University of Washington (& beyond) site update

DELPH-IN Summit
July 14, 2014
Tomar, Portugal

Emily M. Bender, Joshua Crowgey, Antske Fokkens, Michael Wayne Goodman, David, Inman, Woodley Packard, Glenn Slayden,

Grammar Engineering

- Adjectives library nearly done (TJ)
- More work on Thai (Glenn)
- Ling 567 languages: Bardi [bcj], Nanti [cox], Gilaki [glk], Marathi [mar], Meithei [mni], Burmese [mya], Chichewa [nya], Uyghur [uig], Yukaghir [yux]
- ... of which at least two will see some further development for HPSG 2014 papers, if not further: yux [Olga] and cox [David]

2014 567 languages - mapped

lat/long data mostly from wals.info; map by batchgeo.com

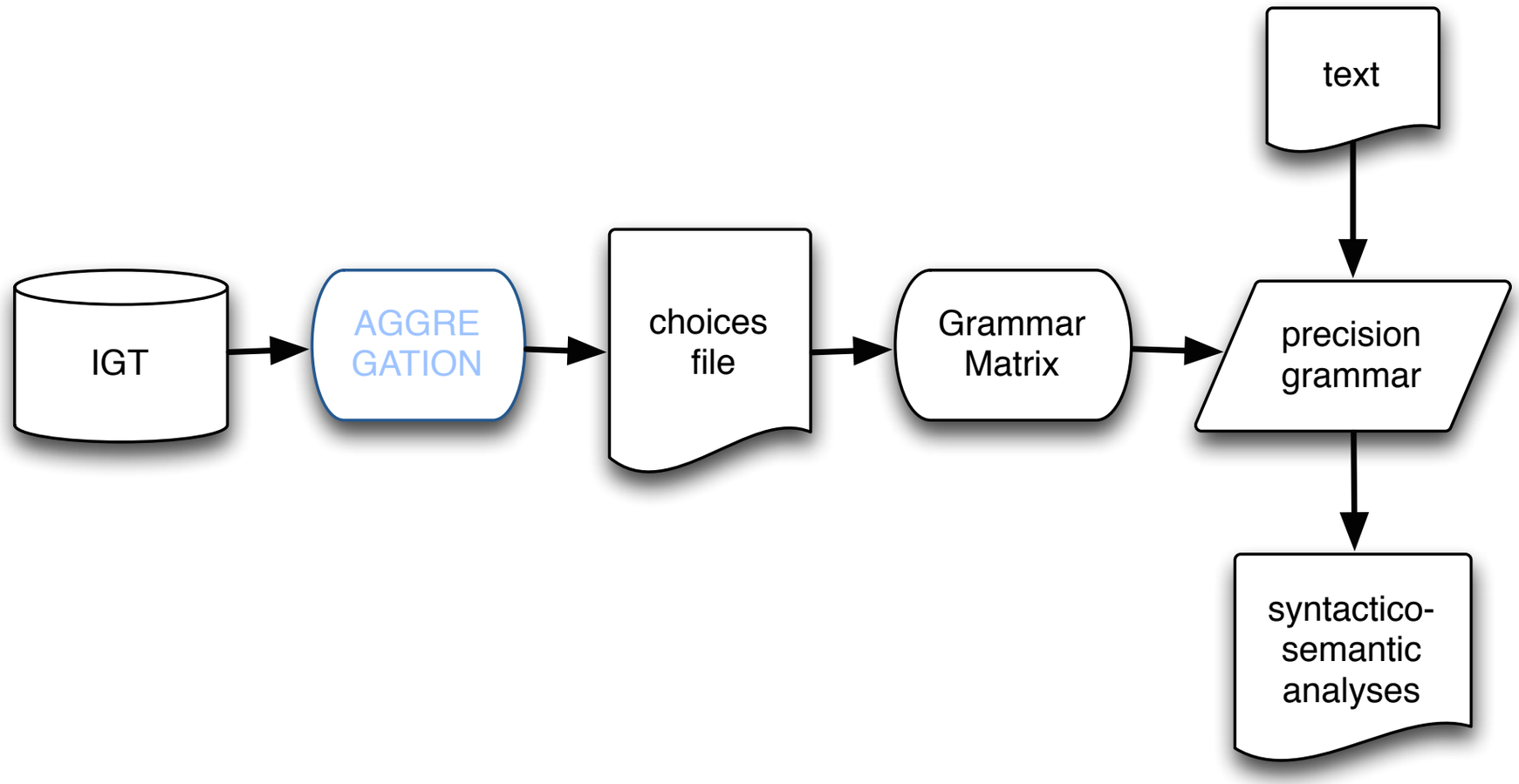
Items with end-to-end output (post transfer rule propagation)

src \ tgt	bcj	cox	eng	frr	glk	mar	mni	mya	nya	nyb	uig	yux	avg
bcj	18	12	7	7	7	6	7	7	12	12	11	6	9.33
cox	10	16	12	9	12	9	8	8	2	10	14	8	9.83
eng	10	16	19	18	18	16	13	18	12	12	16	17	15.42
frr	11	15	18	19	17	16	14	16	12	12	16	16	15.17
glk	11	15	18	16	19	16	14	17	13	12	15	16	15.17
mar	11	14	17	17	15	17	12	15	12	12	15	15	14.33
mni	8	11	12	12	12	12	12	11	10	10	11	11	11.00
mya	8	13	13	11	12	10	8	13	9	9	14	13	11.08
nya	10	14	12	12	12	11	11	11	13	14	13	10	11.92
nyb	11	16	12	12	12	12	11	11	13	15	15	10	12.50
uig	10	14	15	14	15	15	13	15	11	11	15	14	13.50
yux	8	13	15	15	14	13	12	14	9	9	14	15	12.58
avg	10.50	14.08	14.17	13.50	13.75	12.75	11.25	13.00	10.67	11.50	14.08	12.58	

Language CoLLAGE

- Collection of Language Lore Amassed through Grammar Engineering
- Testsuites (IGT), choices files, grammars, write-ups, feedback
- For 50 languages (and counting); of which data for 11 have been curated
- www.delph-in.net/matrix/language-collage/
- LREC 2014 paper

AGGREGATION project



Focus in the past year: Extensive case study of one language (Chintang, ctn)

	Training Data (N = 8863)				Test Data (N = 930)			
choices file	lexical coverage (%)	items parsed (%)	items correct (%)	average readings	lexical coverage (%)	items parsed (%)	items correct (%)	average readings
ORACLE	1165 (13)	174 (3.5)	132 (1.5)	2.17	116 (12.5)	20 (2.2)	10 (1.1)	1.35
BASELINE	1276 (14)	398 (7.9)	216 (2.4)	8.30	41 (4.4)	15 (1.6)	8 (0.9)	28.87
FF-AUTO-NONE	1276 (14)	354 (4.0)	196 (2.2)	7.12	41 (4.4)	13 (1.4)	7 (0.8)	13.92
FF-DEFAULT-GRAM	911 (10)	126 (1.4)	84 (0.9)	4.08	18 (1.9)	4 (0.4)	2 (0.2)	5.00
FF-AUTO-GRAM	911 (10)	120 (1.4)	82 (0.9)	3.84	18 (1.9)	4 (0.4)	2 (0.2)	5.00
MOM-DEFAULT-NONE	1102 (12)	814 (9.2)	52 (0.6)	6.04	39 (4.2)	16 (1.7)	3 (0.3)	10.81
MOM-AUTO-NONE	1102 (12)	753 (8.5)	49 (0.6)	4.20	39 (4.2)	10 (1.1)	3 (0.3)	9.20

Details in ACL 2014 workshop paper

MRS in Applications Seminar: Course goals

- Explore NLP tasks which can be improved with semantic features
- Understand what information is captured by the ERG's MRS output that is relevant to those tasks
- Experience with feature design
- Add MRS features to an existing baseline system, and measure the result
- Experience with error analysis
- Experience with academic writing in CL/NLP

MRS in Applications Seminar: Assignments

- Week 1: Sample MRS output from the ERG
- Weeks 4-5: Target task/baseline system descriptions
  - Required: open-source baseline system to add to, existing gold standard data + evaluation metric (and ideally eval script)
    - Baseline systems generally took an ML approach
- Week 6: Evaluation plan, preliminary error analysis of baseline system output
- Week 8: Feature design
- Weeks 9-10+: Final project, presentation, ACL-style writeup
  - Detailed instructions, resulted in well-written papers

Tasks

- Coref resolution, building on Stanford NLP's sieve-based system (Lee et al 2011):
    - Aldrich, Trimble: Adding ERG-based PNG info; adding ERG-based notion of parallel argument roles (negative result)
    - Carter: Adding handling of cataphora, leveraging focus_d_rel (small improvement)
- Extractive summarization, building on MEAD (Radev et al 2004)
    - Benak: Put more weight on more closely connected keywords (where connections are measured in terms of the MRS graph) (small improvement)

Tasks (cont)

- WSD, building on/inspired by SenseLearner (Mihalcea & Faruque 2004)
    - Packard: MRS features based on predicate symbol, argument type, role +predicate, and role+backed off predicate (small, ephemeral positive results)
- Sentiment analysis, building on the Naive Bayes, n-gram approach of Narayanan et al 2014
    - Gordon & Kramer: Add EDS-style triples as features (small positive result)
    - Accepted to *SEM 2014

Take-aways

- ‘MRS casserole’ approach to adding input from ERG to ML systems seems promising
- Working, reasonably easily modifiable baseline systems were hard to come by
- Emphasis on error analysis was fruitful

ERG-MRS in Ling 573: Question Answering

- Team MRS: Woodley, TJ, Melanie Bolla

- Parse queries, transfer to statement MRS, and regenerate to create new queries for standard IR

- Parse document corpus, compare MRSs to query MRSs and used (heuristic) match score as a reranker feature

- Best performing system of those that used only provided documents (not web search)

Congratulations

- to Dr. Sanghoun Song
- to Antske, whose thesis is now 99.9% complete