University of Washington (& beyond) site update

DELPH-IN Summit
July 2, 2012
Sofia, Bulgaria

Emily M. Bender, Michael Wayne Goodman, Sanghoun Song, David Wax, Joshua Crowgey, Glenn Slayden, Zina Pozen, Megan Schneider, Matt Hohensee, Francesca Gola, Antske Fokkens

Previews

- Negation library now on firm typological foundations
- Semantics of negation in cross-linguistic perspective
- Information structure library
- ICONS
- Discussion/Q&A on using ACE for parsing and generation
- Demo of a new browser-based full-forest treebanking tool
- CLIMB (meta-grammar engineering)
- Phenomena catalogue
- Thesis ideas/feedback (Mike, Zina, Ned)

Grammar engineering for language documentation

- AGGREGATION: funded (at 50% for pilot; 2 year project)
    - Goals: automatically extract answers to Grammar Matrix questionnaire from IGT (ODIN, field projects)
    - Initial steps: modeling verbal morphology (Bender et al 2012; Wax in progress)
- Ling 567:
    - Collaborations with field linguists: Ingush [inh] (Nichols), Hup [jup] (Epps), Sri Lanka Malay [sci] (Nordhoff)
    - Other: Georgian [kat], Bosnian-Serbo-Croatian [bcs]

CARBON Treebank

- NSF funding denied

- Survey indicated strong interest from the field => mailing list to which to advertise similar DELPH-IN resources

- MS project (Schneider) on projecting ERG derivations to PTB-style trees

  - Adapt step-wise, closer and closer to PTB representation (bracketing syntax only, labels, tokenization, flatter structures, empty elements)

  - Train Stanford (or similar) parser on the various versions of the treebank and measure performance on the DDEC evaluation set (Bender et al 2011)

Fokkens et al 2011: Spring Cleaning and Grammar Compression

- Spring cleaning: removing redundant types, i.e., those that are neither *instantiated, introducers of features, or lower bounds* for two other types. (output is tdl)
- Grammar compression: take flopped grammar and remove all types that are neither directly *referenced* in instances nor *glbs* of referenced types
- Applied spring cleaning to a variety of Matrix-derived grammars and found a large proportion of Matrix core types going unused, relatively few language-specific types
- Applied grammar compression to big & medium grammars and found compression rates of 37% (BURGER), 48% (ERG), 51% (Jacy), 52% (GG), and 78% (Wambaya)
- The beginning of a larger enterprise of grammar comparison and exploration...

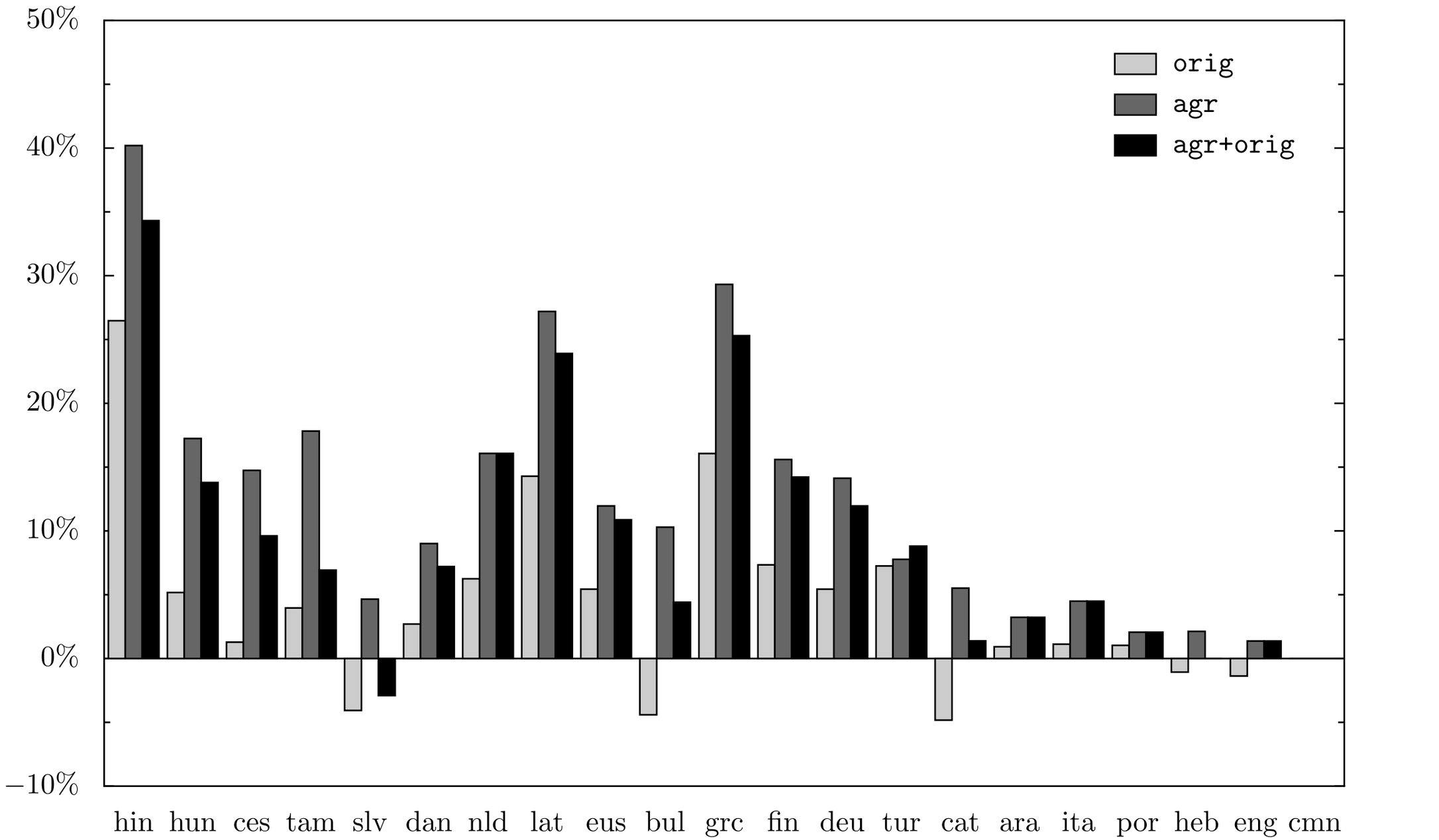
Hohensee 2012: It's only morpho-logical: Modeling agreement in cross-linguistic dependency parsing

- Previous work on language-independent dependency parsing was stymied by morphologically complex languages

- Part of the problem: no modeling of agreement

- Modified MSTParser (McDonald et al, 2006) by adding agreement features:
  - <attr>_agrees, head=<headPOS>, dep=<depPOS>
  - <attr>_disagrees, head=<headPOS>, dep=<depPOS>
  - head_<attr>=value, head=<headPOS>, dep=<depPOS>
  - dep_<attr>=value, head=<headPOS>, dep=<depPOS>
  - ... plus versions of each with dependency label

Hohensee 2012: It's only morpho-logical: Modeling agreement in cross-linguistic dependency parsing



Published at NAACL (Hohensee & Bender 2012)

Gola 2012: An analysis of translation divergence patterns using PanLex translation pairs

- If we could use lexical resources like PanLex to create (or beef up) transfer grammars, massively multi-lingual MT becomes more tractable

- To what extent would naïve transfer rules built on translation pairs run into translation divergence problems?

- Evaluated ita > fra, ita > eng (and at some remove, eng > tha), contrasting low-frequency with high-frequency verbs

- Primary result: translation divergence is more of an issue with high frequency verbs; using PanLex to fill in the “long tail” may be feasible

agree (Slayden and Rarrick)

- Parsing and generation with DELPH-IN grammars, using either LKB or PET config files (tested on ERG, Jacy, Thai and MCG)

- Chart dependencies and DELPH-IN regex morphology, but not yet full REPP support

- Spiffy UI (WPF client app); aside from UI, code base runs in mono (on linux) as well as .NET

- Allows multiple grammars to be loaded simultaneously

- Current (ambitious) work: bidirectional, declarative MRS rewriting for transfer-based MT

agree grammar engineering

Home | Parse | Transfer | Generate | Options

Load | Save | ERG | Thai | Jacy | MCG | old demo | transfer grammar | th-en | en-th

File | Grammars | Funcs

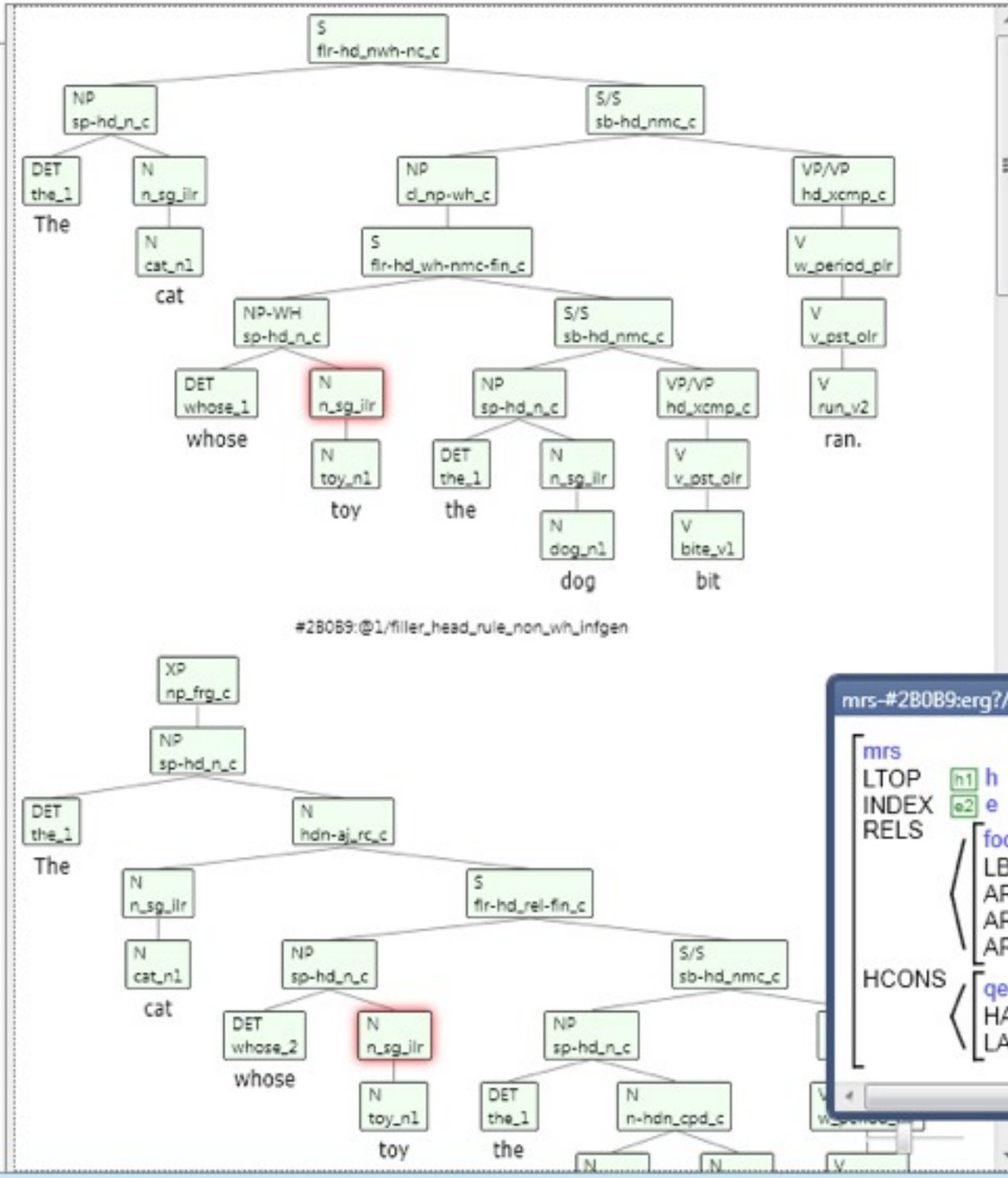
Grammar: erg | Parse derivations explorer | Grammar: jacy | Tfs: 運転 が 吠える

The cat whose toy the dog bit ran.

7 Reset

Rules | Lexical Entries

			license	type	#uses	#a
☑	0	0	the_1	LexicalEntry	7	
☑	1	1	cat_n1	LexicalEntry	7	
☑	1	1	n_sg_ilr	LexicalRule	7	
☑	0	1	sp-hd_n_c	GrammarRule	1	
☑	2	2	whose_1	LexicalEntry	2	
☑	3	3	toy_n1	LexicalEntry	7	
☑	3	3	n_sg_ilr	LexicalRule	7	
☑	2	3	sp-hd_n_c	GrammarRule	7	
☑	4	4	the_1	LexicalEntry	7	
☑	5	5	dog_n1	LexicalEntry	7	
☑	5	5	n_sg_ilr	LexicalRule	4	
☑	4	5	sp-hd_n_c	GrammarRule	4	
☑	6	6	bite_v1	LexicalEntry	4	
☑	6	6	v_pst_olr	MorphologicalRu	4	
☑	6	6	hd_xcmp_c	GrammarRule	3	
☑	4	6	sb-hd_nmc_c	GrammarRule	3	
☑	2	6	flr-hd_wh-nmc-fi	GrammarRule	2	
☑	2	6	cl_np-wh_c	GrammarRule	2	
☑	7	7	run_v2	LexicalEntry	4	
☑	7	7	v_pst_olr	MorphologicalRu	7	
☑	7	7	w_period_plr	MorphologicalRu	7	
☑	7	7	hd_xcmp_c	GrammarRule	4	
☑	2	7	sb-hd_nmc_c	GrammarRule	2	
☑	0	7	flr-hd_nwh-nc_c	GrammarRule	1	
☑	2	2	whose_2	LexicalEntry	5	
☑	6	6	bit_n2	LexicalEntry	2	
☑	6	6	n_sg_ilr	LexicalRule	3	
☑	5	6	n-hdn_cpd_c	GrammarRule	3	



mrs-#2B0B9:erg?/@1/filler_head_rule_non_wh_infgen

```
[ mrs
  LTOP  [ h1 ] h
  INDEX [ e2 ] e
  RELS  < [ focus_d_rel
            LBL  [ h1 ]
            ARG0 [ e3 ] e
            ARG1 [ e2 ]
            ARG2 [ x4 ] x ]
        , [ _the_q_rel
            LBL  [ h5 ] h
            ARG0 [ x4 ]
            RSTR [ h6 ] h
            BODY [ h7 ] h ]
        , [ _cat_n_1_rel
            LBL  [ h6 ]
            ARG0 [ x4 ]
  HCONS < [ qeq
            HARG [ h6 ]
            LARG [ h6 ] ]
        , [ qeq
            HARG [ h10 ]
            LARG [ h10 ] ]
        , [ qeq
            HARG [ h15 ]
            LARG [ h15 ] ]
```

System Log

```
agree-sys (debug), x64 gc=Server, 4 CPUs
erg-loader Begin loading 'erg.gee'.
erg-loader Tdl files hash: 3B6EB723
erg-loader Expanding all type definitions.
erg-loader Loading grammar rule entries.
erg-loader Loading lexical rule entries.
erg-loader Scanning lexical entries.
erg-loader Loading node label entries.
erg-loader Loading start symbol entries.
erg-loader loading definitions for all Entries
erg-loader Check rule compatibility.
erg-loader Loaded 'erg.gee' in 6.4765 s.
jacy-loader Begin loading 'jacy.gee'.
jacy-loader Tdl files hash: BD7723A8
jacy-loader warning: type 'extracted-adj-phras
jacy-loader warning: type 'basic-head-filler-ph
jacy-loader warning: type 'basic-head-filler-ph
jacy-loader warning: type 'gap' already has pa
jacy-loader warning: type 'conj-ref-ind' alread
jacy-loader Expanding all type definitions.
jacy-loader Loading grammar rule entries.
jacy-loader Loading lexical rule entries.
jacy-loader Scanning lexical entries.
jacy-loader Loading node label entries.
jacy-loader Loading start symbol entries.
jacy-loader loading definitions for all Entries
```

System Log | System Info

Ready