

DATA PAPER

Taxonomic and Thematic Relatedness Ratings for 659 Word Pairs

Jon-Frederick Landrigan¹ and Daniel Mirman²¹ Department of Psychology, Drexel University, Philadelphia, PA, United States
Jon.Landrigan@gmail.com² Department of Psychology, Drexel University, Philadelphia, PA, United States and
Moss Rehabilitation Research Institute, Elkins Park, PA, United States
dan@danmirman.org

These data are comprised of taxonomic and thematic relatedness ratings for 300 target words paired with taxonomically related and/or thematically related words for a total of 659 word pairs. The pairs come from 4 prior studies and were normed through surveys provided online using Amazon's Mechanical Turk. Pairs were rated in terms of both their taxonomic similarity and their thematic relationship. The data are provided as comma-separated (.csv) and R data (.rdata) files and can be used to create new studies investigating taxonomic and thematic semantic processing.

Keywords: Semantics; Taxonomic; Thematic; Similarity; Related**Funding statement:** This research was supported by a grant from the National Institutes of Health (R01DC010805) to D.M. and by Drexel University.

(1) Overview

Context

Collection Date(s)

September–November 2014 and June–July 2015.

Background

Understanding how taxonomic relations (based on feature similarity) and thematic relations (based on co-occurrence in events) operate in the mind has been a long standing topic of interest among cognitive psychologists. Investigators have used a variety of tasks to understand how these relations operate and coincide in the mind, including free sorting, forced choice categorization, similarity ratings, and semantic decision paradigms. Recently the neural organization and processing of taxonomic and thematic semantic memory has also become an increasing topic of debate among cognitive scientists [1, 4, 6]. To date however, there has not been a consistent quantitative definition of taxonomic and thematic relatedness. Feature overlap or biological taxonomy is often used to define taxonomic similarity. For thematic relatedness, some researchers have used latent semantic analysis while others have used their expert opinion and thus there has been a lack of consistency in terms of materials across studies [1–3]. This lack of consistency may be a possible reason for the conflicting results regarding the organization and neural mechanisms responsible for taxonomic and thematic processing [1, 2, 6]. In this dataset,

word pairs were collected from four independent studies investigating differences in the processing of taxonomic and thematic semantics and re-normed by obtaining the strength of taxonomic and thematic relationships for each pair of words. This was done in order to provide the field with a common set of normed items for future research.

(2) Methods

Sample

Participants were recruited through Amazon's Mechanical Turk services. All participants were from the United States and were paid for their participation in the surveys (payment varied depending on survey length). Participants were limited to completing only one survey. Participant's responses were only included if they completed 90% or more of the survey, were from the United States of America and spoke English as their primary language. Responses from a total of 157 participants were included in the final data (85 females and 72 males). Ages ranged from 19 to 67 ($M = 34.8$, $SD = 10.63$). All but four participants were native English speakers. The four participants who were not native English speakers learned English by the age of 10.

Materials

The 659 word pairs found in this data set come from four studies conducted by independent research groups. The norming was done through a series of 12 online surveys

(6 taxonomic and 6 thematic surveys). Each survey was comprised of 100 to 300 randomly assigned word pairs and the pairs were randomly ordered within each survey.

Procedures

Through Amazon's Mechanical Turk web services, participants were provided a link to an external survey on either Survey Monkey or Qualtrics websites. Each survey included a brief demographic questionnaire before giving participants the instructions for the survey. There were two different types of surveys – one for taxonomic similarity ratings and one for thematic relatedness ratings – each with different types of instructions.

The instructions for the taxonomic survey were:

“Thank you for participating. In this survey you will be presented with a number of word pairs and asked to rate the similarity of the two words on a scale from 1 (not at all similar) to 7 (very similar). Using the radio buttons below each word pair select 1, 2, 3, 4, 5, 6, or 7 to rate the similarity between the words. Two words are similar if they look alike or belong to the same category. For example, DOTS and STRIPES are similar (both are types of patterns or designs). However, SHIRT and STRIPES would not be similar. Even though STRIPES are often found on SHIRTS, a SHIRT is a type of clothing while STRIPES are not. Another example is ZEBRA and STRIPES, these two words are also not very similar, because they belong to different categories, animal and pattern categories respectively. Please use the full range of the scale (1, 2, 3, 4, 5, 6, or 7) in indicating your responses. Only the buttons below the word pair will work for rating the items. Please make sure to rate all the pairs in the survey.”

The thematic survey instructions were:

“Thank you for participating. In this survey you will be presented with a number of word pairs and asked to rate how connected and or related the two words are on a scale from 1(not related at all) to 7 (very related). Using the radio buttons below each word pair select 1, 2, 3, 4, 5, 6 or 7 to rate the relatedness between the two words. Two words are connected or related if they occur in the same time or place, however, this does not mean they will share similar physical features. For example HELMET and MOTORCYCLE are related (one wears a HELMET while riding a MOTORCYCLE, although they are different shapes and sizes). Whereas CHRISTMAS-TREES and PALM-TREES are not related, because even though they are both trees and share similar features they do not occur in the same time or place. Please use the full range of the scale (1, 2, 3, 4, 5, 6, or 7) in indicating your responses. Only the buttons below the word pair will work for rating the items. Please make sure to rate all the pairs in this survey.”

Each word pair was included in each of the two survey types. Once the surveys were completed, the data were

downloaded and collated, then average taxonomic similarity and thematic relatedness ratings were computed for each word pair. Taxonomic and thematic rating standard deviations were also computed for each of the items. In addition, a difference score was then calculated for each word pair by subtracting the thematic relatedness rating from the taxonomic similarity rating for each word pair. Thus, items with high taxonomic similarity and low thematic relatedness have positive difference scores (6 being the highest); items with negative scores have high thematic relatedness and low taxonomic similarity (–6 being the highest); pairs with difference scores near 0 are approximately equally taxonomically similar and thematically related (this includes both approximately equally high and equally low ratings).

Participant's individual ratings of word pairs are also provided in a separate data table. This table includes a participant id number, their gender, age, native language, the age they learned English if it is not their native language, the country they live in, the rating value and the rating type (i.e. taxonomic or thematic). In total, there are 27317 individual word pair ratings provided in this table. It should be noted that two participants (identified in the individual data-set as id40 and id138) did not report their ages and are denoted as NA in the data-set.

Quality Control

All surveys were distributed through Amazon's Mechanical Turk web services and completed on either Survey Monkey or Qualtrics. Surveys were screened to ensure that participants did not just provide random responses or the same response repeatedly.

Ethical issues

This study had been deemed exempt from IRB review because it involved minimal risk survey procedures without collection of individually identifying information. All participants were paid for their work.

(3) Dataset description

Object name

The dataset appears in the repository as listed below and can be downloaded in the following formats:

Aggregated Data:

TxThmNorms.rdata

TxThmNorms.csv

Individual Rater Data:

IndividualRatingsTxThm.rdata

IndividualRatingsTxThm-1.csv

Data type

Processed data.

Format names and versions

The TxThmNorms and the IndividualRatingsTxThm are provided in the following formats: .rdata and .csv. The rdata files contain single data frames which are analogous to the tables contained in the csv files. For examples of the data sets look to **Table 1** and **Table 2** below.

Ref_Word	Pair_Word	Mean_Rating_Tx	SD_Rating_Tx	Mean_Rating_Thm	SD_Rating_Thm	Difference_Score	Num_Ratings_Tx	Num_Ratings_Thm
BIRD	LAMB	2.95	1.94	2.84	1.92	0.11	21	19
SHOP	MARKET	5.71	1.74	6	1.53	-0.29	21	18
HOOVER	MOWER	3	2.05	2.32	1.6	0.68	21	19
VASE	BUCKET	4.57	1.72	2.63	1.86	1.94	21	19
APPLE	LIME	4.62	1.88	3.79	1.69	0.83	21	19
EXAM	PROGRAMME	3.14	1.96	3.58	1.95	-0.44	21	19
GRAFFITI	POSTER	3.67	1.93	3.95	2.15	-0.28	21	19

Table 1: First 8 rows of the TxThmNormsData table.

ID	Gender	Age	Level_of_Education	First_Language	Age_Learned_English_If_Not_Native	Country_Living_In	Ref_Word	Pair_Word	Rating	Rating_Type
id1	F	19	High School or below	Russian	3	US	BIRD	LAMB	4	Taxonomic
id1	F	19	High School or below	Russian	3	US	SHOP	MARKET	7	Taxonomic
id1	F	19	High School or below	Russian	3	US	HOOVER	MOWER	5	Taxonomic
id1	F	19	High School or below	Russian	3	US	VASE	BUCKET	6	Taxonomic
id1	F	19	High School or below	Russian	3	US	APPLE	LIME	6	Taxonomic
id1	F	19	High School or below	Russian	3	US	EXAM	PROGRAMME	3	Taxonomic

Table 2: First 8 rows of the IndividualRatingsTxThm table.

Data Collectors

N/A.

Language

English (United States of America).

License

Attribution 4.0 International (CC BY 4.0).

Embargo

N/A.

Repository location<http://dx.doi.org/10.7910/DVN/FKTQ4C>**Publication date**

December 22, 2015.

(4) Reuse potential

These aggregated norms have multiple forms of reuse potential. First, they could be used to investigate the organization and processing mechanisms of taxonomic and thematic semantic memory in typical adults and children, and in neurologically impaired individuals. Second, they could be used as comparison/control data for investigating typical and atypical conceptual development and cognitive aging, and for testing semantic processing in acquired neurological impairments such as stroke and dementia, including helping to track disease progression or recovery. Finally, the norms could provide preliminary guidance for studies using picture stimuli with high name agreement (though we recommend directly norming the pictures because pictures can evoke somewhat different performance in semantic tasks [5]). The individual ratings could be used to examine gender or age differences in ratings of taxonomic similarity or thematic relatedness among other possible uses.

Competing Interests

The authors declare that they have no competing interests.

Acknowledgements

We would like to thank Allison Britt and Leah Friedman for their aid in survey development and testing.

References

1. **Jackson, R L, Hoffman, P, Pobric, G and Lambon Ralph, M A** 2015 The Neural Correlates of Semantic Association versus Conceptual Similarity. *Cerebral Cortex*, 20(2010): 8524. DOI: <http://dx.doi.org/10.1093/cercor/bhv003>
2. **Lewis, G A, Poeppel, D and Murphy, G L** 2015 The neural bases of taxonomic and thematic conceptual relations: An MEG study. *Neuropsychologia*, 1–13. DOI: <http://dx.doi.org/10.1016/j.neuropsychologia.2015.01.011>
3. **Maguire, M J, Brier, M R and Ferree, T C** 2010 EEG theta and alpha responses reveal qualitative differences in processing taxonomic versus thematic semantic relationships. *Brain and Language*, 114(1): 16–25. DOI: <http://dx.doi.org/10.1016/j.bandl.2010.03.005>
4. **Mirman, D and Graziano, K M** 2013 Damage to temporo-parietal cortex decreases incidental activation of thematic relations during spoken word comprehension, 50(8): 1990–1997. DOI: <http://dx.doi.org/10.1016/j.neuropsychologia.2012.04.024>
5. **Saffran, E M, Coslett, H B and Keener, M T** 2003 Differences in word associations to pictures and words. *Neuropsychologia*, 41(11): 1541–1546. DOI: [http://dx.doi.org/10.1016/S0028-3932\(03\)00080-0](http://dx.doi.org/10.1016/S0028-3932(03)00080-0)
6. **Schwartz, M F, Kimberg, D Y, Walker, G M, Brecher, A, Faseyitan, O K, Dell, G S, . . . , Coslett, H B** 2011 Neuroanatomical dissociation for taxonomic and thematic knowledge in the human brain. *Proceedings of the National Academy of Sciences of the United States of America*, 108: 8520–8524. DOI: <http://dx.doi.org/10.1073/pnas.1014935108>

Peer review comments: <http://dx.doi.org/10.5334/jopd.24.pr>

How to cite this article: Landrigan, J-F and Mirman, D 2016 Taxonomic and Thematic Relatedness Ratings for 659 Word Pairs. *Journal of Open Psychology Data*, 4: e2, DOI: <http://dx.doi.org/10.5334/jopd.24>

Published: 08 February 2016

Copyright: © 2016 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.