University of Washington site update

DELPH-IN Summit
July 13, 2020
Zoom

Emily M. Bender 🐙, Emma Bateman, Liz Conrad, Kristen Howell, Angie McMillan-Major, Woodley Packard, Lonny Strunk, Olga Zamaraeva, Claude Zhang

567 Languages (SP 2019)

abz	Abui	Alor-Pantar	Kim
mni	Meithei	Sino-Tibetan	Ahn
yaq	Hiaki	Uto-Aztecan	Conrad
sje	Pite Saami	Uralic	Nielsen, Spivey
eng	English	Indo-European	

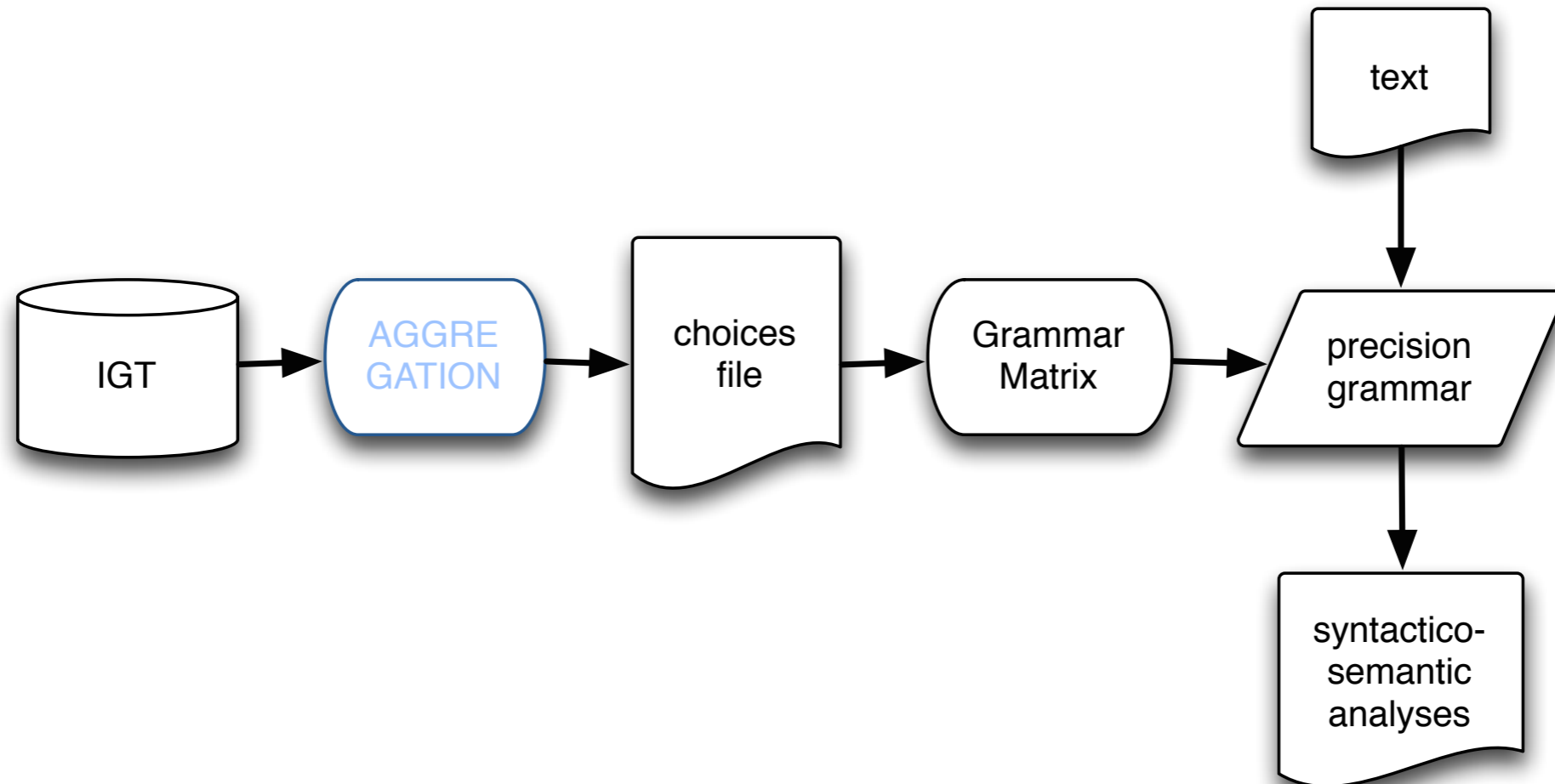# Items with end-to-end output: Final (transfer rule propagation)

	abz	eng	mni	sje	yaq
abz	12	12	9	10	10
eng	13	26	13	21	14
mni	7	13	13	8	8
sje	10	16	9	18	11
yaq	11	15	11	13	15

Big changes for 567 2020: AGGREGATION-based, take 2

- Field languages only

- Start with inference-created choices files

- Again: Lots of useful feedback to Kristen

- BIG choices files hard to work with

- Maybe worth it, especially for leg up in morphology

- Lots of ambiguity, especially from morphological component => Liz Conrad is looking into mitigating this

AGGREGATION project
(Bender, Zamaraeva, Howell, Zhang, Bateman + others)

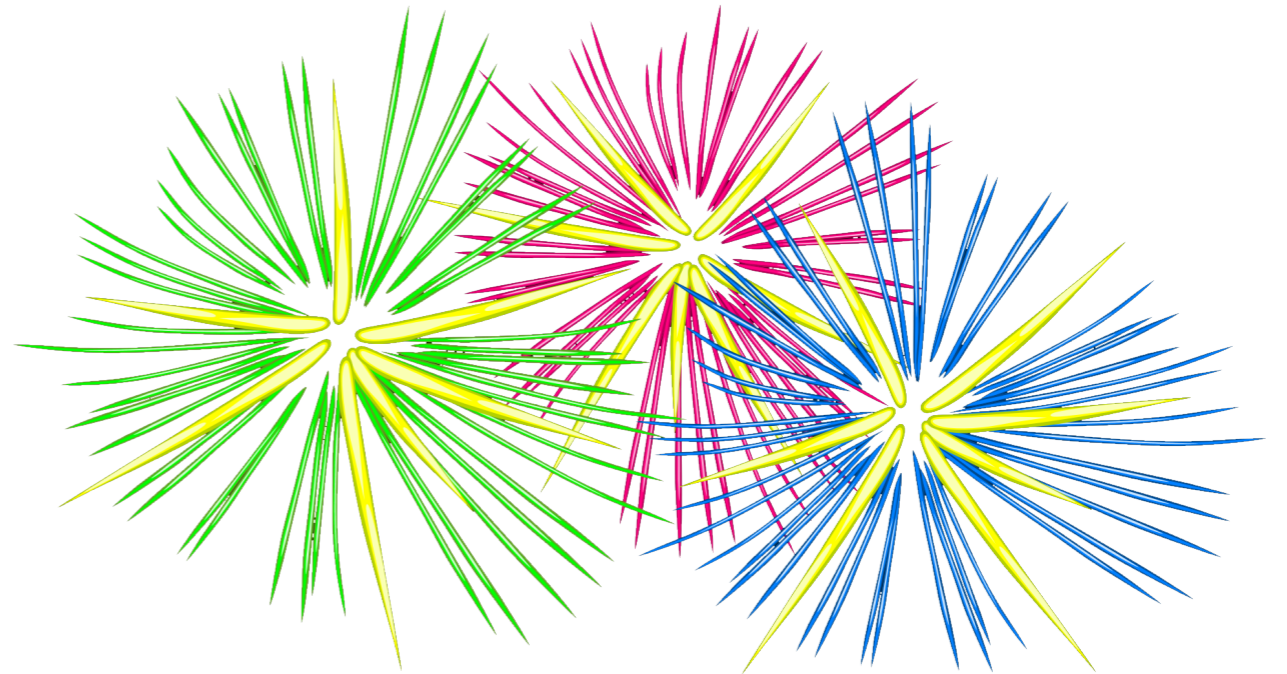# AGGREGATION - Phase II & related

- Integrating and extending the inference code (Howell)

- Large-scale evaluation on dev & held-out languages (Howell)

- Injecting WALS information into inference (Zhang)

- Improvements to MOM GUI (Bateman)

- Planned: Improvements to morphological inference to reduce ambiguity (Conrad)

Grammar engineering: syntax and morphology

- Matrix library for wh-questions (Zamaraeva)

  - Pushing on general infrastructure for non-local features, append lists

- Morphological analyser for Yup'ik (Strunk)

Dr. Kristen Howell

- PhD complete :)

- Title: *Inferring Grammars from Interlinear Glossed Text: Extracting Typological and Lexical Properties for the Automatic Generation of HPSG Grammars*

- Presented in more detail later this session

Data Statements (Bender & Friedman 2018)

- Documentation of what's in a dataset

  - Support reasoning about appropriate deployment of systems trained on a given data set

  - Support appropriate scoping of scientific claims

- May 2020: "LREC" workshop "Data Statements for NLP: Towards Best Practices"

- Currently (with Friedman and McMillan-Major): Analyzing output of that workshop

- At DELPH-IN 2020: Data statement for Redwoods?