Exemplar Theory + Speech Perception

March 30, 2010
Reminders

• On Thursday:
  • Perception Labs due
  • Nasometer Labs will be returned
• Schedule of final project presentations.
  • April 13th + April 15th
• One announcement:
  • We’ll be doing a static palatography demo in Ling 441 on Thursday at 3:30
  • Professional Faculties 114
And Now For Something Completely Different.

• Q: What’s a category?
• A classical answer:
  • A category is defined by properties.
  • All members of the category exhibit the same properties.
  • No non-members of the category exhibit all of those properties.
• This definition of a category goes all the way back to the ancient Greek philosophers.
  • (who were much admired by Noam Chomsky, among others)
Classical Example

- A rectangle (in Euclidean geometry) may be defined as having the following properties:
  
  1. Four-sided, two-dimensional figure (quadrilateral)
  2. Four right angles

This is a rectangle.
Classical Example

• Adding a third property gives the figure a different category classification:

1. Four-sided, two-dimensional figure (quadrilateral)
2. Four right angles
3. Four equally long sides

This is a square.
Classical Example

• Altering other properties does not change the category classification:
  1. Four-sided, two-dimensional figure (quadrilateral)
  2. Four right angles
  3. Four equally long sides

This is still a square.

A. Is red.
Classical Example

- Altering other properties does **not** change the category classification:
  1. Four-sided, two-dimensional figure (quadrilateral)
  2. Four right angles
  3. Four equally long sides **definitive** properties

This is still a square.

A. Is red. **incidental** property
Classical Linguistic Categories

• Phonologists usually define all possible speech sounds in terms of a limited number of definitive properties, known as “distinctive features”. (Chomsky + Halle, 1968)

  \[d\] = [CORONAL, +voice, -continuant, -nasal, etc.]

  \[n\] = [CORONAL, +voice, -continuant, +nasal, etc.]

  ...

• Similar approaches have been applied in syntactic analysis. (Chomsky, 1974)

  Adjectives = [+N, +V]

  Prepositions = [-N, -V]
Prototypes

- The psychological reality of classical categories was called into question by a series of studies conducted by Eleanor Rosch in the 1970s.

- Rosch claimed that categories were organized around privileged category members, known as prototypes.
  - (instead of being defined by properties)

- Evidence for this theory initially came from linguistic tasks:
  1. Semantic verification (Rosch, 1975)
    - Is a robin a bird?
    - Is a penguin a bird?
  2. Category member naming.
Prototype Category Example: “Bird”
Exemplar Categories

• Cognitive psychologists in the late ‘70s (e.g., Medin & Schaffer, 1978) questioned the need for prototypes.
  • Finding: categorization phenomena could be explained without appealing to a category prototype.

• The basic idea:
  • Categories are defined by extension.
  • ⇒ Neither prototypes nor properties are necessary.

• Categorization works by comparing new tokens to all exemplars in memory.
  • Generalization happens on the fly.
A Category, Exemplar-style

“square”
Back to Perception

• When people used to talk about **categorical** perception, they meant perception of **classical categories**.

• A stop is either a [b] or a [g]
  • (no in between)

• Remember: in classical categories, there are:
  • definitive properties
  • incidental properties

• Q: What are the properties that define a stop category?

• The definitive properties must be **invariant**.
  • (shared by all category members)

• So…what are the invariant properties of stop categories?
The Acoustic Hypothesis

- People have looked long and hard for invariant acoustic properties of stops, with little success.
  
  - (and some people are still looking)

- Frequency values of compact (synthetic) bursts cueing different places of articulation, in various vowel contexts.

  (Liberman et al., 1952)
Theoretical Revision

• Since invariant acoustic properties could not be found (especially for velars)…
  • It was assumed that listeners perceived (articulatory) gestures, not (acoustic) sounds.

• Q: What invariant articulatory properties define stop categories?
  • A: If they exist, they’re hard to find.

• Motor Theory Revision #2: Listeners perceive “intended” gestures.

• Note: “intentions” are kind of impossible to observe.
  • But they must be invariant…right?
Another Brick in the Wall

• Another problem for motor theory:
  • Perception of speech sounds isn’t always categorical.

• In particular: vowels are perceived in a more gradient fashion than stops.

• However, vowel perception becomes more categorical when the vowels are extremely short.
• It’s also hard to identify any invariant acoustic properties for vowels.
• Variation is rampant across:
  • tokens
  • speakers
  • genders
  • dialects
  • age groups, etc.
• Variability = a huge problem for speech perception.
More Problems

• Also: infants exhibit categorical perception, too…
  • Even though they don’t know category labels.

• Chinchillas can do it, too!
An Alternative

- It has been proposed that phoneme categories are defined by **prototypes**…
  - which we use to identify vowels in speech.
- One relevant finding: the **perceptual magnet** effect.
- Part 1: play listeners a continuum of synthetic vowels in the neighborhood of [i].
  - Task: judge how much each one sounds like [i].
- Some are better = prototypical
  - Others are worse = non-prototypes
Perceptual Magnets

• Part 2: define either a prototype or a non-prototype as a category center.
• Task: determine whether other vowels on the continuum belong to those categories.

• Result: more same responses when the category center is a prototype.
• Prototype = a “perceptual magnet”
Prototypes, continued

• The perceptual magnet prototypes are usually located at a listener’s average F1 and F2 values for [i].

• 4-month olds exhibit the perceptual magnet effect…
  • but monkeys do not.

• Note: the prototype is the only thing that has to be “invariant” about the category.
  • particular properties aren’t important.

• Testing a prototype model on the Peterson & Barney data yielded 51% correct classification.
  • (Human listeners got 94% correct)

• ⇒ Variability is still hard to deal with.
Flipping the Script

• Another approach to speech perception is to **preserve** all variability that we hear…
  
  • Rather than boiling it down to properties or prototypes.

• In this model, speech categories are defined by **extension**.
  
  • = consist of exemplars

• So, your mental representaton of /b/ consists of every token of /b/ you’ve ever heard in your life.
  
  • …rather than any particular acoustic or articulatory properties.

• Analogy: phonetics field project notes
  
  • (your mind is a pack rat)
Exemplar Theory: Basic Precepts

• Doug Hintzman (1986) sketched out the first working exemplar model of memory. (MINERVA)

1. Listeners store in memory every experience they have in their lifetime.
   • Including all details of those experiences.

2. Specific experiences in memory are known as traces.
   • They are linked to category labels.

3. New experiences are known as probes.
Exemplar Theory: Basic Precepts

- Doug Hintzman (1986) sketched out the first working exemplar model of memory. (MINERVA)

4. When probes are encountered, they activate similar traces in memory.

- Note: amount of activation is proportional to similarity between trace and probe.

- Traces that closely match a probe are activated a lot;
  - Traces that have no similarity to a probe are not activated much at all.
Exemplar Theory: Basic Precepts

• Doug Hintzman (1986) sketched out the first working exemplar model of memory. (MINERVA)

5. The activation of traces in memory are averaged and combined into a perceptual response known as an echo.
Echoes from the Past

• Since the echo is a weighted average of perceptual responses to a probe, it has more general features than either the traces or the probe.

• Inspiration: Francis Galton
The Eyes of the Beholder

• Which of these faces is most attractive?

• And how about these?

• The faces judged most attractive (enclosed in black) exhibited the average:
  • vertical placement of the eyes
  • horizontal distance between the eyes
  • (Pallett et al., 2010)
Exemplar Predictions: Echoes

• The general features of the echo are not stored explicitly in memory…
  • They are just generated on the fly, with each new experience.
• \( \Rightarrow \) Category representations can shift over time.
• Features of the echo will also reflect:
  1. More common traces in memory
     • (since more of them contribute to the average)
     • = frequency effects
  2. Similarities between the probe and the traces in memory.
A (pretend) example: traces = vowels from the Peterson & Barney data set.

Activation of each trace is proportional to distance (in vowel space) from the probe.

Frequency of second formant versus frequency of first formant for ten vowels by 76 speakers.
Exemplar Predictions: Traces

• Point: all properties of all exemplars play a role in categorization…
  • Not just the “definitive” ones.

• Prediction: non-invariant properties of speech categories should have an effect on speech perception.
  • E.g., the room in which a [b] is spoken.
  • The “homefield advantage” effect.

• Is this true?
  • Let’s find out with a continuous word recognition task…
Word Recognition Experiment

- Circle whether each word is a **new** or **old** word in the list.

|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
| 1 |   |   | 9 |   |   | 17|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
| 2 |   |   | 10|   |   | 18|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
| 3 |   |   | 11|   |   | 19|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
| 4 |   |   | 12|   |   | 20|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
| 5 |   |   | 13|   |   | 21|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
| 6 |   |   | 14|   |   | 22|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
| 7 |   |   | 15|   |   | 23|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
| 8 |   |   | 16|   |   | 24|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
Word Recognition Experiment

- Circle whether each word is a **new** or **old** word in the list.

<p>| | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>25.</td>
<td>26.</td>
<td>27.</td>
<td>28.</td>
<td>29.</td>
<td>30.</td>
</tr>
<tr>
<td>33.</td>
<td>34.</td>
<td>35.</td>
<td>36.</td>
<td>37.</td>
<td>38.</td>
</tr>
<tr>
<td>39.</td>
<td>40.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Evidence for Exemplar Storage

• In a “continuous word recognition” task (Palmeri et al., 1993), listeners hear a long sequence of words…
  • some of which are new words in the list, and some of which are repeats.

• Task: decide whether each word is new or a repeat.

• Twist: some repeats are presented in a new voice;
  • others are presented in the old (same) voice.

• Finding: repetitions are identified more quickly and more accurately when they’re presented in the old voice.

• Implication: we store voice + word info together in memory.
Class-Based Data

All Stimuli

Percent Correct

<table>
<thead>
<tr>
<th>Word Type</th>
<th>50%</th>
<th>60%</th>
<th>70%</th>
<th>80%</th>
<th>90%</th>
<th>100%</th>
</tr>
</thead>
<tbody>
<tr>
<td>New</td>
<td></td>
<td></td>
<td></td>
<td>100%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Repeat-Same</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>70%</td>
</tr>
<tr>
<td>Repeat-Different</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>60%</td>
</tr>
</tbody>
</table>
Class-Based Data

Repeated Stimuli

Percent Correct

Within 10 Stimuli

After 10 Stimuli

Repeat-Same

Repeat-Different
More Interactions

• Another task (Nygaard et al., 1994):
  • train listeners to identify talkers by their voices.
  • Then test the listeners’ ability to recognize words spoken in noise by:
    • the talkers they learned to recognize
    • talkers they don’t know

• Result: word recognition scores are much better for familiar talkers.

• Implication: voice properties influence word recognition.

• The opposite is also true:
  • Talker identification is easier in a language you know.
Variability in Learning

• Caveat: it’s best not to let the relationship between words and voices get too close in learning.

• Ex: training Japanese listeners to discriminate between /r/ and /l/.

• Discrimination training on one voice: no improvement. (Strange and Dittman, 1984)

• Bradlow et al. (1997) tried:
  • training on five different voices
  • multiple phonological contexts (onset, coda, clusters)
  • 4 weeks of training (with monetary rewards!)

• Result: improvement!
Variability in Learning

• General pattern:
  • Lots of variability in training → better classification of novel tokens…
  • Even though it slows down improvement in training itself.

• Variability in training also helps perception of synthetic speech. (Greenspan et al., 1988)

• Another interesting effect: dialect identification (Clopper, 2004)

• Bradlow et al. (1997) also found that perception training (passively) improved production skills…
• Japanese listeners performed an /r/ - /l/ discrimination task.

• Important: listeners were told nothing about how to produce the /r/ - /l/ contrast

• …but, through perception training, their productions got better anyway.
Exemplars in Production


• Task 1--Production:
  • A: Listeners read a word (out loud) from a script.
  • B: Listeners hear a word (X), then repeat it.

• Finding: formant values and durations of (B) productions match the original (X) more closely than (A) productions.

• Task 2--Perception: AXB task
  • A different group of listeners judges whether X (the original) sounds more like A or B.

• Result: B productions are perceptually more similar to the originals.
Shadowing: Interpretation

• Some interesting complications:
  • Repetition is more prominent for low frequency words…
  • And also after shorter delays.

• Interpretation:
  • The “probe” activates similar traces, which get combined into an echo.
  • Shadowing imitation is a reflection of the echo.

• Probe-based activation decays quickly.
  • And also has more of an influence over smaller exemplar sets.
XMOD

• Johnson (1997) created an exemplar-based model of speech perception
  • called “XMOD”
  • implemented computationally!
• It’s possible to train XMOD:
  • it stores in memory all experienced tokens of speech
  • and associates each one with a label
• Labels can be linguistic or indexical
  • indexical = (male, female, Canadian, American, old, young, etc.)
XMOD Testing

• A trained XMOD model can be tested with new probes.

• Activation is based on similarity between probes and stored traces in memory.
  • similarity is based on spectral distance between successive slices of probes and traces.

• Activation of traces is proportional to similarity:
  • More similar traces receive more activation.

• Activation percolates up to the labels associated with each trace;
  • The category label with the most activation is the perceptual “winner”.
Vowel Tests

• Human beings correctly identify 94% of the vowels in the Peterson & Barney data set.

• Using only the midpoint formant frequencies…
  • A very basic exemplar model (not XMOD) gets 81% correct…
  • And a very basic prototype model gets 51% right.

• Moral of the story: storing acoustic details in memory makes it easier to deal with variability in speech.
  • Exemplar models are more “robust”

• Another example: silent-center syllable perception.
Silent-Center Syllables

- Rakerd & Verbrugge (1978) found that listeners could identify vowels in CVC syllables even when the centers of them had been edited out and replaced with silence.

- Examples:

- It even works when the gender switches in the middle of the silence!

- For instance:

- One theory: human listeners interpret these vowels with respect to a (gender-free) category abstraction.

- An exemplar model generates abstractions on the fly…

  - Can it correctly categorize these vowels, too?
Simulation Results

• Answer: Yes!

• Exemplar models have shown great promise for advancing the performance of speech recognition technology.
Moral(s) of the Exemplar Story

• Exemplar theory holds that:

1. Linguistic categories are defined by extension.

   • There are therefore no “definitive” or “incidental” properties.

2. Generalizations emerge “on the fly”.

3. These generalizations are more influenced by recent activity;

   • and can change over time.

4. It is beneficial to preserve details (and variability) in mental representations of speech.

   • Rather than reducing linguistic elements to “minimal”, abstract structures.