

# Baumübersetzer in der statistischen Übersetzung

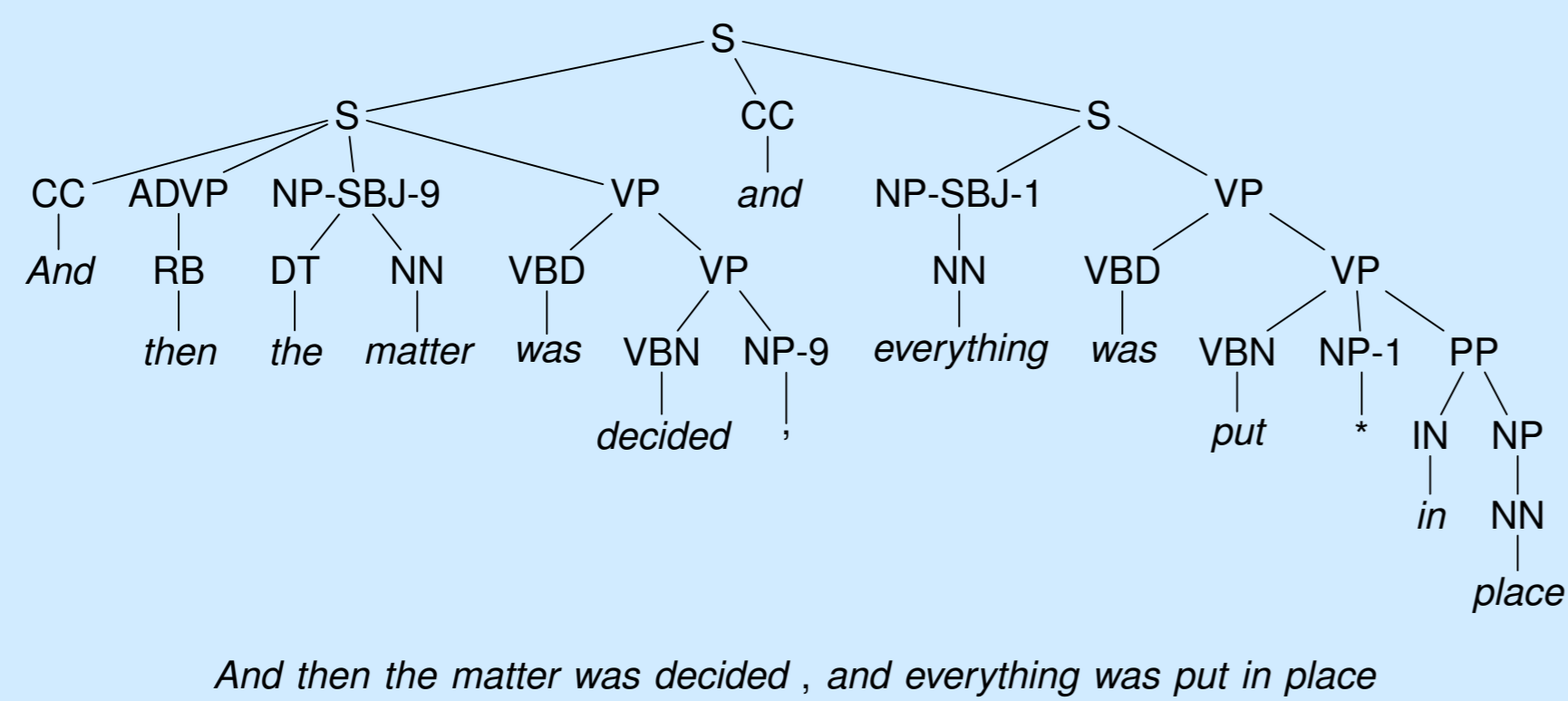
Fabienne Braune, Andreas Maletti, Daniel Quernheim, and Nina Seemann

Institut für maschinelle Sprachverarbeitung, Universität Stuttgart  
 {braunefe,maletti,daniel,seemanna}@ims.uni-stuttgart.de

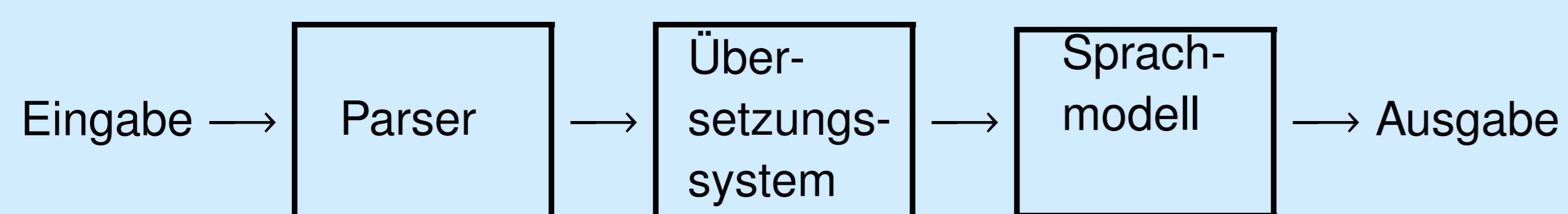
## 1. Motivation

### 1.1 Syntax in der statistischen Übersetzung

- **Syntaxbäume** gute Representation natürlicher Sprachstrukturen



- **Integration syntaktischer Strukturen** in die statistische Übersetzung



- **Synchrone kontextfreie Grammatiken** als Modell:

NP → DT NN / DT NN

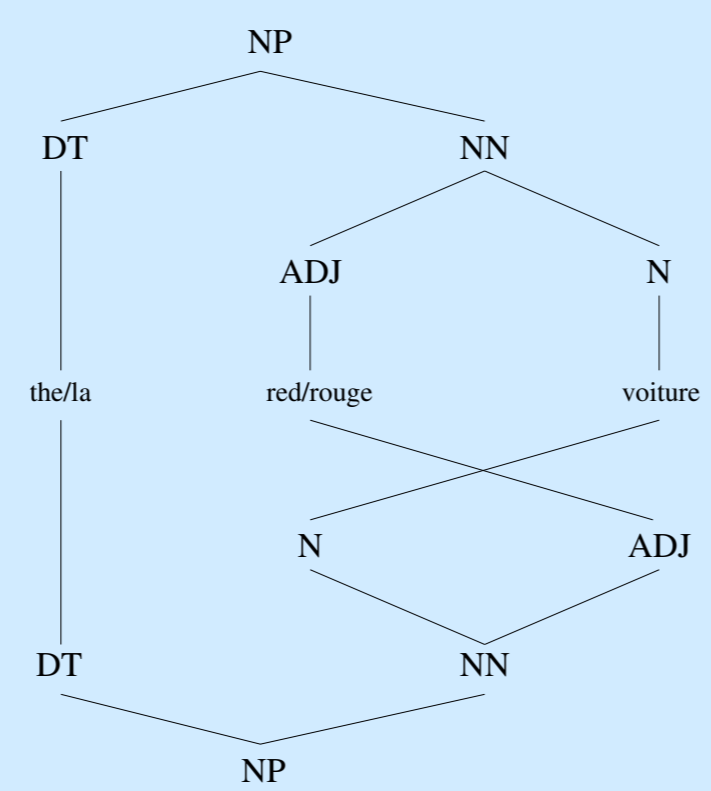
NN → ADJ N / N ADJ

NN → car / voiture

ADJ → red / rouge

DT → the / la

Synchrone Ableitung:

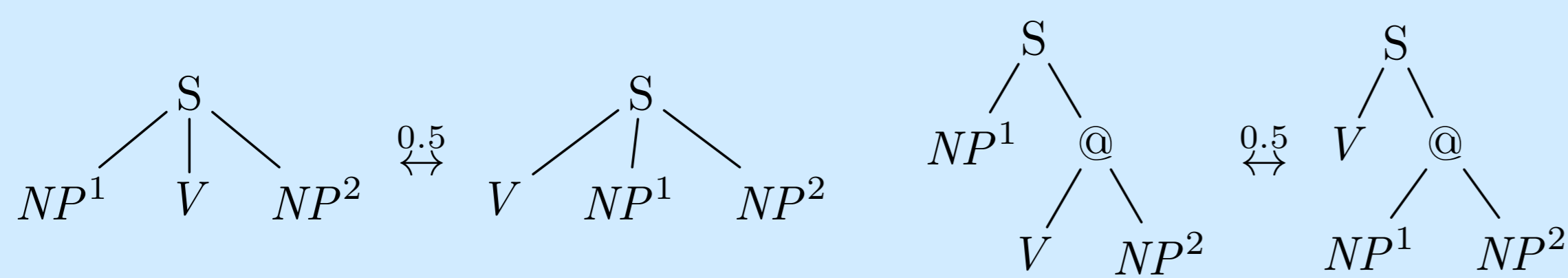


### 1.2 Baumübersetzer in der statistischen Übersetzung

- **Synchrone kontextfreie Grammatiken** (SCFG) häufig zu schwach unten stehende Baumübersetzung als SCFG nicht möglich

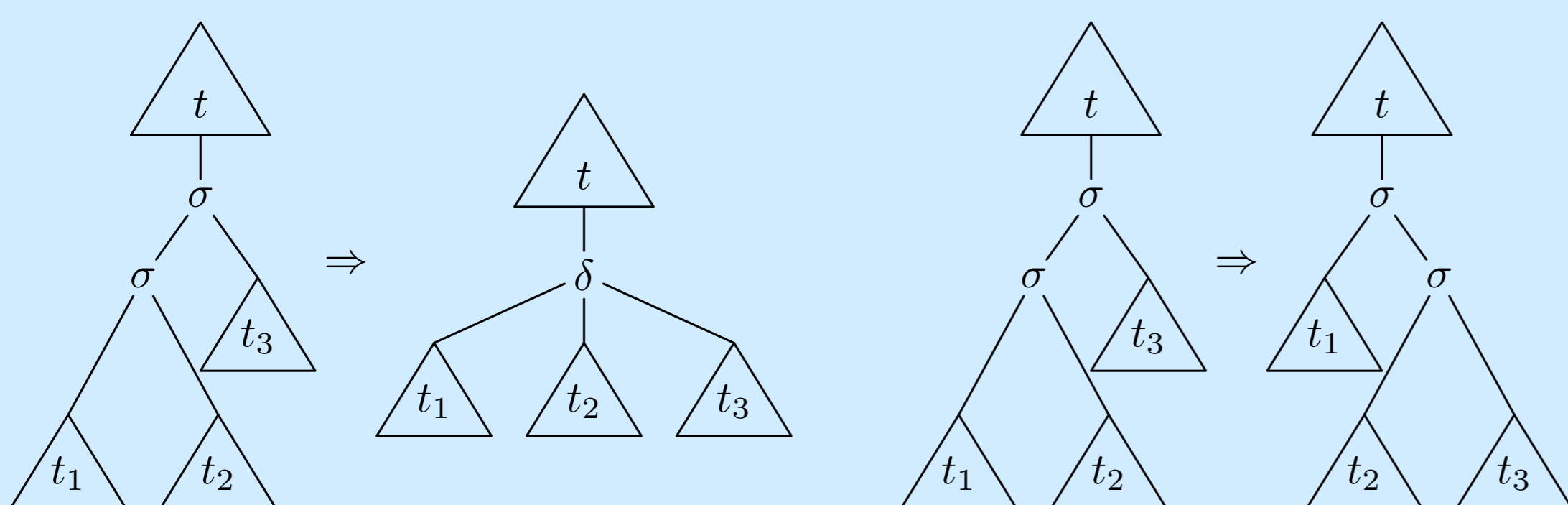
- **Baumübersetzer** (STSG) als Modell für Baumübersetzungen

Beispielregeln:



- **Abstrakte Beispielübersetzung**

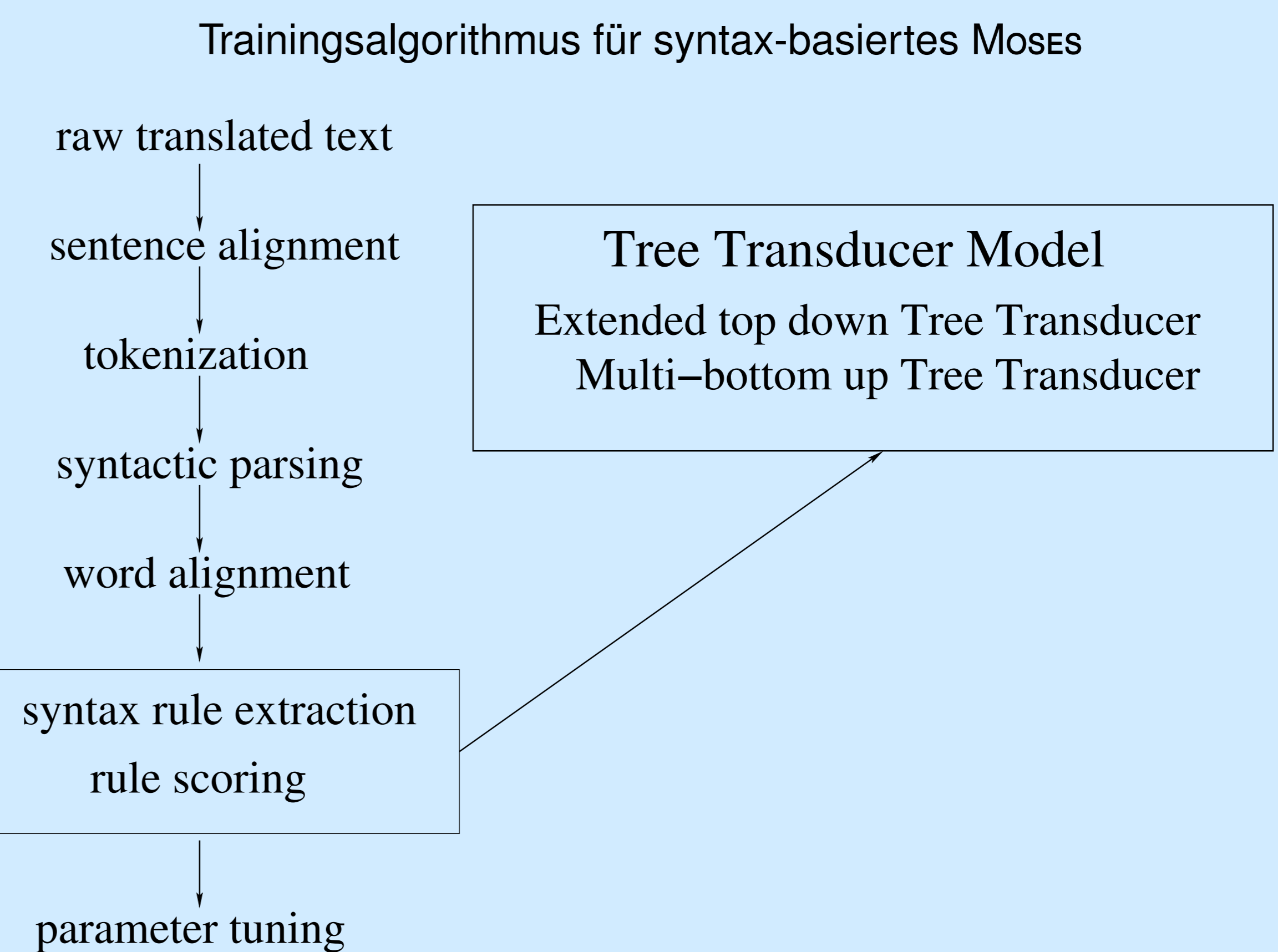
Lokale Rotation:



## 2. Erweiterung des freien Moses Übersetzungsframeworks

- **Moses** : Framework für phrasen-basierte, hierarchische and syntax-basierte maschinelle Übersetzung
- **Integration von Baumübersetzungen** in Moses

- **Änderungen an der Trainingspipeline:**



- **Anpassung des Dekoders**

## 3. Erforschung der grundlegenden Eigenschaften

- a) **Ausdrucksstärke** : geeignet für Anwendung in MT  
 ⇒ Handhabung von lokalen Rotationen  
 – disqualifiziert: top-down tree transducer, SCFG
- b) **Symmetrie** : generisches Modell; sprachunabhängig  
 ⇒ die inverse Übersetzung kann auch berechnet werden  
 – disqualifiziert: top-down und allg. multi bottom-up tree transducer
- c) **Erhaltung der Erkennbarkeit** : effiziente Repräsentation der Sprachen  
 ⇒ Bild und Urbild der Übersetzungen können als erkennbare Baumsprachen repräsentiert werden  
 – disqualifiziert: allg. multi bottom-up tree transducer
- d) **Komposition** : Zusammensetzung einzelner Komponenten  
 ⇒ Vermeidung von Übersetzungskaskaden  
 – disqualifiziert: allg. extended top-down tree transducer

Modell \ Kriterium	a)	b)	c)	d)
synchrone kontextfreie Grammatiken	X	✓	✓	✓
top-down tree transducer	X	X	✓	✓
extended tree transducer (STSG)	✓	✓	✓	X
multi bottom-up tree transducer	✓	X	X	✓
synchrone Baumadjunktionsgrammatik	✓	✓	X	?
synchrone Baumreihen-Substitutionsgrammatik	✓	✓	X	X

## 4. Implementation eines Toolkits für Baumübersetzer

- **Implementation der Baumübersetzermodelle** :  
 – erlaubt empirische Untersuchungen der Eigenschaften  
 – bietet der Community ein weiteres freies Toolkit