

Lernereffekte im deutschen Mittelfeld

Marc Reznicek

4.1.2012

Kolloquium Korpuslinguistik
Humboldt-Universität zu Berlin

- Variation im deutschen Mittelfeld
- Linearisierungsfaktoren
- Modellierung
- Daten und Annotation
- Methode
- Resultate
- Diskussion
- Ausblick

Erwerb syntaktischer Variation



- Lange Tradition von Studien zum Syntaxerwerb
(vgl. Ellis 2009)
- Fokus auf Regelerwerb und Erwerbsstadien
(vgl. Pienemann 2005)
- Bisher nur wenige Studien zur Variation syntaktischer Muster
- Fragestellung:
Wie erwerben Zweisprachler die Kompetenz, konkurrierende sprachliche Muster zielsprachlich zu verwenden?

Das deutsche Mittelfeld



- Topologisches Feldermodell (Drach 1937, Höhle 1986, Pasch et al. 2003)

VF	LSK	MF	RSK	NF
Der Feminismus	hat	den Frauen schon immer	geschadet	durch seine Radikalität

VF	LSK	MF	RSK	NF
	dass	der Feminismus den Frauen schon immer durch seine Radikalität	geschadet hat	

VF	LSK	MF				RSK	
			LSK	MF	RSK		
	dass	der Feminismus	der	sehr radikal	agierte,	den Frauen schon immer	geschadet hat

Variation im deutschen Mittelfeld



- **Scrambling**

Mittelfeldkonstituenten erlauben eine hohe Variabilität konkurrierender Wortstellungen.

(Haider/Rosengreen 2003)

<i>dass</i>	[diese Ansicht] _{AKK} [<i>in Zukunft</i>] [<i>viel mehr Menschen</i>] _{NOM}	<i>zu teilen</i> <i>lernen</i>
-------------	---	-----------------------------------

<i>dass</i>	[<i>viel mehr Menschen</i>] _{NOM} [<i>in Zukunft</i>] [diese Ansicht] _{AKK}	<i>zu teilen</i> <i>lernen</i>
-------------	---	-----------------------------------

<i>dass</i>	[<i>in Zukunft</i>] [<i>viel mehr Menschen</i>] _{NOM} [diese Ansicht] _{AKK}	<i>zu teilen</i> <i>lernen</i>
-------------	---	-----------------------------------

Einflussfaktoren auf die lineare Abfolge



- Wortstellung ist nur teilweise regelbasiert.
- Unterschiedliche Einflussfaktoren auf die Wortstellung sind diskutiert worden (vgl. Siewierska 1997, Uszkoreit 1987)
- **Grammatische Funktion**
Subjekt, Objekt
- **Kasus**
Nom., Akk., Dat.
- **Wortart**
Personalpronomen, Nomen,
Reflexiv
- **Konstituentengewicht**
Anzahl der Wörter, Anzahl der
Silben

Einflussfaktoren auf die lineare Abfolge



- Wortstellung ist nur teilweise regelbasiert.
 - Unterschiedliche Einflussfaktoren auf die Wortstellung sind diskutiert worden (vgl. Siewierska 1997, Uszkoreit 1987)
- **Grammatische Funktion**
Subjekt, Objekt
 - **Kasus**
Nom., Akk., Dat.
 - **Wortart**
Personalpronomen, Nomen, Reflexiv
 - **Konstituentengewicht**
Anzahl der Wörter, Anzahl der Silben
 - **Phrasentyp**
NP, PP, CP
 - **Semantische Rolle**
Agens, Patiens, Recipiens
 - **Informationsstatus**
gegeben, neu
 - **Belebtheit**
Person, Institution, Tier, Materie

Modellierung unterschiedlicher Faktoren



- Meisten Faktoren wurden nur einzeln betrachtet
(vgl. Kurz 2000, Heylen et al. 2005, Bader/Häusler 2010)
- Modellierung des gleichzeitigen Einflusses konkurrierender Faktoren:
 - Möglichkeit I: **Hierarchien**
 - Optimalitätstheorie (Uzkoreit 1987)

 - Möglichkeit II: **Analyse der relativen Faktorenstärke**
 - Quantitative Analyse (Hoberg 1981, Kurz 2000, Heylen et al. 2005, Bader/Häusler 2010)

Spezifische Forschungsfrage



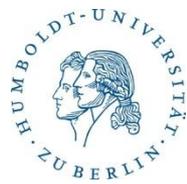
- Zeigen fortgeschrittene Zweistsprachlerner Texte Unterschiede in den Effektstärken für diese Einflussfaktoren als Texte von Muttersprachlern?
- Erste Studie: **Nominale Verbargumente**

Kontrastive Interlanguage Analyse CIA ([Granger 2008](#))

Annahmen:

- Lernaltersprache ist systematisch
- Ausreichend geringe Variation innerhalb der Gruppe
- Transfer & universelle Spracherwerbsphänomene

L1-Zeitungsartikel (Bader & Häusler 2010)



- **Grammatische Funktion** hat großen Effekt

96% **SB-OB** 4% **OB-SB**

- **Kasus** beeinflusst Wortstellung nur wenn zwei Vollnomen beteiligt sind.

DAT_{OBJ} (25%) – SB > **AKK**_{OBJ} (1%) – SB

- **Wortart** hat einen großen Effekt

Pronomen > Vollnomen

- **Konstituentengewicht** hat keinen Einfluss



Lüdeling et al. 2008

- fortgeschrittene DaF-Lerner
- Aufsätze und Zusammenfassungen
- Längsschnitt- & Querschnittsdaten
- ~260.000 Token und am wachsen
- POS, lemma (Treetagger, Schmid 1994)
- Abhängigkeiten geparkt (neu)

untersuchtes Sub-set

- 94 Texte FalkoEssayL2v2.1 (25 L1s)
- 94 Texte FalkoEssayL1v2.1 Kontrollgruppe

<http://www.linguistik.hu-berlin.de/institut/professuren/korpuslinguistik/forschung/falko/standardseite/>

Daten: Zielhypothese



Nicht-kanonische syntaktische Strukturen in Lernaltersprache machen eine Beschreibung mit Standardgrammatiken unmöglich.

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

(FalkoEssayL2v2.0:fk006_2006_08)

Daten: Zielhypothese



Deshalb wird eine minimale grammatische Korrektur (ZH1) explizit ins Korpus integriert .

(Reznicek et al. 2009)

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

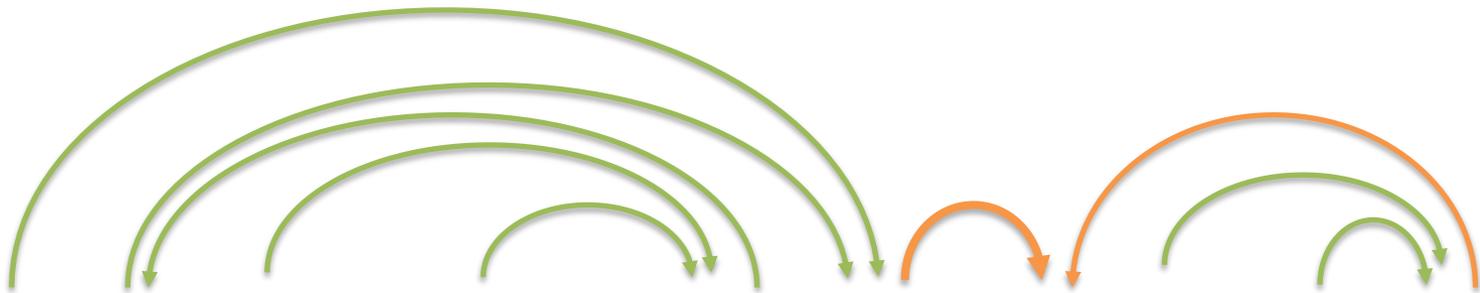
LT: Aber in **die** meisten Fällen **das ist** nicht der Fall.

(FalkoEssayL2v2.0:fk006_2006_08)

Daten: Zielhypothese



Um die ursprüngliche Wortstellung zu behalten, werden die Abhängigkeiten auf Originalstellung zurückbezogen.



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

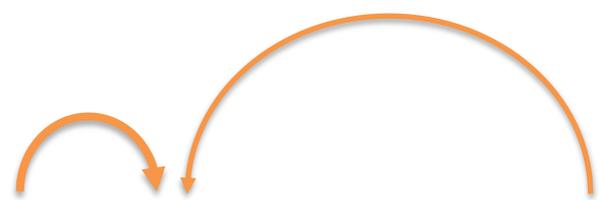
ZH1: Aber in **den** meisten Fällen **ist das** nicht der Fall.

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

Daten: Zielhypothese



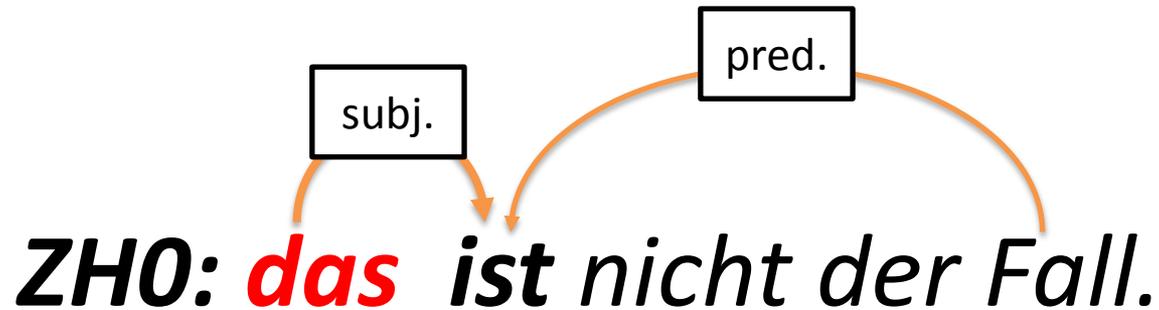
ZH0: Aber in **den** meisten Fällen **das** ist nicht der Fall.



Daten: Zielhypothese



Jede Dependenz wird automatisch mit grammatischer Funktion versehen.



Annotation: Mittelfelder



In allen Äußerungen wurden die Mittelfelder manuell annotiert.

Für die Verbargumente in jedem Mittelfeld wurde annotiert:

- 1) Satztyp (V1, V2, V_{end})
- 2) Abfolge der Satzfunktionen (subj – obj, obj-subj)
- 3) Wortart (Nomen, Pron ...)
- 4) Kasus (Nominativ, Akkusativ, Dativ)
- 5) Konstituentenlänge (in Wörtern, in Silben)

Methode: linear mixed effect model



- Linear mixed effect model (LME) erlauben die Berechnung der Effektstärken unterschiedlicher Faktoren:

(Bates et al. 2011)

$$z = \beta_0 + \beta_1 \overset{\text{Variable}}{\downarrow} x_1 + \beta_2 x_2 + \beta_3 \overset{\text{Effektstärke}}{\downarrow} x_3 + \dots + \beta_k x_{k+1}$$

- → Wahrscheinlichkeit für OB-SB-Reihenfolge mit Subjekt als Vollnomen
- Zufallseffekte: Verb, Text

- LMEs machen zwei unterschiedliche Arten von Einflüssen sichtbar.

a) Haupteffekte

Faktor hat einen direkten Einfluss auf die Zielverteilung

Beispl.: *Pronominalisierung des Objekts begünstigt sein Auftreten an erster Stelle im Mittelfeld mit Effektstärke X.*

b) Interaktionen

Faktor hat nur in Verbindung mit anderen Faktoren einen Einfluss.

Beispl.: *Nur wenn beide nominalen Argumente Vollnomen sind, hat Kasus einen Einfluss. Das gemeinsame Auftreten der Faktoren hat eine Effektstärke von X.*

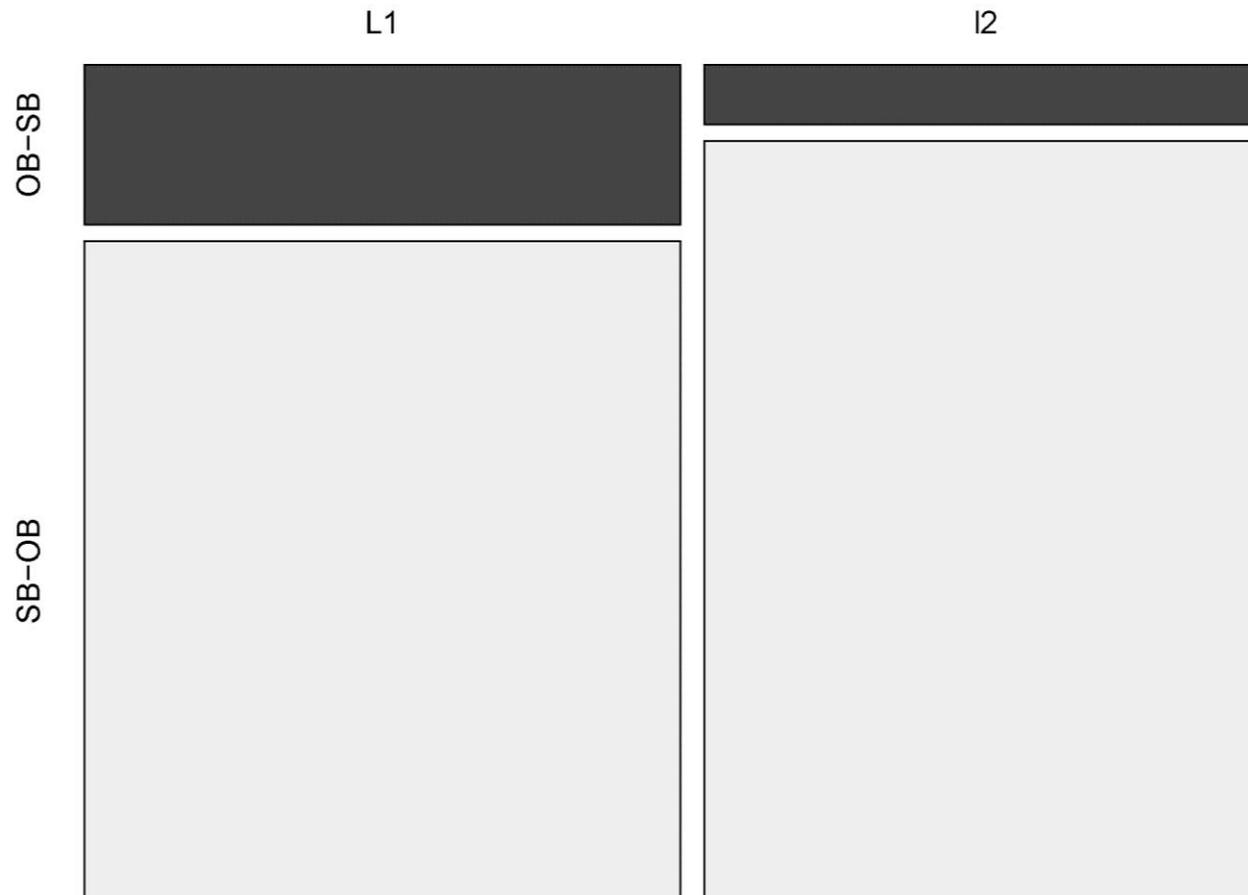
Methode : L1 als Faktor



- Klassische Kontrastive Interlangue-Analyse
 - vergleicht **Abweichungen** zwischen Lernern und Muttersprachlern als **Overuse bzw. Underuse**
- In dieser Arbeit:
 - Muttersprache als weiterer Faktor betrachtet**
 - **Die Haupteffekte sind für die Untersuchung nicht interessant!**
- Wir suchen:
 - Interaktion des Faktors 'Muttersprache' mit weiteren Faktoren.**

Ergebnisse I: χ^2

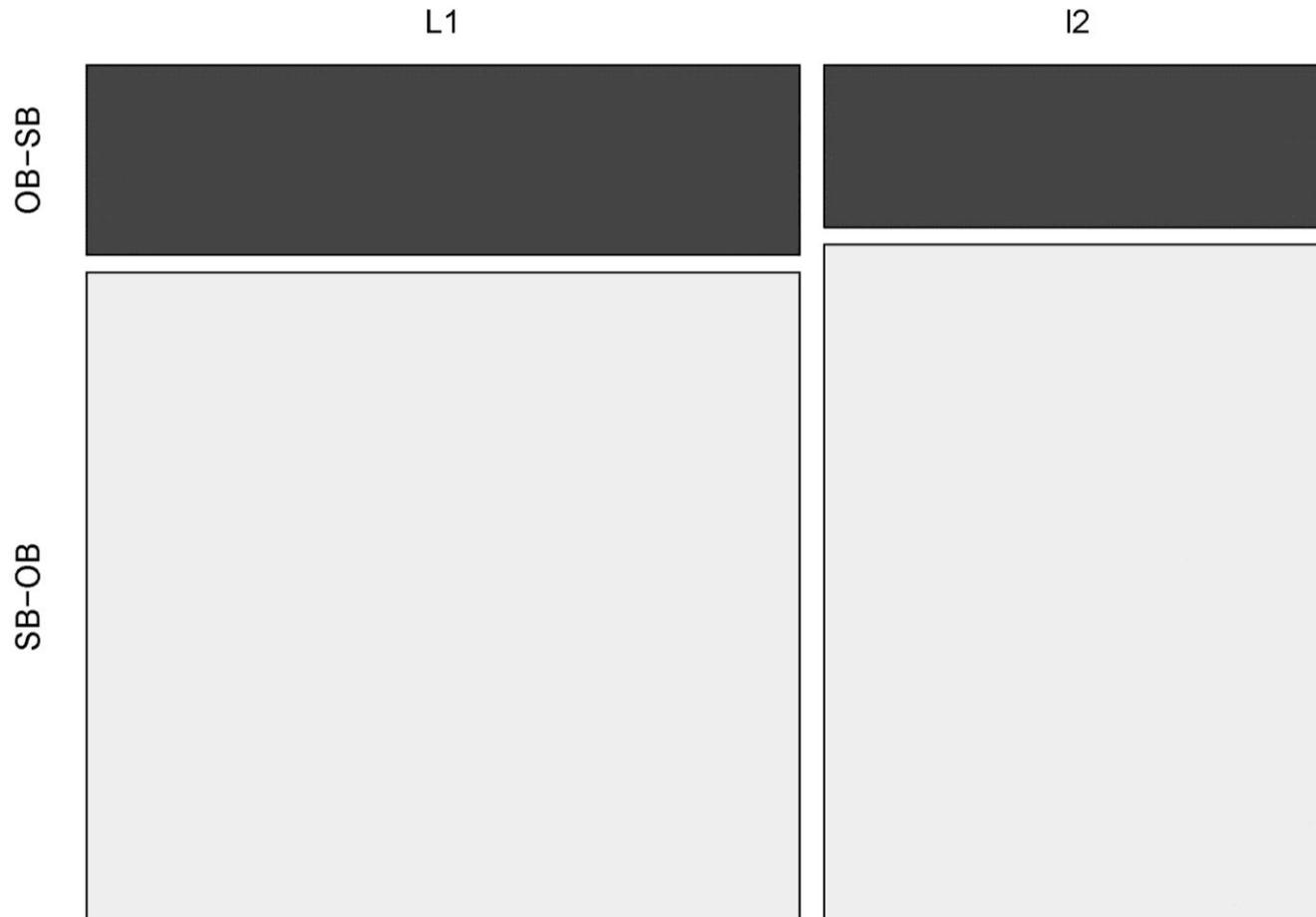
- Lerner benutzen signifikant weniger Objekt-Subjekt Mittelfelder in Nebensätzen (Vend).



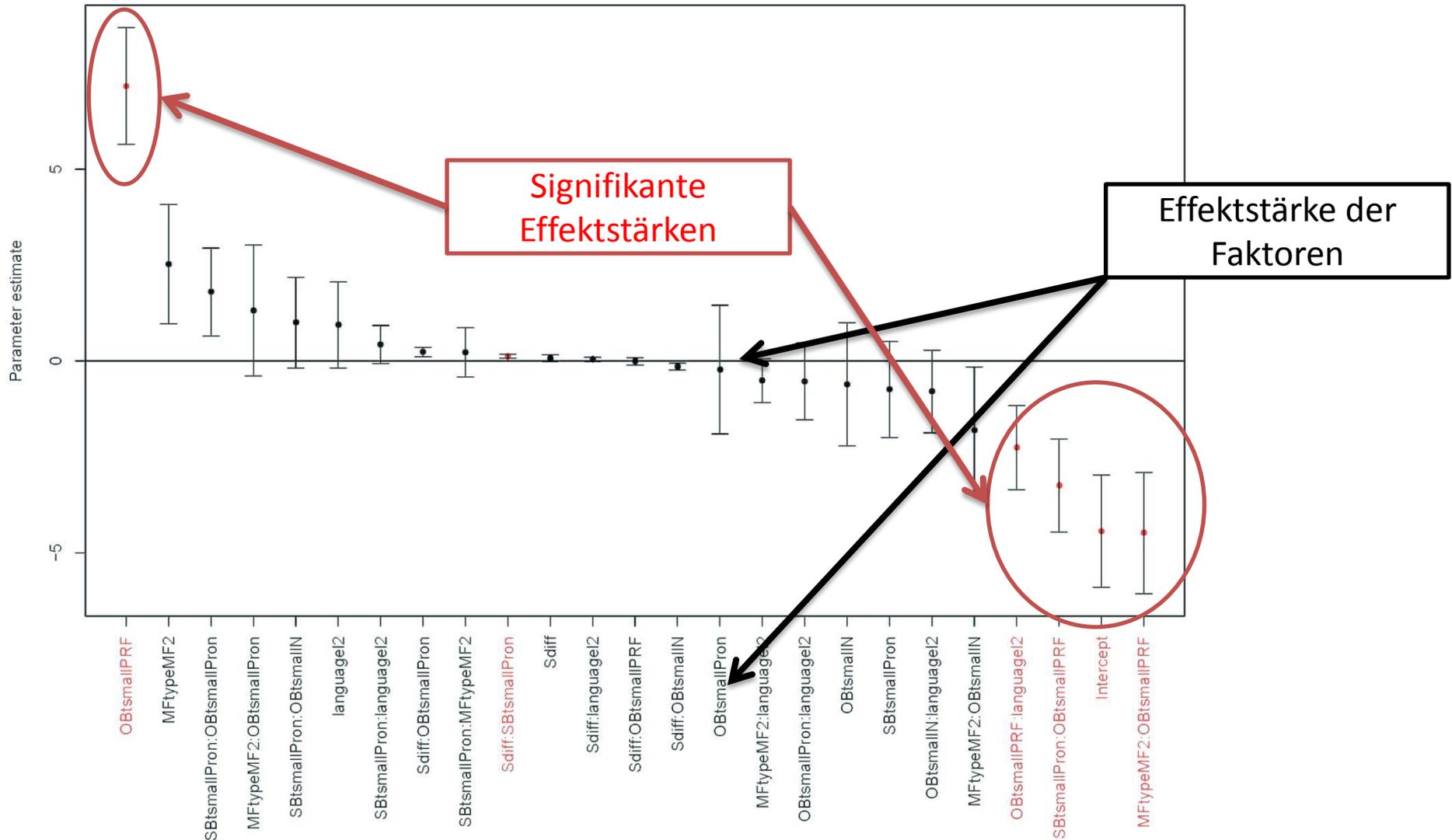
Ergebnisse I: χ^2



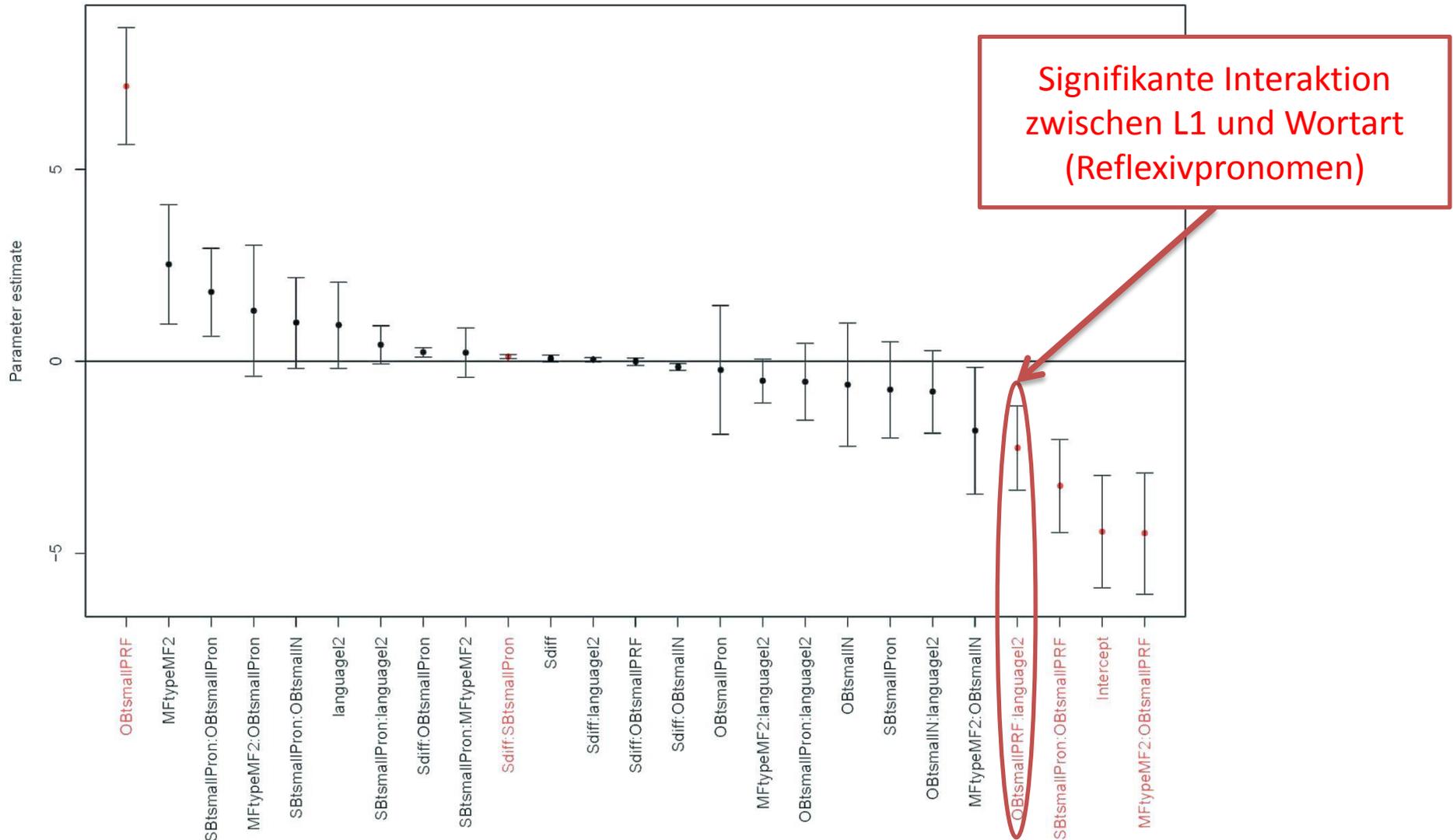
- Das ist aber in Hauptsätzen nicht der Fall.



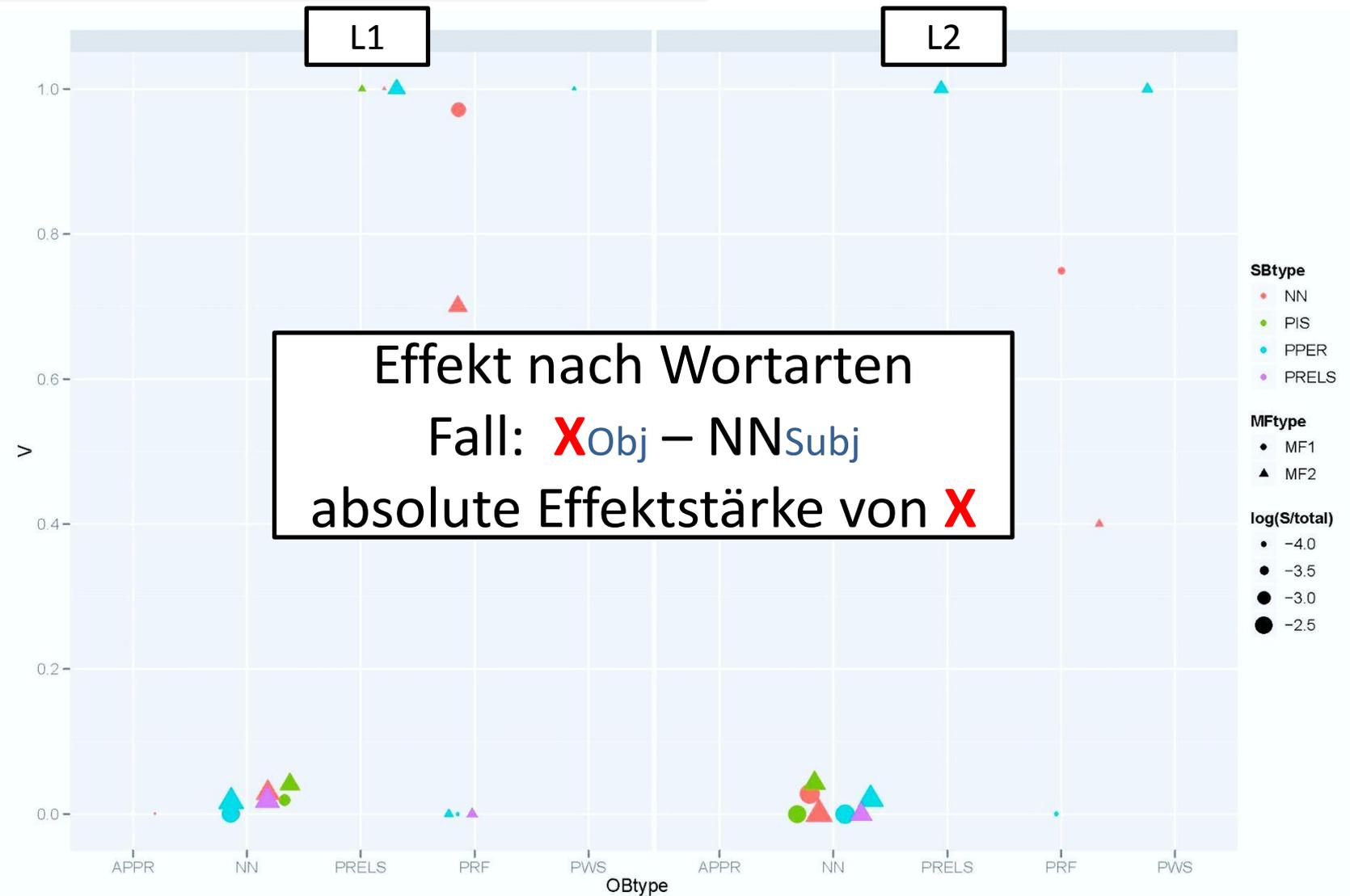
Ergebnisse II: Effekte und Interaktionen



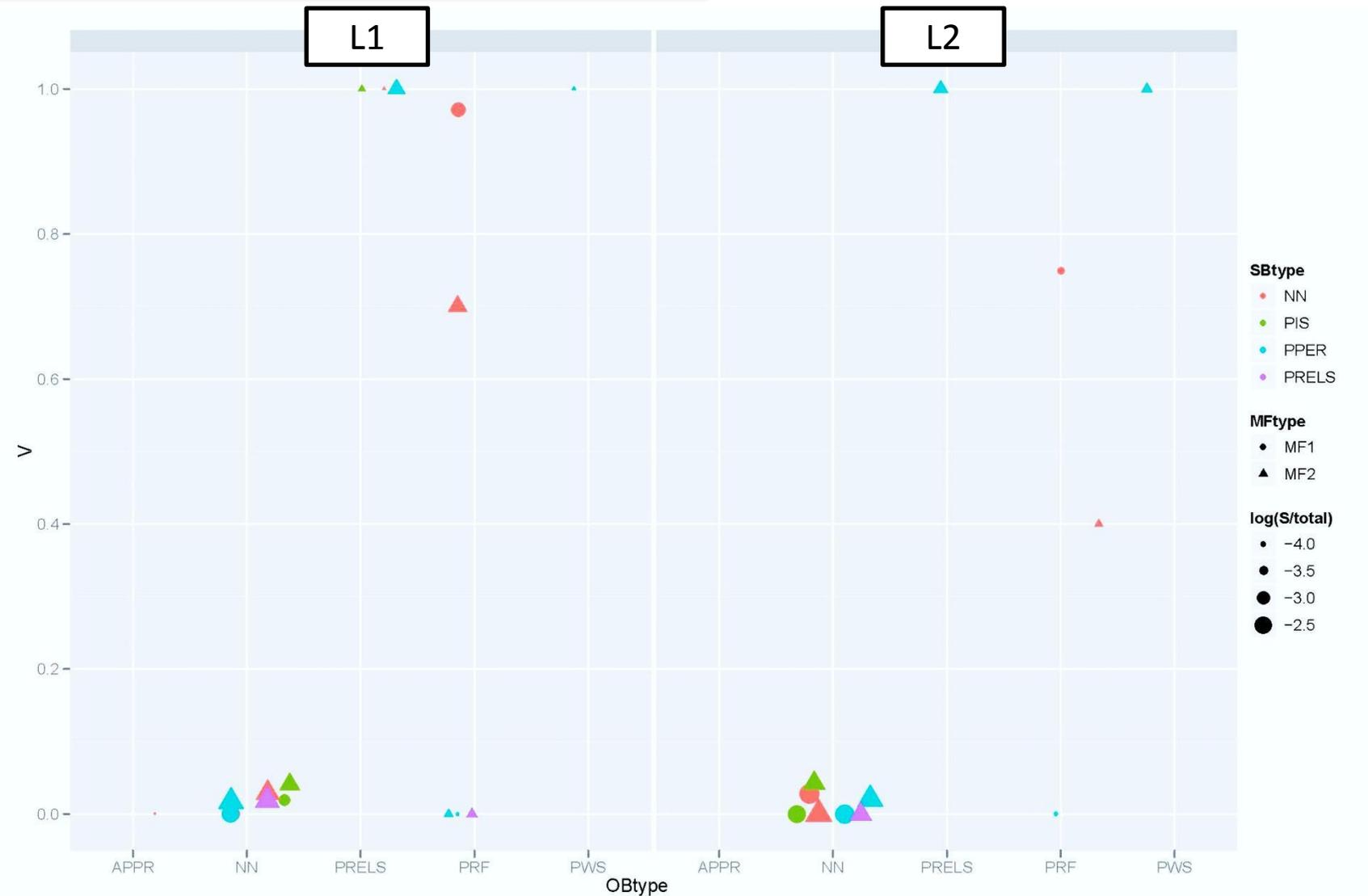
Ergebnisse II: Effekte und Interaktionen



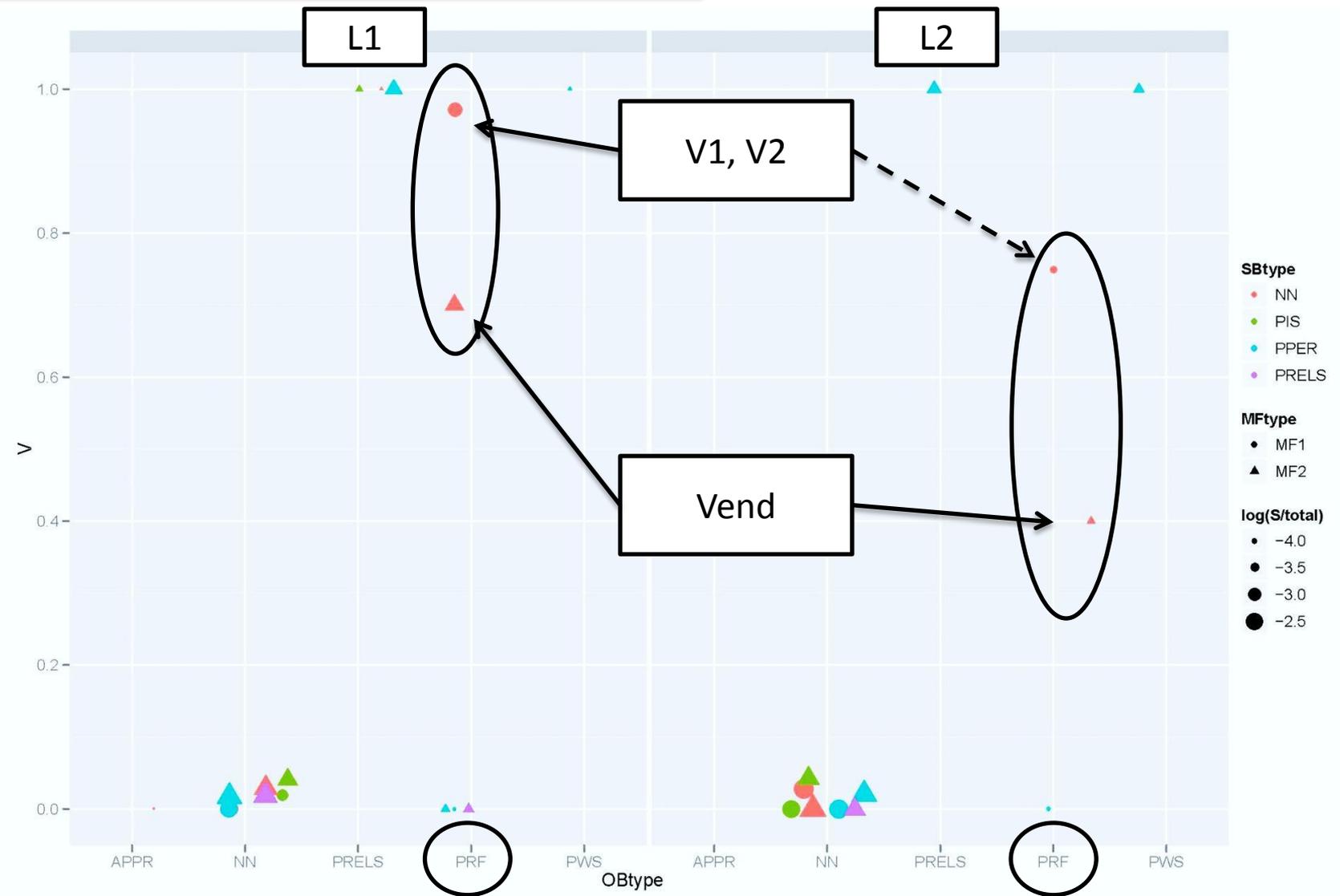
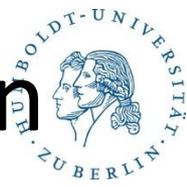
Ergebnisse II: Effekte und Interaktionen



Ergebnisse II: Effekte und Interaktionen



Ergebnisse II: Effekte und Interaktionen



- Die untersuchten Lernertexte zeigen weniger Variation im Gebrauch von Objekt-Subjekt-Mittelfeldern in Nebensätzen.
- Das beruht anscheinend auf einer signifikanten Abweichung hin zu Subjekt-Objekt-Mittelfeldern mit Reflexivpronomen
- **KEIN Effekt** für Kasus oder Konstituentengewicht.
- **Kasus:** Nur sehr wenige Dative in den Daten
- **Gewicht:** Kognitive Belastung → sprachunabhängig
auch in L1-Studien nicht signifikant

VORSICHT!!

- Die Qualität der Parses wurde noch nicht kontrolliert.
 - Automatische Zuweisung von grammatischer Funktion ist auch für Zeitungssprache problematisch
- semi-automatische Korrektur der Parses nötig

Aber Fehler zumindest systematisch über alle Texte

Zusammenfassung:



- Fortgeschrittene Lerner des Deutschen als Fremdsprache zeigen abweichende Variationsmuster für die Abfolge von Verbargumenten im deutschen Mittelfeld
- Das scheint an der nicht-muttersprachlichen **Gewichtung des Einflusses der Faktoren 'grammatische Funktion' und 'Wortart'** im deutschen Mittelfeld zu liegen.

Ausblick:



- Korrektur der Parses
- Einbeziehung weiterer Faktoren
- Ausweitung der Untersuchung auf weitere Konstituententypen im Mittelfeld
 - nicht-nominale Verbargumente
 - Adjunkte

Danke an



Felix Golcher
& die Korplings

- **Bader, Markus; Häussler, Jana (2010):** Word order in German. A corpus study. Exploring the Left Periphery. In: *Lingua* 120 (3), p.717–762.
- **Bates, Douglas; Maechler, Martin; Bolker, Ben (2011):** lme4: Linear mixed-effects models using Eigen and syntax classes. URL: <http://CRAN.R-project.org/package=lme4>
- **Bohnet, Berndt (2010):** Top Accuracy and Fast Dependency Parsing is not a Contradiction. In: *The 23rd International Conference on Computational Linguistics. (COLING 2010)*.
- **Drach, Erich (1937):** Grundgedanken der deutschen Satzlehre. Frankfurt am Main: Diesterweg.
- **Ellis, Rod (2009):** The study of second language acquisition. Oxford [u.a.]: Oxford Univ. Press (= Oxford applied linguistics).
- **Haider, Hubert; Rosengren, Inger (2003):** Scrambling. Nontriggered Chain Formation in OV Languages. In: *Journal of Germanic Linguistics* 15 (03), p.203–267.
- **Heylen, Kris (2005):** A Quantitative Corpus Study of German Word Order Variation. In: Kepser, Stephan; Reis, Marga(eds.): *Linguistic Evidence. Empirical, Theoretical and Computational Perspectives*. Berlin, New York: Mouton de Gruyter (= Studies in generative grammar; 85), p.241–263.
- **Höhle, Tilman N. (1986):** Der Begriff 'Mittelfeld'. Anmerkungen über die Theorie der topologischen Felder. In: Schöne, Albrecht; Stephan, Inge(eds.): *Kontroversen, 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.
- **Kurz, Daniela (2000):** Wortstellungspräferenzen im Deutschen. Master Thesis. Computerlinguistik. Saarbrücken.
- **Lüdeling, Anke; Doolittle, Seanna; Hirschmann, Hagen; Schmidt, Karin; Walter, Maik (2008):** Das Lernerkorpus Falko. In: *Deutsch als Fremdsprache* 45 (2), p.67–73.

- **Pienemann, Manfred (2005):** An introduction to Processability Theory. Parts of this chapter are based on an extended and revised version of my paper "Developmental dynamics in. In: Pienemann, Manfred(ed.): *Cross-linguistic aspects of processability theory*. Amsterdam: Benjamins (= Studies in bilingualism; 30), p.1–73.
- **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. URL: <http://www.linguistik.hu-berlin.de/institut/professuren/korpuslinguistik/forschung/falko> [Stand: 12. Oktober 2010].
- **Schmid, Helmut (1994):** Probabilistic Part-of-Speech Tagging Using Decision Trees. In: *Proceedings of the International Conference on New Methods in Language Processing*, p.44–49.
- **Siewierska Anna (1993):** On the Interplay of Factors in the Determination of Word Order. In: Jacobs, Joachim et al.(eds.): *Syntax*. Berlin, New York: Mouton de Gruyter (= Handbücher zur Sprach- und Kommunikationswissenschaft / Handbooks of Linguistics and Communication Science; 9,1), p.826–846.
- **Uszkoreit, Hans (1987):** Word Order and Constituent Structure in German. Stanford, Calif. (= Center for the Study of Language and Information <Stanford, Calif.>: CSLI lecture notes; 8).
- **Zeldes, Amir; Lüdeling, Anke; Hirschmann, Hagen (2008):** What's hard? Quantitative evidence for difficult constructions in German learner data. In: *Proceedings of QITL 3*. Helsinki.

Alle Quelle am 5.9.2011 geprüft.