University of Washington site update

DELPH-IN Summit
July 19, 2021
Zoom

Emily M. Bender, Liz Conrad, Allison Dods, Kristen Howell, Angie McMillan-Major,
Lonny Strunk, Olga Zamaraeva

567 Languages (WI 2021)

abz	Abui	Alor-Pantar	Dodds, A. Chen
arp	Arapaho	Algic	Barnes, Nguyen
mcb	Matsigenka	Arawakan	Ding, Tu
ttv	Titan	Austronesian	Krug, Imel
wbl	Wahki	Indo-European	Su, C. Chen
yaq	Hiaki	Uto-Aztecan	Shi
sje	Pite Saami	Uralic	Nielsen, Spivey
eng	English	Indo-European	

567 Languages (WI 2020)

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	arp	eng	mcb	sje	ttv	wbl	yaq
abz	15	10	14	11	12	13	10	11
arp	10	10	11	8	8	8	7	7
eng	15	11	26	20	21	20	12	17
mcb	7	5	12	13	11	10	6	11
sje	11	9	18	15	18	15	9	13
ttv	13	10	18	17	17	19	12	14
wbl	7	5	5	6	5	5	8	5
yaq	11	9	18	16	15	14	10	19

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

Transitions

- PhD complete!!

  - Olga Zamaraeva 2021 *Assembling Syntax: Modeling Constituent Questions in a Grammar Engineering Framework*

- MS complete!!

  - Lonny Alaskuk Strunk 2020 *A Finite-State Morphological Analyzer for Central Alaskan Yup'ik*

  - Liz Conrad 2021 *Tracing and Reducing Lexical Ambiguity in Automatically Inferred Grammars*

- Starting PhD: Liz Conrad

- Started MS: Allison Dods

AGGREGATION project
(Bender, Zamaraeva, Howell, Conrad, Dods, Zhang, Bateman + others)
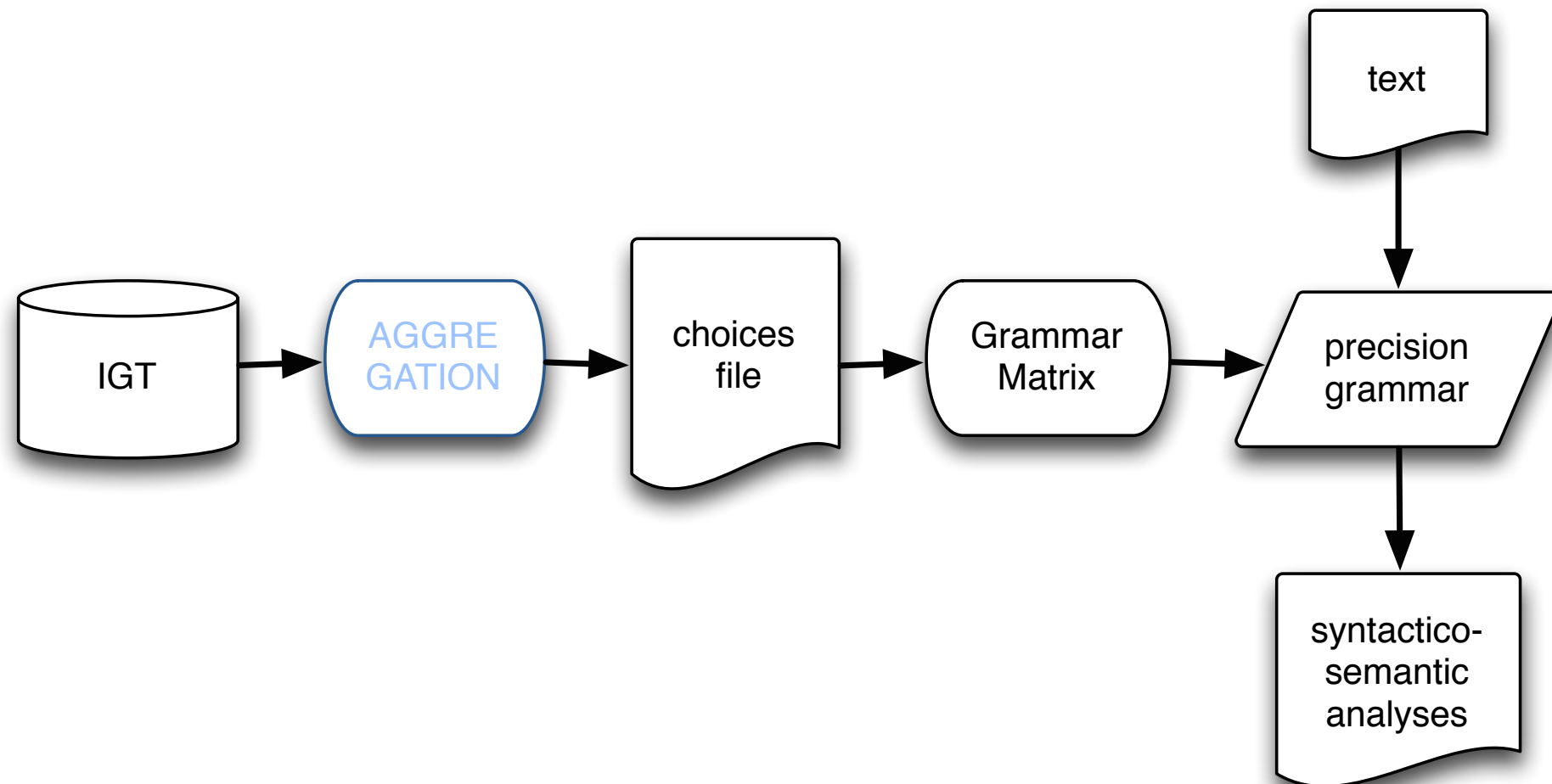
Grammar engineering: syntax and morphology

- Matrix library for wh-questions (Zamaraeva)

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

- Morphological analyser for Yup'ik (Strunk)

Data Statements (Bender & Friedman 2018, McMillan-Major, Bender & Friedman, forthcoming)

- 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"

- Imminent release: schema v2 + How To document based on results of that workshop

- At DELPH-IN 2021: Data statements for treebanks [Day 3]