Once more about syntactical and lexical disambiguation

Ekaterina Ovchinnikova
St. Petersburg State University
Warszawa 20.12.04
Ideograph Project

- St. Petersburg State University, Department of Applied Linguistics (linguistic resources)
- Ideograph Company (programming resources)
Goals

- Formal grammar implementation platform (IdeoLog system)
- Semantic representation of texts in natural language
The sources of semantic and syntactic ambiguity

- Syntactical ambiguity
- Lexical ambiguity
- Semantic interpretation ambiguity
- Reference ambiguity and anaphoric ambiguity
Syntactical ambiguity

Ja byl znakom s nim
I was know_{part} with him
I was sign_{abl} with him

=>

I knew him
vs
I was a sign with him
<table>
<thead>
<tr>
<th>Russian Word</th>
<th>English Translation</th>
</tr>
</thead>
<tbody>
<tr>
<td>лицо RUS.n</td>
<td>face</td>
</tr>
<tr>
<td>человек RUS.n</td>
<td>person</td>
</tr>
<tr>
<td>ПОЗНАТЬ RUS.v</td>
<td>learn, analyze</td>
</tr>
<tr>
<td>ОСМОТРЕТЬ RUS.v</td>
<td>inspect, investigate</td>
</tr>
</tbody>
</table>
Semantic interpretation ambiguity

The portrait picturing the painters wife

vs

The portrait belonging to the painters wife

vs

The portrait painted by the painters wife

[who is also a painter]
Anaphoric ambiguity

Ivan udaril shestom po stolu, i on slomalsja.
Ivan had struck the table with a stick and it broke down.

=>

The table broke down.

vs

The stick broke down.
Levels of analyses in IdeoLog

- Syntactical analyses (formal grammars – AGFL and HPSG)
- Mapping of the lexemes set into the set of synsets in RussNet thesaurus
- Disambiguation in phrases with a parallel usage of the syntactical and lexical information
Syntactical analyses

- Affix Grammar over the Finite Lattice (AGFL)
  - Context free two level grammar
  - Morphology and syntax
  - Finite Lexicon
  - Derivational module (extension of the lexicon)

- Head Driven Phrase Structure Grammar (HPSG)
  - Typed Feature Structures
  - “Head-driven” mechanism
Mapping of the lexemes set into the set of synsets in RussNet

- **RussNet (wordnet-type lexicon)**
  - a lexical semantic database for the Russian language
  - including semantic, semantic-grammatical, semantic-derivational relations
  - using the frequency of the lexemes in the corpora in the RussNet structure
  - valence frames for synsets (!)
Previous analyses with wordnet-type thesaurus

- Mapping of the lexemes chains to the WordNet structures
- Defining the thematic area
- Finding the most likely analysis
- Hypothesis that texts are monothemematic (!)
Our main hypothesis

- Context distribution for lexical meanings
  (valence frames)
Valence frames in RussNet

Information for every valence:

- obligation (boolean)
- direction (active/passive)
- grammatical form of the linguistic object
- semantic class of the linguistic object
- semantic role
Valence frames in RussNet (example) (1)

**RUS-vПОЗНАТЬ.1.изучить**
*learn, analyze*

```xml
<VALENCE_FRAME main_segment="verb_phr">
  <VALENCE active="yes" obligatory="yes" val_type=obj1>
    <morph_data CASE="acc" />
    <sem_data TYPE="wildcat" ID="RUS-nНАУКА"/>
  </VALENCE>
  <VALENCE active="yes" obligatory="no" val_type=subj>
    <morph_data CASE="nom" />
    <sem_data TYPE="wildcat" ID="RUS-nЧЕЛОВЕК"/>
  </VALENCE>
</VALENCE_FRAME>
```
Valence frames in RussNet
(example) (2)

RUS-вОСМОТРЕТЬ.13.изучить
inspect, investigate

<VALENCE_FRAME main_segment="verb_phr">  
  <VALENCE active="yes" obligatory="yes" val_type=obj1>  
    <morph_data CASE="acc" />  
    <sem_data TYPE="wildcat" ID="RUS-нПРЕДМЕТ"/>  
  </VALENCE>  
  <VALENCE active="yes" obligatory="no" val_type=subj>  
    <morph_data CASE="nom" />  
    <sem_data TYPE="wildcat" ID="RUS-нЧЕЛОВЕК"/>  
  </VALENCE>  
</VALENCE_FRAME>
Disambiguation

Syntactic structure

+ Synsets with valence frames for every lexeme

= Lexicalized tree of analyses
Implemented

- Syntactic analyses
- Mapping of the lexemes set into the set of synsets in RussNet
- Disambiguation procedure
Thank you!