarity, frequency of occurrence) that could potentially obscure the interpretation of their results. The MRC Psycholinguistic Database (MRCPD) provides 26 linguistic and psycholinguistic variables for 150,837 English words and has served as a valuable resource for researchers in a variety of disciplines such as artificial intelligence, computer science, and psychology (Coltheart, 1981; Wilson, 1988). One major limitation of the database is that not all words have ratings for all 26 variables. Of interest in the present study is the variable of *imageability* (i.e., how easily and quickly the word evokes a mental image) for which the MRCPD has ratings for only 9,240 (6%) of the total words in its database.

The present study sought to expand the MRCPD database by adding imagery ratings for an additional 106,911 words. This was done by taking imputing imageability ratings for words that are available in the MRCPD database to their synonyms and hyponyms as defined in the Wordnet lexical database (Miller, 1995). The results were validated using three different procedures. (1) Performing regression analyses to compare how well imageability ratings and word frequency of words predict subjects' reaction time in identifying the word. These variables were examined

because previous research reported a strong relationship among them (Baayen, Feldman, & Schreuder, 2006; Green & Brock, 2009). (2) Having human subjects make imageability ratings for both the original MRCPD and the expanded MRCPD+ words. (3) Applying expanded Imageability database on Robust Extraction of Metaphors in Novel Data (REMND) system (Strzalkowski et al., 2013), to assess improvement in performance.

This article is structured as follows: Section 2 presents related to the utility of imageability. Section 3 gives the algorithm of this expansion. Section 4 demonstrates the evaluation of the expansion and its contribution to metaphor extraction. Section 5 shows the expansion of MRCPD+ to other languages.

## 2. Imageability in language research

Words that are highly imageable (e.g., flag, banana, chair) invoke mental images that appeal to our human senses (i.e., reading the word "banana" will automatically lead one to form a mental image of the fruit) relative to those that are non-imageable (e.g., pride, love, virtue). Previous research shows that imageable and non-imageable items affect cognitive processing such that the former items undergo a higher degree of semantic elaboration and are encoded with two traces—lexically and pictorially (Pavio, 1969). As such, those in the marketing and communications industries have capitalized on the memorial benefits of highly

imageable words by increasing the imagerial properties of their message which has been shown to be an effective technique of increasing persuasion (Green & Brock, 2002; Rossiter, 1982).

While the applied utility of imagery research is apparent, researchers have long used imageable and non-imageable words to help shed light on fundamental cognitive processes in humans. Some of the more robust and reliable findings in the field of cognitive psychology have shown that imageable words are: (a) remembered better (Pavio, 1969), (b) acquired at an earlier age (Bird, Franklin, and Howard, 2001), and (c) identified more quickly as an English word (Baayen, Feldman, & Schreuder, 2006). Not surprisingly, highly imageable words are found in metaphors, and consequently, their presence can be used as an important indicator of metaphorical language (Broadwell et al., 2013). An expansion of the MRCPD database, therefore, would surely assist present and future researchers who are interested in automatic identification of metaphors.

### 3. Using WordNet to expand MRCPD

Princeton's WordNet (Miller, 1995) is a large English lexical database with over 150,000 words, hierarchically organized in synsets that capture semantically equivalent words. It is thus reasonable to assume that if one element of a synset has a known imageability score, all other words in this synset should have the same or closely related scores, and can be added to the expanded lexicon with the inherited imageability ratings. It is important to note that a word can have multiple senses defined in Wordnet and some of them may be less commonly used. It may not be appropriate to assign the same imageability ratings to all synsets of the word, because the MRCPD ratings represent words in their most common usage, which is typically captured by the first synset. For example, "dog" has seven senses, including: a domestic dog (the most common sense), a person who is morally reprehensible, as well as "a smooth-textured sausage", which do not have the same degree of imageability. Consequently, we only utilize the top-ranked senses (either top 1<sup>st</sup> or 3) for expansion.

In addition to the synonyms of the top ranked senses, we also propagate imageability scores over the hyponyms linked to the first synset. For example, the hyponyms of the first sense of "dog" include: puppy, basenji, lapdog, poodle, spitz, pooch, doggie, doggy, mutt, etc., each of which is now added to the expanded MRCPD with the same imageability score as "dog". We consider this assignment to be conservative, that is, some of the hyponyms may in fact be more imageable than their parent. We also note that if a word has multiple parents (hypernyms), the inheritance applies only to the closest parent (largest offset).

### 4. Evaluating the Expansion Method

We used three methods to validate our expansion method. In the first method, we assessed whether the expanded set matches certain structural characteristics of the original MRCPD, specifically, correlation between the reaction time and frequency of occurrence, and the degree to which they were predictive of the word's imageability score. The second method of validation relied on judgments made by human subjects. We presented subjects with the original MRC source word and its expansion word and had subjects rate both words on its degree of imageability and computed a correlation coefficient of the ratings. A positive correlation would indicate that there is a linear relationship of the two sets of words such that a word from the MRCPD database that received a high imageable rating would likewise receive a highly imageable rating to its expansion word and vice versa. We also compared the performance accuracy of our metaphor extraction system using the original and the expanded MRCPD.

#### 4.1 Structural validation

Previous research by Hargreaves and Pexman (2012) (see also Baayen et al. 2006) reported a strong negative correlation between a word's imageability rating and length of time (reaction time) it takes human subjects to recognize the word. This means that the *less* imageable the word, the *longer* it takes subjects to correctly identify it as a correct word. Further, this negative relationship also applies if a word's frequency of occurrence. This correlation is indeed observed in the original MRCPD, and we want to see if it holds up in the expansion.

Adopting to the approach of Hargreaves and Pexman (2012), we implemented a multiple regression analysis and randomly sampled a subset of the original MRCPD words and compare the characteristics of those words to the expansion words. This analysis allows one to determine, how strongly each individual predictor (imageability and frequency) accounts for the variance in reaction time, and how strongly both predictors together accounts for reaction time. To get data for a word's reaction time and frequency, we used the English Lexicon Project (ELP, Balota et al., 2007), which has collected ratings for 40,481 words on various measures including reaction time and word frequency (but not imageability). With reaction time and word frequency as the known variables, for all our words, we then assess whether the imageability scores for estimated words are correlated with (predictive of) reaction time and compare those values with the original MRC source words for which the imageability values from which these words were derived. If the expansion method we use is indeed a valid approach, one would expect that the correlation values and relationship among the variables for these two sources of words to be similar.

Three separate samples were evaluated to assess the appropriateness of our expansion method and how well the results generalize depending on type of words for

which values were imputed. The samples differed in the type of expansion word that was assessed: (1) values imputed for most frequent synset, (2) values imputed for the three most frequent synsets, and (3) values imputed for the most frequent synset's hyponyms/hypernyms. For each of the three tests, a new sample of MRCPD source words was selected to ensure that there was no sampling bias and that the words we selected in any one sample were representative of the complete database. For each MRCPD source word, the new expansion word for which we are testing its imputed imageability rating was selected, i.e. each MRC source word and its estimated word was selected and assessed as a pair. Words for which ELP did not have data were dropped from the analysis. The final sample size for the first, second, and third sample is: 245, 199, and 168, respectively. The results of the multiple regression analyses are reported in Table 1.

As expected, we found a negative relationship between imageability on the one hand, and frequency and reaction time, on the other, in all samples, including the expansion. To assess how well both predictors can account for changes in reaction time, the Adjusted  $R^2$  was computed. In all samples, the  $R^2$  value of the estimated words were either highly similar (sample 2) or higher (samples 1 and 3) than that of the MRCPD original word samples, and for both sets, the effect size is large, indicating strong predictive validity of these variables, and the model is statistically significant (see the  $F$  value).

## 4.2 Human validation

We also used human subjects (recruited from the University at Albany psychology participant pool) to

validate the samples by assigning imageability to a subset of words. The total sample size was 200 (100 MRCPD original words and 100 expansion words) for all three samples. The sample size was reduced because we were concerned with subjects becoming fatigued and a diminishment of motivation as the task progressed. The number of subjects completing each sample was 22. The subjects judged the imageability of the words on a 7-point Likert scale. As a measure of quality control, for each item, we averaged the response for each word across all subjects and compared each subject's individual response to the mean response of that item. A response that was greater than 3 standard deviations from the mean of that item was removed because their response was not representative of the population of interest. This resulted in the exclusion of 7 trials from the second sample and 2 trials from the third sample. Before computing the correlation coefficient of the imageability ratings between the MRCPD original words and the expansion, a scatterplot was constructed so as to examine whether there are any outliers (Cohen, Cohen, West, & Aiken, 2002). One pair of item was removed from the second sample and another pair was removed from the third sample. The correlation coefficients for the three samples are: .570, .339, and .355, all of which were statistically significant,  $p < .05$ , two-tailed. These correlations show that (1) the expansion method we report here is valid for all three samples, and (2) the expansion method is more robust when considering the top synset (the first sample) than compared to the other two samples.

Sample	Word Set	Predictors	Standardized $\beta$	Adj. $R^2$	$F$
(1): Top Synset	MRCPD original	Imageability	-.291*	.492	117.35*
		Frequency	-.624*		
	Estimated expansion	Imageability	-.197*	.584*	169.96*
		Frequency	-.747*		
(2): Top 3 Synsets	MRCPD original	Imageability	-.351*	.497*	96.87*
		Frequency	-.594*		
	Estimated expansion	Imageability	-.101*	.493*	95.44*
		Frequency	-.709*		
(3): Hyponyms	MRCPD original	Imageability	-.330*	.509*	85.67*
		Frequency	-.649*		
	Estimated expansion	Imageability	-.138*	.625*	137.61*
		Frequency	-.796*		

Note. All values are statistically significant,  $*p < .05$ , two-tailed

Table 1: Multiple regression analyses using imageability and log frequency to predict log reaction time reaction time

	System Extracted Metaphors	Precision	Recall
Original MRCPD used	139	66.1%	57.1%
Expanded MRCPD+ used	163	66.3%	67.5%

Table 2. REMND performance with/without imageability expansion

### 4.3 Performance of Metaphor Extraction

Word’s imageability plays an important role in REMND system for linguistic metaphor detection (Broadwell et al., 2013). To test how well the expanded imageability database increases the accuracy of our metaphor detection system, we selected 389 passages containing both literal and metaphorical expressions from the domains of governance and economic inequality. The expressions were judged by human subjects recruited through Amazon Mechanical Turk, using a 7-point Likert scale for its metaphoricity and imageability. At least 30 judgments (or subjects) rated each item (inter-rater consistency > 0.7). In the set, 161 passages were marked as metaphors and 228 were marked as literal.

Table 2 compares the performance of our metaphor system (REMND) (vs. the judgments of human subjects) on this test data when using the original MRCPD and when using the expanded MRCPD+. We note that the system is able to find 24 additional metaphors, which makes for a marked improvement in recall without losing precision. This, and other similar tests we run, validates the utility of the expanded lexicon in practical applications.

### 5. Expanding Imageability database into other languages

Creating an extended imageability lexicon for other languages can be accomplished by following the same expansion and validation process as outlined above for English; however, there are at least two additional challenges to overcome. The first challenge is the availability of the initial imageability lexicon in the target language. For example, there is a (smaller) version of MRCPD for Spanish (Sebastián Gallés, 2000), but none we are aware of for Russian or Farsi, the other two language we are targeting in our work. The second challenge is the availability of sufficiently large Wordnet database for that language. For example, while Spanish imageability lexicon contains 6,233 words, the small size of Spanish Wordnet (30,485 synsets vs. 117,000 synsets in English Wordnet) means we can only get expand

imageability ratings over an additional 2,000 words, a tiny fraction of English expansion.

A way around these issues is to translate the expanded English MRCPD+ database into another language through a mechanical process, e.g., Google Translate, with scores averaged in case of many-to-one translations. Table 3 shows the size of translated MRCPD+ in 3 target languages using this approach. We are now in process of validating these lexicons, and the initial evidence from the REMND system suggests that they perform as expected.

### 6. Conclusion

The present study expanded the imageability ratings for the words in the MRCPD database from 9,240 to 116,151 by imputing imageability ratings for the MRCPD source words to its synonyms and hyponyms identified using WordNet. We also converted expanded MRCPD database into other languages. The resulting expanded resource has been validated and we hope would become a valuable addition to language resources available for research.

### 7. Acknowledgement

This research is supported by the Intelligence Advanced Research Projects Activity (IARPA) via Department of Defense US Army Research Laboratory contract number W911NF-12-C-0024. The U.S. Government is authorized to reproduce and distribute reprints or Governmental purposes notwithstanding any copyright annotation thereon. Disclaimer: The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of IARPA, DoD/ARL, or the U.S. Government.

### 8. References

Baayen, R. H., Feldman, L. B., & Schreuder, R. (2006). Morphological influences on the recognition of

Language	Size of translated MRCPD+
Spanish	110,671
Russian	91,925
Farsi	79,201

Table 3. Imageability lexicons translated from English MRCPD+

- monosyllabic monomorphemic words. *Journal of Memory and Language*, 55, 290-313.
- Balota, D.A., Yap, M.J., Cortese, M.J., Hutchinson, K.A., Kessler, B., Loftis, B., Neely, J.H., Nelson, D. L., Simpson, G.B., & Treiman, R. (2007). The English lexicon project. *Behavioral Research Methods*, 39(3): 445-459.
- Bird, H., Howard, D., & Sue, F. (2001). Age of acquisition and imageability ratings for a large set of words, including verbs and function words. *Behavioral Research Methods, Instruments, and Computers*, 33(1): 73-79.
- Broadwell, G.A., Boz, U., Cases, I., Strzalkowski, T., Feldman, L., Taylor, S., Shaikh, S., Liu, T., & Cho, K. (2013). Using Imageability and Topic Chaining to Locate Metaphors in Linguistic Corpora. In *Proceedings of International Conference on Social Computing, Behavioral-Cultural Modeling, & Prediction*. 2013. Washington D.C.
- Cohen, J., Cohen, P., West, S.G., & Aiken, L.S. (2002). *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences* (3<sup>rd</sup> ed.). Mahwah, NJ: Lawrence Erlbaum Associates Publishers.
- Coltheart, M. (1981). The MRC Psycholinguistic Database. *Quarterly Journal of Experimental Psychology*, 33A: 497-505.
- Google Scholar. (2012). Retrieved from <http://scholar.google.com>.
- Green, M. C., & Brock, T.C. (2002). In the mind's eye: Transportation-imagery model of narrative persuasion. In Melanie C. Green, Jeffrey J. Strange, and Timothy C. Brock (Eds.), *Narrative impact: Social and cognitive foundations* (315-341). Mahwah, NJ: Lawrence Erlbaum Associates Publishers.
- Hargreaves, I.S., & Pexman, P.M. (2012). Does richness lose its luster? Effects of extensive practice on semantic richness in visual word recognition. *Frontiers in Human Neuroscience*, 6: 1-11.
- Miller, G. A. (1995). WordNet: A Lexical database for English. *Communications of the ACM*, 38(11): 39-41.
- Pavio, A. (1969). Mental imagery in associative learning and memory. *Psychological Review*, 76(3): 241-263.
- Ratcliff, R. (1993). Methods for dealing with reaction time outliers. *Psychological Bulletin*, 33: 510-532.
- Rossiter, J. R. (1982). Visual imagery: Applications to Advertising. In Andrew Mitchell (Ed.), *Advances in Consumer Research Vol. 9: Association for Consumer Research*: 101-106.
- Sebastián Gallés, Núria (coord.), M. Antònia Martí Antonín, Manuel Francisco Carreiras Valiña, and Fernando Cuetos Vega. LEXESP, Léxico Informatizado del Español Programa CORCO. Barcelona: Edicions de la Universitat de Barcelona, 2000.
- Strzalkowski, T., Broadwell, G.A., Taylor, S., Feldman, L., Shaikh, S., Liu, T., Yamrom, B., Cho, K., Boz, U., Cases, I., and Elliot, K. (2013.) Robust Extraction of Metaphor from Novel Data. In *Proceedings of Workshop on Metaphor in NLP, NAACL*. 2013 Atlanta.
- Wilson, M. D. (1988). The MRC Psycholinguistic Database: Machine Readable Dictionary, Version 2. *Behavioural Research Methods, Instruments and Computers*, 20(1): 6-11.

## Appendix: a fragment of MRCPD+

The table below shows a small fraction of the expanded MRCPD English lexicon and Russian Lexicon (with English Translation). Scores above 0.67 (on a scale of 0-1) are considered high imageability; it is one of the thresholds we use for locating metaphorical expressions.

Original MRCPD Word	Synset	Hyponym	Imageability Score	POS
ego	egotism	n/a	0.447	n
fruit	n/a	seed	0.839	n
quake	earthquake	submarine earthquake	0.661	n
kill	n/a	slaughter	0.687	v
religion	faith	christianity	0.620	n
acidity	sour	tartness	0.707	n
adversity	hardship	distress	0.610	n
alien	foreigner	importee	0.671	n
alternative	option	preference	0.383	n

Word	Translated from	POS	Imageability Score
вариант	option	n	0.440
самолюбие	egotism/conceit/self-importance...	n	0.511
землетрясение	quake/earthquake/seism/temblor	n	0.661
религия	religion	n	0.620
вера	faith/believe	n	0.523
тайна	mystery/arcanum	n	0.597
кислый	sour	n	0.707