Autonomous Knowledge Capture: Systems that Build Themselves

Paul Cohen
USC Information Sciences Institute
## Autonomous Knowledge Capture

| 1970s: knowledge engineers feeling their way | 1980s: knowledge engineering tools | 1990s: ontologies, models, assembly of knowledge, experts building their own systems | 21st century knowledge systems that build themselves |
How will knowledge systems build themselves?

- Lenat and Feigenbaum’s answer: Cyc will “go critical” and start to read, augmenting its knowledge
- My answer: Systems will develop conceptually and linguistically much as children do
- Web mining (???)

- What do these answers have in common and what are the scientific bets?
  - Embodiment, dynamics vs. disembodied descriptions, logic
  - Learning by doing vs. learning in the abstract
  - Representational homogeneity vs. heterogeneity

- How do children, office assistants learn
- Robot Baby
## Human Language Milestones

<table>
<thead>
<tr>
<th>Age</th>
<th>Lexicon and Semantics</th>
<th>Grammar</th>
</tr>
</thead>
<tbody>
<tr>
<td>Birth - 1 year</td>
<td>Prefers sounds of native language</td>
<td>Develops sensitivity to natural phrasal units</td>
</tr>
<tr>
<td></td>
<td>Detects words in speech</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Preverbal gestures</td>
<td></td>
</tr>
<tr>
<td>1 - 2 years</td>
<td>Says first words. Vocabulary grows to several hundred words</td>
<td>Combine 2 words in telegraphic speech. Starts to generate 3-word</td>
</tr>
<tr>
<td></td>
<td></td>
<td>sentences with subject, verb, object</td>
</tr>
<tr>
<td>3 - 5 years</td>
<td>Understand metaphor based on sensory comparisons. Coins new words for</td>
<td>Forms sentences with adult grammatical categories. Masters complex</td>
</tr>
<tr>
<td></td>
<td>concepts</td>
<td>grammatical constructs</td>
</tr>
<tr>
<td>6 - 10 years</td>
<td>At school entry, vocab.~10K words Grasps meanings given only definitions.</td>
<td>Masters more complex constructs such as passive and infinitive</td>
</tr>
<tr>
<td></td>
<td>Appreciate multiple meanings, metaphor, humor</td>
<td></td>
</tr>
</tbody>
</table>
How the lexicon is learned

• Shared attention: By 12-15 months kids are expert at attending to what adults are attending to, often the referent of speech.
• Whole-object assumption: Word probably refers to the object not a part thereof
• Nouns first, though some verbs, function words, descriptors
• Fast mapping: Very fast learning of some aspect of word meaning based on very few, incidental exposures (e.g., “bring me the chromium tray”…sufficient for kids to learn “chromium” is a color). Note: No explicit instruction!
• Syntactic bootstrapping: “This is the chromium one” Sentence structure alerts child to word class of chromium.
• Always grounded, rarely via instruction, almost never via word definitions. Young kids never ask for definitions, don’t understand them
## How office assistants learn

### Expert - novice differences

<table>
<thead>
<tr>
<th>Reason for learning</th>
<th>Experts</th>
<th>Novices</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Learn for the future,</td>
<td>Learn for a particular task</td>
</tr>
<tr>
<td></td>
<td>learn for others</td>
<td></td>
</tr>
</tbody>
</table>

### Availability of schemas

<table>
<thead>
<tr>
<th></th>
<th>Experts</th>
<th>Novices</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Learn new as a variant of old, copy-edit</td>
<td>Still trying to learn the basic schemas</td>
</tr>
<tr>
<td>Unusual cases</td>
<td>Unusual cases</td>
<td>May not recognize that a case is unusual</td>
</tr>
</tbody>
</table>

### How opportunities arise

<table>
<thead>
<tr>
<th></th>
<th>Experts</th>
<th>Novices</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Actively test variations</td>
<td>Mightn’t even be able to generate variations</td>
</tr>
</tbody>
</table>

### What is learned

<table>
<thead>
<tr>
<th></th>
<th>Experts</th>
<th>Novices</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Procedures + reasons</td>
<td>Procedures</td>
</tr>
</tbody>
</table>

### How active

<table>
<thead>
<tr>
<th></th>
<th>Experts</th>
<th>Novices</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Knows what she knows</td>
<td>Overestimates what she knows</td>
</tr>
</tbody>
</table>

### Metacognitive aspects

<table>
<thead>
<tr>
<th></th>
<th>Experts</th>
<th>Novices</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
As humans execute **procedures** they construct **non-procedural domain knowledge**

The density of *grounded* concepts around procedures grows as we try procedures in new contexts, invent new variants, and learn constraints, preferences, and the background knowledge to express them.
Learning by doing

“I wish someone had told me five years ago, when I started work, that learning about procedures comes first and the facts, justifications, and other knowledge comes later. I thought I was learning backwards and there was something wrong with me.” -DB
How we learn

There is a present scene which often is generated by performing a task

Shared attention helps identify a subject in the scene

Learner has mental model of the scene / task under discussion

Some part of the model is given, some part is new. Cooperative teacher or information source introduces new part, helps connect it to given part

The learner is grounded in the physical world, so new concepts can make reference to physical and dynamical features of the scene.
How we learn

There is a *present scene* which often is generated by performing a task

Shared attention helps identify a subject in the scene

Learner has mental model of the scene / task under discussion

Some part of the model is *given*, some part is *new*. Cooperative teacher or information source introduces new part, helps connect it to given part

The learner is grounded in the physical world, so new concepts can make reference to physical and dynamical features of the scene.
Lexical learning

Model

Person presents

Object
Name
Shape
Color

Associations

“This is..”

“ball”

“blue”

Scene

This is a blue ball

Language
Lexical learning challenges

What kinds of models and how to learn them

Extracting words from speech

Extracting aspects of the scene to which words refer

Person presents

Object
Name
Shape
Color

“This is..”

Which associations?

“ball”

“blue”
Clustering and association to learn word meanings

Dozens of sensory episodes

Clustered by dynamics yields prototypes

Associative learning sensory aspects of word meanings

“The robot bumped into the wall”
“It followed the other one”
"The robot pushed the green block”

Distributional clustering

Contact

Not punctual

bump hit push move

Cohen – USC Information Sciences Institute and University of Massachusetts, Amherst
"The robot was spinning to its right"
"It's turning right until it faces forward"
"Robot turns clockwise ninety degrees"

"It's moving backwards"
"The robot is backing up"
"The robot is reversing"
Many instances of this activity look qualitatively like this instance. Can we cluster them?

Translational vel.

Rotational vel.

2 of 7 sonars
Clustering by Dynamics: Similarity

• Clustering methods are unsupervised, and group together elements based on their similarity

High intragroup similarity  
Low intergroup similarity

• How to judge the similarity of multivariate time series?
Dynamic time warping as a similarity measure
(Oates 1998)

Best warped fit is the similarity of A and B

Polynomial time, easily extended to multivariate series
Multivariate series in each cluster are averaged to produce a *prototype* for the cluster.
Clustering and Association
Unsupervised learning of word meanings

Dozens of sensory episodes

Cluster by dynamics yields prototypes

Associative learning sensory aspects of word meanings

"The robot bumped into the wall"
"It followed the other one"
"The robot pushed the green block"

Distributional clustering

Contact
Not punctual

bump hit push move

T.V.
Hierarchical clustering of words by distributional statistics

Words in syntactic classes appear in similar contexts; e.g., “The big X ate…” For clustering, words are similar if they appear in similar contexts.

Interior nodes are often labeled by linguists with syntactic or semantic class labels.
Words and word classes obtained by distributional clustering in Oates Cohen Eyler-Walker experiment

[verb]

No direct object

[verb]

Direct object

No contact

Contact

Punctual

Not punctual

TV TV

move

[verb]

turn start stop avoid follow bump hit push

Cohen – USC Information Sciences Institute and University of Massachusetts, Amherst
Representational requirements

What do we require of representations for them to capture all aspects of what words mean and how they are used?

• Allows variations in scale (short vs. long “push”)
• Preserves the “shapes” of experiences (for prediction and recounting)
• Captures the dynamics of situations
• Compositional (this part refers to forward movement, that part refers to the pushee)
• Denoting (this part refers to…)
• $Pr(\text{utter word} \mid \text{semantic features})$ is gradual function of the features
The trouble with sensory prototypes – they don’t denote objects and they aren’t compositional

Turn toward the cup on your left
The trouble with sensory prototypes – they don’t denote objects and they aren’t compositional.
Perceptual system for the robot

State description at time $t$

- (green C)
- (red A)
- (left-of A C)
- (right-of C A)
- (behind C A)
- (in-front A C)
- (approach R A)
- ...

Given by us, and learned

Scene

Feature detectors

deictic markers

Cohen – USC Information Sciences Institute and University of Massachusetts, Amherst
• Oates, 2001, learned word models from continuous speech and scenes, then learned the semantics of ~40 words in a corpus of ~100 words

• Steels "Talking Heads" project: evolve a lexicon, negotiate meanings of words

• Regier's algorithms learned meanings of prepositions

• Many linguists ground semantics in physical primitives

• Maps for Verbs theory grounds semantics of some verbs in a dynamical model
Four of Eighteen Movies

- 10:02
- 11:12
- 14:08
- 16:43

Cohen – USC Information Sciences Institute and University of Massachusetts, Amherst
Maps for Verbs

VR = V(A) - V(B)

VR > 0

VR = 0

VR < 0

E(AB) < 0

E(AB) = 0

E(AB) > 0

D(AB) > 0

D(AB) = 0

D(AB) < 0

Distance of AB unit > 0

Cohen – USC Information Sciences Institute and University of Massachusetts, Amherst
Procedure

• Each of 44 adult subjects saw 18 movies
• Each was asked to write about
  – 1. What are the balls doing in this movie? (Give your overall impression of what was happening between them, the ‘gist’)  
  – 2. What is the red ball doing?  
  – 3. What is the blue ball doing?  
  – 4. Can you think of any words to describe the tone or the mood of the movie? (e.g., the balls are friendly/ not friendly)
• We canonicalized all words – pushed, push, pushing -> pushing
• We selected 65 verbs, those used by at least 10 subjects
Procedure 2

65 of 155 verbs were use by at least 10 subjects: advancing, annoying, approaching, attaching, attacking, avoiding, backing, beating, bouncing, bullying, bumping, catching, charging, chasing, circling, coming, controlling, defending, dominating, escaping, fighting, floating, following, forcing, getting, giving, guiding, helping, hitting, kissing, knocking, leading, leaving, letting, looking, losing, nudging, pursuing, placing, playing, propelling, pushing, repeating, repelling, resisting, responding, rolling, running, shoving, slamming, slowing, sneaking, standing, standing-ones-ground, staying, stopping, striking, tagging, teasing, touching, traveling, trying, waiting, wanting, winning
Procedure – Hierarchical agglomerative clustering based on Euclidean distance between ranked word frequency vectors

<table>
<thead>
<tr>
<th>Movie 0</th>
<th>advancing</th>
<th>annoying</th>
<th>approaching</th>
<th>waiting</th>
<th>wanting</th>
<th>winning</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>4</td>
<td>6</td>
<td>0</td>
<td></td>
<td>54</td>
<td>61</td>
</tr>
<tr>
<td>Movie 1</td>
<td>17</td>
<td>31</td>
<td>30</td>
<td></td>
<td>14</td>
<td>11</td>
</tr>
<tr>
<td>Movie 17</td>
<td>51</td>
<td>47</td>
<td>37</td>
<td></td>
<td>9</td>
<td>21</td>
</tr>
</tbody>
</table>
Procedure – Label the movies with five features

1. Red doesn’t react to blue, contact begins gently vs. red is reactive, contact begins not gently.

2. Red looks purposeful or not. Harass-wander is wrong

3. Gentle start/end vs. not

Subtrees A and B identical when all 155 verbs are used

Pilot results with 3 - 4 year-olds are unclear.
Push
Agent: Must be animate
Patient: Animate or inanimate

Contact X,Y
Trans. Vel. +
D(Trans. Vel. +) < 0

Push: A then B then C
Manner: as acceleration increases, more like “hit”

With such a representation one might understand the denotations of words, infer the entailments (prediction), judge the appropriateness of words, use the words grammatically, etc.
How will knowledge systems build themselves?

- Lenat and Feigenbaum’s answer: Cyc will “go critical” and start to read, augmenting its knowledge
- My answer: Systems will develop conceptually and linguistically much as children do

- What do these answers have in common and what are the scientific bets?
  - Embodiment, dynamics vs. disembodied descriptions, logic
  - Learning by doing vs. learning in the abstract
  - Representational homogeneity vs. heterogeneity

- Synthesis??
Challenge Problem: Language learning and the enrichment or deepening of knowledge through experience (thanks Michael Witbrock)

Cyc knows a lot about distance, including:

- A salient fact about touches directly is that iff some X and some Y touch each other or share a part then the shortest distance between X and Y is zero feet.

Can these representations enrich each other and the sum of the knowledge?

Does grounding help?
Suppose the concept *collision* exists in the Cyc knowledge base. It can be found by matching the following description to its axioms:

- Before this event, translational velocity is positive, afterwards it is roughly zero
- Before this event, a distance-measuring device returns a positive number, afterwards it returns zero
- Before this event the bump sensor is low, afterwards it is high

Then any assertion about collision can be conjectured to hold for these prototypes, e.g., “bump sensor” measures (in part) change of state from not touching to touching
The prototype can also generate conjectures about collision, such as, “every collision is preceded by a reduction of distance to zero” (Consider gentle push)

This is a conjecture about the meaning of the word “collision”, generated by the machine! It’s trying to figure out what a word means!
Why must programs learn what representations mean?
"It's easy … to imagine symbol tokens interacting causally in virtue of their syntactic structures. The syntax of a symbol might determine the causes and effects of its tokenings in much the way that the geometry of a key determines which locks it will open. … The syntax of a symbol determines its causal role in a way that respects its content" – Fodor, 1987

"If you take care of the syntax, the semantics will take care of itself."

– Haugeland, 1981
It was only because of the intimate relationship between Lisp and Mathematics that the mutation operators … turned out to yield a high “hit rate” of viable, useful new math concepts. … Of course we can never directly mutate the meaning of a concept, we can only mutate the structural form of the concept as embedded in some representation scheme. Thus there is never a guarantee that we aren’t just mutating some ‘implementation detail’ that is a consequence of the representation, rather than some genuine part of the concept’s intentionality.  

(Lenat and Brown, AAAI, 1983, p.237)
"Taking care of the syntax" is a hard design problem

Unlike locks, where all the pieces fit one way, once and for all, components of computer programs can interact in unforeseen and incorrect ways.

We must check that conclusions make sense because the machine cannot, being disassociated from semantics: **Semantic babysitting**

Tasks that depend on semantics—e.g., merging knowledge bases, automatic knowledge acquisition, explanation—are difficult or impossible.

I want programs that don’t need semantic babysitting, programs that really know what their representations mean.
Stiff-neck, fever, headache suggest spinal meningitis
What do these symbols mean?

Stiff-neck, fever, headache suggest spinal meningitis.

This is the lay term for meningitis caused by the "meningococcus," Neiserria meningitidis. ...All meningitis is "spinal," since the infectious agent circulates freely through the medium of the spinal fluid between the brain and spinal cord. Meningococcal meningitis is highly contagious by the respiratory route (you get it by breathing air that contains the germs, which are freely shed from the mucous membranes). It is particularly easy to transmit if conditions are crowded, for example in army barracks or college dormitories. Preventative antibiotics are routinely given to all contacts of the identified patient to prevent their colonization with the germ.

...Smoking is a risk factor for catching the disease. Meningococcal disease is especially frightening to doctors, because no matter how ardently we search for it, every once in a while we will miss a case, with fatal results. This is because the disease can start so insidiously, mimicking a trivial viral illness, with few if any definite symptoms.

This is the lay term for meningitis caused by the "meningococcus," Neisseria meningitidis.

All meningitis is "spinal," since the infectious agent circulates freely through the medium of the spinal fluid between the brain and spinal cord. Meningococcal meningitis is highly contagious by the respiratory route (you get it by breathing air that contains the germs, which are freely shed from the mucous membranes). It is particularly easy to transmit if conditions are crowded, for example in army barracks or college dormitories. Preventative antibiotics are routinely given to all contacts of the identified patient to prevent their colonization with the germ.

Smoking is a risk factor for catching the disease. Meningococcal disease is especially frightening to doctors, because sometimes no antecedent cause is found for it, every once in a while we will miss a case, with fatal results. This is because the disease can start so insidiously, mimicking a trivial viral illness, with few if any

Stiff-neck, fever, headache suggest spinal meningitis

Suggestion of spinal meningitis suggests spinal tap to confirm

Syntactic, in the "mind of the machine"
Grammar development: First three years

As with lexicon, comprehension precedes production.
Telegraphic (two-word) utterances give way to three-word sentences that have subject, verb, object order of the native language, around 3 years of age.

The child acquires some fairly complex regular and irregular forms, interestingly, in a fixed order (source: Brown, 1973)

- Verb present progressive (He singing)
- Prepositions on, in
- Noun plural
- Verb irregular past tense (He ran, It broke)
- Noun possessive (Daddy’s hat)
- Uncontractible “be” form (Are kitties sleepy)
- Articles “a” and “the”
- Verb regular past tense “-ed” (He kicked it)
- Verb present tense, third person singular (He likes it)
- Verb present tense, third person irregular (He has [not have] it)
How grammar is acquired

This is a subject of vigorous debate! Positions include:

• Universal grammar: Children are born with a universal grammar in their heads, individual languages parameterize the grammar differently, grammar learning = parameter learning
• Semantic bootstrapping: basic grammatical categories come from identifying semantic word classes. A nativist version of this says the word classes are innate, the words that belong in each are learned
• Various machine learning/ neural network/ grammatical inference / ngram model approaches
You never take out the garbage!

Sheesh, look at that disgusting mess…
An early half-wit system: Eliza (Weizenbaum)

You never take out the garbage!

Sheesh, look at that disgusting mess...

“Would you like me to take out the garbage?”

“Would you like me to X?”

“You never X”
You never take out the garbage!

Sheesh, look at that disgusting mess...

Yes! Why must I do it all the time?

“Would you like me to X?"

“You never X”

Would you like me to take out the garbage?
You never take out the garbage!

Sheesh, look at that disgusting mess...

Yes! Why must I do it all the time?

"You never X"

"Would you like me to X?"

Would you like me to take out the garbage?

"Why must I Z"

"Why do you feel you must Z?"

Why do you feel you must do it all the time?
Which one is the half-wit?

This one knows what's being discussed.

You never take out the garbage!

This one doesn’t, but can fake it.

Would you like me to take out the garbage?

Yes! Why must I do it all the time?

Why do you feel you must do it all the time?
Half-wit systems: Which half?

You never take out the garbage!

Language: lexical and grammatical knowledge

Physical grounding:
This is a garbage can.
This is how we take out the garbage. This stinks.

Social context
Goals and beliefs
Autonomy, trustworthiness
All of which are learned

“You never X”
“Would you like me to X?”
X / take out the garbage
“Would you like me to take out the garbage?”

Built by hand
Half-wit systems: Which half?

You never take out the garbage!

Language: lexical and grammatical knowledge

Physical grounding:
This is a garbage can.
This is how we take out the garbage. This stinks.

Social context
Goals and beliefs
Autonomy, trustworthiness
All of which are learned

"You never X"
"Would you like me to X?"
"Would you like me to take out the garbage?"

The easy half!
How can we build the other half?

- Language: lexical and grammatical knowledge
- Physical grounding:
  - This is a garbage can.
  - This is how we take out the garbage. This stinks.
- Social context
- Goals and beliefs
- Autonomy, trustworthiness
  - All of which are learned

Tough! You’re the robot

You never take out the garbage!