

SALSA

Proseminar Lexikalische
Semantik
WS 05

Benjamin Roth
Randolf Altmeyer

Überblick

- Was ist Salsa?
- Corpus vor und nach der Annotation
- Annotationsschritte

Was ist Salsa?

- **S**Aarbrücken **L**exical **S**emantics **A**nnotation
- Manuelle Rollenannotation eines deutschen Corpus (TIGER)
- Verwendung von FrameNet-Frames und Erstellung eigener, neuer Frames (Unknown)
- Vorgehen: Lemma-weise (FN: Frame-weise)

Ziele

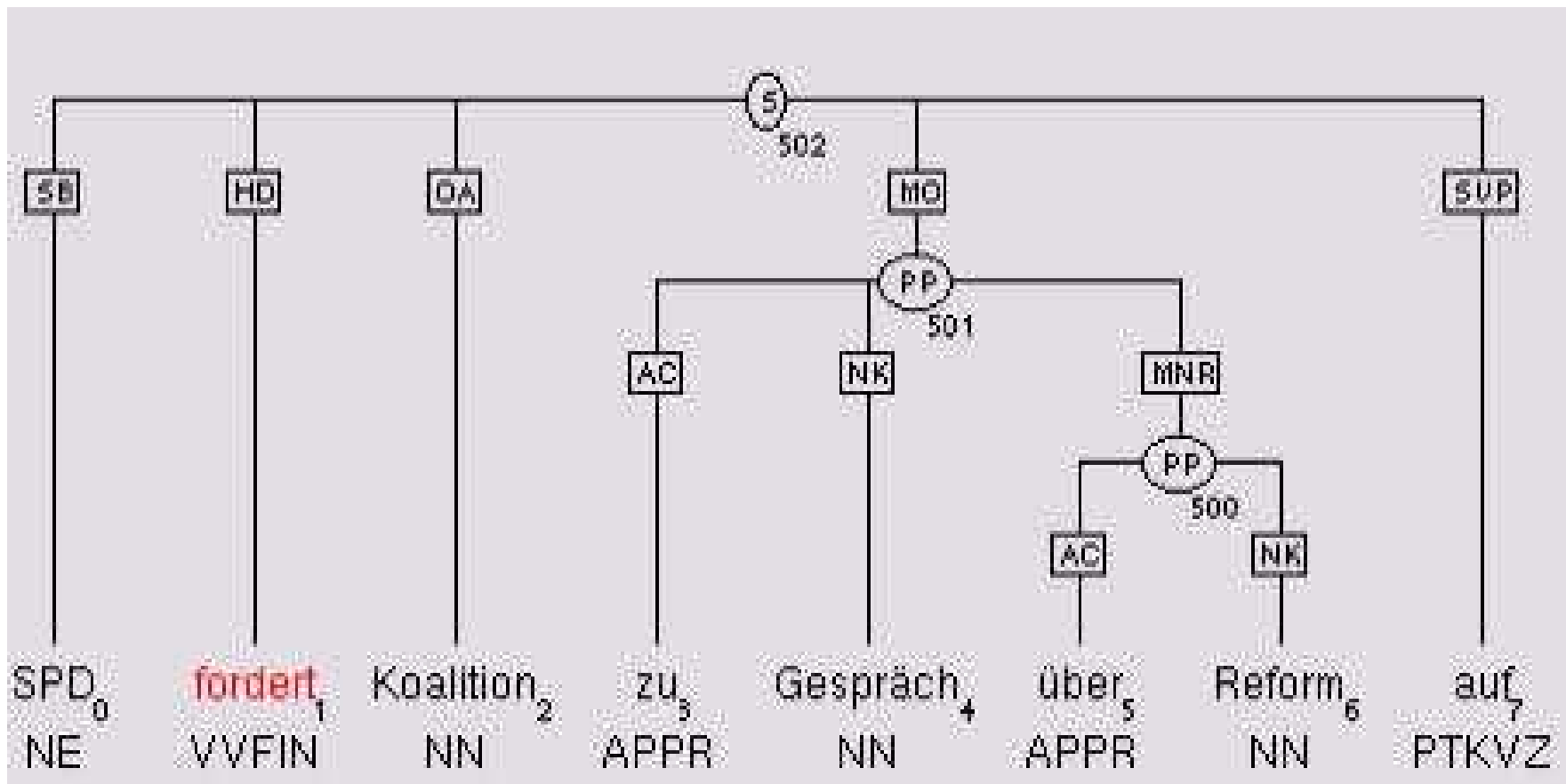
- semantisch annotierter Corpus
- automatische Annotation neuer, großer Corpora
- semantische Disambiguierung
- (Informationsextraktion, Machinelle Übersetzung, Question answering, ...)

Überblick

- Was ist Salsa?
- Corpus vor und nach der Annotation
- Annotationsschritte

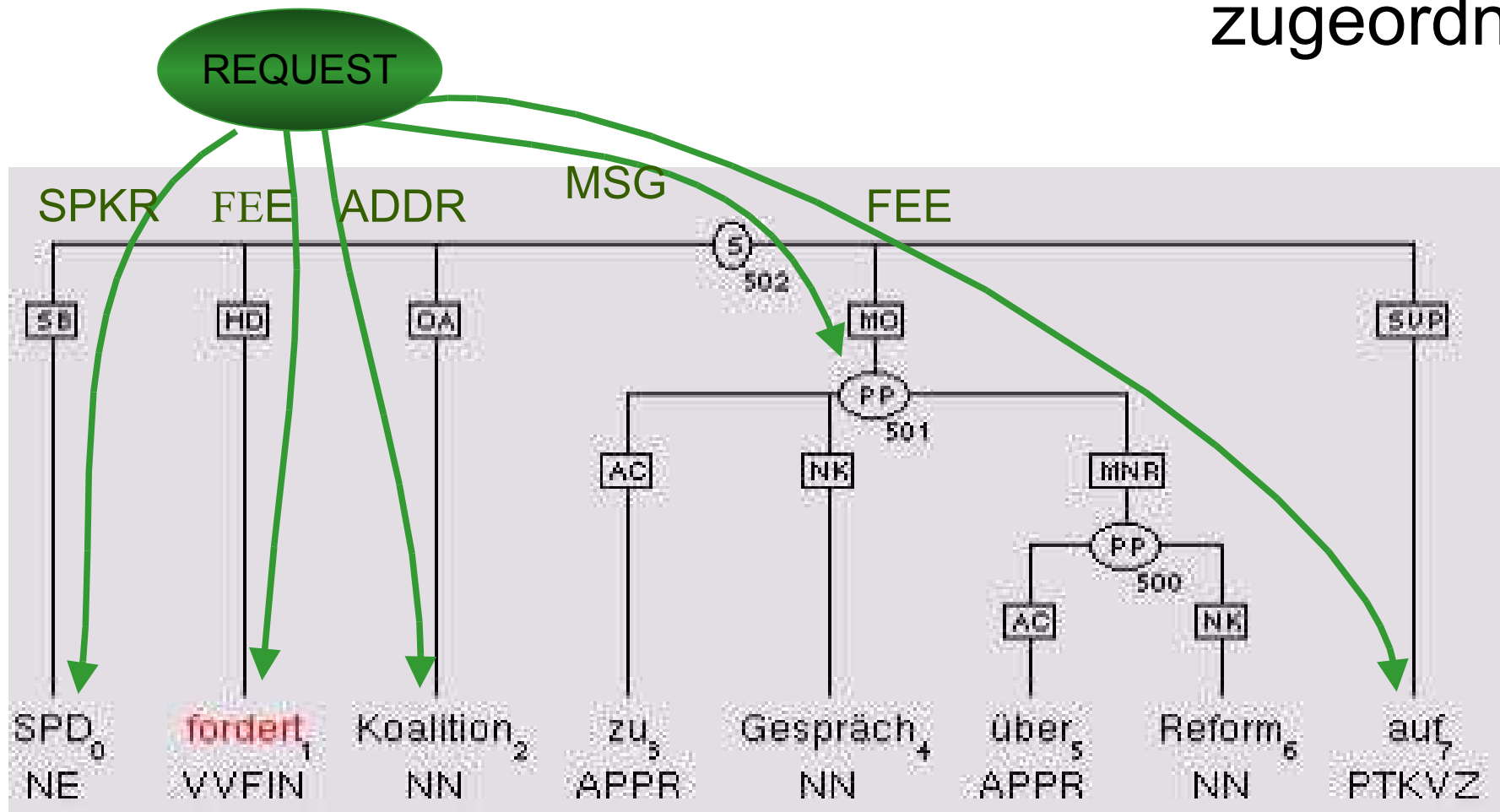
Corpus vor der Annotation

- ohne semantische Annotation
- syntaktisch vorannotiert



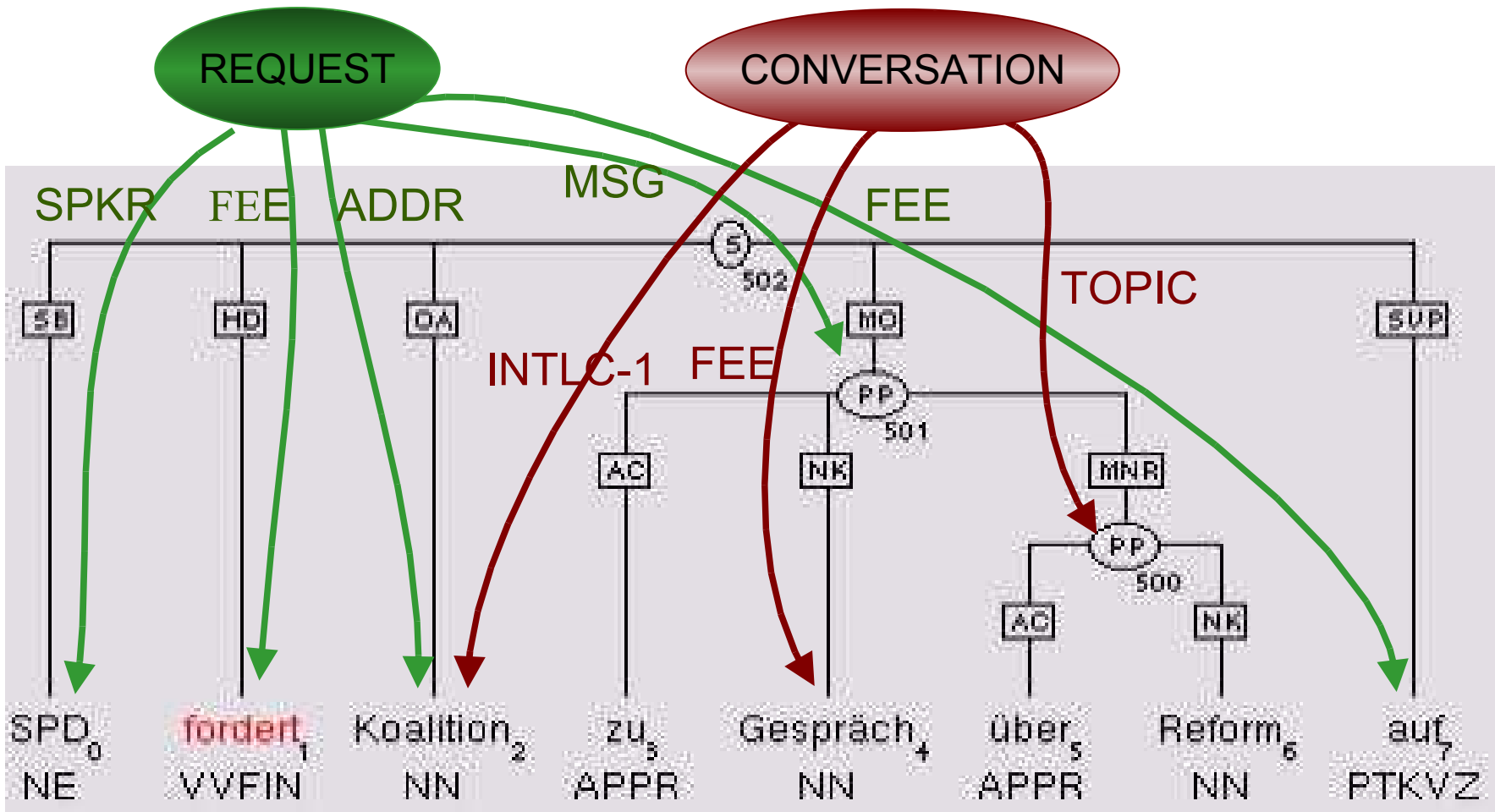
Corpus nach der Annotation

- evozierte Frames
- Rollen sind Wörtern oder komplexen Konstituenten zugeordnet



Corpus nach der Annotation

- pro Satz können mehrere Frames evoziert werden



Überblick

- Was ist Salsa?
- Corpus vor und nach der Annotation
- **Annotationsschritte**

Annotationsschritte

(1) Extraktion

- alle Sätze mit einem bestimmten Lemma werden automatisch zu einem Subcorpus zusammengefasst
- Probleme bei:
 - ♦ abtrennbaren Präfixverben
 - ♦ Synkretismen (z.B. „führen“ beim Lemma „fahren“)

Annotationsschritte

(2) Suche nach möglichen Frames

- Sichtung der ersten 20 Sätze
- Auswahl dafür passender FN-Frames oder Erstellung eigener Frames (Unknowns)
- -> WIKI (which Frames und Suggestions)

Annotationsschritte

(3) Verteilung der Subcorpora

- an je 2 Annotatoren

(4) Annotation

- passende Frames und Rollen werden für jede Instanz des Lemmas mit Hilfe des Tools zugewiesen
- -> DEMO (Tool, einfache Fälle)

Annotationsschritte

- Fragen zu schwierigeren Fällen werden ins Wiki eingetragen und besprochen:

→ Was gehört zum FE?

„Der Strand, genannt Clifton Beach.“

„Wie die Polizei mitteilte , wurden acht Protestanten festgenommen .“

Annotationsschritte

→ mehrere Frames oder Frameelemente passen

=> Unterspezifikation

„Der Staat verteilt die Mittel an die Betriebe.“

„Außerirdische greifen an, meldete die FAZ.“

→ Metapher

=> Source-Frame, Target-Frame

„Sollte Postminister Wolfgang Böttsch umgekippt sein?“

→ kein Frame passt -> neuer Unknown

Annotationsschritte

(5) Merging und Adjudikation

- Vergleich der zwei Annotationen