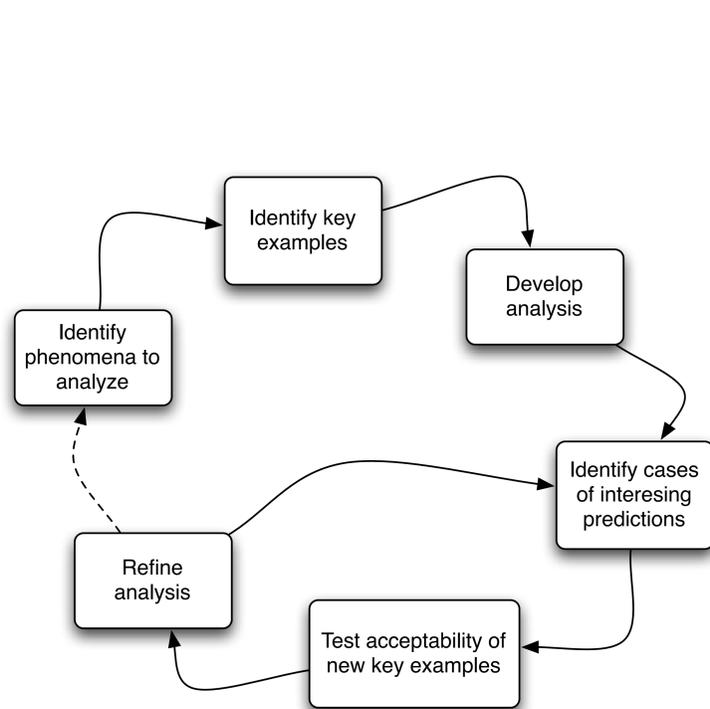
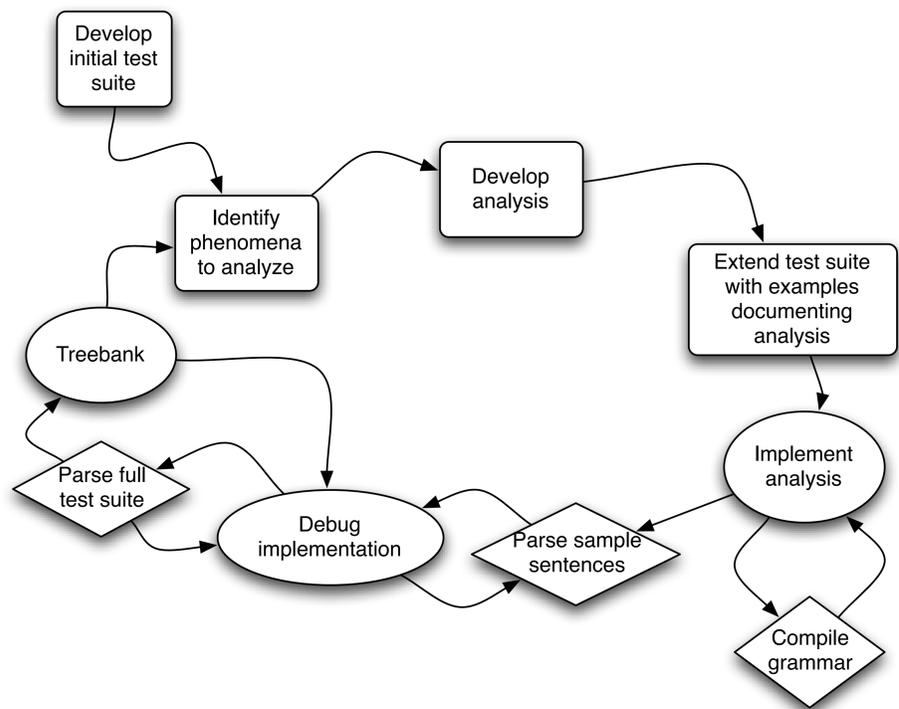


Multilingual Grammar Engineering with the LinGO Grammar Matrix

Emily M. Bender
University of Washington



Workflow in syntactic research without computer assistance
(Bender et al., 2011)



Workflow in syntactic research with computer assistance
(Bender et al., 2011)

1. Grammar Engineering

- Grammar engineering is the practice of encoding linguistic grammars in machine readable form
- Useful for both linguistic hypothesis testing and practical applications
- Requires explicit, precise definition of grammatical constraints
- Exposes interactions between analyses

2. The LinGO Grammar Matrix

<http://www.delph-in.net/matrix>

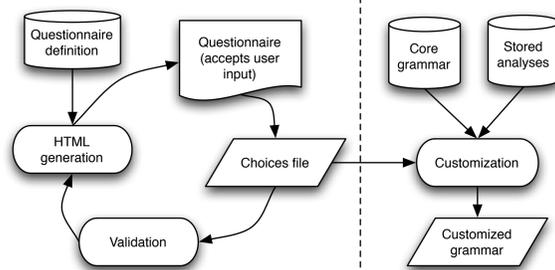
- A starter kit for the creation of implemented HPSG (Pollard and Sag, 1994) grammars, with Minimal Recursion Semantics (Copestake et al., 2005)
- Shared core grammar (Bender et al., 2002)
- Customization system, allowing users to specify both general typological information as well as defining lexical types, lexical rules and lexical entries (Bender et al., 2010)
- The customization system output is a working grammar fragment, mapping surface strings to semantic representations

3. Course goals

<http://courses.washington.edu/ling567>

- Hands-on grammar engineering experience
- A new perspective on natural language syntax and the interconnectedness of linguistic phenomena
- Deeper understanding of the syntax/semantics relationship
- Experience working with descriptive materials
- Practice building and debugging an extensible system
- Computational techniques for developing and testing formalizations

Elicitation of typological information



Grammar customization system (figure from Bender et al. 2010)

Grammar creation

4. Course assignments (weekly labs)

- W1 Practice with grammar of English
- W2-4 Develop test suite, work with customization system
- W5-8 Extend grammar to handle phenomena not covered by the customization system
- W9 Adapt grammars for MT system
- W10 "Multilingual MT extravaganza"

5. Lab write ups

- Description of phenomena
- Glossed examples
- Analyses applied
 - Through customization system
 - Through hand-editing
- Coverage & overgeneration (over test suite)

6. Phenomena covered

- Major constituent & NP-internal word order
- Person/number/gender & tense/aspect/mood
- Agreement, Case, Direct-inverse marking
- Non-verbal predicates
- Basic adnominal & adverbial modifiers
- Sentential negation, yes-no questions
- Coordination
- Basic lexical types, clausal complements
- Information status & information structure

7. Additional software used

- LKB (Copestake, 2002): grammar development environment
- [incr tsdb()] (Oepen, 2001): competence & performance profiling, regression testing
- emacs: source file editing
- svn: version control

8. Generalizable skills

- Test suite development
- Regression testing
- Version control
- Debugging

References

- Emily M. Bender, Scott Drellishak, Antske Fokkens, Laurie Poulson, and Safiyyah Saleem. 2010. Grammar customization. *Research on Language & Computation*, pages 1–50.
- Emily M. Bender, Dan Flickinger, and Stephan Oepen. 2002. The grammar matrix: An open-source starter-kit for the rapid development of cross-linguistically consistent broad-coverage precision grammars. In John Carroll, Nelleke Oostdijk, and Richard Sutcliffe, editors, *Proceedings of the Workshop on Grammar Engineering and Evaluation at COLING 2002*, pages 8–14. Taipei, Taiwan.
- Emily M. Bender, Dan Flickinger, and Stephan Oepen. 2011. Grammar engineering and linguistic hypothesis testing: Computational support for complexity in syntactic analysis. In Emily M. Bender and Jennifer E. Arnold, editors, *Language from a Cognitive Perspective: Grammar, Usage and Processing*, pages 5–29. CSLI Publications, Stanford, CA.
- Ann Copestake. 2002. *Implementing Typed Feature Structure Grammars*. CSLI Publications, Stanford, CA.
- Ann Copestake, Dan Flickinger, Carl Pollard, and Ivan A. Sag. 2005. Minimal recursion semantics: An introduction. *Research on Language & Computation*, 3(4):281–332.
- Stephan Oepen. 2001. [incr tsdb()] — Competence and performance laboratory. User manual. Technical report, Saarbrücken, Germany.
- Carl Pollard and Ivan A. Sag. 1994. *Head-Driven Phrase Structure Grammar*. Studies in Contemporary Linguistics. The University of Chicago Press and CSLI Publications, Chicago, IL and Stanford, CA.

9. Acknowledgments

This material is based upon work supported by the National Science Foundation under Grant No. 0644097. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.