

Die Repräsentation und Auflösung von ambigen Wortbedeutungen in der Computerlinguistik

Übung 2: SALTO

PD Dr. Sabine Schulte im Walde

27. März 2010

1 Aufgabe

- Sie definieren Wortbedeutungsinventare und nutzen ein Korpus-Annotations-Werkzeug, um entsprechend der Definitionen Wort-Vorkommen mit ihren Bedeutungen zu annotieren.
- Abgabe: Lösungen bis zum 18. April 2010 als pdf per Email an schulte@ims.uni-stuttgart.de schicken.

2 Daten

Sie arbeiten mit einem Ausschnitt des *TIGER Corpus*:

<http://www.ims.uni-stuttgart.de/projekte/TIGER/TIGERCorpus/>.

Der Ausschnitt enthält drei kleine Teil-Korpora mit 23 Vorkommen von *bestehen*, 15 Vorkommen von *glatt*, und 20 Vorkommen von *Leiter*.

3 Werkzeug

SALTO:

- Siehe Foliensatz.
- Siehe Software auf *SALSA*-Homepage:
<http://www.coli.uni-saarland.de/projects/salsa/>

4 Übungen

4.1 SALTO und Bedeutungsinventare

1. Vorbereitung:
Machen Sie sich mit dem Werkzeug vertraut.
2. (9 Punkte) Definieren Sie Bedeutungsinventare für die drei Wörter *bestehen*, *glatt* und *Leiter*. Nutzen Sie dafür sowohl die Korpus-Daten, auf die Sie über SALTO Zugriff haben, als auch Online-Wörterbücher und eigene Intuitionen. Erstellen Sie mindestens zwei und höchstens fünf Bedeutungen pro Wort. Definieren Sie jede Bedeutung kurz in eigenen Worten.

3. (6 Punkte) Annotieren Sie die Vorkommen der drei Wörter in den Korpora unter SALTO mit ihren Wortbedeutungen.
4. (5 Punkte) Erstellen Sie eine Frequenzliste der Wortbedeutungen in den jeweiligen Korpora für die drei Wörter. Beschreiben Sie die Frequenzlisten. Gibt es dominante Wortbedeutungen, und wenn ja, welche sind das?