LLMs: the Linguistic Perspective

Micha Elsner
Department of Linguistics

THE OHIO STATE UNIVERSITY
Chomsky on LLMs

We have to ask here a certain question. Is it engineering, or is it science? Engineering, in the sense of just trying to build something that's useful, or science, in the sense that it's trying to understand something about elements of the world...

We can ask that question, is it useful? Yeah, it's pretty useful. I use Google Translator. So, on engineering grounds it's kinda worth having, like a bulldozer. **Does it tell you anything about human language? Zero, nothing.** And in fact, it's very striking. From the very beginning it's just totally remote from science.

-Noam Chomsky, in 2019 interview with Lex Fridman
Chomsky doesn’t speak for everyone

These models have opened the space of plausible linguistic theories, allowing us to test principles beyond the ones that have traditionally concerned linguists. They allow us to finally develop compelling theories of the interplay of structure and statistics. And they appear to solve many problems that generative syntacticians have worried about, but without using any of their theoretical tools and constructs. Large language models rewrite the philosophy of approaches to language.

- Steve Piantadosi, “Modern language models refute Chomsky’s approach to language” (2023)
Debates and perspectives

**Applications:** engineering tasks that linguists care about

**Philosophical puzzles:** is it fair to say LLMs “understand”? 

**LLMs as cognitive models:** how are LLMs like humans? Can they help us understand cognition?

**Where do we go from here?**
Engineering for linguists

Documentation and education for endangered languages

(Elsner and Needle “Translating a low-resource language using GPT-3 and a human-readable dictionary” 2023)

Finding examples in corpora (smart search)

Interpreting brain recordings
Inuktitut: a polysynthetic language

Long words composed of many content and functional morphemes:

\[
\text{saviggirunnaqtutit}
\]

savik- ggiq -runnaq -tutit
snow.knife -bring -ABIL -2S

“You can bring your knife.”

An inexperienced speaker cannot easily use individual morpheme definitions, or even the gloss, to understand such a word.
Existing dictionary: human-readable text in English

Example

saviggirunnaqtutit
« You can bring your knife. »

Related Morpheme:

savlk
« (1) metal; steel; iron (2) snow-knife »

-ggir-
« to bring someone or something along »

-runnaq-
« to be able to perform a certain action »

-tutit
« you »
The Inuktitut word saviggirunnaqtutit is made up of the following parts:

savik: (1) metal; steel; iron (2) snow-knife; or, to be fitted with a metal point (harpoon; spear)
ggiq: to bring someone or something along, as in "Don't bring your gun along."
runnaq: to be able to perform a certain action, as in "Could you find that out if he/she arrived"
tutit: you, as in "you sleep"

In English, saviggirunnaqtutit means roughly: **you are able to bring a snow-knife**
<table>
<thead>
<tr>
<th>Baselines</th>
<th>BLEURT (Sellam et al. 2020)</th>
<th>Translation</th>
</tr>
</thead>
<tbody>
<tr>
<td>No definitions</td>
<td>13</td>
<td>We are thankful.</td>
</tr>
<tr>
<td>Concatenate</td>
<td>43</td>
<td>(1) metal; steel; iron (2) snow-knife, to be fitted with a metal point (harpoon; spear) to bring someone or something along to be able to perform a certain action you</td>
</tr>
</tbody>
</table>

GPT-3 does **not** know Inuktitut.

Text-to-text generation is required.
<table>
<thead>
<tr>
<th>Systems</th>
<th>BLEURT (Sellam et al. 2020)</th>
<th>Translation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Definition only</td>
<td>48</td>
<td>are you able to bring a snow-knife?</td>
</tr>
<tr>
<td>Definition + usage example</td>
<td>51</td>
<td>you are able to bring a snow-knife</td>
</tr>
<tr>
<td>Ask for multiple answers (average)</td>
<td>46</td>
<td>you are able to bring metal along</td>
</tr>
<tr>
<td>Ask for multiple answers (oracle best)</td>
<td>59</td>
<td>you are able to bring metal along</td>
</tr>
<tr>
<td>Grammar lesson</td>
<td>48</td>
<td>you are able to bring a snow-knife</td>
</tr>
<tr>
<td>Chain-of-thought</td>
<td>43</td>
<td>you can make it</td>
</tr>
</tbody>
</table>

Definitions help!

Asking for multiple answers is useful if the user has contextual cues for which one is right.

GPT-3 is not great at applying grammatical information to concrete examples.
Engineering for linguists

Even Chomsky agrees this area has a bright future.

Linguists will use these new technologies to do interesting things.

But do LLMs qualify as a scientific breakthrough in linguistics beyond their applications?
Do LLMs really “understand” language?

Human language is **grounded** (sensory experience of the physical world) and largely **referential** (mapping between language and physical entities).

“Vector models are purely language-internal, solipsistic models of meaning.”

-Lazaridou, Bruni and Bertoni “Is this a wampimuk? Cross-modal mapping between distributional semantics and the visual world” (2014)
The skeptics: the case of the hyper-intelligent octopus

Bender and Koller “Climbing towards NLU: On Meaning, Form, and Understanding in the Age of Data” (2020)

How is your day going?

Still stuck on an island. You?
The skeptics: the case of the hyper-intelligent octopus

How is your day going?

It’s sunny and I’m eating coconuts.

Bender and Koller “Climbing towards NLU: On Meaning, Form, and Understanding in the Age of Data” (2020)
The skeptics: the case of the hyper-intelligent octopus

Bender and Koller “Climbing towards NLU: On Meaning, Form, and Understanding in the Age of Data” (2020)

Have you tried building a coconut catapult? Here’s the plan… let me know if it works.

It works! // It doesn’t, sorry.
Claim and counter-claim

“We take meaning to be the relation between the form and something external to language… a model of natural language that is trained purely on form will not learn meaning.” - Bender and Koller (2020)

On the other hand, some philosophers believe speech communities create meaning:

“What we should worry about is not grounding but instead whether LMs are part of a relevant linguistic community. Unlike in the case of grounding… prima facie considerations tell in favor of counting LMs as part of our linguistic community, just because it looks like they are part of our speech community: they speak and interact in linguistically ordinary ways.” - Mandelkern and Linzen “Do Language Models Refer?” (2023)
Concept networks as creators of meaning

“We see a complicated network of similarities overlapping and criss-crossing: similarities in the large and in the small… likewise the kinds of number, for example, form a family. Why do we call something a “number”? Well, perhaps because it has a direct affinity with several things that have hitherto been called “number”; and this can be said to give it an indirect affinity with other things that we also call “numbers”. And we extend our concept of number, as in spinning a thread we twist fibre on fibre.”

-Wittgenstein “Philosophical Investigations” (1946)

Understanding can be partial: a student can count to 10 but perhaps not to 100, or perhaps does not understand the idea of infinity.
Interim conclusion: LLMs “understand” in a partial way

We can usefully look for similarities between LLMs and human language processing:

    To discover what kinds of language are learnable

    To understand the role of prediction in language processing

    To look for latent linguistic structures such as phrases

But we have to watch out for dissociations!
Do LLMs do what humans are doing?

Yes: latent syntactic structure and other linguistic interpretations

(Manning et al. “Emergent linguistic structure in artificial neural networks trained by self-supervision” 2020, Linzen and Baroni “Syntactic structure from deep learning” 2021, and others)

No: uncertainty and reading time prediction

(Oh and Schuler “Why Does Surprisal From Larger Transformer-Based Language Models Provide a Poorer Fit to Human Reading Times?”, “Transformer-Based LM Surprisal Predicts Human Reading Times Best with About Two Billion Training Tokens”, both 2023)
Core assumption

Human language processing in real time depends on **predictability** of words in sequence.

Readers look longer and harder at difficult words.

<table>
<thead>
<tr>
<th>Sentence</th>
<th>Are</th>
<th>tourists</th>
<th>enticed</th>
<th>by</th>
<th>these</th>
<th>attractions</th>
<th>threatening</th>
<th>their</th>
<th>very</th>
<th>existence?</th>
</tr>
</thead>
<tbody>
<tr>
<td>$r_{1,w}$</td>
<td>Fixate</td>
<td>Fixate</td>
<td>Fixate</td>
<td>Skip</td>
<td>Fixate</td>
<td>Fixate</td>
<td>Fixate</td>
<td>Skip</td>
<td>Fixate</td>
<td>Fixate</td>
</tr>
<tr>
<td># skips</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>3</td>
<td>0</td>
</tr>
</tbody>
</table>
Standard result: better LMs mean better fits

For decades, there was a robust correlation between:

- LMs’ fit to data (high probability / low perplexity of text)
- LMs’ fit to human reading time (find the same words difficult as humans)

Fossum and Levy “Sequential vs. Hierarchical Syntactic Models of Human Incremental Sentence Processing” (2012)
This isn’t true anymore

← Probability of data

Oh and Schuler 2023
Training data size, not model size, is critical
Superhuman world knowledge

“On the subset of named entities [such as Elvis Presley], which emerged as the top two subsets across all corpus-by-LM combinations, the larger LM variants show systematically higher SSEs due to underprediction… For language models, named entity terms that typically consist of multiple tokens have high mutual information, making it easy for them to accurately predict subsequent tokens given the first (e.g., Presley given Elvis), resulting in especially lower surprisal estimates for larger LM variants.”
World knowledge from text instead of grounding

Due to rich grounding, humans learn common-sense inference without the need for massive amounts of text.

Life experience enables reasoning about world scenarios ("The snow had piled up on the drive so high that they couldn’t get the car out. When Albert woke up, his father handed him a ____") which small LLMs cannot do.

Where to?

Philosophically fair to say LLMs “understand” some kinds of language.

Representations are **networks of concepts** which correspond in some respects to human cognitive representations used by speech communities…

But also differ, especially where grounding is most important.

Interpretation of what LLMs are doing can help us to understand human language processing, but we shouldn’t assume the two are isomorphic.
Where to?

LLMs already demonstrate the *existence* of algorithms that learn linguistic structure from raw data.

This is a critical counter-claim to traditional arguments that language is unlearnable without very strong priors.

But it doesn’t prove that humans learn the way LLMs do!

Especially with the huge differences in data required, this is likely not true.