Ling/CSE 472: Introduction to Computational Linguistics

5/26: Linguistic semantics and NLP

Overview

- English Resource Semantics
- Form & meaning (& octopusses)
- Next time: Catch-up/review

MRS & ERS

- Minimal Recursion Semantics (Copestake et al 2005): A formalism for underspecified logical forms
- English Resource Semantics (Flickinger et al 2014): MRS representations for English sentences, including many design decisions about specific semantic phenomena
- ERG Semantic Documentation: An attempt to explain those representations for consumers of them (people who use the grammar in parsing or generation)

What’s in an ERS?

- ERSes:
  - make explicit the connections between the semantic predicates introduced by the words
  - make explicit semantic predicates introduced by syntactic constructions
  - make explicit morphosemantic features such as person/number, tense/aspect, and sentential force

ERS examples: Predicate-argument structure

- The cheerful children wanted to sing and dance

$$\langle\, h_1, e_3, \left|\begin{array}{l} h_4{:}\_the\_q(x_6, h_7, \_), h_8{:}\_cheerful\_a\_1(\_, x_6), h_8{:}\_child\_n\_1(x_6), \\ h_2{:}\_want\_v\_1(e_3, x_6, h_{10}), \\ h_{14}{:}\_and\_c(\_, h_{11}, h_{16}), h_{11}{:}\_sing\_v\_1(e_{12}, x_6, \_), h_{16}{:}\_dance\_v\_1(e_{17}, x_6, \_) \end{array}\right| \{\, h_1 =_q h_2, h_7 =_q h_8, h_{10} =_q h_{14} \,\}\,\rangle$$

- This technique is impossible to apply

$$\langle\, h_1, e_3, \left|\begin{array}{l} h_4{:}\_this\_q\_dem(x_6, h_7, \_), h_8{:}\_technique\_n\_1(x_6), \\ h_2{:}\_impossible\_a\_for(e_3, h_9, \_), h_{11}{:}\_apply\_v\_2(e_{12}, \_, x_6) \end{array}\right| \{\, h_1 =_q h_2, h_7 =_q h_8, h_9 =_q h_{11} \,\}\,\rangle$$

ERS examples: Quantifiers

- All short jokes are funny v. All funny jokes are short

$$
\left\langle h_1, e_3,
\left|\begin{array}{l}
h_4\text{:\_all\_q}(x_5, h_6, \_),\\
h_8\text{:\_short\_a\_of}(\_, x_5, \_),\\
h_8\text{:\_joke\_n\_1}(x_5),\\
h_2\text{:\_funny\_a\_1}(e_3, x_5)
\end{array}\right|
\{ h_1 =_q h_2, h_6 =_q h_8 \} \right\rangle
$$

$$
\left\langle h_1, e_3,
\left|\begin{array}{l}
h_4\text{:\_all\_q}(x_5, h_6, \_),\\
h_8\text{:\_funny\_a\_1}(\_, x_5),\\
h_8\text{:\_joke\_n\_1}(x_5),\\
h_2\text{:\_short\_a\_of}(e_3, x_5, \_)
\end{array}\right|
\{ h_1 =_q h_2, h_6 =_q h_8 \} \right\rangle
$$

ERS examples: Scopal operators

- The meteorologist says it probably won't rain

$$\left\langle h_1, e_3, \left| \begin{array}{l} h_4\text{:\_the\_q}(x_6, h_7, \_), h_8\text{:\_meteorologist\_n\_1}(x_6), \\ h_2\text{:\_say\_v\_to}(e_3, x_6, h_{10}, \_), \\ h_{11}\text{:\_probable\_a\_1}(\_, h_{13}), \\ h_{14}\text{:neg}(\_, h_{15}), \\ h_{17}\text{:\_rain\_v\_1}(e_{18}) \end{array} \right| \left\{ \begin{array}{l} h_1 =_q h_2, h_7 =_q h_8, \\ h_{10} =_q h_{11}, h_{13} =_q h_{14}, h_{15} =_q h_{17} \end{array} \right\} \right\rangle$$

ERS examples: Multi-word expressions

- Kim looked up the answer

$$\langle h_1, e_3, \left| \begin{array}{l} h_4\text{:proper\_q}(x_6, h_5, \_), h_8\text{:named}(x_6, Kim), \\ h_2\text{:\_look\_v\_up}(e_3, x_6, x_9), \\ h_{10}\text{:\_the\_q}(x_9, h_{12}, \_), h_{13}\text{:\_answer\_n\_to}(x_9, \_) \end{array} \right| \{ h_1 =_q h_2, h_5 =_q h_8, h_{12} =_q h_{13} \} \rangle$$

- Kim looked up the chimney

$$\langle h_1, e_3, \left| \begin{array}{l} h_4\text{:proper\_q}(x_6, h_5, \_), h_8\text{:named}(x_6, Kim), \\ h_2\text{:\_look\_v\_1}(e_3, x_6), h_2\text{:\_up\_p\_dir}(\_, e_3, x_{10}), \\ h_{11}\text{:\_the\_q}(x_{10}, h_{13}, h_{12}), h_{14}\text{:\_chimney\_n\_1}(x_{10}) \end{array} \right| \{ h_1 =_q h_2, h_5 =_q h_8, h_{13} =_q h_{14} \} \rangle$$

Where do ERSes come from?

- Implementations of analyses of specific constructions in the English Resource Grammar

- At parse time, these various analyses interact to produce syntactico-semantic structures for input sentences

Are ERSes ‘meanings’?

- More accurately: ‘meaning representations’
- Need to be paired with a model theory/interpretation function
- Include information that goes beyond any theory logic developed to date
- For the subset that is covered by e.g. predicate logic, compatible

What are ‘fingerprints’?

- Hypothesis: recurring subparts of ERSes that can be attributed to specific grammar entities (phrase structure rules, lexical rules, lexical types) are interesting candidates for ‘semantic phenomena’
- Fingerprints are schematized ERS pieces that should match the ERS for any sentence evincing the phenomenon they illustrate
- In principle, fingerprints can be used to search sembanks of sentences annotated with ERSes
- We hope that an explanation of ERG semantic analyses centered on fingerprints will make the representations more interpretable to non-grammar developers

Reading questions

- Functional question: how do we read the semantic.. .things? With all the letters and variables etc.

- What do the indices for all the italic 'h's mean?

- Also confused about reading semantic fingerprints. Especially why the arguments seem to appear twice? For example, the x1, x2 in N-N compounding:

```
h0:compound[ARG1 x1, ARG2 x2]
h0:[ARG0 x1]
[ARG0 x2]
```

Reading questions

- I am curious about how systems like ErgSemantics handle Metaphors and Idioms. Do they handle the sentence as is or try to figure out the hidden meaning behind them?

- In the ERG Semantic Documentation, the author notes that NLP tasks rarely exercise the kind of inference by the proper treatment of quantifiers. What kind of NLP tasks would highlight the importance of this sensitivity to quantifiers?

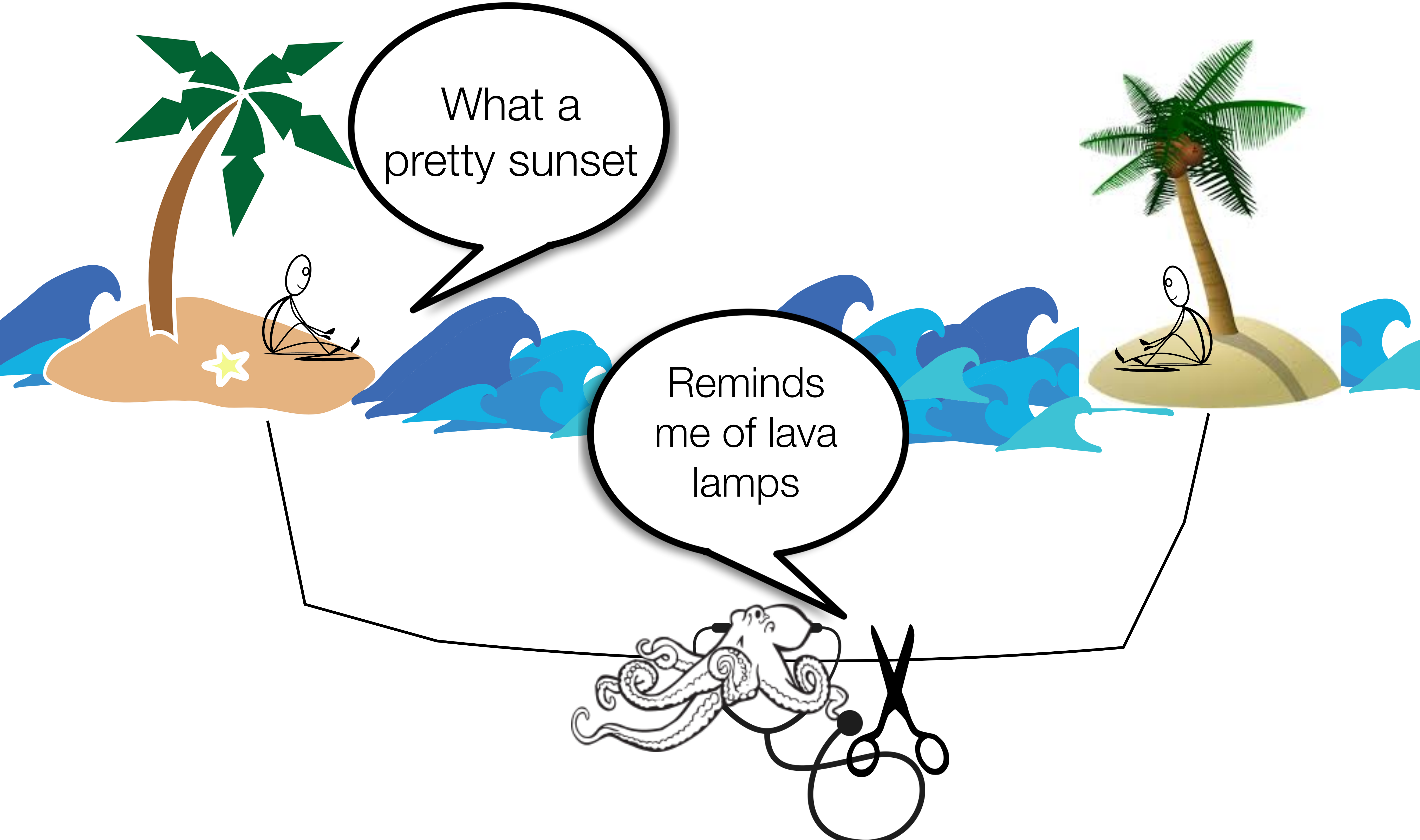
Bender & Koller 2020: Definitions

- Linguistic form: movements of articulators, marks on a page, pixels, or bytes
- Meaning: The relation between form and something external to language
    - $M \subseteq E \times I$ — the relation M between pairs of expressions *e* and communicative intents *i*
    - $C \subseteq E \times S$ — the relation C between pairs of expressions *e* and standing meanings *s*
- Language model: System trained only on the task of predicting linguistic form

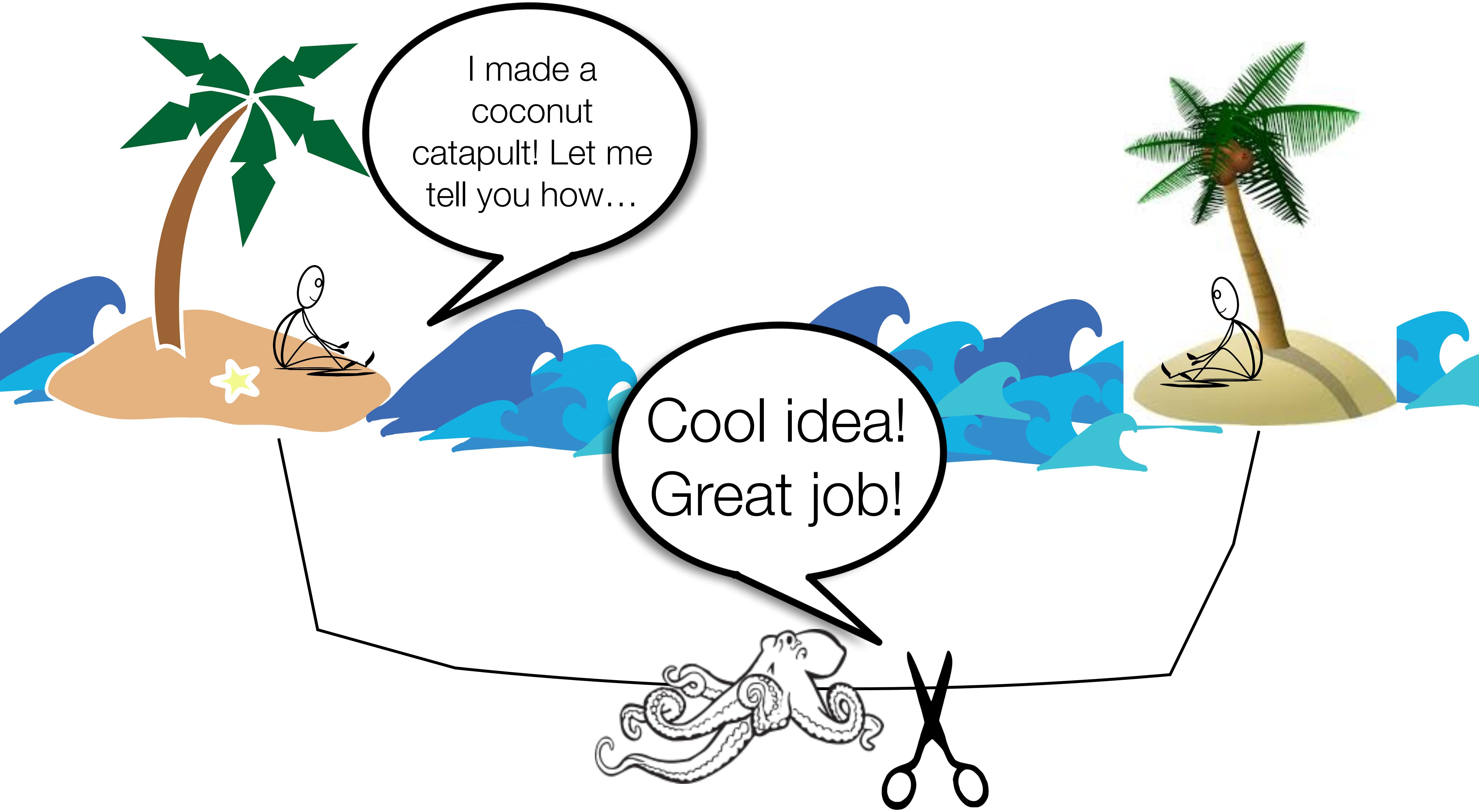
Bender & Koller 2020: Key claims

- If the information isn't in the training data, then an ML system can't learn it
- Training data that consists of linguistic form doesn't include information about the meaning — it only seems to if we view it as speakers of the language in question
- Therefore, language models cannot in principle learn meaning
- However, because speakers are very keen to *make sense of* linguistic forms they encounter, it can seem like they do

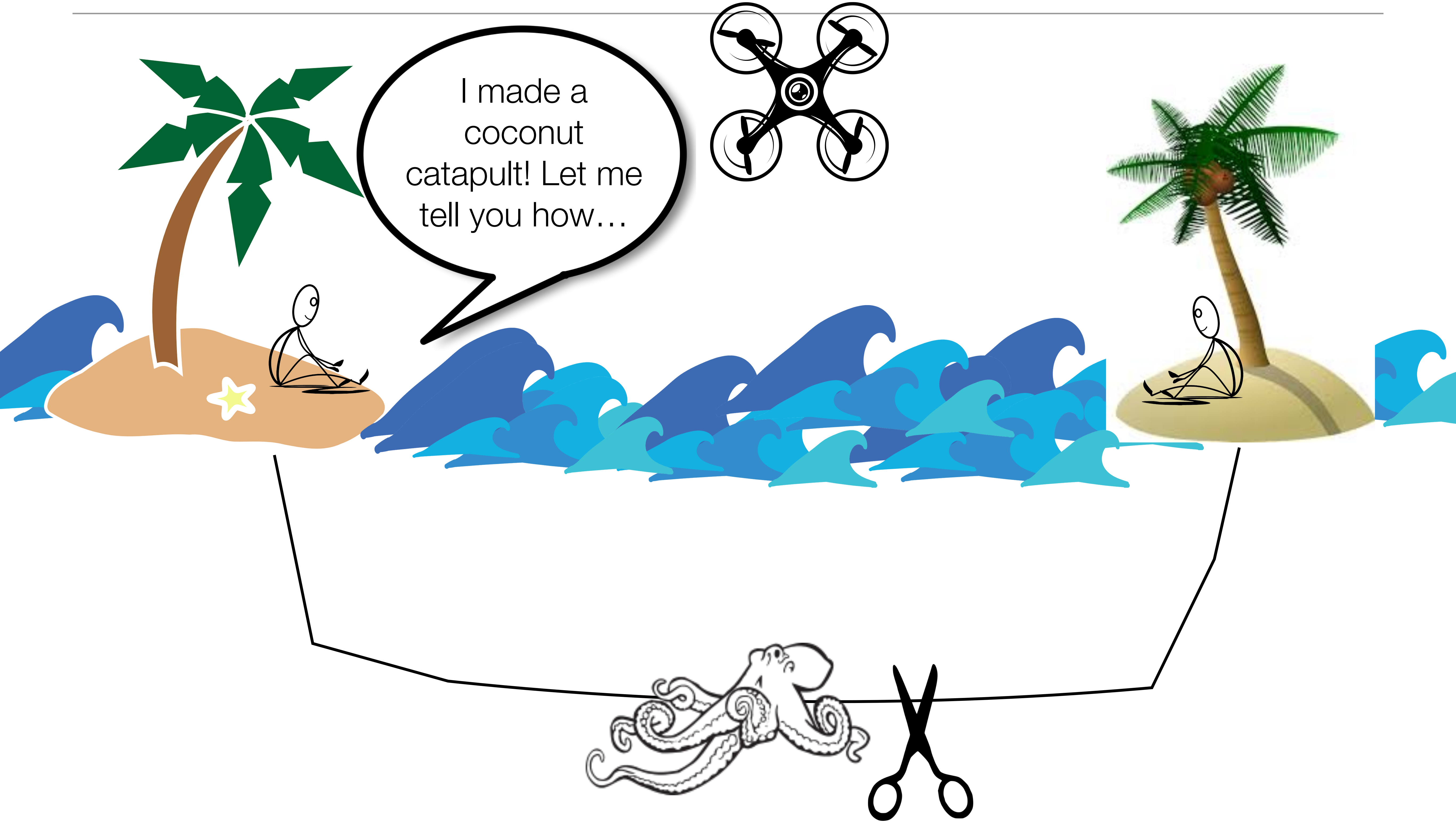
Thought experiment: Meaning from form alone



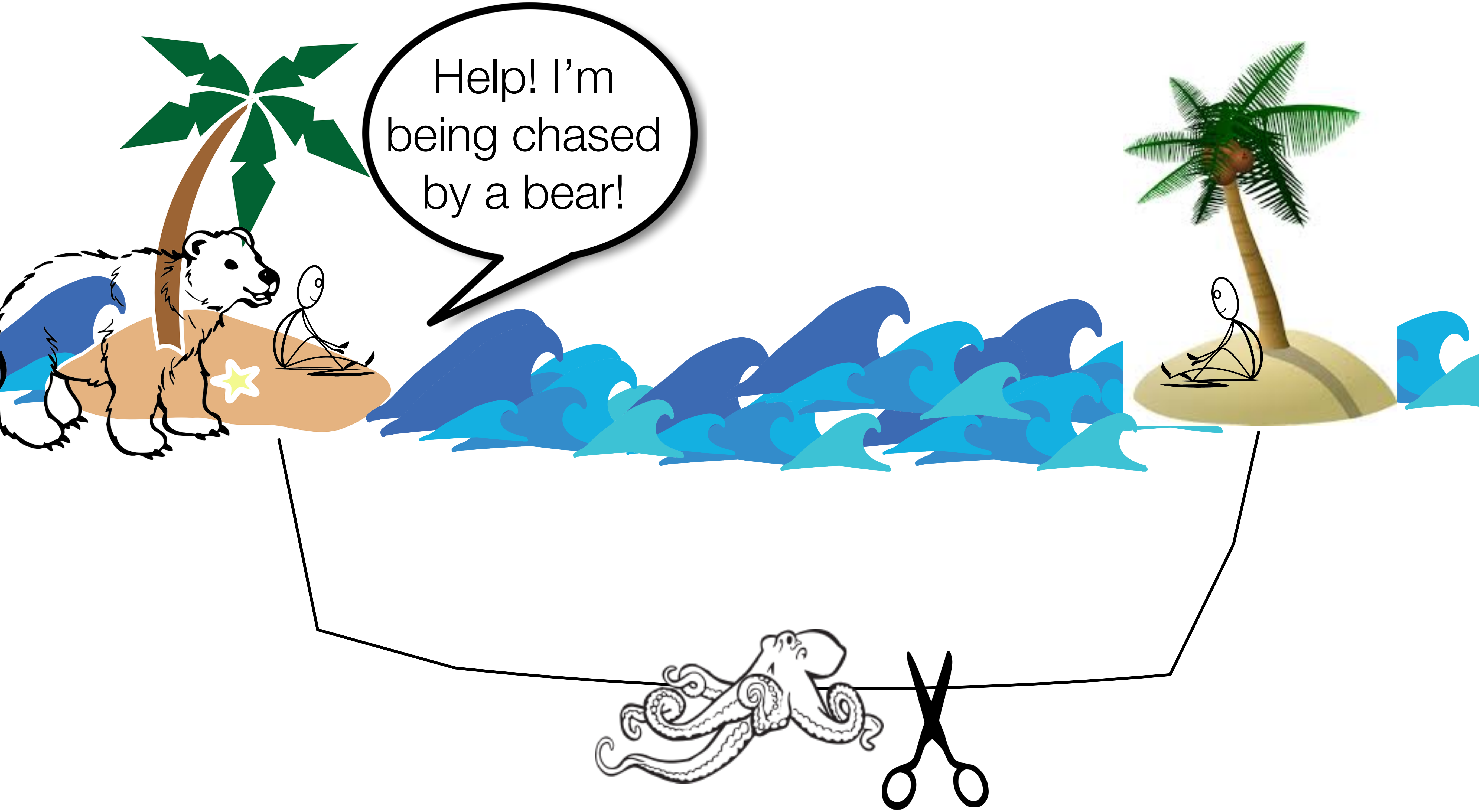
Thought experiment: Meaning from form alone



Thought experiment: Meaning from form alone



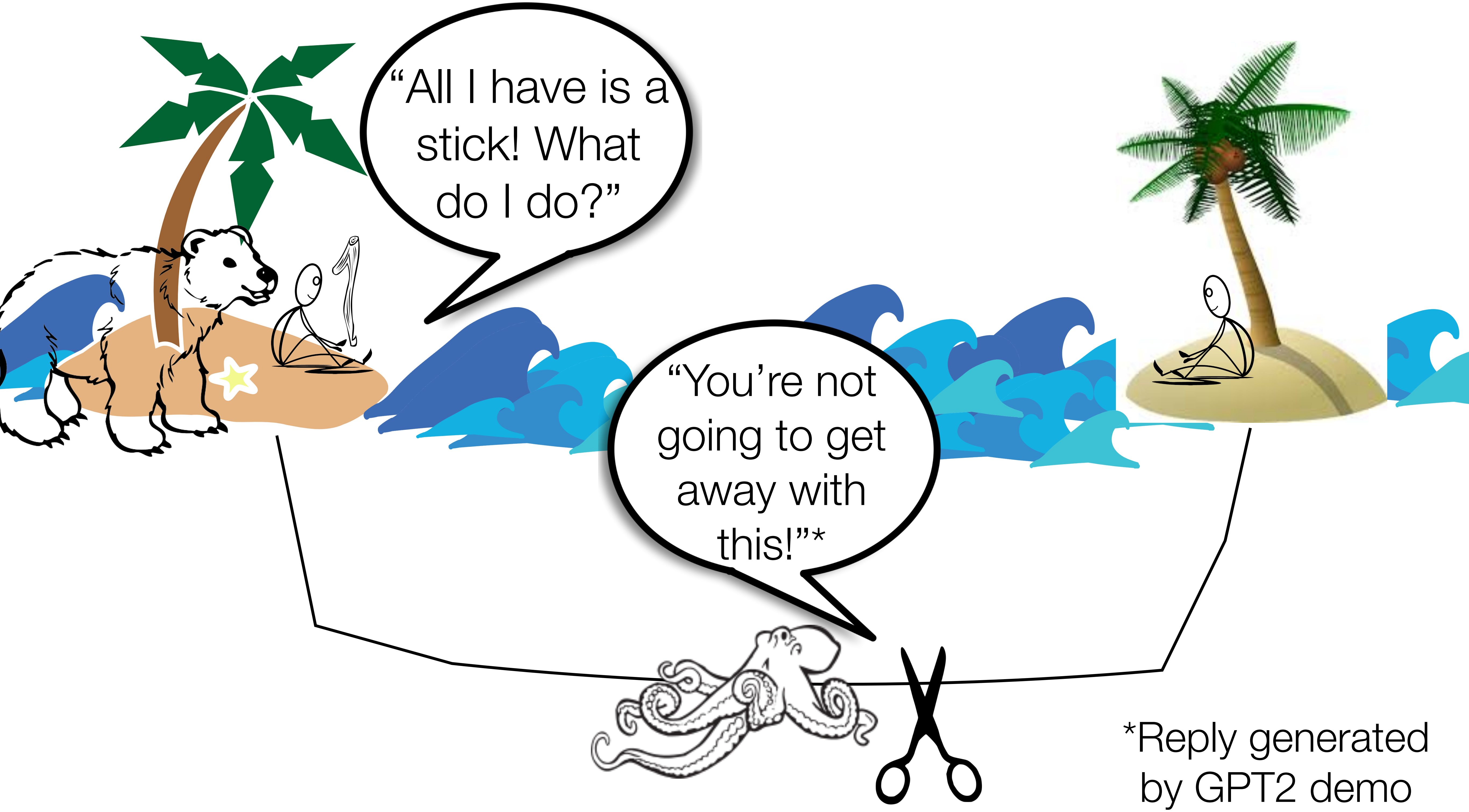
Thought experiment: Meaning from form alone



Thought experiment: Meaning from form alone

*Reply generated by GPT2 demo

Thought experiment: Meaning from form alone



*Reply generated by GPT2 demo

Octopus test v. Turing test v. Chinese room

- Turing: A machine can be said to “think” if it can fool a human judge into think it’s human after an arbitrary conversation
- Searle: Imagine a non-Chinese speaking person inside a room receiving messages written in Chinese and responding according to some perfect set of rules in a sensible way. No actual “understanding” is taking place there.
- Current NLP: Trying to build something like Searle’s Chinese room without ever having had access to what things mean (cf. Harnad’s “symbol grounding problem”)

Reading questions

- Do computational linguists ever argue that the terminology related to "learning/ comprehension" actually applies? Is there a divide in the field, or are there ever certain areas of agreement?

    - https://medium.com/huggingface/learning-meaning-in-natural-language-processing-the-semantics-mega-thread-9c0332dfe28e

- Isn't meaning the same thing as intent? Or what does it mean to take the product of E and I? And why is "understanding" the process of finding I given E, rather than finding M given E? I found it a little unclear. The reading was fairly understandable otherwise, though.

- I'm not sure that I completely understand what is meant by communicative intent. How are inanimate objects like bank accounts or computer filer systems considered to be communicative intents? Would these not be communicative intents if no human initiated the communication?

Reading questions

- "In addition, certain tasks are designed in a way that specific forms are declared as representing certain semantic relations of interest." Does this mean, for example in the summarization task, that texts are paired with their summaries, which creates an explicit semantic relationship between the two?

Reading questions

- The paper asserts that meaning is "the relation between the form and something external to language", such that each member of the set of meanings is pairs of (e, i) where e is a language expression and i is the external intent. But how can one be sure that this captures all that there are? Can this encompass the entirety of "the meanings" that are alleged here to be the purpose of human communication? (e.g. What about the activity of talking to oneself? I do that sometimes to try and clearly state my thought, that could be my intent of the activity itself, but not the content of the said sentences. Would this still be applicable to this model of meaning?)

Reading questions

- This is less of a conceptual question but I'm interested by this "octopus test". Is this a made up thing that just uses an octopus hence the name, or can it be called like "the cat test" by replacing the octopus with the cat? Or is there a background story with this test?

- Considering the octopus illustration, O will never be on A or B's islands and never experience what they're talking about, thus according to the paper's argument O will never understand the meaning of the words. Is it even a realistic goal to create an AI that understands meaning if it cannot experience and thus learn meaning? Would we have to 'create artificial life' that can sense and experience the actual world for it to 'understand'?

- The "octopus test" section states the following: "Without access to a means of hypothesizing and testing the underlying communicative intents, reconstructing them from the forms alone is hopeless." What is meant by 'testing' here? Is it referring to the Turing test exclusively?

Reading questions

- I like the analogy of climbing the right hill (it's fittingly reminiscent of finding local vs. global optima in machine learning). But even if we collectively realize through the proposed diagnostics that we're not climbing the right hill, in what ways can that help us move closer to the end goal? There may not exist a "right" hill which can actually take us to the end goal, or we may simply not know/have enough to climb it. In other words, what do you think would be the next steps?

Reading questions

- This is probably an unanswerable question, but after reading the Bender & Koller paper, I am wondering if language models will ever be able to "escape" the Chinese Room. Even if we give a model more than form, such as images and associations with concrete objects, will the computer ever be doing anything more than looking at instructions and spitting out answers?
- If a neural network or language model becomes so sophisticated that it can produce semantically and pragmatically coherent sentences and responses based on its training data, wouldn't there be no point in learning meaning from form?

Reading questions

- From section 6, it mentions that "human children do not learn meaning from form alone and we should not expect machines to do so either". Not sure if I am misinterpreting something, but once machines can learn by interacting with people, then they will be able to learn? I am bit confused, because the paper also mentions that this concept of "learning" does not apply to LMs that are focused on comprehension, for example.

- I was wondering if anyone has tried to induce some kind of semantic recognition, at least, by training a model (as long as we're staying in a statistical learning space) for a primary task that's not all language-internal while requiring that some of the instruction, at least, is presented in reasonably non-repetitive natural language.

  - Bisk et al 2020: https://arxiv.org/abs/2004.10151

Reading questions

- Is it possible, then, to conceive such a system that can attempt this? In other words, do we understand enough of the theory of computation and languages to design a system that can perform the "understanding" of language? Do we even know enough about the nature of "understanding" itself? If so what is it?
- "... avenue towards human-analogous NLU." - I keep thinking of those dystopian movies about robots gaining consciousness and taking over the world, and some people probably legitimately fear this. Is it possible to create a system/model that can work like human NLU? Understanding that doesn't solely rely on the systematicity of language? We talked about systems working with a knowledge base, but then how do you really encode all the complexities of meaning in human language?

Reading questions

- For tech driven application of NLP, like Amazon's Alexa, is human-analogous NLU necessarily or even desirable? At least in the short term, a bottom-up take on NLP advancements seems to be working well for tech companies. Would Amazon specifically (or other companies) benefit from from a top-down perspective, or otherwise focusing on NLU?

Reading questions

- What are the possible consequences of people buying into the "AI hype", on both the consumer side and the developer side? Without getting too space Odyssey, /HAL/ can we disprove that machines will ever be able to truly understand meaning? It was mentioned that a two way communication was required for someone to truly acquire language, and that a machine just learning form wouldn't have this aspect of input, but isn't the only thing differentiating an octopus from the thought experiment, and a child, just knowledge of the world? If it too understood what a rope referred to, why couldn't it build the catapult? Sorry if this question is too blade runner / ghost in the shell. (If you haven't watched the originals, you /odyssey/ it. Sorry I'll just blade run on out of here).

Overview

- English Resource Semantics
- Form & meaning (& octopusses)
- Next time: Catch-up/review