

# Maschinelle Analyse und Übersetzung natürlicher Sprachen

Lehrstuhl Grundlagen der Programmierung

Forschungslinie – Einführung in die Forschung  
SS 2014

NLP = natural language processing

## Forschung:

- ▶ Maschinelle Analyse und Übersetzung natürlicher Sprachen (syntax-based machine translation)
  
  
  
  
  
  
  
  
  
  
- ▶ Weighted Tree Automata and Logics

## Forschung:

- ▶ Maschinelle Analyse und Übersetzung natürlicher Sprachen (syntax-based machine translation)
  - ▶ theoretisch und praktisch
  
- ▶ Weighted Tree Automata and Logics

## Forschung:

- ▶ Maschinelle Analyse und Übersetzung natürlicher Sprachen (syntax-based machine translation)
  - ▶ theoretisch und praktisch
  - ▶ DFG-Projekt beantragt
  
- ▶ Weighted Tree Automata and Logics

## Forschung:

- ▶ Maschinelle Analyse und Übersetzung natürlicher Sprachen (syntax-based machine translation)
  - ▶ theoretisch und praktisch
  - ▶ DFG-Projekt beantragt
  - ▶ praktisches System VANDA in Haskell, SHK-Stellen
  
- ▶ Weighted Tree Automata and Logics

## Forschung:

- ▶ Maschinelle Analyse und Übersetzung natürlicher Sprachen (syntax-based machine translation)
  - ▶ theoretisch und praktisch
  - ▶ DFG-Projekt beantragt
  - ▶ praktisches System VANDA in Haskell, SHK-Stellen
  - ▶ Kooperation mit St Andrews, Schottland
  
- ▶ Weighted Tree Automata and Logics

## Forschung:

- ▶ Maschinelle Analyse und Übersetzung natürlicher Sprachen (syntax-based machine translation)
  - ▶ theoretisch und praktisch
  - ▶ DFG-Projekt beantragt
  - ▶ praktisches System VANDA in Haskell, SHK-Stellen
  - ▶ Kooperation mit St Andrews, Schottland
- ▶ Weighted Tree Automata and Logics
  - ▶ theoretisch, mit Motivation aus NLP

## Forschung:

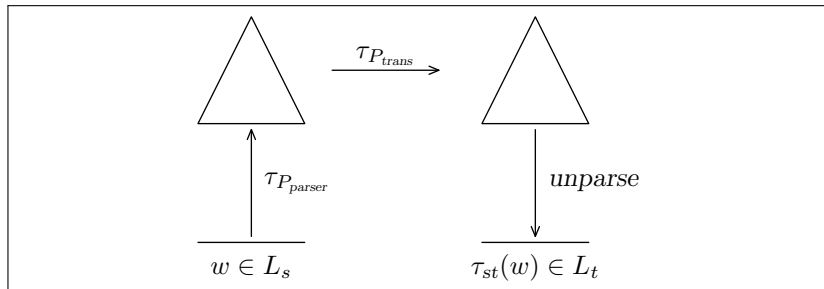
- ▶ Maschinelle Analyse und Übersetzung natürlicher Sprachen (syntax-based machine translation)
  - ▶ theoretisch und praktisch
  - ▶ DFG-Projekt beantragt
  - ▶ praktisches System VANDA in Haskell, SHK-Stellen
  - ▶ Kooperation mit St Andrews, Schottland
- ▶ Weighted Tree Automata and Logics
  - ▶ theoretisch, mit Motivation aus NLP
  - ▶ bilaterale Forschung mit Leipzig und Szeged (Ungarn)



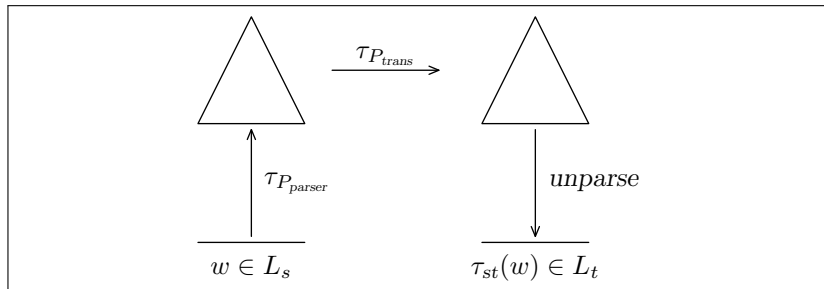
## Forschung:

- ▶ Maschinelle Analyse und Übersetzung natürlicher Sprachen (syntax-based machine translation)
  - ▶ theoretisch und praktisch
  - ▶ DFG-Projekt beantragt
  - ▶ praktisches System VANDA in Haskell, SHK-Stellen
  - ▶ Kooperation mit St Andrews, Schottland
  
- ▶ Weighted Tree Automata and Logics
  - ▶ theoretisch, mit Motivation aus NLP
  - ▶ bilaterale Forschung mit Leipzig und Szeged (Ungarn)
  - ▶ DFG-Graduiertenkolleg "Quantitative Logics and Automata" (Beginn: Okt. 2012), Stipendien

## Syntax-based Machine Translation

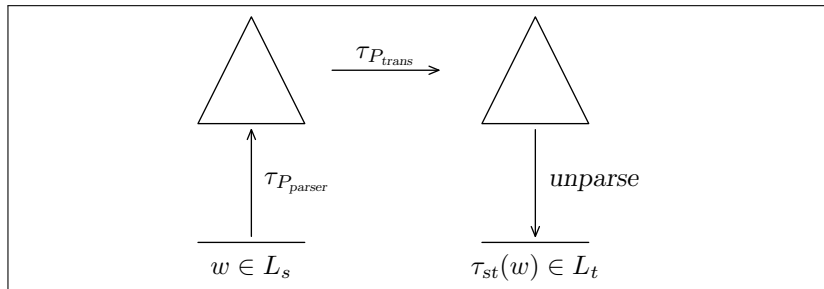


## Syntax-based Machine Translation



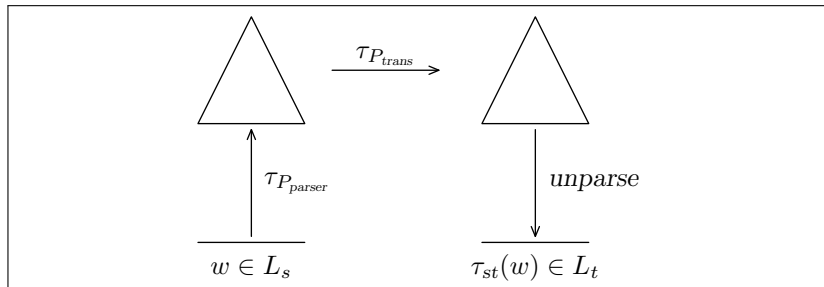
- probabilistic parsing

## Syntax-based Machine Translation



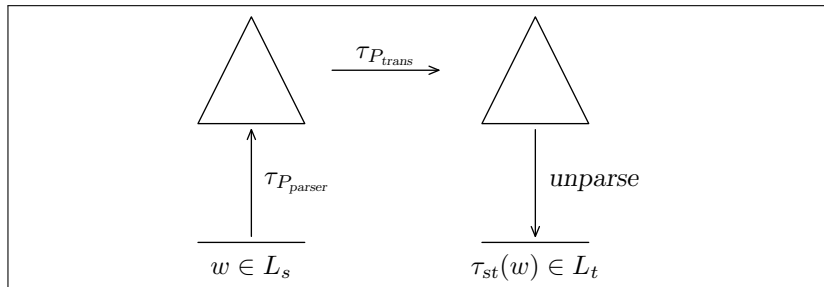
- ▶ probabilistic parsing
- ▶ k-best

## Syntax-based Machine Translation



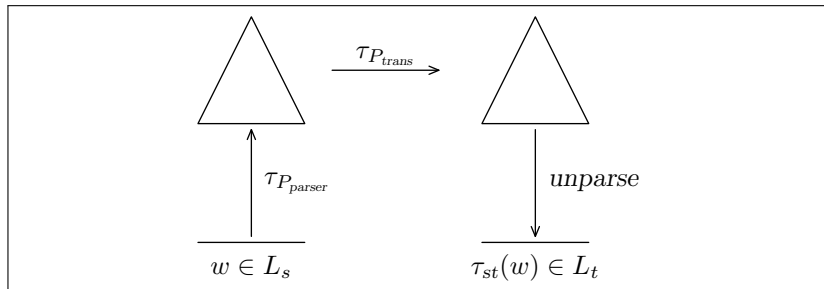
- ▶ probabilistic parsing
- ▶ k-best
- ▶ rule extraction

## Syntax-based Machine Translation



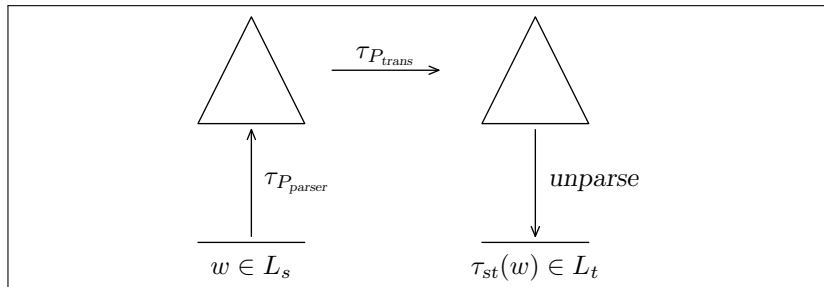
- ▶ probabilistic parsing
- ▶ k-best
- ▶ rule extraction
- ▶ training

## Syntax-based Machine Translation



- ▶ probabilistic parsing
- ▶ k-best
- ▶ rule extraction
- ▶ training
- ▶ preservation of recognizability

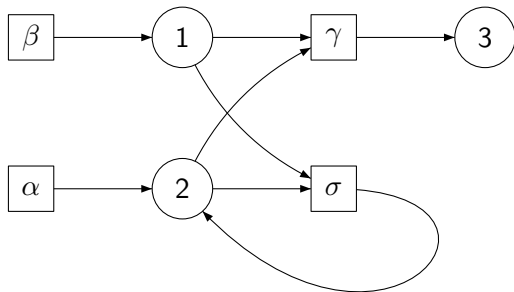
## Syntax-based Machine Translation



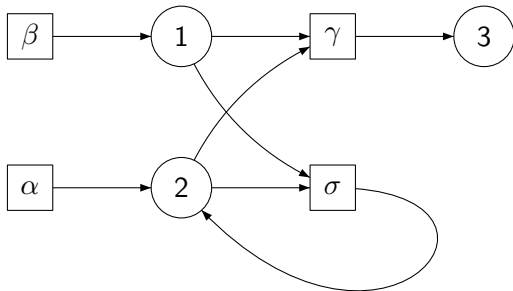
- ▶ probabilistic parsing
- ▶ k-best
- ▶ rule extraction
- ▶ training
- ▶ preservation of recognizability
- ▶ compositionality of the formalism



## Weighted Tree Automata

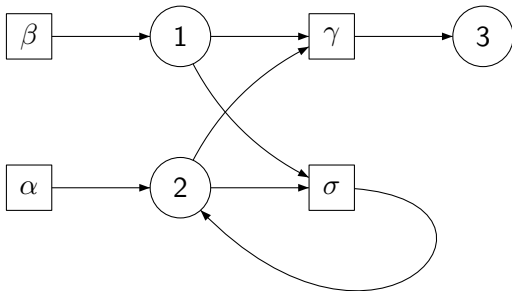


## Weighted Tree Automata



unfoldings (starting at 3):

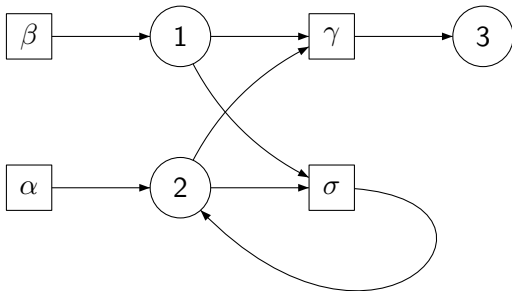
## Weighted Tree Automata



unfoldings (starting at 3):

$\gamma(\beta, \alpha)$

## Weighted Tree Automata

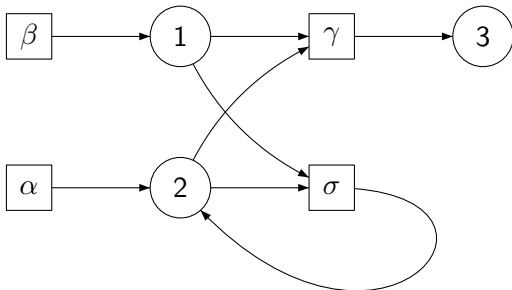


unfoldings (starting at 3):

$\gamma(\beta, \alpha)$

$\gamma(\beta, \sigma(\beta, \alpha))$

## Weighted Tree Automata



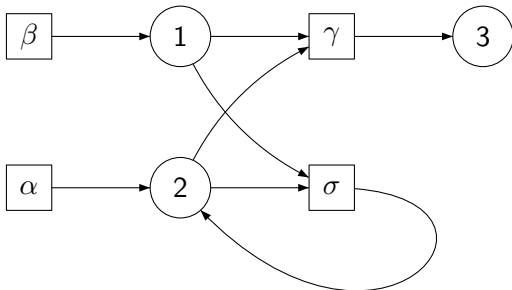
unfoldings (starting at 3):

$\gamma(\beta, \alpha)$

$\gamma(\beta, \sigma(\beta, \alpha))$

$\gamma(\beta, \sigma(\beta, \sigma(\beta, \alpha)))$

## Weighted Tree Automata



unfoldings (starting at 3):

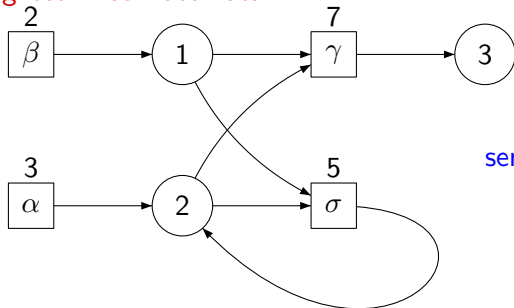
$\gamma(\beta, \alpha)$

$\gamma(\beta, \sigma(\beta, \alpha))$

$\gamma(\beta, \sigma(\beta, \sigma(\beta, \alpha)))$

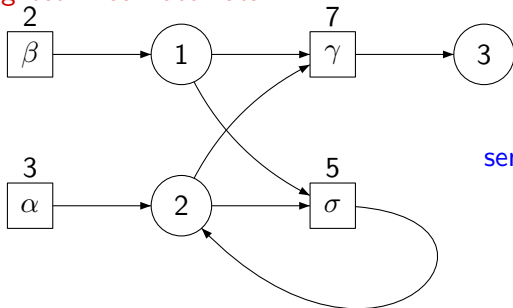
tree language  $\gamma(\beta, [\sigma(\beta, ]^n \alpha [)]^n )$  for every  $n \geq 0$

## Weighted Tree Automata



semiring  $(\mathbb{N}, +, \cdot, 0, 1)$

## Weighted Tree Automata



semiring  $(\mathbb{N}, +, \cdot, 0, 1)$

weighted tree language:

$$\gamma(\beta, [\sigma(\beta, ]^n \alpha [)]^n) \mapsto 7 \cdot 2 \cdot (5 \cdot 2)^n \cdot 3$$