

This document is downloaded from DR-NTU, Nanyang Technological University Library, Singapore.

| | |
|-----------|---|
| Title | Automatic Identification of Basic-Level Categories |
| Author(s) | Mills, Chad; Bond, Francis; Levow, Gina-Anne |
| Citation | Mills, C., Bond, F., Levow, G.-A. (2018). Automatic Identification of Basic-Level Categories. The 9th Global WordNet Conference (GWC 2018). |
| Date | 2018 |
| URL | http://hdl.handle.net/10220/44917 |
| Rights | © 2018 The author(s). This is the author created version of a work that has been peer reviewed and accepted for publication by The 9th Global WordNet Conference (GWC 2018). It incorporates referee's comments but changes resulting from the publishing process, such as copyediting, structural formatting, may not be reflected in this document. The full-text is available at: [http://compling.hss.ntu.edu.sg/events/2018-gwc/pdfs/GWC2018_paper_5.pdf]. |

Automatic Identification of Basic-Level Categories

Chad Mills

University of Washington
Seattle, WA, United States
chills@uw.edu

Francis Bond

Nanyang Technological University
Singapore
bond@ieee.org

Gina-Anne Levow

University of Washington
Seattle, WA, United States
levow@uw.edu

Abstract

Basic-level categories have been shown to be both psychologically significant and useful in a wide range of practical applications. We build a rule-based system to identify basic-level categories in WordNet, achieving 77% accuracy on a test set derived from prior psychological experiments. With additional annotations we found our system also has low precision, in part due to the existence of many categories that do not fit into the three classes (superordinate, basic-level, and subordinate) relied on in basic-level category research.

1 Introduction

WordNet organizes concepts into a hierarchy of hypernyms and hyponyms (Miller 1995). While WordNet also identifies other information, such as meronymy, one interesting property that is not currently captured is which concepts represent basic-level categories.

This is an important and valuable property to capture. Brown (1958) first noted that, although there are many terms that could be used to refer to an object at different levels of abstraction, “it often happens that a hierarchy develops in both directions from a middle level of abstraction.” Rosch et al. (1976) called this the basic-level, identifying psychological advantages basic-level categories have as well as psychological tests to find these concepts in a hierarchy. Examples of basic-level categories include *table*, *car*, *tree*, *bird*, *guitar*, *shirt*, *fish*, and *apple* (Rosch et al. 1976).

Unfortunately, though, the process of identification does not scale well and only dozens of

these concepts have been identified in the psychology literature (Rosch et al. 1976, Markman and Wisniewski 1997).

While there has been little work to automate the identification of basic-level categories (discussed in Section 2), knowing the basic-level has been shown to be valuable. Knowing the basic-level helps with word sense disambiguation (Legrand 2006), image searches (Rorissa and Iyer 2008), ad targeting (Wang et al. 2015), accurately measuring the readability of a text (Lin et al. 2009), making search result entity cards more easily consumable (Wang et al. 2015), linking together different domain-specific information classification systems (Green 2006), and user-centered design of image-browsing interfaces (Rorissa and Iyer 2008). We also believe it could help with having a common set of words to work from in building WordNets for other languages, as well as language grounding and many other problem areas.

Given the wide variety of demonstrated applications of this information as well as the opportunity for application in other areas, we attempt to automate the identification of basic-level categories.

We specifically look at heuristics to identify the basic-level noun categories in the Princeton WordNet of English (Fellbaum 1998), hereinafter PWN. One author assigned this task as a project in a class he taught in 2010 and 2011. This work builds on the various techniques students used and combines them with novel rules into a rule-based system to identify basic-level categories.

2 Related Work

2.1 Basic-level categories

Interest in basic-level concepts spans many disciplines, including philosophy (Rand 1966), psy-

chology (Rosch et al. 1976), library and information science (Green 2006), computer science (Wang et al. 2015), and others. While different disciplines have come up with very different theories to explain essentially the same underlying phenomena, they each bear many resemblances given the similarity in phenomena described.

While philosophy provides the foundation on which much of the work is based, and the field even has some work specifically on basic-level categories, the most numerous work on basic-level categories has been in psychology following the work of Rosch et al. (1976).

Rosch et al. (1976) distinguished between three levels of categories: basic-level, superordinate (hypernyms of the basic-level), and subordinate (hyponyms of the basic level). They found many properties of these categories, such as that basic-level categories are the most inclusive level at which a concrete picture of the category as a whole can be formed.

Markman and Wisniewski (1997) offer what may be a more fundamental and clear definition of the basic-level as being the level with the most alignable differences. An alignable difference is a difference in degree rather than kind; for example, cars and motorcycles have a different number of wheels (alignable) but a car carries a jack and a motorcycle does not (non-alignable). *Car* and *motorcycle* here are both taken to be basic-level categories, while *vehicle* is a superordinate and *coupe* is a subordinate. The various subordinates of car (*coupe*, *sedan*, etc.) vary in a handful of ways, but they have more similarities than differences. Cars and motorcycles, on the other hand, have many more differences and many of these are alignable (number of wheels, type of seat, steering controls, acceleration controls, etc.). According to (Markman and Wisniewski 1997), this abundance of alignable differences is a clear indicator that *car* and *motorcycle* are basic-level.

There has been a wide variety of additional research in this area within psychology showing a range of properties, applications, and even several potential issues with basic-level categories. Though before the concept was well-established, Brown (1958) noticed that children learn some middle level of concepts first, which Rosch et al. (1976) later showed was true of basic-level categories. Rosch et al. (1976) also showed basic-level category membership is verified fastest, objects are named faster at the basic-level, and objects are preferentially named with their basic-level category. Studies have shown children learn basic-level categories first, then subordinates,

then superordinates (Jónsdóttir and Martin 1996), with children not even considering a novel noun to potentially be a superordinate until around age 7 (Golinkoff et al. 1995).

At the same time, there are some limitations to these advantages. Adult experts in a domain may be so fluent with the subordinate level in that domain that some of the advantages of the basic-level over the subordinate level become greatly diminished (Tanaka and Taylor 1991). Still, even here the boundary between basic and superordinate concepts is an important one with qualitative differences in how they are represented, such as superordinate concepts (e.g. *furniture*) often referring to groups of entities and basic-level (e.g. *table*) referring to individuals (Murphy and Wisniewski 1989). Some interesting corner cases have also been found with abnormal exemplars, for example with *penguin* having the basic-level advantages but *bird* being the clear basic-level category for most birds (Jolicoeur et al. 1984).

Despite these and other limitations, though, there has been a surprisingly broad variety of research into applications of basic-level categories, as discussed to motivate the problem in Section 1, showing that a system identifying the basic-level would be valuable.

2.2 Identifying basic-level categories at scale

There has been very little work specifically on detecting basic-level categories at scale. The experiments in psychology have around a dozen examples of basic-level categories (Rosch et al. 1976, Markman and Wisniewski 1997).

There have only been a few efforts to use this data to learn patterns and extrapolate to a broader set of basic-level categories, all working with PWN, though some of the psychology literature also points out attributes of basic-level categories that may be helpful.

Farwell (2009) started with all nouns and did some filtering of superordinates and subordinates by depth in the hierarchy. This was followed by a voting scheme to pick the best candidate on each path from the top of the hierarchy to a leaf node, considering how short the word is, how frequently the word is used, and how many words are in the synset all as positive features while having few hyponyms and fewer relationships with other synsets more broadly as negative features (Green 2006). There was no effort to reconcile results from nearby paths down the hierarchy, though, and the list of basic-level categories generated

was fed into a downstream system to map information systems together, with no evaluation of the categories themselves.

Another effort focused on word sense disambiguation, with Izquierdo et al. (2007) using a simpler approach that filtered out the lower levels of the hierarchy and searched up the hypernym tree exclusively looking for a synset with a large number of PWN relations. These features were already included by (Green 2006), and here as well the evaluation was only performed on the applied system and an evaluation was not performed on this basic-level category identification system as such. Izquierdo et al. (2007) did make one important distinction, though, between basic-level categories and the similarly-named base concepts. Base concepts are a set of concepts core to many relations and tend to occur relatively high in the hierarchy (Izquierdo et al. 2007). On the other hand, while there is certainly overlap, basic-level concepts tend to occur closer to the middle of the hierarchy and tend to have less relations (Izquierdo et al. 2007).

Lin et al. (2009) attempted to identify basic-level categories by looking for words that are shorter than their hyponyms and where the word is frequently contained within its hyponyms as a compound. Again this was only evaluated in the application of measuring text readability, and like the other experiments they used all the available data for forming the rules without holding aside any data for an independent evaluation.

3 Data

We are aware of two major lists of basic-level categories as well as corresponding superordinates and some subordinates.

The original experiments that started much of the work in this area (Rosch et al. 1976) include nine superordinate taxonomies for their first two experiments. For the three of these superordinates falling in the biological taxonomy, the experimental results showed the presumed superordinate level (*tree, fish, bird*) is actually the basic-level. So, for these three groups the taxonomy was shifted down one level (e.g. basic to supordinate) and new superordinates (*plant, animal, animal*) were added to ensure the experimental results were accounted for. Additionally, eight additional basic-level categories were used in their later experiments 3-4 (Rosch et al. 1976), so these were also added. Markman and Wisniewski (1997) also provide a large list of superordinates,

basic-level categories, and subordinates, though there is overlap with the aforementioned list.

A summary of the lists is shown in Table 1.

| Level | Rosch | Markman | Combined |
|---------------|-------|---------|-----------|
| Superordinate | 8 | 24 | 24 |
| Basic-level | 29 | 80 | 92 |
| Subordinate | 45 | 25 | 68 |

Table 1: Categories with known classification by level

This is the data used for training and evaluating our system. The details of how the data is split up for that purpose is discussed in Section 5.

4 Our Approach

We start with 29 student projects each independently trying to solve this problem, cataloging the types of approaches and rules considered and then combining a slightly-constrained set of these, as well as novel rules, into a combined system.

While the goal is to produce one system by evaluating the collective set of rules, some boundaries are needed to constrain this. For example, one student only considered words also appearing in the ‘adventure’ category of the Brown corpus (Francis and Kucera 1964), a small, categorized corpus of English, which restricts the project beyond the goals of this work. We therefore start with a general approach common to most solutions (Section 4.1), describe the relevant rules (Section 4.2), experimenting to determine which Filtering Rules are more and less effective (Section 5.1), and then combine the more effective rules into a combined system before experimenting with a set of Voting Rules (Section 5.2).

4.1 General Approach

We start with all noun synsets in PWN. The available gold standard labels discussed in Section 3 are all nouns, though it is worth noting some research has indicated it is likely possible to extend the basic-level to other parts of speech (Lemaitre and Heller 2013).

We then take the labeled data from the psychology literature discussed in Section 3, manually map each of the categories to the closest PWN synset when one exists, and the goal becomes to extrapolate from these to other PWN synsets that are also at the basic-level and not at the superordinate or subordinate levels. In the psychology experiments (Rosch et al. 1976, Markman and Wisniewski 1997) this was done with words

whose senses were disambiguated by context, so we operate at the sense level. For our purposes, category and synset will be used interchangeably.

The students were identifying words, not synsets, though each student had to try to map words to synsets to use PWN features before producing a final list of words from there, losing the synset distinctions. For this work, we treat the basic-level as operating at the sense level and ensure our labels for training and evaluation are on PWN synsets to remove this unnecessary complexity.

Essentially everything the students did to identify basic-level categories can be generalized as one of two approaches:

1. filtering out nouns that are not basic-level or
2. on a particular path from the root to a leaf node in the hypernym/hyponym hierarchy, score each node and choose the optimal one as the basic-level on that path

We adopt both of these approaches, first applying a set of Filtering Rules to remove synsets unlikely to be basic-level and then choosing at most one per path based on a set of Voting Rules.

There were a few other extensions students considered, such as taking the top 2000 results with a provided sorting function, but since we do not want to assume a particular number of basic-level categories we do not incorporate these approaches into our system. Many students also deduped their final list, dealt with lemmatization, chose which word in a synset to use to represent the synset, and other issues that are not necessary when operating at the synset level and thus were omitted here.

4.2 Rules

We have cataloged the rules students used, along with our own novel rules, generalizing them and parameterizing rules where possible to enable experimenting with different thresholds. Note some rules focus on a word since students were not working on synsets, so for these rules we follow the convention most students followed in mapping words to synsets by taking the first lemma in the synset as the word for applying these rules.

The list of Filtering Rules is shown in Table 2, and the Boolean Voting Rules used for voting schemes to pick the best synset left in a chain after filtering are shown in Table 3. Parameter ranges used by students, or examples in cases where there are long lists of parameters, are shown after the rule. Ranges are given in interval notation to avoid boundary condition ambiguity.

| Filtering Rules |
|--|
| 1. Filter words with a set of suffixes (-ing, -ment, ... [59 total]) |
| 2. Filter words with a set of prefixes (un-, th-) |
| 3. Filter words of length n or greater [7, 16] |
| 4. Filter words of length n or fewer [1, 4] |
| 5. Filter space-separated compound words |
| 6. Filter hyphenated words ('-') |
| 7. Filter joined compounds (e.g. racetrack) |
| 8. Filter words with numbers |
| 9. Filter words with symbols |
| 10. Filter words with more adjective than noun senses |
| 11. Filter words with more adverb than noun senses |
| 12. Filter words with over 1 more verb than noun sense |
| 13. Filter words that are not substrings in immediate subordinate nodes |
| 14. Filter words containing any word at a higher level |
| 15. Filter stopwords |
| 16. Filter plural words |
| 17. Filter words with no vowels |
| 18. Filter words with over n vowels [1] |
| 19. Filter capitalized words |
| 20. Filter synsets with average depth $((\min+\max)/2, \text{recursive})$ outside the range a to b [4.2, 9) |
| 21. Filter synsets with hyponym depth $(\min+\max)/2$ outside the range a to b [1.1, 2.2) |
| 22. Filter synsets with $\text{avg_depth}/(\text{avg_depth}+\text{avg_height})$ outside the range a to b [.74, .91] |
| 23. Filter the top n levels of the hierarchy [2-7] |
| 24. Filter nodes with n levels below them (5) |
| 25. Filter synsets with an average depth $((\max+\min)/2)$ of $\leq n$ (5.4) |
| 26. Filter the bottom n levels of the hierarchy [1, 3] |
| 27. Filter synsets n or more levels deep [9, 15] |
| 28. Filter siblings of synsets with 0 hyponyms |
| 29. Filter nouns with a to b hyponyms [0,2], [5,inf) |
| 30. Filter synsets in the Brown corpus with frequency $< n$ (1-10) |
| 31. Filter synsets in the Brown corpus with frequency $> n$ (40) |
| 32. Filter all synsets under abstraction.n.06 |
| 33. Filter all synsets except those under set S (combinations of physical_entity.n.01, thing.n.08, substance.n.01, process.n.01) |
| 34. Filter all words in the CHILDES corpus |

- | |
|---|
| 35. Filter words in the CMU Pronouncing Dictionary with > 9 phonemes |
| 36. Filter all synsets with n or more siblings having no hyponyms |
| 37. Filter all synsets with at least p percent of siblings having no hyponyms |
| 38. Filter synsets with less than n siblings |
| 39. Filter words not in the Childes corpus |

Table 2: Filtering rules

| Voting Rules |
|---|
| 40. Top frequency in the chain (sum of lemma frequencies in synset) |
| 41. Top frequency in the chain in SEMCOR and frequency $\leq n$ (60) |
| 42. Word length between a and b [3, 7] |
| 43. Synset is of depth a to b in the hierarchy [6, 10] |
| 44. The word appears in Dolch's Word List |
| 45. The word appears in compound nouns |
| 46. Maximum % of children including the term as a compound in the chain |
| 47. The synset has hyponyms |
| 48. The highest value in the chain for (frequency in brown + 1)/15 + (compounds in hyponym subtree containing word + 1)/5 |
| 49. Highest frequency in Brown + Gutenberg corpora combined in the chain |
| 50. Maximum word length in chain |
| 51. Maximum number of meronyms in the chain |
| 52. Minimum word length in chain |

Table 3: Voting rules

Several resources are used in the Rules listed in Table 2 and Table 3. The Brown corpus (Francis and Kucera 1964) is a one million word corpus of American English. The CHILDES corpus (MacWhinney 2000) is a collection of transcripts of early language acquisition. The CMU Pronouncing Dictionary (Weide 1998) is a machine-readable English pronunciation dictionary which maps words to phonetic translations. SEMCOR (Landes et al. 1998) is a PWN sense-tagged corpus. Dolch’s Word List (Dolch 1948) is a list of 510 words commonly spoken by kindergarteners. The Gutenberg Corpus is a subset of the public domain books available on Project Gutenberg (Gutenberg n.d.) and made available by the Natural Language Toolkit (Loper and Bird 2002).

5 Experiments

For the purpose of evaluation, we mapped the gold standard labels mentioned in Section 3 to

synsets in PWN. Some categories, such as *green seedless grapes* and *double knit pants*, did not have corresponding PWN synsets and were discarded. The labels also included four superordinates under which the psychology experiments and PWN had substantial incompatibilities, and these were also discarded. For example, whereas one superordinate in psychology experiments was taken to be *exercise equipment* (Markman and Wisniewski 1997), the three basic-level categories underneath this mapped to very different hypernym trees in PWN: *sports equipment*, *exercise device*, and even *athletic facility*.

We then divided the mapped categories into a train, development, and test set. This division was done manually at the superordinate level rather than completely randomly because there are several hypernyms with many basic-level categories labeled underneath them and having those split across sets may result in reporting better-than-real-world results as a result of learning location-specific patterns. Instead, splits have been made manually at higher levels in the hypernym hierarchy, though the available labels leave out significant portions of the PWN hypernym hierarchy so this is still imperfect. The number of categories in each set is shown in Table 4:

| | Train | Dev | Test | Total |
|---------------|--------------|------------|-------------|--------------|
| Superordinate | 7 | 8 | 9 | 24 |
| Basic-level | 29 | 24 | 25 | 78 |
| Subordinate | 10 | 22 | 18 | 50 |
| Total | 46 | 54 | 52 | |

Table 4: Summary of the labels for the experiments

5.1 Filtering Rule Experiments

Our first step was to set parameters on each individual Filtering Rule (Table 2) on the train set and select the promising rules based on their performance on the development set. The filtering rules are designed to provide accurate filtering to remove many non-basic categories before applying voting rules where the system can be more robust to errors by combining multiple rules. Filtering rules were tuned on the train set to not filter out any basic-level categories but to filter out as many superordinates and subordinates as possible.

Of the 39 proposed filtering rules, only 15 could be tuned to avoid filtering out basic-level categories while also filtering out subordinate or superordinate categories. These rules then generalized poorly to the development set, with only 5

rules performing at that same standard, though another 3 rules were kept which had worked on the train set and which did not filter anything out in the development set.

We also considered rules that filtered out a small number of basic-level categories in the train set while also filtering out a large number of non-basic-level categories, but these made even more mistakes on the development set and the mistakes did not overlap well. As a result, we left the decisions with imperfect filters to the Voting Rule portion of the system.

The final Filtering Rules chosen are shown with their parameters in Table 5.

| |
|---|
| 1. Filter words with suffixes -ment or -age |
| 10. Filter words with more adjective than noun senses |
| 19. Filter capitalized words |
| 21. Filter synsets with hyponym depth $(\min + \max) / 2$ outside the range [1,3.5] |
| 23. Filter the top 6 levels of the hierarchy |
| 24. Filter nodes with 7 levels below them |
| 36. Filter all synsets with 65 or more siblings having no hyponyms |
| 37. Filter all synsets with at least 92% percent of siblings having no hyponyms |

Table 5: Chosen Filtering Rules with Parameters

5.2 Voting-Rule Experiments

The Voting Rules (Table 3) are applied to categories not already filtered by Filtering Rules. These are applied along each chain from the bottom to the top of the hypernym hierarchy. Like Filtering Rules, these rules are also applied to a category although evaluated in the context of a chain.

Using a greedy search starting with the most accurate Voting Rules, we identified a set of the rules which together enabled high accuracy on the development set. This combination is listed in Table 6.

| |
|--|
| 40. Top frequency in the chain (sum of lemma frequencies in synset) |
| 47. The synset has hyponyms |
| 49. Highest frequency in Brown + Gutenberg corpora combined in the chain |
| 51. Maximum number of meronyms in the chain |

Table 6: Selected Voting Rules

We determined that by using these rules together, and only selecting categories with three of these Voting Rules being fulfilled, high accuracy

could be obtained on the development set. This does limit the number of basic-level categories that can be selected to one in each chain from the bottom to the top of the hypernym hierarchy. However, with three of the four rules only being fulfilled for one node in the chain, it is possible not to select a basic-level category in some chains.

6 Evaluation

Our system’s overall performance on the test data is listed in Table 7.

| | Accuracy |
|---------------|------------|
| Superordinate | 100% |
| Basic-level | 84% |
| Subordinate | 44% |
| Overall | 77% |

Table 7: System Effectiveness

Accuracy is measured as the percentage of categories filtered (or not filtered) correctly based on the test data. Our system did well at filtering out superordinates, made a moderate number of mistakes filtering out basic-level categories, and was least successful at filtering out subordinates.

Just as when tuning the Filtering Rules on the development set, there was a substantial degradation in performance when extrapolating to the test set. Results on the development set, including subsystem breakdowns, are shown in Table 8.

The Filtering Rules provide the most substantial portion of the impact as measured on the development set with 77% accuracy, while the Voting Rules improved accuracy by 17 points to 94%. Comparing this to the results on the test data from Table 7, though, the system only performed as well as the Filtering Rules component did on its own on the development set. The generalization to the test data was better than expected for superordinates, worse than expected for the basic-level, and substantially worse than expected for subordinates.

| | Filtering Rules | Filtering + Voting Rules |
|---------------|-----------------|--------------------------|
| Superordinate | 62% | 88% |
| Basic-level | 100% | 92% |
| Subordinate | 68% | 100% |
| Total | 77% | 94% |

Table 8: Accuracy on Development Set by Subsystem

The labels used here are sparse and non-random, collected from psychological research papers which had experimental reasons to control the ratios of subordinates to basic-level and superordinate categories. As an additional measure of system performance, we manually annotated a random set of 250 categories predicted to be at the basic-level by our system to estimate the precision of the system. Unfortunately, the estimated precision was only 10.4% (26 of 250). This annotation was done by two annotators using as a standard the property from Rosch et al. (1976) that basic-level categories are the most inclusive level at which a concrete picture of the category as a whole can be formed (previously mentioned in Section 2.1). This was chosen because it was a simple mental test to perform unlike many of the other properties and it was a pattern observed in all of the basic-level categories from that experiment (Rosch et al. 1976). That same experiment was one of the underlying sources of our labeled data. The inter-annotator agreement is 92% and the kappa score is 59%. Disputes were resolved through discussion.

Our system predicts 13,082 synsets are basic-level. Using our accuracy on the basic-level as a measure of recall, combined with our estimated precision, we estimate that there are around a total of 1,620 basic-level categories in PWN. This is a quantity we have not previously seen estimated.

Judging by the examples predicted as basic-level in our estimate of precision, there are some systematic errors in cases where the system predicts a category is basic-level but it turns out not to be. The most interesting type of mistake accounted for just over half of the mistakes. In this case, the categories were not basic-level but also were not clearly superordinates or subordinates either, at least as described in the psychology literature. In the psychology experiments, the focus is primarily on physical objects and organisms. Rather, there were many examples where the word was a noun describing an action (e.g. violence), denoting a relation (e.g. proportion), or denoting a role in a more complicated semantic frame (e.g. defalcation). It is possible we are being too restrictive in our labeling, but it appears to us that there are many nominal categories which describe things that belong to classifications other than superordinate, basic-level, and subordinate. This suggests the low precision is not just due to the prevalence of subordinates relative to basic-level categories (which is an issue). In addition, though,

much of the imprecision may be due to phenomena outside our limited theoretical label space.

We are making our list of predicted basic-level categories available for download at http://e22pii.com/research/files/GWC2018/predicted_basic_level_categories_synsets.txt. The labels we used, mapping labeled words in the psychology literature to PWN synsets, is also available at <http://e22pii.com/research/files/GWC2018/labels.txt>

7 Conclusion

We built a rule-based system to automatically identify basic-level categories using PWN. We were effective at including most basic-level categories and excluding superordinates, but not as effective at excluding subordinates.

We were 77% accurate overall at classifying our limited test data derived from psychological experiments. However, we have evidence that suggests these labels are based on a simplistic view that divides categories into 3 groups which do not appear to cover the full range of phenomena described by nouns. Outside of this limited test data, we manually annotated a sample of our system's predicted basic-level categories and found low precision with the majority of the mistakes outside of these three groups. This suggests that for greater broad-coverage accuracy it may be necessary to model cross-part-of-speech relationships and other phenomena that do not fit nicely in the existing label space.

In the future, we hope to refine and scale a labeling process using mechanical turk to build a larger and less-biased training set. We hope to rely on several of the tests in the psychology research, although modeling additional phenomena may require extending these tests. Additionally, we hope to build a machine learning-based system to turn the many weak rules we have into features that can help improve system performance, as well as to evaluate the system on a much larger test set with this rule-based system as a baseline.

References

- Roger Brown (1958). "How shall a thing be called?" *Psychological review* **65**(1): 14.
- Edward William Dolch (1948). *Problems in reading*, Garrard Press.
- D. Dorr B. Habash N. Helmreich S. Hovy E. Green R. Levin L. Miller K. Mitamura T. Rambow O. Farwell (2009). "Interlingual annotation of multilingual text

- corpora and FrameNet." *TRENDS IN LINGUISTICS STUDIES AND MONOGRAPHS* **200**: 287-318.
- Christiane Fellbaum (1998). *WordNet : an electronic lexical database*. Cambridge, Mass, MIT Press.
- N Francis and H Kucera (1964). "Manual of Information to Accompany A Standard Corpus of Present-Day Edited American English, for use with Digital Computers (Department of Linguistics, Brown University, Providence, USA)." *icame. uib. no/brown/bcm. html* (accessed 12 October 2010).
- Roberta Michnick Golinkoff, Margaret Shuff-Bailey, Raquel Olguin and Wenjun Ruan (1995). "Young children extend novel words at the basic level: Evidence for the principle of categorical scope." *Developmental Psychology* **31**(3): 494.
- Rebecca Green (2006). "Vocabulary alignment via basic level concepts. Final Report 2003 OCLC/ALISE Library and Information Science Research Grant Project." *Dublin, OH: OCLC Online Computer Library, Inc. Retrieved May 14*: 2008.
- Project Gutenberg (n.d.). from <http://www.gutenberg.org>.
- Rubén Izquierdo, Armando Suárez and German Rigau (2007). *Exploring the automatic selection of basic level concepts*. Proceedings of RANLP, Citeseer.
- Pierre Jolicoeur, Mark A Gluck and Stephen M Kosslyn (1984). "Pictures and names: Making the connection." *Cognitive psychology* **16**(2): 243-275.
- María K Jónsdóttir and Randi C Martin (1996). "Superordinate vs basic level knowledge in aphasia: A case study." *Journal of Neurolinguistics* **9**(4): 261-287.
- Shari Landes, Claudia Leacock and Randee I Tengi (1998). "Building semantic concordances." *WordNet: an electronic lexical database* **199**(216): 199-216.
- Steve Legrand (2006). "Word Sense Disambiguation with Basic-Level Categories." *Advances in Natural Language Processing. Ed. Alexander Gelbukh, Research in Computing Science* **18**: 71-82.
- Guillaume Lemaitre and Laurie M Heller (2013). "Evidence for a basic level in a taxonomy of everyday action sounds." *Experimental brain research* **226**(2): 253-264.
- Shu-Yen Lin, Cheng-Chao Su, Yu-Da Lai, Li-Chin Yang and Shu-Kai Hsieh (2009). "Assessing text readability using hierarchical lexical relations retrieved from WordNet." *Computational Linguistics and Chinese Language Processing* **14**(1): 45-84.
- Edward Loper and Steven Bird (2002). *NLTK: The natural language toolkit*. Proceedings of the ACL-02 Workshop on Effective tools and methodologies for teaching natural language processing and computational linguistics-Volume 1, Association for Computational Linguistics.
- Brian MacWhinney (2000). *The CHILDES project: The database*, Psychology Press.
- Arthur B Markman and Edward J Wisniewski (1997). "Similar and different: The differentiation of basic-level categories." *Journal of Experimental Psychology: Learning, Memory, and Cognition* **23**(1): 54.
- George A. Miller (1995). "WordNet: A Lexical Database for English." *Communications of the ACM*. **38**(11): 39.
- Gregory L Murphy and Edward J Wisniewski (1989). "Categorizing objects in isolation and in scenes: What a superordinate is good for." *Journal of Experimental Psychology: Learning, Memory, and Cognition* **15**(4): 572-586.
- Ayn Rand (1966). "Introduction to Objectivist Epistemology (II)." *The Objectivist*(08/66).
- Abebe Rorissa and Hemalata Iyer (2008). "Theories of cognition and image categorization: What category labels reveal about basic level theory." *Journal of the American Society for Information Science and Technology* **59**(9): 1383-1392.
- Eleanor Rosch, Carolyn B Mervis, Wayne D Gray, David M Johnson and Penny Boyes-Braem (1976). "Basic objects in natural categories." *Cognitive psychology* **8**(3): 382-439.
- James W Tanaka and Marjorie Taylor (1991). "Object categories and expertise: Is the basic level in the eye of the beholder?" *Cognitive psychology* **23**(3): 457-482.
- Zhongyuan Wang, Haixun Wang, Ji-Rong Wen and Yanghua Xiao (2015). *An Inference Approach to Basic Level of Categorization*. Proceedings of the 24th ACM International on Conference on Information and Knowledge Management, ACM.
- Robert L Weide (1998). "The CMU pronouncing dictionary." URL: <http://www.speech.cs.cmu.edu/cgibin/cmudict>.