NLG and NLU

Hercules Dalianis
E-mail: hercules@kth.se
Tel: +46 70 568 13 59
http://www.dsv.su.se/~hercules

Differences speech text

- Differences NLP and NLG
- The computer’s production of text
- Applications of text generation

Natural language parsing

- Input of sentences/queries/text/
- Lexical analysis/morphological analysis
- Syntactic analysis
- Semantic analysis
- Pragmatic analysis

Similarities between NLU and NLG

- Same
  - Grammar and lexicon
  - Knowledge representation

DCG grammar for generation and parsing of text

```
text --> s.
text --> text,s.
s --> np,vp.
np --> det,noun.
vp --> tr_verb,np.
det --> [a].
det --> [every].
noun --> [man].
noun --> [woman].
tr_verb --> [loves].
```
DCG-Prolog

- DCG-Definite Clause Grammar
- Prolog programming language
- DCG-parsing-generation
- Demo

Discourse

- Discourse is the linguistic term for spoken or written text.
- A discourse contains a number of coherent sentences.

Spoken language

- Spoken language is passing, faster and less planned that written language, that is printed and exists for a while.
- The grammar is simple and lacks complicated constructions.

- Spoken language is loosely connected and contains incomplete sentences.
- Repeatings are common in spoken language
- Spoken language uses less number of words than written language.
- There are no punctuation marks in spoken language

- Spoken language is filled with small words that the speaker adds and that does not have any larger meaning for the content: hmm, ehh, you know.
- Spoken language has more dimensions than written language, dimensions as intonation, careless pronounced words and body language.

- Body language convey more information that the speakers spoken language. It is difficult to to decide what a word is in spoken language.
- All this dimension makes it difficult to analyze spoken language.
Written language

- Written language is more strictly grammatically than spoken language.
- It uses more words than spoken language.
- In the text there are punctuation marks, that informs when a sentence stops and when next sentence starts, Query marks that marks questions, citation marks.

- All these punctuation marks remove some of the ambiguities in written languages and makes it easier to analyze and formalize.
- Written language is more explicit than spoken language.

Hercules teaches all NLG-courses

∀(X) (nlg-course(X) -> teach(hercules, X)) (Predicate logic)

Demo

- Syntax
- Logic

NLP/NLU

- Problems
- Ill-formed input, spelling errors
- Not syntactic input
- Pragmatic problem
- Ambiguous input
- Dialogue