Why it’s hard for robots to learn word meanings

Paul Cohen
Department of Computer Science
How Children Learn Word Meanings

"Here is how they do it: Young children can parse adult speech (or sign) into distinct words. They think of the world as containing entities, properties, events, and processes; most important, they see the world as containing objects. They know enough about the minds of others to figure out what they are intending to refer to when they use words. They can generalize…They can also make sense of pronouns and proper names, which refer to distinct individuals, not to categories…Months after their first words, [children exploit]… the syntactic and semantic properties of the utterances that new words belong to. This enables the learning of many more words, including those that could only be acquired through this sort of linguistic scaffolding."  – Paul Bloom
The Environment of Child Language Acquisition

Child and adult are embedded in a scene the child can parse and understand.

Most dialog with kids is about something, usually present or immediately inferable things.

Mental state is perhaps the most interesting inferable thing.

Not like corpus-based language learning.

“Don’t you want your cookie?”
“No”
“You are so stubborn!”
<< kid learns some of what stubborn means >>
Associative learning of word meanings

“stubborn” =

“ Don’t you want your cookie? ”
“ No”
“ You are so stubborn!”
Referential ambiguity:
An objection to an empiricist account

“Poor thing!” (the dog)

“Do you think it’s full?” (the cup)

“Isn’t it hot?” (the coffee)

“OMG look at that!” (the whole scene)

“Quick, take a picture!” (someone outside of the scene)

“Do you suppose it hurts?” (???)

Heuristics (e.g., whole object assumption), parents work hard, kids are good mind readers, and above all the empirical fact that we can figure out the referents in each sentence given the scene.
Robots learn word meanings: Early attempts
The agents look at the scene on the whiteboard in front of them.

One agent says a word which distinguishes one object in the scene from all the others.

If the other agent doesn't know the meaning of the word, the game fails.

If one agent doesn't know the meaning of a word used by another, it can't pick out the object that the other agent is talking about. But it can try to work out a property - such as a color or size - that identifies one object in the scene, and learn that as the meaning of the unknown word.

Luc Steels
the talking heads

One agent says a word which distinguishes one object in the scene from all the others.

The hearer checks the word against its own vocabulary.

The hearer understands the word 'wabaku' to mean 'red'.

As the agents learn each other's words, they begin to be able to communicate. In this example, the hearer has successfully recognised that the word 'wabaku' denotes a particular shade of red, and has been able to identify the object that the speaker has in mind.

Luc Steels
Associative learning issues

What should this representation be?  How to represent unobservables like speaker's intentions?  How best to collect the data
The scene is often dynamic

Formal representations of scenes  Sentences

“Don’t you want your cookie?”
“No”
“You are so stubborn!”
Robot Baby
Associating Words with Sensory Patterns
(with Tim Oates, 1998 - 2001)
Assessment of Robot Baby word learning

**Good**

- Making sensory experience part of word meaning – what it "feels like" to bump something
- Associating sensory prototypes with classes of words, as classes of words have meanings, too
- Ability to represent temporal "story" is good for verb meanings

**Not So Good**

- There's more to the meaning of a word than the sensory experience associated with it

  ![Forward](https://example.com/forward.png) ![Backward](https://example.com/backward.png)

  But can't learn that forward and backward are *antonyms*

- Sensory prototypes don't individuate and denote things and relations. Where's the *object of the verb*, the thing the robot bumped or hit?

  ![Don't bump the block](https://example.com/don't-bump-the-block.png)
Welcome

Wubble World is a fun place to meet new friends and play creative games. Wubbles are curious and intelligent creatures, as you'll find out when you teach your very own Wubble about the world it lives in. You and your Wubble can compete with other players and win cool accessories.

Enter now to customize your Wubble and begin playing!
Welcome to Wubble World:

go to the green rectang
How Wubbles Learn the Meanings of Words

"Go to the blue cylinder"

1. parse

2. retrieve a semantic representation of a word or phrase

3. score correspondences between retrieved representation and scene

4. act or ask

5. given positive feedback, update semantic representation
Assessment of Wubble World word learning

Good

Simple, gradual learning method based on positive instances only. Learns multiple word meanings simultaneously. Converges fast to correct meaning.

Demonstrates that kids can teach wubbles word meanings in an enjoyable web-based game

Syntactic bootstrapping - parse first, guess at part of speech, create new feature vector for it

Not So Good

Unclear whether it will work for richer semantic representations

Even these smart, geeky kids were bored after two hours (NB: all agreed that wubbles that learn and respond are better than usual fare)

Very poor semantics, very few parts of speech, what about syntactically ambiguous words?

What about verbs?

What about the speaker's mental state? “I want the red one…No, the other one.”
Checkpoint

- Most learning of word meanings in AI is superficial in the sense of associating sensory features with words.
- You can’t do much with those learned meanings: antonymy, synonymy, comparing word senses, metaphor, etc.
- Some of those learned meanings don’t even function as representations; don’t individuate or denote, aren’t composable, etc.

- Games are probably a good data collection method.
- Association isn’t a bad way to learn.
- If we had better features we could learn deeper word meanings.

- If words are just names for concepts then word learning is largely concept learning: deeper meaning = associations w/ more concepts.
- Word meanings are acquired gradually and deepen over time.
- Deep understanding doesn’t require a big, fancy vocabulary.
Desiderata for concept representations
“you are so stubborn!”

“stubborn”
These probably aren’t arbitrary arrangements, they have a syntax, e.g., type constraints.

They do work, e.g., comparing concepts, finding similar concepts, inference – prediction and explanation -- and QA, analogy, metaphor.

We haven’t found an off-the-shelf source of these concepts or semantic primitives.
Pr(utter word | semantic features) is gradual function of the features

Clustering experiences in metric space and associating words with joint probability distributions Pr(utter word | dimensions) seem to work
Representations of verbs should “tell the story”

<table>
<thead>
<tr>
<th>jump over</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>InBackOf(W,D)</td>
<td>11111111</td>
</tr>
<tr>
<td>NegativeVerticalMotion(W)</td>
<td>111111</td>
</tr>
<tr>
<td>Away(W,ceiling)</td>
<td>11111111</td>
</tr>
<tr>
<td>Towards(W,floor)</td>
<td>11111111</td>
</tr>
<tr>
<td>Away(W,D)</td>
<td>1111111111</td>
</tr>
<tr>
<td>Below(D,W)</td>
<td>111111111</td>
</tr>
<tr>
<td>Above(W,D)</td>
<td>111111111</td>
</tr>
<tr>
<td>PositiveVerticalMotion(W)</td>
<td>11111111</td>
</tr>
<tr>
<td>Away(W,floor)</td>
<td>11111111</td>
</tr>
<tr>
<td>Towards(W,ceiling)</td>
<td>11111111</td>
</tr>
<tr>
<td>Jump(W)</td>
<td>11111111</td>
</tr>
<tr>
<td>Towards(W,D)</td>
<td>11111111</td>
</tr>
<tr>
<td>SurfaceMotion(W)</td>
<td>11111111</td>
</tr>
<tr>
<td>Motion(W)</td>
<td>11111111</td>
</tr>
<tr>
<td>Forward(W)</td>
<td>11111111</td>
</tr>
<tr>
<td>Collision(W,floor)</td>
<td>11111111</td>
</tr>
<tr>
<td>On(W,floor)</td>
<td>11111111</td>
</tr>
<tr>
<td>InFrontOf(W,D)</td>
<td>11111111</td>
</tr>
</tbody>
</table>

But they also should be denoting and afford propositional reasoning
Softbots learn word meanings: Current work
Unsupervised Learning of Associations Between Sentence and Static Scene Graphs

Daniel Hewlett
Verb models for Wubble World
Wesley Kerr

Time

249 propositions
Verb models for Wubble World

These are the seven Allen Relations

Coincidence without cause is rare

Are these rare coincidences or real patterns?

A fluent is an interval during which a proposition is true
Illustrative results

The algorithm learned an 18-fluent pattern corresponding to jump-over with no intrusions from the other 221 noise fluents.

<table>
<thead>
<tr>
<th>Fluent</th>
<th>Pattern</th>
</tr>
</thead>
<tbody>
<tr>
<td>InBackOf(W,D)</td>
<td>1111111111</td>
</tr>
<tr>
<td>NegativeVerticalMotion(W)</td>
<td>11111111111</td>
</tr>
<tr>
<td>Away(W, ceiling)</td>
<td>11111111111</td>
</tr>
<tr>
<td>Towards(W, floor)</td>
<td>11111111111</td>
</tr>
<tr>
<td>Away(W, D)</td>
<td>11111111111</td>
</tr>
<tr>
<td>Below(D, W)</td>
<td>11111111111</td>
</tr>
<tr>
<td>Above(W, D)</td>
<td>11111111111</td>
</tr>
<tr>
<td>PositiveVerticalMotion(W)</td>
<td>11111111111</td>
</tr>
<tr>
<td>Away(W, floor)</td>
<td>11111111111</td>
</tr>
<tr>
<td>Towards(W, ceiling)</td>
<td>11111111111</td>
</tr>
<tr>
<td>Jump(W)</td>
<td>11111111111</td>
</tr>
<tr>
<td>Towards(W, D)</td>
<td>11111111111</td>
</tr>
<tr>
<td>SurfaceMotion(W)</td>
<td>1111111111111111111111</td>
</tr>
<tr>
<td>Motion(W)</td>
<td>1111111111111111111111</td>
</tr>
<tr>
<td>Forward(W)</td>
<td>1111111111111111111111</td>
</tr>
<tr>
<td>Collision(W, floor)</td>
<td>1111111111111111111111</td>
</tr>
<tr>
<td>On(W, floor)</td>
<td>1111111111111111111111</td>
</tr>
<tr>
<td>InFrontOf(W, D)</td>
<td>1111111111111111111111</td>
</tr>
</tbody>
</table>

![Graph showing the breakdown of learned composite fluents](image.png)
Omniscient Word Learning

Heider and Simmel, 1944
(e.g., http://www.youtube.com/watch?v=sZBKer6PMtM)
Omniscient Word Learning
Wes Kerr, Jim O’Donnell

Each agent is controlled by a Soar-based BDI system so we have access to

- position and velocity
- dynamics
- perceptions of each agent
- which controllers were “on”
- which goals were “on”
- affective state
- inferences given state

Protocol
1) Run the agents in social interactions, recording all of the above (the “semantic trace”)
2) Make movies and show them to humans, who annotate them
3) Bring the annotations into correspondence with the semantic trace to figure out the meanings of words
Why it is hard for robots to learn word meanings

Not sensing and perception, learning methods, lack of data, lack of humans to interact with, or lack of interesting environments

It’s the lack of conceptual foundations for word meanings that connect to sensing and perception, are learnable from data, and that support all the kinds of reasoning one does with linguistically-expressed concepts