Linguistics, Living Its Best Life

Emily M. Bender
University of Washington
@emilymbender

Max Planck Institute for Psycholinguistics
January 6, 2021

If linguistics is living its best life…

- The future of linguistics is:

1. Broadly inclusive

2. Integrative

3. Computationally aided

4. Impactful in the world

My journey into and through linguistics

- Undergraduate (BA Linguistics) at UC Berkeley
    - Construction grammar, Japanese linguistics circle, BeRP project RA (the BeRP Wizard)
    - One year at Tohoku University (Sendai, Japan), exposed to P&P
    - One computer science class, inspired first compling project
- Off to Stanford, ready to create “Generalized Bay Area Grammar”!

My journey into and through linguistics

- Graduate studies (MA, PhD Linguistics) at Stanford
    - HPSG, LFG, Minimalism: having multiple toolkits helps us ask more questions!
    - RA on the LinGO project: grammar engineering
    - Sociolinguistics, language acquisition, semantics
    - No “big questions” in syntax proper resonated with me
- Dissertation: *Syntactic Variation and Linguistic Competence: The Case of AAVE Copula Absence*

My journey into and through linguistics

- No luck finding syntax or sociolinguistics positions
- Grammar engineering experience led to industry position led to UW position in computational linguistics
- At UW:
  - Multilingual grammar engineering and applications to language documentation
  - Computational sociolinguistics
  - Computational semantics
  - Ethics and societal impact of natural language processing (NLP)

If linguistics is living its best life, the future of linguistics is **broadly inclusive**

A broadly inclusive linguistics values

- All languages
- All language modalities
- All language varieties
- All speakers
- ‘Applied’ as well as ‘theoretical’ topics

We’re not there yet

- In NLP, English is so unmarked, it rarely even gets named (cf. the #BenderRule; Bender 2019)

- In sociolinguistics in many English-speaking areas, white people are so unmarked, they don’t get named (Lanehart 2009, see also the #HollidayRule)

- Work on lesser-studied languages has access to less shared knowledge among scholars, influencing what topics are considered ‘publishable’ (DiCanio 2019)

- Across most subfields, spoken languages are so unmarked we rarely use the phrase spoken languages

- Much theoretical work relies on an idealized notion of ‘monolingual native speaker’ which maps poorly to most language experience in the world

Proposed value shifts

- Work that involves primary data collection should be seen as more prestigious

- Work that involves less idealized data collection circumstances should be understood as more difficult

- Work that focuses only on English or other well-studied languages should be understood as narrow in its scope

- Work that focuses only on speakers from dominant groups should be understood as narrow in its scope

- No language, modality, or speaker population should get to be unmarked

If linguistics is living its best life, the future of linguistics is **integrative**

Integrated language use

- Most linguists tend to look at individual, specific aspects of language knowledge or use

- But as language users, we handle phonology, morphology, syntax, semantics, pragmatics in every sentence

- As language perceivers, our structural language processing is carried out in tandem with and informed by our knowledge of which language variants carry which social signals

- “Denotational” meaning and “social” meaning aren’t cleanly separable, in use or change (e.g. McConnell-Ginet 1984, Beltrama 2020)

- And language use is also incremental language learning and incremental language change

Socio-psycholinguistics

- Listener perception of speaker gender informs listener perception of /s/ and /ʃ/ fricatives in American English (Strand & Johnson 1996, Bouavichith et al 2019)

- Visual exposure to stuffed toys (kangaroo or kiwi) informs listener perception of the vowel space among New Zealand English speakers (Hay & Drager 2010)

- Deliberate change to grammatical gender for French profession nouns succeeded only after ideological change (Burnett & Bonami 2019)

- Listener understanding of speaker political ideology influences listener perception of projection of politically-relevant presuppositions (American English & American political context; Mahler 2020)

Towards scaled-up and integrated models of language (Bender & Good 2010)

- Challenges:
  - Data acquisition and annotation: how to curate sufficiently large datasets?
  - Data mining: how to find relevant subsets of large datasets for a given problem?
  - Complexity: test and refine theories for scalability

Towards scaled-up and integrated models of language (Bender & Good 2010)

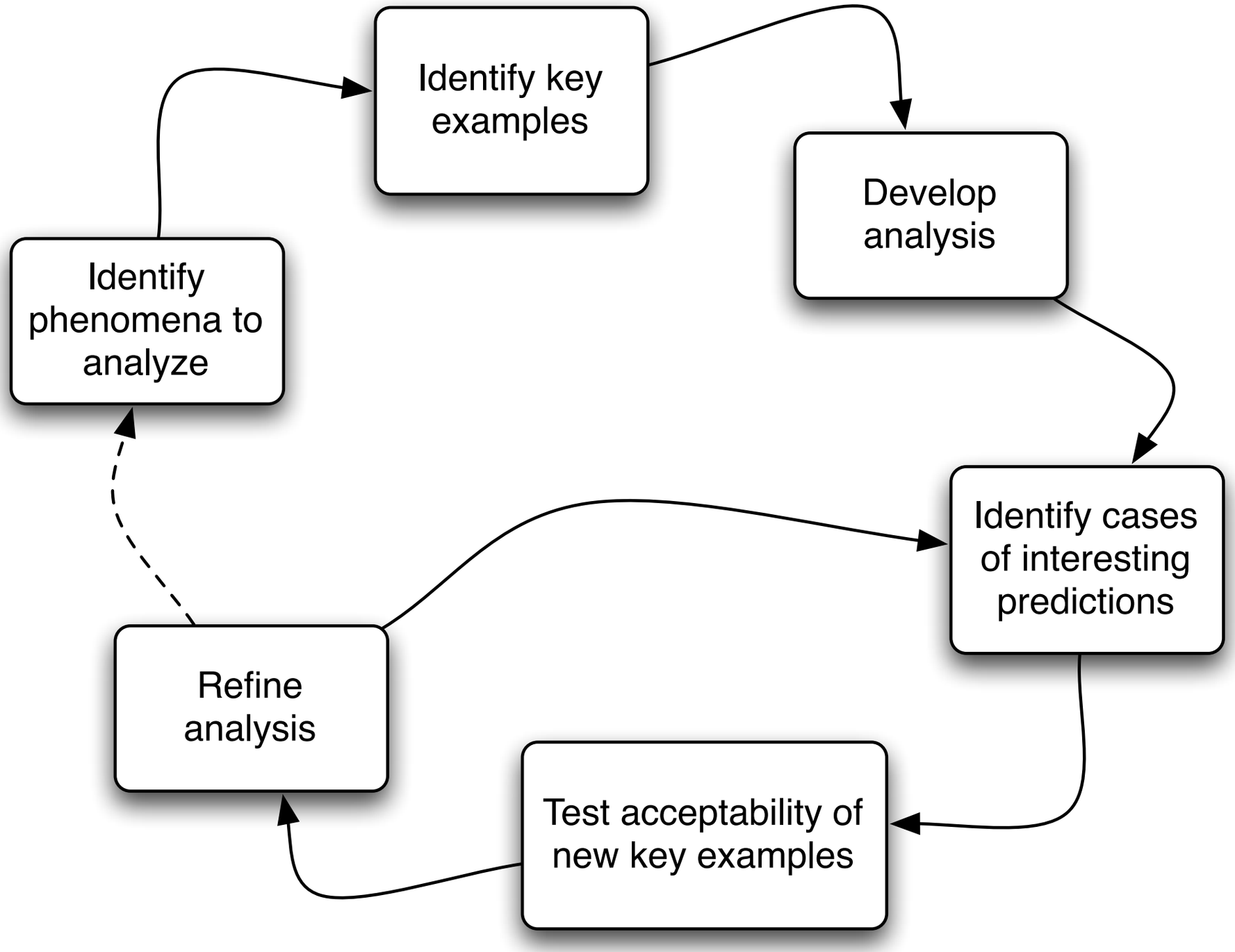
- Build a culture of data sharing and standards for doing so
    - E.g. Leipzig Glossing Rules (Bickel et al 2008), OLAC meta-data (Bird & Simons 2001), Universal Dependencies project (Nivre et al 2020), Xigt (Goodman et al 2015)
- Support standards-compliant tool development
    - Help ‘OWLs’ (ordinary working linguists) make standards-compliant, shareable datasets
- Create a culture of responsible data sharing
- Leverage computational methods

If linguistics is living its best life, the future of linguistics is **computationally-aided**

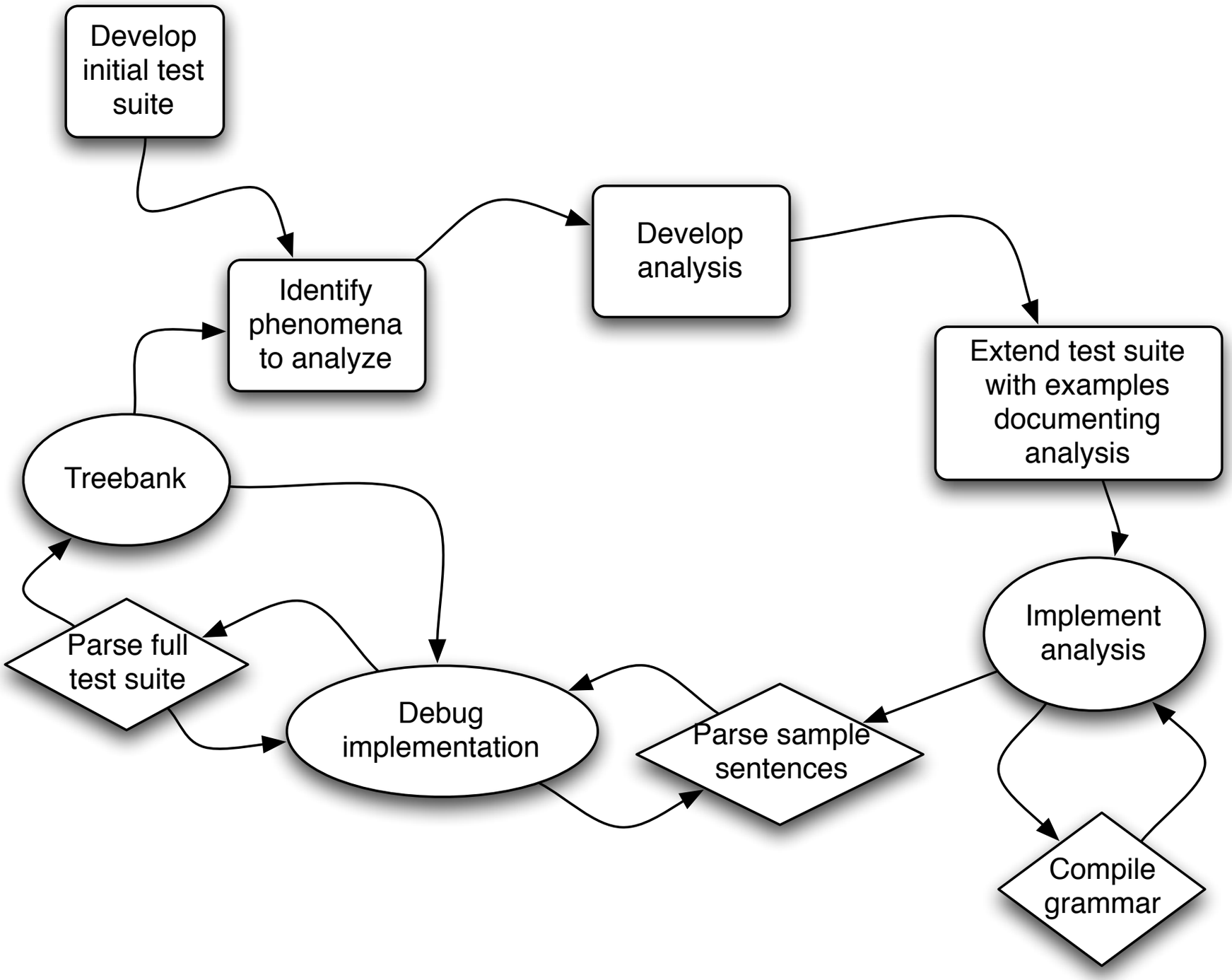
Computers are tools that allow us to scale up linguistic investigations

- Run various language production & processing experiments
- Search through larger datasets for examples, variants, counter-examples
- Test hypotheses rapidly and consistently against larger testsuites
- Verify interaction of components of models

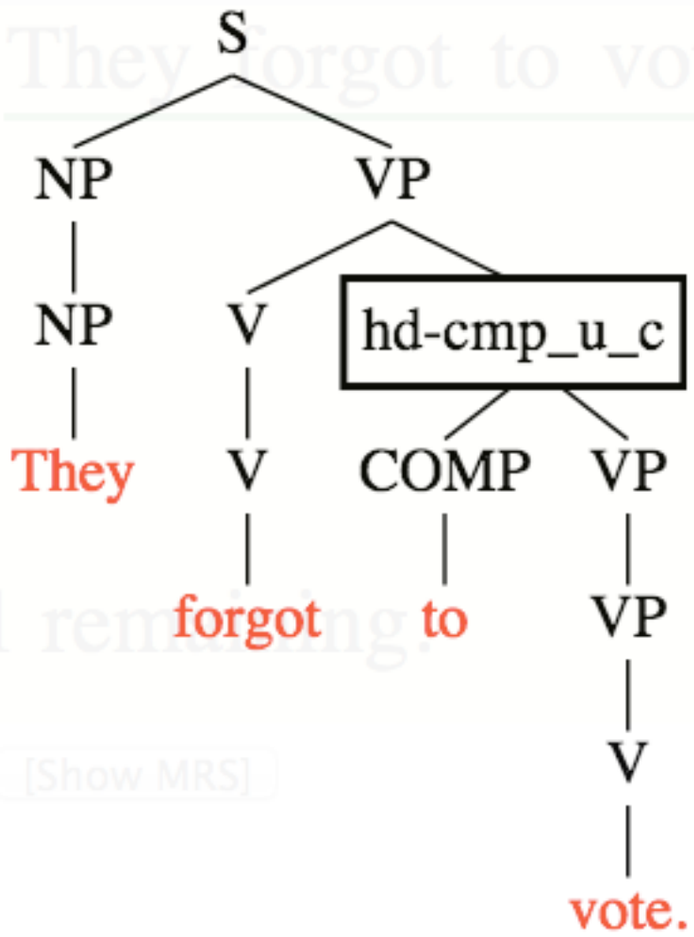
Example: Pen and paper syntax
(Bender et al 2011)

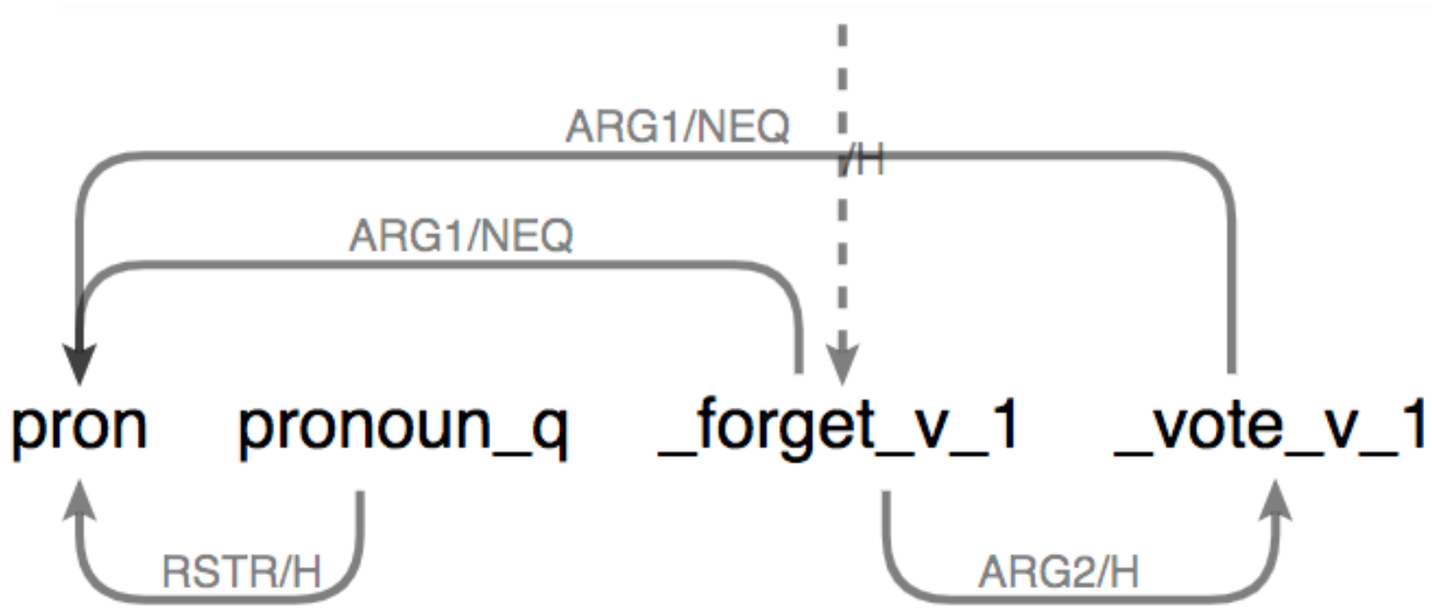


Example: Syntax with grammar engineering
(Bender et al 2011)

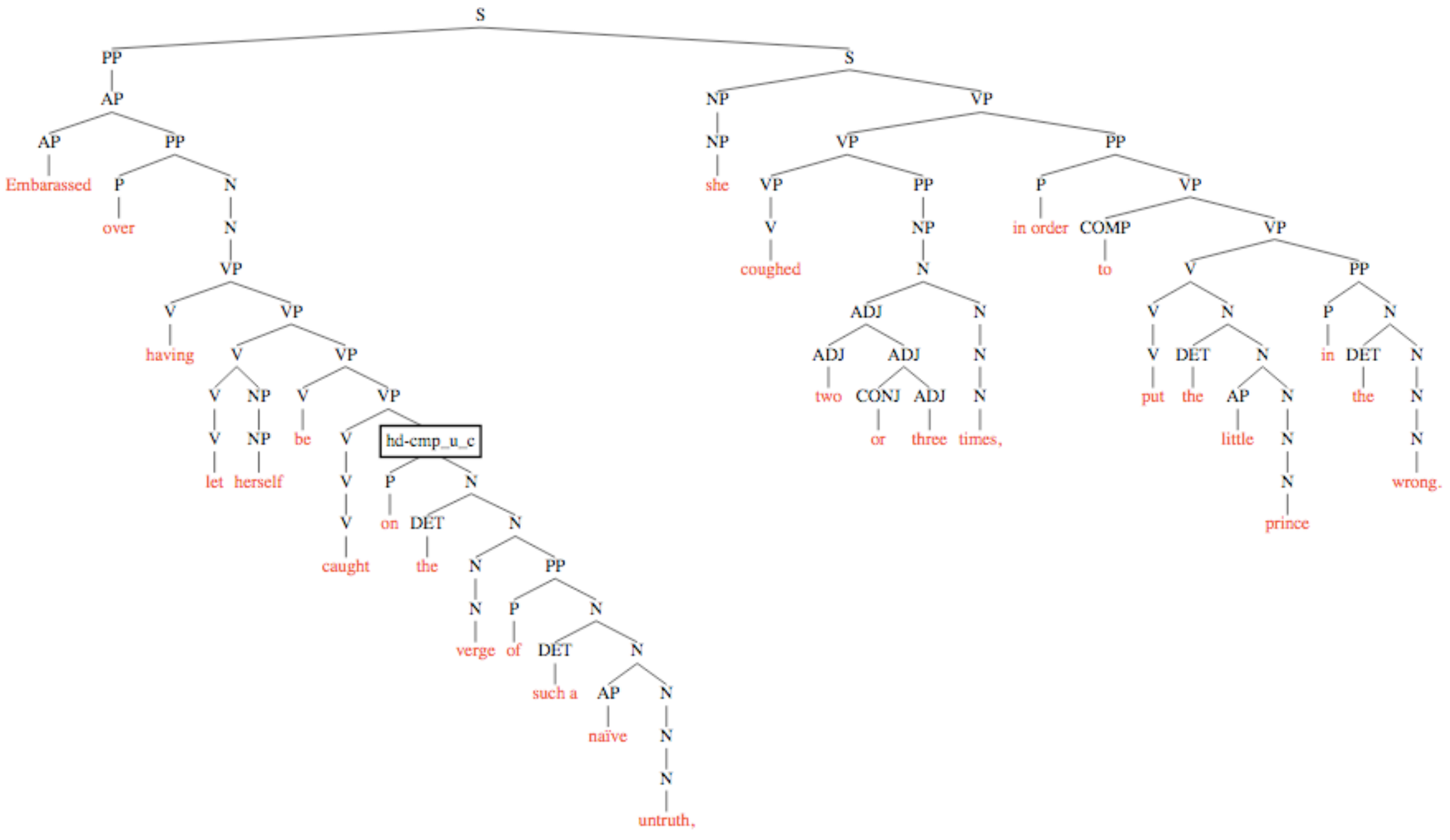


Examples from the English Resource Grammar
(Flickinger 2000, 2011)





ERG: Examples

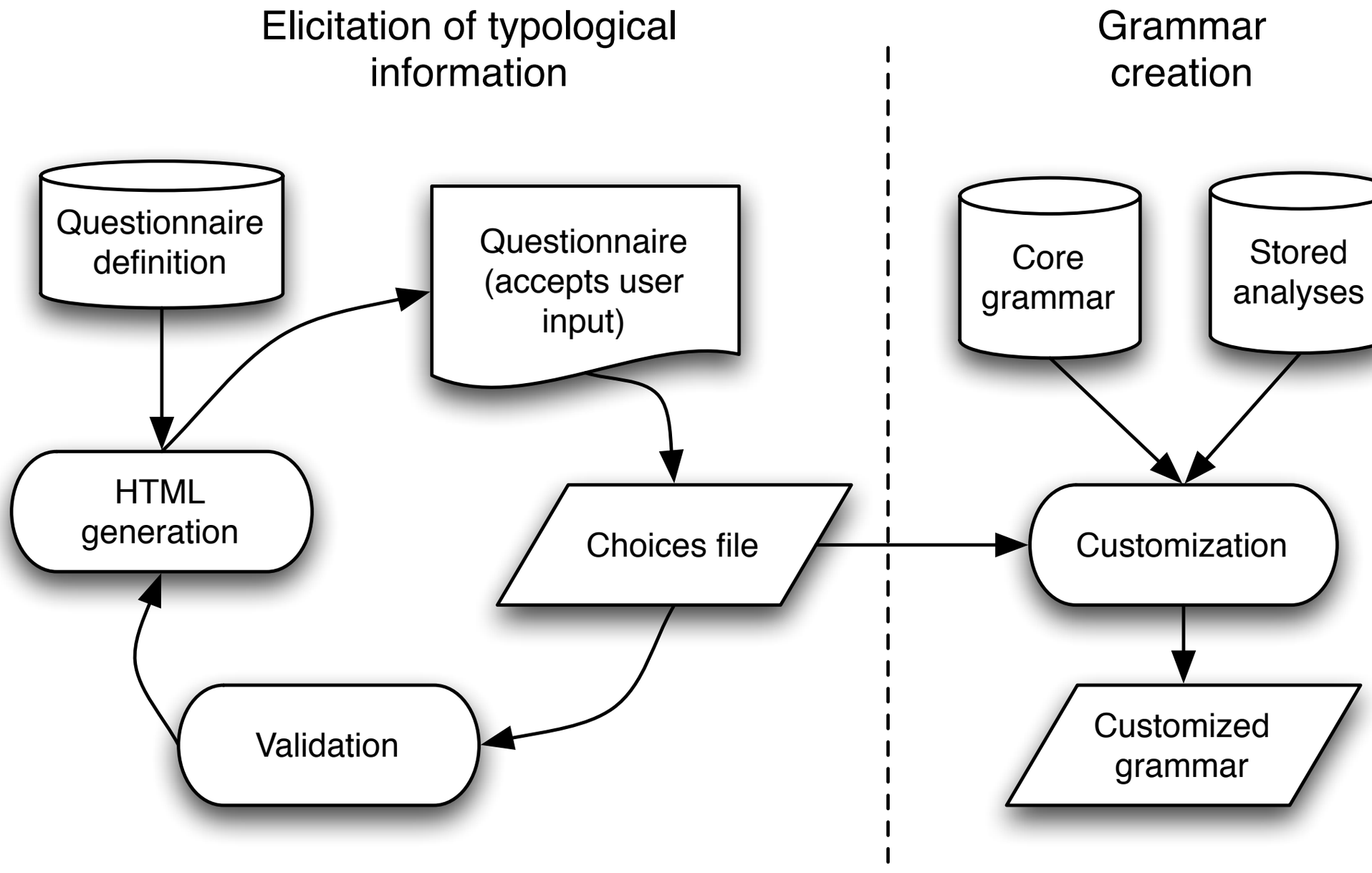


ERG: Examples

INDEX: **e2**
RELS:
h1:subord_rel(ARG0: **e4**,ARG1: **h5**,ARG2: **h6**)
h7:"_embarassed/JJ_u_unknown_rel"(ARG0: **e8**,ARG1: **i9**)
h7:_over_p_rel(ARG0: **e10**,ARG1: **e8**,ARG2: **x11**)
h12:udef_q_rel(ARG0: **x11**,RSTR: **h13**,BODY: **h14**)
h15:nominalization_rel(ARG0: **x11**,ARG1: **h16**)
h16:"_let_v_1_rel"(ARG0: **e17**,ARG1: **i18**,ARG2: **h19**)
h20:pron_rel(ARG0: **x21**)
h22:pronoun_q_rel(ARG0: **x21**,RSTR: **h23**,BODY: **h24**)
h25:"_catch_v_1_rel"(ARG0: **e26**,ARG1: **i27**,ARG2: **x21**,ARG3: **h28**)
h25:parg_d_rel(ARG0: **e29**,ARG1: **e26**,ARG2: **x21**)
h30:_on_p_rel(ARG0: **e31**,ARG1: **x21**,ARG2: **x32**)
h33:_the_q_rel(ARG0: **x32**,RSTR: **h34**,BODY: **h35**)
h36:"_verge_n_1_rel"(ARG0: **x32**)
h36:_of_p_rel(ARG0: **e37**,ARG1: **x32**,ARG2: **x38**)
h39:_such+a_q_rel(ARG0: **x38**,RSTR: **h40**,BODY: **h41**)
h42:"_naïve/JJ_u_unknown_rel"(ARG0: **e43**,ARG1: **x38**)
h42:"_untruth_n_1_rel"(ARG0: **x38**)
h44:pron_rel(ARG0: **x3**)
h45:pronoun_q_rel(ARG0: **x3**,RSTR: **h46**,BODY: **h47**)
h48:"_cough_v_1_rel"(ARG0: **e2**,ARG1: **x3**)
h48:loc_nonsp_rel(ARG0: **e49**,ARG1: **e2**,ARG2: **x50**)
h51:udef_q_rel(ARG0: **x50**,RSTR: **h52**,BODY: **h53**)
h54:card_rel(CARG: "2",ARG0: **e56**,ARG1: **x50**)
h57:_or_c_rel(ARG0: **e58**,L-INDEX: **e56**,R-INDEX: **e59**,L-HNDL: **h54**,R-HNDL: **h60**)
h60:card_rel(CARG: "3",ARG0: **e59**,ARG1: **x50**)
h57:"_times_n_1_rel"(ARG0: **x50**)
h62:"_in+order+to_x_rel"(ARG0: **e63**,ARG1: **h64**,ARG2: **h65**)
h66:"_put_v_1_rel"(ARG0: **e67**,ARG1: **x3**,ARG2: **x68**,ARG3: **h69**)
h70:_the_q_rel(ARG0: **x68**,RSTR: **h71**,BODY: **h72**)
h73:"_little_a_1_rel"(ARG0: **e74**,ARG1: **x68**)
h73:"_prince_n_of_rel"(ARG0: **x68**,ARG1: **i75**)
h76:_in_p_rel(ARG0: **e77**,ARG1: **x68**,ARG2: **x78**)

The LinGO Grammar Matrix

(Bender et al 2002, Drellishak 2009, Bender et al 2010)



The LinGO Grammar Matrix: Combining typological breadth with syntactic depth (Bender 2016)

Table 1: Libraries in the Grammar Matrix and their typological sources.

Library	Citation	Typological sources
Coordination	Drellishak & Bender (2005)	Payne (1985); Stassen (2000); Drellishak (2004)
Person	Drellishak (2009)	Cysouw (2003); Siewierska (2004)
Number	Drellishak (2009)	Corbett (2000)
Gender	Drellishak (2009)	Corbett (1991)
Agreement	Drellishak (2009)	Corbett (2006)
Case	Drellishak (2009)	Comrie (1989); Dixon (1994)
Direct-inverse	Drellishak (2009)	Givón (1994)
Argument Optionality	Saleem (2010); Saleem & Bender (2010)	Ackema et al. (eds.) (2006); Dryer (2008)
Tense	Poulson (2011)	Comrie (1985); Dahl (1985); Bybee et al. (1994), *inter alia*
Aspect	Poulson (2011)	Comrie (1976); Dahl (1985); Bybee et al. (1994), *inter alia*
Sentential Negation	Crowgey (2012)	Dahl (1979); Dryer (2005)
Information Structure	Song (2014)	Féry & Krifka (2009); Buring (2010), *inter alia*
Adjectives	Trimble (2014)	Stassen (2003, 2013); Dixon (2004); Dryer (2013a), *inter alia*

The Grammar Matrix: Extensions

- Fokkens 2014: Meta-grammar engineering to facilitate multi-path exploration of grammar development

- Zamaraeva in prep: Incorporating extensive language-specific test-driven development into typologically-motivated library development

- AGGREGATION Project: Automatically creating grammars from IGT collections (Bender et al 2013, 2014, Zamaraeva et al 2017, 2019, Howell 2020)

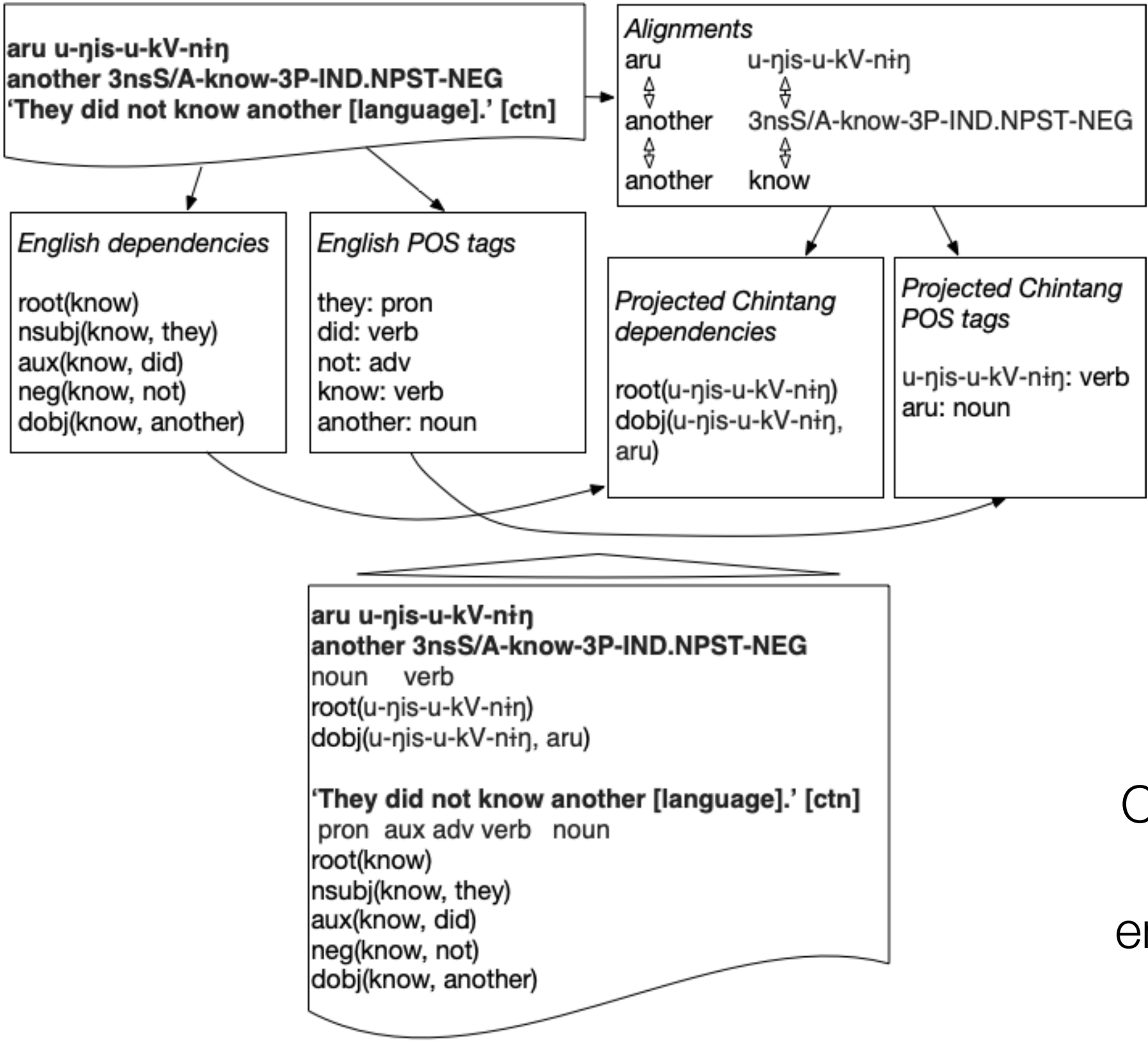
AGGREGATION Project: Motivation & overview

- Precision grammars are potentially useful for endangered language documentation (Bender et al 2012)

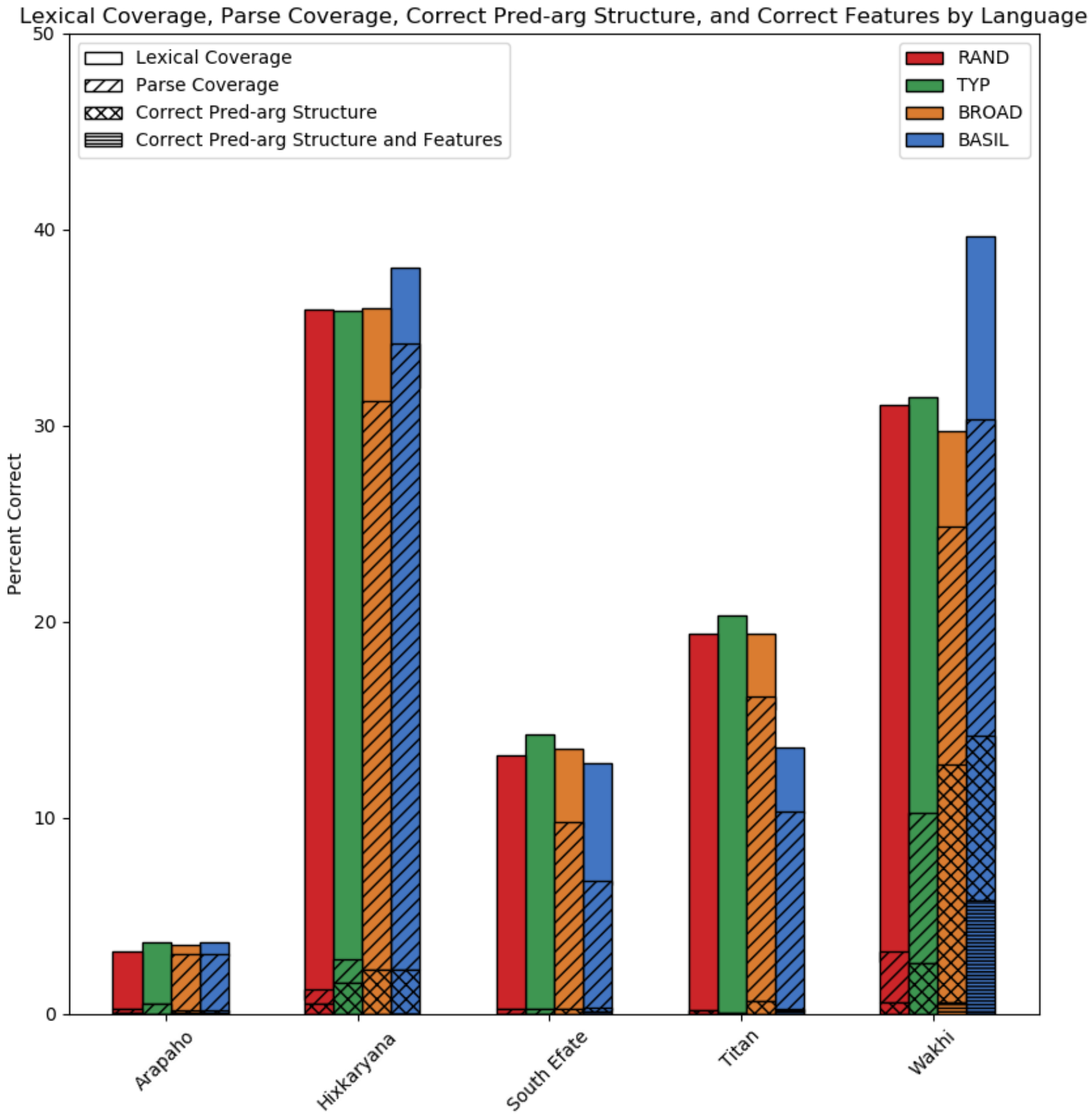
- Field linguists produce extremely rich annotations in the form of interlinear glossed text

- The Grammar Matrix provides a mapping from grammar specifications to precision grammars

- Can we infer sufficiently accurate and complete grammar specifications from IGT?

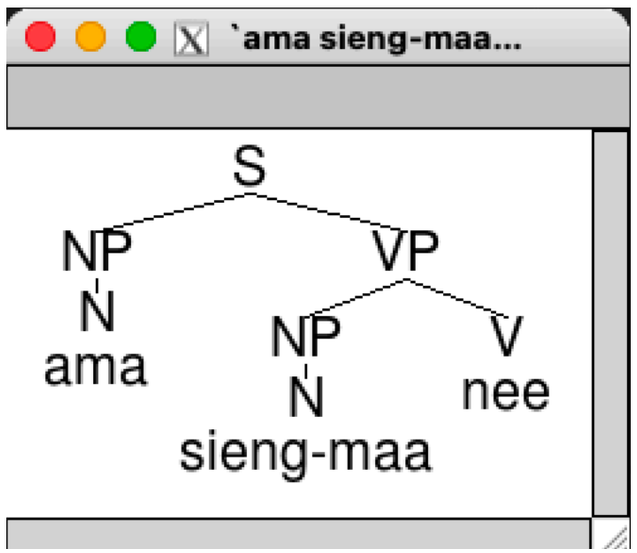


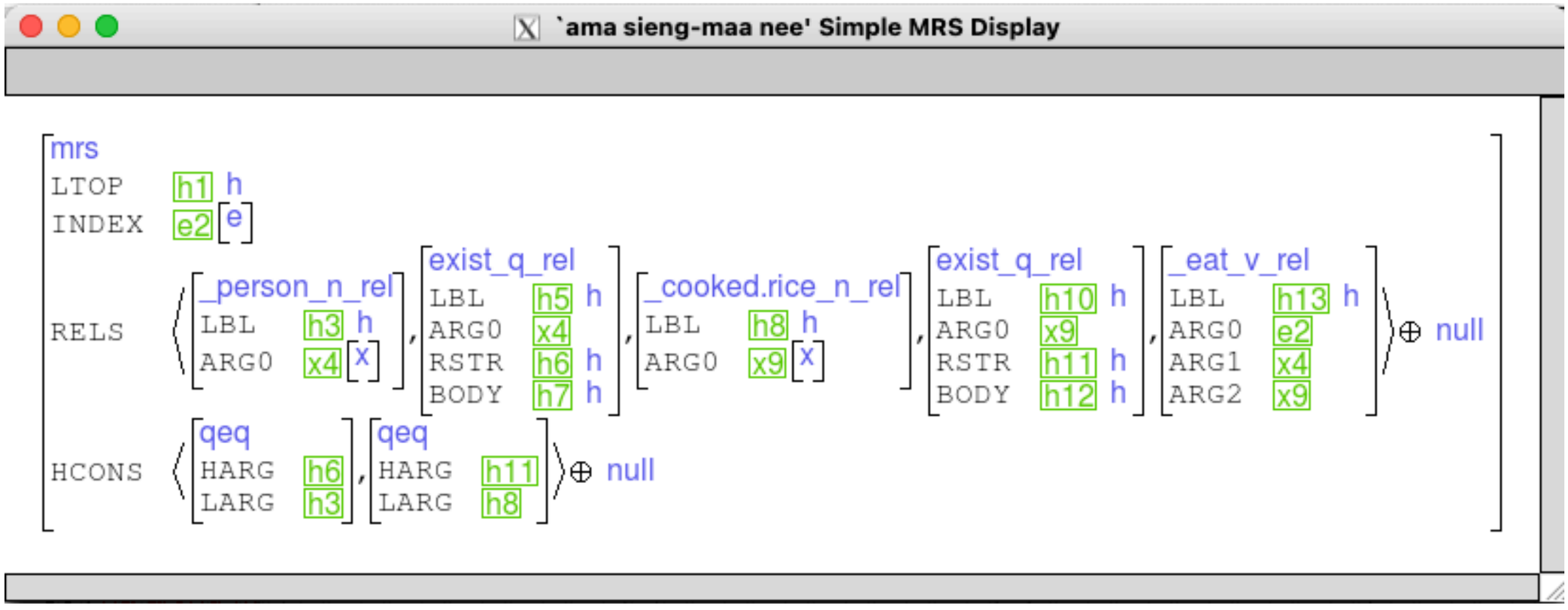
Chintang data from
Bickel et al 2009
enriched w/INTENT
(Georgi 2016)
Fig from
Howell 2020


Lexical Coverage, Parse Coverage, Correct Pred-arg Structure, and Correct Features by Language
Lexical Coverage
Parse Coverage
Correct Pred-arg Structure
Correct Pred-arg Structure and Features
RAND
TYP
BROAD
BASIL
Percent Correct
50
40
30
20
10
0
Arapaho
Hixkaryana
South Efate
Titan
Wakhi

(Howell 2020)

Sample parse for Abui [abz] sentence from auto-generated grammar (data from Kratochvíl)





Data-driven typology

- Autotypologizing databases (Bickel & Nichols 2002):
  - track existing categories and add new ones as needed
  - record exemplar sentences with each language/category entry
  - typology emerges, without pre-determined (‘etic’) categories
  - supports data-driven explorations of e.g. cross-linguistic applicability of semantic-role types (Bickel et al 2014)
- World Atlas of Language Structures (Dryer & Haspelmath 2013)
  - facilitate collaboration among many linguists on areal typology and correlations among features (Comrie et al 2013)

Mining word vectors

- Distribution in text reflects word meaning (Harris 1954, Firth 1957)

- Vector-space representations of word meaning can be used to trace patterns of lexical semantic change: Hamilton et al (2016) find that higher frequency words change senses more slowly and more polysemous words change sense more quickly, in English, French, German and Chinese

- Vector-space representations can be used to show stereotypes and biases as encoded in text corpora (Bolukbasi et al 2016, Caliskan et al 2017; see also Blodgett et al 2020)

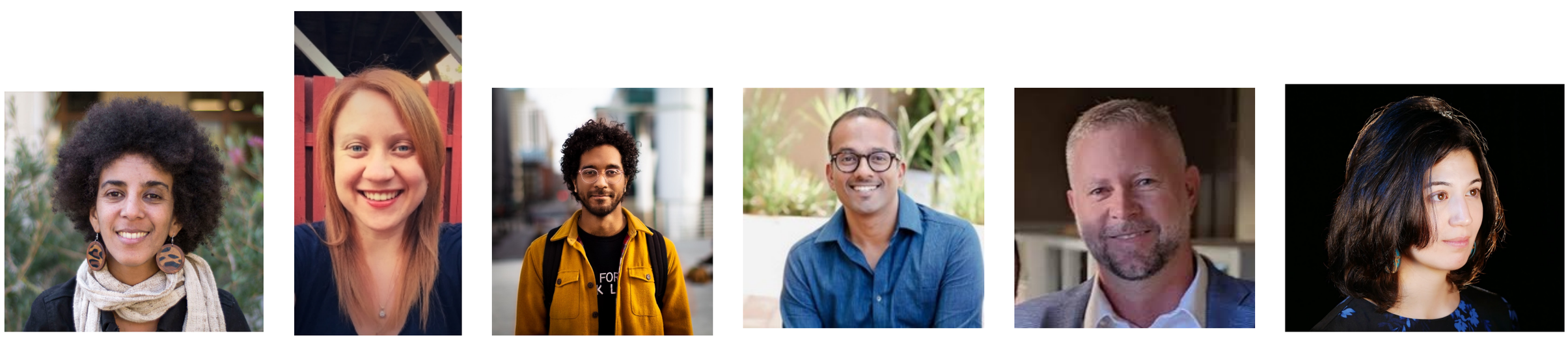
- … and used to study those biases and stereotypes (e.g. Herbelot et al 2012, Mohan 2020)

A cautionary note: Bigger isn't always better

- Scaling up theoretical models requires computational support, to handle larger datasets

- But very large datasets can also be used to drive natural language processing applications

- Broad deployment of language technology + the trend towards ever larger training datasets brings risks at a new scale



On the Dangers of Stochastic Parrots 🦜

(Bender et al 2021)

---

1. Environmental cost of training runs (and experimentation) (Strubell et al 2019, Henderson et al 2020)

2. Training data that is too large to document (GPT-3: 570GB), but sampled from sources known to overrepresent hegemonic viewpoints and frozen in time

3. Large anguage models (LLMs) encode and amplify biases from training data (Bolukbasi et al 2016, Caliskan et al 2017, Zhao et al 2017, Blodgett et al 2020)

4. LLMs produce seemingly coherent text and can bulldoze NLP tasks meant to test for language understanding (Bender and Koller 2020)

5. LLMs can be used to auto-populate message boards used to recruit extremists (McGuffie and Newhouse 2020)

6. LLM-driven NLP research shuts out researchers and languages with fewer resources

If linguistics is living its best life, the future of linguistics is **impactful in the world**

Language documentation work by and in service of speaker communities

- Leonard (2017, 2020) calls for decolonizing the notion of ‘language’ to promote language *reclamation* (rather than the frame of *revitalization*) by centering Indigenous notions of language.

- Gaby & Woods (2020) note the discrepancy of priorities between outsider linguists and Indigenous communities: “outsider linguists are far more likely to document the paradigm of case-marked pronouns than how a name is bestowed upon a baby, for example, or the song that lulls that baby to sleep” (p.e270)

- See also Bird (2020) on moving away from techno-solutionism in NLP for endangered languages

    - What can the academy do to make space for students and early-career researchers, member of Indigenous communities and otherwise, to take these approaches?

Linguistic analysis combatting racism

- Rickford & King (2016): how linguists can “help vernacular speakers be better heard in courtrooms and beyond”

- Wassink (2019): linguists can inform practice for speech-language pathologists to more accurately diagnose and treat speakers of non-standard varieties

- Voigt et al (2017): linguistic analysis of officer/community member interactions can document (at scale) differential treatment based on perceived race of community member

- Rosa & Flores (2017), Flores (2016): developing an understanding of the social co-construction of language and race can help move beyond harmful ‘deficit’ framings of e.g. Latinx bilinguals in the US.

Ethics and NLP and the linguist's-eye view

- Found that my CS colleagues were interested in information (sentiment, world knowledge, etc) encoded in language

- Where, as a linguist, I was focused on the language itself and how it shapes the information encoded

- Potential for harm when text is mistaken for an objective representation of the world (cf Speer 2017)

- Potential for harm when technology is developed only for speakers of prestige varieties

- Led to: Data statements proposal (Bender & Friedman 2018), stochastic parrots papers (Bender et al 2021)

Proposed value shifts

- Understand that theoretical work derives its value from the ways in which it supports current and future applied work

- Situate traditional academic conceptualizations of language as one-among-many, especially when working with communities on language projects

- Value public scholarship

- Expand curricula to include lessons how to apply what we know to the social world around us

- Value work that engages with communities and social justice, including aspects of the work that are grounded in the specifics of the particular situation

Research will become more accurate and more valuable when disciplinary understandings of rigor and impact go beyond restrictive notions that dominate linguistics and academia today, when researchers acknowledge their subjectivities, and when the discipline comes to see social impact as an inherent part of research and a valued contribution to scholarship, not as an optional addendum.

(Charity Hudley, Mallinson & Bucholtz 2020, p.e221)

If linguistics is living its best life…

- The future of linguistics is:

1. Broadly inclusive
2. Integrative
3. Computationally aided
4. Impactful in the world

References

Beltrama, A. (2020). Social meaning in semantics and pragmatics. *Language and Linguistics Compass*, 14(9):e12398.

Bender, E. M. (2016). Linguistic typology in natural language processing. *Linguistic Typology*, 20:645–660.

Bender, E. M. (2019). The #benderrule: On naming the languages we study and why it matters. *The Gradient*. `https://thegradient.pub/the-benderrule-on-naming-the-languages-we-study-and-why-it-matters/`.

Bender, E. M., Crowgey, J., Goodman, M. W., and Xia, F. (2014). Learning grammar specifications from IGT: A case study of chintang. In *Proceedings of the 2014 Workshop on the Use of Computational Methods in the Study of Endangered Languages*, pages 43–53, Baltimore, Maryland, USA. Association for Computational Linguistics.

Bender, E. M., Drellishak, S., Fokkens, A., Poulson, L., and Saleem, S. (2010). Grammar customization. *Research on Language & Computation*, pages 1–50. 10.1007/s11168-010-9070-1.

Bender, E. M., Flickinger, D., and Oepen, S. (2002). The grammar matrix: An open-source starter-kit for the rapid development of cross-linguistically consistent broad-coverage precision grammars. In Carroll, J., Oostdijk, N., and Sutcliffe, R., editors, *Proceedings of the Workshop on Grammar Engineering and Evaluation at the 19th International Conference on Computational Linguistics*, pages 8–14, Taipei, Taiwan.

Bender, E. M., Flickinger, D., and Oepen, S. (2011). Grammar engineering and linguistic hypothesis testing: Computational support for complexity in syntactic analysis. In Bender, E. M. and Arnold, J. E., editors, *Language from a Cognitive Perspective: Grammar, Usage and Processing*, pages 5–29. CSLI Publications, Stanford, CA.

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., Gebru, T., McMillan-Major, A., and et al (2021). On the dangers of stochastic parrots: Can language models be too big? 🦜 . In *Proceedings of FAccT 2021*.

Bender, E. M., Ghodke, S., Baldwin, T., and Dridan, R. (2012). From database to treebank: Enhancing hypertext grammars with grammar engineering and treebank search. In Nordhoff, S. and Poggeman, K.-L. G., editors, *Electronic Grammaticography*, pages 179–206. University of Hawaii Press, Honolulu.

Bender, E. M. and Good, J. (2010). A grand challenge for linguistics: Scaling up and integrating models. White paper contributed to NSF's SBE 2020 initiative.

Bender, E. M., Goodman, M. W., Crowgey, J., and Xia, F. (2013). Towards creating precision grammars from interlinear glossed text: Inferring large-scale typological properties. In *Proceedings of the 7th Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities*, pages 74–83, Sofia, Bulgaria. Association for Computational Linguistics.

Bender, E. M. and Koller, A. (2020). Climbing towards NLU: On meaning, form, and understanding in the age of data. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 5185–5198, Online. Association for Computational Linguistics.

Bickel, B., Comrie, B., and Haspelmath, M. (2008). The Leipzig glossing rules: Conventions for interlinear morpheme-by-morpheme glosses. Max Planck Institute for Evolutionary Anthropology and Department of Linguistics, University of Leipzig.

Bickel, B., Gaenszle, M., Rai, N. K., Lieven, E., Banjade, G., Bhatta, T. N., Paudyal, N., Pettigrew, J., Rai, I. P., Rai, M., Schikowski, R., and Stoll, S. (2009). Audiovisual corpus of the Chintang language, including a longitudinal corpus of language acquisition by six children, plus a trilingual dictionary, paradigm sets, grammar sketches, ethnographic descriptions, and photographs.

Bickel, B. and Nichols, J. (2002). Autotypologizing databases and their use in fieldwork. In *Proceedings of the International LREC Workshop on Resources and Tools in Field Linguistics*, Las Palmas.

Bickel, B., Zakharko, T., Bierkandt, L., and Witzlack-Makarevich, A. (2014). Semantic role clustering: An empirical assessment of semantic role types in non-default case assignment. *Studies in Language. International Journal sponsored by the Foundation Foundations of Language*, 38(3):485–511.

Bird, S. (2020). Decolonising speech and language technology. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 3504–3519, Barcelona, Spain (Online). International Committee on Computational Linguistics.

Bird, S. and Simons, G. (2001). The OLAC metadata set and controlled vocabularies. In *Proceedings of the ACL 2001 Workshop on Sharing Tools and Resources*.

Blodgett, S. L., Barocas, S., Daumé III, H., and Wallach, H. (2020). Language (technology) is power: A critical survey of "bias" in NLP. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 5454–5476, Online. Association for Computational Linguistics.

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 Lee, D. D., Sugiyama, M., Luxburg, U. V., Guyon, I., and Garnett, R., editors, *Advances in Neural Information Processing Systems 29*, pages 4349–4357. Curran Associates, Inc.

Bouavichith, D. A., Calloway, I. C., Craft, J. T., Hildebrandt, T., Tobin, S. J., and Beddor, P. S. (2019). Bidirectional effects of priming in speech perception: Social-to-lexical and lexical-to-social. *The Journal of the Acoustical Society of America*, 145(3):1911–1911.

Burnett, H. and Bonami, O. (2019). Linguistic prescription, ideological structure, and the actuation of linguistic changes: Grammatical gender in French parliamentary debates. *Language in Society*, 48(1):65–93.

Caliskan, A., Bryson, J. J., and Narayanan, A. (2017). Semantics derived automatically from language corpora contain human-like biases. *Science*, 356(6334):183–186.

Charity Hudley, A., Mallinson, C., and Bucholtz, M. (2020). Toward racial justice in linguistics: Interdisciplinary insights into theorizing race in the discipline and diversifying the profession. *Language*, 96(4):e200–e235.

Comrie, B., Dryer, M. S., Gil, D., and Haspelmath, M. (2013). Introduction. In Dryer, M. S. and Haspelmath, M., editors, *The World Atlas of Language Structures Online*. Max Planck Institute for Evolutionary Anthropology, Leipzig.

DiCanio, C. (2019). Linguistic common ground as privilege. Guest blog post on LanguageLog, accessible at `https://languagelog.ldc.upenn.edu/nll/?p=41758`.

Drellishak, S. (2009). *Widespread But Not Universal: Improving the Typological Coverage of the Grammar Matrix*. PhD thesis, University of Washington.

Dryer, M. S. and Haspelmath, M., editors (2013). *WALS Online*. Max Planck Institute for Evolutionary Anthropology, Leipzig.

Firth, J. R. (1957). A synopsis of linguistic theory, 1930-1955. *Studies in linguistic analysis*.

Flickinger, D. (2000). On building a more efficient grammar by exploiting types. *Natural Language Engineering*, 6(1 Special Issue on Efficient Processing with HPSG):15 – 28.

Flickinger, D. (2011). Accuracy v. robustness in grammar engineering. In Bender, E. M. and Arnold, J. E., editors, *Language from a Cognitive Perspective: Grammar, Usage and Processing*, pages 31–50. CSLI Publications, Stanford, CA.

Flores, N. (2016). A tale of two visions: Hegemonic whiteness and bilingual education. *Educational Policy*, 30(1):13–38.

Fokkens, A. S. (2014). *Enhancing Empirical Research for Linguistically Motivated Precision Grammars*. PhD thesis, Department of Computational Linguistics, Universität des Saarlandes.

Gaby, A. and Woods, L. (2020). Toward linguistic justice for Indigenous people: A response to Charity Hudley, Mallinson, and Bucholtz. *Language*, 96(4):e268–e280.

Goodman, M. W., Crowgey, J., Xia, F., and Bender, E. M. (2015). Xigt: Extensible interlinear glossed text for natural language processing. *Language Resources and Evaluation*, 49:455–485.

Hamilton, W. L., Leskovec, J., and Jurafsky, D. (2016). Diachronic word embeddings reveal statistical laws of semantic change. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1489–1501, Berlin, Germany. Association for Computational Linguistics.

Harris, Z. S. (1954). Distributional structure. *Word*, 10(2-3):146–162.

Hay, J. and Drager, K. (2010). Stuffed toys and speech perception. *Linguistics*, 48(4):865–892.

Henderson, P., Hu, J., Romoff, J., Brunskill, E., Jurafsky, D., and Pineau, J. (2020). Towards the systematic reporting of the energy and carbon footprints of machine learning. *arXiv preprint arXiv:2002.05651*.

Herbelot, A., von Redecker, E., and Müller, J. (2012). Distributional techniques for philosophical enquiry. In *Proceedings of the 6th Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities*, pages 45–54, Avignon, France. Association for Computational Linguistics.

Howell, K. (2020). *Inferring Grammars from Interlinear Glossed Text: Extracting Typological and Lexical Properties for the Automatic Generation of HPSG Grammars*. PhD thesis, University of Washington.

Lanehart, S. L. (2009). Diversity and intersectionality. In *Texas Linguistic Forum*, volume 53(7).

Leonard, W. (2017). Producing language reclamation by decolonising 'language'. *Language Documentation and Description*, 14:15–36.

Leonard, W. Y. (2020). Insights from Native American Studies for theorizing race and racism in linguistics (Response to Charity Hudley, Mallinson, and Bucholtz). *Language*, 96(4):e281–e291.

Mahler, T. (2020). The social component of the projection behavior of clausal complement contents. *Proceedings of the Linguistic Society of America*, 5(1):777–791.

McConnell-Ginet, S. (1984). The origins of sexist language in discourse. In White, S. J. and Teller, V., editors, *Discourses in Reading and Linguistics*, pages 123–135. New York Academy of Sciences, New York.

McGuffie, K. and Newhouse, A. (2020). The radicalization risks of GPT-3 and advanced neural language models. Technical report, Center on Terrorism, Extremism, and Counterterrorism, Middlebury Institute of International Studies at Monterrey. `https://www.middlebury.edu/institute/sites/www.middlebury.edu.institute/files/2020-09/gpt3-article.pdf`.

Mohan, P. (2020). An analysis of gender bias in K-12 assigned literature through comparison of non-contextual word embedding models. Master's thesis, University of Washington.

Nivre, J., de Marneffe, M.-C., Ginter, F., Hajič, J., Manning, C. D., Pyysalo, S., Schuster, S., Tyers, F., and Zeman, D. (2020). Universal Dependencies v2: An evergrowing multilingual treebank collection. In *Proceedings of the 12th Language Resources and Evaluation Conference*, pages 4034–4043, Marseille, France. European Language Resources Association.

Rickford, J. R. and King, S. (2016). Language and linguistics on trial: Hearing Rachel Jeantel (and other vernacular speakers) in the courtroom and beyond. *Language*, 92(4):948–988.

Rosa, J. and Flores, N. (2017). Unsettling race and language: Toward a raciolinguistic perspective. *Language in Society*, 46(5):621–647.

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.

Strand, E. A. and Johnson, K. (1996). Gradient and visual speaker normalization in the perception of fricatives. In *Natural Language Processing and Speech Technology: Results of the 3rd KONVENS Conference*, pages 14–26, Bielefed, Germany.

Strubell, E., Ganesh, A., and McCallum, A. (2019). Energy and policy considerations for deep learning in NLP. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 3645–3650, Florence, Italy. Association for Computational Linguistics.

Voigt, R., Camp, N. P., Prabhakaran, V., Hamilton, W. L., Hetey, R. C., Griffiths, C. M., Jurgens, D., Jurafsky, D., and Eberhardt, J. L. (2017). Language from police body camera footage shows racial disparities in officer respect. *Proceedings of the National Academy of Sciences*, 114(25):6521–6526.

Wassink, A. B. (2019). Where sociolinguistics and speech science meet: The physiological and acoustic consequences of underbite in a multilectal speaker of African-American English. In Buchstaller, I. and Blake, R., editors, *The Routledge Companion to the Work of John Rickford*. Routledge, New York.

Zamaraeva, O. (in prep). *Modeling Constituent Questions for a Grammar Engineering Resource*. PhD thesis, University of Washington.

Zamaraeva, O., Howell, K., and Bender, E. M. (2019). Handling cross-cutting properties in automatic inference of lexical classes: A case study of Chintang. In *Proceedings of the 3rd Workshop on the*

Use of Computational Methods in the Study of Endangered Languages, volume 1 Papers, pages 28–38, Honolulu, Hawai'i.

Zamaraeva, O., Kratochvíl, F., Bender, E. M., Xia, F., and Howell, K. (2017). Computational support for finding word classes: A case study of Abui. In *Proceedings of the 2nd Workshop on the Use of Computational Methods in the Study of Endangered Languages*, pages 130–140, Honolulu. Association for Computational Linguistics.