Making Sense out of Food

Brent Kievit-Kylar (bkievitk@indiana.edu)
Cognitive Science Program, IU Bloomington, IN 47405 USA

Yong-Yeol Ahn (yyahn@indiana.edu)
School of Informatics and Computing, IU Bloomington, IN 47405 USA

Peter M. Todd (pmtodd@indiana.edu)
Cognitive Science Program, IU Bloomington, IN 47405 USA

Abstract

In this paper we explore the application of a novel data collection scheme for multi-sensory information to the question of whether different sensory domains tend to show similar relations between objects (along with some unique variance). Our analyses—hierarchical clustering, MDS mapping, and other comparisons between sensory domains—support the existence of common representational schemes for food items in the olfactory, taste, visual, and tactile domains. We further show that the similarity within different sensory domains is a predictor for Rosch (1975) typicality measures. We also use the relative importance of sensory domains to predict the overall similarity between pairs of words, and compare subjective similarities to objective similarities based on physical sensory properties of the foods, showing a reasonable match.

Keywords: *Multi-sensory; data collection; typicality.*

Introduction

While humans are primarily visual creatures (Barton 1998, 1995) we rely on all of our senses to function in the real world. If early humans judged whether food had gone bad only from sight without use of smell, they would have had a lower survival rate. The use of multisensory information is ingrained in our world representations so deeply that it is often encountered in pre-conscious tasks such as priming (Pecher 1998). But how much distinctive information do the different sensory domains provide about objects? Are exceptional objects in one sensory domain unexceptional in others, or do the different senses tend to provide largely overlapping information about objects? Addressing these questions and understanding the structure of multimodal sensory representations may provide critical insights for building better semantic space models, understanding language acquisition, and modeling memory phenomena including priming. Here we take an initial step by introducing a crowdsourcing framework for collecting multi-sensory object information, and ways of analyzing it.

In previous work, Kievit-Kylar & Jones (2011) showed that carefully collected visual information could be used as a successful predictor for people's judgments of overall similarity between objects, and that this predictor captured variance different from that supplied by semantic models based on text corpora analysis (e.g., Dumais et al, 1997, Jones et al 2006, Lund and Burgess 1996) and featural information (e.g., so-called McRae features that people generate to describe objects—McRae 2005). Similarly, multi-modal information from objective measures of the visual, gustatory, and olfactory modalities along with subjective semantic and featural representations has been shown to have significant cross-modal predictive power (Kievit-Kylar & Jones 2012a,b): Information about an object in one sensory modality can provide significant information on what that object's representation is in another modality. By combining information about an object across multiple modalities, the prediction of the unknown modality improves further.

Unfortunately, collecting objective similarity measures based on physical features in various sensory domains is a difficult and expensive task, requiring specialized equipment for smell, taste, and touch information. Also, the resulting measures computed by collecting this information do not necessarily reflect the same sort of information available to and used by humans when they make their own similarity judgments (e.g., due to nonlinearities of senses as well as potential mismatch between the features that can be detected by humans versus machines). Here we use a novel technique based on a fluency and grouping task to collect subjective similarity information across multiple sensory domains. This data is used to test the hypothesis that, overall, different sensory modalities tend to conserve the same similarity relations among a set of objects, coding overlapping information. At the same time, the unique variance contained in the details of those sensory modalities is critical to understanding the relationships of these objects.

To show this, we use cross-modal data we collected about different types of food. The category of food is useful for this exploration because foods are fundamental objects for humans, and people have rich multi-sensory conceptions of various foods in terms of modalities including visual, olfactory, taste, and tactile (we did not include aural). We then compare the subjective representations obtained from people between sensory domains as well as to existing objective data within domains (e.g., comparing how similar people judge the smell between two objects with how much their composition of volatile chemicals overlaps) to assess the extent of shared information across sensory domains for foods.

Procedures

This experiment was performed using the crowdsourcing platform Amazon Mechanical Turk. Turk users were led to a custom web page from which to perform the task:

http://www.indiana.edu/~semantic/fluency/fluency.html

After providing consent, participants were asked to perform a traditional verbal fluency task (Henley 1969) in which they were given two minutes to list as many foods as they could think of. Each word entered cleared from the screen to avoid cuing. Spelling was cleaned up in post-processing. In the next phase, participants were given a trial practice run with the word layout tool shown in Figure 1. The layout tool was seeded with words chosen to represent all of the sensory domains we used so as to not bias any in particular. The words used were: fragrant, woody, sweet, salty, rough, smooth, red, green. The order of these words was randomized for participants.

In the word layout tool, participants were allowed to move each word by clicking and dragging with the mouse. Words that were dragged close to each other were then considered grouped by the system and this was indicated by showing a connecting line. If items a and b were considered in the same group, they were connected with a line. Far a and b to be in the same group there has to exist a set of items c, containing at least a and b and such that every item in c is within a Euclidean length l of at least one other item in c (where l was given as a fraction of the screen width, or the square root of .03 times the width of the display area). Participants were asked to move the words around the display of the layout tool to indicate which words were similar to each other. Their goal was to place the words into groups to represent this similarity in the same fashion as multidimensional scaling (MDS—Kruskal & Wish, 1978). The practice phase lasted until all items had been moved, and the participant selected the "finish" button.

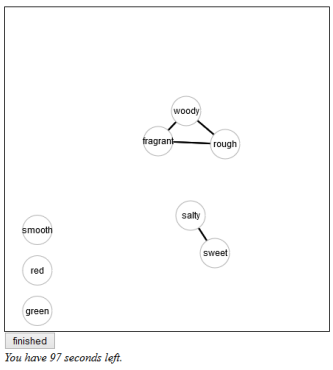

Figure 1: Example of clustering as seen by participants when using the word layout tool.

After the practice round, each participant was entered into three more sensory modality rounds. These rounds were selected by the program and assigned in a random order from the five categories "smell", "taste", "feel", "look", "overall" (where "overall" was intended to lead participants to judge the overall similarity between items). (We limited domains seen per participant to three to avoid fatigue.) The current sensory modality was indicated above the layout tool as shown in Figure 2 using an icon (nose, tongue, hand, eye, and blank), the relevant verb in large bold text, and a brief sentence priming that sensory domain, as follows:

***Smell**: Think about what it would be like to drive along a highway and smell the incredible odor of a skunk.*

***Taste**: Think about what it would be like to lick a penny.*

***Feel**: Think about what it would be like to rub sandpaper on your cheek.*

***Look**: Think about what it would be like to view a beautiful landscape.*

***Overall**: Think about all of the foods that you have eaten.*

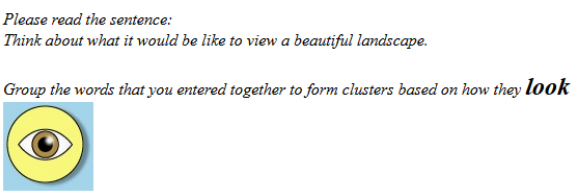

Figure 2: Example of sensory cue as seen by participants.

At the beginning of each layout tool round, the words given to the participants to sort were the foods that they themselves had entered during the initial verbal fluency segment of the experiment. These words were given lined up on the left side of the screen in the order in which they had been entered (similar to Figure 1). The participant then sorted the items according to their similarity in the indicated sensory modalities until each item had been moved at least once and the participant clicked the "finished" button. After the three sensory sorting rounds, the participant was given a completion code and reimbursed at standard MTurk payment rates. In total, 110 participants completed the task.

Hypothesis

We hypothesize that the similarity spaces generated by the participants will show common structure across the sensory modalities. That is, if two items are close in one sensory modality, they will tend to be close in another sensory modality. At the same time, we also hypothesize that there are important outliers that will provide more insights on the multi-sensory information integration.

Results

The resulting data set contained 8,609 food instances. A total of 475 word substitutions were generated to correct spelling and lemmatize the data. This left 736 unique words, 294 words of which occurred more than 2 times

Similarity

One of the primary goals of this data collection is to measure similarities between pairs of food items within each sensory domain. Getting a consensus similarity metric from the individual participant results required a technique for combining their data. First, we pre-processed all word sets to standardize case and remove words not used by at least three participants. The remaining words were then reviewed by hand to identify spelling mistakes and standardize language (e.g. normalize pluralization, compound vs. separate words, and overly specific identifiers).

We define the similarity of a given pair of food items as follows:

$$sim(f_1, f_2) = \frac{\sum_s joined(f_1, f_2, s)}{\sum_s used(f_1, f_2, s)}$$

where s is the set of all participants, joined(a,b,s) is 1 iff a and b were in the same group as defined by the participant s, and used(a,b,s) is 1 iff a and b were both entered by s.

This simple similarity measure represents the fraction of participants who had connected two words together over the number of participants who had used both words. While a measure based on the actual on-screen distance between pairs of items may have provided finer detail, the lines shown to participants (Fig. 1) indicated binary thresholds for similarities that may have been what was most important to them as they positioned items; the distances between groups or between individual items within a group may not have mattered much during their layout process. It is important to note that this is a ***similarity*** *metric, with higher values meaning the two words are considered more similar, as opposed to a distance metric where lower values (smaller distances) between words indicate they are more similar. Both metrics will be used in the following sections.*

One problem with this technique is that a single participant could greatly influence a similarity measure between uncommon words. To avoid this, a threshold of at least three participants all having used the same pair of words (regardless of whether they grouped them together or not) was required for the similarity measure, otherwise the measure was given a value of zero. Similarity between a word and itself was assigned the value 1.

Hierarchical Clustering

Hierarchical clustering was performed using the MultiDendrgrams software (Fernández & Gómez 2008) with unweighted average distance clustering. As tree graphs with hundreds of nodes are very difficult to read, only the most frequently used food items (by over 50 participants) are shown for each separate sensory modality (see http://www.indiana.edu/~semantic/fluency/imgs/dendrogram.png for image).

Results

Overall, each dendrogram displays a similar, intuitive pattern of clusters, as hypothesized. In each of the sensory modalities, fruits tend to be grouped, as do vegetables, meats, dairy, and baked goods. It is mainly in the details of the variations that the differences in the sensory modalities exert themselves, such as tomato moving in with the fruits in the smell category, or onion and potato following the round fruits in the visual sensory modality.

One way to see these differences is to follow a single food item through the different sensory modalities and compare its nearest neighbor in each—for instance, "french fry." In the visual domain, the fry is one of the immediate neighbors of potato chip, as both of these food items have similar colors and one dimension longer than the others. The smell domain also places the potato chip closest to the french fry presumably owing to the potato and oil scents. In the tactile (feel) modality, the greasiness of the fry pulls pizza in as the nearest neighbor. Under taste, subjects used knowledge that a fry is made out of potatoes to place the potato as the nearest neighbor. As an overall measure, the potato chip is again the closest neighbor to the french fry which may partially be due to the proximity of these two items within more sensory modalities, or may be due to the relative strengths of these modalities when defining this food pair. This is the first evidence for our main hypothesis.

MDS Layout

We use classical Multidimensional Scaling (MDS) to visualize the relationships between food items in 2D space. This is a mathematical version of what each participant manually and intuitively did with the layout tool for three sensory modalities. Here we combined all of the similarity information from participants in all five modalities using the similarity measure described above and performed MDS on the resulting averaged similarity spaces within each sensory domain for all of the food items.

Visualizations of this form and others have been shown to be extremely useful for understanding large data sets, and the Word2Word visualization engine (Kievit-Kylar & Jones 2012a) provides a comprehensive package for generating such visualizations over semantic or network types of data. Words or concepts are treated as nodes in this model, and similarities between words are represented as edges with thickness relative to the strength of the similarity. The resulting MDS layouts for each sensory domain are shown in Figure 3, where the overall similarity in shape of each is what should be focused on at this level (see also http://www.indiana.edu/~semantic/fluency/imgs/MDS.png).

Results

Similarly to the dendrograms from hierarchical clustering, the results of MDS show that these sensory modalities contain highly overlapping information (along with unique variance). The overall shape of the space within each of the sensory categories divides into fruits, vegetables, meats, and other foods. The dispersion of the

different domains is also an interesting phenomenon observable in these plots. While some domains, such as "overall," tend to have similarity relations that are well agreed upon across participants, and thus tighter clusters, visual or tactile information has many different types of features (within each sensory domain) that participants could be using to determine similarity, and the resulting MDS space is therefore more dispersed.

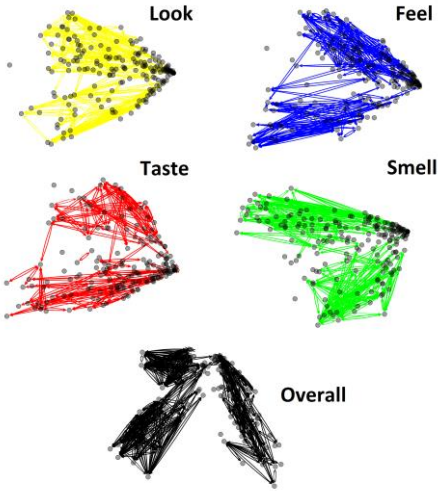

Figure 3: MDS layout on each sensory domain separately.

Comparing Sensory Domains

To determine more quantitatively how representations in one sensory modality match those in other modalities, we need a metric that can compute overall similarity between two similarity matrices. There are many techniques to do this (Procrustes analysis, sum-squared error of pairs, Pearson or Spearman correlation of rows, etc.). For the purposes of this paper, a simple sum-squared error of pairs will be used. This is justified insofar as the data generation techniques we use produce their own normalization (i.e., data are in common ranges). The precise definition of this similarity metric is as follows:

$$sim(m_1, m_2) = \sum_{w_1} \sum_{w_2} (sim(w_1, w_2, m_1) - sim(w_1, w_2, m_2))^c$$

where the two-argument sim function on the left is the similarity between two sensory modalities m_1 and m_2 and the three-argument sim function is the similarity between the first and the second word w_1 and w_2 within a given sensory modality m. The distances between the sensory modalities using this measure are shown in Table 1 and displayed as an MDS space in Figure 4. To perform MDS on this table, distance space must be converted into similarity space, by a simple power inversion such that each distance d was transformed into $(1\text{-}n)^c$ with the constant c set to spread the values into a reasonable range (here c=40).

Table 1: Similarity between sensory domains.

	Smell	*Taste*	*Feel*	*Look*
Smell	*1*	*.66*	*.59*	*.58*
Taste	*.66*	*1*	*.65*	*.64*
Feel	*.59*	*.65*	*1*	*.63*
Look	*.58*	*.64*	*.63*	*1*

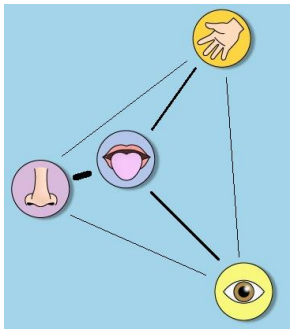

Figure 4: Visualization of similarity of sensory domains.

Results

As before, this analysis showed that the different sensory domains were highly interrelated. Taste and smell were the most strongly correlated sensory modalities, most likely due to the strong interconnection between these two senses when consuming food (Roach, 2013). The least related domains were visual and olfactory, reflecting their disparate bases. Visual information is effective at a greater range (finding an item) but is not necessarily linked directly to chemical composition (can consuming that item harm us), which can be much more easily detected through olfaction.

Rosch Typicality

Another use of the data described in this paper is to determine the relative importance of different sensory domains in human ratings of how typical different items are for their respective categories. Rosch (1975) collected a set of human generated typicality ratings for a number of different categories, including fruits and vegetables. Of the rated vegetables, 27 were matched with items our participants generated over 5 times, and 29 matches were found for the fruit category. To compare Rosch's ratings to our data, we can approximate a measure of typicality within each sensory modality by averaging the similarity of a particular item with all other items within that category. Thus the typicality of an apple as a fruit, within the taste domain, is the mean of our participants' taste-similarity ratings of apple with every other fruit they generated. Each word can then be plotted in terms of its Rosch typicality versus our participants' mean similarity for each sensory modality. This is shown for the smell modality in Figure 5.

This figure indicates that the olfactory domain is a moderate predictor of typicality of fruit (r=-.36) but not of

vegetables (r=-.19). For the latter, the tactile and visual sensory domains are better predictors (r=-.35 and -.33).

Item Level Domain Importance

The data set described in this paper gives a unique look at the similarity between sensory domains for a set of varied food items. We can also explore, on a per-item level, how items relate between domains. How does the similarity of a particular object to its neighbors in one sensory domain compare to those relationships in another domain?

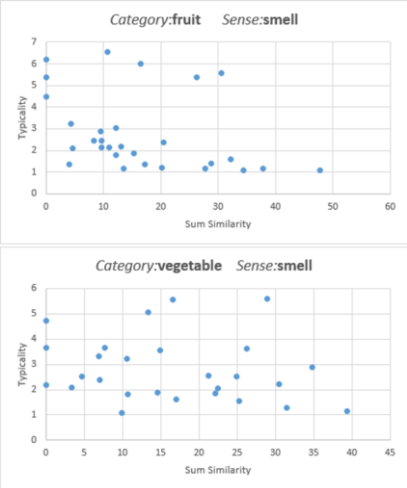

Figure 5: Rosch typicality versus mean similarity.

To plot relative similarity of a word in its use between domains, the "overall" domain was used as a base. For every word and every domain, we performed the following operation: The row representing the similarity of the object w to all other objects within the domain d was extracted. The row representing the similarity of the word w within the "overall" domain was also extracted. A sum-squared difference similarity metric was run to compare these two vectors, resulting in a single value for each word/domain pair, which we call the Domain Importance Value (DIV). Note that a high DIV means high sum-square error and thus indicates a large difference between the "overall" domain and the sensory domain in question (or a low importance for that sensory domain in the overall domain).

To visualize this information, an intensity value was chosen for each item such that lighter colors are used to represent items that are more dissimilar between the target domain and the "overall" domain. The points representing items were then displayed in an MDS distribution (over the "overall" similarity space) as shown in Figure 6 (see also http://www.indiana.edu/~semantic/fluency/imgs/ILDI.png for a larger version).

Results

Once again, there is a consistency across the four sensory domains, but this time in their inconsistencies with the "overall" similarity space. Darker points (more similar to the "overall" domain) tend to occur toward the center of the MDS layout, while the same foods (in different sensory domains) that are inconsistent with the "overall" domain (lighter points) are scattered around the periphery.

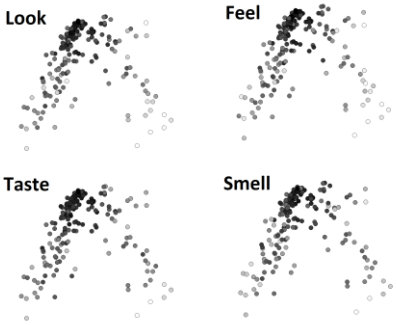

Figure 6: Items shown with intensity relative to their similarity of use between the "overall" domain and each individual sensory domain.

Comparison to Objective Feature Similarity

How well do the subjective similarity spaces generated by our participants in each sensory domain match those constructed from objective sensory features of the food items? We were only able to collect objective data about a sufficient number of the food items in our participant data for the sensory modalities of olfaction (from flavor compound information in Ahn et al., 2011) and vision (color and texture information in Meule & Blechert, 2012). Overlaps of words were found between the data sets, and similarity measures were computed between the participant generated similarities and the objective feature similarities.

Figure 7 shows the olfaction data with a visual representation of the Procrustes analysis between subjective and objective similarity spaces. Thus, MDS was used on both data sets independently and then optimal parameters were used to tune scale, rotation, and translation to align the two layouts. The participant similarities are shown with red connections and the objective similarities in green. Both sets of words are laid out independently with MDS and aligned optimally. Black lines connect the same word in one set with itself in another. Thus shorter lines mean better matches between the sets.

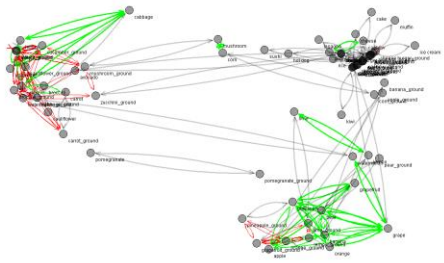

Figure 7: Procrustes visualization of subjective versus objective similarities for the olfactory domain.

While olfaction similarities had extremely high correlations, visual similarities matched significantly less well when full space matches like Procrustes were used, and the structure of the olfactory data caused the MDS process to break down (too sparse a similarity set).

On a per-item basis, the trend is far clearer. To compute this for a given item, the vector row of that item corresponding to the similarity of that item with every other item according to the participant measure was compared to the row of that item corresponding to the objective ground truth for that item. Because the two vectors of similarity values were generated in different ways (the former by the mean overlap similarity measure defined earlier, and the latter by cosine similarity), we used a Pearson similarity measure to compare them. The average Pearson similarity was then computed for each item within each sensory domain. The mean similarity per item between the subjective and objective similarity spaces was .51 for the olfactory domain and .5 for the visual domain, indicating a good match between subjective and objective similarity spaces for foods for these two senses.

Conclusions

Overall, these results support the hypothesis that the representation of a particular item in different sensory modalities will typically have strong overlap in terms of the similarities to other items. Moreover, subjective and objective similarity structures match up reasonably well (which also helps to validate the introduced data collection technique). Ongoing extensions include testing how well the cross-modal representation consistency can predict age of acquisition of different words—will those items that have consistent representations in different sensory domains be easier to learn by children?

Food items were used in this study because foods elicit sensory responses in many different modalities. However, by selecting this category, some modalities are increased in importance and others neglected (in particular sound was not considered a relevant sense to elicit in this domain). Food was also chosen due to the availability of rich objective information about items in various sensory domains (such as nutrition or olfactory information), which allows comparisons with base-line measures on the items. The experimental procedure we developed could however be used for any item category for which researchers can obtain a relevant set of objective sensory features.

References

Ahn, Y.-Y., Ahnert, S.E., Bagrow, J.P. & Barabási, A.-L. (2011). Flavor network and the principles of food pairing. Scientific Reports 1, 196. DOI:10.1038/srep00196.

Barton, R., Purvis, A., & Harvey, P. (1995). Evolutionary radiation of visual and olfactory brain systems in primates, bats and insectivores. Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences, 348(1326), 381-392.

Barton, R. A. (1998). Visual specialization and brain evolution in primates. Proceedings of the Royal Society of London. Series B: Biological Sciences, 265(1409), 1933-1937.

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

Fernández, A., & Gómez, S. (2008). Solving non-uniqueness in agglomerative hierarchical clustering using multidendrograms. Journal of Classification, 25(1), 43-65.

Henley, N. M. (1969). A psychological study of the semantics of animal terms. Journal of Verbal Learning and Verbal Behavior, 8(2), 176-184.

Jones M. N., Kintsch W., and Mewhort D. J. (2006). High-dimensional semantic space accounts of priming. Journal of memory and language, 55(4), 534-552

Kievit-Kylar, B., & Jones, M. N. (2011). The semantic Pictionary project. In L. Carlson, C. Holscher, & T. Shipley (Eds.), Proceedings of the 33rd Annual Conference of the Cognitive Science Society (pp. 2229-2234). Austin, TX: Cognitive Science Society.

Kievit-Kylar, B., & Jones, M. N. (2012a). Visualizing multiple word similarity measures. Behavior research methods, 44(3), 656-674.

Kievit-Kylar, B., & Jones, M. N. (2012b). Cross modal inference in distributional models of semantics. Paper presented at the 22nd Meeting of the Canadian Society for Brain, Behavior, & Cognitive Science, Kingston, ON.

Kruskal, J.B., & Wish, M. (1978). Multidimensional Scaling (Quantitative Application in the Social Sciences, 07-011). Beverly Hills and London: Sage Publications.

Lund K. & Burgess C. (1996). Producing high-dimensional semantic spaces from lexical co-occurrence. Behavior Research Methods, Instruments, & Computers, 28(2), 203-208.

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, 37(4), 547-559

Meule, A., & Blechert, J. (2012). food.pics: A picture database for the study of eating and appetite. Obesity Facts 5 (Suppl. 2), 20.

Pecher D., Zeelenberg R., & Raaijmakers J. G. (1998). Does pizza prime coin? perceptual priming in lexical decision and pronunciation. Journal of Memory and Language, 38(4), 401-418

Roach, M. (2013). Gulp: Adventures on the alimentary canal. New York: W.W. Norton.

Rosch, E. (1975). Cognitive representations of semantic categories. Journal of experimental psychology: General, 104(3), 192-233.