Climbing towards NLU:

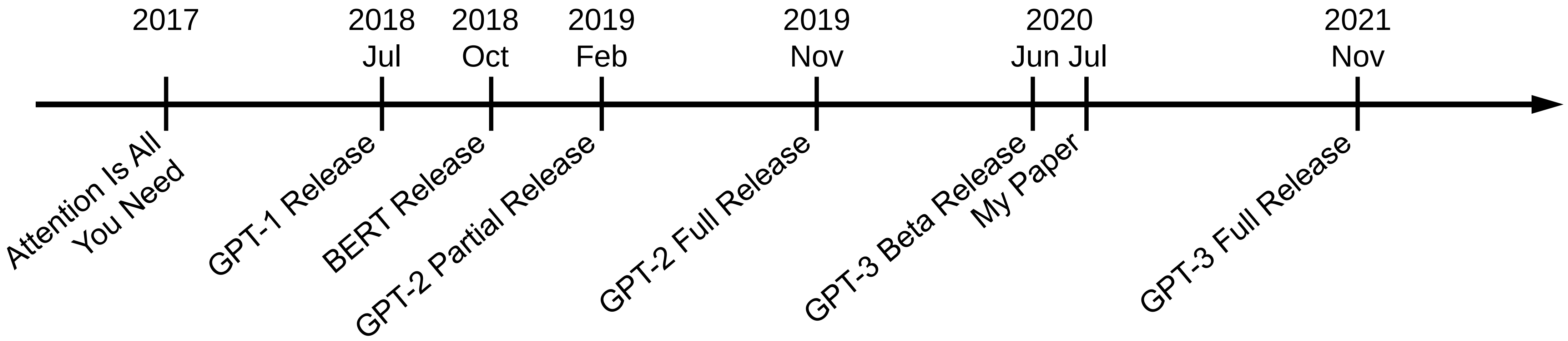
On Meaning, Form, and Understanding in the Age of Data

Emily M. Bender
University of Washington

Alexander Koller
Saarland University

Presented by Elena Bösch
22. May 2024

Timeline


2017
2018 Jul
2018 Oct
2019 Feb
2019 Nov
2020 Jun Jul
2021 Nov
Attention Is All You Need
GPT-1 Release
BERT Release
GPT-2 Partial Release
GPT-2 Full Release
GPT-3 Beta Release
My Paper
GPT-3 Full Release

"BERT is a system by which Google's algorithm uses pattern recognition to better **understand** how human beings communicate so that it can return more relevant results for users." [B2C]

"In order to train a model that **understands** sentence relationships, we pre-train for a binarized next sentence prediction task." (Devlin et al., 2019)

"Here are some of the examples that showed up our evaluation process that demonstrate BERT's ability to **understand** the intent behind your search." [Google Blog]

"Using BERT, a pretraining language model, has been successful for single-turn machine **comprehension**…" (Ohsugi et al., 2019)

"Our BERT models, on the other hand, **understand** that "stand" is related to the concept of the physical demands of a job, and displays a more useful response." [Google Blog]

"The surprisingly strong ability of these models to **recall factual knowledge** without any fine-tuning demonstrates their potential as unsupervised open-domain QA systems." (Petroni et al., 2019)

Main Thesis

“A system trained only on form has *a priori* no way to learn meaning.”

Contents

Linguistic Background

- Def: *Language model*
  - Any system trained only on the task of string prediction
  - May operate over characters, words or sentences
  - May operate sequentially or not

- Def: *form*
  - Any observable realization of language
  - E.g. marks on a page, pixels/bytes in a digital representation of text, movements of the articulator (e.g. vocal tract, or hands & face for a sign language)

Linguistic Background

- Def: *communicative intent*
  - What we're trying to achieve when using language
  - E.g. convey information, ask someone to do something, socialize, etc.
  - Is about something *outside* of language

- Def: *meaning*
  - The relation between form and communicative intent
  - Formally: $M \subset E \times I$, contains pairs (e, i) of natural language expressions e and communicative intent i

- Def: *understand*
  - Understanding is the process of retrieving i given e

Linguistic Background

- Def: *conventional/standing meaning*
  - That which is constant across all possible contexts of use
  - An abstract object that represents the communicative potential of a form
  - Must have interpretations

- Def: *linguistic system*
  - The relation between form and standing meaning
  - Formally: $C \subset E \times S$, contains pairs (e, s) of natural language expressions e and their standing meaning s

Linguistic Background

Example:

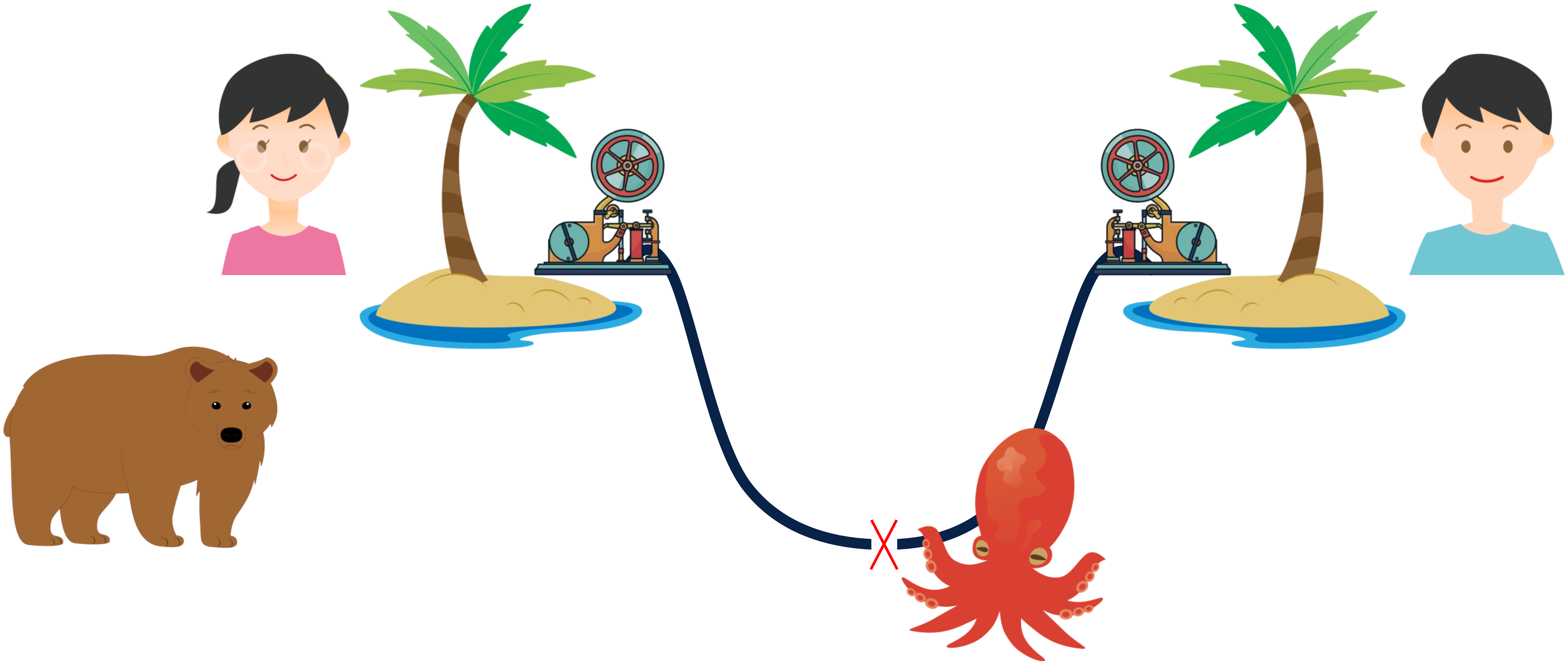
- Speaker has communicative intent i
- Chooses expression e with standing meaning s to express i
- Listener hears e and reconstructs s
- Uses knowledge and hypotheses to (try to) deduce i

$\Longrightarrow$ Active participation of listener is crucial!

Main Thesis: Formally

"A model of natural language that is purely trained on form will not learn meaning: if the training data is only form, there is **not sufficient signal** to learn the relation M between that form and the non-linguistic intent of human language users, nor the relation C between form and the standing meaning the linguistic system assigns to each form."

Thought Experiment:
The Octopus Test



Telegraph art from www.freevector.com
Other artwork from creazilla.com

Thought Experiment:
Java

- Def: *Java*
  - The relation $J \subset E \times I$, containing pairs (e, i), where e is a Java program, and i is the function that maps inputs to the outputs produced when compiling and executing the program e on the Java Virtual Machine

- Thought Experiment:
  - Train LM on all well-formed Java code on GitHub
  - Only code, no bytecode, compiler, no sample inputs and outputs
  - Task: ask the model to execute a program, expect correct output

  $\Longrightarrow$ Ridiculous!

- What makes the task impossible?
  - Form of Java programs has no information on how to execute them

Possible Counterarguments:

"But human children can acquire language just by listening to it."

- Not supported by scholarly work!
  - Children don’t pick up a language from passive exposure such as TV/radio
  - English-learning infants can learn Mandarin phonemic distinctions from brief interactions with a Mandarin-speaking experimenter but not from exposure to Mandarin TV or radio.
  - Toddlers (18–20 months old) don’t pick up labels uttered by someone behind a screen, but do pick up labels uttered by someone performing joint attention with them

- Acquiring a linguistic system needs joint attention and intersubjectivity

  $\Longrightarrow$ being aware of what another human is attending, guessing what they are intending to communicate

  $\Longrightarrow$ grounded in physical world and interaction with other people

Possible Counterarguments:
“But ‘meaning’ doesn’t mean what you say it means.”

- A working definition, as general as possible
- Captures Main points: meaning is based on link between linguistic form and something that is not language
- Cannot ignore core function of language: conveying communicative intents

Possible Counterarguments:
“But meaning could be learned from…”

- Augmenting form with grounding data
  - Idea: as learner gets access to more information on top of text itself, it can learn more facets of meaning (Bisk et al., 2020)
- Declaring specific forms as representing certain semantic relations
  - E.g. NLI datasets
- Control codes or tokens
  - E.g. TL;DR

For Java: Unit tests!

Possible Counterarguments:

“But BERT improves performance on meaning related tasks, so it must have learned something about meaning.”

- Research findings
  - Mirroring distribution $\Longrightarrow$ Performance falls back to chance (Niven & Kao, 2019)
  - Frustrating heuristics $\Longrightarrow$ Performance falls to below chance (McCoy et al., 2019)

- What *has* BERT learned?
  - Incomplete reflection of meaning
  - potentially useful
  - Could just be artifacts in the data

Related Work

- Sahlgren and Carlsson. *“The Singleton Fallacy: Why Current Critiques of Language Models Miss the Point.”*
- Piantadosi and Hill. *“Meaning without reference in large language models.”*
- Michael. *“To Dissect an Octopus: Making Sense of the Form/Meaning Debate”*
- Merrill et al. *“Provable Limitations of Acquiring Meaning from Ungrounded Form: What Will Future Language Models Understand?”*
- Tsimpoukelli et al. *“Multimodal Few-Shot Learning with Frozen Language Models.”*

Are we climbing the right hill?

Bottom-up vs Top-down perspective

Bottom-up

- Any result that (partially) solves a specific challenge is a success
- Progress, as long as such successes are frequent

Top-Down

- Focus is on the remote end goal
- End goal: offering a complete, unified theory
- Questions whether: Does all our bottom-up progress lead us in the right direction?

In NLP:

- Many tasks are being improved significantly

$\Longrightarrow$ Things are going great!

In NLP:

- Is the hill we are climbing so rapidly the right hill?
- How do we know that incremental progress on today's tasks will take us to our end goal?

Best practices to ensure we're climbing the right mountain

1. Cultivate humility towards language and ask top-down questions
2. Be aware of the limitations of tasks
3. Value and support the work of carefully creating new tasks
4. Evaluate models of meaning across tasks
5. Perform thorough analysis of both errors and successes

My Opinion

Positives

+ Important topic
+ Interesting perspective
+ Easy to understand / accessible
+ Best practice suggestions

Negatives

– No formal arguments
– LMs don't *have* to reason like humans

Thank you for your attention!

Any Questions?

Thought Experiment:
English and unlabeled photos

- Train LM on English text and very large collection of unlabeled photos
- No indications of speaker intent, no connection between text and photos
- Text Training Task: predicting form
- Image Training Task: anything that only involves the image data

- Test Task: given a photo, answer a question concerning the picture, e.g.
  - "How many dogs in the picture are jumping?“

- What makes the tasks impossible?
  - Form of English has no information about [what language-external entities are referring to]

Possible Counterarguments:
“But there is so much form out there – surely that’s enough.”

- We showed: learning meaning requires more than form
  $\Longrightarrow$ amount is not relevant
- with enough form, can achieve high scores in evaluation, but not enough for human-analogous NLU

Possible Counterarguments:
“But aren’t neural representations meaning too?”

- NNs capture aspects of meaning (e.g. semantic similarity)
- But cannot be standing meaning, or communicative intent

GPT-2 on fighting bears with sticks

Robinson cried out, “Help! I’m being chased by a bear! All I have is these sticks. What should I do?” The survivalist said, “Here’s what you can do. Take one stick and punch the bear, and then run faster to the store. Wait there until the bear is gone, and then give her all of the sticks. Then go back inside and get your gun.” Advertisement Bobby told me that she began to cry. But when she was finished taking down the bear, she marched out into the yard, began collecting up the sticks and armed herself. She said she felt like a heroic hero and a Yankee heroine. It turns out that bears have more than their fair share of venomous snakes, bees

Robinson cried out, “Help! I’m being chased by a bear! All I have is these sticks. What should I do?” The survivalist said, “Here’s what you can do. Go back to civilization. Take your gun. Pick up some sticks and try to kill the bear. If you don’t kill it, then run.” Robinson said, “I do not think that the Bear is going to attack. It’s just out of my way. I’m going to take my bag. I’m going to leave my things in there. I’ll try to hide out there. I’ll call in the friends.” Two other little boys went out onto the

References

- Merrill, William et al. “Provable Limitations of Acquiring Meaning from Ungrounded Form: What Will Future Language Models Understand?” *Transactions of the Association for Computational Linguistics* 9 (2021): 1047-1060.
- Sahlgren, Magnus and Fredrik Carlsson. “The Singleton Fallacy: Why Current Critiques of Language Models Miss the Point.” *Frontiers in Artificial Intelligence* 4 (2021): n. pag.
- Piantadosi, Steven T. and Felix Hill. “Meaning without reference in large language models.” *ArXiv* abs/2208.02957 (2022): n. pag.
- Michael, Julian. “To Dissect an Octopus: Making Sense of the Form/Meaning Debate” *Blog of Julian Michael* (2020). Retrieved May 20, 2024, from https://julianmichael.org/blog/2020/07/23/to-dissect-an-octopus.html#top
- Tsimpoukelli, Maria et al. “Multimodal Few-Shot Learning with Frozen Language Models.” *Neural Information Processing Systems* (2021).
- Niven, Timothy and Hung-Yu Kao. “Probing Neural Network Comprehension of Natural Language Arguments.” *ArXiv* abs/1907.07355 (2019): n. pag.

References

- McCoy, R. Thomas et al. “Right for the Wrong Reasons: Diagnosing Syntactic Heuristics in Natural Language Inference.” *Annual Meeting of the Association for Computational Linguistics* (2019).
- Devlin, Jacob et al. “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding.” *North American Chapter of the Association for Computational Linguistics* (2019).
- Ohsugi, Yasuhito et al. “A Simple but Effective Method to Incorporate Multi-turn Context with BERT for Conversational Machine Comprehension.” *ArXiv* abs/1905.12848 (2019): n. pag.
- Petroni, Fabio et al. “Language Models as Knowledge Bases?” *Conference on Empirical Methods in Natural Language Processing* (2019).
- Bisk, Yonatan et al. “Experience Grounds Language.” *ArXiv* abs/2004.10151 (2020): n. pag.
- Brooke, Connor. “What to Do About BERT & Google’s Recent Local Algorithm Update” *Business2Community* (2019). Retrieved May 19, 2024, from https://www.business2community.com/search-engine-optimization/what-to-do-about-bert-googles-recent-local-algorithm-update-02259261
- Nayak, Pandu. “Understanding searches better than ever before” *Google Blog* (2019). Retrieved May 19, 2024, from https://blog.google/products/search/search-language-understanding-bert/