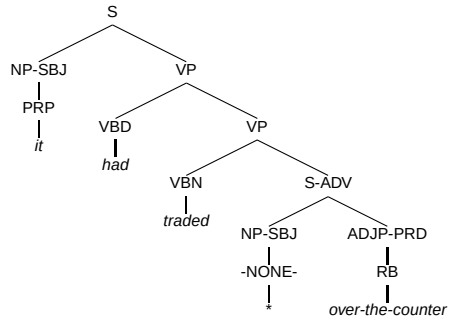


$$\begin{bmatrix} \texttt{TOP} & h_1 \\ \texttt{INDEX} & e_2 \\ \texttt{RELS} & \left\langle \begin{bmatrix} \textit{prpstn\_m\_rel} \\ \texttt{LBL} & h_1 \\ \texttt{MARG} & h_3 \end{bmatrix} \begin{bmatrix} \textit{def\_q\_rel} \\ \texttt{LBL} & h_4 \\ \texttt{ARG0} & x_5 \\ \texttt{RSTR} & h_6 \\ \texttt{BODY}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

Language Technology

(In Less Than Twenty Minutes)

Stephan Oepen

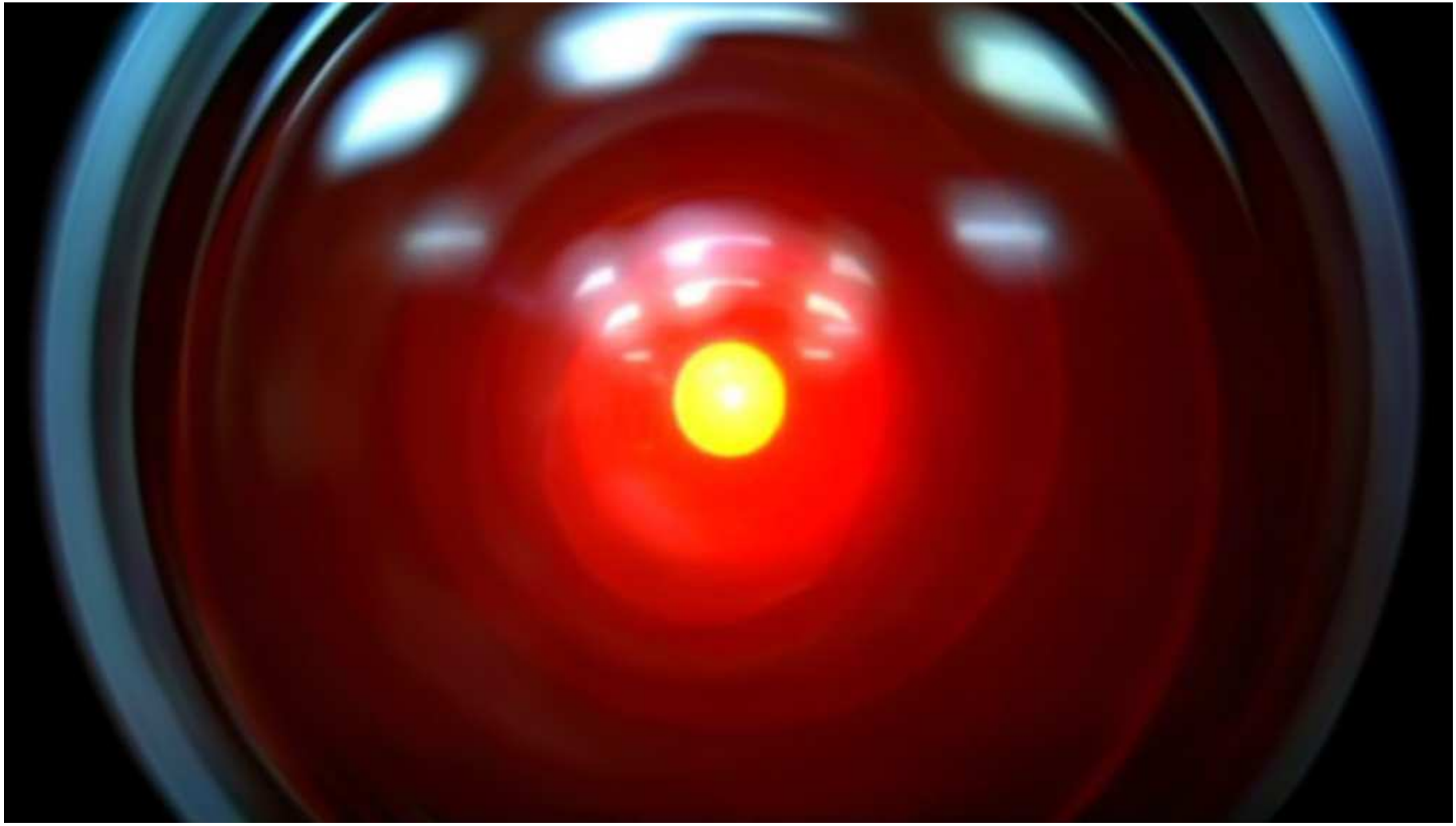
Universitetet i Oslo & CSLI Stanford

oe@ifi.uio.no

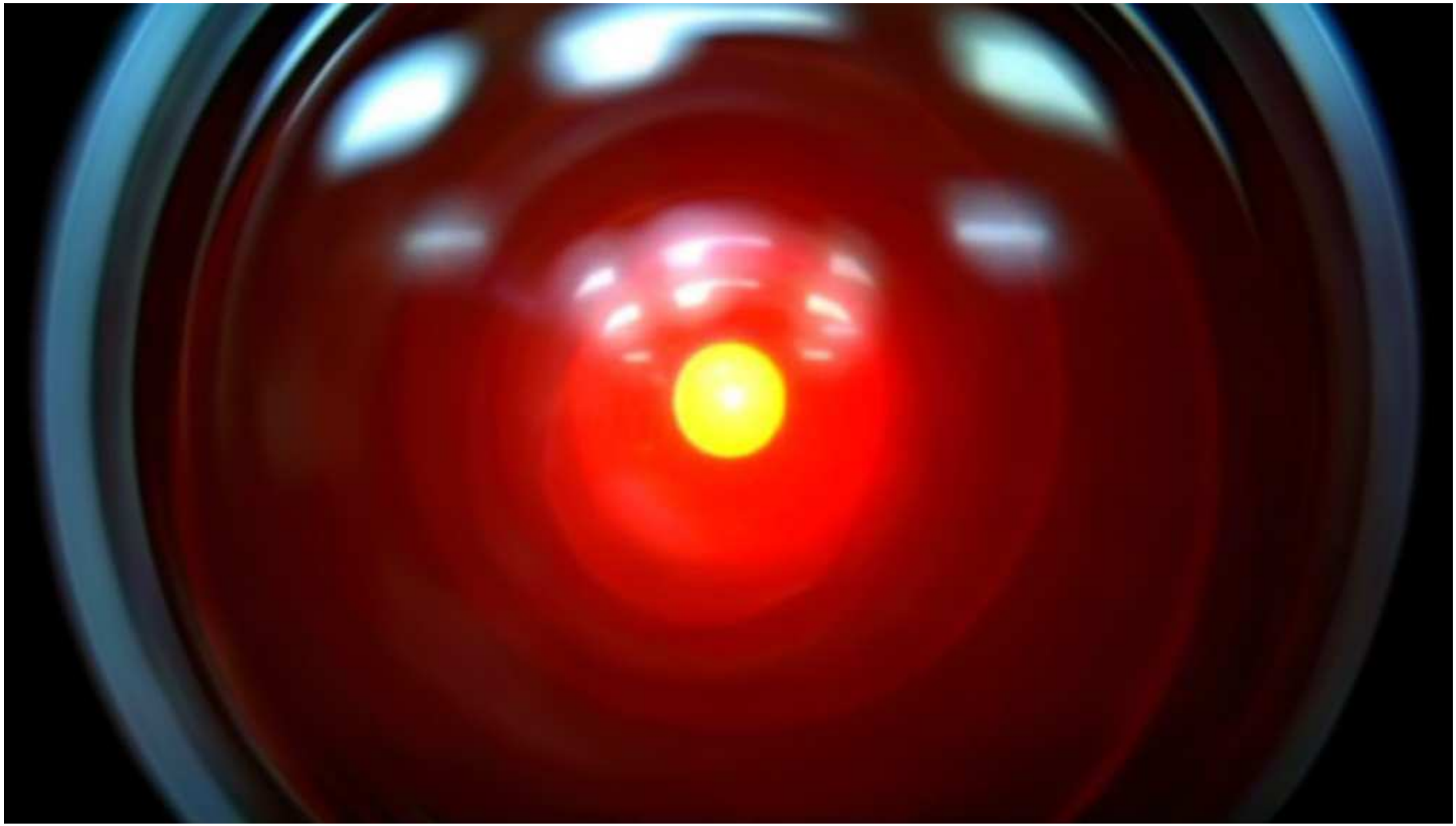
(Holmen Fjord Hotel, April 15, 2008)

So, What Actually is Language Technology?

So, What Actually is Language Technology?



So, What Actually is Language Technology?



(2001: A Space Odyssey; HAL 9000; 1968)

No, Really, What is Language Technology?

No, Really, What is Language Technology?

... teaching computers our language. (Returning Alien Resident, 2000)

No, Really, What is Language Technology?

... teaching computers our language. (Returning Alien Resident, 2000)

We Understand™. Unlike other solutions based on keyword or phrase recognition, YY Software's product actually understands customer e-mails and Web interaction. (Marketing Blurb, 2000)

No, Really, What is Language Technology?

... teaching computers our language. (Returning Alien Resident, 2000)

We Understand™. Unlike other solutions based on keyword or phrase recognition, YY Software's product actually understands customer e-mails and Web interaction. (Marketing Blurb, 2000)

... the scientific study of human language — specifically of the system of grammar and the ways in which it is used in communication — using mathematical models and formal procedures that can be realized and validated using computers; a cross-over of many disciplines. (Stanford Researcher, 2002)

No, Really, What is Language Technology?

... teaching computers our language. (Returning Alien Resident, 2000)

We Understand™. Unlike other solutions based on keyword or phrase recognition, YY Software's product actually understands customer e-mails and Web interaction. (Marketing Blurb, 2000)

... the scientific study of human language — specifically of the system of grammar and the ways in which it is used in communication — using mathematical models and formal procedures that can be realized and validated using computers; a cross-over of many disciplines. (Stanford Researcher, 2002)

→ (young) interdisciplinary science: language, cognition, computation;

→ (again) culturally and commercially relevant for 'knowledge society'.

Families of Language Processing Tasks

Speech Recognition and Synthesis

Summarization & Text Simplification

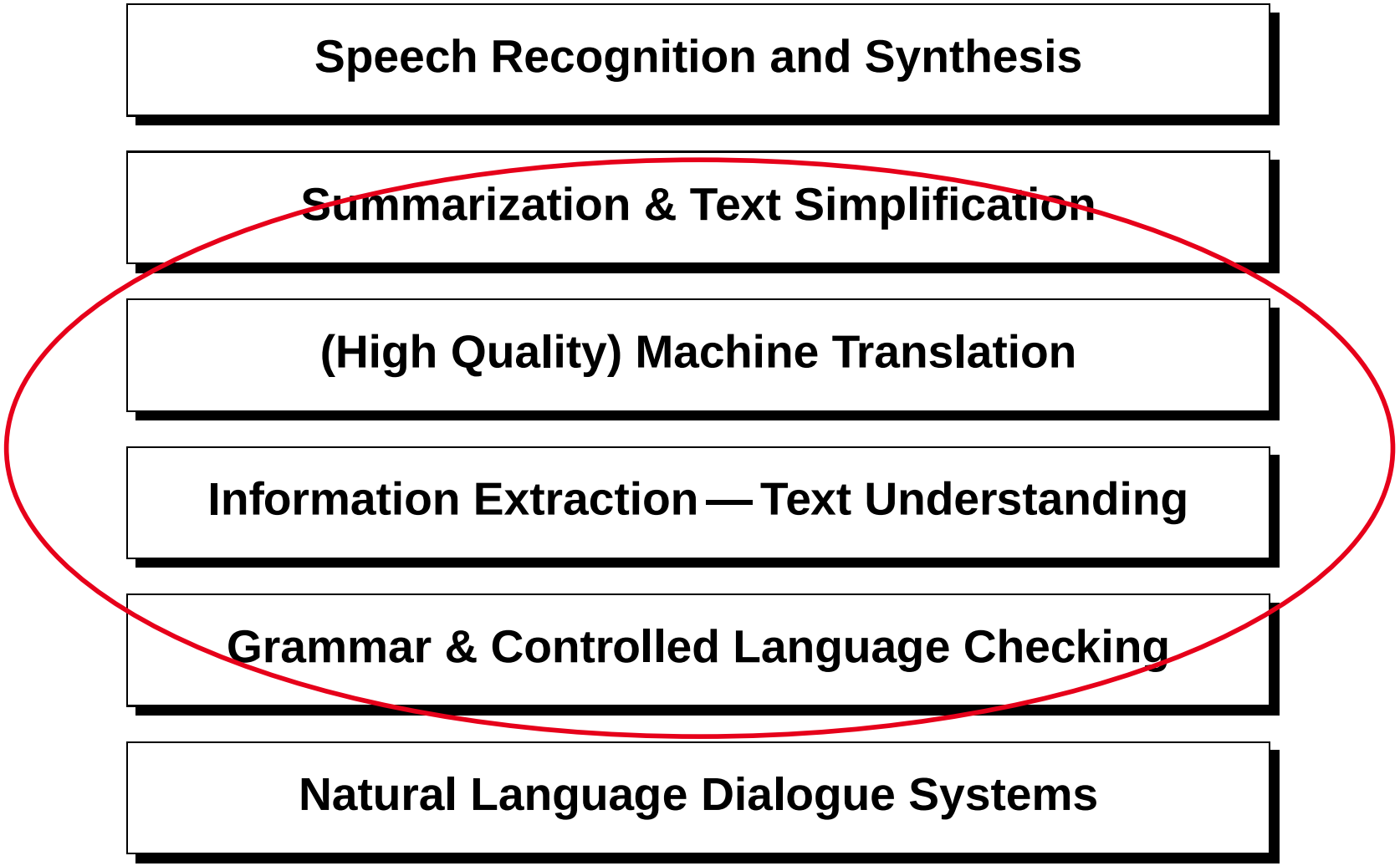
(High Quality) Machine Translation

Information Extraction — Text Understanding

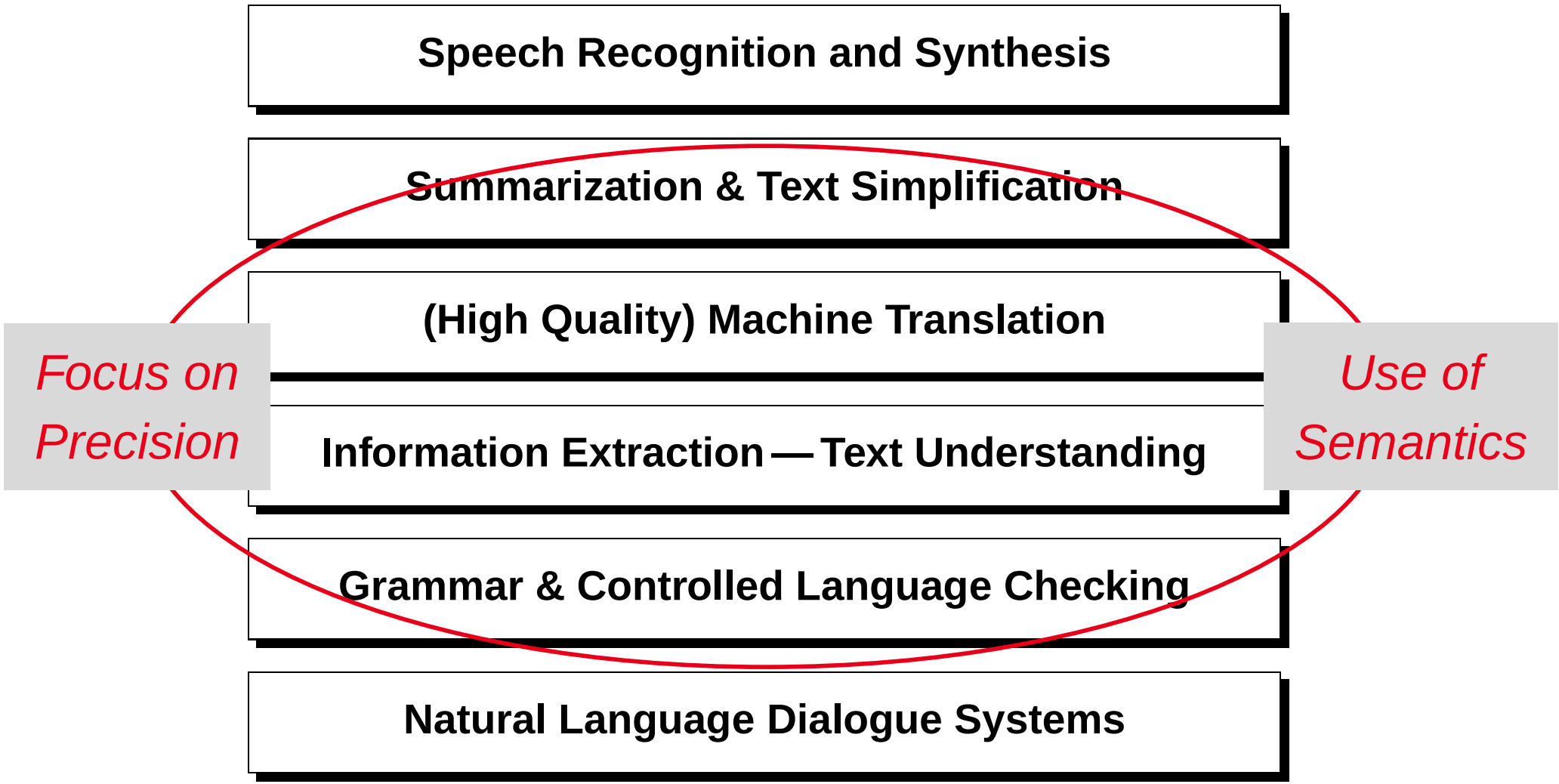
Grammar & Controlled Language Checking

Natural Language Dialogue Systems

Families of Language Processing Tasks



Families of Language Processing Tasks



What Makes Natural Language a Hard Problem?

```
|< |Den andre veien mot Bergen er kort.| --- 16 x 52 x 112 = 112
|> |That other path against Bergen is short.| [0.70] <0.03> (0:1:0).
|> |That other path towards Bergen is short.| [0.70] <0.03> (0:0:0).
|> |That second path against Bergen is short.| [0.65] <0.03> (2:1:0).
|> |That second path towards Bergen is short.| [0.65] <0.03> (2:0:0).
|> |That other road against Bergen is short.| [0.62] <0.03> (0:3:0).
|> |That other road towards Bergen is short.| [0.62] <0.03> (0:2:0).
...
|> |The second path against Bergen is short.| [0.18] <0.03> (3:1:0).
|> |The second path towards Bergen is short.| [0.18] <0.03> (3:0:0).
|> |That second path against Bergen is a card.| [0.17] <0.02> (8:1:0).
|> |That second path towards Bergen is a card.| [0.17] <0.02> (8:0:0).
|> |That other path against Bergen is cards.| [0.17] <0.03> (5:1:0).
|> |That other path towards Bergen is cards.| [0.17] <0.03> (5:0:0).
...
|> |Short is that other road, against Bergen.| [-0.37] <0.03> (0:3:2).
|> |Short is that second road, towards Bergen.| [-0.42] <0.03> (2:2:2).
```

What Makes Natural Language a Hard Problem?

```
|< |Den andre veien mot Bergen er kort.| --- 16 x 52 x 112 = 112
|> |That other path against Bergen is short.| [0.70] <0.03> (0:1:0).
|> |That other path towards Bergen is short.| [0.70] <0.03> (0:0:0).
|> |That second path against Bergen is short.| [0.65] <0.03> (2:1:0).
|> |That second path towards Bergen is short.| [0.65] <0.03> (2:0:0).
|> |That other road against Bergen is short.| [0.62] <0.03> (0:3:0).
|> |That other road towards Bergen is short.| [0.62] <0.03> (0:2:0).
...
|> |The second path against Bergen is short.| [0.18] <0.03> (3:1:0).
|> |The second path towards Bergen is short.| [0.18] <0.03> (3:0:0).
|> |                                                                :0).
|> |                                                                :0).
|> |                                                                ).
|> |                                                                ).
...
|> |                                                                :2).
|> |                                                                2:2).
```

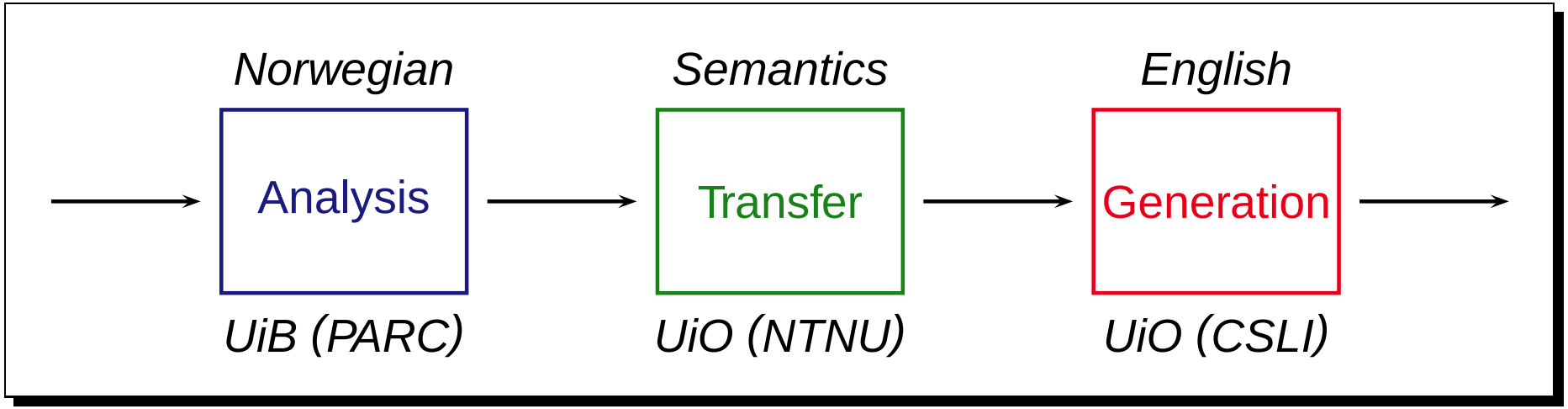
Scraped Off the Internet

the road to the other bergen is short .

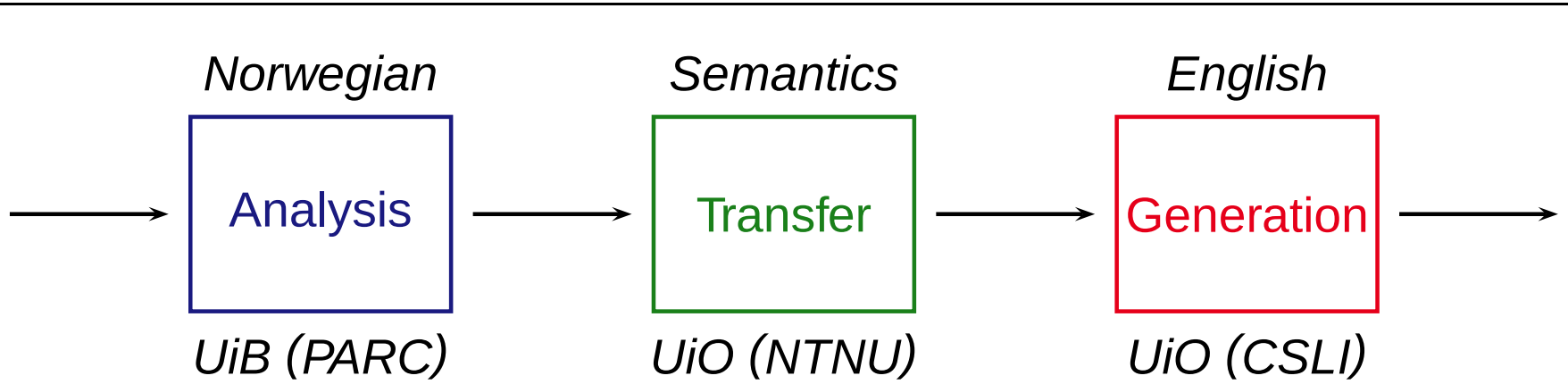
Den other roads against Boron Gene are short.

Other one autobahn against Mountains am abrupt.

An Example: Machine Translation in LOGON



An Example: Machine Translation in LOGON

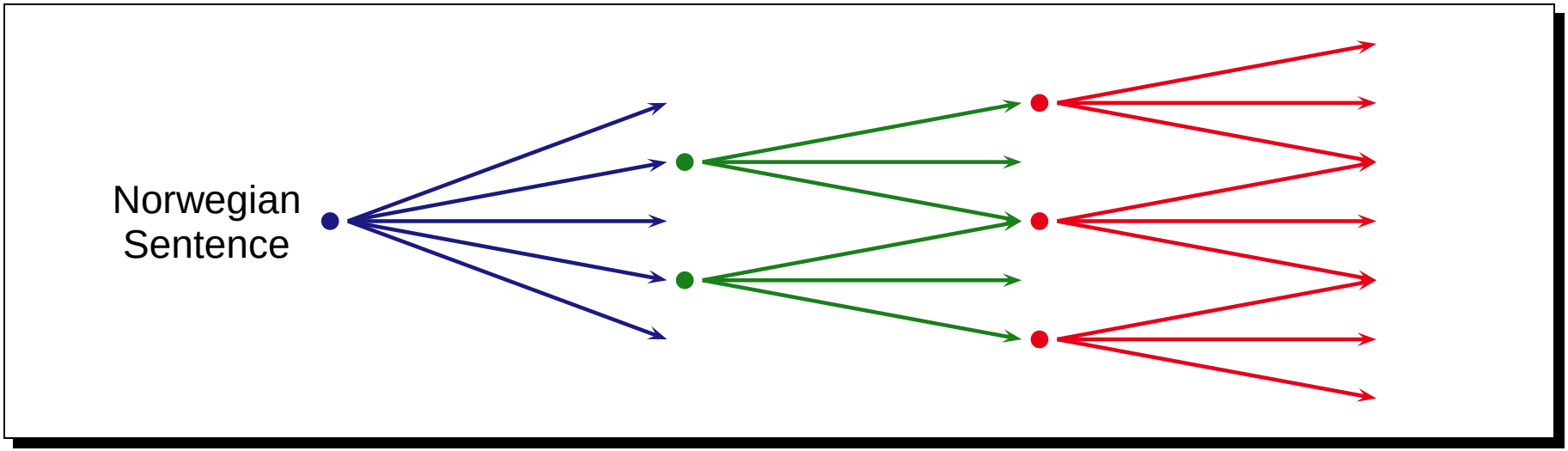


Some LOGON Highlights

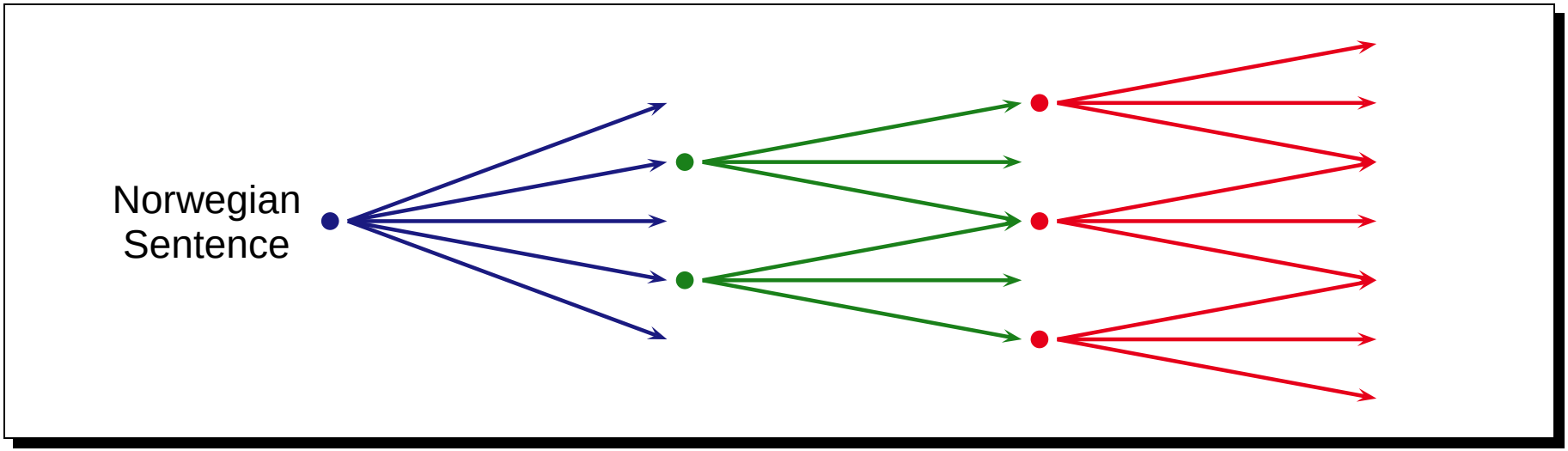
- Re-usable, mono-lingual precision grammars as linguistic back-bone;
- abstract from language-internal idiosyncrasies by semantic transfer;

→ limited domain and vocabulary: fully competitive with state-of-the-art.

Ambiguity Management: Stochastic Processes



Ambiguity Management: Stochastic Processes

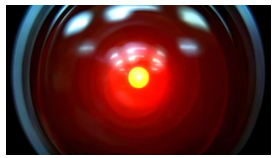


Stochastic Elements in LOGON

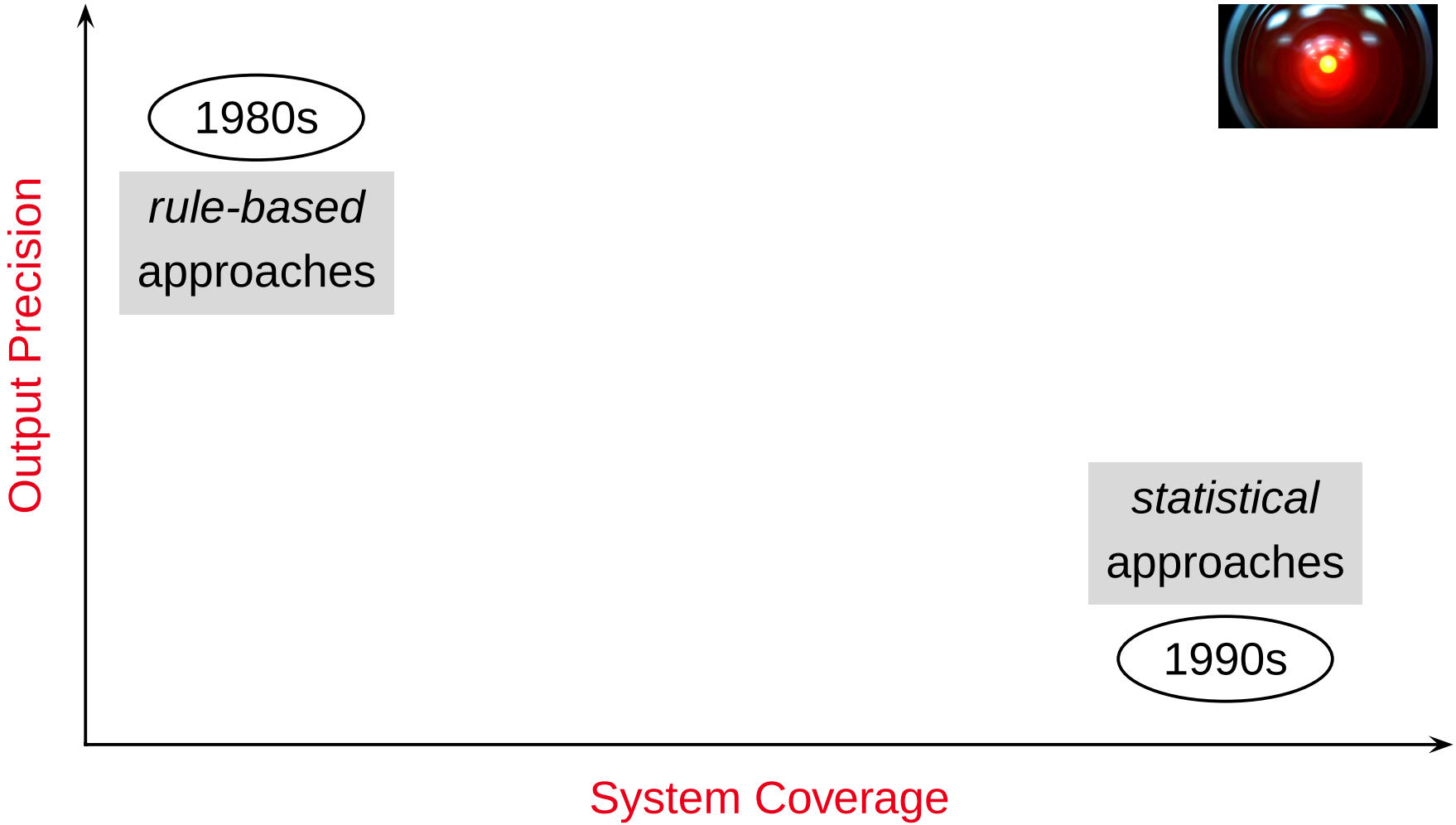
- At each stage, rank alternate hypotheses — finally, re-rank globally;
- computationally intensive: specialized algorithms, grid computation;

$\longrightarrow$ hybrid MT: linguistic back-bone combined with advanced statistics.

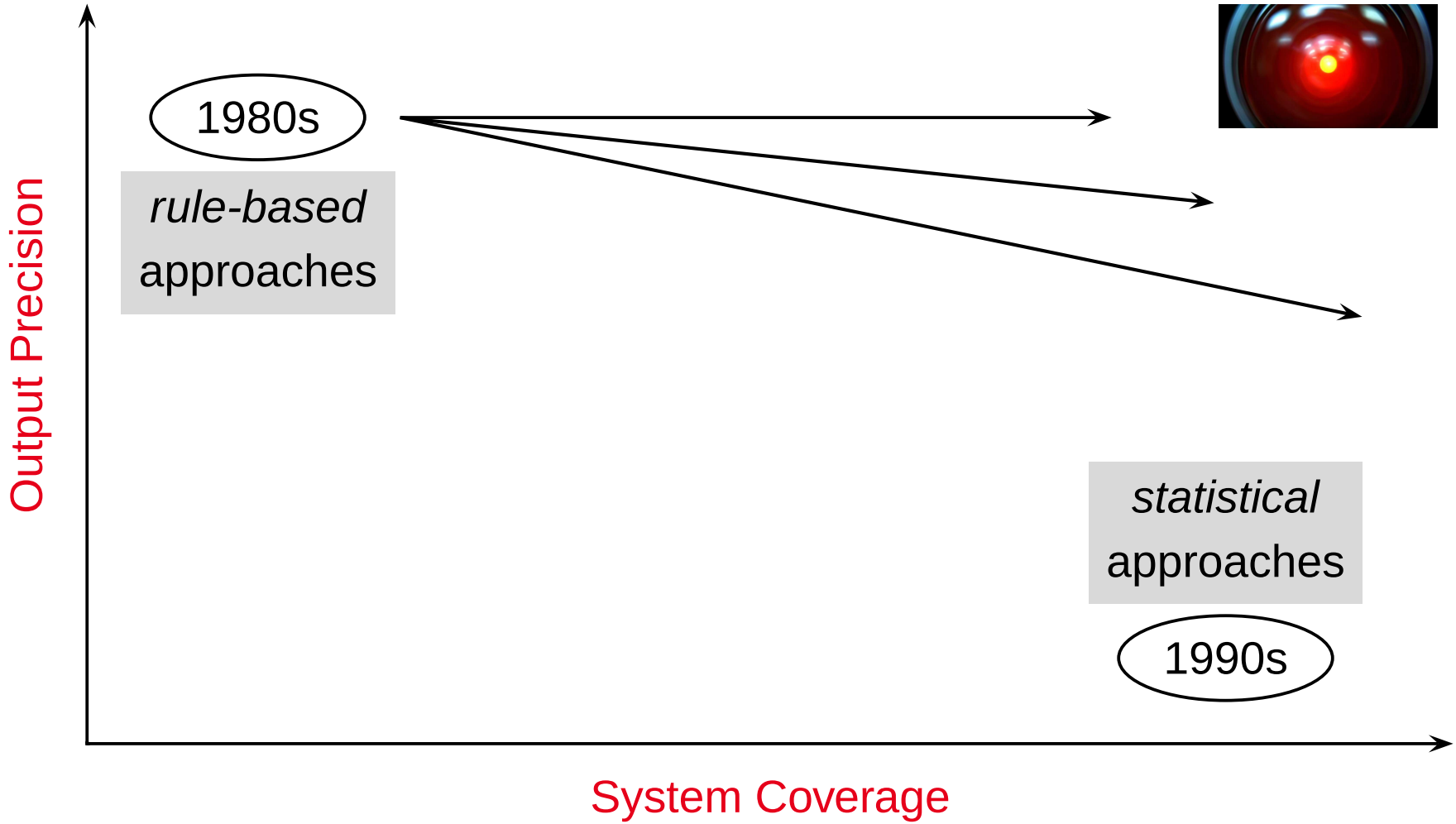
Our Vision: Balancing Coverage and Precision



Our Vision: Balancing Coverage and Precision



Our Vision: Balancing Coverage and Precision



The IFI (Logic and) Natural Languages Group

Language Technology

Doctoral Fellow	Liv Ellingsen	Soft Grammatical Constraints
Professor	Jan Tore Lønning	Computational Semantics
Professor	Stephan Oepen	Constraint-Based Processing
Associate Professor	Erik Velldal	Machine Learning
Doctoral Fellow	Gisle Ytrestøl	Incremental Parsing
Three MSc Candidates Finishing in 2008		

Computational Logic

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

Outlook: Research, Collaborations, Projects (1 of 2)

English Resource Grammar — DELPH-IN Repository

- Language Technology requires sustained, incremental development;
- de-facto standard: open-source repository of lingware and software;

$+$ co-founder of world-wide initiative, jointly with Stanford since 1993.

Machine Translation

- LOGON framework already in active use in Germany, Japan, and US;
- originally funded by NFR; currently pending approval of KUNSTI II;

$\longrightarrow$ harden core technology, scale up; maybe jointly with FP7 EuroMatrix.

Outlook: Research, Collaborations, Projects (2 of 2)

Web-Scale Parsing (*ESSENCE*)

- Semantic analysis of scholarly literature: 'who did what to whom';
- Wikipedia: map entities & relations; subjects, techniques, people;

$\longrightarrow$ 'twin' project with DFKI (Germany); submission to FriTek in 2008.

Cognitive Language Technology (*CoLT*)

- incremental parsing; connect computational and psycholinguists;

$\longrightarrow$ consortium just passed stage-one FET Open (IFI as coordinator).

Ontology Learning

$\rightsquigarrow$ Relate NL semantics to existing ontology; acquire new knowledge.

Outlook: Research, Collaborations, Projects (2 of 2)

Web-Scale Parsing (*ESSENCE*)

- Semantic analysis of scholarly literature: 'who did what to whom';
- Wikipedia: map entities & relations; subjects, techniques, people;

→ 'twin' project … riTek in 2008.

Gisle Ytrestøl (Doctoral Fellow)

Additional MN-Funded Position(s)

C… *T*)

- incremental parsing; connect computational and psycholinguists;

→ consortium just passed stage-one FET Open (IFI as coordinator).

Ontology Learning

⇝ Relate NL semantics to existing ontology; acquire new knowledge.

Another (Albeit Somewhat Dubious) Vision

[Dave] *Open the pod bay doors, HAL.*

[HAL] *I'm sorry Dave, I'm afraid I can't do that.*

[Dave] *Dave: What's the problem?*

[HAL] *I think you know what the problem is just as well as I do.*

[Dave] *What are you talking about, HAL?*

[HAL] *This mission is too important for me to allow you to jeopardize it.*

. . .

[HAL] *Dave, this conversation can serve no purpose anymore. Goodbye.*

Some LOGON Sample Translations (Version 0.9)

1 *Velkommen til Jotunheimen!*

Welcome to Jotunheimen.

1037 *På vestbredden lå det der tre setre nesten ved siden av hverandre.*

On the west bank, 3 mountain pastures lay there almost beside each other.

1048 *Vil du ikke gå så langt, er Besstrondrundhø et utmerket alternativ.*

If you don't want to go so far, Besstrondrundhø is an excellent alternative.

1376 *Den toppen er et fint turmål om du bor på Bessheim eller Gjendesheim.*

That summit, a nice trip tongue is if you stay at Bessheim or Gjendesheim.