

Negation Detection

WUW

Verneinungsvorhersager

Unser Ziel

In einem Korpus Sätze zu finden, die Verneinungen enthalten, sowie Skopus und Fokus der jeweiligen Verneinung zu finden.

Unser Ziel

In einem Korpus Sätze zu finden, die Verneinungen enthalten, sowie Skopus und Fokus der jeweiligen Verneinung zu finden.

Otto hat die Schere nicht in den Müll geworfen

Unser Ziel

In einem Korpus Sätze zu finden, die Verneinungen enthalten, sowie Skopus und Fokus der jeweiligen Verneinung zu finden.

Otto hat die Schere nicht in den Müll geworfen

Unser Ziel

In einem Korpus Sätze zu finden, die Verneinungen enthalten, sowie Skopus und Fokus der jeweiligen Verneinung zu finden.

Otto hat die Schere nicht in den Müll geworfen

Skopus

Unser Ziel

In einem Korpus Sätze zu finden, die Verneinungen enthalten, sowie Skopus und Fokus der jeweiligen Verneinung zu finden.

Otto hat die Schere nicht in den Müll geworfen

The diagram shows the sentence "Otto hat die Schere nicht in den Müll geworfen". The word "nicht" is underlined with a green line, and a green arrow points from the word "Fokus" above it to "nicht". A green line is drawn under the entire sentence, with the word "Skopus" centered below it.

Fokus

Skopus

Unser Ziel

In einem Korpus Sätze zu finden, die Verneinungen enthalten, sowie Skopus und Fokus der jeweiligen Verneinung zu finden.

Otto hat die Schere nicht in den Müll geworfen

Skopus

Unser Ziel

Annotation der Sätze mit

1. Negation
2. Negationsskopus
3. Negationsfokus

Unsere Daten

"Der Hund von Baskerville" – A. C. Doyle

Unsere Daten

"Der Hund von Baskerville" – A. C. Doyle

- im Tiger XML Format

Unsere Daten

"Der Hund von Baskerville" – A. C. Doyle

- im Tiger XML Format
- manuell annotierte Negationen zur Evaluation

Unsere Daten

"Der Hund von Baskerville" – A. C. Doyle

- im Tiger XML Format
- manuell annotierte Negationen zur Evaluation
- 50% davon fürs Training
50% zum Testen

Verwendete Werkzeuge

Verwendete Werkzeuge

- Abhängigkeitsparser (Matetools)

Verwendete Werkzeuge

- Abhängigkeitsparser (Matetools)
- XML-Parser (Beautiful Soup, lxml)

Unsere Vorgehensweise

Unsere Vorgehensweise

- Cuewords aus den vorannotierten Trainingsdateien extrahiert

Unsere Vorgehensweise

- Cuewords aus den vorannotierten Trainingsdateien extrahiert
- Mithilfe des Abhängigkeitsparsers die Negationstargets ermittelt

Unsere Vorgehensweise

- Cuewords aus den vorannotierten Trainingsdateien extrahiert
- Mithilfe des Abhängigkeitsparsers die Negationstargets ermittelt
- Regeln manuell abgeleitet

WUW

- Import von Tiger XML
- Signalwort-Erkennung
- Skopus-Erkennung
- Fokus-Erkennung
- Export nach Tiger XML

Cuewords

atemlos ADV
außer ADJA
außerordentlichen ADJA
außerordentlicher ADJA
außerordentliches ADJA
bewusstlos ADV
dies PDS
es PPER
fruchtlosen ADJA
gefahrlos ADJD
hilflos ADJD
kein PIAT
keinePIAT
keinen PIAT
keiner PIAT
keiner PIS
keinerlei PIAT
keines ADJA
keinesfalls ADV
lautlos ADJD

nein ADV
nein NN
nein PTKANT
nicht PTKNEG
nichts NN
nichts PIAT
nichts PIS
nichts PTKANT
nie ADV
niemals ADV
niemand PIS
niemandem PIS
niemanden PIS
nirgends ADJA
nirgends NN
nirgendwoADV

...

Splitwords

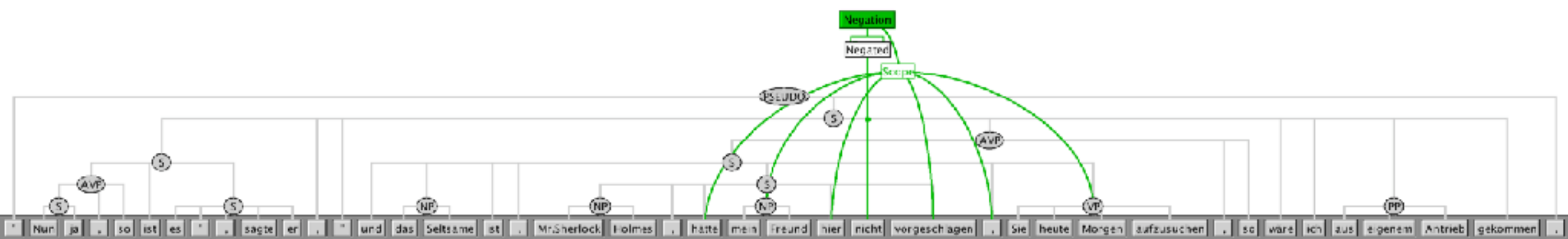
s15_50	unheimliches	ADJA
15_50_s0	un	
15_50_s1	heimliches	
15_50_s0	TARGET	
15_50_s1	NEGATED	
15_50_s1	SCOPE	

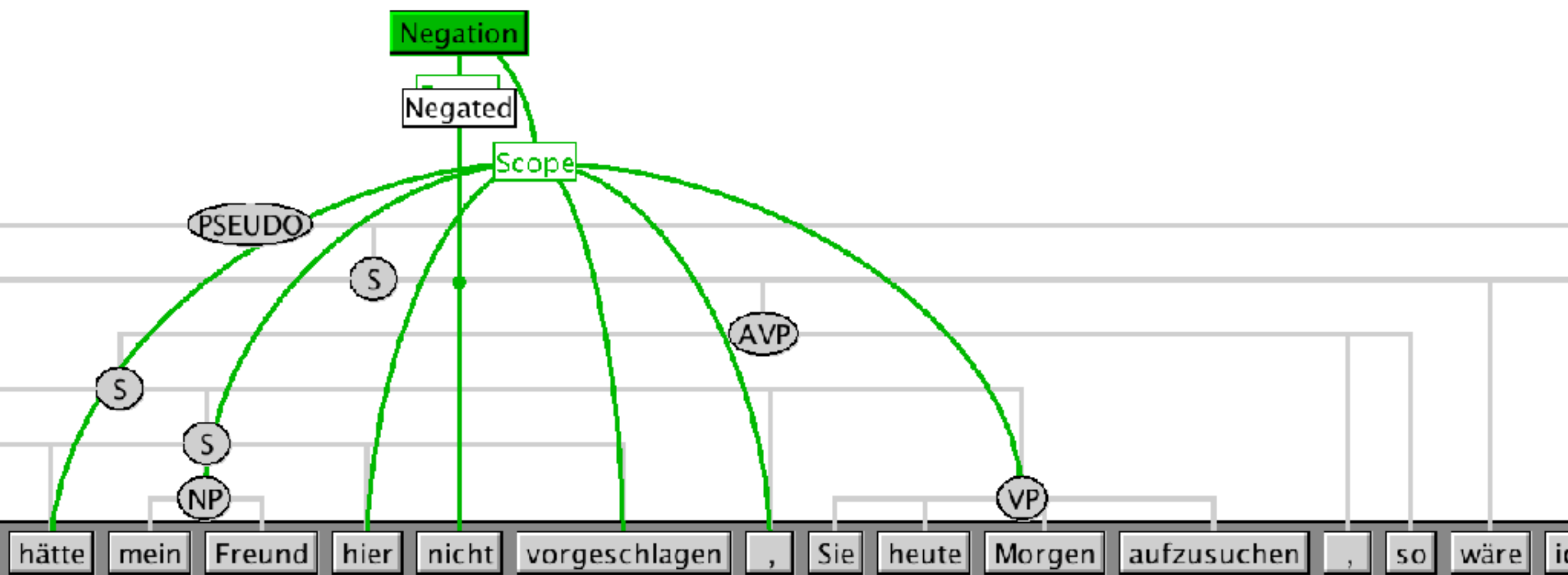
Scope & Focus

s19_41	nichts	PIS
s19_41	TARGET	nichts
s19_46	NEGATED	bringen
s19_543	SCOPE	['mich', 'dazu', 'bringen', ',', ['an', ['so', 'etwas'], ['zu', 'glauben']]]

Algorithmus in Pseudocode

```
for satz in corpus:
    if cue in satz:
        save cue;
        for scope_rule in rules:
            if scope_rule matches satz:
                save scope
                for focus_rule in rules:
                    if focus_rule matches satz:
                        save focus
```



Vergleich der Ergebnisse

Frame: s113_f1

Target [Gold]: ['nicht']

Target [Test]: ['nicht']

Focus [Gold]: ['viel']

Focus [Test]: ['viel']

Negated [Gold]: ['']

Negated [Test]: ['']

Scope [Gold]: ['Leben', 'Noch', 'britische', 'das', 'ich', 'viel', 'weiß', 'über']

Scope [Test]: ['Leben', 'Noch', 'britische', 'das', 'ich', 'viel', 'weiß', 'über']

Vergleich der Ergebnisse

Frame: s113_f2

Target [Gold]: ['nicht']

Target [Test]: ['nicht']

Focus [Gold]: ['alltäglichen']

Focus [Test]: ['alltäglichen']

Negated [Gold]: ['']

Negated [Test]: ['']

Scope [Gold]: ['Routine', 'Stiefel', 'alltäglichen', 'einen', 'es', 'gehört', 'seiner', 'verlieren', 'zu', 'zur']

Scope [Test]: ['Routine', '', 'alltäglichen', '', 'es', 'gehört', '', '', '', 'zur']

Vergleich der Ergebnisse

Frame: 33_f2

Target [Gold]: ['nicht']

Target [Test]: ['nicht']

Focus [Gold]: ['entfesselt']

Focus [Test]: ['unsrem']

Negated [Gold]: ['']

Negated [Test]: ['']

Scope [Gold]: [',', 'Familie', 'Leiden', 'Ruin', 'die', 'entfesselt', 'entsetzlich', 'erneut', 'haben', 'jene', 'so', 'unserer', 'unsrem', 'werden', 'zu', 'zugesetzt', 'üblen']

Scope [Test]: ['', '', '', 'Ruin', '', 'entfesselt', '', 'erneut', '', '', '', '', 'unsrem', 'werden', 'zu', '', '']

Evaluation

Trainingsset

Testset

Evaluation

Trainingsset

90%

Cuewords

Testset

Evaluation

Trainingsset

90%

Cuewords

47%

Targets

Testset

Evaluation

Trainingsset

90%

Cuewords

47%

Targets

30%

Focus

Testset

Evaluation

Trainingsset

90%

Cuewords

47%

Targets

30%

Focus

86%

Scope

Testset

Evaluation

Trainingsset

90%

Cuewords

47%

Targets

30%

Focus

86%

Scope

Testset

68%

Cuewords

Evaluation

Trainingsset

90%

Cuewords

47%

Targets

30%

Focus

86%

Scope

Testset

68%

Cuewords

26%

Targets

Evaluation

Trainingsset

90%

Cuewords

47%

Targets

30%

Focus

86%

Scope

Testset

68%

Cuewords

26%

Targets

27%

Focus

Evaluation

Trainingsset

90%

Cuewords

47%

Targets

30%

Focus

86%

Scope

Testset

68%

Cuewords

26%

Targets

27%

Focus

67%

Scope

Probleme

- Wenig Manpower
- Unerwartete Probleme

Probleme

- Wenig Manpower
- Unerwartete Probleme
- ✓ Lokal ausführbar

Probleme

- Wenig Manpower
- Unerwartete Probleme
- ✓ Lokal ausführbar
 - Webservice mit API

Probleme

- Wenig Manpower
- Unerwartete Probleme
- ✓ Lokal ausführbar
 - Webservice mit API
 - GUI oder App

Verbesserungs-Ideen

Verbesserungs-Ideen

- Unterschiedliche Regelsets je nach Domäne des zu analysierenden Textes



Bitte keine Fragen