

# Toward a stable ERG 2018 release

Dan Flickinger

Stanford University and  
Center for Advanced Study, Oslo

DELPH-IN Diderot

Paris, 17 June 2018

# Improvements

- Annotations
- Token mapping
- Syntax
- Semantics
- Platforms and applications



# Annotations

- Supplied full-forest treebanks for Redwoods profiles, including the first five sections of the WSJ.
- Added profiles for WeSearch user-generated content (wlb03, wnb03), and for Sherlock Holmes *Speckled Band* story, as well as test suites for control, and for relative clauses (Huddleston & Pullum, Ch.12).
- Improved the well-formedness and consistency of the MRSs, aiming for more consistency with an updated version of the semantic algebra.



# Token mapping

- Upgraded to use GML 1.0, particularly relevant for the WeScience corpus.
- Improved support for ‘strong’ brackets (manually inserted) to signal phrase boundaries that should not be crossed.

*we [ saw a man ] with a telescope.*

- Maintained ‘weak’ brackets motivated by, for example, hyphens
- Changed RNAME value on rules from string to type, to allow ‘weak’ bracket rules to constrain which rules must apply, as for named entities:

*The [ New York ] [ Stock Exchange ]*



# Syntax

- Enabled extraction from within NPs

*Which people did you send pictures of to your friends?*

- Changed the attachment order of pre- and post-nominal modifiers, so now the pre-modifiers attach before post-modifiers

*The top contender in the race disappeared.*

- Added two new types of modifiers of verbal projections

indefinite NPs: *She walked out of the casino a rich woman.*

gapped clauses of saying: *They will, we suspect, leave early.*

- Added the ‘do-be’ construction

*The only thing we said she had to do was finish the assignment.*



# Semantics

- Moved information-structure constraints from RELS to ICONS in MRSs, including for focus movement ('topicalization') and passivization.
- Simplified the inventory of role labels, notably for conjunction relations, now just two-place, where ARG1 and ARG2 replace the old L-HNDL, R-HNDL, L-INDEX, R-INDEX roles.
- Improved SEM-I consistency.
- Better approximated semantic algebra constraints, moderately revised



# Platforms and applications

- Added support for ubertagging with PET and ACE, thanks to Rebecca Dridan (and Woodley Packard)
- Added robust parsing mode for ACE using csaw PCFG, thanks to Woodley Packard, building on Yi Zhang's Jigsaw foundations.
- Added support for 'agree' parser/generator, thanks to Glenn Slayden
- Expanded the inventory of mal-rules for grammar-checking.
- Added 'transfer' rules and support for generation from first order logic.
- Added support for robust 'bridging' rules (disabled by default).



# Release process

- Make the grammar better
- Compile grammar for PET and ACE
- Batch-check lexicon, and rebuild SEM-I
- Update treebanked profiles for Redwoods (800K words)
- Check semantic well-formedness of Redwoods
- Rebuild parse selection models (.mem files)
- Update quick-check paths for LKB, PET, ACE
- Generate from MRSs for linguistic test suites
- Update Linguistic Type DB for Redwoods
- Parse English Wikipedia in first-only mode (WikiWoods)
- Retrain PCFG model for robust parsing in 'hybrid' mode
- Retrain ubertagging model for efficiency via lexical pruning

