How to Analyze Natural Language with Transparent Intensional Logic?

Vojtěch Kovář, Aleš Horák, Miloš Jakubíček

Faculty of Informatics, Masaryk University Brno
Botanická 68a, 602 00 Brno, Czech Republic
{xkovar3, hales, xjakub}@fi.muni.cz

RASLAN 2010

Outline

Motivation

- Goal: Translating natural language into TIL formulas
  - base for complex semantic processing
  - $\longrightarrow$ automatic inference
  - $\longrightarrow$ better representation of language for computers
- Transparent Intensional Logic (TIL)
  - higher-order logical system
  - extended type hierarchy
  - represents NL expressions as constructions
  - expressive enough for all NL nuances

Related work

- English
  - translating into predicate logic (Butler and Miyao, 2010)
  - not expressive enough
  - TIL?
- Czech
  - tectogrammatical layer of language description (ÚFAL MFF UK)
  - not really a logical formalism

Procedure

- Automatic generating constructions from a Czech parser
  - according to Normal Translation Algorithm (Horák)
  - syntactic analysis → syntactic trees → TIL constructions
  - so far aiming at simple sentences
- Checking automatic constructions
  - by human experts
  - providing feedback to parser developers
- Outputs
  - low-cost corpus of TIL formula
  - improving automatic analysis

Current Work

- Extending the lexicon of til types and schemata
  - the prototype implementation is just for few example verbs
  - we need to increase corpus coverage
  - $\longrightarrow$ exploiting Verbalex verb valency lexicon for Czech
  - $\longrightarrow$ planned usage of WordNet for obtaining NP types
  - so far handling NPs is simplified
- User interface for checking constructions
  - as a web application (`corpora.fi.muni.cz/til`)
  - user can accept/reject the particular construction
  - keeping logs using the `git` versioning system

1 vzdát se$_1$, vzdávat se$_1$, zříci se$_3$, zříkat se$_3$ ≈

-frame: **AG**<person:1>$^{\text{obl}}_{\text{kdo1}}$ **VERB**$^{\text{obl}}$ **ACT**<role:1>$^{\text{obl}}_{\text{čeho2}}$

-example: *zřekl se svého úřadu (pf)*

-synonym: abdikovat$_1$, odstoupit$_2$, odstupovat$_2$, složit$_{10}$, skládat$_{10}$, vypovědět$_8$, vypovídat$_8$

-use: prim

-reflexivity: refl (vzdát se, vzdávat se); tantum (zříci se, zříkat se)

2 skládat$_{10}$, složit$_{10}$, vypovídat$_8$, vypovědět$_8$ ≈

-frame: **AG**<person:1>$^{\text{obl}}_{\text{kdo1}}$ **VERB**$^{\text{obl}}$ **ACT**<role:1>$^{\text{obl}}_{\text{co4}}$

-example: *složil funkci (pf)*

-synonym: abdikovat$_1$, odstoupit$_2$, odstupovat$_2$, vzdát se$_1$, vzdávat se$_1$, zříci se$_3$, zříkat se$_3$

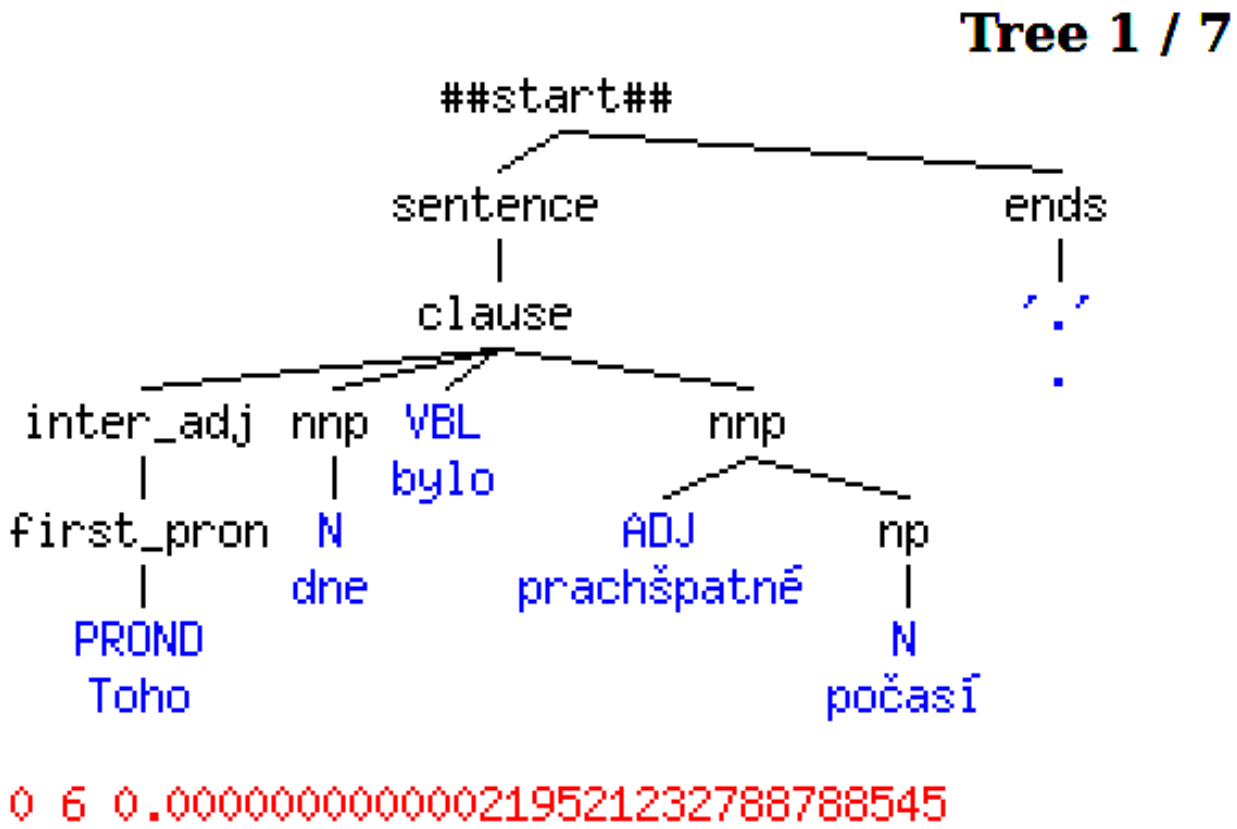
-use: fig (skládat, složit); prim (vypovídat, vypovědět)

-reflexivity: no

```
vzdát
otriv T_VOBJ_I
otriv T_VOBJ

skládat
otriv T_VOBJ_I
otriv T_VOBJ

skládat
hPTc2r\{z\} :exists:V(v):V(v):and:V(v)=[\#0,V(w)]
                          and [[awt(\#1),V(w)],V(v)]
hPTc4r\{na\} :exists:V(v):V(v):and:V(v)=[\#0,V(w)]
                           and [[awt(\#1),V(w)],V(v)]
hPTc3 :exists:V(v):V(v):and:V(v)=[[\#0,try(\#1)],V(w)]
hPTc4 :exists:V(v):V(v):and:V(v)=[[\#0,try(\#2)],V(w)]
```


Tree 1 / 7
##start##
sentence
ends
clause
'.'
.
inter_adj
nnp
VBL
nnp
bylo
first_pron
N
ADJ
np
dne
prachšpatné
PROND
N
Toho
počasí
0 6 0.00000000000021952123278878545

Korpus TIL konstrukcí (neznalek)

vybrat jiný soubor vět

<< předchozí strana | následující strana >>

Zobrazeny konstrukce 1 - 100 / 183

00170.1: $\lambda w_1 \lambda t_2 [\mathbf{P}_{t_2}, [\mathbf{Onc}_{w_1}, \lambda w_3 \lambda t_4 (\exists\, x_5)(\exists\, i_6)(\exists\, i_7)([\mathbf{Does}_{w_3 t_4}, i_7, [\mathbf{Imp}_{w_3}, x_5]] \wedge [\mathbf{den}_{w_3 t_4}, i_6] \wedge x_5 = [\mathbf{být}, i_6]_{w_3} \wedge [\mathbf{prachšpatný}_{w_3 t_4}, i_7] \wedge [\mathbf{počasí}_{w_3 t_4}, i_7])], \mathbf{Anytime}] \ldots \pi$

? OK X

Toho dne bylo prachšpatné počasí .

00191.1: $\lambda w_1 \lambda t_2 [\mathbf{P}_{t_2}, [\mathbf{Onc}_{w_1}, \lambda w_3 \lambda t_4 (\exists\, x_5)([\mathbf{Does}_{w_3 t_4}, \mathrm{On}, [\mathbf{Imp}_{w_3}, x_5]] \wedge x_5 = \mathbf{myslit}_{w_3})], \mathbf{Anytime}] \ldots \pi$

? OK X

myslil si .

00279.1: $\lambda w_1 \lambda t_2 [\mathbf{Prog}_{w_1 t_2}, \lambda w_3 \lambda t_4 (\exists\, x_5)(\exists\, i_6)([\mathbf{Does}_{w_3 t_4}, \mathrm{Já}, [\mathbf{Imp}_{w_3}, x_5]] \wedge [\mathbf{kouzelný}_{w_3 t_4}, i_6] \wedge [\mathbf{hůlka}_{w_3 t_4}, i_6] \wedge x_5 = [\mathbf{mít}, i_6]_{w_3})] \ldots \pi$

? OK X

I kouzelnou hůlku mám .

00302.1: $\lambda w_1 \lambda t_2 [\mathbf{P}_{t_2}, [\mathbf{Onc}_{w_1}, \lambda w_3 \lambda t_4 (\exists\, x_5)([\mathbf{Does}_{w_3 t_4}, \mathrm{Ono}, [\mathbf{Perf}_{w_3}, x_5]] \wedge x_5 = \mathbf{stát}_{w_3})], \mathbf{Anytime}] \ldots \pi$

? OK X

To se stalo !

Conslusions

- First steps in translation NL sentences into TIL formulas
  - we get good data in a cheap way
  - feedback from human annotators
  - $\longrightarrow$ improving the analysers
- Future work
  - improving the lexicons
  - improving the parser(s) with regard to the human experts feedback