

Valency Frames of Homonymous and Polysemous Verbs in Verbalex

Vít Baisa

Natural Language Processing Centre
Masaryk University
Brno

RASLAN 2010, Karlova studánka

- 1 Introduction
- 2 Valency Lexicon Verbalex
- 3 Frames of Polysemous and Homonymous Verbs
- 4 Results
- 5 Future Work
- 6 Conclusion

Introduction

- homonymy vs. polysemy
- local context
- Verbalex

Structure of Verbalex

- synset (6256)
 - atakovat:1, útočit:2, dorážet:3, napadnout:6
- frame (mačkat:1) and slot (19247)
 - AG^{kdo1}_{person:1} + VERB + OBJ^{co4}_{object:1} + (PART^{v čem6}_{hand:1})
- subsynset (19247)
- semantic roles
 - I: ABS, ISUB, AG, KNOW, PAT, VERB, ... (29)
 - II: abstraction:1, person:1, artifact:1, body part:1, ... (10³)
- other constraints
 - prepositional cases, animality, POS, obligation, ...

Frames of Polysemous and Homonymous Verbs

- humans disambiguate with context
- Verbalex describes local contexts of verbs
- is Verbalex suitable for verb sense disambiguation?
- 10425 lemmas
- 4683 homonymous and polysemous lemmas

Processing

- python API for Verbalex structure
- for all senses of homonymous and polysemous verbs we checked all their pairs and their frames
- several criteria for identity of frames

Results

number of all possible identities

$$\sum_{v \in V} \sum_{i, j \in S_v, i < j} \min(|F_v^i|, |F_v^j|)$$

quantity	%	identity criterion
891	4.67	absolute identity
2203	11.5	identity only for obligatory semantic roles
3343	17.5	identity of semantic role names
605	3.17	absolute identity (only prf. or imprf. variants)
190795	10 ³	all possible identities

Examination of Identical Frames

- have a look at identical frames
- we can observe three types of problems
 - too fine-grained distinction of senses
 - invalid verb in subsynset
 - some error in frame

Too fine-grained distinction of senses

akreditovat:1

pověřit někoho zastupováním

velvyslanec akreditoval novináře

vs.

akreditovat:2

udělit akreditaci/pověření/souhlas

akreditoval novináře

$AG_{\text{person}:1}^{\text{kdo1}} + \text{VERB} + PAT_{\text{person}:1}^{\text{koho4}}$

Invalid verb in a subsynset

chtít:4, přát si:3, zamýšlet:1

mít něco v úmyslu a chtít to uskutečnit

architekt zamýšlí výstavbu nové haly

vs.

chtít:7

silně po něčem toužit; vyjádřit touhu po něčem

chtěl pochvalu

$$AG_{person:1}^{kdo1} + VERB + ACT_{act:2}^{co4}$$

Error in frame

dokázat:1

příkladem, vysvětlením nebo zkouškou doložit platnost nebo pravdivost; názorně předvádět; podat důkaz o něčem, projevit něco skutkem

dokázal, že to platí

vs.

dokázat:2

podařit se někomu něco, být v něčem úspěšný

dokázal, že to zvládne

$$\text{AG}_{\text{person}:1}^{\text{kdo}1} + \text{VERB} + \text{INFO}_{\text{info}:1}^{\text{že}}$$

Future works

- manual annotations of identical frames
- correct errors in Verbalex
- find realisations of all frames (i.e. *proof* of their validity) in corpora

Conclusion

- number of identical frames is very low
- Verbalex has promising potential for verb sense disambiguation
- we can evaluate this potential
- it is at least 98.25 %