

Ausgangspunkt

Erste Beobachtung:

Textkorpora verfügen in der Regel über keine semantische relationale und ontologische Annotation. Daher lassen sich sprachliche Muster nur schwer disambiguieren.

Zweite Beobachtung:

Mithilfe von Textkorpora lassen sich formale und morphosyntaktische Daten quantitativ erheben.

CQL-Abfrage: `[lemma="Debatte"] [tag="ART \. (Def|Indef)|PRO. (Dem|Poss). Attr"] . Gen. * "] [tag="N. * "]`

Ergebnisse: *Debatte der Gegenwart* | *Debatte der Historiker* | *Debatte der Hasssprache*

Theoretische Grundlagen

Semantische Valenz in der nominalen Argumentstruktur

Substantive verfügen über die Fähigkeit, als Prädikate zu fungieren.

Eine Argumentstruktur ist „die semantische Repräsentation einer Proposition und besteht [...] aus genau einem Prädikat und der ihr zugeordneten Liste von Argumenten [...]“ (Engelberg, 2019)

Die semantische Valenz „reflektiert den Umstand, dass Wörter (als Valenzträger) bestimmte Kontextpartner mit bestimmten Bedeutungsmerkmalen (als Valenzpartner) fordern, andere Kontextpartner mit anderen Bedeutungsmerkmalen aber ausschließen. Die semantische Valenz regelt somit die Besetzung von Leerstellen mit Klassen von Partnern, die semantisch durch bestimmte Bedeutungsmerkmale festgelegt sind.“ (Helbig 1992, S. 8)

Theory of Norms and Exploitations (TNE) | Corpus Pattern Analysis (CPA)

Um die Bedeutungsnuancen lexikalischer Einheiten festlegen zu können, schlägt Hanks (2013) einen Ansatz (TNE) vor, auf dem die Herangehensweise CPA beruht. Die Grundidee besteht darin, dass durch Abstraktionen der Bedeutungsmerkmale und die daran anlehende Erstellung lexikalischer Sets das semantische — auch kombinatorische — Verhalten von Wörtern leichter zu fassen ist.

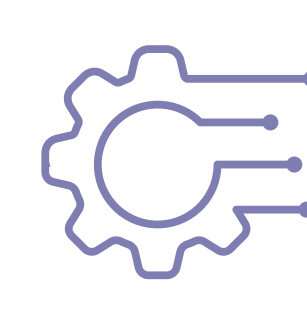
PORTLEX, MultiGenera und MultiComb

PORTLEX



Mehrsprachiges Wörterbuch zur Valenz des Nomens mit Beschreibung der syntaktisch-semantischer Schnittstelle.

MultiGenera und MultiComb



Automatische Generatoren von Argumentstrukturen des Nomens mittels einer Filterung von Bedeutungsmerkmalen.

MultiTools



Erstellung von semantisch annotierten Vokallisten mit Rückgriff auf Wordnet.

Forschungsfrage

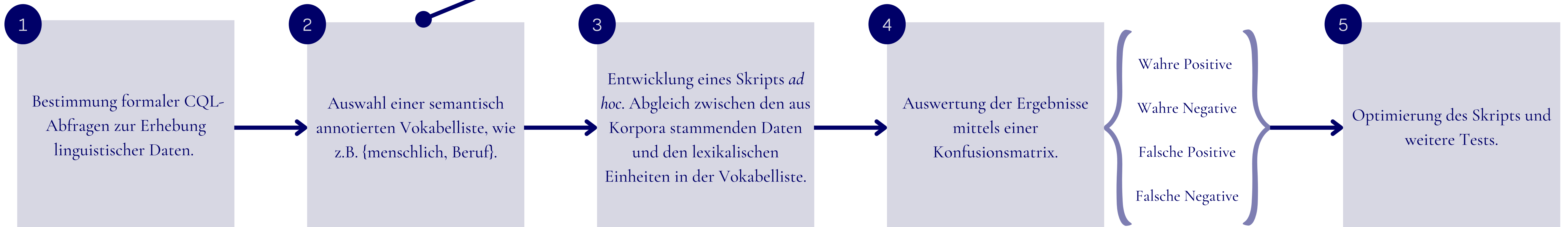
Hypothese:

Automatisch getaggte Daten mit semantischer — kategorialer und relationaler (Engel, 2004) — Information tragen zur Operationalisierung und Beschreibung von Argumentstrukturmustern bei.

Forschungsfragen:

- 1 Ist die Entwicklung eines (semi)automatischen Annotationssystems und seine künftige Integration in Systeme zur Bearbeitung von Texten sinnvoll?
- 2 Wie kann ein semantisches Annotationstool — zunächst als Prototyp (hier vorgestellt) — erstellt werden?

Methode



Vorläufige Ergebnisse



Simulation: Genitivrealisierungen in der Funktion einer Subjektivergänzung

Auf einen zweckgemäß funktionierenden Prototyp hinweisende Faktoren

Lexikalische Einheiten, die die Funktion anderer Ergänzungen erfüllen, werden nicht annotiert. Das Annotationssystem annotiert Genitivrealisierungen, die als Subjektivergänzungen fungieren (z.B. *Debatte der Republikaner*). Andere Ergänzungen (wie z.B. die Objektivergänzung, *Debatte der Frage*) werden nicht getaggt. Dieser Standpunkt entspricht dem Zweck unserer Untersuchung. Auch lexikalische Einheiten, die der Rolle einer Angabe entsprechen, werden nicht markiert (*Debatte der Woche*), was auch das Ziel ist.

Zu verbessernde Aspekte

Lexikalische Einheiten, die in der ausgewählten Vokabelliste nicht vorhanden sind, werden nicht getaggt. Das ist der Fall der meisten Komposita bei Nominalphrasen wie *Debatte der Parteivorsitzenden*.
 Lösungsvorschlag: Anwendung regulärer Ausdrücke wie `*vorsitzende`
 Eigennamen, Akronyme und Abkürzungen werden nicht annotiert.
 Partizipien I und II, die als Nomina fungieren, werden nicht anerkannt und nicht getaggt.
 Lösungsvorschlag: Einbeziehung von Partizipien in die Vokabelliste oder Umsetzung regulärer Ausdrücke wie `*nden`
 Lexikalische Einheiten, die reguläre Polysemie aufweisen, werden nicht annotiert. Das ist der Fall von Strukturen, die mindestens zwei Lesarten aufweisen: z.B. *Bericht der Zeitung*.
 { Lesart: Der Bericht ist in der Zeitung zu finden.
 { Lesart: Es wird über die Zeitung (als Institution) berichtet. }