

Measuring Coverage of a Valency Lexicon Using Full Syntactic Analysis

Miloš Jakubíček, Vojtěch Kovář, Aleš Horák



NLP Centre, Faculty of Informatics, Masaryk University Brno
Botanická 68a, 602 00 Brno, Czech Republic

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

RASLAN 2009, 5. 12. 2009

Outline

- 1 Verbalex
- 2 Synt
- 3 Extraction of Valencies

Verbalex

- large valency lexicon for Czech
- syntactic annotation
- two-level semantic annotation
- uses the notion of *Complex Valency Frames*
- need to evaluate its coverage and speed up building of the lexicon

Verbalex

- large valency lexicon for Czech
- syntactic annotation
- two-level semantic annotation
- uses the notion of *Complex Valency Frames*
- need to evaluate its coverage and speed up building of the lexicon

Verbalex

- large valency lexicon for Czech
- syntactic annotation
- two-level semantic annotation
- uses the notion of *Complex Valency Frames*
- need to evaluate its coverage and speed up building of the lexicon

Verbalex

- large valency lexicon for Czech
- syntactic annotation
- two-level semantic annotation
- uses the notion of *Complex Valency Frames*
- need to evaluate its coverage and speed up building of the lexicon

Verbalex

- large valency lexicon for Czech
- syntactic annotation
- two-level semantic annotation
- uses the notion of *Complex Valency Frames*
- need to evaluate its coverage and speed up building of the lexicon

Synt

- probabilistic parser for Czech with context-free backbone
- output in the form of phrasal or dependency trees or flat syntactic structures
- key idea: estimate the Verbalex coverage by employing the extraction of structures for automatic retrieval of verb valencies

Synt

- probabilistic parser for Czech with context-free backbone
- output in the form of phrasal or dependency trees or flat syntactic structures
- key idea: estimate the Verbalex coverage by employing the extraction of structures for automatic retrieval of verb valencies

Synt

- probabilistic parser for Czech with context-free backbone
- output in the form of phrasal or dependency trees or flat syntactic structures
- key idea: estimate the Verbalex coverage by employing the extraction of structures for automatic retrieval of verb valencies

How?

- parse sentences in the DESAM corpus
- divide sentences into clauses and extract noun or prepositional phrases and other structures covered by the BRIEF format
- example:
 - ; extracted from sentence: Nenadálou finanční krizi musela podnikatelka řešit jiným způsobem .
 - řešit <v>hTc4,hTc7
 - (*The businessman had to **solve** the sudden financial crisis in another way.*)
 - An accusative and instrumental valency has been found.
- main problem: differentiation between obligatory valencies (arguments) and additional non-obligatory modifiers (adjuncts)

How?

- parse sentences in the DESAM corpus
- divide sentences into clauses and extract noun or prepositional phrases and other structures covered by the BRIEF format

- example:

; extracted from sentence: Nenadálou finanční krizi musela podnikatelka řešit jiným způsobem .
řešit <v>hTc4,hTc7

*(The businessman had to **solve** the sudden financial crisis in another way.)*

An accusative and instrumental valency has been found.

- main problem: differentiation between obligatory valencies (arguments) and additional non-obligatory modifiers (adjuncts)

How?

- parse sentences in the DESAM corpus
- divide sentences into clauses and extract noun or prepositional phrases and other structures covered by the BRIEF format
- example:
 - ; extracted from sentence: Nenadálou finanční krizi musela podnikatelka řešit jiným způsobem .
 - řešit <v>hTc4,hTc7
 - (*The businessman had to **solve** the sudden financial crisis in another way.*)
 - An accusative and instrumental valency has been found.
- main problem: differentiation between obligatory valencies (arguments) and additional non-obligatory modifiers (adjuncts)

How?

- parse sentences in the DESAM corpus
- divide sentences into clauses and extract noun or prepositional phrases and other structures covered by the BRIEF format
- example:
 - ; extracted from sentence: Nenadálou finanční krizi musela podnikatelka řešit jiným způsobem .
 - řešit <v>hTc4,hTc7
 - (*The businessman had to **solve** the sudden financial crisis in another way.*)
 - An accusative and instrumental valency has been found.
- main problem: differentiation between obligatory valencies (arguments) and additional non-obligatory modifiers (adjuncts)

Results I

indicator	covered	total	%
verb coverage	2,957	3,685	80.24
valency coverage	5,348	9,430	56.71
valency coverage with consideration of error analysis	5,348	6,397	83.60

Table: Coverage of the Verbalex valency lexicon on the annotated DESAM corpus

Results II

indicator	number of missing valencies	%
noun phrases (<i>case only valencies</i>)	499	11.00
prepositional phrases (<i>preposition+case valencies</i>)	3,142	76.97
other (<i>subordinated clauses, infinitives etc.</i>)	491	12.03
total	4,082	100

Table: Error analysis of missing valencies.