Grounded language understanding: Overview

Christopher Potts

Stanford Linguistics

CS224u: Natural language understanding
Associated materials

1. Code
   a. Notebook: colors_overview.ipynb
   b. Homework and bake-off: hw_colors.ipynb

2. Core reading: Monroe et al. 2017

In the 1967 Stanley Kubrick movie *2001: A Space Odyssey*, the spaceship’s computer HAL can:

- display graphics;
- play chess; and
- conduct natural, open-domain conversations with humans.

How well did the filmmakers do at predicting what computers would be capable in 2001?
HAL

Graphics

HAL

Jurassic Park (1993)

Slide idea from Andrew McCallum
HAL

Chess

HAL

Deep Blue (1997)

Slide idea from Andrew McCallum
HAL Dialogue

HAL

David Bowman: Open the pod bay doors, HAL.

HAL: I’m sorry, Dave, I’m afraid I can’t do that.

David: What are you talking about, HAL?

HAL: I know that you and Frank were planning to disconnect me, and I’m afraid that’s something I cannot allow to happen.

2014

Slide idea from Andrew McCallum
Siri

You: Any good burger joints around here?

Siri: I found a number of burger restaurants near you.

You: Hmm. How about tacos?

Apple: [Siri remembers that you asked about restaurants. so it will look for Mexican restaurants in the neighborhood. And Siri is proactive, so it will question you until it finds what you’re looking for.]
Siri

Colbert: For the love of God, the cameras are on, give me something?

Siri: What kind of place are you looking for? Camera stores or churches?

[...] 

Colbert: I don’t want to search for anything! I want to write the show!

Siri: Searching the Web for “search for anything. I want to write the shuffle.”

Slide idea from Marie de Marneffe
Levinson’s (2000) analogy
Levinson’s (2000) analogy

“We interpret this sketch instantly and effortlessly as a gathering of people before a structure, probably a gateway; the people are listening to a single declaiming figure in the center. […] But all this is a miracle, for there is little detailed information in the lines or shading (such as there is). Every line is a mere suggestion […]. So here is the miracle: from a merest, sketchiest squiggle of lines, you and I converge to find adumbration of a coherent scene […].
Levinson’s (2000) analogy

“We interpret this sketch instantly and effortlessly as a gathering of people before a structure, probably a gateway; the people are listening to a single declaiming figure in the center. […] But all this is a miracle, for there is little detailed information in the lines or shading (such as there is). Every line is a mere suggestion […]. So here is the miracle: from a merest, sketchiest squiggle of lines, you and I converge to find adumbration of a coherent scene […].

“The problem of utterance interpretation is not dissimilar to this visual miracle. An utterance is not, as it were, a veridical model or “snapshot” of the scene it describes […]. Rather, an utterance is just as sketchy as the Rembrandt drawing.”
Indexicality

1. I am speaking.
2. We won. [A team I’m on; a team I support; . . . ]
3. I am here [office; Stanford; . . . planet earth; . . . ]
4. We want to go here. [pointing at a map]
5. We went to a local bar after work.
6. three days ago, tomorrow, now
Context dependence

Where are you from?
Context dependence

*Where are you from?*

- **Connecticut.** (Issue: birthplaces)
- **The U.S.** (Issue: nationalities)
- **Stanford.** (Issue: affiliations)
- **Planet earth.** (Issue: intergalactic meetings)
Context dependence

I didn’t see any.
Context dependence

- Are there typos in my slides?

I didn’t see any.
Context dependence

- Are there typos in my slides?
- Are there bookstores downtown?

I didn’t see any.
Context dependence

- Are there typos in my slides?
- Are there bookstores downtown?
- Are there cookies in the cupboard?

*I didn’t see any.*
Context dependence

- Are there typos in my slides?
- Are there bookstores downtown?
- Are there cookies in the cupboard?
- . . .

I didn’t see any.
Routine pragmatic enrichment

how big is the contextually restricted domain of students?  what’s the additional contextual restriction?

false for most students?  who’s the speaker?

Many students met with me yesterday.

what’s the time of utterance?  but perhaps many met with the speaker at other times?
SHRDLU (Winograd 1972)
SHRDLU (Winograd 1972)

Full-grounded system that parses the user’s input, maps it to logical form, interprets that logical form in its world, tries to take appropriate action, and generates responses.

http://hci.stanford.edu/winograd/shrdlu/

One project did succeed. Terry Winograd’s program SHRDLU could use English intelligently, but there was a catch: the only subject you could discuss was a micro-world of simulated blocks.
SHRDLU (Winograd 1972)

1. Person: WHAT DOES THE BOX CONTAIN?
2. Computer: THE BLUE PYRAMID AND THE BLUE BLOCK.
3. Person: WHAT IS THE PYRAMID SUPPORTED BY?
4. Computer: THE BOX.
5. Person: HOW MANY BLOCKS ARE NOT IN THE BOX?
6. Computer: FOUR OF THEM.
7. Person: IS AT LEAST ONE OF THEM NARROWER THAN THE ONE WHICH I TOLD YOU TO PICK UP?
9. Person: IS IT SUPPORTED?
10. Computer: YES, BY THE TABLE.
Humans and situated language learning

Children learn language

1. with incredible speed
2. despite relatively few inputs
3. by using cues from
   ▶ contrast inherent in the forms they hear
   ▶ social cues
   ▶ assumptions about the speaker’s goals
   ▶ regularities in the physical environment.

Frank et al. 2012; Frank and Goodman 2014
Consequences for NLU

1. Human children are the best agents in the universe at learning language, and they depend heavily on grounding.

2. Problems that are intractable without grounding are solvable with the right kinds of grounding.

3. Deep learning is a flexible toolkit for reasoning about different kinds of information in a single model, so it’s led to conceptual and empirical improvements in this area.

4. We should seek out (and develop) data sets that include the right kind of grounding.
Datasets

1. Stanford English Colors in Context Corpus [link]
2. Stanford Chinese Colors in Context Corpus [link]
3. OneCommon [link]
4. Edinburgh Map Corpus [link]
5. Cards Corpus [link]
6. Deal or No Deal? [link]
7. CraigsListBargain [link]
8. ALFRED [link]
9. CrossTalk [link]
10. Room-to-Room [link]
References


