

SIG 8: Grammar Engineering Workflow

Le Tuan Anh

Division of Linguistics and Multilingual Studies
School of Humanities and Social Sciences
Nanyang Technological University

6 August 2015

Background

Area of focus

- Descriptive Linguistics
- Computational Linguistics

Background

Area of focus

- Descriptive Linguistics
- Computational Linguistics

Descriptive Linguistics

- Reference Grammar
- Dictionary
- Reference Corpus

Background

Area of focus

- Descriptive Linguistics
- Computational Linguistics

Descriptive Linguistics

- Reference Grammar
- Dictionary
- Reference Corpus

Computational Linguistics

- Computer-readable Grammar
- Computer-readable Lexicon
- Computer-readable Corpus (treebank, etc.)

Practices

Descriptive Linguistics

- Toolbox
- Praat
- Text file
- LibreOffice Calc
- LibreOffice Writer
- Text file

Practices

Descriptive Linguistics

- Toolbox
- Praat
- Text file
- LibreOffice Calc
- LibreOffice Writer
- Text file

Computational Linguistics

- HPSG formalism for grammars
- WordNet
- TSDB++

HPSG Workflow

Aspect

- Linguistic analysis
- Implementation

HPSG Workflow

Aspect

- Linguistic analysis
- Implementation

Toolkits

- Grammar engineering: LKB/Agree, etc.
- TSDB++/gTest, etc. for treebanking, testing
- ACE/PET for production

Problems

- Logon is too big
- LKB uses outdated technology
- Documentation of language as well as of our practices

Thoughts

Our workflow

- Start with a small grammar from the Matrix
- Create testsuite
- LKB-TSDB++
- Repeat

A few thoughts

- Separate linguistics analysis from formalism
- Shareable linguistic resources
- Better tools

- Port LKB to a better platform
- Port Agree to Linux (use Mono)
- HPSG IDE

Thank you

Thank you.