

Meaning Representation in Natural Language Categories

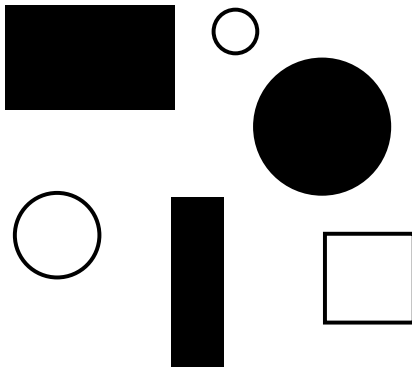
Trevor Fountain

School of Informatics
The University of Edinburgh

11 June 2010

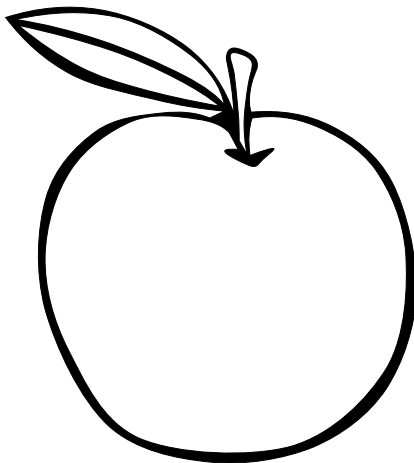
What is Categorization?

How do people assign objects to categories?



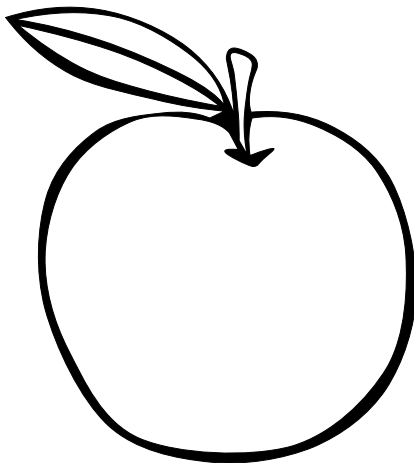
What is Categorization?

Why does it matter?



What is Categorization?

Why does it matter?



is_a_fruit

is_edible

is_perishable

grows_on_trees

...

Theories of Categorization

Classical Theory

- ▶ List of features which are both *necessary* and *sufficient*
- ▶ Items are placed in a category *iff* they possess all requisite features.

| |
|------------------------|
| FRUIT |
| <i>is_edible</i> |
| <i>is_sweet</i> |
| <i>grows_on_plants</i> |
| <i>contains_seeds</i> |

Theories of Categorization

Classical Theory

- ▶ List of features which are both *necessary* and *sufficient*
- ▶ Items are placed in a category *iff* they possess all requisite features.

| |
|------------------------|
| FRUIT |
| <i>is_edible</i> |
| <i>is_sweet</i> |
| <i>grows_on_plants</i> |
| <i>contains_seeds</i> |

What about tomatoes? Seedless grapes?

Theories of Categorization

Prototype Theory

- ▶ A *schema* in which features are weighted by importance
- ▶ Categorization is based on similarity to the schema.

| FRUIT | |
|------------------------|-----|
| <i>is_edible</i> | 0.1 |
| <i>is_sweet</i> | 0.2 |
| <i>grows_on_plants</i> | 0.9 |
| <i>contains_seeds</i> | 0.7 |

Theories of Categorization

Exemplar Theory

- ▶ List of previously encountered exemplars
- ▶ Categorization is based on similarity to each stored exemplar.

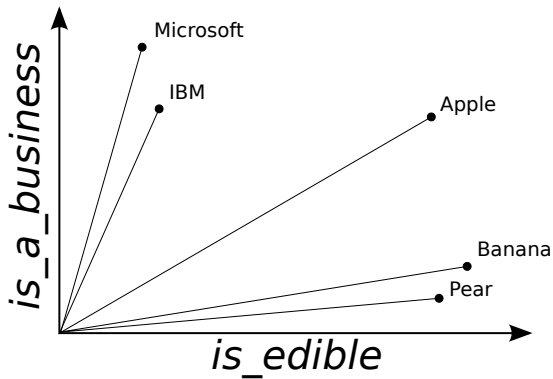
| |
|---------------|
| FRUIT |
| <i>Apple</i> |
| <i>Orange</i> |
| <i>Pear</i> |
| <i>Banana</i> |

Meaning Representation

- ▶ *Object vs. Word* categorization
- ▶ How do we represent the meaning of a word?
- ▶ Predicate logic? Lambda calculus? Vector spaces?

Similarity

in Feature Space



Meaning Representation

Why focus on the representation?

- ▶ Traditional categorization uses real-world features.
- ▶ In Natural Language Categorization these are features of the words' referents.
- ▶ ...but in a prototype or exemplar model features are only used to compute similarity.

Similarity

in Co-occurrence

Document 1

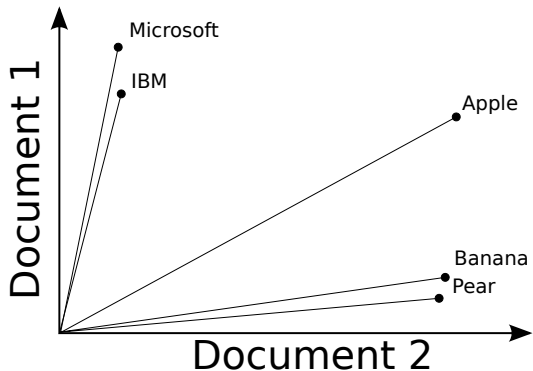
Tech companies Google, **IBM**, **Apple** and **Microsoft** head the world's top 100 most valuable brands, according to a new global survey on...

Document 2

Online Plant Nursery featuring exotic fruit trees including **apple**, **orange** and **pear** for your garden or orchard.

Similarity

in Co-occurrence



Question

Question: Can we approximate real-world features with co-occurrence counts, at least vis-a-vis similarity?

Outline

Introduction

- Categorization
- Representation
- Similarity

Tasks

- Category Naming
- Exemplar Generation
- Typicality Rating

Data

- Goal
- Norms
- Mechanical Turk
- Representations

Experiments

- Design
- Results

Tasks

Category Naming

Given a word, predict the category to which it belongs.

- ▶ 'apple' belongs to the category FRUIT.
- ▶ 'Microsoft' belongs to the category CORPORATION.

Tasks

Exemplar Generation

Given a category, output a set of words that exemplify it.

- ▶ FRUIT is exemplified by 'apple', 'orange', 'grape', etc.
- ▶ CORPORATION is exemplified by 'Microsoft', 'IBM', 'Apple', etc.

Tasks

Typicality Rating

Given a category-exemplar pair, rate how 'typical' the exemplar is among members of the category.

| | FRUIT | CORPORATION |
|-------------|-------|-------------|
| 'banana' | 0.70 | 0.02 |
| 'apple' | 0.95 | 0.40 |
| 'Microsoft' | 0.01 | 0.99 |

Data

What do we need?

Goal:

- ▶ List of target words grouped into categories
- ▶ Single label for each category
- ▶ Typicality rating for each exemplar within each category
- ▶ Vector representations for each word in both feature & corpus space

Data

Target Words

Use feature norms of McRae et al. 2005

- ▶ 541 nouns with human-annotated features
- ▶ No category labels
- ▶ No typicality ratings

Data

Mechanical Turk

[Your Account](#)[HITS](#)[Qualifications](#)**179,260 HITS**
available now[Sign In](#)[All HITS](#) | [HITS Available To You](#) | [HITS Assigned To You](#)Search for containing that pay at least \$ for which you are qualified **Timer:** 00:00:00 of 60 minutes [Want to work on this HIT?](#) [Want to see other HITS?](#) **Total Earned:** Unavailable**Total HITS Submitted:** 0

Judge Machine Translation output (for speakers of English)

Requester: WMT Admin**Reward:** \$0.01 per HIT**HITS Available:** 385**Duration:** 60 minutes**Qualifications Required:** Total approved HITS is greater than 5, HIT approval rate (%) is greater than 85, Location is US

Judge Edited Machine Translation Outputs

Instructions:

Data

Mechanical Turk

amazon **mechanical turk**
beta Artificial Intelligence

[Your Account](#)[HITS](#)[Qualifications](#)

179,260 HITS
available now

[Sign In](#)

[All HITS](#) | [HITS Available To You](#) | [HITS Assigned To You](#)

Search for containing that pay at least \$ for which you are qualified

Timer: 00:00:00 of 60 minutes [Want to work on this HIT?](#) [Want to see other HITS?](#) **Total Earned:** Unavailable

Total HITS Submitted: 0

Judge Machine Translation output (for speakers of English)

Requester: WMT Admin

Reward: \$0.01 per HIT

HITS Available: 385

Duration: 60 minutes

Qualifications Required: Total approved HITS is greater than 5, HIT approval rate (%) is greater than 85, Location is US

Judge Edited Machine Translation Outputs

Instructions:

Data

Mechanical Turk

[Your Account](#)[HITS](#)[Qualifications](#)**179,260 HITS**
available now[Sign In](#)[All HITS](#) | [HITS Available To You](#) | [HITS Assigned To You](#)Search for containing that pay at least \$ for which you are qualified **Timer:** 00:00:00 of 60 minutes [Want to work on this HIT?](#) [Want to see other HITS?](#) **Total Earned:** Unavailable**Total HITS Submitted:** 0

Judge Machine Translation output (for speakers of English)

Requester: WMT Admin**Reward:** \$0.01 per HIT**HITS Available:** 385**Duration:** 60 minutes**Qualifications Required:** Total approved HITS is greater than 5, HIT approval rate (%) is greater than 85, Location is US

Judge Edited Machine Translation Outputs

Instructions:

Data

Mechanical Turk

[Your Account](#)[HITS](#)[Qualifications](#)**179,260 HITS**
available now[Sign In](#)[All HITS](#) | [HITS Available To You](#) | [HITS Assigned To You](#)Search for containing that pay at least \$ for which you are qualified **Timer:** 00:00:00 of 60 minutes [Want to work on this HIT?](#) [Want to see other HITS?](#) **Total Earned:** Unavailable**Total HITS Submitted:** 0

Judge Machine Translation output (for speakers of English)

Requester: WMT Admin**Reward:** \$0.01 per HIT**HITS Available:** 385**Duration:** 60 minutes**Qualifications Required:** Total approved HITS is greater than 5, HIT approval rate (%) is greater than 85, Location is US

Judge Edited Machine Translation Outputs

Instructions:

Data

Category Naming

*In this HIT you are given a series of words and asked to label each one with the category to which best belongs. For example, you might assign "apple" the category "fruit", or decide that "computer" is a member of the category "device." Do **not** come up with a single category for entire group – the words are not necessarily related to one another. If you can, try to come up with category labels that are only a single word; for example, don't use "musical instrument" when "instrument" will do. I have filled in a few examples; you should complete the rest.*

| | Exemplar | Category |
|----------|-----------------|-----------------|
| EXAMPLE: | pizza | food |
| EXAMPLE: | calculator | device |
| | accordion | |
| | balloon | |
| | clarinet | |
| | sailboat | |
| | lime | |
| | whale | |
| | umbrella | |
| | buffalo | |
| | dishwasher | |
| | goldfish | |

Table: An example category naming task. For each exemplar, participants are asked to generate an appropriate category label.

Data

Typicality Rating

In this HIT you are given a set of words belonging to a single category and asked to rank how 'typical' each is of the category on a scale of 1 to 7. For example, if the category was "Car" you might assign the following typicality ratings to the words "Ford", "Saturn", and "Citroën":

EXAMPLE:

| Car | Rating | | | | | | |
|---------|--------|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| Saturn | | | | | | x | |
| Ford | | | | | | | x |
| Citroën | | x | | | | | |

YOUR TASK:

| Instrument | Rating | | | | | | |
|-------------|--------|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| accordion | | | | | | | |
| flute | | | | | | | |
| drum | | | | | | | |
| guitar | | | | | | | |
| harpsichord | | | | | | | |
| kazoo | | | | | | | |

Table: An example typicality rating task. For each exemplar in the given category participants are asked to rate how 'typical' that exemplar is among other members of the category.

Data

What do we need?

Goal:

- ▶ List of target words grouped into categories
- ▶ Single label for each category
- ▶ Typicality rating for each exemplar within each category
- ▶ **Vector representations for each word in both feature & corpus space**

Data

Representations

| | | | | | |
|-----------------------------------|-------|---------------------|------------------------|---------------------|-----|
| | | <i>has_4_legs</i> | <i>used_for_eating</i> | <i>is_a_pet</i> | ... |
| Feature Norms | TABLE | 12 | 9 | 0 | ... |
| | DOG | 14 | 0 | 15 | ... |
| | | Document 1 | Document 2 | Document 3 | ... |
| Latent Semantic Analysis (LSA) | TABLE | 0.02 | 0.98 | -0.12 | ... |
| | DOG | 0.73 | -0.02 | 0.01 | ... |
| | | Topic 1 | Topic 2 | Topic 3 | ... |
| Latent Dirichlet Allocation (LDA) | TABLE | 0.02 | 0.73 | 0.04 | ... |
| | DOG | 0.32 | 0.01 | 0.02 | ... |
| | | <i>subj-of-walk</i> | <i>subj-of-eat</i> | <i>obj-of-clean</i> | ... |
| Dependency Vectors (DV) | TABLE | 0 | 3 | 28 | ... |
| | DOG | 36 | 48 | 19 | ... |

Experiment

Design

2x3x4 design:

- ▶ Exemplar vs. Prototype
- ▶ Category Naming vs. Exemplar Generation vs. Typicality Rating
- ▶ Feature Norms vs. LSA vs. LDA vs. DV

Experiments

Results

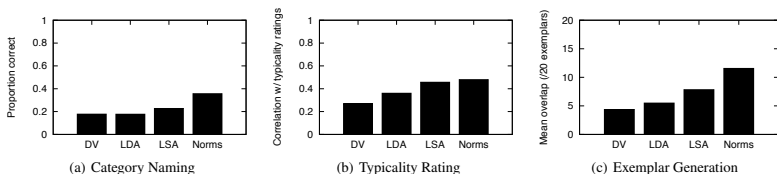


Figure 1: Performance of exemplar model using feature norms and data-driven meaning representations.

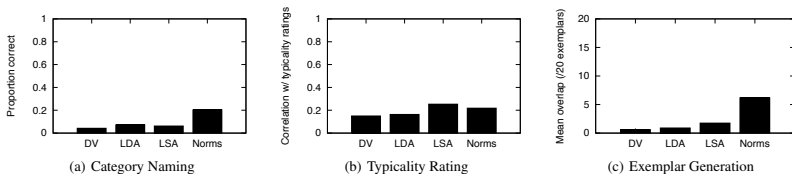


Figure 2: Performance of prototype model using feature norms and data-driven meaning representations.

Questions?

| | | | | | |
|------------|-------------|----------------|-----------|-------------|------------|
| INSTRUMENT | keyboard | FURNITURE | chair | HOUSING | apartment |
| REPTILE | rattlesnake | CONTAINER | bin | VEHICLE | bike |
| CLOTHING | jeans | STRUCTURE | building | VEGETABLE | carrot |
| HARDWARE | drill | APPLIANCE | stove | BIRD | seagull |
| HOUSE | cottage | PLANT | vine | TOOLS | hammer |
| EQUIPMENT | football | UTENSIL | ladle | THING | doll |
| TOY | surfboard | KITCHEN | dish | RODENT | rat |
| BUG | beetle | HOME | house | FRUIT | grapefruit |
| MAMMAL | horse | OBJECT | door | ACCESSORIES | necklace |
| STORAGE | cabinet | BUILDING | apartment | ANIMAL | cat |
| DEVICE | stereo | TRANSPORTATION | van | FOOD | bread |
| GARMENT | coat | FISH | trout | ENCLOSURE | fence |
| INSECT | grasshopper | SPORTS | helmet | COOKWARE | pan |
| WEAPON | bazooka | | | | |

Table: Category labels with most typical exemplars produced by participants in category naming and typicality rating study.