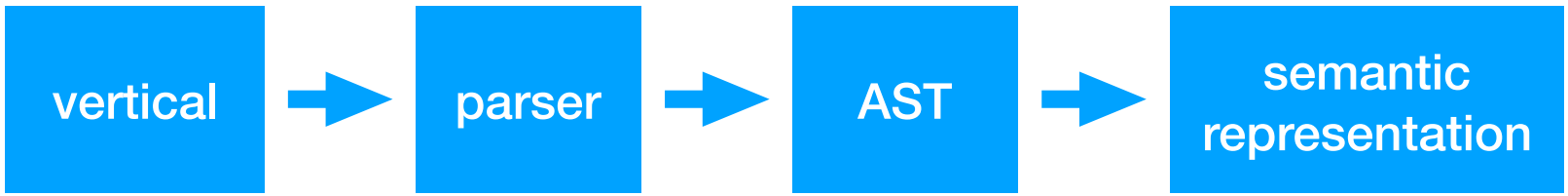
Multilinguality Adaptations of Natural Language Logical Analyzer

Marek Medveď, Terézia Šulganová, and Aleš Horák

Natural Language Processing Centre Faculty of Informatics, Masaryk University
Botanická 68a, 602 00 Brno, Czech Republic

December 1, 2017

Workflow



Introduction

- semantic representation of one expression through multiple languages should be (structurally) very similar
- AST was designed as language independent tool that from a syntactic tree sentence representation can create its logical representation
- input to the AST processor from standard plain text sentences, we use the SET parser
- we describe the SET and AST modifications that allow flexible multilingual setup of the whole pipeline

Tagset Translation

- grammatical agreement test
- Czech attributive tagset - carries lot of information about the grammatical case, number, gender, person

Tagset translation

Example (Subject-predicate agreement rule)

system supplies an inexplicit subject formed with a personal pronoun within correct number, gender and person from the form of the main

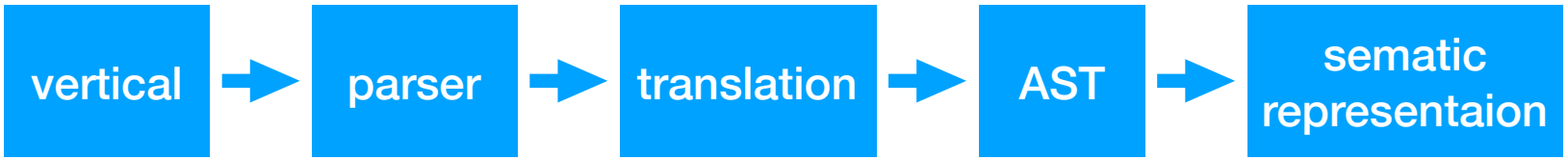
Tagset Translation

English Pen TreeBank tags			*Czech attributive tags*		
word	*lemma*	*tag*	*word*	*lemma*	*tag*
Some	some	DT	Some	some	k3
agents	agent	NNS	agents	agent	k1gInP
are	be	VBP	are	be	k5mInP
mobile	mobile	JJ	mobile	mobile	k2gId1
,	,	,	,	,	kIx,
other	other	JJ	other	other	k2gId1
agents	agent	NNS	agents	agent	k1gInP
are	be	VBP	are	be	k5mInP
static	static	JJ	static	static	k2gId1
.	.	SENT	.	.	kIx.

Tagset translation (Rules)

- pronouns:
  - masculine gender (gM) for pronouns: he, his, himself (similarly for feminine and neuter genders)
  - personal pronoun (xP) for words with tag PP,
  - possessive pronoun (xO) for word with CDZ, PPZ;
- conjunctions:
  - coordinative type for and, but, for, nor, or, so, yet,
  - subordinate type otherwise.

Workflow

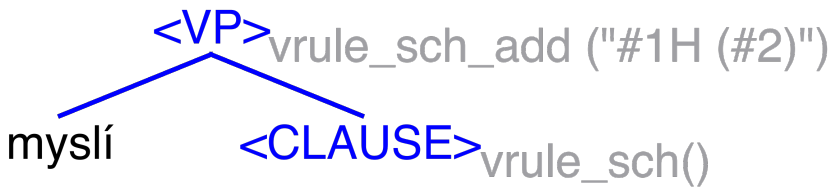


Set modifications

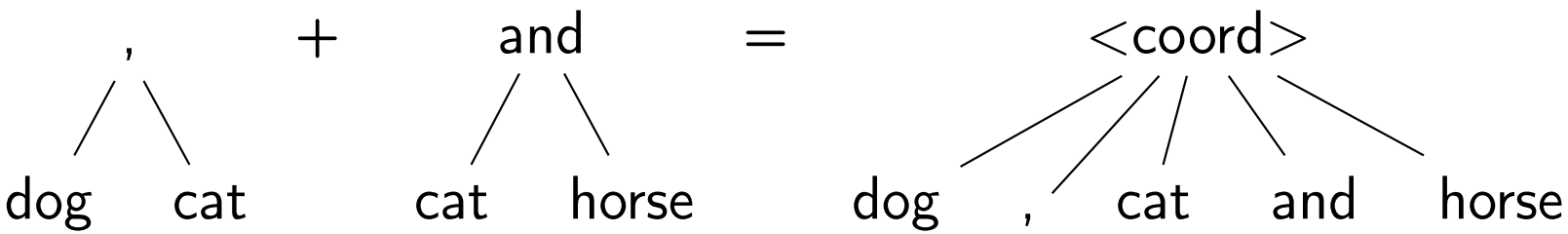
- Action LABELDEP
- Structured Clauses and Coordinations
- Action LABELTOP
- Coordination Nodes Morphology

Action LABELDEP

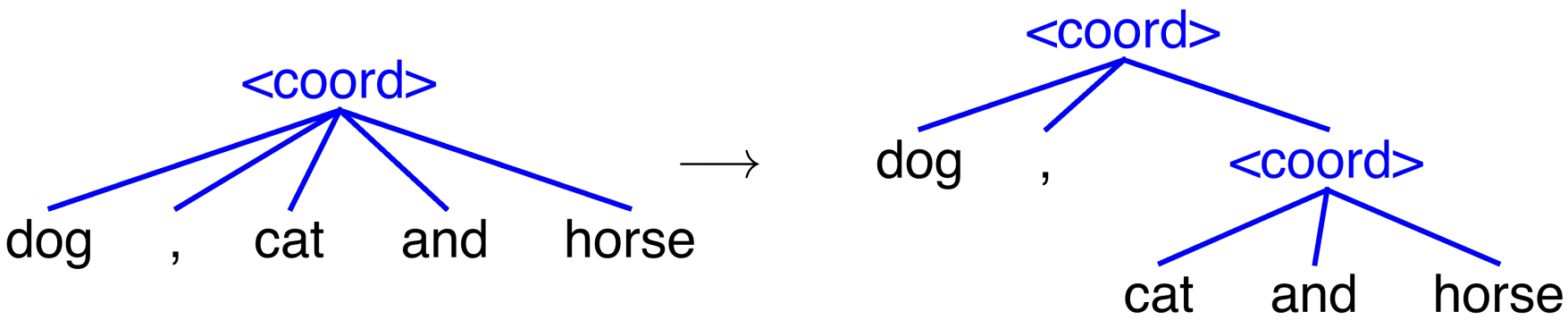
```
TMPL: verbfin ... comma $CONJ ... verbfin ... rbound
MARK 2 5 7 <clause> HEAD 3 DEP 0 PROB 50
LABEL vrule_sch()
LABELDEP vrule_sch_add("#1H 2H")
$CONJ(tag): k3.*yR k3.*xQ k8.*xS
```



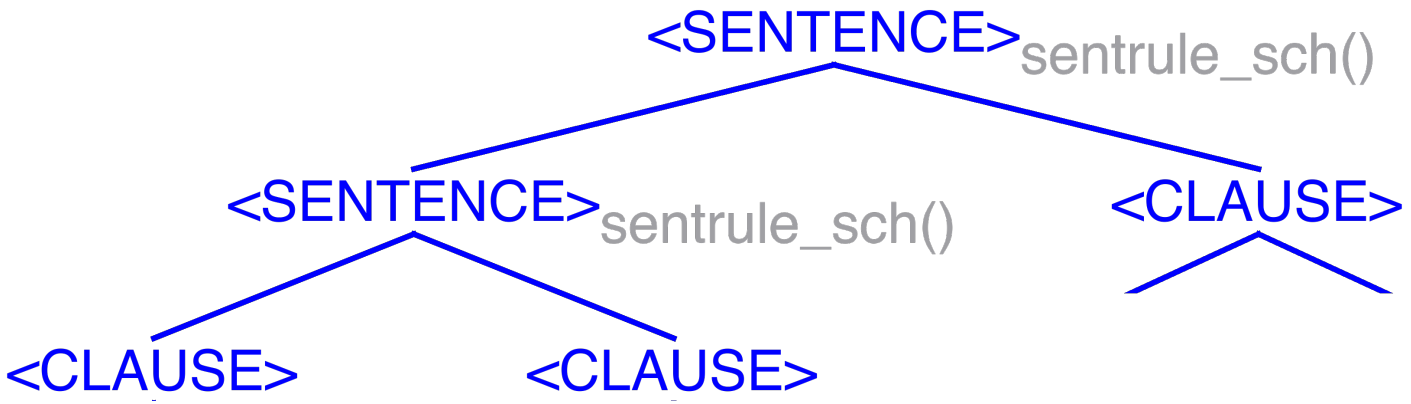
Structured Clauses and Coordinations



Structured Clauses and Coordinations



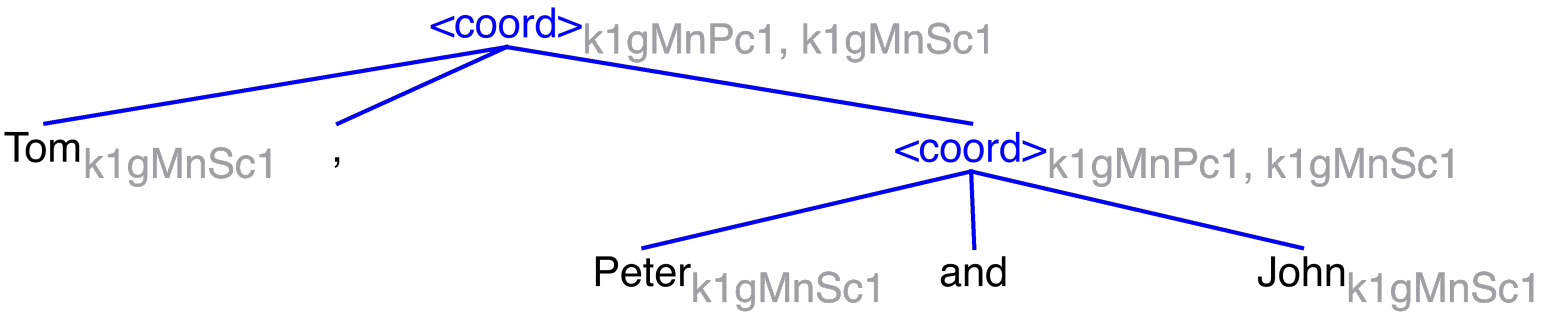
Action LABELTOP



Coordination Nodes Morphology

[A skier or a climber]$_{plural}$ are risking their lives.

Each club member is [a skier or a climber.]$_{singular}$



AST modifications

- lexical item types
- verb valency lexicon
- prepositional types
- sentence schema lexicon

Lexical item types

`stop`

`/k5/otriv` $((o(oo_{\tau\omega})(oo_{\tau\omega}))\omega)$

`/k5/otriv`

$(((o(oo_{\tau\omega})(oo_{\tau\omega}))\omega)\iota)$

Verb Valency Lexicon

```
stop
hPTc4 :exists:V(v):V(v):and:V(v)=[[#0,try(#1)],V(w)]
hPTc2r{at} :exists:V(v):V(v):and:V(v) subset [#0,V(w)] and [#1,V(v)]
```

Prepositional Type Lexicon

```
of
0 hP hH
in
0 hL hW
to
0 hP hL hA hW
for
0 hP hW hH
with
0 hT hP
on
0 hT hL hA hW
```

Sentence Schema Lexicon

```
("when";","): "lwt(tense_temp(awt(#2),awt(#1)))"
```

Sentence Schema Lexicon

$$
\begin{aligned}
&\lambda w_1 \lambda t_2 [\mathbf{P}_{t_2}, \\
&\quad [\mathbf{Onc}_{w_1}, \lambda w_3 \lambda t_4 (\exists x_5)(\exists i_6)(\exists i_7)(\\
&\qquad [\mathbf{Does}_{w_3 t_4}, \\
&\qquad\quad i_7, \\
&\qquad\quad [\mathbf{Perf}_{w_3}, x_5] \\
&\qquad] \\
&\qquad \wedge\ [\mathbf{Ostravsko}_{w_3 t_4}, i_6] \\
&\qquad \wedge\ x_5 = [\mathbf{navštívit}, i_6]_{w_3} \\
&\qquad \wedge\ [\\
&\qquad\quad [\mathbf{Of}, \\
&\qquad\qquad \mathbf{ministr}, \\
&\qquad\qquad \lambda w_8 \lambda t_9 \lambda x_{10} (\\
&\qquad\qquad\quad [\mathbf{doprava}_{w_8 t_9}, x_{10}] \\
&\qquad\qquad\qquad \wedge\ [\mathbf{ČR}_{w_8 t_9}, x_{10}] \\
&\qquad\qquad) \\
&\qquad\quad]_{w_3 t_4}, \\
&\qquad\quad i_7 \\
&\qquad]) \\
&\quad], \mathbf{Anytime} \\
&] \ldots \pi
\end{aligned}
$$

Conclusion

- latest developments of the pipeline used for logical analysis
- changes aimed at promoting multilingual setup
- English used as the first tested transfer language

Future work

- test the setup with more languages
- wider scale comparisons of logical structure sharing between different language environments