Linking VerbaLex with FrameNet: Case Study for the Indicate Verb Class

Jiří Materna

Centre for Natural Language Processing, FI MU Brno

December 4, 2010
Motivation

- Current frame-based lexical resources of Czech:
  - VerbaLex
  - Vallex
- FrameNet – lexical resource of English based on Frame Semantics.
- Similar projects for German, Spanish, Japanese, etc.
- Aim of the work:
  1. to build a FrameNet-like lexicon of Czech
since 1990 at the Berkeley university.
- based on frame semantics (C. J. Fillmore)
- exemplification in corpus data
- 8 frame-to-frame relations
- lexical Unit (LU) – pairing of a word with its meaning
- Frame Elements (FEs) – semantic roles belonging to the frame
- currently, FrameNet includes about:
  - 10,000 lexical units (6,100 fully annotated)
  - 850 semantic frames
  - 135,000 sentences annotated in the corpus
**Apply_heat**

**Definition:**
A Cook applies heat to Food, where the Temperature_setting of the heat and Duration of application may be specified. A Heating_instrument, generally indicated by a locative phrase, may also be expressed. Some cooking methods involve the use of a Medium (e.g. milk or water) by which heat is transferred to the Food. A less semantically prominent Food or Cook is marked Co_participant.

**Usage:**
Sally FRIED an egg in butter.
Sally FRIED an egg in a teflon pan.
Ellen FRIED the eggs with chopped tomatoes and garlic.
Apply_heat

Frame elements:
Container – The Container holds the Food to which heat is applied.
Cook – The Cook applies heat to the Food.
Food – Food is the entity to which heat is applied by the Cook.
Heat_instr – This FE identifies the entity that directly supplies heat to the Food.
Temp – This FE identifies the Temperature_setting of the Heating_instrument for the Food.

Lexical units:
bake.v, barbecue.v, blanch.v, boil.v, braiandse.v, broil.v, brown.v, char.v, coddle.v, cook.v, deep fry.v, fry.v, grill.v, microwave.v, parboil.v, poach.v, roast.v, saute.v, scald.v, sear.v, simmer.v, singe.v, steam.v, steep.v, stew.v, toast.v
20. **THERE** Existence **WERE** Existence **ONLY** Sole_instance **TWO** Cardinal_numbers **MEN** People **IN** Locative_relation the **ROOM** Building_subparts, Sir **Henry** and **Stapleton**. They **SAT** Posture with their profiles towards me **ON** Locative_relation either **SIDE** Part_orientational of the round table. Both of them were **SMOKING** Ingest_substance cigars, and **COFFEE** Food and wine were in front of them. **Stapleton** was **TALKING** Chatting with animation, but the baronet **LOOKED** Appearance pale and distrait. Perhaps the **THOUGHT** Cognition of that lonely **WALK** Self_motion across the ill-omened **MOOR** Biological_area was weighing heavily upon his mind.

Clear Sentences   Turn Colors Off

[X] **THERE** **WERE** only two men **in** the **room**, Sir Henry and Stapleton.
[X] There were **ONLY** two men in the **room**, Sir Henry and Stapleton.
[X] There were only two men **IN** the **room**, Sir Henry and Stapleton.
[X] They **SAT** with their profiles towards me **ON** either side of the round table.
[X] Both of them were **SMOKING** cigars, and coffee and wine were in front of them.
VerbaLex

- lexicon of Czech verb valencies developed at Faculty of Informatics MU
- directly connected to WordNet
- Deep (semantic) valencies on two levels:
  - 1st level – 29 semantic roles (similar to Fillmore’s semantic cases)
  - 2nd level – part of the top ontology from EuroWordNet
- currently more than 26,000 frames
Figure: VerbaLex frame example: to mine something
Linking VerbaLex with FrameNet

Figure: Semi-automatic linking tool.
The case study

- *Indicate* – one of 111 semantic classes defined in VerbaLex
- 136 verb senses
- 27 CZWN synsets
- 119 valency frames
<table>
<thead>
<tr>
<th>FrameNet frame name</th>
<th>VerbaLex frames</th>
</tr>
</thead>
<tbody>
<tr>
<td>Telling</td>
<td>30</td>
</tr>
<tr>
<td>Reasoning</td>
<td>22</td>
</tr>
<tr>
<td>Reveal_secret</td>
<td>14</td>
</tr>
<tr>
<td>Gesture</td>
<td>11</td>
</tr>
<tr>
<td>Expressing_publicly</td>
<td>7</td>
</tr>
<tr>
<td>Forgiveness</td>
<td>5</td>
</tr>
<tr>
<td>Communication</td>
<td>4</td>
</tr>
<tr>
<td>Sign</td>
<td>3</td>
</tr>
<tr>
<td>Others</td>
<td>9</td>
</tr>
<tr>
<td>None</td>
<td>9</td>
</tr>
</tbody>
</table>

Table: Assigned FrameNet frames.
Problems in the frame linking

- Inadequacy of frame definitions in the corresponding semantic domain or area.
- Insufficient coverage of the domain in Berkeley FrameNet (i.e. English lexical units and corresponding frames have not been defined yet).
- (too fine-grained senses in FrameNet)
frame element is missing:

*Jeho pohled nám naznačil, že nemluví pravdu.*

AG(co1;<quality:1>) VERB PAT(komu3;<person:1>)
INFO(const;<info:1>)

FrameNet frame **Sign:**
Indicated (INFO)
Indicator (AG)
Degree
frame element is too general

*Ta zpěvačka demonstrovala svou lásku ke zvířatům.*

AG(kdo1;<person:1>) VERB ACT(co4;<act:2>) PAT(komu3;<animal:1>)

**FrameNet frame** Expressing_publicly:
Communicator (AG)
Content
Medium
frame element is too specific

*Ten pes cenil zuby na kočku za stromem.*

AG<animal:1> VERB DPHR<zuby> PAT
<person:1|animal:1>

FrameNet frame **Gesture:**
Adressee (PAT)
Body_part (DPHR)
Communicator (AG) ?
Indicated_entity
Message
Conclusions

- 99% VerbaLex frames linkable to FrameNet
- 82% VerbaLex frames directly linkable without any modifications of their FEs
- linking VerbaLex with FrameNet is a good starting point for building FrameNet in Czech
Thank you for your attention.