Parsing System with Contextual Constraints

Vladimír Kadlec

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

xkadlec@fi.muni.cz

RASLAN 2007

Outline

Motivation

- Why not to use CF rules only
  - free word order – combinatoric expansion
- Concepts
  - CF rules generated by rules with combinatoric constructs
  - and supplemented by contextual constraints
- **Requires effective analyser able to solve constraints**

Motivation

- Why not to use CF rules only
  - free word order – combinatoric expansion
- Concepts
  - CF rules generated by rules with combinatoric constructs
  - and supplemented by contextual constraints
- **Requires effective analyser able to solve constraints**

Motivation

- Why not to use CF rules only
  - free word order – combinatoric expansion
- Concepts
  - CF rules generated by rules with combinatoric constructs
  - and supplemented by contextual constraints
- **Requires effective analyser able to solve constraints**

Motivation

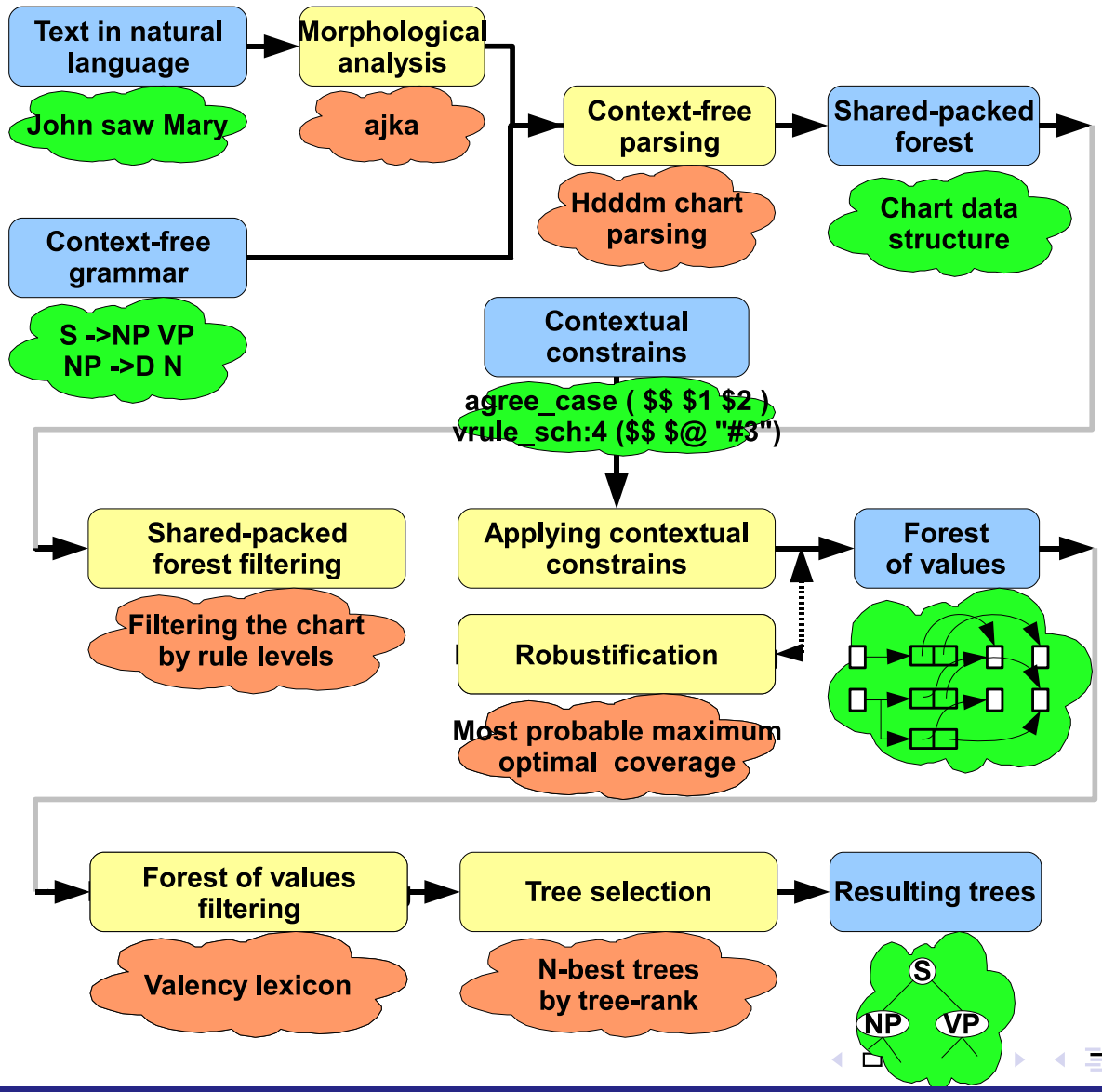
- Why not to use CF rules only
  - free word order – combinatoric expansion
- Concepts
  - CF rules generated by rules with combinatoric constructs
  - and supplemented by contextual constraints
- **Requires effective analyser able to solve constraints**

Motivation

- Why not to use CF rules only
  - free word order – combinatoric expansion
- Concepts
  - CF rules generated by rules with combinatoric constructs
  - and supplemented by contextual constraints
- Requires effective analyser able to solve constraints

Motivation

- Why not to use CF rules only
  - free word order – combinatoric expansion
- Concepts
  - CF rules generated by rules with combinatoric constructs
  - and supplemented by contextual constraints
- **Requires effective analyser able to solve constraints**


Text in natural language
John saw Mary
Morphological analysis
ajka
Context-free grammar
S ->NP VP
NP ->D N
Context-free parsing
Hdddm chart parsing
Shared-packed forest
Chart data structure
Contextual constrains
agree_case ($$ $1 $2)
vrule_sch:4 ($$ $@ "#3")
Shared-packed forest filtering
Filtering the chart by rule levels
Applying contextual constrains
Robustification
Most probable maximum optimal coverage
Forest of values
Forest of values filtering
Valency lexicon
Tree selection
N-best trees by tree-rank
Resulting trees
S
NP
VP

Head-driven parsing with dependent dot move

- Based on *Satta and Stock, 1989*.
- Bottom-up approach.
- Bidirectional parsing.
- Start at the head of the given grammar rule.
- Head positions are crucial for the method.
- Comparable with state of the art algorithms.

Head-driven parsing with dependent dot move

- Based on *Satta and Stock, 1989*.
- Bottom-up approach.
- Bidirectional parsing.
- Start at the head of the given grammar rule.
- Head positions are crucial for the method.
- Comparable with state of the art algorithms.

Head-driven parsing with dependent dot move

- Based on *Satta and Stock, 1989*.
- Bottom-up approach.
- Bidirectional parsing.
- Start at the head of the given grammar rule.
- Head positions are crucial for the method.
- Comparable with state of the art algorithms.

Head-driven parsing with dependent dot move

- Based on *Satta and Stock, 1989*.
- Bottom-up approach.
- Bidirectional parsing.
- Start at the head of the given grammar rule.
- Head positions are crucial for the method.
- Comparable with state of the art algorithms.

Head-driven parsing with dependent dot move

- Based on *Satta and Stock, 1989*.
- Bottom-up approach.
- Bidirectional parsing.
- Start at the head of the given grammar rule.
- Head positions are crucial for the method.
- Comparable with state of the art algorithms.

Head-driven parsing with dependent dot move

- Based on *Satta and Stock, 1989*.
- Bottom-up approach.
- Bidirectional parsing.
- Start at the head of the given grammar rule.
- Head positions are crucial for the method.
- Comparable with state of the art algorithms.

Optimizing Heads for Parsing

- A rule instantiation strategy (*Oepen and Callmeier, 2000*), only binary rules.
- Training heads on the corpus – optimal heads for given input.
- Heuristic methods.
- Three times faster parser (for "free").

Optimizing Heads for Parsing

- A rule instantiation strategy (*Oepen and Callmeier, 2000*), only binary rules.
- Training heads on the corpus – optimal heads for given input.
- Heuristic methods.
- Three times faster parser (for "free").

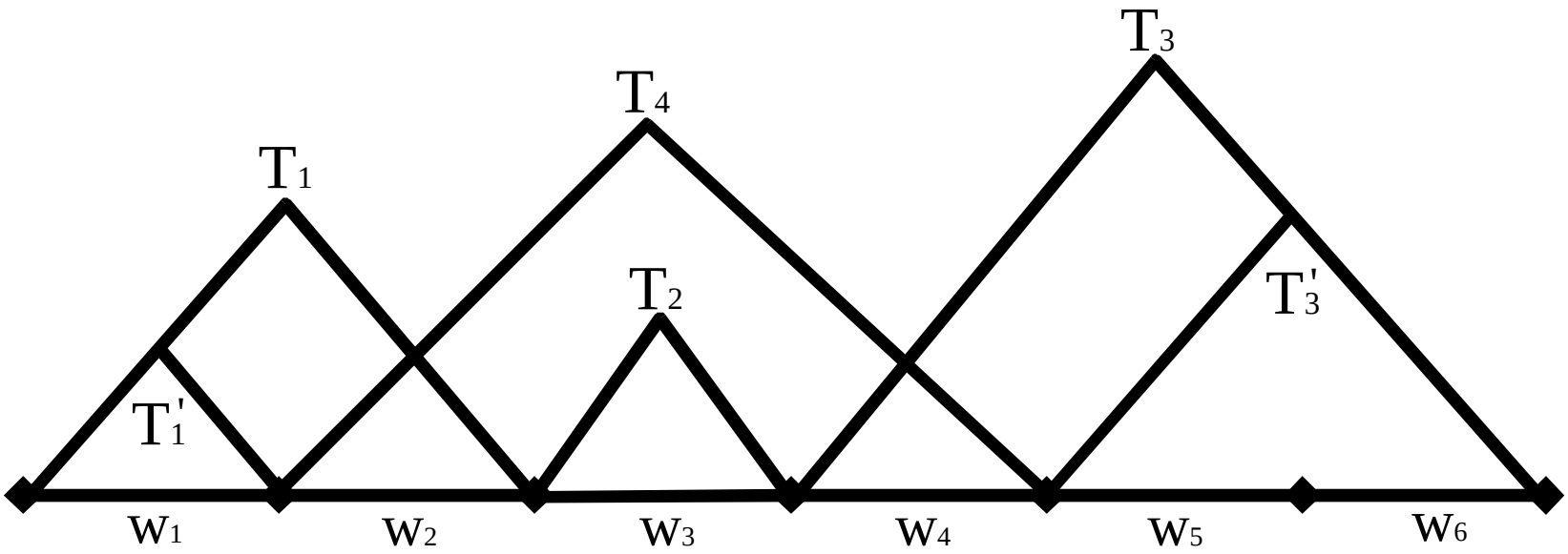
Optimizing Heads for Parsing

- A rule instantiation strategy (*Oepen and Callmeier, 2000*), only binary rules.
- Training heads on the corpus – optimal heads for given input.
- Heuristic methods.
- Three times faster parser (for "free").

Optimizing Heads for Parsing

- A rule instantiation strategy (*Oepen and Callmeier, 2000*), only binary rules.
- Training heads on the corpus – optimal heads for given input.
- Heuristic methods.
- Three times faster parser (for "free").

Example



Coverages $C = (T_1, T_2, T_3)$ and $C' = (T'_1, T_4, T'_3)$.

Robust stochastic parsing using optimal maximum coverage

- Coverage
  - Sequence of non-overlapping, possibly partial, derivation trees.
  - Concatenation of the leaves of these trees corresponds to the whole input sentence.
- Properties
  - A full parse is a coverage.
  - It always exists — a trivial coverage.

Robust stochastic parsing using optimal maximum coverage

- Coverage
  - Sequence of non-overlapping, possibly partial, derivation trees.
  - Concatenation of the leaves of these trees corresponds to the whole input sentence.
- Properties
  - A full parse is a coverage.
  - It always exists — a trivial coverage.

Robust stochastic parsing using optimal maximum coverage

- Coverage
  - Sequence of non-overlapping, possibly partial, derivation trees.
  - Concatenation of the leaves of these trees corresponds to the whole input sentence.
- Properties
  - A full parse is a coverage.
  - It always exists — a trivial coverage.

Robust stochastic parsing using optimal maximum coverage

- Coverage
  - Sequence of non-overlapping, possibly partial, derivation trees.
  - Concatenation of the leaves of these trees corresponds to the whole input sentence.
- Properties
  - A full parse is a coverage.
  - It always exists — a trivial coverage.

Robust stochastic parsing using optimal maximum coverage

- Coverage
  - Sequence of non-overlapping, possibly partial, derivation trees.
  - Concatenation of the leaves of these trees corresponds to the whole input sentence.
- Properties
  - A full parse is a coverage.
  - It always exists — a trivial coverage.

Robust stochastic parsing using optimal maximum coverage

- Coverage
  - Sequence of non-overlapping, possibly partial, derivation trees.
  - Concatenation of the leaves of these trees corresponds to the whole input sentence.
- Properties
  - A full parse is a coverage.
  - It always exists — a trivial coverage.

Contextual Constraints and Semantic Actions

1. Bottom-up evaluation.
2. Limited domain for arguments and results.

Contextual Constraints and Semantic Actions

1. Bottom-up evaluation.
2. Limited domain for arguments and results.

Representation of Values

- NP-complete problem in general *Barton, Berwick and Ristad, 1987*
- Solution – allow only limited number of features.
- Packed-shared forest like representation.
- The limitation is not possible in all cases.

Representation of Values

- NP-complete problem in general *Barton, Berwick and Ristad, 1987*
- Solution – allow only limited number of features.
- Packed-shared forest like representation.
- The limitation is not possible in all cases.

Representation of Values

- NP-complete problem in general *Barton, Berwick and Ristad, 1987*
- Solution – allow only limited number of features.
- Packed-shared forest like representation.
- The limitation is not possible in all cases.

Representation of Values

- NP-complete problem in general *Barton, Berwick and Ristad, 1987*
- Solution – allow only limited number of features.
- Packed-shared forest like representation.
- The limitation is not possible in all cases.

synt Project

- The target language is Czech,
- Manually created meta-grammar (250 meta-rules),
- CF rules automatically generated (2800 CF rules),
- Semantic actions and contextual constraints,
- Head-driven chart parser + evaluation of the constraitns,
- WWW front-end.

synt Project

- The target language is Czech,
- Manually created meta-grammar (250 meta-rules),
- CF rules automatically generated (2800 CF rules),
- Semantic actions and contextual constraints,
- Head-driven chart parser + evaluation of the constraitns,
- WWW front-end.

synt Project

- The target language is Czech,
- Manually created meta-grammar (250 meta-rules),
- CF rules automatically generated (2800 CF rules),
- Semantic actions and contextual constraints,
- Head-driven chart parser + evaluation of the constraitns,
- WWW front-end.

synt Project

- The target language is Czech,
- Manually created meta-grammar (250 meta-rules),
- CF rules automatically generated (2800 CF rules),
- Semantic actions and contextual constraints,
- Head-driven chart parser + evaluation of the constraitns,
- WWW front-end.

synt Project

- The target language is Czech,
- Manually created meta-grammar (250 meta-rules),
- CF rules automatically generated (2800 CF rules),
- Semantic actions and contextual constraints,
- Head-driven chart parser + evaluation of the constraitns,
- WWW front-end.

synt Project

- The target language is Czech,
- Manually created meta-grammar (250 meta-rules),
- CF rules automatically generated (2800 CF rules),
- Semantic actions and contextual constraints,
- Head-driven chart parser + evaluation of the constraitns,
- WWW front-end.

Conclusions and Future Directions

Conclusions

- Language independent parsing system.
- Fast head-driven CF algorithm.
- Effective evaluation of semantic actions and contextual constraints.
- Robust extension.

Current Development

- Create dependecies directly by semantic actions.
- Checking with verb valency lexicon VerbaLex.

Conclusions and Future Directions

Conclusions

- Language independent parsing system.
- Fast head-driven CF algorithm.
- Effective evaluation of semantic actions and contextual constraints.
- Robust extension.

Current Development

- Create dependecies directly by semantic actions.
- Checking with verb valency lexicon VerbaLex.

Conclusions and Future Directions

Conclusions

- Language independent parsing system.
- Fast head-driven CF algorithm.
- Effective evaluation of semantic actions and contextual constraints.
- Robust extension.

Current Development

- Create dependecies directly by semantic actions.
- Checking with verb valency lexicon VerbaLex.

Conclusions and Future Directions

Conclusions

- Language independent parsing system.
- Fast head-driven CF algorithm.
- Effective evaluation of semantic actions and contextual constraints.
- Robust extension.

Current Development

- Create dependecies directly by semantic actions.
- Checking with verb valency lexicon VerbaLex.

Conclusions and Future Directions

Conclusions

- Language independent parsing system.
- Fast head-driven CF algorithm.
- Effective evaluation of semantic actions and contextual constraints.
- Robust extension.

Current Development

- Create dependecies directly by semantic actions.
- Checking with verb valency lexicon VerbaLex.

Conclusions and Future Directions

Conclusions

- Language independent parsing system.
- Fast head-driven CF algorithm.
- Effective evaluation of semantic actions and contextual constraints.
- Robust extension.

Current Development

- Create dependecies directly by semantic actions.
- Checking with verb valency lexicon VerbaLex.