

# Analyse von Phrasenstrukturen

05/06/2024 13:39:44

[FAQ Article Print](#)

<b>Category:</b>	FAQ Clarin::Tools	<b>Votes:</b>	0
<b>State:</b>	public (all)	<b>Result:</b>	0.00 %
<b>Language:</b>	de	<b>Last update:</b>	11:20:17 - 03/17/2015 (Europe/Berlin)

## Keywords

Syntax , Phrasenstrukturen

## Symptom (public)

Ich suche ein Werkzeug zur Erkennung und Analyse von Phrasenstrukturen, mit dem ich beispielsweise Rechts- und Linkserweiterungen von Nominalphrasen analysieren und ihre Häufigkeit quantitativ auswerten kann.

## Problem (public)

## Solution (public)

Wenn Phrasenstrukturen von bislang unannotierten Texten analysiert werden sollen, ist eine Reihe von Arbeitsschritten notwendig:

- Bestimmung der relevanten Charakteristika, nach denen die Phrasenstrukturen durchsucht werden
- bestimmte Wortreihenfolgen, Wortarten, Syntax, Eigennamen, usw.
- Automatische Annotation der Texte, die Sie durchsuchen möchten, in WebLicht.

Hier muss man durch Ausprobieren die Werkzeuge finden, die zum einen gute Ergebnisse auf Ihren Daten liefern, das betrifft zum einen Satzgrenzen-Erkennung, Wortgrenzenerkennung, Wortarten, Eigennamen, ggf. Morphologie und Lemmas. Zum anderen muss der Parser, der die syntaktische Struktur annotiert, nicht nur möglichst korrekt sein, sondern gleichzeitig sollte sein theoretisches Modell die Strukturen, die Sie suchen, möglichst gut abbilden.

Die in WebLicht annotierten Texte kann man im negra-export-Format exportieren und in das TIGERSearch-Abfragewerkzeug importieren. Darin kann man Suchanfragen erstellen, die mit möglichst gutem Recall und Precision die gesuchten Muster auffinden.

Bei automatischer Annotation findet man nicht 100% aller Treffer, aber die Methode eignet sich gut für einen Überblick, eine quantitative Analyse und für das Finden neuer Muster. TIGERSearch erlaubt leider keinen direkten Export der Treffer.