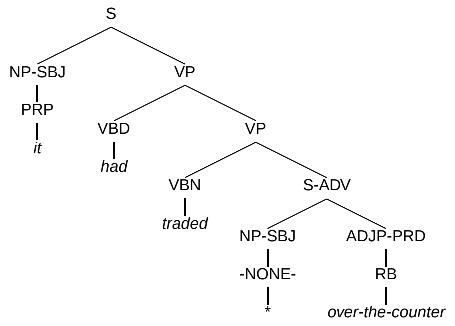


$$\begin{bmatrix} \texttt{TOP} & h_1 \\ \texttt{INDEX} & e_2 \\ \texttt{RELS} & \left\langle \begin{bmatrix} \textit{prpstn\_m\_rel} & \\ \texttt{LBL} & h_1 \\ \texttt{MARG} & h_3 \end{bmatrix} \begin{bmatrix} \textit{def\_q\_rel} & \\ \texttt{LBL} & h_4 \\ \texttt{ARG0} & x_5 \\ \texttt{RSTR} & h_6 \\ \texttt{BODY} & h_7 \end{bmatrix} \begin{bmatrix} \textit{\_dog\_n\_rel} & \\ \texttt{LBL} & h_8 \\ \texttt{ARG0} & x_5 \end{bmatrix} \begin{bmatrix} \textit{\_bark\_v\_rel} & \\ \texttt{LBL} & h_9 \\ \texttt{ARG0} & e_2 \\ \texttt{ARG1} & x_5 \end{bmatrix} \right\rangle \\ \texttt{HCONS} & \langle h_3 =_q h_9, h_6 =_q h_8 \rangle \end{bmatrix}$$

Oslo Status Updates

(In Fifteen Minutes)

Stephan Oepen

Universitetet i Oslo

`oe@ifi.uio.no`

(DELPH-IN Summit — July 2, 2010)

The IFI Language Technology Group

Table of Contents

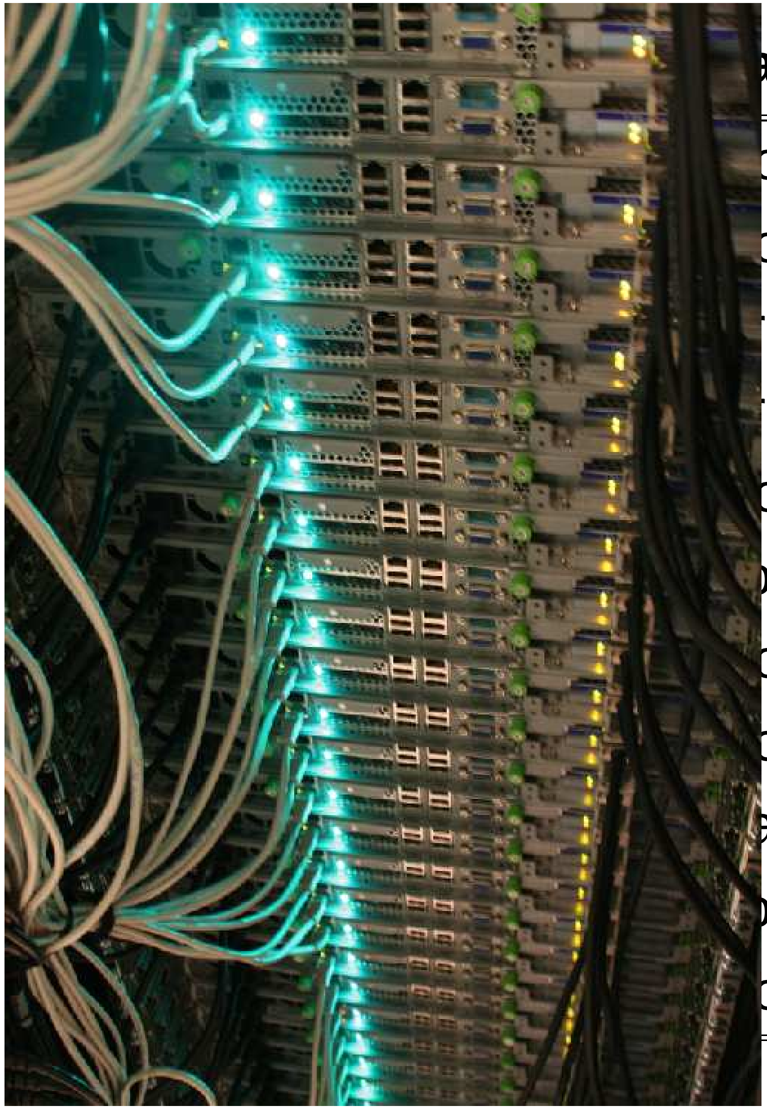
Gordana Ilić Holen	Doctoral Fellow	Coreference Resolution
Elisabeth Lien	Doctoral Fellow	Textual Inference
Jan Tore Lønning	Professor	Computational Semantics
Stephan Oepen	Professor	Grammar-Based Processing
Woodley Packard	Doctoral Fellow	Joint Disambiguation
Erik Velldal	Post-Doctoral Fellow	Classification
Gisle Ytrestøl	Doctoral Fellow	Incremental Parsing
Aleksander Øhrn	Adjunct Professor	Information Retrieval
Lilja Øvrelid	Associate Professor	Data-Driven NLP
NN	Post-Doctoral Fellow	Parser Adaptation
NN	Doctoral Fellow	High-Quality Research

The IFI Language Technology Group

Table of Contents

Gordana Ilić Holen	Doctoral Fellow	Cor *Spring 2009* ution
Elisabeth Lien	Doctoral Fellow	Te *Fall 2009* ce
Jan Tore Lønning	Professor	Computational Semantics
Stephan Oepen	Professor	Grammar-Based Processing
Woodley Packard	Doctoral Fellow	Joi *Spring 2010* on
Erik Velldal	Post-Doctoral Fellow	(*Fall 2009* ı
Gisle Ytrestøl	Doctoral Fellow	Incremental Parsing
Aleksander Øhrn	Adjunct Professor	Inf *Spring 2010* val
Lilja Øvrelid	Associate Professor	Da *Fall 2010* LP
NN	Post-Doctoral Fellow	Pa *Fall 2010* ion
NN	Doctoral Fellow	High *Fall 2010* earch

The IFI Language Technology Group



able of Contents

oral Fellow	Coreference Resolution
oral Fellow	Textual Inference
ofessor	Computational Semantics
ofessor	Grammar-Based Processing
oral Fellow	Joint Disambiguation
octoral Fellow	Classification
oral Fellow	Incremental Parsing
ct Professor	Information Retrieval
ate Professor	Data-Driven NLP
octoral Fellow	Parser Adaptation
oral Fellow	High-Quality Research

WikiWoods: Syntacto-Semantic Analysis of Wikipedia

General Idea

- Enabling technology: Wikipedia as a corpus and a knowledge source;
- e.g. research in linguistics, lexical acquisition, ontology learning, etc.

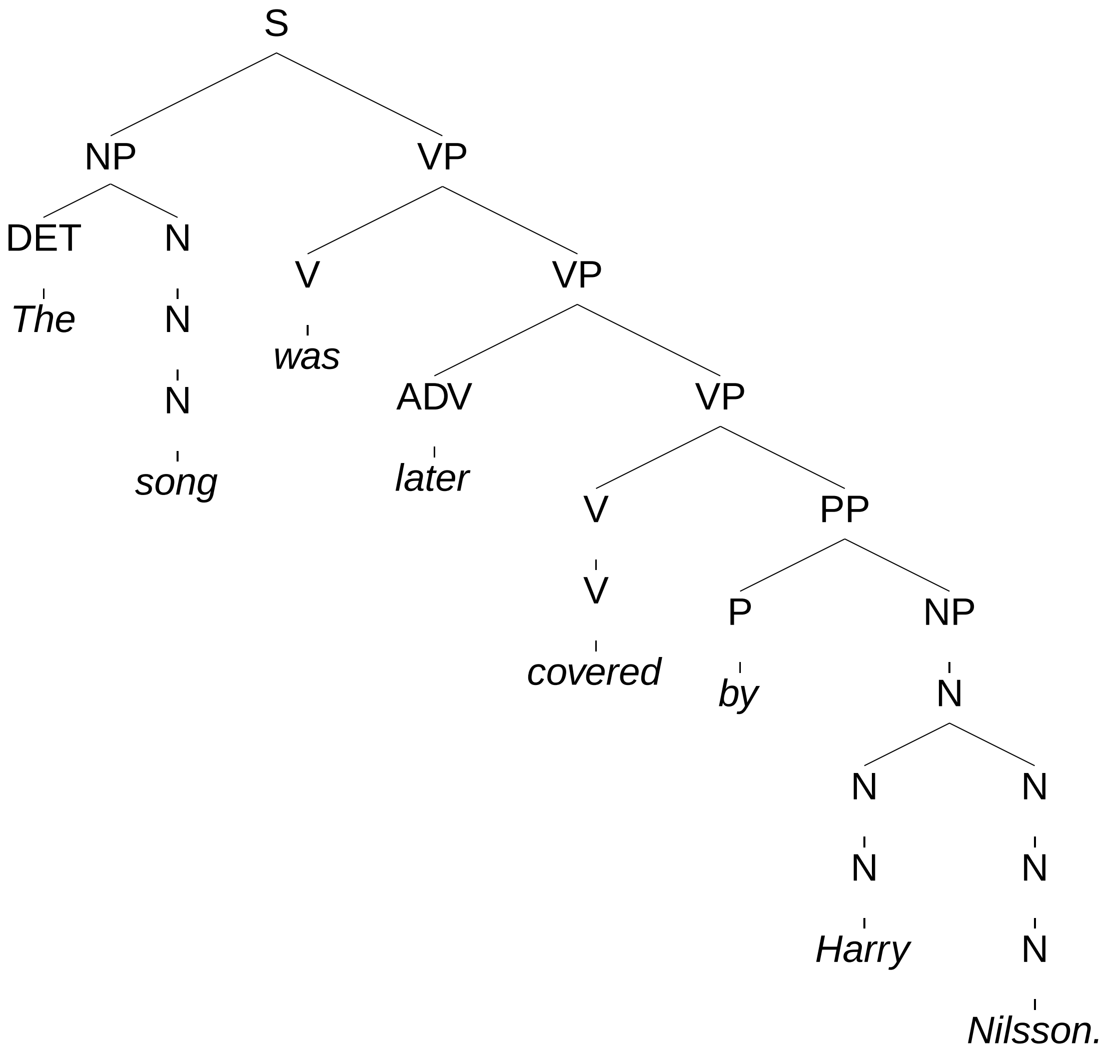
Approach & Technology

- Semi-automated ‘deep’ linguistic annotation, from pre-existing parser;
- gold-standard annotation of domain-specific subset: ~250,000 words.

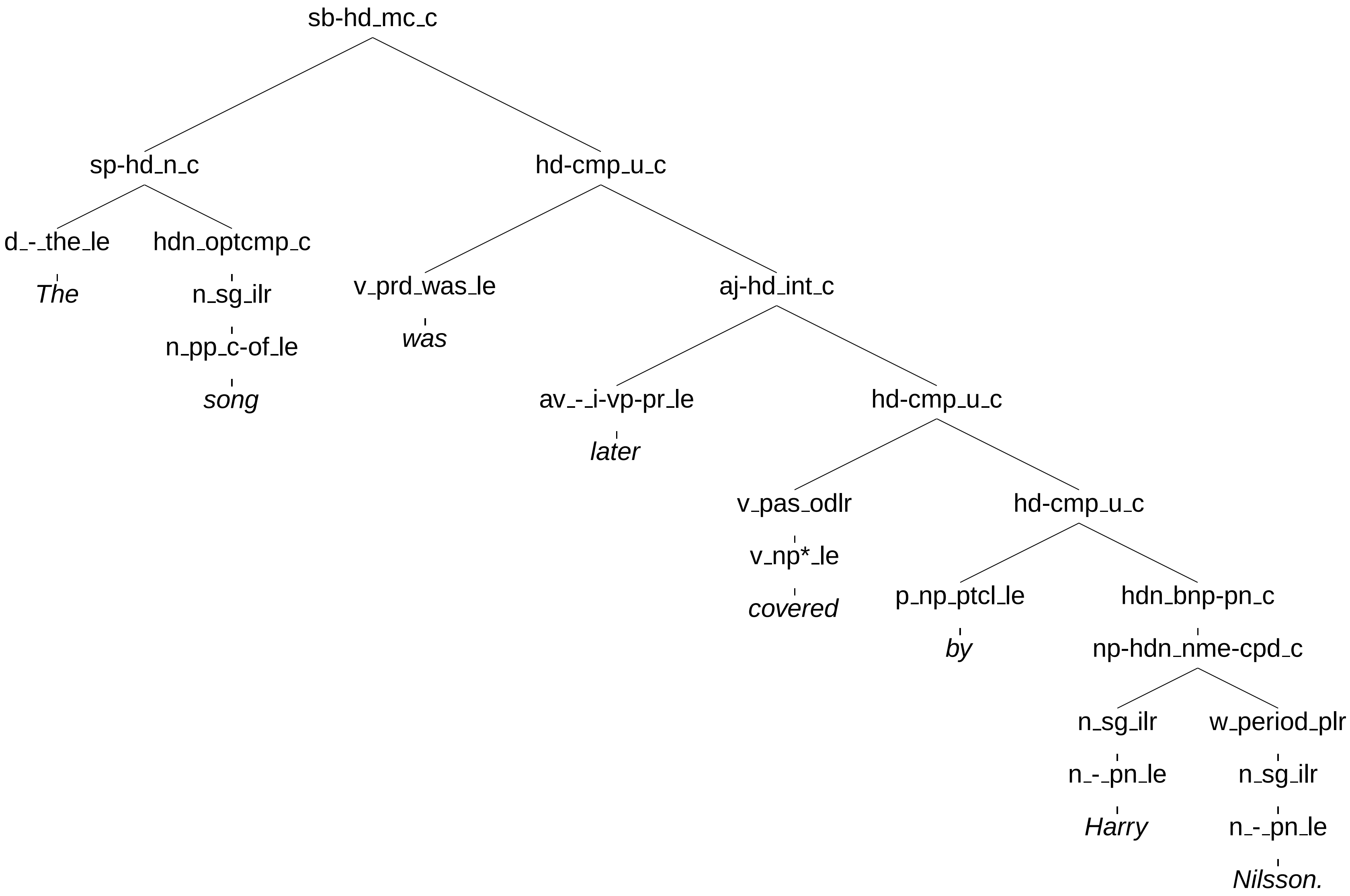
More Information (Download Site)

`http://www.delph-in.net/wikiwoods`

Syntactic Annotation: ‘Classic’ Constituent Tree



Syntactic Annotation: HPSG Derivation



Semantic Annotation: Predicate – Argument Structure

The song was later covered by Harry Nilsson.

$$
\begin{array}{l}
\langle\, h_1, \\
\left|\begin{array}{l}
h_3\text{:\_the\_q}(x_5, h_6, h_4), h_7\text{:\_song\_n\_of}(x_5\{\text{PERS } 3, \text{NUM } \textit{sg}\}, \_\_), \\
h_9\text{:\_cover\_v\_1}(e_2\{\text{SF } \textit{prop}, \text{TENSE } \textit{past}, \text{MOOD } \textit{ind}\}, x_{11}, x_5), \\
h_9\text{:\_later\_a\_1}(\_\_, e_2), \\
h_{16}\text{:compound\_name}(\_\_, x_{11}, x_{17}), \\
h_{19}\text{:proper\_q}(x_{17}, h_{20}, h_{21}), h_{22}\text{:named}(x_{17}\{\text{PERS } 3, \text{NUM } \textit{sg}\}, \textit{Harry}), \\
h_{13}\text{:proper\_q}(x_{11}, h_{14}, h_{15}), h_{16}\text{:named}(x_{11}\{\text{PERS } 3, \text{NUM } \textit{sg}\}, \textit{Nilsson})
\end{array}\right| \\
\{\, h_{20} =_q h_{22}, h_{14} =_q h_{16}, h_6 =_q h_7 \,\}\,\rangle
\end{array}
$$

Semantic Annotation: Predicate – Argument Structure

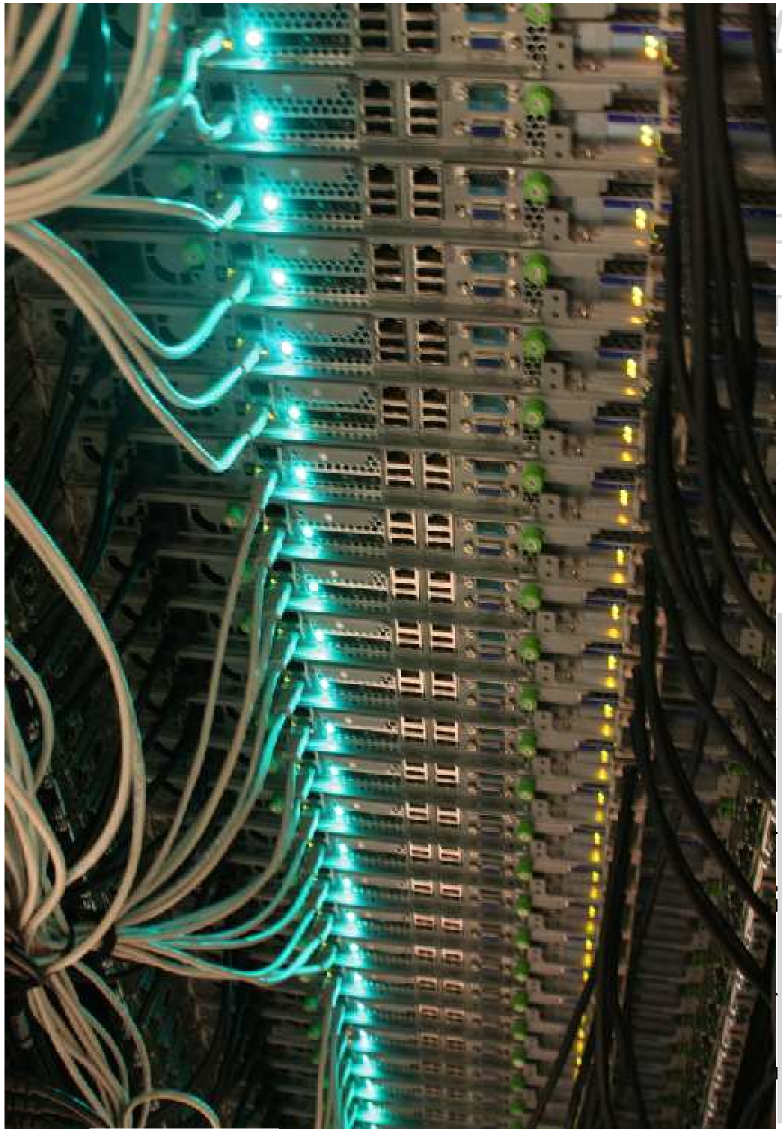
The song was later covered by Harry Nilsson.

$$
\begin{array}{l}
\langle\, h_1, \\
\quad \left|\begin{array}{l}
h_3\text{:\_the\_q}(x_5, h_6, h_4), h_7\text{:\_song\_n\_of}(x_5\{\text{PERS } 3, \text{NUM } sg\}, \_\_), \\
h_9\text{:\_cover\_v\_1}(e_2\{\text{SF } prop, \text{TENSE } past, \text{MOOD } ind\}, x_{11}, x_5), \\
h_9\text{:\_later\_a\_1}(\_\_, e_2), \\
h_{16}\text{:compound\_name}(\_\_, x_{11}, x_{17}), \\
h_{19}\text{:proper\_q}(x_{17}, h_{20}, h_{21}), h_{22}\text{:named}(x_{17}\{\text{PERS } 3, \text{NUM } sg\}, Harry), \\
h_{13}\text{:proper\_q}(x_{11}, h_{14}, h_{15}), h_{16}\text{:named}(x_{11}\{\text{PERS } 3, \text{NUM } sg\}, Nilsson)
\end{array}\right| \\
\quad \{\, h_{20} =_q h_{22}, h_{14} =_q h_{16}, h_6 =_q h_7 \,\}\, \rangle
\end{array}
$$

→ 1.3 million content articles, 55 million utterances, ~900 million tokens;

→ ~85 % parsing coverage, ~83 % of analyses totally or nearly correct.

Semantic Annotation: Predicate – Argument Structure



ter covered by Harry Nilsson.

ong_n_of(x_5{PERS 3, NUM *sg*}, _),

TENSE *past*, MOOD *ind*}, x_{11}, x_5),

$_{11}$, x_{17}),

h_{22}:named(x_{17}{PERS 3, NUM *sg*}, *Harry*),

~120,000 cpu hours (six days);

~130 gigabytes compressed data;

→ *subject extraction present in one of 15 utterances;*

→ *~90 % in relative clauses.*

WeSearch: Parsing User-Generated Content

Scalable & Adaptable Parsing (with the ERG)

- Closely investigate trade-offs: robustness – precision – efficiency;
- parser adaptation across genres and domains: degrees of formality;
- interplay of PoS tagging, supertagging, chart pruning, and others.

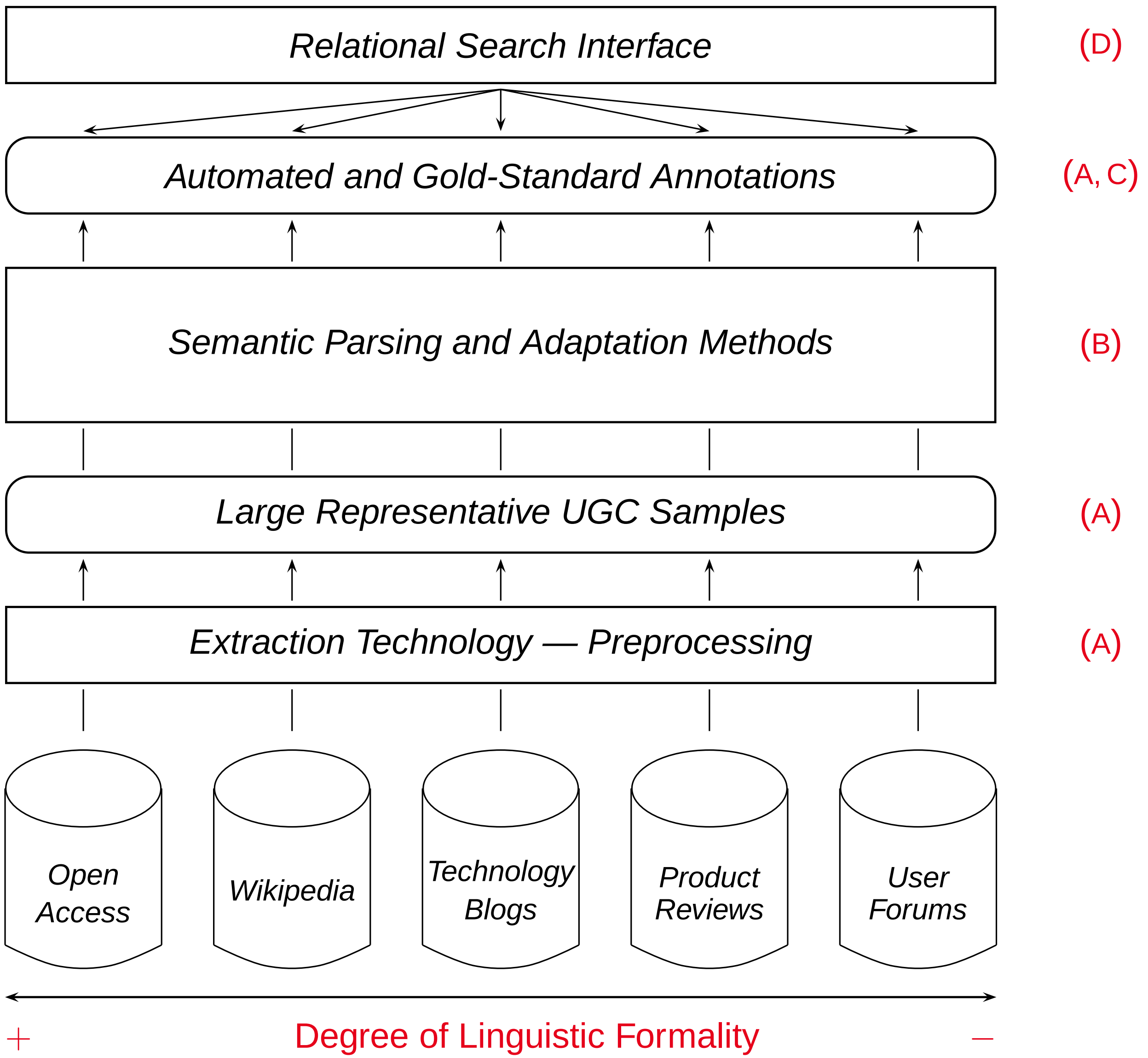
Semantic Interface Corroboration

- Document a growing body of *stable* aspects of semantic analyses;
- extend MRS test suite; joint workshops (with Stanford & Cambridge).

Relational Search Interface

- ‘Semantic’ search interface to select content (joint work with DFKI).

WeSearch: The Big Picture



Relational Search Interface
(D)
Automated and Gold-Standard Annotations
(A, C)
Semantic Parsing and Adaptation Methods
(B)
Large Representative UGC Samples
(A)
Extraction Technology — Preprocessing
(A)
Open Access
Wikipedia
Technology Blogs
Product Reviews
User Forums
+
Degree of Linguistic Formality
−

Discussion Topic: DELPH-IN HPC Portal

Available Infrastructure

- Substantial national HPC resources accessible with relative ease;
- large subset of DELPH-IN tools 'packaged' for batch use (LOGON);
- HPC group at UiO experienced in providing bio-informatics 'portal'.

'Deep' Parsing Portal at UiO

- Reduce technology barriers: on-line demonstrators *and* processing;
- unified, Web-based point of entry; balance ease of use and flexibility;
- common & user-provided data sets, pre-defined processes & formats;

? which services (if any) should DELPH-IN aim to package this way?

Finally, Various Short-Term Activities

2010 *Paris* Release of LOGON Tree

- Primary goal: reference snapshot to accompany WikiWoods release;
- synchronize code, fix a few known bugs (Antonio, Berthold, Montse);
- co-developers: update 'your' components, e.g. the various grammars;
- clarify licensing conditions across the LOGON tree (with Francis);
- schedule: code freeze on August 16; public release by August 31.

Miscellaneous

- Velldal, Øvrelid, & Oepen (2010): successful in CoNLL Shared Task;
- syntactic 'scope' resolution for hedges; though not using ERG parses.