



Language Independent Probabilistic Context-Free Parsing Bolstered by Machine Learning

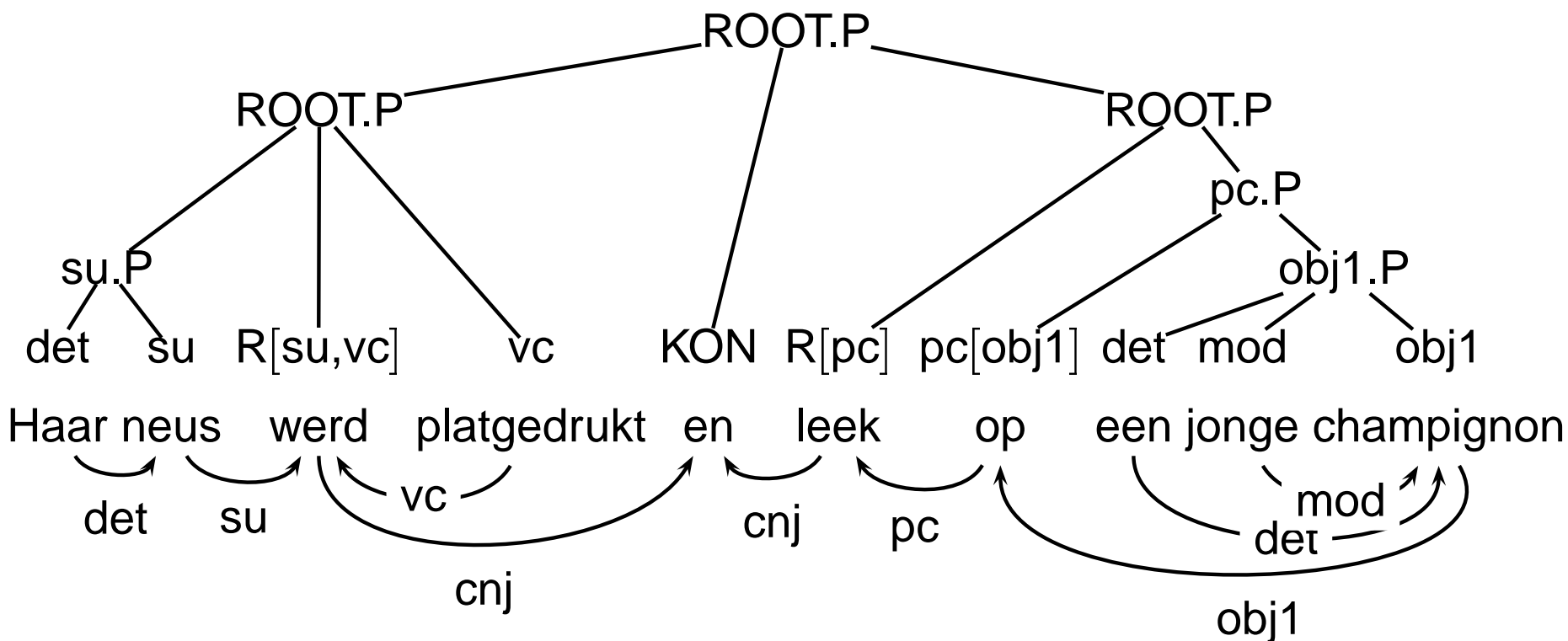
Michael Schiehlen & Kristina Spranger

Institut für Maschinelle Sprachverarbeitung
Universität Stuttgart

Outline

- General approach: convert dependency structure to constituency structure and use plain PCFG
- insert information on subcategorisation into the grammar (automatically from dependency relations)
- which names for phrasal categories?

From Dep. to Const. Structure



Improvements after Submission

- Markovization of PCFG rules (minor improvements)
- language-dependent manual determination of phrasal categories for Chinese, Czech, German, Slovene, Spanish (major improvements)

Tagging Approach

- Dependency Parsing as Tagging: use MaxEnt-tagger to assign head–relation pairs to individual tokens
- heads in ‘nth-tag’ representation, e.g.
 - $< NN$ for the last token with POS tag NN
 - $>> NN$ for the second NN to the right
- Combination of PCFG-Parsing and Tagging: use parser output as an additional feature

Performance of Combination

| AR | CH | CZ | DA | DU | GE | JA | PO | SL | SP | SW | TU |
|----|----|----|----|----|----|----|----|----|----|----|----|
| | | | | | | | | | | | |
| | | | | ● | | | | | | | |
| | | ● | | | | | | ● | | | |
| | | | | ○ | ● | | | | | | |
| | | | | | | ● | | | | | |
| ● | ● | | ● | | | | ● | | | | |
| | | | | | | | | | ● | ● | |
| | | | ○ | | | | | | | | ● |
| | | | | | ○ | ○ | | | | ○ | ○ |
| | ○ | ○ | | | | | ○ | | | | |
| | | | | | | | | | | | |
| ○ | | | | | | | | ○ | ○ | | |