The Semantic Pictionary Project

Brent Kievit-Kylar (bkievitk@indiana.edu)
Cognitive Science Program
Indiana University, Bloomington, Indiana USA

Michael N. Jones (jonesmn@indiana.edu)
Department of Psychological and Brain Sciences
Indiana University, Bloomington, Indiana USA

Abstract

Here we describe the Semantic Pictionary Project—a set of online games and tools designed to collect large amounts of structured data about the object characteristics and perceptual properties of word referents. The project hinges on the use of encoding-decoding games and a set of creation tools to capture data using online crowdsourcing. We describe the architecture of the basic tools behind the games, the structure of the resulting data, and how this information may be integrated into existing statistical semantic models. We also describe two validations using data collected from one of the tools (2D Geon Pictionary) demonstrating typicality effects in the metrics of raw Geon objects created by subjects, and unique variance in the predictions of word pair metrics over currently used linguistic and property data.

Keywords: Geon; natural language processing; crowdsourcing; semantic space models; embodied cognition.

Introduction

Humans learn about the meanings of words and larger discourse units from repeated experience with both linguistic and perceptual information. However, current models of lexical semantics focus only on learning from linguistic structure using statistical abstraction algorithms. Part of the problem is the lack of realistic structured data containing information about the perceptual structure of word referents. Text is plentiful, but usable object structure data currently are not. Given the ideological movement towards models of embodied lexical representation, model development is currently being held up due to a lack of structured human data containing configural object and property information about concrete word referents. Here, we describe the NSF-funded *Semantic Pictionary Project*, an online approach to the problem of data capture that makes use of the paradigms of crowdsourcing and online gaming to gather data containing the perceptual structure of word referents. The online games may be played at www.SemanticPictionary.org.

Statistical semantic models (SSMs; e.g., Landauer & Dumais, 1997) have recently been attacked as implausible cognitive models because they learn from only linguistic information and are not grounded in perception and action, contrary to the literature on embodied cognition, and this limits their ability to account for human behavior on semantic tasks (for a review of the debate, see de Vega, Graesser, & Glenberg, 2008). The inadequacy of SSMs as cognitive models punctuates the current movement in cognitive science towards models of embodied cognition. There is a growing body of both behavioral and neuroimaging research demonstrating that when humans process words (in isolation or in context) they automatically activate sensorimotor information about the perceptual features of the word's referent, how it is commonly used, and physical contexts in which it has been experienced (for a review, see Riordan & Jones, 2010). A large number of behavioral experiments also demonstrate convincing evidence that sensorimotor experience becomes an inseparable part of a word's lexical representation, including information about object features (color, shape, motion, etc.). Perceptual information is an inherent part of the semantic organization of the human lexicon, but much of this information cannot be learned from statistics in a text corpus—it must be learned from multisensory experience.

Perceptually grounded SSMs are now emerging in the cognitive science literature (e.g., Andrews, Vigliocco, & Vinson, 2009; Recchia & Jones, 2010; Steyvers, 2010). As a proxy for sensorimotor perception, these new integrative models use norms of human-generated properties (e.g., McRae et al., 2005). These norms are collected by asking hundreds of subjects to produce the physical properties (internal and external parts), appearance, sounds, smells, tastes, functional properties, categorical membership, etc. for concrete nouns and event verbs based on multisensory experience. A property vector for a word is then created by aggregating across subjects. For example, the property <has_4_legs> will have a high probability for *dog* and *cow*, but a low probability for *centipede*, and a zero probability for *strawberry*. However, <is_red> is a highly salient property of *strawberry* and not for *dog*.

The development of perceptually grounded SSMs is currently being held up by a lack of data. The overall goal of the Semantic Pictionary Project is to collect large amounts of object and property data online using a combination of crowdsourcing and our new encoding-decoding games, and to make the large datasets available to researchers to develop superior grounded semantic models.

The Semantic Pictionary Paradigm

The Semantic Pictionary paradigm is a two-stage task with self-correction built in. In the first stage, subjects are presented with a target word selected from the high-concreteness/early-AoA nouns from the MRC Psycholinguistic Database. Subjects are then provided a tool to make a representation of this noun, with the goal of making a depiction such that another subject could guess what word is being represented (similar to the popular Pictionary game). That representation is then given to a different subject to attempt to recover the initial label. Success at recovering the initial label is in indication of a valid and meaningful encoding by the first subject. The data created by this paradigm can then be used in various modeling applications.

Words can be encoded in a variety of ways. The goal is to build a symbolic representation of the word's referent in a constrained domain. Example domains could be other words, physical shape, smell, or sound. In each domain, a specially constructed tool is used by subjects to generate the feature set representing that word, and the feature set can then be given to a different set of subjects to verify.

Crowdsourcing

Crowdsourcing is a paradigm that has recently emerged to use aggregate groups of humans to solve problems online that are impossible for computers to currently solve. Crowdsourcing takes advantage of crowd wisdom (Surowiecki, 2004) to capture data that only humans can currently produce at a massive scale. GWAPs ("games with a purpose") take advantage of crowdsourcing and the amount of human computation currently wasted on online games to capture data for practical purposes. For example, humans spend approximately 10 billion hours each year playing solitaire online. Facebook's Farmville game allows users to grow virtual crops and form social relations with other players online—Farmville sees about 68 million users each day. GWAPs harness the power of human computation for data labeling using an entertaining game. The original GWAP was called the "ESP Game" (von Ahn, 2006; now the "Google Image Labeler"). The ESP game used online human computation to solve the problem of labeling images and image components on the web. All of the Semantic Pictionary GWAPs are available online and are also linked to social media sites such as Facebook to collect massive amounts of structural data with crowdsourcing.

Semantic Pictionary Games

There are two broad classes of data representation games we employ: Property Pictionary and Geon Pictionary. Property Pictionary is a class of games in which the subject encodes the target as a set of constrained verbal features that describe it. Property Pictionary can be thought of as an online crowdsourcing version of McRae et al.'s (2005) feature generation task (originally collected in the laboratory), with the addition of a decoding phase in which different subjects attempt to guess the target word given a generated feature vector. Geon Pictionary is purely nonverbal. When presented with a target word to encode, the subject uses an editor to create an object model the referent using a constrained set of Geons (Biederman, 1987) in either two- or three-dimensional space. The Geon object constructed from a target word is then provided to different subjects to guess the target word given the image.

The two classes of games were selected to compliment information learned well by corpus-based models. Geon Pictionary collects information about object structure, color, viewpoint variance, component connectivity, etc. that do not seem to be possible to mine from language (see Riordan & Jones, 2011). Property Pictionary capitalizes on verbal feature generation to produce high-level descriptions of words including physical properties (internal and external parts), appearance, sounds, smells, tastes, functional properties, categorical membership, affordances, etc. not captured by the low-level Geon shape descriptions or the distributional structure of natural language.

In the following sections, we briefly describe the architectures of each of these GWAP tools as well as the structure of the data they collect and how it may be integrated into existing statistical semantic models. Then we turn to an analysis of data collected through the 2D Geon Pictionary game. We demonstrate that the information captured by the game can reproduce standard semantic typicality effects, and contains unique variance in semantic similarity used by humans but that is currently missing from linguistic structure and verbal properties.

Property Pictionary

Property production norms have proven extremely valuable in a variety of semantic experiments, and in cognitive models of semantic representation and processing. However, these databases are currently limited to a few hundred concepts. By taking lessons from McRae et al.'s (2005) original study and string normalizations, a crowdsourcing GWAP can potentially produce a database like McRae's spanning thousands of words in 1-2 years. In addition, we will have "goodness of transmission" values for features from the encoding/decoding paradigm that were not possible in McRae's original lab-collected database. Verbally coded features contain perception and action information at a higher level than the Geon shape description, and both are needed to evaluate perceptual integration in SSMs.

In Property Pictionary, subjects are assigned to be encoders or decoders. In the encoding phase, the subject is presented with a target word (e.g., DOG) and is asked to generate N descriptive features such that a decoder could guess the target word from the features. Subjects gain points as encoders the more people who can correctly guess the target word from their feature encoding. When a certain number of words have been encoded, subjects then progress to the decoding phase, guessing the target word that is represented by the feature encoding produced by another encoder for a different word. In this fashion, we are able to quantify the diagnosticity of produced features. An encoding of DOG = [+has_wings, +is_made_of_metal, ...] will not only be very infrequent, but will also have a very low probability of anyone else guessing the target word given this encoded pattern.

We have conducted pilot tests of Property Pictionary using both traditional psychology subject pool players, and using subjects via Amazon's Mechanical Turk crowdsourcing site. In the pilot collection phase, we used the same concrete nouns in McRae et al.'s (2005) original laboratory study, and built interactive checks at the input phase that originally had taken a significant portion of time to manually recode after data collection in McRae et al.'s norms. For example, if a subject typed in "has four legs" or "is four legged" the input system would remap in real-time by suggesting the equivalent recoded label, e.g., "do you mean <has_4_legs> ?"

The details of the Property Pictionary pilot norms have been described elsewhere (Recchia & Jones, 2011), so we will just briefly summarize here. The online Property Pictionary version of the feature norms was remarkably well correlated with the original McRae et al. (2005) laboratory-collected norms. The verbal features generated by our online subjects were very similar to those generated by McRae et al.'s subjects in the laboratory setting. The correlations between words' feature vectors in the online version and McRae et al.'s original database produced a mean correlation .83 (SD = .08). In addition, the semantic similarity among words in each of the norms were highly correlated. If one creates a word-by-word correlation matrix within each of the norms and then computes the correlations between rows of the two matrices, the mean correlation is .96 (SD = .03). The online version of the norms also had high similarities to other production characteristics of the original norms; e.g., # of features, # of distinguishing features, # of visual-motor/forms, # of tactile features, etc. The remainder of this paper will focus on validating the Geon Pictionary data.

Geon Pictionary

Geon Pictionary games require the subject to produce an object image representing the referent of the target word using a constrained set of components and attachments. If given DOG as a target word, the subject essentially draws a picture of a dog by selecting from a set of primitive geons, adjusting shape, color, orientation, size, and attachment structure of the components using our geon editor. This image is then provided to a second subject to guess at what target word the encoder is representing with the image. The system is designed such that we maximize the potential representable objects while at the same time having a compact and constrained enough description that meaningful comparisons can be made between objects. The data structure of the resulting image is stored as a tree-based representation of object configurations and properties, and we have several similarity algorithms available to determine the similarity among geon objects. The tree-based object is represented as a phrase structure grammar, so the visual object may be recoded to a text-based sentence. This allows corpus-based models to integrate the statistical information from the visual object while bypassing the problem of providing the models with vision.

The Semantic Pictionary website has two- and three-dimensional versions of the Geon Pictionary game. The three-dimensional version is necessarily tree attachment based (to preserve the object structure as viewpoint is rotated). The two-dimensional version has two versions. The tree-based version requires that geons be attached to one another at specified attachment points to construct a hierarchical tree-based representation and phrase grammar. The 'no-tree' version is unconstrained with regards to attachment points between geons—this allows much faster production of images, but object similarities are reliant on vector superposition since the object representation is flat rather than hierarchically structured. We next describe the three-dimensional version of Geon Pictionary in detail, and then the restricted two-dimensional versions more briefly.

Three-Dimensional Geon Pictionary

A final object generated with the 3D Geon Pictionary tool is a tree of Geon objects with properties and their connection or attachment constraints. Each Geon has the following properties:

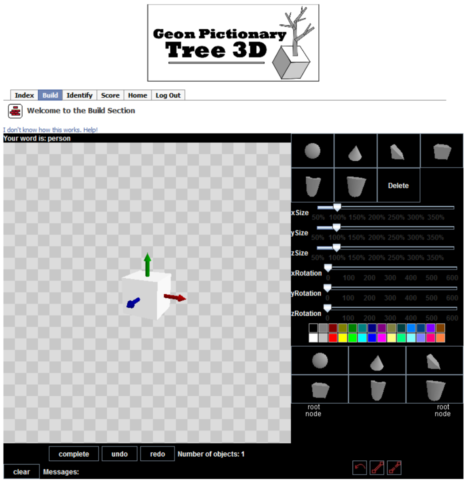
Geon Type: Chosen from (cube, sphere, cyliner, cone, handle).

Size: Scaling in the X, Y and Z axis in set increments (from 50% to 350% of one unit in steps of 1%).

Rotation: Rotation around the X, Y and then Z planes in set increments (from 0 to 6 radians in steps of .01 radians)

Color: Chosen from a reduced color set (original MSPaint)

The first Geon is set as the root of the model and each Geon added is then attached to the root at specified attachment points. An attachment point is defined as a pair of points, one defining the location on the parent and one the location on the child. The child is then moved such that these two attachment points are aligned in the same three-dimensional point. The points defining potential attachments are the 27 points formed by a bounding rectangle around the Geon (3 potential values for the X, Y and Z axis).



Welcome to the Build Section

Your word is: person

Number of objects: 1

Messages:

Figure 1. The build interface of the 3D Geon tool.

The basic web interface is shown in Figure 1. The 3D rendering is done with an in-lab developed 3D rendering tool so no external libraries are required (this greatly facilitates web distribution). The model can be rotated by dragging the mouse in the X or Y plane. Particular Geons can be selected by clicking on them and deselected by a second click. The rendering window is also used in moving Geons to new parent nodes. Controls allow the subject to manipulate the color, scaling, rotation, and geon type of a selected object or group of objects.

Geons may be added to existing Geons (at default connections points). Connected Geons can also manage their attachment points. Since each Geon (except the root) has exactly one parent, but may have multiple children, we decided to show and let the subjects manipulate the connection between the selected node and its parent. Selecting appropriate attachment points is the most difficult task for subjects and to facilitate this we have provided two manipulation techniques. The connection can be modified by use of either radio buttons representing the connection in the X, Y and Z axis, or by selecting the point by clicking on the appropriate location on a 3D wireframe model of a cube that shares its orientation with the Geon on the rendering screen. Figure 2 shows a rendered object and the attachment point cubes.

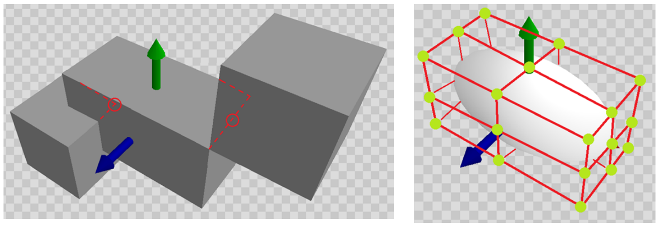


Figure 2. A rendered object and attachment point cubes.

The web environment is written in php and provides three paths of interface, Amazon's Mechanical Turk, Facebook application, and our own web domain www.SemanticPictionary.org. The php code provides a login system to manage users, a set of written and video instructions on how to use the Geon tools, the actual building interface, the identification interface and a scoring system.

Two-Dimensional Geon Pictionary

The two dimensional version of Geon Pictionary tool is designed to have a very similar look and feel to its three dimensional equivalent. There are two primary differences between these two tools. The first is that whereas the three dimensional system only allows subjects to select stepped values for rotation and scale, the two dimensional system allows arbitrary values. The connection interface for the two dimensional system also allows subjects to click and drag Geons (and all descendents) and will automatically select the attachment point that would most closely represent the released location, which makes object production much faster. In addition to the tree-based version of the 2D game, there is also a freeform no-tree version. In this version, the tree structure requirement has been removed. Primitive instances can be added to the scene in arbitrary locations. While this version makes it faster for the subject to produce an object, the resulting data structure is flat and requires different types of similarity algorithms to analyze.

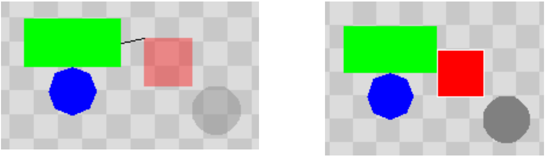
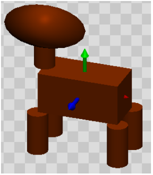


Figure 3. 2D Geon Pictionary click-and-drag interface.

Encoding, Decoding, and Extrapolation

For storage and transfer, models are encoded in a simple shorthand coding. In this coding, key symbols are used to describe the tree structure and properties of instances. This encoding is short and easy to transfer between systems. The model is decomposed into a text-based encoding through a simple rule set. Each primitive instance can be decomposed into a sentence unit of the following form:

[SENTENCE] = An [OBJECT] made up of a [ROOT OBJECT DESCRIPTION] [OBJECT DESCRIPTION] = [COLOR], [SCALE], [GEON] rotated [ROTATION] whose [CHILD ATTACHMENT POINT] is attached to the [PARENT ATTACHMENT POINT] of a [CHILD OBJECT DESCRIPTION] and a [CHILD OBJECT DESCRIPTION] ...



A horse is made up of a brown, wide, Cube whose left, top, front, is attached to the top, of a brown, narrow, shallow, Cylinder whose right, top, front, is attached to the top, of a brown, narrow, shallow, Cylinder whose left, top, back, is attached to the top, of a brown, narrow, shallow, Cylinder whose right, top, back, is attached to the top, of a brown, narrow, shallow, Cylinder whose right, bottom, is attached to the bottom, of a brown, narrow, shallow, Cylinder whose center, is attached to the bottom, of a brown, wide, Sphere

Figure 4. An example of a horse model converted to natural text description.

Vector Encoding of Object Structure

Though the models can be decomposed into natural language and read by any natural language engine, purpose built translators for particular NLP models are likely to improve performance. We next describe an encoding algorithm for the BEAGLE semantic model (Jones & Mewhort, 2007) to make use of the Geon models directly.

BEAGLE uses a set of two holographic vectors to represent each word in a language. The first vector is the environmental vector; this is the static representation of the word in the universe (sampled from a Gaussian distribution). The second vector is the lexical vector, which stores the relational information learned by the system through interaction with the corpus. After learning, word relations can be extracted though holographic operations on sets of vectors such as cosine for similarity. Algorithms may then be applied to convert a model into a single holographic vector usable by the BEAGLE model.

Each property value (color, scale, Geon type, rotation, attachment points) is assigned a randomly generated permutation of dimension equivalent to the language model. After these are assigned, they will remain constant throughout all encodings. A primitive instance is then encoded as the point-wise sum of the relevant property vectors. Property vectors will be calculated in two different ways depending on whether the property is continuous or drawn from a small set. For those drawn from a small set such as Geon type, attachment point and possibly color, the property vector will simply be the natural language environmental vector for that word. This is useful since those environmental vectors will already have relational meaning from previous or post experience with supplemental corpora (we would expect most of the color names and many of the shape names to occur in common English text).

Those values from continuous sets such as scale and rotation can be encoded with frequency-encoded vectors where vector values are chosen from a distribution reflecting the value of the property (higher values for example, may shift a distribution). A model vector can then be calculated from the vectors for each of its primitive instances. To do this, each child is permuted by a static random permutation and then added point-wise to its parent representation.

Information Structure in 2D Geon Pictionary

We assess the structure contained in 2D geon representations constructed by groups of subjects in two tasks. In the first task, subjects were asked to generate geon representations of the concrete nouns from Rosch’s (1975) study of semantic typicality. In studies of typicality effects, stimuli are normally words. Here, we evaluate the structure of the geon representations of those words using the above described similarity algorithm applied to the geons. In the second task, we had subjects produce geon representations for words from the original McRae et al. (2005) norms, and we assess the information contained in the geon representations of those words compared to the McRae et al. feature vectors and a corpus-based co-occurrence metric.

Semantic Typicality Effects

Figure 5 shows the similarity structure among words from Rosch’s (1975) high, medium, and low typicality conditions. In verification experiments, subjects are typically faster to verify that two high typicality exemplars are members of the same category (e.g., *robin-sparrow*) than medium (*hawk-chicken*) or low (*penguin-ostrich*) typicality exemplars. To compute similarity between geons in Figure 5, each possible color, shape, rotation and scale is assigned a random Gaussian vector (these values could be taken from a learned training run on a corpus). The vector representation for a given geon is simply the sum of the part vectors. The tree's holographic vector is then the root instances holographic vector added to the holographic vector of its children with a present random permutation added at each level. As is shown in Figure 5, members of a semantic category that are rated as being more typical exemplars tend to look more like one another in their geon encodings as well. This effect is stable over all typicality bins (right panel), but also the individual categories (left panel). These results suggest that at least part of typicality structure can be encoded in how subjects describe word referents using our Geons, and this information would be represented in our natural language or vector representations as unique variance to be used to enrich statistical semantic models that typically only have linguistic structure from which to make inferences.

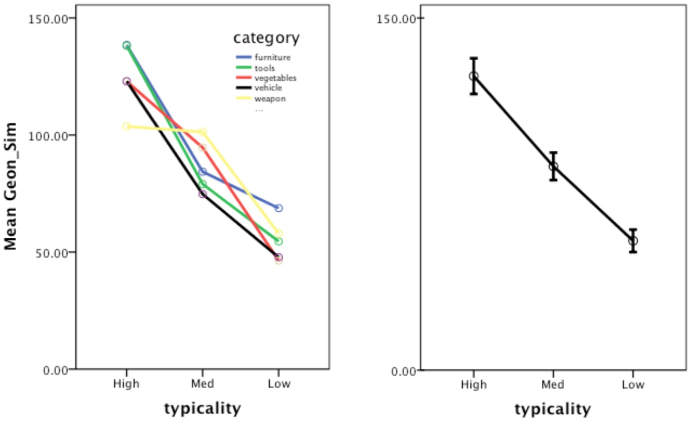


Figure 5. Typicality effects in 2D Geon Pictionary

Predicting Word Pair Similarities

Using the same similarity metric applied to the 2D geon representations, we computed the pairwise similarities between words from the McRae et al. (2005) norms (different group of subjects than produced the typicality data). These pairwise similarities were then entered into a hierarchical regression to predict similarities between words in WordNet using the JCN metric; JCN has been shown to give the best approximation to human judgments of semantic similarity between words (Jones & Mewhort, 2007). Included in the regression was cosine similarity from the McRae norms and pointwise mutual information (PMI) between the word pair in the TASA corpus.

Table 1. Hierarchical regression predicting WordNet pairs.

Model	R	F Δ	Partial R
PMI	.158	377.11	.158
PMI + McRae	.344	1556.27	.096, .305
PMI + McRae + Geon	.358	144.42	.096, .302, .084

all $p < .001$.

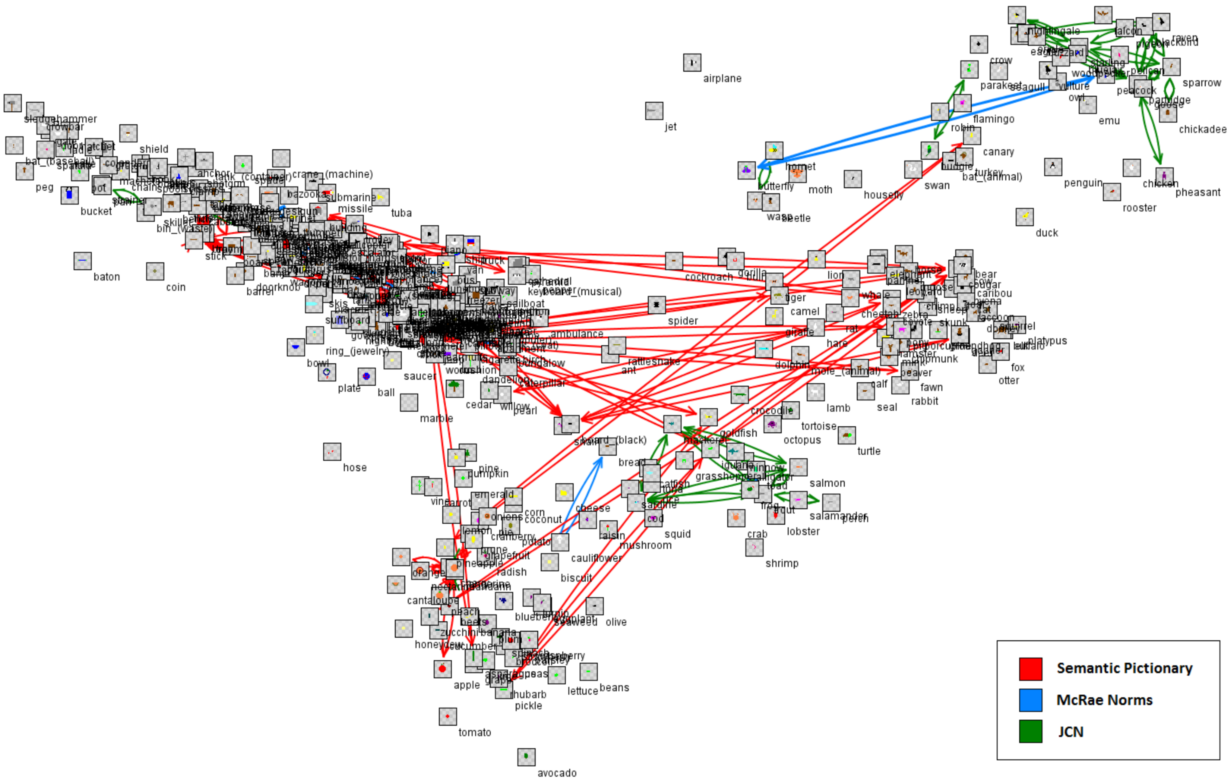


Figure 6. Similarity structure that is unique to the text, verbal feature, and geon object representations of words.

As Table 1 shows, there is a considerable amount of redundancy in the three variables when predicting variance in WordNet similarities. However, each also contains a significant portion of unique variance not accounted for by the others. We entered geon similarity to the regression equation as the last step to stack chance against it. However, as Table 1 demonstrates, similarity between the geon representations of the words predicts a significant portion of variance that is not accounted for by the text-based or verbal feature measures. Figure 6 shows this structure more clearly. The MDS plot is arranged so that proximities are based on similarity from the McRae et al. (2005) norms. The red lines show strong connections between items found by their geon similarity that are not seen by the other metrics. Qualitatively, this includes a considerable amount of shape structure (e.g., the similarity between *pizza* and *coin*), color (*pickle-grasshopper*) material (green plants and wood/metal), symmetry/asymmetry, internal consistency, etc. This information is important to human semantic organization, but is neither learned by the text-based models nor is it well represented in standard verbal feature generation norms.

Acknowledgements

This research was supported by grants from Google Research and NSF BCS-1056744 to MNJ. We would like to thank Ken McRae and Gabe Recchia for their guidance.

References

Andrews, M., Vigliocco, G., & Vinson, D. (2009). Integrating experiential and distributional data to learn semantic representations. *Psyc Review, 116,* 463-498.

Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review, 94,* 115-147.

deVega, M., Graesser, A., & Glenberg, A. (2008). *Symbols and Emboiment: Debates on Meaning and Cognition.* NY: Oxford Press.

Jones, M. N., & Mewhort, D. J. K. (2007). Representing word meaning and order information in a composite holographic lexicon. *Psychological Review, 114,* 1-37.

Jones, M. N., & Recchia, G. (2010). You can't wear a coat rack: A binding framework to avoid illusory feature migrations in perceptually grounded semantic models. *Proceedings of the 32nd Annual Cognitive Science Society.*

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

McRae, K., Cree, G. S., Seidenberg, M. S., & McNorgan, C. (2005). Semantic feature production norms for a large set of living and nonliving things. *Behavior Research Methods, Instruments, & Computers*, 37, 547-559.

Recchia, G., & Jones, M. N. (2011). Crowdsourcing large-scale semantic feature norms. Paper presented at the Midwest Cognitive Science Conference, MSU.

Riordan, B., & Jones, M. N. (2010). Redundancy in perceptual and linguistic experience: Comparing feature-based and distributional models of semantic representation. *Topics in Cognitive Science.*

Rosch, E. H. (1975). Cognitive representations of semantic categories. *Journal of Experimental Psychology: General, 104,* 192-233.

Steyvers, M. (2010). Combining feature norms and text data with topic models. *Acta Psychologica, 133,* 234-243.

Von Ahn, L. (2006). Games with a purpose. *Computer*.