



Wie erwerben fortgeschrittene Lerner des Deutschen die Linearisierung im Mittelfeld?

Wortstellung im topologischen Feldermodell (u.a. Höhle 1986)

Vorfeld	linke Satzklammer	Mittelfeld					rechte Satzklammer	Nachfeld
Später	konnte	er	sich	nicht einmal mehr	nach ihr		umdrehen	ohne zu lachen.

Linearisierungsfaktoren

Das deutsche Mittelfeld ist bestimmt durch

- wenige Regeln (u.a. Engel 1996,Eisenberg 2006)
- viele Linearisierungsfaktoren (u.a. Siewierska 1997, Uszkoreit 1987)

- | | |
|---|---|
| <ul style="list-style-type: none">Grammatische Funktion
Subjekt., dir. Objekt., ind. ObjektKasus
Nominativ, Akkusativ, DativWortart
Personalpronomen, Vollnomen, ReflexivpronomenKonstituentengewicht
Wortanzahl, Silbenanzahl | <ul style="list-style-type: none">Satztyp
Hauptsatz, NebensatzSemantische Rolle
Agens, Patiens, BenefactorInformationsstatus
gegebene Information, neue InformationBelebtheit
Person, Institution, Tier, Materie |
|---|---|

→ Wie erfolgreich reproduzieren fortgeschrittene Lerner des Deutschen den unterschiedlichen Einfluss dieser Faktoren?

Design

Contrastive Interlanguage Analysis (CIA) (u.a. Granger 2008)

Variation von Subjekt-Objekt vs. Objekt-Subjekt

- Falko Aufsatzkorpus (Lüdeling et al 2008)
 - 92 Lernertexte (1088 Mittelfelder)
 - 95 Muttersprachlertexte (1511 Mittelfelder)
- Annotation
 - manuell
 - Mittelfelder
 - Satztyp
 - automatisch
 - Wortart & Kasus: automatische Annotation durch rfTagger (Schmid & Laws 2008)
 - Konstituentengewicht (automatisch)
 - Satzfunktionen (Mate Tools , siehe "Zielhypothesen parsen")

Zielhypothesen parsen:

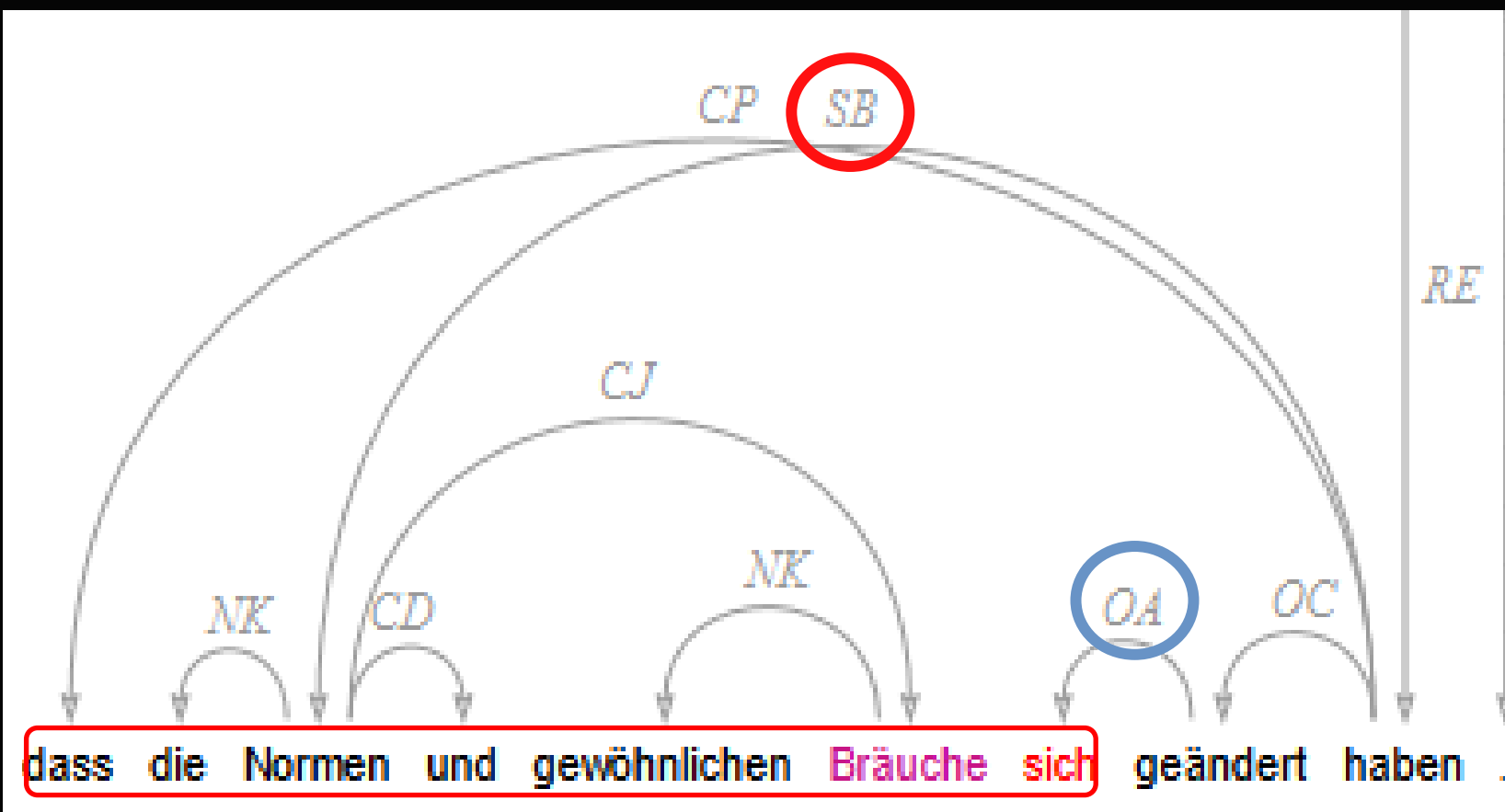
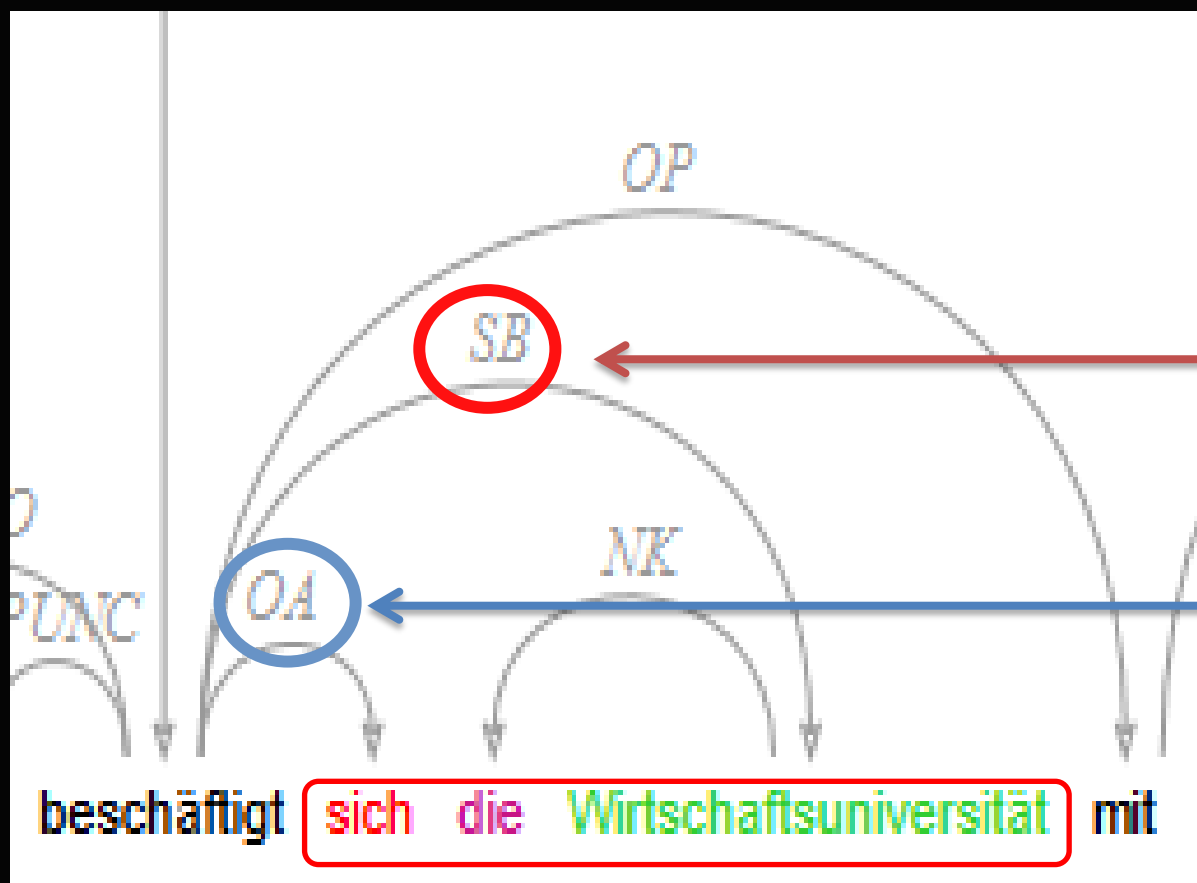
Grammatische Funktion → syntaktisch tief annotierten Daten

- Problem I:** Parser wie die Mate Tools (Bohnet 2010) sind nicht auf Lernertexte (LT) trainiert.
- Lösung I:** Für jeden Satz wird eine grammatische Zielhypothese (ZH1) gepart (vgl. Reznicek et al 2010).
- Problem II:** Grammatische Zielhypothesen können die Linearisierung ändern.
- Lösung II:** Die Auswertung findet auf einer weiteren Zielhypothese (ZH0) ohne "Bewegungen" statt.

ZH0: Aber in **den** meisten Fällen **das** ist nicht der Fall.

ZH1: Aber in **den** meisten Fällen ist **das** nicht der Fall.

LT: Aber in **die** meisten Fällen **dass** ist nicht der Fall



Mixed Effects Models

- Die gleichzeitigen Effektstärken der unterschiedlichen Faktoren lassen sich durch ein *linear mixed effects model* (Bates et al. 2011) miteinander vergleichen.
- Das Modell optimalsiert die Effektstärken in einer linearen Funktion für alle Faktoren bis die beste Vorhersage für die wirklichen Daten erreicht ist.
- Das Modell verlangt zwei Typen von Faktoren:
 - a) fixed effects: die untersuchten Faktoren
 - b) random effects: Lerner, Verb

Unterscheidung zwischen L1 oder L2 wurde ebenfalls als *fixed effect* modelliert.

→ Lernereffekte sind daher Interaktionen von L1/L2 und anderen Faktoren.

$$z = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_{k+1}$$

Ergebnisse

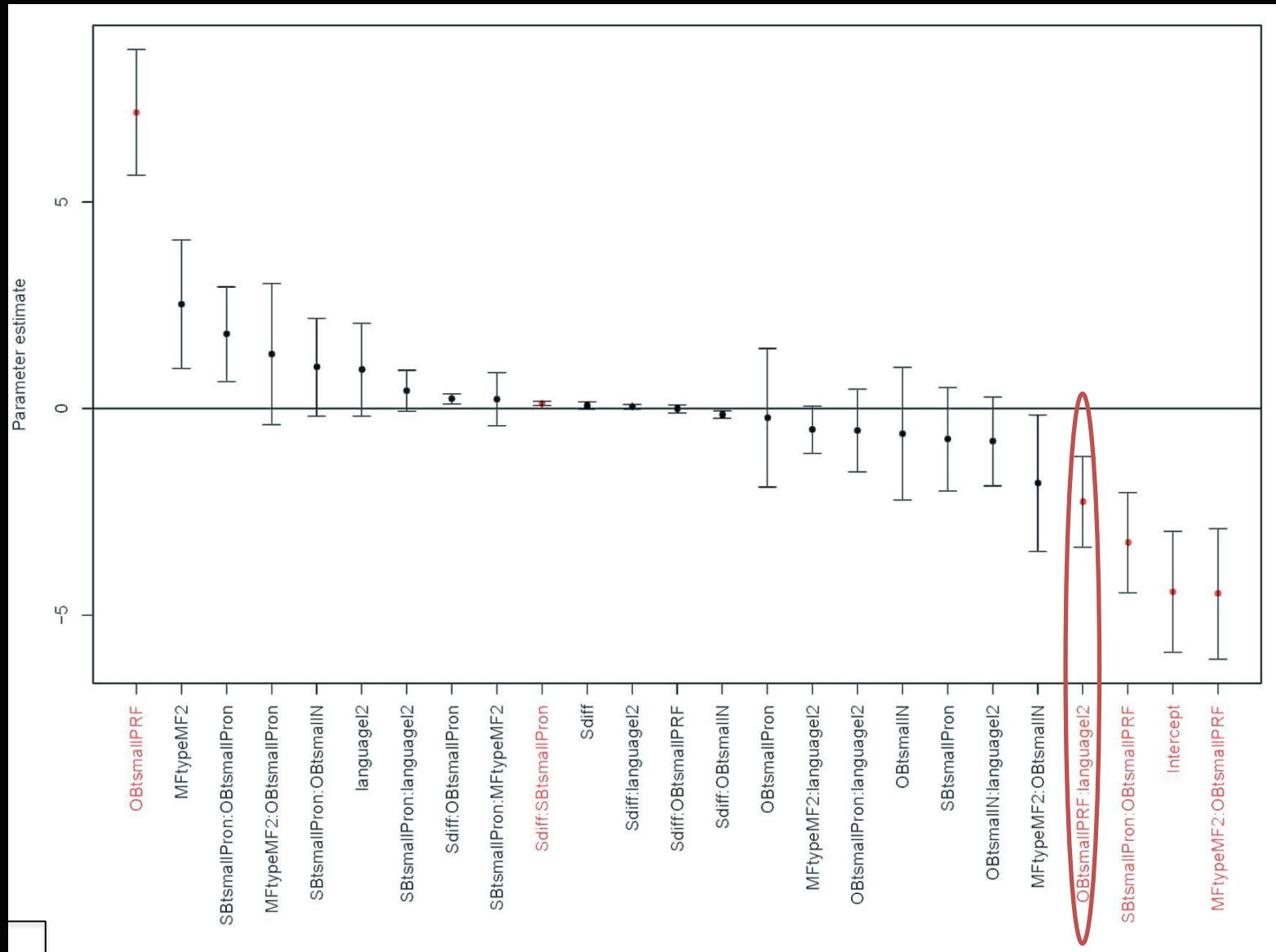
Satztypen



Nebensatz Hauptsatz

- Lerner produzieren signifikant weniger Objekt-Subjekt-Mittelfelder als Muttersprachler.
- ABER nur in Nebensätzen**

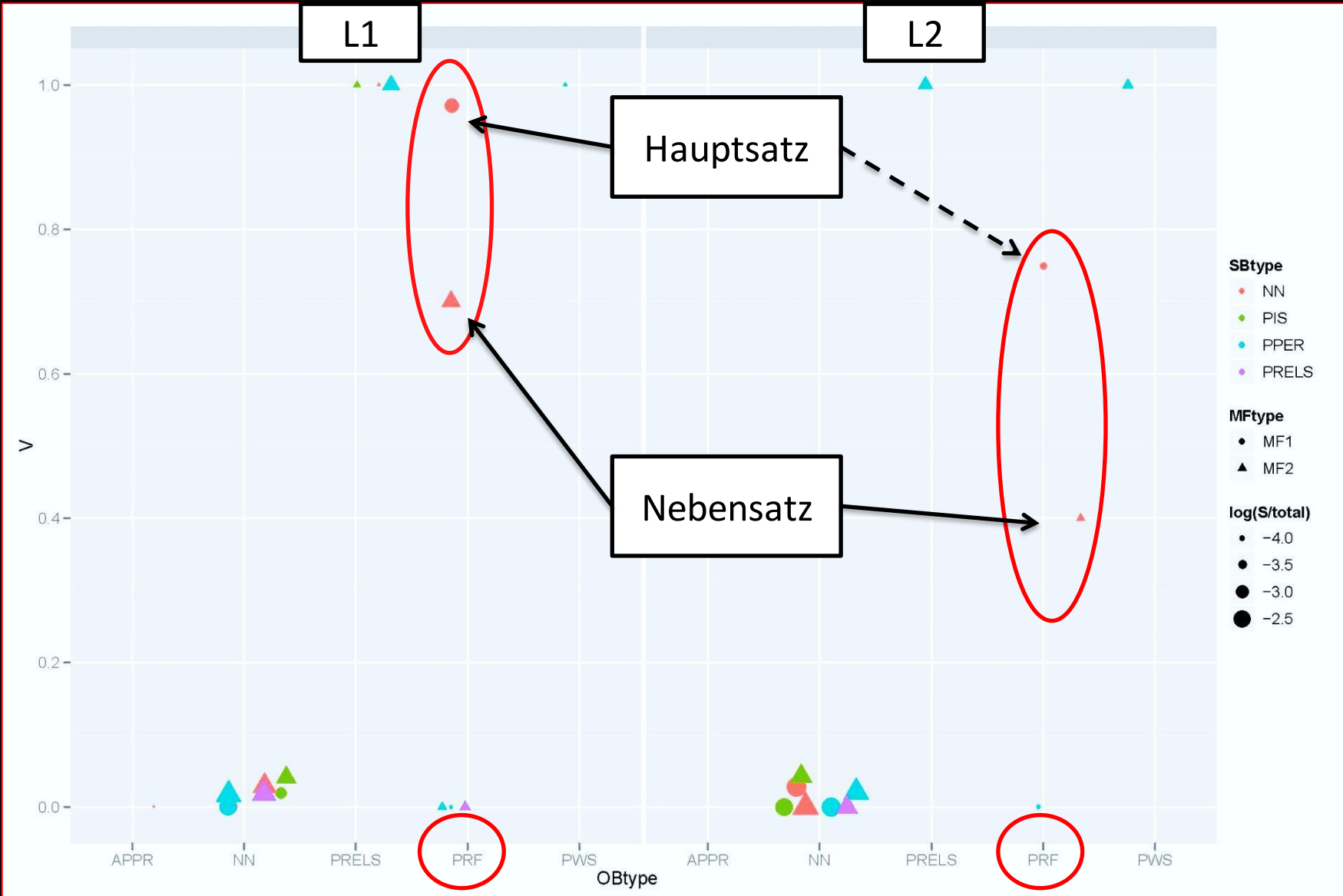
Übersicht für Faktoren



Einfluss bei festem NP-Subjekt

- Lerner unterschätzen die Bedeutung der Variation beim Reflexivpronomen "sich".
- Deckt sich mit Ergebnissen aus Overuse/Underuse-Studien in Zeldes et al. 2008.

Übersicht für Faktoren



Einfluss bei festem NP-Subjekt

- Keine der übrigen untersuchten Faktoren zeigt eine Interaktion mit der L1/L2-Unterscheidung

Diskussion

- Methode ist auf die Daten scheinbar anwendbar.
- Einzelne Linearisierungsfaktoren (Wortart) und deren Interaktion (Satztyp) scheinen auch für fortgeschrittene DaF-Lerner problematisch zu sein.
- Weitere Faktoren müssen untersucht werden
- PROBLEME:**
 - Qualität der Parses ist unbekannt
 - Daten zeigen nur wenig Varianz (sparse data).
 - Bisher wurde nur ein Teil der Daten ausgewertet
 - Texttopik- und Registereffekte noch nicht berücksichtigt.

Literatur

Bates, Douglas; Maechler, Martin; Bolker, Ben (2011): lme4. Linear mixed-effects models using Eigen and Eigen. URL: <http://CRAN.R-project.org/package=lme4> [Stand: 05. September 2011]. ↯ Bohnet, Berndt (2010): Top Accuracy and Fast Dependency Parsing is not a Contradiction. In: *The 23rd International Conference on Computational Linguistics. (COLING 2010)*. ↯ Eisenberg(2006): Der Satz. 3., durchges. Aufl. ↯ Engel(1996): Deutsche Grammatik. 3., korr. Aufl. ↯ Granger(2008): Learner corpora. In: Lüdeling/Kytö(Hrsg.): *Corpus linguistics*, 259–275. ↯ Höhle, Tilman N. (1986): Der Begriff 'Mittelfeld'. Anmerkungen über die Theorie der topologischen Felder. In: Schöne, Albrecht;Stephan, Inge(eds.): *Kontraversen, alte und neue. Akten des VII. Kongresses der Internationalen Vereinigung für germanische Sprach- und Literaturwissenschaft*. Tübingen: Niemeyer (= Kontroversen, alte und neue; 6), p.329–340. ↯ Lüdeling et al(2008): Das Lernerkorpus Falko. In: *Deutsch als Fremdsprache* 45 (2), 67–73. Schmid, Helmut; Laws, Florian (2008): Estimation of conditional probabilities with decision trees and an application to fine-grained POS tagging. In: Scott, Donia(ed.): *22nd International Conference on Computational Linguistics. Coling 2008*. COLING. Stroudsburg, Pa: Association for Computational Linguistics, p.777–784. ↯ Reznicek, Marc; Walter, Maik; Schmidt, Karin; Lüdeling, Anke; Hirschmann, Hagen; Krummes, Cedric; Andreas, Thorsten (2010): Das Falko-Handbuch. Korpusaufbau und Annotationen. Version 1.0. Berlin: Institut für deutsche Sprache und Linguistik, Humboldt-Universität zu Berlin. ↯ Siewierska(1993): On the Interplay of Factors in the Determination of Word Order. In: Jacobs et al. (Hrsg.): *Syntax*(= Handbücher zur Sprach- und Kommunikationswissenschaft, 826–846. . ↯ Uszkoreit, Hans (1987): Word Order and Constituent Structure in German. Stanford.(= Center for the Study of Language and Information lecture notes; 8). ↯ Zeldes et al(2008): What's hard? In: *Proceedings of QILT 3*.