

Test suites, [incr tsdb()]

Ling 567

January 14, 2014

Overview

- Questions from Lab 1
- Evaluation and computational linguistics
- Evaluation and precision grammars
- Test suites and precision grammars
- Our test suites
- Features of [incr tsdb()]
- Look at Lab 2 instructions

Ask more questions!

- This class is not designed so that you can complete the work on your own with the information provided.
- I'm *relying* on you to ask questions, and not spend lots of time stuck. The 10 minute rule is for real!

Questions from Lab 1: HPSG and analyses

- What's the deal with HOOK, C-CONT, XARG?
- Is C-CONT similar to ARG-ST?
- What's KEYREL?
- What is LIST in RELS?
- What is TARGET? ICONS-KEY?
- Why isn't the formalism more like what's in the 566 textbook?
- How to understand rules as feature structures, not productions.
- Why do we have head-spec and head-subj rules?

Questions from Lab 1: Grammar Matrix

- In general, this grammar is extremely elaborate compared to what we studied, so I am wondering how we are going to navigate through this across the quarter. In general, I am wondering how we'd be modifying this system without actually understanding it completely.
- What is the relationship between LinGO/DELPH-IN/the Grammar Matrix/the LKB?
- What leads you to edit/change matrix.tdl?

Questions from Lab 1: LKB software

- Can I get the LKB to tell me which constraint is being violated?
- How do I get the LKB to generate?
- How are the files related, and is there some global file that says so?
- Can the LKB do morphology?

Questions from Lab 1: Grammar engineering

- It seems with something this complex that the tiniest change can have massive cascading effects. Are there best practices for designing a language hierarchy to minimize this?
- How do I know which rule and which constraint to look for?
- Is this chain of identities assignment like debugging grammars?

Overview

- Questions from Lab 1
- Evaluation and computational linguistics
- Evaluation and precision grammars
- Test suites and precision grammars
- Our test suites
- Features of [incr tsdb()]
- Look at Lab 2 instructions

Evaluation and Computational Linguistics

- Why is evaluation so prominent in computational linguistics?
- Why is it not so prominent in other subfields of linguistics?
- What about CS?

Intrinsic v. extrinsic evaluation

- Intrinsic: How well does this system perform its own task, including generalizing to new data?
- Extrinsic: To what extent does this system contribute to the solution of some problem?
- Examples of intrinsic and extrinsic evaluation of parsers?

Test data

- Test suites
 - Hand constructed examples
 - Positive and negative examples
 - Controlled vocabulary
 - Controlled ambiguity
 - Careful grammatical coverage

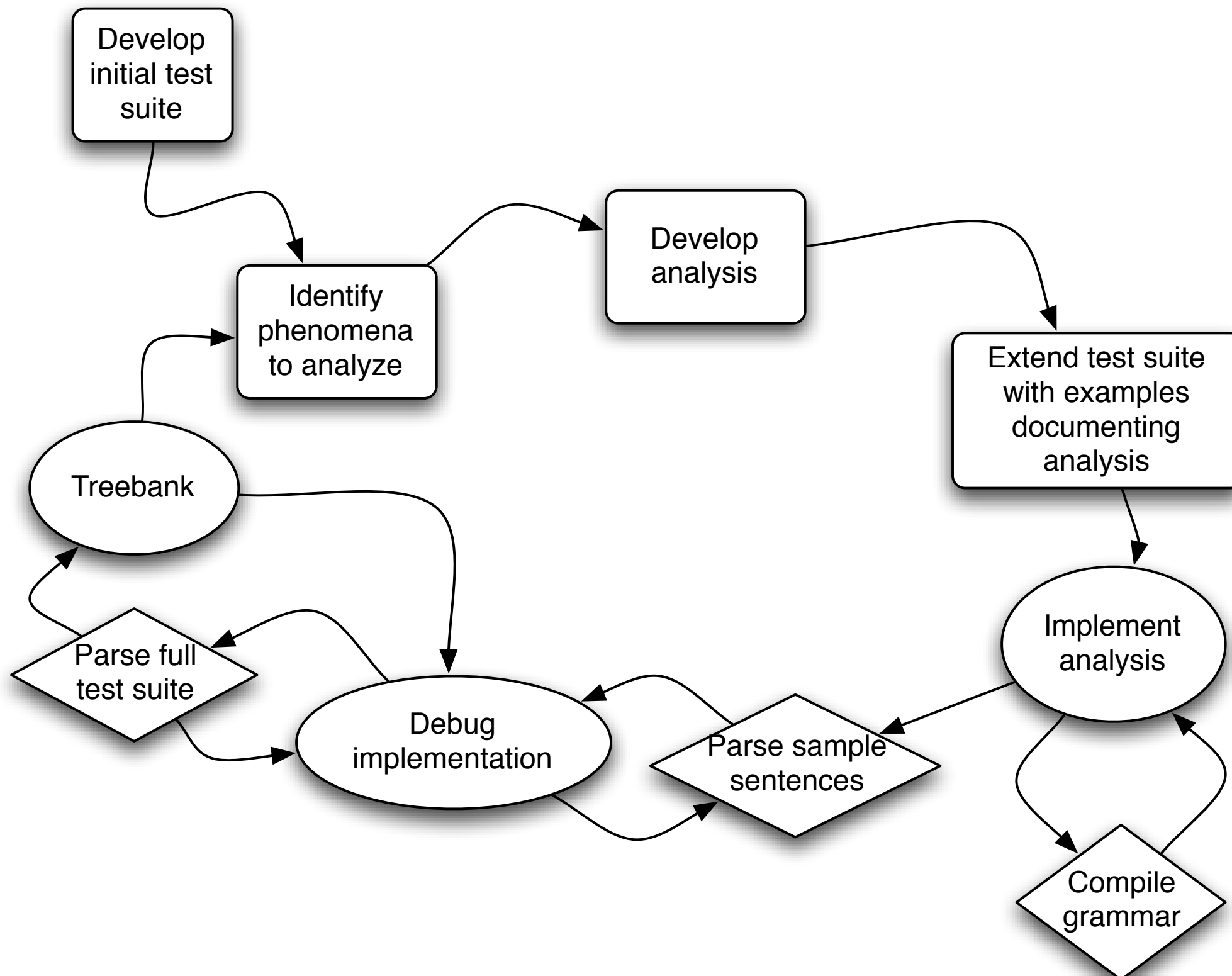
Test data

- Test corpora
 - Naturally occurring
 - More open vocabulary
 - Haphazard ungrammatical examples
 - Application-focused
- Which test data for which purposes?

Uses of test data

- How far do I have left to go?
 - Internal metric
 - Objective comparison of different systems
- Where have I been?
 - Regression testing
 - Documentation

Grammar engineering workflow



Evaluating precision grammars

- Coverage over some corpus
 - Which corpus?
 - Challenges of lexical acquisition
- Coverage of phenomena
 - How does one choose phenomena?
- Comparison across languages

Levels of adequacy

- grammaticality
- “right” structure
- “right” dependencies
- “right” full semantics
- only legit parses (how can you tell?)
- some set of parses including the preferred one
- preferred parse only/within first N

Our test suites

- Map out territory we hope to cover
- Include both positive and negative examples
- Serve as an exercise in understanding the description of the language
 - IGT format
 - Creating examples where necessary

[incr tsdb()] basics

- [incr tsdb()] stores test suite profiles as (plain text) relational databases: Each is a directory with a fixed set of files in it.
- Most files are empty.
- A profile that has not been processed has only two non-empty files: item (the items to be processed) and relations (always the same)
- Once the profile has been processed, the result of the processing is stored in some of the other files (in particular, parse and result)

[incr tsdb()] basics

- A test suite *skeleton* consists of just the item and relations files and can be used to create new test suite profiles
- [incr tsdb()] allows the user to compare two profiles to see how they differ
- It can also produce graphs plotting summary data from many profiles to visualize grammar evolution over time
- -> Demo

Overview

- Questions from Lab 1
- Evaluation and computational linguistics
- Evaluation and precision grammars
- Test suites and precision grammars
- Our test suites
- Features of [incr tsdb()]
- Look at Lab 2 instructions