Site Update Saarbrücken/DFKI

Hans Uszkoreit & Yi Zhang

LT-Lab, DFKI

July 29th, 2013

Deependance

- Running 2012-2014
- Extends the typed feature structure logic to incorporate feature value disjunctions
- `TDL` variant to allow explicit specification of named disjunctions and preference distribution

Without disjunctions		With named disjunctions

`koffer_n1 := noun-lex &`
`[ MORPH [ AGR [ NUM sg,`
`PERS 3rd ],`
`GEND masc ] ].`
	`koffer_n2 := noun-lex &`
	`[ MORPH [ AGR [ NUM pl,`
	`PERS 3rd ],`
	`GEND gend ] ].`
		`koffer_n := noun-lex &`
		`[ MORPH [ AGR [ NUM %1(sg|pl){0.6,0.4},`
		`PERS 3rd ],`
		`GEND %1(masc|gend) ] ].`

Deependance (cont.)

- What we have been working on . . .
  - Implementation of the new formalism in C++ (**depfs**)
  - Tomabechi directed graph unifier + Kasper unification of disjunctive feature descriptions
  - Getting rid of the named disjunctions during grammar compilation
  - (Somewhat) backward compatible with DELPH-IN grammars
- What we will work on . . .
  - Data-driven (transition-based?) dependency parsing interoperating with TFS
  - Porting of a small/core grammar from the existing DELPH-IN grammars
  - Learning of lexica and stochastic models from DeepBank(s)

Deependance (cont.)

From the application perspective

- Relation extraction
  - “weighted” selective unpacking for task-oriented parse disambiguation
- Computer-assisted language learning
  - Dependency-driven sentence linearization model for generation (Rui & Walter’s talks yesterday)

DeepBank

- A beta release (v0.9) in November 2012 at TLT
- In cooperation with Dan and Stephan, soon to be ready for the release v1.0
- Besides the raw tsdb profiles, also available in various “popular” formats
  - PTB-style .mrg formats for labeled constituent trees
  - CoNLL-2009 formats for both syntactic (derivation-derived) and semantic (MRS-derived) bilexical dependencis
- Coverage gap closed by robust paring with approximating PCFGs
- Soon to be available through META-SHARE

Statistical English-Spanish Tree-to-Tree (Deep) MT

- Ongoing M.Sc. project by Philip John Gorinski at CoLi UdS
- Aims to improve translation quality by training on parallel treebanks parsed with deep grammars (ERG & SRG on Europarl)
- Automatically align parallel treebanks by
  - word alignment information
  - structure similarity in HPSG analyses
- Automatically extract synchronous-LTAG for translation
- Relatively low coverage ($\sim$12%) due to data spareseness
- ...

Chinese Grammar

- Collaboration with NTU (Francis, Zhenzhen)
- Collaborating Xiangli Wang (Japan Patent Information Organization) on a phenomena test suite for Mandarin

Crowdsourcing Grammar Documentation

- An experiment to create documentation for an existing large-scale grammar (GG)
- As a course project (GE SS2012), a group of six CoLi students worked towards a collection of loosely organized documentation for the resource, taking different approaches
  - from publications
  - from treebanks/test-suites
  - from rules and type hierarchy
  - from lexicon
- `http://www.coli.uni-saarland.de/~yzhang/wiki/doku.php?id=gg`