Exploring and Extending Czech WordNet and VerbaLex

RASLAN 2009

Zuzana Nevěřilová

December 4–6, 2009

Describing VerbaLex

Exploring VerbaLex

Extending VerbaLex

Conclusion

VerbaLex entry consists of

- ▶ verb *frame*
- ▶ verb lemmata
- ▶ 2 layers of semantic information:
    - ▶ semantic role (AG, PAT, LOC. . .)
    - ▶ relation to WordNet *hypernym*

VerbaLex entry consists of

- verb *frame*
- verb lemmata
- 2 layers of semantic information:
    - semantic role (AG, PAT, LOC...)
    - relation to WordNet *hypernym*

VerbaLex entry consists of

- ▶ verb *frame*
- ▶ verb lemmata
- ▶ 2 layers of semantic information:
  - ▶ semantic role (AG, PAT, LOC...)
  - ▶ relation to WordNet *hypernym*

Example of a Frame

```xml
<frame_entry frame_index='1'>
  <frame_lemmata>
    <lemma  sense='1' aspect='impf'>žádat</lemma>
  </frame_lemmata>
  <synonym_lemmata>
    <lemma aspect='impf' sense='8'>chtít</lemma>
    <lemma aspect='impf' sense='7'>vyžadovat</lemma>
  </synonym_lemmata>
  <gloss>ned: chtít:8 ned: vyžadovat:7</gloss>
  <example>ned: žádám tě o omluvu</example>
  <use>prim</use>
  <frame_slots>
    <slot number='1' functor='AG' type='obl' class='person:1'>
      <form type='direct_case' case='kdo1' />
    </slot>
    <slot number='2' type='obl' functor='VERB'/>
    <slot number='3' functor='PAT' type='obl' class='person:1'>
      <form type='direct_case' case='koho4' />
    </slot>
    <slot number='4' functor='ACT' type='obl' class='act:2'>
      <form type='prepos_case' prepos_lemma='o' case='co4' />
    </slot>
  </frame_slots>
</frame_entry>
```

XML + XSL = RDF... and Prolog

VerbaLex and WordNet are converted to RDF through a XSL transformation.

Frame ID is generated.

XML + XSL = RDF... and Prolog

VerbaLex and WordNet are converted to RDF through a XSL transformation.

Frame ID is generated.

SWI Prolog module `rdf_db` is used.

⋮

```
Parsed "verbalex2.rdf" in 141.67 sec; 219,200 triples
Parsed "wncze.utf8.rdf" in 5.98 sec; 96,690 triples
```

Syntactic Analysis and Semantic Role Assignment

```
?- find_roles('přicestovat',['ministr','zastávka'],FrameID,Roles).
FrameID = 'http://nlp.fi.muni.cz/verbalex#pRicestovat_1_2',
Roles = [ (ministr, 'AG', kdo1, obl), (zastávka, 'LOC', čeho2, opt)] ;
```

Syntactic Analysis and Semantic Role Assignment

```
?- find_roles('přicestovat',['ministr','zastávka'],FrameID,Roles).
FrameID = 'http://nlp.fi.muni.cz/verbalex#pRicestovat_1_2',
Roles = [ (ministr, 'AG', kdo1, obl), (zastávka, 'LOC', čeho2, opt)] ;
```

- ▶ appropriate verb meaning recognition
- ▶ frame identification
- ▶ semantic roles assignment
- ▶ grammatical information (cases)

Evaluation

Manual evaluation of 17 clauses from newspaper texts (idnes.cz)

- ▶ in 6 cases no appropriate frame was found
- ▶ in 8 cases a frame with verb + one slot was found
- ▶ no frame with verb + more slots was found

Evaluation

Manual evaluation of 17 clauses from newspaper texts (idnes.cz)

- ▶ in 6 cases no appropriate frame was found
- ▶ in 8 cases a frame with verb + one slot was found
- ▶ no frame with verb + more slots was found

Evaluation

Manual evaluation of 17 clauses from newspaper texts (idnes.cz)

▶ in 6 cases no appropriate frame was found
▶ in 8 cases a frame with verb + one slot was found
▶ no frame with verb + more slots was found

Evaluation

Following problems occur:

- ▶ word from the list not found in CZWN (CZWN is small and does not contain proper names)
- ▶ verb (with appropriate meaning) not found in VerbaLex
- ▶ no suitable frame for the list of words (constraints on grammatical information, constraints on CZWN hypernym)

Evaluation

Following problems occur:

- ▶ word from the list not found in CZWN (CZWN is small and does not contain proper names)
- ▶ verb (with appropriate meaning) not found in VerbaLex
- ▶ no suitable frame for the list of words (constraints on grammatical information, constraints on CZWN hypernym)

Evaluation

Following problems occur:

- ▶ word from the list not found in CZWN (CZWN is small and does not contain proper names)
- ▶ verb (with appropriate meaning) not found in VerbaLex
- ▶ no suitable frame for the list of words (constraints on grammatical information, constraints on CZWN hypernym)

Extending VerbaLex

VerbaLex is frame-based knowledge representation.

There are no relations among the frames.

Extending VerbaLex

VerbaLex is frame-based knowledge representation.

There are no relations among the frames.

Frame-based knowledge representations consist at least of:

- ▶ preconditions
- ▶ effects
- ▶ decompositions

(see FrameNet)

Generating New Propositions

```
?- find_roles('přicestovat',['ministr','zastávka'],FrameID,Roles).
FrameID = 'http://nlp.fi.muni.cz/verbalex#pRicestovat_1_2',
Roles = [ (ministr, 'AG', kdo1, obl), (zastávka, 'LOC', čeho2, opt)] ;

?- find_effect('přicestovat',['ministr','zastávka'],FrameID,Roles).
FrameID = 'http://nlp.fi.muni.cz/verbalex#nachAzet_se_1_1',
Roles = [ (ministr, 'ENT', kdo1, obl), (zastávka, 'LOC', čem6, opt)] .
```

- ▶ new frame identification
- ▶ change of roles assignment (AG → ENT)
- ▶ change of grammatical information (čeho2 → čem6)

Generating New Propositions

```
?- find_roles('přicestovat',['ministr','zastávka'],FrameID,Roles).
FrameID = 'http://nlp.fi.muni.cz/verbalex#pRicestovat_1_2',
Roles = [ (ministr, 'AG', kdo1, obl), (zastávka, 'LOC', čeho2, opt)] ;

?- find_effect('přicestovat',['ministr','zastávka'],FrameID,Roles).
FrameID = 'http://nlp.fi.muni.cz/verbalex#nachAzet_se_1_1',
Roles = [ (ministr, 'ENT', kdo1, obl), (zastávka, 'LOC', čem6, opt)] .
```

▶ new frame identification

▶ change of roles assignment (AG → ENT)

▶ change of grammatical information (čeho2 → čem6)

Generating New Propositions

```
?- find_roles('přicestovat',['ministr','zastávka'],FrameID,Roles).
FrameID = 'http://nlp.fi.muni.cz/verbalex#pRicestovat_1_2',
Roles = [ (ministr, 'AG', kdo1, obl), (zastávka, 'LOC', čeho2, opt)] ;

?- find_effect('přicestovat',['ministr','zastávka'],FrameID,Roles).
FrameID = 'http://nlp.fi.muni.cz/verbalex#nachAzet_se_1_1',
Roles = [ (ministr, 'ENT', kdo1, obl), (zastávka, 'LOC', čem6, opt)] .
```

- ▶ new frame identification
- ▶ change of roles assignment (AG → ENT)
- ▶ change of grammatical information (čeho2 → čem6)

Inference Rules

```
<proposition rdf:about="#pRicestovat_1_1_effect_1">
  <action rdf:resource="#pRicestovat_1_1"/>
  <effect rdf:resource="#nachAzet_se_1_1"/>
  <mapping>
    <map>
      <from rdf:resource="#AG"/>
      <into rdf:resource="#ENT"/>
    </map>
  </mapping>
  <mapping>
    <map>
      <from rdf:resource="#LOC"/>
      <into rdf:resource="#LOC"/>
    </map>
  </mapping>
</proposition>
```

Inference Rules

```
<proposition rdf:about="#pRicestovat_1_1_effect_1">
  <action rdf:resource="#pRicestovat_1_1"/>
  <effect rdf:resource="#nachAzet_se_1_1"/>
  <mapping>
    <map>
      <from rdf:resource="#AG"/>
      <into rdf:resource="#ENT"/>
    </map>
  </mapping>
  <mapping>
    <map>
      <from rdf:resource="#LOC"/>
      <into rdf:resource="#LOC"/>
    </map>
  </mapping>
</proposition>
```
How to add precondition, effect, decomposition rules efficiently?

Classification of Verb Frames

We supposed that verbs of a semantic class have same frame structure.

Classification of Verb Frames

We supposed that verbs of a semantic class have same frame structure.

The pattern (AG, kdo1), (PAT, komu3), (INFO, co4) is contained in 29 frames:

avizovat, hlásit, ohlásit, ohlašovat, oznámit, oznamovat, dát, dávat, doplňovat, doplnit, donášet, donést, nahlásit, nahrát, nahrávat, přehrát, přehrávat, odůvodňovat, odůvodnit, sdělit, sdělovat, vyřídit, vyřizovat, podtrhnout, podtrhovat, vypíchnout, vypichovat, postavit, povídat, povědět, vyprávět, vypravovat, říct, říci, říkat, tvrdit, tajit, utajit, utajovat, zamlčet, zamlčovat, zapírat, zapřít, zatajit, zatajovat, vyklápět, vyklopit, vypátrat, vyzkoumat, zjišťovat, zjistit, vypozorovat, vysledovat

Classification of Verb Frames

We supposed that verbs of a semantic class have same frame structure.

The pattern (AG, kdo1), (PAT, komu3), (INFO, co4) is contained in 29 frames:

avizovat, hlásit, ohlásit, ohlašovat, oznámit, oznamovat, dát, dávat, doplňovat, doplnit, donášet, donést, nahlásit, nahrát, nahrávat, přehrát, přehrávat, odůvodňovat, odůvodnit, sdělit, sdělovat, vyřídit, vyřizovat, podtrhnout, podtrhovat, vypíchnout, vypichovat, postavit, povídat, povědět, vyprávět, vypravovat, říct, říci, říkat, tvrdit, tajit, utajit, utajovat, zamlčet, zamlčovat, zapírat, zapřít, zatajit, zatajovat, vyklápět, vyklopit, vypátrat, vyzkoumat, zjišťovat, zjistit, vypozorovat, vysledovat

The pattern (AG, kdo1), (LOC, čem6), (COM, ˙) is contained in 10 frames:

cestovat, jezdit, putovat, hlásat, šířit, oznámit, oznamovat, rozhlásit, rozhlašovat, rozšířit, rozšiřovat, roznášet, nacházet, najít, nalézat, naleznout, nalézt, odhalit, odhalovat, vypátrat, vypozorovat, účinkovat, vystoupit, vystupovat, zapíjet, zapít. . .

Classification of Verb Frames

We supposed that verbs of a semantic class have same frame structure.

The pattern (AG, kdo1), (PAT, komu3), (INFO, co4) is contained in 29 frames:

avizovat, hlásit, ohlásit, ohlašovat, oznámit, oznamovat, dát, dávat, doplňovat, doplnit, donášet, donést, nahlásit, nahrát, nahrávat, přehrát, přehrávat, odůvodňovat, odůvodnit, sdělit, sdělovat, vyřídit, vyřizovat, podtrhnout, podtrhovat, vypíchnout, vypichovat, postavit, povídat, povědět, vyprávět, vypravovat, říct, říci, říkat, tvrdit, tajit, utajit, utajovat, zamlčet, zamlčovat, zapírat, zapřít, zatajit, zatajovat, vyklápět, vyklopit, vypátrat, vyzkoumat, zjišťovat, zjistit, vypozorovat, vysledovat

The pattern (AG, kdo1), (LOC, čem6), (COM, ˙) is contained in 10 frames:

cestovat, jezdit, putovat, hlásat, šířit, oznámit, oznamovat, rozhlásit, rozhlašovat, rozšířit, rozšiřovat, roznášet, nacházet, najít, nalézat, naleznout, nalézt, odhalit, odhalovat, vypátrat, vypozorovat, účinkovat, vystoupit, vystupovat, zapíjet, zapít...

It is necessary to ask the right questions

Contribution to NLP and Future Work

Nowadays: feedback for VerbaLex (and CZWN)

evaluation of more texts

Contribution to NLP and Future Work

Nowadays: feedback for VerbaLex (and CZWN)

evaluation of more texts

In the future: discourse analysis

all objects mentioned in the discourse, their roles, properties, location etc.

Thank you

`mailto:xpopelk@aurora.fi.muni.cz`