

Oslo Status Updates

(In Fifteen Minutes)

Stephan Oepen

Universitetet i Oslo

oe@ifi.uio.no

(DELPH-IN Summit — July 20, 2009)

Broad Families of Language Processing Tasks

Speech Recognition and Synthesis

Summarization & Text Simplification

Information Extraction — Text Understanding

(High Quality) Machine Translation

Grammar & Controlled Language Checking

Natural Language Dialogue Systems



Broad Families of Language Processing Tasks

Speech Recognition and Synthesis

Summarization & Text Simplification

Information Extraction — Text Understanding

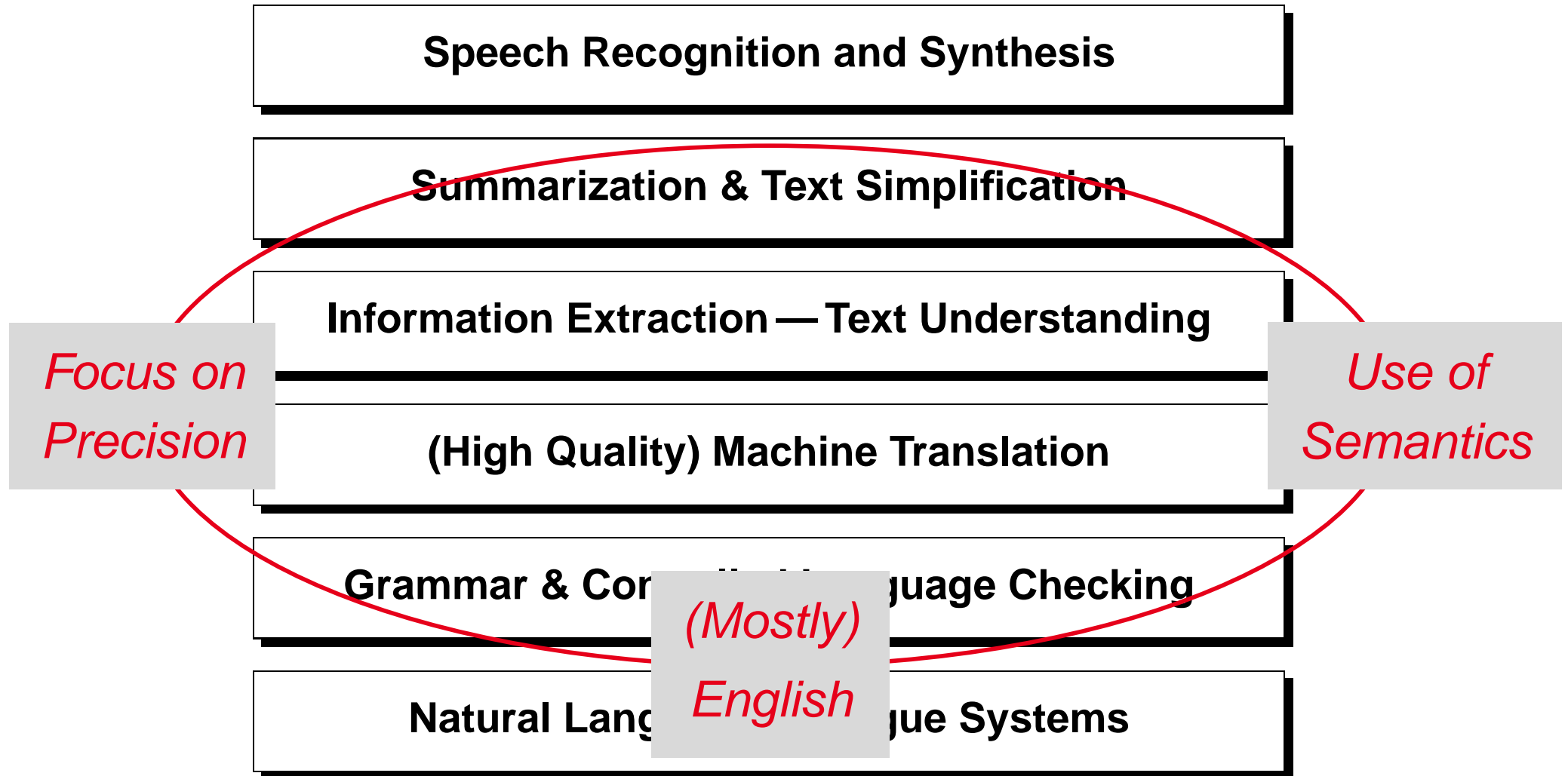
(High Quality) Machine Translation

Grammar & Controlled Language Checking

Natural Language Dialogue Systems



Broad Families of Language Processing Tasks



The UiO (Logic and) Language Technology Group

Language Technology

Liv Ellingsen	Doctoral Fellow	Soft Grammatical Constraints
Gordana Ilić Holen	Doctoral Fellow	Coreference Resolution
Elisabeth Lien	Doctoral Fellow	(Maybe) Ontology Learning
Jan Tore Lønning	Professor	Computational Semantics
Stephan Oepen	Professor	Unification-Based Processing
Erik Velldal	Post-Doctoral Fellow	Parse Selection
Gisle Ytrestøl	Doctoral Fellow	Incremental Parsing
NN	Doctoral Fellow	High-Quality Research
NN	Associate Professor	Data-Driven NLP

Computational Logic

Asbjørn Brændeland	Assistant Professor	Functional Programming
Herman Ruge Jervell	Professor	Proof Theory



The UiO (Logic and) Language Technology Group

Language Technology

Liv Ellingsen	Doctoral Fellow	Soft Grammatical Constraints
Gordana Ilić Holen	Doctoral Fellow	Core <i>Spring 2009</i> tion
Elisabeth Lien	Doctoral Fellow	(Maybe <i>Fall 2009</i> earning
Jan Tore Lønning	Professor	Computational Semantics
Stephan Oepen	Professor	Unification-Based Processing
Erik Velldal	Post-Doctoral Fellow	P <i>Fall 2009</i> on
Gisle Ytrestøl	Doctoral Fellow	Incremental Parsing
NN	Doctoral Fellow	High <i>Winter 2009</i> arch
NN	Associate Professor	De <i>Fall 2010</i> LP

Computational Logic

Asbjørn Brændeland	Assistant Professor	Functional Programming
Herman Ruge Jervell	Professor	Proof Theory



The UiO (Logic and) Language Technology Group

All Working with DELPH-IN Resources

Liv Ellings		al Constraints
Gordana Ilić	Doctoral Fellow	Coreference Resolution
Elisabeth Lien	Doctoral Fellow	(Maybe) Ontology Learning
Jan Tore Lønning	Professor	Computational Semantics
Stephan Oepen	Professor	Unification-Based Processing
Erik Velldal	Post-Doctoral Fellow	Parse Selection
Gisle Ytrestøl	Doctoral Fellow	Incremental Parsing
NN	Doctoral Fellow	High-Quality Research
NN	Associate Professor	Data-Driven NLP

Computational Logic

Asbjørn Brændeland	Assistant Professor	Functional Programming
Herman Ruge Jervell	Professor	Proof Theory



Mid- to Long-Term Personnel Developments



Recent and Ongoing Activities (1 of 2)

English Resource Grammar (Jointly with LinGO)

- Re-worked tokenization and preprocessing: REPP & chart mapping;
- fully support PoS-based unknown word handling (parsing in PET);
- reliable characterization, PREDs, and CARGs (also in [incr tsdb()]);
- better support for 'unknown' predicates in generation (using the LKB).

The LOGON Tree (A DELPH-IN Distribution)

- Migrated to SVN (also for ERG, LKB, and [incr tsdb()] repositories);
- added additional grammars and resources (Cheetah, HAG, KRG);
- new release, codenamed *Barcelona*, available by end of the month.



Recent and Ongoing Activities (2 of 2)

WeScience: Parsing Scholarly Literature

- Domain-specific selection from Wikipedia: one hundred NLP articles;
- new 10,000-item ERG treebank as part of *Barcelona* release (0907);
- next: text extraction and parsing of research literature (NORA & ARC).

Large-Scale HPC Utilization

- Improved collaboration with UiO grid computing group (~5000 cpus);
- joint project: adapt and scale up DELPH-IN tools for HPC usage;
- packaging, scripting and distribution of standard tasks (with SLURM);
- low-level profiling of PET and TADM; activate MPI support in TADM.



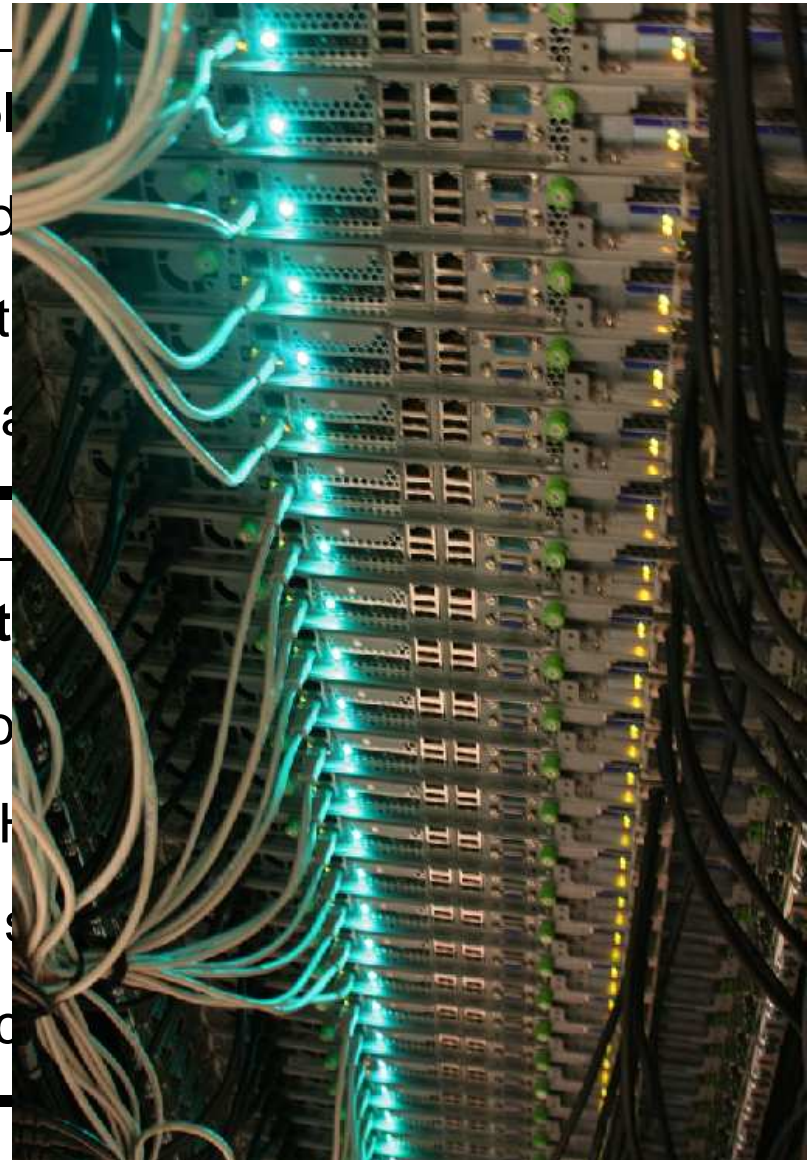
Recent and Ongoing Activities (2 of 2)

WeScience: Parsing School

- Domain-specific selection from Wikipedia
- new 10,000-item ERG treebank as part
- next: text extraction and parsing of resear

Large-Scale HPC Ut

- Improved collaboration with UiO grid co
- joint project: adapt and scale up DELPH
- packaging, scripting and distribution of s
- low-level profiling of PET and TADM; ac



Forthcoming and Planned Activities

Large-Scale Parsing (with Erik Velldal and others)

- Parse selection: new features, feature selection, self-training, et al.
- CFG approximations: pruning, disambiguation, robustness, et al.

ESSENCE (with CoLi Saarbrücken: *SOLID*)

- Semantic parsing of scholarly texts: towards 'deep' IE and search;
- ontologies: relate NL semantics to knowledge representation (OWL).

COLT (FP7 Consortium)

- Incremental parsing; connect computational and psycholinguistics.

