



From Linguist in NLP to Humanist in AI: How a Linguist's Perspective on Data Has Informed My Work on Ethics in NLP

Emily M. Bender
University of Washington

Widening NLP @ ACL 2019
28 July 2019

@emilymbender

My journey into computational linguistics

- Discovered linguistics freshman year of university; AB (UC Berkeley), MA, PhD (Stanford) all in Linguistics
- First programming language: Logo (4th grade)
- First programming class: CS 60A @ Cal, in Scheme
 - Concurrently: Morphology with Sharon Hargus & TA David A. Peterson
 - First compiling project: Bantu morphological analyzer in Scheme



My journey into computational linguistics

- Grad school: Introduction to computational linguistics (Martin Kay), phenomenology (Terry Winograd)
 - RAship in grammar engineering, with Ivan Sag and Dan Flickinger
 - Dissertation (2001): *Syntactic Variation and Linguistic Competence: The Case of AAVE Copula Absence*
- No luck on the job market as syntactician or sociolinguist
- Short stint in industry (YY Technologies) as a grammar engineer for Japanese
- Laid off in first dot-com bust @7 months pregnant

My journey into computational linguistics

- While at YY, started the Grammar Matrix, in connection with Project Deep Thought
- After a couple more years of temporary positions, hired by UW Linguistics to start the CLMS program
- At the time: strong language group in EE working on MT & ASR (Mari Ostendorf, Jeff Bilmes, Katrin Kirchhoff)
- CSE had AI/IE folks, who worked with language data

Language *per se* vs.
Information encoded in language

Outline

- The field as I found it in 2003 and where it is now
- Current issues
- What a linguist can bring
- There's more to NLP than SOTA
- Towards more interdisciplinary, multilingual and ethical NLP
- How we can do better

A linguist's eye view of the recent history of NLP

- 2000-2015: Machine learning “versus” rule-based systems (aka linguistics)
 - Machine learning in the service of building better NLP applications
 - But also and increasingly: NLP as a proving ground for ML
 - Role for linguistics in feature engineering
- 2015-now: Deep learning
 - NLP as proving ground for DL
 - No need for feature engineering: off-load understanding how to represent data to the machines
 - End-to-end everything
 - But also: work asking what is it that the big models are learning?

What's the problem?

- End-to-end, task-focused research entails always looking through the window
- Miss that language itself has structure
- A language is a general purpose communication tool; a whole pile of systems trained on end-to-end tasks won't be



What's the problem?

- Languages have structure
- ... which varies (within bounds) across languages
- Linguistically naïve \neq language independent (Bender 2009)
 - See also Typ-NLP Workshop on Thursday



Aside: the #BenderRule

- “Always state the name of the language you are working on, even if it is English”
- Coined by (at least) Nathan Schneider, Yuval Pinter, Rob Munro & Andrew Caines



Emily M. Bender

@emilymbender



Dear Computer Scientists,

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

That is all.

-Emily

9:32 AM - 26 Nov 2018

255 Retweets 1,132 Likes



Aside: the #BenderRule

- I invite you to join me in asking authors, if they don't specify, which language(s) they tested their systems on
- Why does this matter, if we always know it's English unless otherwise specified?
- Status quo: Work on non-English is “language specific”, work on English is “NLP”
- But English is just one language, like any other and not representative of all!
 - A window with its own specific pattern of raindrops



How is English non-representative?

- It's a spoken language, not a signed language
- It has a well-established, long-used, roughly phone-based orthographic system
- ... with white space between words
- ... using (mostly) only lower-ascii characters
- It has relatively little morphology and thus fewer forms of each word
- It has relatively fixed word order
- English forms might 'accidentally' match database field names, ontology entries, etc.
- It has massive amounts of training data available (like the 3.3B tokens used to train BERT (Devlin et al 2019))

What's the problem?

- If we're always looking through the window, we miss the variation within languages
- Sociolinguists have found that variation correlates with speaker demographic characteristics, speech situation & more (e.g. Labov 1966)
- Speakers attach social meaning to linguistic variation and use it to construct & project identities (e.g. Eckert and Rickford 2001)
- Sociolinguistically naïve NLP will miss these realms of meaning
- Sociolinguistically naïve NLP won't work equally well for all users (even in high resource languages)



What's the problem?

- If we're always looking through the window, we risk mistaking the scene on the other side for "ground truth"
- Work on learning world knowledge or "common sense" from corpora conflates what people say about the world with ground truth
 - "Black sheep" problem (Meg Mitchell, pc)
 - Poor performance of sentiment analyzers because of toxic discourse about immigration in the US (Speer, 2017)



How can linguists/linguistics help?

Understanding the structure of language

- Not just for rule-based systems!
- Feature engineering (where applicable)
- Design of ancillary tasks (see Smith 2017)
- Error analysis
- Design of annotation schemes + expert annotation:
 - Without it, we can't know if we've solved the problem
 - The field should value this work (see Heinzerling 2019)

How can linguists/linguistics help?

Understanding variation in language

- Where might our assumptions fail for a different language?
- How do we ensure that deployed models work equitably
 - For all users
 - For all indirect stakeholders (see Friedman & Hendry 2019)

How can linguists/linguistics help?

Understanding relationship between form & meaning

- Form: text, speech, sign (+ paralinguistic information like gesture or tone)
- Conventional/standing meaning: logical form (or equivalent) that the linguistic system pairs with that form
- Communicative intent of the speaker: what they are publicly committed to by uttering that form (+ additional plausibly deniable inferences)
- Relationship between communicative intent & the world, e.g.:
 - True assertion, mistaken assertion, lie, accidentally true assertion, social act related to construction of social world, question about the interlocutor's beliefs, ...

How can linguists/linguistics help?

Stepping off the SOTA treadmill

- Linguistics encourages us to:
 - understand our data
 - be interested in the linguistic form itself — and see the raindrops as distinct from the view on the other side
- Language is always changing, but on a very different time scale than current NLP



When the SOTA is all that counts

- SOTA chasing encourages a frenetic pace, especially in combination with arXiv
- We lose researchers who can't just drop everything to stay up working all night
- We don't have time for “research slow” (see Kan 2018), or to understand **how** and **why** systems work as they do (Niven & Kao 2019)
- Which SOTA? Just for English? Multilingual? Reproducible? (see Fokkens 2017)

Interdisciplinarity

- NLP/CL is at the intersection of: linguistics, CS, statistics, EE, ...
- NLP/CL connects with: biomedical informatics, computational social science, data science, ...
- Being interdisciplinary is about ***cooperation***, not ***competition***
- We are working on problems that require multiple kinds of expertise to solve, and we'll get there by learning from each other

Towards promoting interdisciplinarity in NLP

- Tutorials at NAACL 2012 and ACL 2018: “100 things you always wanted to know about linguistics, but were afraid to ask ... for fear of being told 1000 more”
- *Linguistic Fundamentals for Natural Language Processing: 100 Essentials from Morphology and Syntax* (Morgan & Claypool, 2013)
- *Linguistic Fundamentals for Natural Language Processing II: 100 Essentials from Semantics and Pragmatics* (Morgan & Claypool, forthcoming 2019)

Towards promoting interdisciplinarity in NLP

- COLING 2018 PC activities (with Leon Derczynski)
 - Paper types, including “computer assisted linguistic analysis”
 - Review forms emphasizing error analysis and hypothesis testing
 - 9 Best Paper awards, across different categories
- For details, see the COLING 2018 PC blog: <http://coling2018.org/category/pc-blog/>

Towards promoting interdisciplinarity in NLP

- UW's Computational Linguistics Master of Science (CLMS) curriculum design:
 - 3 of 9 courses are in linguistics (exceptions for those who already have ling degrees)
 - cross-cutting themes emphasize multilinguality, ambiguity resolution and ethical considerations
 - recruit cohorts with diverse training and promote collaborative learning
 - prerequisite: introduction to linguistics

Towards more multilingual NLP

- Bender 2009 “Linguistically naïve != language independent”
- Bender 2011 Dos & don’ts for language independent NLP, including:

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.



Towards more ethical NLP: Data Statements (Bender & Friedman 2018)

- CLMS advisory board member Lesley Carmichael suggested we should include ethics in the curriculum (late 2015 or early 2016)
- After trying & failing to find someone to teach it, decided to try myself:
 - Ling 575: Ethics and NLP, WI 2017
http://faculty.washington.edu/ebender/2017_575/
- While preparing that course, fortuitously met Batya Friedman (UW iSchool)
- Guest lecture by Friedman on value sensitive design (<https://vsdesign.org/>)

Data Statements for NLP

Proposed Schema: Long Form

- A. Curation Rationale
- B. Language Variety
- C. Speaker Demographic
- D. Annotator Demographic
- E. Speech Situation
- F. Text Characteristics
- G. Recording Quality
- H. Other
- I. Provenance Appendix

Why NLP Needs Data Statements

- Systems trained on naturally occurring text learn the biases held by the authors of the text (*pre-existing bias*)
 - Word embeddings pick up gender (e.g. Bolukbasi et al 2016) and race/ethnicity bias (e.g. Speer 2017)
 - Machine learning systems can amplify the biases they learn (e.g. Zhao et al 2017)
- Systems trained on one subpopulation don't work as well for others (*emergent bias*)
 - POS tagging (Hovy and Søgaard, 2015; Jørgensen et al., 2015); ASR engines (Tatman, 2017)

How do data statements help?

- Emergent bias: Procurers, consumers and advocates can check whether a system is trained on appropriate data for its deployed use case
- Emergent bias: As a field, we can track what speaker populations are underserved
- Pre-existing bias: Knowing what kind of texts a system is trained on can be key to working out the source of bias, as in Speer's (2017) study of word embeddings and sentiment analysis

Data statements alone won't 'solve' bias, but if we do not make a commitment to data statements or a similar practice for making explicit the characteristics of datasets, then we will single-handedly undermine the field's ability to address bias.

Suggested actions

- Write (and look for) data statements :)
- As a reviewer, value work that
 - Explores NLP for lower resource languages
 - Provides careful error analysis
 - Provides careful success analysis
- Value the interdisciplinary nature of our field
 - Learn enough of the other pillars to engage in meaningful collaboration



Suggested actions

- Step off the SOTA treadmill
 - If you're worried about being scooped, there's probably a more interesting question you could be pursuing
- But how do we change we the field, so that we can succeed as individuals with fewer, more thoughtful publications?



Suggested actions

- Where you get the opportunity, value analytical work in addition to (or even above) ‘SOTA’
 - Avoid using ‘technical’ to mean ‘involves math/programming’
- Advocate for reviewing structures that value crosslinguistic and/or analytical work (see COLING 2018)
- When people don’t state the language they’re working on, ask :)
 - Feel free to blame this awkward asking-the-obvious question on me
- Engage broadly with emerging conversation about ethics and NLP and ethics and AI

Thank you!

- Slides available online: <http://faculty.washington.edu/ebender/slides.html>
- Twitter: @emilymbender



References

- Bender, E. M. (2009). Validating analyses against data: Why syntax needs large scale validation. Paper presented at LSA Symposium on Computational Linguistics in the Service of Linguistic Analysis. San Francisco, CA, January 2009.
- Bender, E. M. (2011). On achieving and evaluating language independence in NLP. *Linguistic Issues in Language Technology*, **6**, 1–26.
- Bender, E. M. (2013). *Linguistic Fundamentals for Natural Language Processing: 100 Essentials from Morphology and Syntax*. Morgan & Claypool.
- Bender, E. M. and Friedman, B. (2018). Data statements for natural language processing: Toward mitigating system bias and enabling better science. *Transactions of the Association for Computational Linguistics*, **6**, 587–604.
- Bender, E. M. and Lascarides, A. (forthcoming 2019). *Linguistic Fundamentals for Natural Language Processing II: 100 Essentials from Semantics and Pragmatics*. Morgan & Claypool.
- Bolukbasi, T., Chang, K.-W., Zou, J. Y., Saligrama, V., and Kalai, A. T. (2016). Man is to computer programmer as woman is to homemaker? Debiasing word embeddings. In D. D. Lee, M. Sugiyama, U. V. Luxburg, I. Guyon, and R. Garnett, editors, *Advances in Neural Information Processing Systems 29*, pages 4349–4357. Curran Associates, Inc.
- Eckert, P. and Rickford, J. R., editors (2001). *Style and Sociolinguistic Variation*. Cambridge University Press, Cambridge.
- Friedman, B. and Hendry, D. G. (2019). *Value Sensitive Design: Shaping Technology with Moral Imagination*. MIT Press.

- Heinzerling, B. (2019). NLP's Clever Hans moment has arrived. Blog post, available at <https://bheinzerling.github.io/post/clever-hans/>, accessed July 25, 2019.
- Hovy, D. and Søgaard, A. (2015). Tagging performance correlates with author age. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, pages 483–488. Association for Computational Linguistics.
- Jørgensen, A., Hovy, D., and Søgaard, A. (2015). Challenges of studying and processing dialects in social media. In *Proceedings of the Workshop on Noisy User-generated Text*, pages 9–18. Association for Computational Linguistics.
- Kan, M.-Y. (2018). Research fast and slow. Keynote talk, COLING 2018; video available at <https://vimeo.com/287505750>.
- Labov, W. (1966). *The Social Stratification of English in New York City*. Center for Applied Linguistics, Washington, DC.
- Niven, T. and Kao, H.-Y. (2019). Probing neural network comprehension of natural language arguments. *Proceedings of ACL 2019*.
- Smith, N. A. (2017). Squashing computational linguistics. Keynote talk, ACL 2017; video available at <https://vimeo.com/234958746>.
- Speer, R. (2017). Conceptnet numberbatch 17.04: better, less-stereotyped word vectors. Blog post, <https://blog.conceptnet.io/2017/04/24/conceptnet-numberbatch-17-04-better-less-stereotyped-word-vectors/>, accessed 6 July 2017.
- Tatman, R. (2017). Gender and dialect bias in YouTube's automatic captions. In *Proceedings of the First ACL Workshop on Ethics in Natural Language Processing*, pages 53–59. Association for Computational Linguistics.
- Zhao, J., Wang, T., Yatskar, M., Ordonez, V., and Chang, K.-W. (2017). Men also like shopping: Reducing gender bias amplification using corpus-level constraints. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 2941–2951. Association for Computational Linguistics.