

# Algorithms for AI and NLP (INF4820 — Welcome)

(defun ! (n) (if (equal n 0) 1 (\* n (! (- n 1)))))

**Erik Velldal and Stephan Oepen**

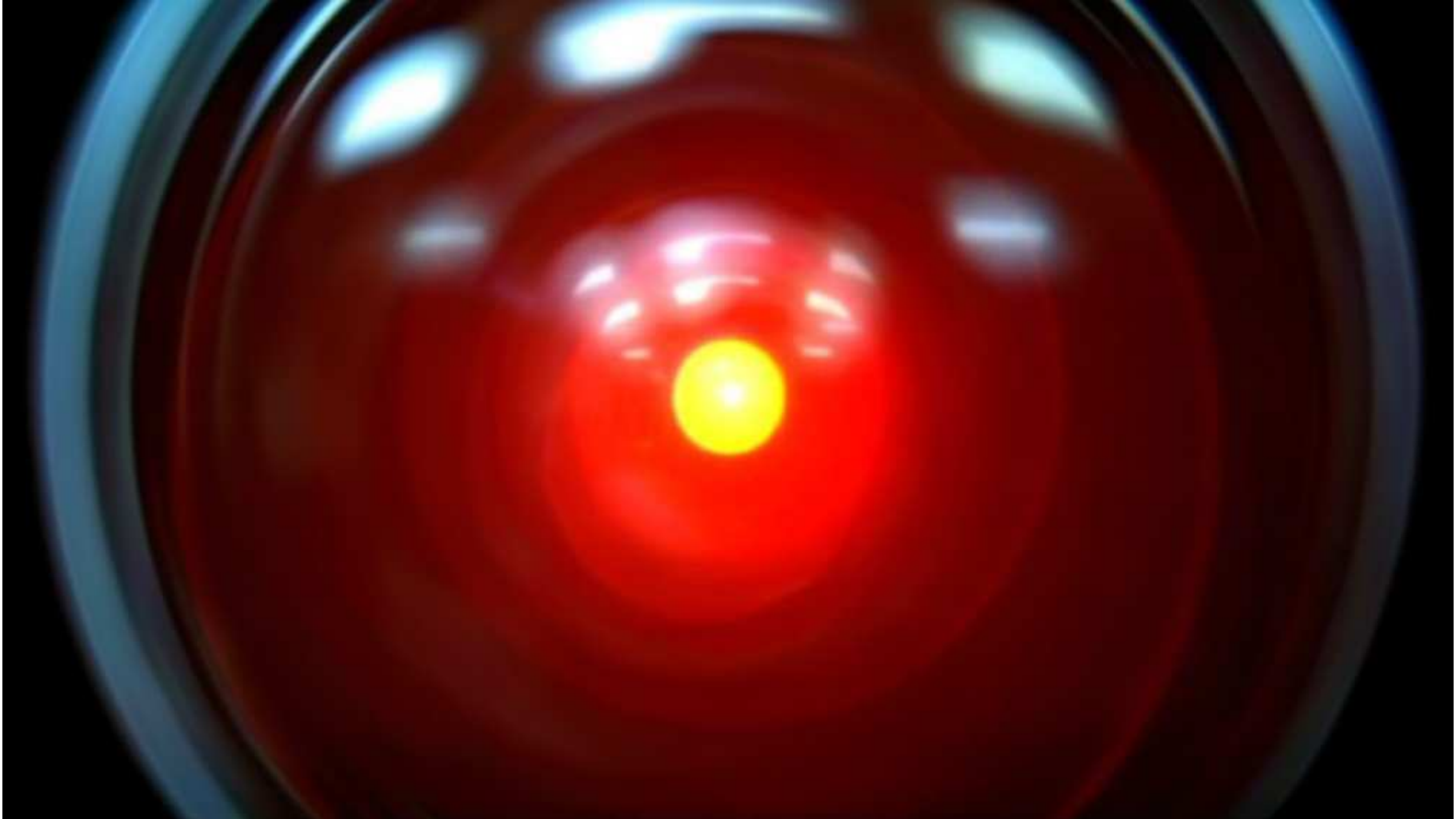
Universitetet i Oslo

{ erive | oe }@ifi.uio.no

# So, What Actually is Natural Language Processing?



# So, What Actually is Natural Language Processing?



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



INF4820 — 25-AUG-09 (oe@ifi.uio.no)

Natural Language Understanding (2)

# So, What Actually is Natural Language Processing?



- (young) interdisciplinary science: language, cognition, computation;
- (again) culturally and commercially relevant for 'knowledge society'.



# Well, Who is Actually Working on This?

*In the next three to five years, [...] mobile devices [...] will become prevalent. [...] Desired technologies will soon replace menus and graphic user interfaces with natural-language interfaces. — People so much want to speak English to their computer.* (Steve Ballmer, December 2005)



# Well, Who is Actually Working on This?

*In the next three to five years, [...] mobile devices [...] will become prevalent. [...] Desired technologies will soon replace menus and graphic user interfaces with natural-language interfaces. — People so much want to speak English to their computer.* (Steve Ballmer, December 2005)

*IBM has unveiled the details of its plans to build a computing system that can understand complex questions and answer with enough precision and speed to compete on America's favorite quiz show, Jeopardy!.* (IBM Press Release, April 27, 2009)



# 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

**Speech Recognition and Synthesis**

**Summarization & Text Simplification**

**(High Quality) Machine Translation**

**Information Extraction — Text Understanding**

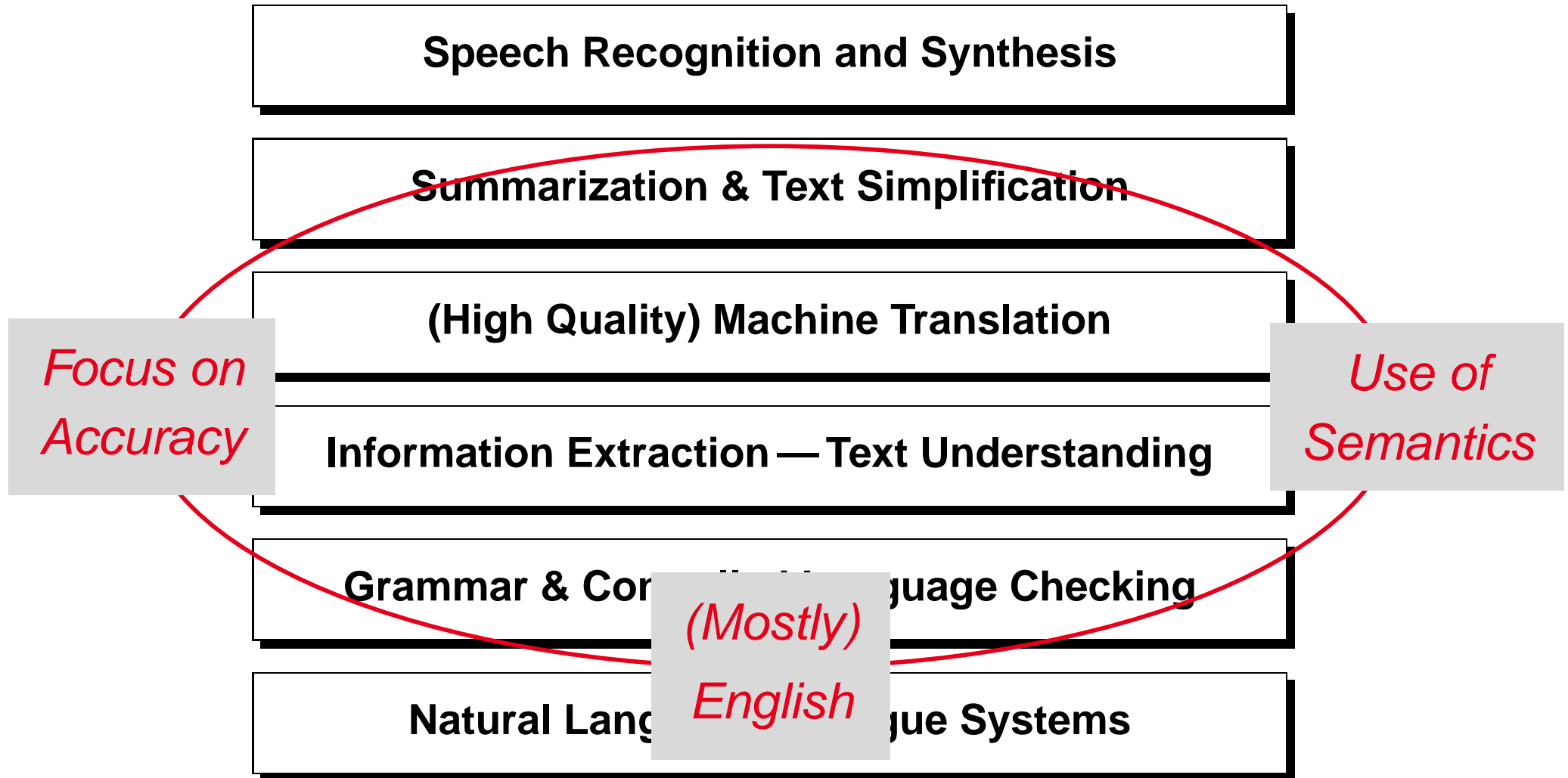
**Grammar & Controlled Language Checking**

**Natural Language Dialogue Systems**





# Families of Language Processing Tasks



# What Makes Natural Language a Hard Problem?

- < Den andre veien mot Bergen er kort. --- 12 x 30 x 25 = 25
- > The other path towards Bergen is short. {0.58} (1:1:0).
- > The other road towards Bergen is short. {0.56} (1:0:0).
- > The second road towards Bergen is short. {0.55} (2:0:0).
- > That other path towards Bergen is a card. {0.54} (0:1:0).
- > That other road towards Bergen is a card. {0.54} (0:0:0).
- > The second path towards Bergen is short. {0.51} (2:1:0).
- > The other road against Bergen is short. {0.48} (1:2:0).
- > The second road against Bergen is short. {0.48} (2:2:0).
- ...
- > Short is the other street towards Bergen. {0.33} (1:4:0).
- > Short is the second street towards Bergen. {0.33} (2:4:0).
- ...



# What Makes Natural Language a Hard Problem?

- < Den andre veien mot Bergen er kort. --- 12 x 30 x 25 = 25
- > The other path towards Bergen is short. {0.58} (1:1:0).
- > The other road towards Bergen is short. {0.56} (1:0:0).
- > The second road towards Bergen is short. {0.55} (2:0:0).
- > That other path towards Bergen is a card. {0.54} (0:1:0).
- > That other road towards Bergen is a card. {0.54} (0:0:0).
- > The second path towards Bergen is short. {0.51} (2:1:0).
- > The other road against Bergen is short. {0.48} (1:2:0).
- > Th ... )).

## Scraped Off the Internet

- ..
- > Sh the road to the other bergen is short . (0).
- > Sh Den other roads against Boron Gene are short. (1:0).
- .. Other one autobahn against Mountains am abrupt.



# More, and More, and More Ambiguity

## Speech Recognition

<i>its</i>	<i>hard</i>	<i>to</i>	<i>wreck</i>	<i>a</i>	<i>nice</i>	<i>beach</i>
<i>it</i>	<i>'s</i>	<i>hard</i>	<i>to</i>	<i>recognize</i>		<i>speech</i>

## Morphology

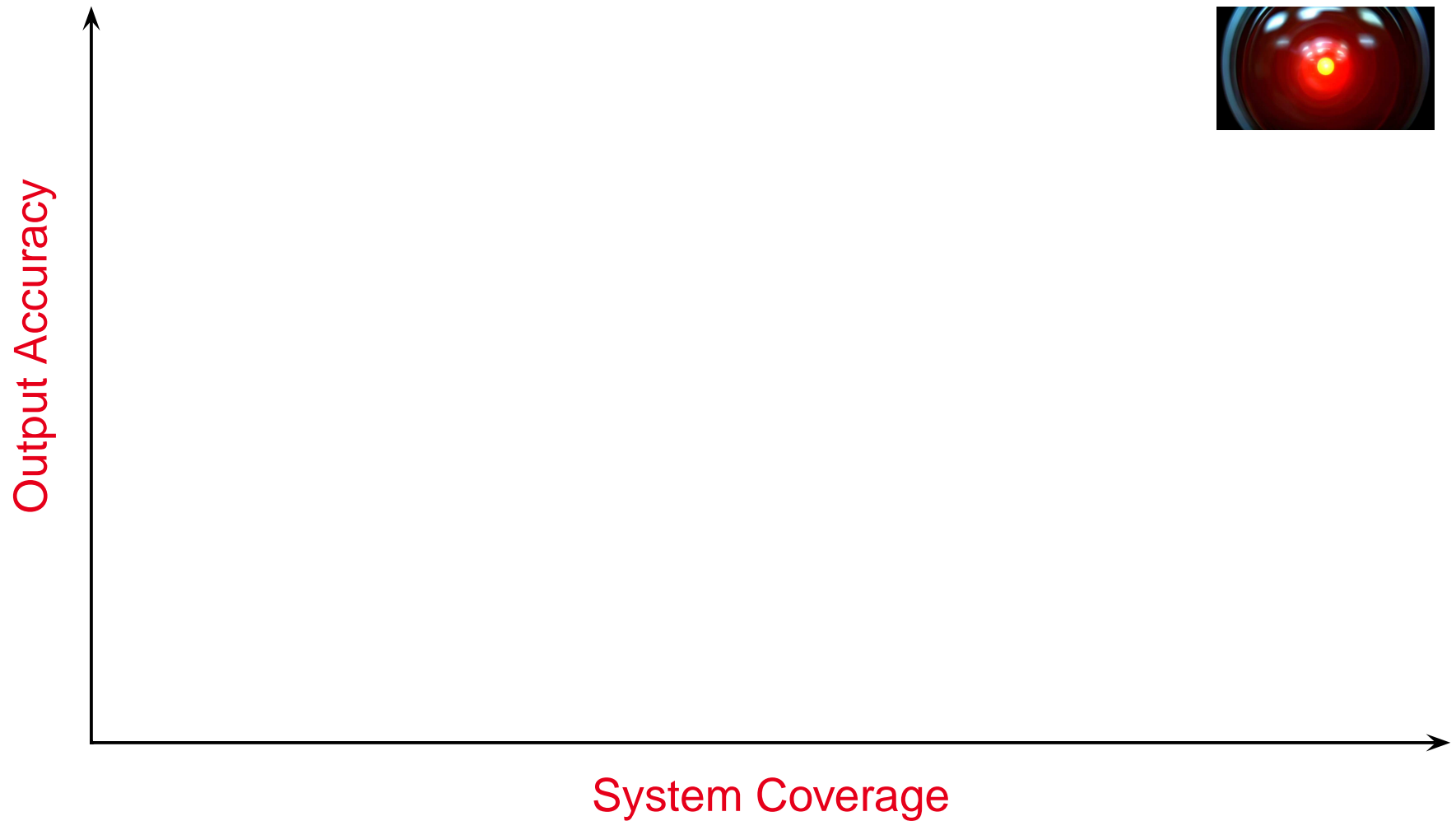
- *fisker* *fisk*<sub>N</sub> + plural vs. *fiske*<sub>V</sub> + present vs. *fisker*<sub>N</sub> + singular;
- *brus-automat* vs. *bru-sau-tomat*; *vinduene* vs. *vin-duene*; et al.

## Semantics

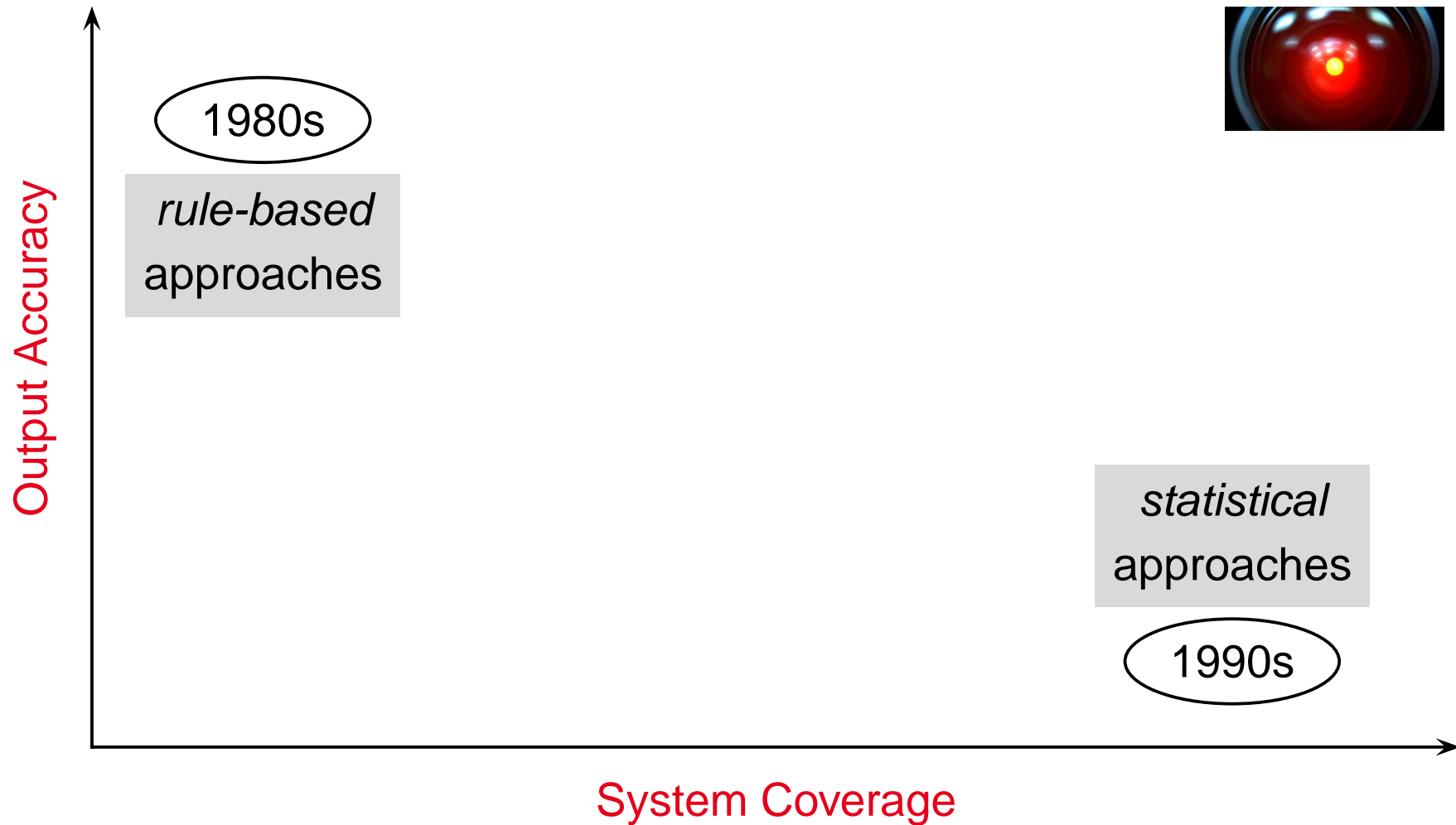
- *All Norwegians speak two languages.*  $\exists l_1, l_2 \forall n \dots$  vs.  $\forall n \exists l_1, l_2 \dots$



# The Holy Grail: Balancing Coverage and Accuracy



# The Holy Grail: Balancing Coverage and Accuracy



# The Holy Grail: Balancing Coverage and Accuracy

