Linguistic Logical Analysis of Direct Speech

Aleš Horák, Miloš Jakubíček, Vojtěch Kovář

Centrum zpracování přirozeného jazyka
Fakulta Informatiky, Masarykova Univerzita
Botanická 68a, 602 00 Brno

{hales, jak, xkovar3}@fi.muni.cz

RASLAN 2012
December 7, 2012

Outline

Normal Translation Algorithm in TIL

- specification of translation from natural language sentences to constructions of Transparent Intensional Logic
- logical analysis based on syntactic rules
- describes analysis of all main linguistic phenomena
- the TIL project:
  - 2010 – analysis of simple sentences in past, present and future tense containing selected verbs
  - 2011 – analysis of relative time-related subordinate sentences
  - 2012 – analysis of complex sentences with temporal events including direct speech

Direct and Indirect Discourse

- Direct speech:
  Waiter said: "Are you ready to order, sir?"
  Mr Smith replied: "Yes. I'll have the beef stew for starters and my wife would like tomato soup."
- Indirect speech
  The waiter asked, whether Mr Smith is ready to order.
  He replied, that he would have the beef stew for starters and his wife would like tomato soup.

Synt

- chart parser with a PCFG backbone
- in-programmed actions (per rule defined) that increase grammar expressivity (checking e. g. grammatical agreement) on top of a PCFG
- small metagrammar that auto-generates a large full grammar
- grammar rules separated into priority levels
- phrase-structure trees, dependency graph or selected syntactic structures as output

Syntactic analysis of Direct Speech

```
clause → clause ':'  direct_speech
clause → direct_speech clause
clause → direct_speech clause ',' direct_speech

direct_speech → '"' sentence '"'
9:direct_speech → '"' non_sentence '"'
non_sentence → /[^"]+/
```

Logical analysis in Synt

- works on top of the parsing module
- applied on each parse tree selectively
- implemented in the form of fast TIL actions

Direct Speech Kinds

- Peter said: “Hand me the book.”
- Peter asked: “Hand me the ...”
- Peter thought: “The unicorn!”
- Peter screamed: “Aaaargh!”

Logical analysis of Direct Speech

Peter said: "Hand me the book."

$$\lambda w_1 \lambda t_2 \Big[\mathbf{P}_{t_2}, \Big[\mathbf{Onc}_{w_1}, \lambda w_3 \lambda t_4 (\exists x_5)(\exists c_6)(\exists i_7) \Big(\Big[\mathbf{Does}_{w_3 t_4}, i_7, [\mathbf{Perf}_{w_3}, x_5] \Big] \wedge \wedge [\mathbf{Peter}_{w_3 t_4}, i_7] \wedge x_5 = [\mathbf{say}, c_6]_{w_3} \wedge \wedge c_6 = {}^0 \Big[\lambda w_8 \lambda t_9 (\exists x_{10})(\exists i_{11}) \Big(\Big[\mathbf{Does}_{w_8 t_9}, Ty, [\mathbf{Perf}_{w_8}, x_{10}] \Big] \wedge \wedge x_{10} = [\mathbf{hand\_sb\_st}, Já, i_{11}]_{w_8} \wedge [\mathbf{book}_{w_8 t_9}, i_{11}] \Big) \Big] \Big) \Big], \mathbf{Anytime} \Big] \ldots \pi \tag{1}$$

Peter/$(o\iota)_{\tau\omega}$; say/$((o(o\pi)(o\pi))_{\omega} *_n)$; hand_sb_st/$((o(o\pi)(o\pi))_{\omega}\iota\iota)$;
book/$(o\iota)_{\tau\omega}$;

Sample anaphora resolution in Saara

Type	*Id*	*Word/Phrase*	*Reference*
sentence	sent1	Peter said: “Hand me the book.”	
clause	m1	Peter said	
np	m2	Peter	
clause	m3	_ Hand me the book	
pron_pers_zero	m_zerosubj1	_	
pron_pers_strong	m4	me	m2
np	m5	book	

Sample anaphora resolution in Saara

Type	*Id*	*Word/Phrase*	*Reference*
sentence	sent1	Peter said: "Hand me the book."	
clause	m1	Peter said	
np	m2	Peter	
clause	m3	_ Hand me the book	
pron_pers_zero	m_zerosubj1	_	
pron_pers_strong	m4	me	m2
np	m5	book	

$$Já = i_7$$

Logical analysis of non-sentences

Peter screamed: "Aaaargh!"

$$\lambda w_1 \lambda t_2 \Big[\mathbf{P}_{t_2}, \Big[\mathbf{Onc}_{w_1}, \lambda w_3 \lambda t_4 (\exists x_5)(\exists c_6)(\exists i_7) \Big(\Big[\mathbf{Does}_{w_3 t_4}, i_7, [\mathbf{Perf}_{w_3}, x_5] \Big] \wedge \wedge [\mathbf{Peter}_{w_3 t_4}, i_7] \wedge x_5 = [\mathbf{scream}, c_6]_{w_3} \wedge \wedge c_6 = {}^{00}\text{“Aaaargh”} \Big) \Big], \mathbf{Anytime} \Big] \ldots \pi$$

Peter$/(o\iota)_{\tau\omega}$; scream$/((o(o\pi)(o\pi))_{\omega} * n)$; “Aaaargh”$/\iota$

Corpus of Direct Speech

- built from the czTenTen corpus by the following CQL query:

```
<s/> containing
(<s>
[word!="\""]* [k!="k1"] "\"" [word!="\""]+
[k="k5"] [word!="\""]+ "\"" [word!="\""]*
</s>)
```

- 20,000 sentences