INF5820

Language technological applications

H2008
Jan Tore Lønning/Stephan Oepen
jtl@ifi.uio.no/oe@ifi.uio.no

Today

- Examples of applications
- Approaches, methods
- Course content
- Practicalities
- Beginning machine translation

NLP applications

1. Translation
2. Dialogue
3. Voice control
4. Information processing
5. Speech ←→ text
6. Language support

1. Translation

- Fully-automatic text-translation: Systran, Google
- Speech-translation: commercial
- Aid for professional translators: trados
- From ca 1950, predating other NLP and AI: IBM 1954 press release

2. Dialogue systems

- Automatic number information
- Route schedule information
- Shopping
- Simple: say one or press one now!

3. Voice control

- Mobile phone
- Entertainment
- Navigation,
  - not perfect yet
- Fridge?
- Washer?
- Vacuum?

4. Information processing

- Document/web page search (information retrieval):
  - Language enhanced
- Factoid questions: PowerSet
- Summarization

5. Speech ←→ Text

- Speech synthesis: ATT
- Speech recognition: Dragon
- Part of applications

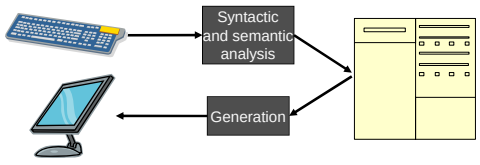
6. Language support tools

- Spell checkers
- Grammar checkers
- Translation help
- Foreign language learning tools

Today

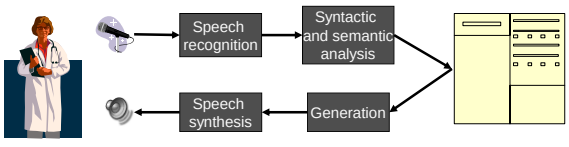
- Examples of applications
- Approaches, methods
- Course content
- Practicalities
- Beginning machine translation

Communicating with the computer



- The model of the computer as communicatior:
  - Analysis
  - Process
  - Generate/synthesis

Oral communication



- The model of the computer as communicatior:
  - Analysis: speech, grammar, semantics, pragmatics
  - Process
  - Generate/synthesis: content, grammar, speech

The communicating computer

- This models fits many applications
- The processing step varies:
  - Translation
  - Find an answer
  - Carry out an order
- Some applications are parts of this model/process, e.g. speech recognition
- Some applications don't fit this picture at all: spelling correction

Theoretical formal approaches

- Build a declarative model using
  - Linguistics
  - Logic
- Algorithms
- Symbolic
- Example:
  - S → NP VP
  - parsing

Theoretical less formal

- E.g., pragmatics:
  - Gricean implicatures
  - Dialogue handling: turntaking
- Heuristic algorithms

Empirical methods

- Learn from examples
- Generalize
- Examples:
  - Tagging
  - Speech recognition
  - Statistical MT

Hybrid methods

- Empirical methods need constructed categories
- Symbolic methods need empirical methods to handle ambiguity

Today

- Examples of applications
- Approaches, methods
- Course content
- Practicalities
- Beginning machine translation

A. Machine translation

1. MT overview
2. Statistical MT
3. Rule-based MT with semantic transfer
4. Hybrid methods
5. MT evaluation

A. Machine translation

1. MT overview
   1. What is MT, why is it difficult?
   2. History
   3. Approaches
2. Statistical MT
3. Rule-based MT with semantic transfer
4. Hybrid methods
5. MT evaluation

A. Machine translation

1. MT overview
2. Statistical MT
   - Prerequisite:
     Some statistical NLP, HMM tagging
3. Rule-based MT with semantic transfer
4. Hybrid methods
5. MT evaluation

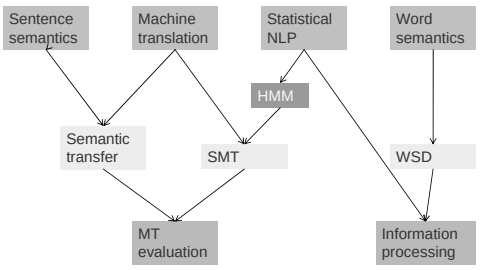
A. Machine translation

1. MT overview
2. Statistical MT
3. Rule-based MT with semantic transfer
   - Prerequisite:
     Sentence semantics
4. Hybrid methods
5. MT evaluation

B. Other applications

1. Word sense disambiguation (WSD)
   - Prerequisite:
     - word semantics
     - Some simple statistical NLP
2. Information processing

Dependencies



Proposed order

1. MT
2. Word semantics
3. Simple statisticsl NLP
4. WSD
5. Information processing
6. SMT
7. Sentence semantics
8. Semantic transfer
9. MT evaluation

Today

- Examples of applications
- Approaches, methods
- Course content
- Practicalities
- Beginning machine translation

INF5820

- http://www.uio.no/studier/emner/matnat/ifi/INF5820/index.xml

INF5820

- Bygger på INF4820 (kan tas samtidig)
- Alternerer med INF5830 Teoretisk datalingvistikk

Oblig.er

- http://www.uio.no/studier/emner/matnat/ifi/HUMIT4722MN/
- 4 obligatoriske innleveringer (godkjent-ikke godkjent)
- Semesteroppgave: A–F

Oppgave 1

- Legge frem en historisk artikkel 11.9
- Levere 1-2 sider referat. 9.9

Oppgaver videre

- Sett 2: 1.10
- Sett 3: 22.10
- Sett 4: 12.11 (Prosjektet del 1)
- Innlevering prosjekt 12.12

Undervisning

- Første del av semesteret: 2 ganger i uka
- Andre del: 1 gang i uka + prosjekt

Introducing MT

1. Why is machine translation a problem?
2. Traditional approaches:
   1. Direct
   2. Interlingua
   3. transfer
3. Empirical approaches:
   1. SMT
   2. Example-based MT (EBMT)
4. Some history
5. The LOGON approach
6. Evaluation

Eksempel: freetranslation.com

↓

The construction of the park lasted for a number of years. The area east of the two Frogner ponds had already by the turn of the century been opened to the public.

freetranslation.com :
Konstruksjonen av parken vart for et antall år. Områdesom øst av den to Frogner damer hatt allerede ved det vender av århundret vært åpnet til offentligheten.

Development

- The construction of the park lasted for a number of years. The area east of the two Frogner ponds had already by the turn of the century been opened to the public.
- freetranslation.com :
  Konstruksjonen av parken vart for et antall år. Områdesom øst av den to Frogner damer hatt allerede ved det vender av århundret vært åpnet til offentligheten.
- Google:
  Konstruksjonen av parken varte i en årrekke. Området øst for de to Frogner dammer hadde allerede ved århundreskiftet blitt åpnet for allmennheten.

Eksempel 2

- http://www.systransoft.com/index.html
- http://www.systranet.com/systran/net
- http://www.heeg.de/~uta/iX/art.htm
- www.dn.se

Not only word-by-word:

1. Grammatical structure - ambiguity
2. Structure is not always preserved
3. Lexical choice
4. More than sentence meaning

1. Ambiguous structure

De satte pris på dyrene
De uppskattade djuren
gir oss kjærlighet, vennskap og innimellom beskyttelse.
ger oss kärlek, vänskap och ibland skydd.

- Behov for fullstendig grammatisk analyse
- Fra 1950-tallet: Utvikling av grammatikker egnet for dataprosessering

2. Structure not preserved

Han heter Paul.
His name is Paul.
Il s'appelle Paul.

He likes to swim.
Er schwimmt gern.

3. Lexical choice

Jeg skar av makens tomme!
Jag skär av makens tumme!

The box is in the pen.
(Y. Bar-Hillel 1960)

- Problemet er "AI-hardt" – kan ikke vente fullgod løsning

4. Beyond sentence meaning

- Larger units, paragraphs
- Tracking the referent, No: den/det
- Metaphors, idioms
- Changre,
- Rhime, rythm
- Deliberate ambiguity, humor
- …