What we can do with what we know about how language works

Emily M. Bender
University of Washington
“I’m a linguist, so I know how language works…”

Clarion West presents
WISHFUL THINKING AND AI
Ted Chiang & Dr. Emily M. Bender in Conversation

A fundraiser for Clarion West: Join us for an evening with two of Time Magazine’s 100 Most Influential People in AI as they discuss the hype, realistic expectations, and whose voices are missing in this rapidly changing field.

Friday, November 10 at 7:30 PM
Town Hall Seattle
Linguists are in the business of knowing (and finding out) how language works

• As an abstract system
• In the process of language learning
• In face to face communication
• In interactions with societal structures
• ...
SYNTACTIC VARIATION AND LINGUISTIC COMPETENCE: 
THE CASE OF AAVE COPULA ABSENCE

A DISSERTATION 
SUBMITTED TO THE DEPARTMENT OF LINGUISTICS 
AND THE COMMITTEE ON GRADUATE STUDIES 
OF STANFORD UNIVERSITY 
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS 
FOR THE DEGREE OF 
DOCTOR OF PHILOSOPHY
RA work on the English Resource Grammar
(Flickinger 2000, 2011)
RA work on the English Resource Grammar (Flickinger 2000, 2011)
The Grammar Matrix: 
Computational methods for combining typological breadth & syntactic depth 
(Bender et al 2002, 2010; Zamaraeva et al 2022)

Table 1: The Grammar Matrix libraries with selected typological sources

<table>
<thead>
<tr>
<th>Library</th>
<th>Citation(s)</th>
<th>Selected typological sources</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coordination</td>
<td>Drellishak and Bender 2005</td>
<td>Payne 1985; Stassen 2000; Drellishak 2004</td>
</tr>
<tr>
<td>Polar Questions</td>
<td>Bender and Flickinger 2005</td>
<td>–</td>
</tr>
<tr>
<td>Person, Number, Gender</td>
<td>Drellishak 2009</td>
<td>Cysouw 2003; Siewierska 2004; Corbett 2000</td>
</tr>
<tr>
<td>Agreement</td>
<td>Drellishak 2009</td>
<td>Corbett 2006</td>
</tr>
<tr>
<td>Case; Direct-Inverse</td>
<td>Drellishak 2008, 2009</td>
<td>Givón 1994</td>
</tr>
<tr>
<td>Argument Optionality</td>
<td>Saleem and Bender 2010; Saleem 2010</td>
<td>Ackema et al. 2006; Dryer 2013a</td>
</tr>
<tr>
<td>Tense</td>
<td>Poulson 2011</td>
<td>Comrie 1985; Dahl 1985</td>
</tr>
<tr>
<td>Aspect</td>
<td>Poulson 2011</td>
<td>Comrie 1976; Bybee et al. 1994</td>
</tr>
<tr>
<td>Lexicon</td>
<td>Drellishak and Bender 2005; Trimble 2014</td>
<td>Dixon 2004</td>
</tr>
<tr>
<td>Morphotactics</td>
<td>O’Hara 2008; Goodman 2013</td>
<td>–</td>
</tr>
<tr>
<td>Sentential Negation</td>
<td>Crowgey 2012, 2013</td>
<td>Dahl 1979; Dryer 2013b</td>
</tr>
<tr>
<td>Information Structure</td>
<td>Song 2014</td>
<td>Féry and Krifka 2008; Büring 2009</td>
</tr>
<tr>
<td>Adjectives; Copulas</td>
<td>Trimble 2014</td>
<td>Dixon 2004; Stassen 1997, 2013</td>
</tr>
<tr>
<td>Evidentials</td>
<td>Haeger 2017</td>
<td>Aikhenvald 2004; Murray 2017</td>
</tr>
<tr>
<td>Nominalized Clauses</td>
<td>Howell et al. 2018</td>
<td>Noonan 2007</td>
</tr>
<tr>
<td>Clausal Modifiers</td>
<td>Howell and Zamaraeva 2018</td>
<td>Thompson et al. 1985</td>
</tr>
<tr>
<td>Valence Change</td>
<td>Curtis 2018b,a</td>
<td>Haspelmath and Müller-Bardew 2001</td>
</tr>
<tr>
<td>Adnominal Possession</td>
<td>Nielsen and Bender 2018; Nielsen 2018</td>
<td>Payne and Barshi 1999; Heine 1997</td>
</tr>
<tr>
<td>Clausal Complements</td>
<td>Zamaraeva et al. 2019b</td>
<td>Noonan 2007</td>
</tr>
</tbody>
</table>
The Grammar Matrix: Computational methods for combining typological breadth & syntactic depth (Bender et al. 2002, 2010; Zamaraeva et al. 2022)

http://www.delph-in.net/matrix/customize/matrix.cgi
Computational linguistics

- Faculty director, UW CLMS (since program inception in 2005)
  - https://www.compling.uw.edu/
- NAACL Board (2012-2019; Chair 2016-2017)
- ACL Executive Board (2016-2017 & 2022-2025; President 2024)
POV: A linguist living in machine-learning land

- Emphasis is on getting the computer to ‘learn’ directly from data

- ‘Better’ if the researcher never looks at the data at all

- Devaluing of domain expertise (the ‘expensive’ people the machine is supposed to replace)

- => Hierarchies of knowledge (Raji et al 2021, Gebru & Torres 2023)

- Frequent thought while reviewing papers: These people don’t know the first thing about how language works!
But compiling is an interdisciplinary field

T1: 100 Things You Always Wanted to Know about Linguistics But Were Afraid to Ask*
*... for fear of being told 1000 more

Emily M. Bender
Morning session, 9am-12:30pm
But compiling is an interdisciplinary field

Tutorials

Tutorials will be held on July 15th, 2018. All tutorials will run for a half-day at the times noted below

**T1: 100 Things You Always Wanted to Know about Semantics & Pragmatics But Were Afraid to Ask**

*Emily M. Bender*

09:00 – 12:30

Location: 216, MCEC
But comping is an interdisciplinary field
Dear Computer Scientists,

"Natural Language" is *not* a synonym for "English".

That is all.
-Emily

9:32 AM · Nov 26, 2018
4 Prescriptions for Typologically-Informed NLP

This section provides a brief list, in etiquette-book format, of suggestions for producing more typologically-informed (and thus hopefully more language-independent) NLP research. These suggestions are aimed at both authors and reviewers, or rather, at members of the NLP community in both their author and reviewer guises.

**Do state the name of the language that is being studied, even if it’s English.** Acknowledging that we are working on a particular language foregrounds the possibility that the techniques may in fact be language-specific. Conversely, neglecting to state that the particular data used were in, say, English, gives false veneer of language-independence to the work.

(Bender 2011: On Achieving and Evaluating Language-Independence in NLP)
The #BenderRule origins, part 2

At NAACL 2019 (and elsewhere) made a pest of myself: “What language was this on?”

Alex O’Connor (gone to mastodon) · Jun 3, 2019
@uberalex · Follow
Replying to @emilymbender and @seb_ruder

Is there a formal statement of the Bender rule? Asking for future use.

@emilymbender@dair-community.social on Mastodon
@emilymbender · Follow

"Always name the language(s) you're working on."

That's really the bare minimum. I'd really like to encourage people to go much further and do data statements: aclweb.org/anthology/pape...

4:57 PM · Jun 3, 2019

(Bender 2019)
Similar precepts hold in other areas of linguistics

• Case in point, the Holliday Rule:

Nicole Holliday (mixedlinguist@bsky.social)
@mixedlinguist

Introducing the Holliday rule for ling papers, summarized as "It's alright to say they're white" 😅. If you got participants, you should ask their race(s), tell us what they said, even if they're white! h/t @kirbyconrod for the idea, & @emilymbender for pioneering this kinda rule!

11:04 AM · Dec 2, 2020

22 Reposts 7 Quotes 153 Likes 5 Bookmarks
POV: A linguist living in machine-learning land

• With the release of BERT (Devlin et al 2019) by Google in 2018, started to see repeated claims of BERT and similar language models “understanding”

• But a language model (better: corpus model, Veres 2022) models the distribution of word forms in text

  • Distributional similarity gives a clue to semantic similarity (Harris 1954, Firth 1957)


  • … and nothing like understanding

• => Interminable Twitter arguments and then finally the octopus paper
Bender & Koller 2020: Working definitions

- **Form**: marks on a page, pixels or bytes, movements of the articulators

- **Meaning**: relationship between linguistic form and something external to language

  - $M \subseteq E \times I$ : pairs of expressions and communicative intents

  - $C \subseteq E \times S$ : pairs of expressions and their standing meanings

- **Understanding**: given an expression $e$, in a context, recover the communicative intent $i$
Thought experiment: Meaning from form alone

What a pretty sunset

Reminds me of lava lamps

A

B
Thought experiment: Meaning from form alone

I made a coconut catapult! Let me tell you how…

Cool idea! Great job!
Thought experiment: Meaning from form alone

Help! I’m being chased by a bear!
Thought experiment: Meaning from form alone

All I have is a stick! What do I do?

The bear is chasing me!*

*Reply generated by GPT2 demo
Thought experiment: Meaning from form alone

All I have is a stick! What do I do?

You’re not going to get away with this!*

*Reply generated by GPT2 demo
So how do babies learn language?

• Interaction is key: Exposure to a language via TV or radio alone is not sufficient (Snow et al 1976, Kuhl 2007)

• Interaction allows for joint attention: where child and caregiver are attending to the same thing and mutually aware of this fact (Baldwin 1995)

• Experimental evidence shows that more successful joint attention leads to faster vocabulary acquisition (Tomasello & Farrar 1986, Baldwin 1995, Brooks & Meltzoff 2005)

• Meaning isn’t in form; rather, languages are rich, dense ways of providing cues to communicative intent (Reddy 1979). Once we learn the systems, we can use them in the absence of co-situatedness.
Meanwhile...

• In 2016, CLMS Advisory Board member Lesley Carmichael suggested that the curriculum should include something on ethics.

• First taught “Ethics and NLP” (now “Societal Impacts of Language Technology”) seminar in Winter 2017, asking:

  • What ethical considerations arise in the design and deployment of NLP technologies?

  • Which of these are specific to NLP (as opposed to AI or technology more generally?)

  • What best practices can/should NLP developers deploy in light of the ethical concerns identified?
What can go wrong? Just about everything…

• Language models pick up bias from training data & amplify it

  • Biases come out in search engine results (Noble 2018), machine translation output (Caliskan & Lewis 2021), & more

• Systems trained for prestige varieties don’t work as well for non-standard/stigmatized varieties (e.g. Wassink et al 2022, Sap et al 2019)

• … but better coverage of language varieties enables more surveillance (Bender & Grissom II, 2024)

• Data annotation processes often involve exploitative labor practices (Fort et al 2011; see also news coverage in 2023 by Billy Perrigo and Karen Hao)
What can we do?

• Value sensitive design (Friedman & Hendry 2019): collection of methodologies for
  • Identifying stakeholders (direct and indirect)
  • Eliciting stakeholder values
  • Designing to support stakeholder values

• Values of transparency and recourse + knowledge of sociolinguistic variation
  => Data Statements for Natural Language Processing

• See also: Gebru et al 2021, Mitchell et al 2019, Arnold et al 2019, Stoyanovich & Howe 2019
Data Statements for Natural Language Processing

- Bender & Friedman 2018, Bender et al 2021, McMillan-Major et al 2023, McMillan-Major & Bender forthcoming

https://techpolicylab.uw.edu/data-statements/
2020: Origin story of the Stochastic Parrots paper

- Started off as a Twitter DM conversation, with Dr. Timnit Gebru:

Timnit Gebru
@timnitGebru

Hi Emily, I'm wondering if you've written something regarding ethical considerations of large language models or something you could recommend from others? I'm only getting to learn about this via the GPT-2 conversations that you, Anima etc were having and the resources you've been sharing

and if you haven't written something yet I would by customer #1 of anything you write on this end

This is something I'm trying to advocate for at Google

sent them some of your tweets as well haha

so many conversations about how we're not leading in large language models and should

GPT-3 is so impressive etc

and each time I'm like ANNDD see what Emily has to say

I can think of three other angles around ethical implications of GPT-3 and the like:

1) Carbon cost of creating the damn things (see Strubell et al at ACL 2019)
2) AI hype/people claiming it's understanding when it isn't (Bender & Koller at ACL 2020)
3) Deepfakes/random generated text that no one is accountable for but which is interpreted as meaningful.

Sep 8, 2020, 4:54 PM
Bender, Gebru et al 2021
On the Dangers of Stochastic Parrots:
Can Language Models be too big?


- **Hutchinson**: Hutchinson et al 2019, 2020, 2021

- **Díaz**: Lazar et al 2017, Díaz et al 2018
History & context of this paper

- October 2020: Paper goes through Google-internal “pup-approve” before submission to FAccT

- Late November: Google asks Dr. Gebru to either retract the paper or remove the Google co-authors’ names from it

- Dr. Gebru pushes back, asking for information on what exactly was being objected to and objecting to how she & her team were being treated

- December 2, 2020: Google fires Dr. Gebru

- Dr. Margaret Mitchell starts documenting what happened to Dr. Gebru and calling on people within Google to apologize & fix systems

- February 19, 2021: Google fires Dr. Mitchell
History & context of this paper

- Google’s actions led to intense media interest, both about their treatment of Dr. Gebru (and eventually Dr. Mitchell) and about our research

  - Selected media coverage

- Someone leaked the “pub-approve” version of the paper to Reddit

- FAccT 2021 primary reviewers were complete before the media story broke (preserving anonymous review)

- FAccT 2021 acceptances announced on December 22, 2020

- Camera ready due January 22, 2021
Risks of large language models (LLMs)

• Environmental impact (including environmental racism angle)

• Financial inaccessibility (impact on participation in research)

• Unmanageably large training data (no, the Web isn’t representative)

• Reproduction and amplification of bias

• Misdirected research effort

• Harms of synthetic media
Stochastic Parrots

An LM is a system for haphazardly stitching together linguistic forms from its vast training data, without any reference to meaning: a stochastic parrot.
Stochastic Parrots coda (2024 update)

• "How do you feel now that your predictions have come true?"

• Those weren't predictions, they were warnings!

• What we didn’t predict/notice at the time:
  • Exploitative labor practices
  • Just how enthusiastic people would be about synthetic text
  • Pollution of the information ecosystem (Shah & Bender 2023)
  • The transition to treating LLMs as “everything machines”, i.e. an “unscoped technology” (Gebru & Torres 2023)
Stochastic Parrots

An LM is a system for haphazardly stitching together linguistic forms from its vast training data, without any reference to meaning: a stochastic parrot.
Stochastic parrots and what we know about how language works

• Language as a system of signs (pairings of form and meaning; de Saussure 1959)

• Language use as a joint activity, involving intersubjectivity (Clark 1996, Baldwin 1995)

• => Non-understanding of LLMs is obvious

• But there’s a huge need to share this understanding

• … and increasingly so as venture capital is pushing the use of LLMs in every aspect of modern life
Public scholarship: AI hype takedowns

• Started with Twitter threads, then blog posts

• Mostly focused on media coverage of “AI” claims


=> Watching for and calling out implausible claims

=> Close reading of how people are talking about “AI” and where the discourse goes astray
Public scholarship: AI hype takedowns

On NYT Magazine on AI: Resist the Urge to be Impressed

22 min read · Apr 17, 2022

2.2K

The Real Problem with A.I. with Emily Bender

Factually! with Adam Conover | EP162

Privacy Share Subscribe
An accidental podcast (with Alex Hanna)

https://www.dair-institute.org/maiht3k/
November 2022: (Galactica then) ChatGPT
Now LLM talk wasn’t just in tech circles

• Intense media attention, now from the non-tech press

• Approached by Liz Weil for her *New York Magazine* piece

• Policymakers (city, federal, UN, IMF)
Can feel like saying the same thing over and over again, but:

• Each new media contact is a chance to reach a new audience

• Continually working on how to make the message resonate

• Continually learning from other scholars and activists in the “AI” ethics space

• Continually asking: What specifically am I bringing to this conversation
  • based on what I know about how language works
Linguistics is good for more than combatting AI hype

• AI hype is all over the news right now, but it’s far from the only arena where the world can benefit from what we know about how language works

• Public scholarship can take many forms, and doesn’t have to involve engaging on social media or talking with reporters

• Sometimes the relevant public is a small, tight-knit community

• Sometimes it’s teachers and/or parents

• There’s a long tradition of public scholarship in linguistics
Some of my linguistics heroes

• Dr. Anne Charity Hudley, redefining spaces from the K-12 classroom, to college campuses, to the LSA to make them truly inclusive

• Dr. Wesley Leonard, developing approaches to language reclamation and language science (teaching) that center an Indigenous perspectives

• Dr. John Rickford & Dr. Sharese King, finding interventions for when the legal system can’t “hear” vernacular speakers

• Dr. Kirby Conrod, providing practical pronoun guidance, showing folks they can in fact get this right

• Dr. Kelly Elizabeth Wright, using sociolinguistics to dismantle institutionalized racism
An explicitly political activity

• My date of entry into the ethics and NLP space is not coincidental (2016)
  
  • I was asking myself: How can I use my expertise to do something meaningful?
  
• Explicitly political moves can be scary

• But scholarship is inherently political—to say otherwise is also a (covert) political move
An explicitly political activity

• The more privilege one has, the more easily one can lay claim to the apocryphal “view from nowhere”

• … and the more tempting it is to stay there.

• But it’s worth it to take the leap

What are you doing or will you do with what you know about how language works?
References


