LIVEPERSON

Typologically Grounded Grammar Inference

Presenter: Kristen Howell, DSML manager at LivePerson
AGGREGATION Collaborators: Emily M. Bender (PI), Fei Xia, Olga Zamaraeva (and many others)

How can we generate linguistically rich grammars for typologically diverse, local* languages?

*small, primarily-oral languages, often Indigenous or endangered, including the original and emerging languages of Africa, Asia, Australia, the Americas, the Pacific, and the minority languages of Europe (Bird, 2022)

Outline of talk

- Precision grammars: what are they and what are they good for?
- Automatic grammar generation
- Grammar inference with BASIL
- Conclusion

Precision Grammars

Precision Grammars
… what are they?

- Machine readable
- Collection of syntactic rules
  - DELPH-IN JRF (Copestake, 2002)
- Can be loaded into software to parse sentences
- Output syntactic and semantic representations
  - HPSG (Pollard and Sag, 1994)
  - MRS (Copestake et al., 2005)

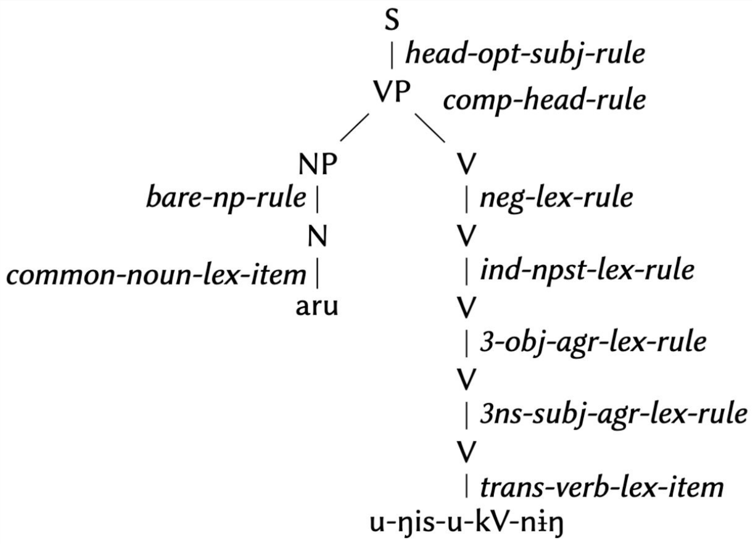
Syntactic rules

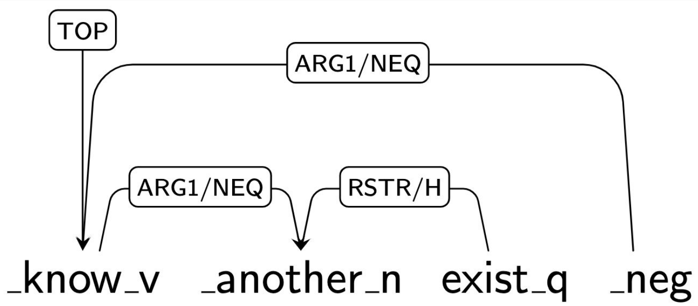
```
verb-pc5_lrt2-lex-rule := cont-change-only-lex-rule &
                          verb-pc5-lex-rule-super &
  [ C-CONT [ HOOK [ XARG #xarg,
                    LTOP #ltop,
                    INDEX #ind ],
             RELS <! event-relation & [ PRED "neg_rel",
                                        LBL #ltop,
                                        ARG1 #harg ] !>,
             HCONS <! qeq & [ HARG #harg,
                              LARG #larg ] !> ],
    SYNSEM.LKEYS #lkeys,
    DTR.SYNSEM [ LKEYS #lkeys,
                 LOCAL [ CONT.HOOK [ XARG #xarg,
                                     INDEX #ind,
                                     LTOP #larg ],
                         CAT.HEAD verb ] ] ].


verb-pc14_lrt1-suffix :=
%suffix (* -nɨŋ)
verb-pc14_lrt1-lex-rule.
```

Parses

Aru unisokonɨŋ.
`They did not know another [language].'
(Chintang [ctn]; Bickel et al., 2013a)





_know_v (ARG0 {SF *prop*, tense *npst*, aspect *ind*}, ARG1 {per *3*, num *ns* }, ARG2 {per *3*})

Precision Grammars
… what are they good for?

- Treebanking
  - Oepen et al. (2002)
- Data exploration
  - Letcher and Baldwin (2013)
  - Bouma et al. (2015)
- Developing grammar checkers
  - da Costa et al. (2016)
- Developing automatic tutors
  - Hellan et al. (2013)

So why don’t we see more of these grammars?

- They are very time consuming to write by hand
- Grammar engineering toolkits streamline this a lot, but there is still a lot of manual work
    - Grammar Matrix (Bender et al., 2002, 2010; Zamaraeva et al., forth)
    - CoreGram (Müller, 2015)
    - ParGram (Butt and King, 2002)

Grammar customization with the Grammar Matrix Bender et al. (2002, 2010), Zamaraeva et al. (forth)

Web-based questionnaire

- What's the morphosyntactic exponence of the negation construction you'd like to model?
- Is the negation morpheme bound or free?

Grammar Specification

```
section=general
  language=Chintang
  iso-code=ctn

section=sentential-negation
  neg-exp=1
  infl-neg=on
  neg-aux=on

section=morphology
  verb-pc14_name=verb-pc14
  verb-pc14_order=suffix
  verb-pc14_inputs=verb-pc1, verb-pc3, ...,
      verb-pc14_lrt1_feat1_name=negation
      verb-pc14_lrt1_feat1_value=plus
      verb-pc14_lrt1_feat1_head=verb
      verb-pc14_lrt1_lri1_inflecting=yes
      verb-pc14_lrt1_lri1_orth=-nɨŋ
```

Grammar

```
verb-pc5_lrt2-lex-rule := cont-change-only-lex-rule &
                          verb-pc5-lex-rule-super &
  [ C-CONT [ HOOK [ XARG #xarg,
                    LTOP #ltop,
                    INDEX #ind ],
             RELS <! event-relation & [ PRED "neg_rel",
                                        LBL #ltop,
                                        ARG1 #harg ] !>,
             HCONS <! qeq & [ HARG #harg,
                              LARG #larg ] !> ],
    SYNSEM.LKEYS #lkeys,
    DTR.SYNSEM [ LKEYS #lkeys,
                 LOCAL [ CONT.HOOK [ XARG #xarg,
                                     INDEX #ind,
                                     LTOP #larg ],
                         CAT.HEAD verb ] ] ].

verb-pc14_lrt1-suffix :=
%suffix (* -nɨŋ)
verb-pc14_lrt1-lex-rule.
```

Can we automate more of this process?

Automatic Grammar Generation

Grammar Induction from strings

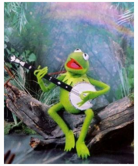
- Based on text alone
- Produces noisy representations
- Only partially align with trees created by linguists
- Identifying constituents and their categories
  - Klein and Manning (2001, 2002)
  - Bod (2009)
- Using neural nets
  - Hewitt and Manning (2019)
- Paired with an image
  - Shi et al. (2019)
  - Zhao and Titov (2020)

Grammar Induction from strings paired with meaning representations

- Based on text and some semantic representation
  - logical form (Kate et al., 2005; Kate and Mooney, 2006)
  - semantic dependency parse (Buys and Blunsom, 2017; Chen et al., 2018)
- Will produce semantic representations similar to those included in the input

Grammar Extraction

- Based on treebanks
- Will produce parse trees and/or semantic representations with roughly the same level of expressivity as those in the input
- Walk the trees, collect rules, prune
  - Krotov et al. (1994, 1998)
- May convert from one formalism to another
  - Xia (1999)
  - Hockenmaier and Steedman (2007)



KERMIT (Zanzoto et al., 2020) uses BERT (Devlin et al., 2019) to generate syntactic parses of the same form as those in training data, but does not generate a human readable grammar

Grammar Inference

- Can generate formal linguistic grammars without a treebank
- Based on text with partial grammatical information
  - Interlinear Glossed Text (Hellan, 2010; Bender et al. 2014)
  - POS, agreement, predicate-argument structure, clause type (Indurkhya, 2020)
- And some external source of linguistic knowledge
  - Grammar Matrix (Bender et al. 2014)
  - TypeGram (Hellan 2010)
  - Minimalist axioms, merge (Indurkhya, 2020)

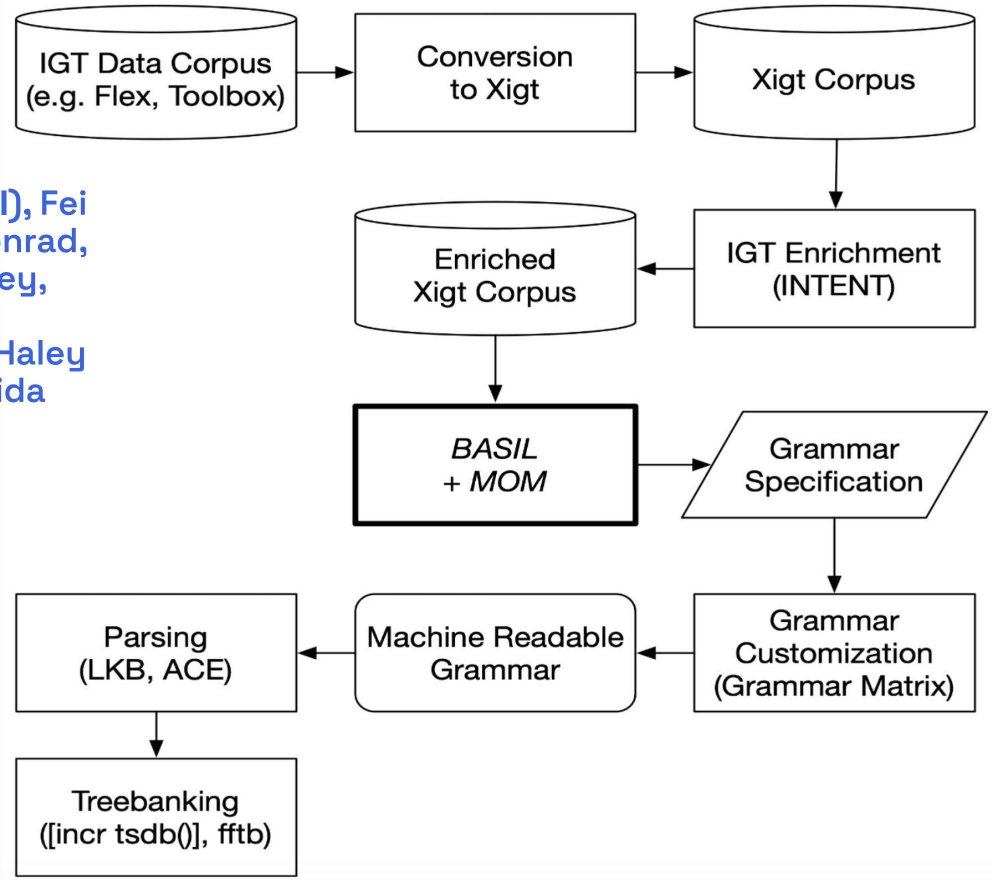
Where does typology fit in

- The external source of linguistic knowledge used in grammar inference should be cross-linguistically robust
- Inferring typological features is useful for defining grammar specifications
- A cross-linguistically generalizable inference system should be developed with typological diversity in mind

The AGGREGATION Project
Grammar Inference with BASIL

AGGREGATION

Collaborators: Emily M. Bender (PI), Fei Xia, Olga Zamaraeva, Elizabeth Conrad, Michael Goodman, Joshua Crowgey, David Wax, Ryan Georgi, Michael Lockwood, Swetha Ramaswamy, Haley Lepp, Claude Zhang, Tifa de Almeida



Grammar Inference with BASIL
Building Analyses from Syntactic Inference in Local Languages
Howell (2020), Howell and Bender (2022)

Lexicon
- Intransitive and transitive verbs*
- Common Nouns*
- Auxiliaries
- Pronouns
- Determiners
- Coordinating conjunctions
- Negation words
- Case-marking adpositions

Morphology
- Noun and verb lexical classes*

Syntax
- Word order
- Case system and case frame
- Argument optionality and argument marking on verbs
- Sentential negation
- Coordination

Syntactico-semantic Features
- Person, Number, Gender
- Tense, Aspect, Mood

*Using the MOM morphotactic inference system (Wax, 2014; Zamaraeva 2016; Zamaraeva et al. 2017)

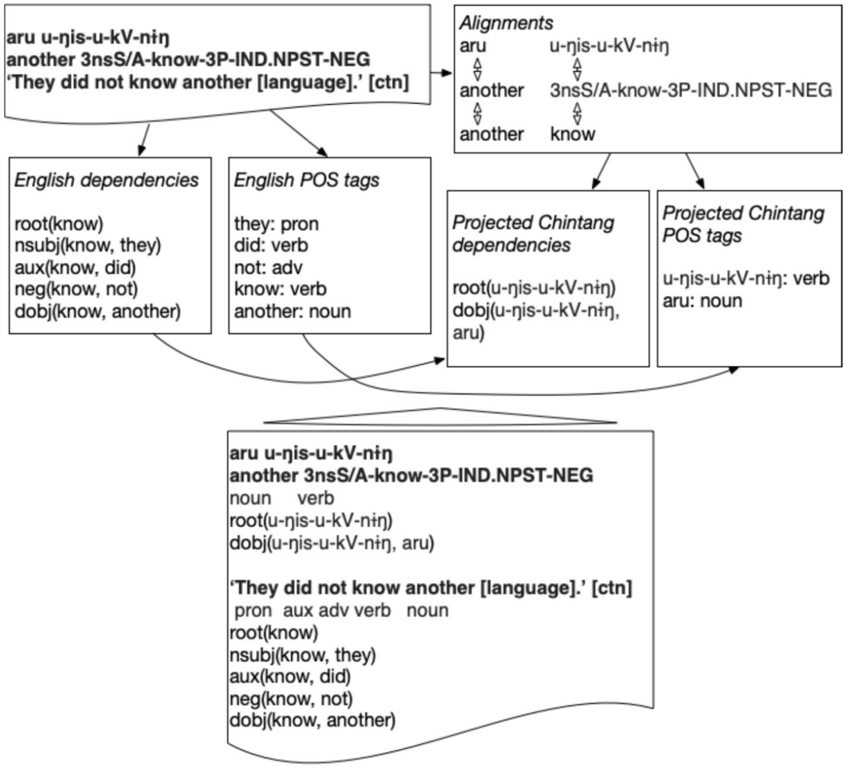
Mapping Features from WALS

- The WALS feature database is sparse, but the SIGTYP 2020 shared task (Bjerva et al. 2020) is a step towards increasing the available information
- de Almeida et al. (2019) found that 20 WALS features can be mapped directly to Grammar Matrix features
  - This accounts for 8.5% of the Grammar Matrix's specifications
- Zhang et al. (2019) did preliminary work to incorporate WALS features into inference

Grammar Inference with BASIL

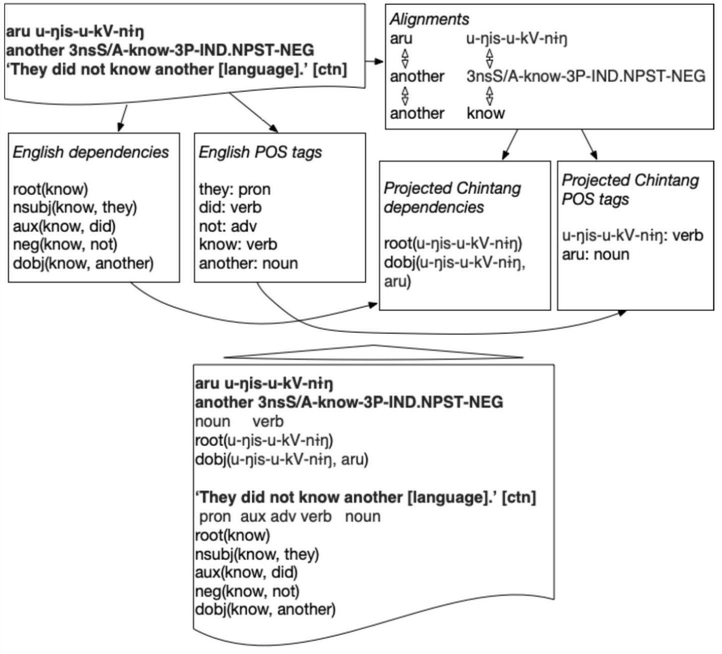
Building Analyses from Syntactic Inference in Local Languages

- Largely heuristic
- Leverages linguistic annotations
  - Morpheme segmentation
  - POS tags
  - Morphosyntactic features
- Leverages enriched IGT data (Georgi 2016)
  - Projected dependency parses
  - Projected POS tags



Generating a Grammar Specification

Enriched IGT



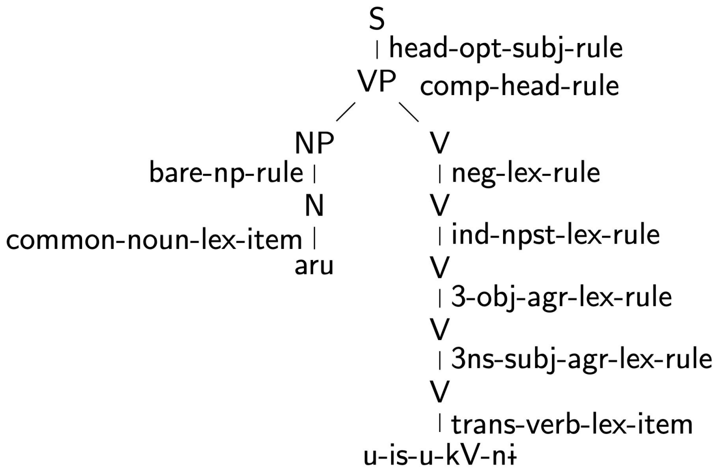
Grammar Specification

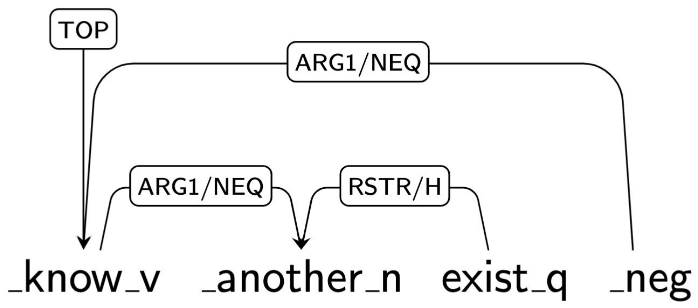
```
section=general
  language=Chintang
  iso-code=ctn

section=sentential-negation
  neg-exp=1
  infl-neg=on
  neg-aux=on

section=morphology
  verb-pc14_name=verb-pc14
  verb-pc14_order=suffix
  verb-pc14_inputs=verb-pc1, verb-pc3, ...,
      verb-pc14_lrt1_feat1_name=negation
      verb-pc14_lrt1_feat1_value=plus
      verb-pc14_lrt1_feat1_head=verb
      verb-pc14_lrt1_lri1_inflecting=yes
      verb-pc14_lrt1_lri1_orth=-nɨŋ
```

The Grammar Matrix generates a custom grammar, which produces rich syntactic and semantic representations





_know_v (ARG0 {SF *prop*, tense *npst*, aspect *ind*}, ARG1 {per *3*, num *ns* }, ARG2 {per *3*})

Languages for Development and Evaluation

red = development, blue = consulted, green = held out

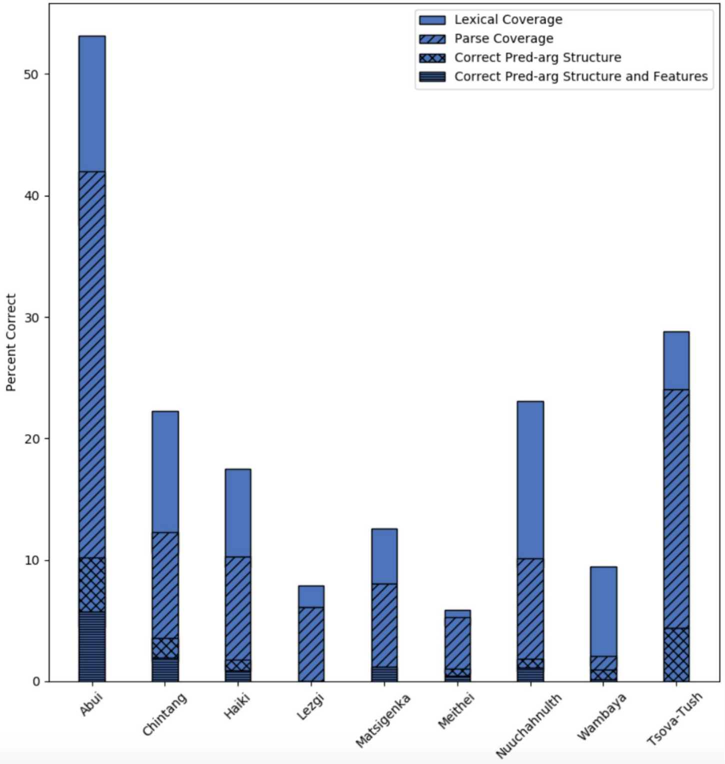
	Language	iso	Corpus	Resource
	Development			
1	Abui	abz	Kratochvíl 2019	Kratochvíl 2007
2	Chintang	ctn	Bickel et al. 2013b	Schikowski 2013
3	Matsigenka	mcb	Michael et al. 2013	Michael 2008
4	Nuuchahnulth	nuk	Inman 2019b	Inman 2019a
5	Wambaya	wmb	Nordlinger 1998	Nordlinger 1998
6	Haiki	yaq	Harley 2019	Sanchez et al. 2015
				Dedrick and Casad 1999
7	Lezgi	lez	Donet 2014b	Donet 2014a
8	Meithei	mni	Chelliah 2019	Chelliah 2011
9	Tsova-Tush	bbl	Hauk 2016–2019	Hauk and Harris forthcoming
				Hauk 2020
	Consulted			
10	Bardi	bcj	Bowern 2012	Bowern 2012
11	Ik	ikx	Schrock 2014	Schrock 2014
12	Old Javanese	jav	Acri 2018	
13	Yup'ik	esu	Miyaoka 2012	Miyaoka 2012
14	Basque	eus	Xia et al. 2016	de Urbina 1989
15	Dutch	nld	Xia et al. 2016	Booij 2002
16	Finnish	fin	Xia et al. 2016	Sulkala and Karjalainen 1992
17	Greek	ell	Xia et al. 2016	Holton et al. 2012
18	Hausa	hau	Xia et al. 2016	Newman 2000
19	Hungarian	hun	Xia et al. 2016	Kenesei et al. 2002
20	Indonesian	ind	Xia et al. 2016	Sneddon et al. 2012
21	Italian	ita	Xia et al. 2016	Monachesi 1996
22	Japanese	jpn	Siegel et al. 2016	Siegel et al. 2016
			Xia et al. 2016	Hinds 1986
23	Korean	kor	Xia et al. 2016	Sohn 1994
24	Mandarin	cmn	Xia et al. 2016	Li and Thompson 1989
25	Polish	pol	Xia et al. 2016	
26	Russian	rus	Xia et al. 2016	
27	Turkish	tur	Xia et al. 2016	Kornfilt 1997
	Held Out			
28	Arapaho	arp	Cowell 2018	Cowell and Moss Sr 2011
29	Hixkaryana	hix	Meira 2020	
30	South Efate	erk	Thieberger 2006a	Thieberger 2006b
31	Titan	ttv	Bowern 2019	Bowern 2011
32	Wakhi	wbl	Kaufman et al. 2020	

Development Corpora

- Iterative development with 9 languages from 7 language families
  - Primarily from FLEx and Toolbox corpora
- Consulted an additional 18 languages to total 19 families

	Language	ISO 639-3	Family	Source Type	Number of IGT
	Development				
1	Abui	abz	Trans-New Guinea	Toolbox	1568
2	Chintang	ctn	Sino-Tibetan	Toolbox	9785
3	Matsigenka	mcb	Arawakan	FLEx	349
4	Nuuchahnulth	nuk	Wakashan	FLEx	641
5	Wambaya	wmb	Mirndi	Book	818
6	Haiki	yaq	Uto-Aztecan	FLEx	2235
7	Lezgi	lez	Nakh-Daghestanian	FLEx	1168
8	Meithei	mni	Sino-Tibetan	FLEx	955
9	Tsova-Tush	bbl	Nakh-Daghestanian	FLEx	1601
	Consulted				
10	Bardi	bcj	Nyulnyulan	Book	178
11	Ik	ikx	Eastern Sudanic	Book	201
12	Old Javanese	jav	Austronesian	Toolbox	308
13	Yup'ik	esu	Eskimo-Aleut	Book	217
14	Basque	eus	Basque	ODIN	1033
15	Dutch	nld	Indo-European	ODIN	3543
16	Finnish	fin	Uralic	ODIN	3123
17	Greek	ell	Indo-European	ODIN	2065
18	Hausa	hau	Afro-Asiatic	ODIN	2504
19	Hungarian	hun	Uralic	ODIN	2077
20	Indonesian	ind	Austronesian	ODIN	1699
21	Italian	ita	Indo-European	ODIN	3513
22	Japanese	jpn	Japonic	ODIN	6655
				Book	116
23	Korean	kor	Korean	ODIN	5383
24	Mandarin	cmn	Sino-Tibetan	ODIN	5045
25	Polish	pol	Indo-European	ODIN	2691
26	Russian	rus	Indo-European	ODIN	4161
27	Turkish	tur	Altaic	ODIN	2617

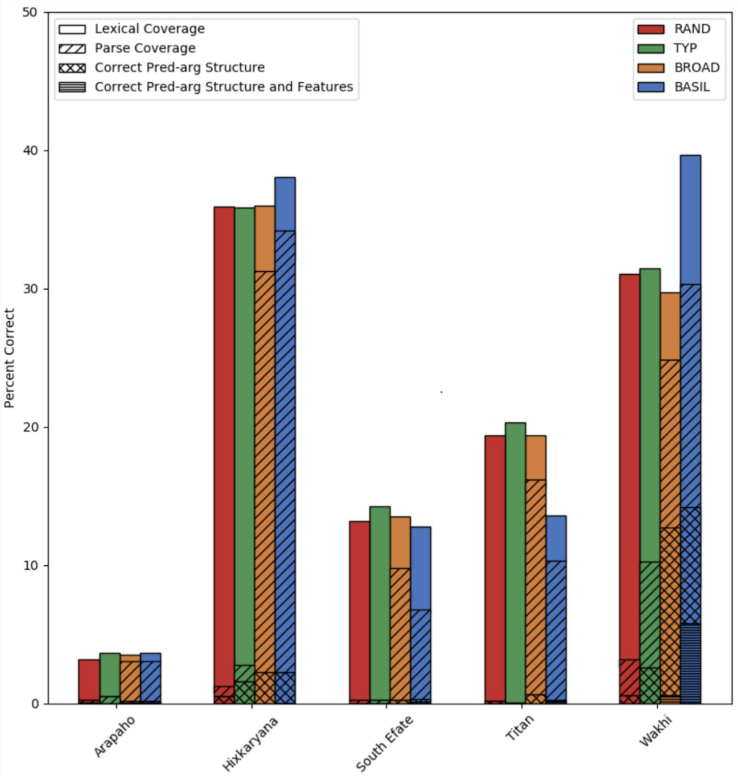
Iterative evaluation on development languages



Language [iso]	Lexical Coverage (%)	Parse Coverage (%)	Correct Pred-Arg Structure (%)	Correct Pred-Arg Strugure and Features (%)	Ambiguity
Abui [abz]	53.19	41.96	10.19*	5.73*	2195
Chintang [ctn]	22.29	12.24	3.58*	1.94*	5562
Haiki [yaq]	17.49	10.29	1.79*	0.89*	161
Lezgi [lez]	7.88	6.08	0.00*	0.00*	10419
Matsigenka [mcb]	12.61	8.02	1.15	1.15	2333
Meithei [mni]	5.86	5.24	1.05	0.42	3722
Nuuchahnulth [nuk]	23.09	10.14	1.87	1.09	265
Wambaya [wmb]	9.41	2.08	0.98	0.12	4
Tsova-Tush [bbl]	28.79	24.05	4.35*	0.00*	3418

Evaluation on Held-out Language Families

	Language	ISO 639-3	Family	Source Type	Number of IGT
1	Arapaho	arp	Algic	Toolbox	5000
2	Hixkaryana	hix	Cariban	Toolbox	5749
3	South Efate	erk	Austronesian	Toolbox	1875
4	Titan	ttv	Austronesian	Toolbox	1799
5	Wakhi	wbl	Indo-European	FLEx	683



Use in grammar engineering class

- Inferred grammars have been used as a starting point for students in a grammar engineering class at the University of Washington for 4 years
- Gives them parse coverage from the start
  - This makes some illustrative examples more discoverable
- Allows bigger lexicons with less work
  - Enables using corpora from linguists
- Saves time so they can get more out of the customization system
- Correcting the inferred morphology can be tedious

Current work in the AGGREGATION Project

- Reduce ambiguity in inferred grammars (Conrad, 2021)
- Improving root detection for morphological inference (Tara Wueger)
- New inference for adnominal possession (Allison Dods)
- New inference for valence-changing morphology (Yi-Chien Lin)
- Processing more data and fixing Matrix bugs (Tom Liu)

Conclusion

Conclusion

- Precision grammars are useful for treebanking, data exploration, language learning
- They are difficult to build
- Grammar inference leverages typological knowledge to generate these grammars automatically
- Grammars inferred with BASIL are small, but the parses they produce are linguistically rich

Thank you!

Thank you to all former and current contributors to the AGGREGATION Project, on whose work this research is built.

We are deeply indebted to speaker communities who provided linguistic data for documentation and the field linguists who shared it with us.

This material is based upon work supported by the National Science Foundation under Grant No. BCS-1561833 (PI Bender).

References

de Almeida, Tifa, Youyun Zhang, Kristen Howell, and Emily M. Bender. 2020. Feature Comparison across Typological Resources. Poster presented at the Second Workshop on Computational Research in Linguistic Typology.

Bender, Emily M, Dan Flickinger, and Stephan Oepen. 2002. The grammar matrix: An open-source starterkit for the rapid development of cross-linguistically consistent broad-coverage precision grammars. In Proceedings of the Workshop on Grammar Engineering and Evaluation at the 19th International Conference on Computational Linguistics, pages 8–14,Taipei.

Bender, Emily M, Sco Drellishak, Antske Fokkens, Laurie Poulson, and Safiyyah Saleem. 2010. Grammar customization. Research on Language & Computation, 8(1):23–72. 10.1007/s11168-010-9070-1.

Bender, Emily M, Joshua Crowgey, Michael Wayne Goodman, and Fei Xia. 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. Association for Computational Linguistics.

Bird, Steven. 2022. Local languages, third spaces, and other high-resource scenarios. In Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pages 7817–7829, Dublin, Ireland. Association for Computational Linguistics.

Johannes Bjerva, Elizabeth Salesky, Sabrina J. Mielke, Aditi Chaudhary, Giuseppe G. A. Celano, Edoardo Maria Ponti, Ekaterina Vylomova, Ryan Cotterell, and Isabelle Augenstein. 2020. SIGTYP 2020 Shared Task: Prediction of Typological Features. In Proceedings of the Second Workshop on Computational Research in Linguistic Typology, pages 1–11, Online. Association for Computational Linguistics.

Bod, Rens. 2009. From exemplar to grammar: A probabilistic analogy-based model of language learning. Cognitive Science, 33(5):752–793.

Bouma, Gosse, JM van Koppen, Frank Landsbergen, JEJM Odijk, Ton van der Wouden, and Matje van de Camp. 2015. Enriching a descriptive grammar with treebank queries. In Proceedings of the Fourteenth International Workshop on Treebanks and Linguistic Theories (TLT14), volume 14, pages 13–25.

Butt, Miriam and Tracy Holloway KING. 2002. Urdu and the Parallel Grammar Project In Proceedings of the 3rd Workshop on Asian Language Resources and International Standardization, volume 12 of COLING ’02, Stroudsburg. Association for Computational Linguistics.

Buys, Jan and Phil Blunsom. 2017. Robust incremental neural semantic graph parsing. In Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pages 1215–1226, Vancouver, Canada. Association for Computational Linguistics.

Chen, Yufei, Weiwei Sun, and Xiaojun Wan. 2018. Accurate SHRG-based semantic parsing. In Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pages 408–418, Melbourne, Australia. Association for Computational Linguistics.

Conrad, Elizabeth. 2021. Tracing and reducing lexical ambiguity in automatically inferred grammars. Master’s thesis, University of Washington.

Copestake, Ann. 2002a. Definitions of typed feature structures. In Stephan Oepen, Dan Flickinger, Junichi Tsujii, and Hans Uszkoreit, editors, Collaborative Language Engineering, pages 227–230. CSLI Publications, Stanford.

Copestake, Ann, Dan Flickinger, Carl Pollard, and Ivan A Sag. 2005. Minimal Recursion Semantics: An introduction. Research on Language and Computation, 3(2-3):281–332.

da Costa, Luis Morgado, Francis Bond, and Xiaoling He. 2016. Syntactic well-formedness diagnosis and errorbased coaching in computer assisted language learning using machine translation. In Proceedings of the 3rd Workshop on Natural Language Processing Techniques for Educational Applications (NLPTEA2016), pages 107–116.

References continued…

Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.

Georgi, Ryan. 2016. From Aari to Zulu: Massively Multilingual Creation of Language Tools Using Interlinear Glossed Text. Ph.D. thesis, University of Washington.

Hellan, Lars. 2010. From descriptive annotation to grammar specification. In Proceedings of the Fourth Linguistic Annotation Workshop, pages 172–176, Uppsala. Association for Computational Linguistics.

Hellan, Lars, Tore Bruland, Elias Aamot, and Mads H Sandøy. 2013. A grammar sparrer for Norwegian. In Proceedings of the 19th Nordic Conference of Computational Linguistics (NODALIDA 2013); May 22-24; 2013; Oslo University; Norway. NEALT Proceedings Series 16, 085, pages 435–439. Linköping University Electronic Press.

Hewitt, John and Christopher D Manning. 2019. A structural probe for finding syntax in word representations. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), pages 4129–4138, Minneapolis. Association for Computational Linguistics.

Hockenmaier, Julia and Mark Steedman. 2007. CCG-bank: A corpus of CCG derivations and dependency structures extracted from the Penn Treebank. Computational Linguistics, 33(3):355–396.

Howell, Kristen. 2020. Inferring Grammars from Interlinear Glossed Text: Extracting Typological and Lexical Properties for the Automatic Generation of HPSG Grammars. Ph.D. thesis, University of Washington.

Howell, Kristen and Emily M. Bender. 2022. Building Analyses from Syntactic Inference in Local Languages: An HPSG Grammar Inference System. Northern European Journal of Language Technology, 8(1). doi:10.3384/nejlt.2000-1533.2022.4017. https://nejlt.ep.liu.se/article/view/4017/3515.

Indurkhya, Sagar. 2020. Inferring Minimalist grammars with an SMT-solver. In Proceedings of the Society for Computation in Linguistics, volume 3.

Kate, Rohit J and Raymond J Mooney. 2006. Using string-kernels for learning semantic parsers. In Proceedings of the 21st International Conference on Computational Linguistics and the 44th annual meeting of the Association for Computational Linguistics, pages 913–920. Association for Computational Linguistics.

Kate, Rohit J, Yuk Wah Wong, and Raymond J Mooney. 2005. Learning to transform natural to formal languages. In Proceedings of the 1st AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment, pages 1062–1068.

Klein, Dan and Christopher D Manning. 2001. Natural language grammar induction using a constituent-context model. In Advances in neural information processing systems 14, pages 35–42.

Klein, Dan and Christopher D Manning. 2002. A generative constituent-context model for improved grammar induction. In Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics, pages 128–135, Philadelphia, Pennsylvania, USA. Association for Computational Linguistics.

Krotov, Alexander, Robert Gaizauskas, and Yorick Wilks. 1994. Acquiring a stochastic context-free grammar from the Penn Treebank. In Proceedings of the Irish Conference on NLP, Dublin.

Krotov, Alexander, Mark Hepple, Robert Gaizauskas, and Yorick Wilks. 1998. Compacting the Penn Treebank Grammar. In Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics (ACL-1998), pages 699–703, Montreal.

References continued...

Letcher, Ned and Timothy Baldwin. 2013. Constructing a phenomenal corpus: Towards detecting linguistic phenomena in precision grammars. In Proceedings of the Workshop on High-level Methodologies for Grammar Engineering at ESSLLI 2013, pages 25–36.

Müller, Stefan. 2015. The CoreGram project: Theoretical linguistics, theory development and verification. Journal of Language Modelling, 3(1):21–86.

Oepen, Stephan, Kristina Toutanova, Stuart Shieber, Chris Manning, Dan Flickinger, and Thorsten Brants. 2002. The LinGO Redwoods treebank. Motivation and preliminary applications. In Proceedings of the 19th International Conference on Computational Linguistics, Taipei.

Pollard, Carl and Ivan A Sag. 1994. Head-Driven Phrase Structure Grammar (Studies in Contemporary Linguistics). University of Chicago Press, Chicago.

Shi, Haoyue, Jiayuan Mao, Kevin Gimpel, and Karen Livescu. 2019. Visually grounded neural syntax acquisition. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, pages 1842–1861, Florence. Association for Computational Linguistics.

Wax, David. 2014. Automated grammar engineering for verbal morphology. Master's thesis, University of Washington.

Xia, Fei. 1999. Extracting tree adjoining grammars from bracketed corpora. In Proceedings of 5th Natural Language Processing Pacific Rim Symposium (NLPRS-1999), Beijing.

Zamaraeva, Olga. 2016. Inferring morphotactics from interlinear glossed text: combining clustering and precision grammars. In Proceedings of the 14th SIGMORPHON Workshop on Computational Research in Phonetics, Phonology, and Morphology, pages 141–150.

Zamaraeva, Olga, František Kratochvíl, Emily M Bender, Fei Xia, and Kristen Howell. 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.

Zamaraeva, Olga, TJ Trimble, Kristen Howell, Michael Wayne Goodman, Antske Fokkens, Guy Emerson, Chris Curtis, and Emily M Bender. Forthcoming. 20 years of the Grammar Matrix: Cross-linguistic hypothesis testing of increasingly complex interactions. Journal of Language Modeling.

Zhang, Youyun, Tifa de Almeida, Kristen Howell, and Emily M. Bender. 2020. Using Typological Information in WALS to Improve Grammar Inference. Poster presented at the Second Workshop on Computational Research in Linguistic Typology.

Zanzoto, Fabio Massimo, Andrea Santilli, Leonardo Ranaldi, Dario Onorati, Pierfrancesco Tommasino, and Francesca Fallucchi. 2020. KERMIT: Complementing transformer architectures with encoders of explicit syntactic interpretations. In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), pages 256–267.

Zhao, Yanpeng and Ivan Titov. 2020. Visually grounded compound PCFGs. In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), pages 4369–4379, Online. Association for Computational Linguistics.