



# **THE GERMAN LEARNER MIDDLE FIELD**

## **LINEARISATION-FACTORS OF VERBAL ARGUMENTS IN THE FALKO ADVANCED LEARNER CORPUS**

Marc Reznicek

Humboldt-Universität zu Berlin

Tübingen-Berlin-Meeting  
Universität Tübingen  
05.12.2011

# Overview

---

- Acquiring linguistic variation
- The German middle field
- Variation in the German middle field
- Modeling
- Falko
- Annotation
- Analysis
- Results
- Outlook

# Acquiring linguistic variation

---

- Long tradition of syntax acquisition research (see Ellis 2009)

# Acquiring linguistic variation

---

- Long tradition of syntax acquisition research (see Ellis 2009)
- Focus mainly on acquisition of word order rules and acquisition stages (e.g. Pienemann 2005 )

# Acquiring linguistic variation

---

- Long tradition of syntax acquisition research (see Ellis 2009)
- Focus mainly on acquisition of word order rules and acquisition stages (e.g. Pienemann 2005 )
- Only few studies on acquisition of variation in syntactic patterns

# Acquiring linguistic variation

---

- Long tradition of syntax acquisition research (see Ellis 2009)
- Focus mainly on acquisition of word order rules and acquisition stages (e.g. Pienemann 2005 )
- Only few studies on acquisition of variation in syntactic patterns
- Research question:  
**How do second language learners acquire the competence for using those competing structures?**

# German topological field model

---

- Topological field model for German (Drach 1937, Höhle 1986, Pasch et al. 2003)

prefield	lsb	MF	rsb	post field
Der Feminismus	hat	den Frauen schon immer	geschadet	durch seine Radikalität
<i>The feminism-NOM</i>	<i>has</i>	<i>the women-ACC</i>	<i>damaged</i>	<i>with its radicality</i>

lsb: left sentence bracket

rsb: right sentence bracket

# German topological field model

---

- Topological field model for German (Drach 1937, Höhle 1986, Pasch et al. 2003)
- Verb-Second Rule (V2)

prefield	lsb	MF	rsb	post field
Der Feminismus	hat	den Frauen schon immer	geschadet	durch seine Radikalität
<i>The feminism-NOM</i>	<i>has</i>	<i>the women-ACC</i>	<i>damaged</i>	<i>with its radicality</i>

lsb: left sentence bracket

rsb: right sentence bracket



# Variation in the German middle field

---

- **scrambling:**

Constituents in the middle field allow a variety of competing word orders (Haider/Rosengreen 2003)

<i>dass</i>	[ <b>diese Ansicht</b> ] <sub>AKK</sub> [ <i>in Zukunft</i> ] [ <b>viel mehr Menschen</b> ] <sub>NOM</sub>	<i>zu teilen</i> <i>lernen</i>
<i>that</i>	[ <i>those opinions</i> ] <sub>ACC</sub> [ <i>in the future</i> ] [ <i>a lot more people</i> ] <sub>NOM</sub>	<i>to share learn</i>

(dew07\_2007\_09\_v2.1)

<i>dass</i>	[ <b>viel mehr Menschen</b> ] <sub>NOM</sub> [ <i>in Zukunft</i> ] [ <b>diese Ansicht</b> ] <sub>AKK</sub>	<i>zu teilen</i> <i>lernen</i>
<i>dass</i>	[ <i>in Zukunft</i> ] [ <b>viel mehr Menschen</b> ] <sub>NOM</sub> [ <b>diese Ansicht</b> ] <sub>AKK</sub>	<i>zu teilen</i> <i>lernen</i>

# Factors in middle field word order

---

- Word order is not strictly rule based

# Factors in middle field word order

---

- Word order is not strictly rule based
- A variety of **influencing factors** for word orders have been discussed (e.g. Siewierska 1997, Uszkoreit 1987)

# Factors in middle field word order

---

- Word order is not strictly rule based
- A variety of **influencing factors** for word orders have been discussed (e.g. Siewierska 1997, Uszkoreit 1987)
- **grammatical function**  
subject., dir. object., ind. object
- **case**  
nominative, accusative, dative
- **part-of-speech**  
personal pronoun, full noun, reflexive
- **weight**  
amount of word, amount of syllables

# Factors in middle field word order

---

- Word order is not strictly rule based
- A variety of **influencing factors** for word orders have been discussed (e.g. Siewierska 1997, Uszkoreit 1987)

- **grammatical function**

subject., dir. object., ind. object

- **case**

nominative, accusative, dative

- **part-of-speech**

personal pronoun, full noun, reflexive

- **weight**

amount of word, amount of syllables

- **phrase type**

noun phrase, prepositional phrase, clause

- **semantic role**

agent, patient, recipient

- **information status**

given, new

- **agentivity**

person, institution, animal, material

# Modeling competing factors

---

- Most factors have been looked at one at a time  
(see Kurz 2000, Heylen et al. 2005, Bader/Häusler 2010)

# Modeling competing factors

---

- Most factors have been looked at one at a time  
(see Kurz 2000, Heylen et al. 2005, Bader/Häusler 2010)
- For modeling of simultaneous influence of competing factors

# Modeling competing factors

---

- Most factors have been looked at one at a time  
(see Kurz 2000, Heylen et al. 2005, Bader/Häusler 2010)
- For modeling of simultaneous influence of competing factors
- Possibility I: **Hierarchies**
  - Optimality theory (Uzkoreit 1987)



# Modeling competing factors

---

- Most factors have been looked at one at a time  
(see Kurz 2000, Heylen et al. 2005, Bader/Häusler 2010)
- For modeling of simultaneous influence of competing factors
- Possibility I: **Hierarchies**
  - Optimality theory (Uzkoreit 1987)
- Possibility II: **Relative factor strength analysis**
  - Quantitative analysis (Hoberg 1981, Kurz 2000, Heylen et al. 2005, Bader/Häusler 2010)

# L1 results for news paper articles

---

(Bader & Häusler 2010)

- Grammatical function has a strong effect
  - 96% SB-OB    4% OB-SB

# L1 results for news paper articles

---

(Bader & Häusler 2010)

- Grammatical function has a strong effect
  - 96% SB-OB    4% OB-SB
- Case influences word order in NN-NN combinations

SB – ACC<sub>OBJ</sub> (99%) > SB – DAT<sub>OBJ</sub> (75%)

# L1 results for news paper articles

---

(Bader & Häusler 2010)

- Grammatical function has a strong effect
  - 96% SB-OB 4% OB-SB
- Case influences word order in NN-NN combinations

SB – ACC<sub>OBJ</sub> (99%) > SB – DAT<sub>OBJ</sub> (75%)

- Part-of-Speech has a strong effect
  - pronouns > full nouns

# L1 results for news paper articles

---

(Bader & Häusler 2010)

- Grammatical function has a strong effect
  - 96% SB-OB 4% OB-SB
- Case influences word order in NN-NN combinations

SB – ACC<sub>OBJ</sub> (99%) > SB – DAT<sub>OBJ</sub> (75%)

- Part-of-Speech has a strong effect
  - pronouns > full nouns
- Constituent-weight has no effect

# Research Question:

---

Do second language learner texts show a difference in effect strength for those factors than native speaker texts?

# Research Question:

---

Do second language learner texts show a difference in effect strength for those factors than native speaker texts?

- Contrastive Interlanguage analysis CIA (Granger 2008)
  - Assumption
    - learner language is systematic
    - variation in the group
    - transfer & generell language acquisition processes

# Data : Falko learner corpus of German



Lüdeling et al. 2008

- advanced learners of German B1+
- essays and summaries
- cross-sectional & longitudinal data
- ~260.000 tokens, growing
- automatically annotated POS, lemma  
(Treetagger, Schmid 1994)
- dependency parsed (NEW) (Bohnet 2010)



# Data : Falko learner corpus of German



Lüdeling et al. 2008

- advanced learners of German B1+
- essays and summaries
- cross-sectional & longitudinal data
- ~260.000 tokens, growing
- automatically annotated POS, lemma  
(Treetagger, Schmid 1994)
- dependency parsed (NEW) (Bohnet 2010)

## sub set used

- 94 texts learners of German (25 L1s)
- 94 text German controll group

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

# Data : Target hypotheses

---

Non-canonical syntactic structures in learner texts (LT) make a description with standard grammars impossible.

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

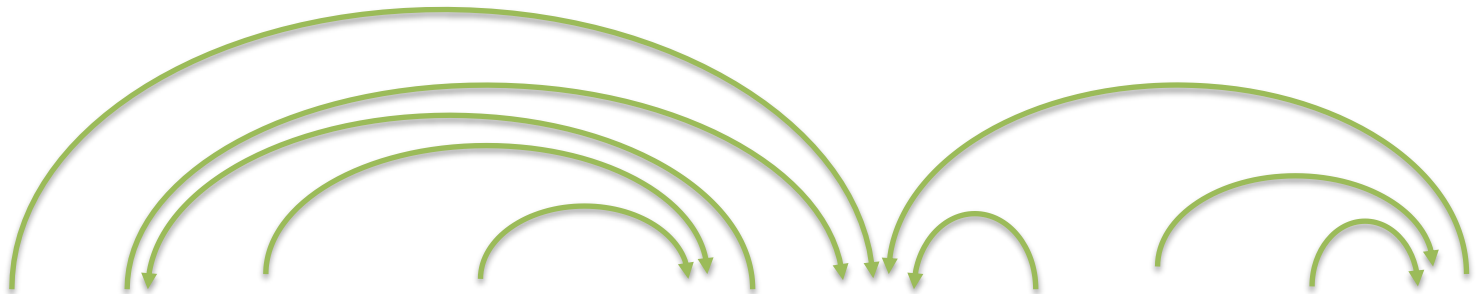
(FalkoEssayL2v2.0:fk006\_2006\_08)

*But unfortunately such percentages define the value of universities.*

# Data : Target hypotheses

---

Therefore a minimal grammatical correction (TH1) is explicitly included into the corpus (Reznicek et al. 2009)



**TH1:** *Aber in **den** meisten Fällen **ist** **das** nicht der Fall.*

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

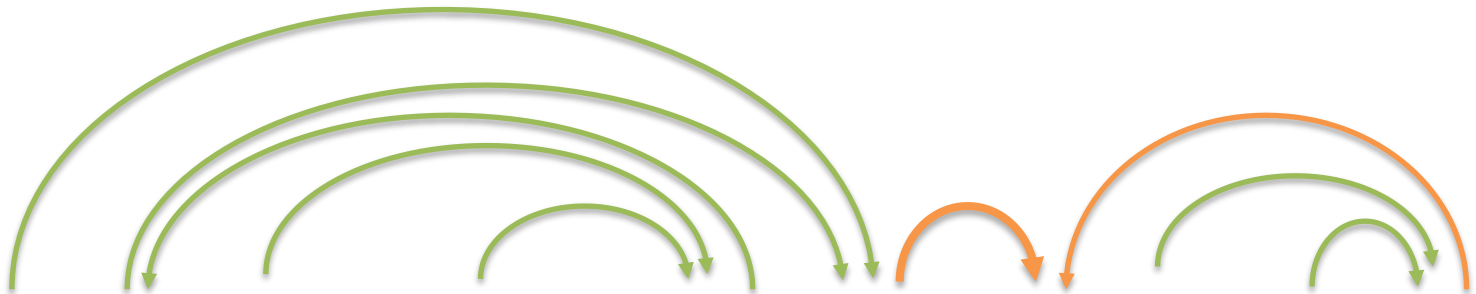
(FalkoEssayL2v2.0:fk006\_2006\_08)

But in the-FEM most cases-MASC that is not the case.

# Data : Target hypotheses

---

To conserve the original word order, dependencies are mapped back on original **sites**.



**TH0:** Aber in **den** meisten Fällen **das** ist nicht der Fall.

**TH1:** Aber in **den** meisten Fällen **ist das** nicht der Fall.

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

But in the-FEM most cases-MASC that is not the case.

# Data : Target hypotheses

---

*TH0: Aber in **den** meisten Fällen **das** ist nicht der Fall.*

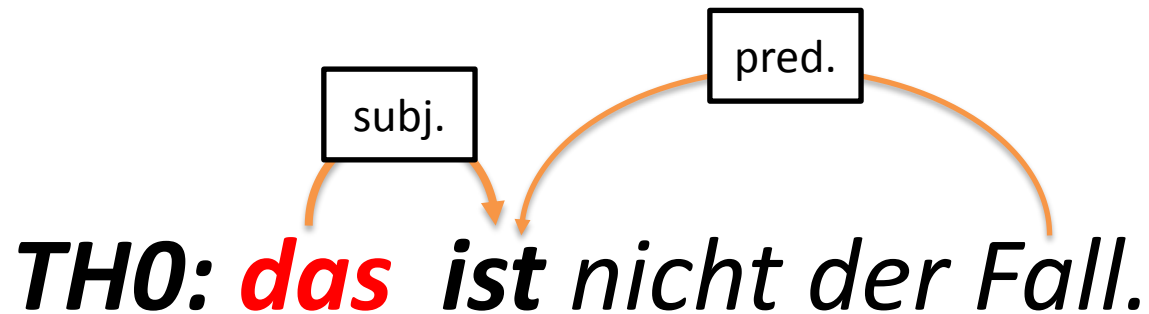


But in the-FEM most cases-MASC that is not the case.

# Data : Target hypotheses

---

Each dependency is automatically labeled with the sentence function.



...that is not the case.

# Annotation : middle fields

---

- In all utterances the middle fields have been manually annotated.

# Annotation : middle fields

---

- In all utterances the middle fields have been manually annotated.
- For each middle field following information has been extracted



# Annotation : middle fields

---

- In all utterances the middle fields have been manually annotated.
- For each middle field following information has been extracted
- Only for verb arguments
  - 1) clause type** (main clause, subordinate clause)
  - 2) verb argument order** (obj-sub, sub-obj)
  - 3) part-of-speech** (noun, pron, prf, prep)
  - 4) case** (nom, acc, dat)
  - 5) length of constituent in tokens**
  - 6) length of constituents in syllables**

# method: linear mixed effect model

---

linear mixed effect model to calculate the effect strength of different factors:

(Bates et al. 2011)

$$z = \beta_0 + \beta_1 \mathbf{x}_1 + \beta_2 \mathbf{x}_2 + \beta_3 \mathbf{x}_3 + \dots + \beta_k \mathbf{x}_{k+1}$$

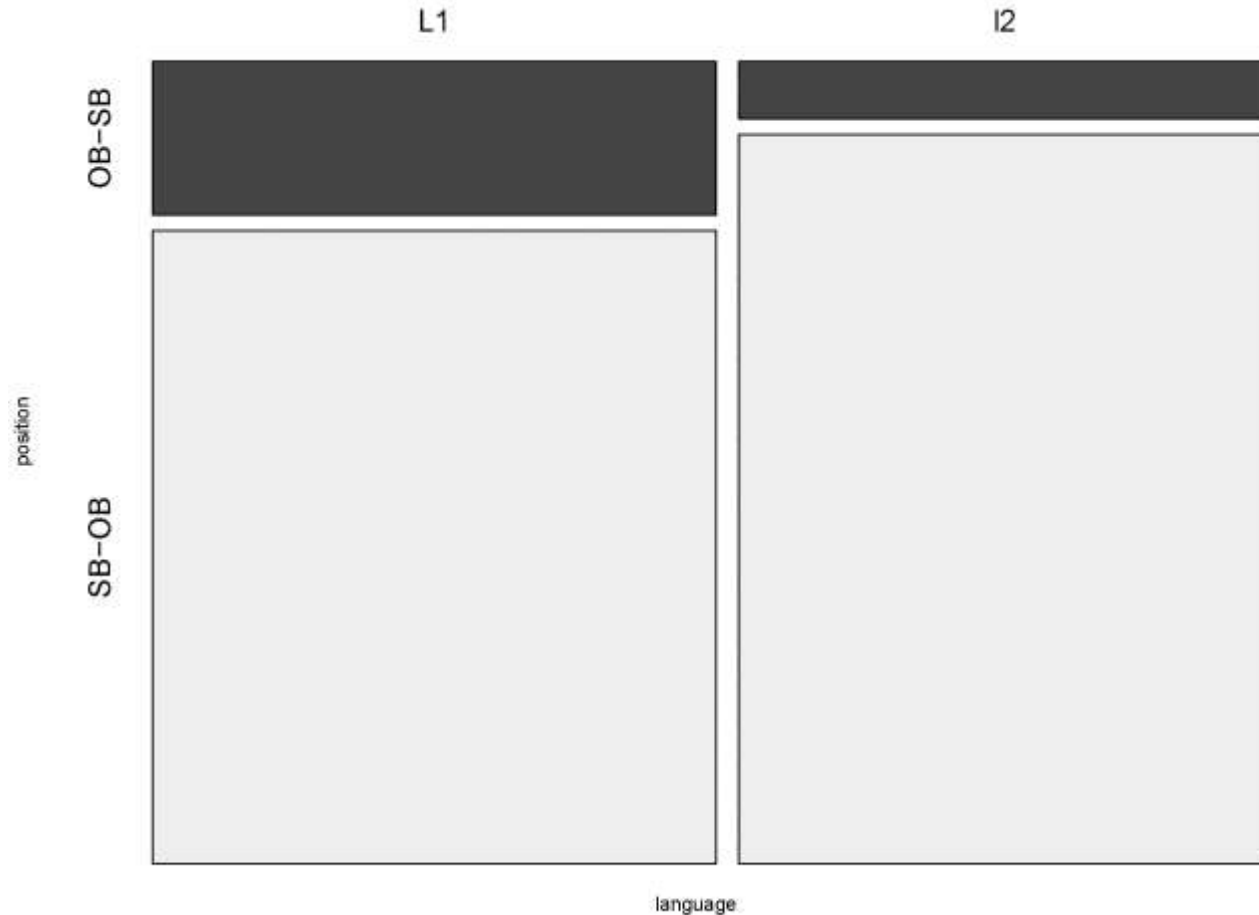
variable                      effect strength  
↓                                      ↓

→ probabilities for OB-SB-order with subject as full noun  
random effects: verb, text

# results I: $\chi^2$

---

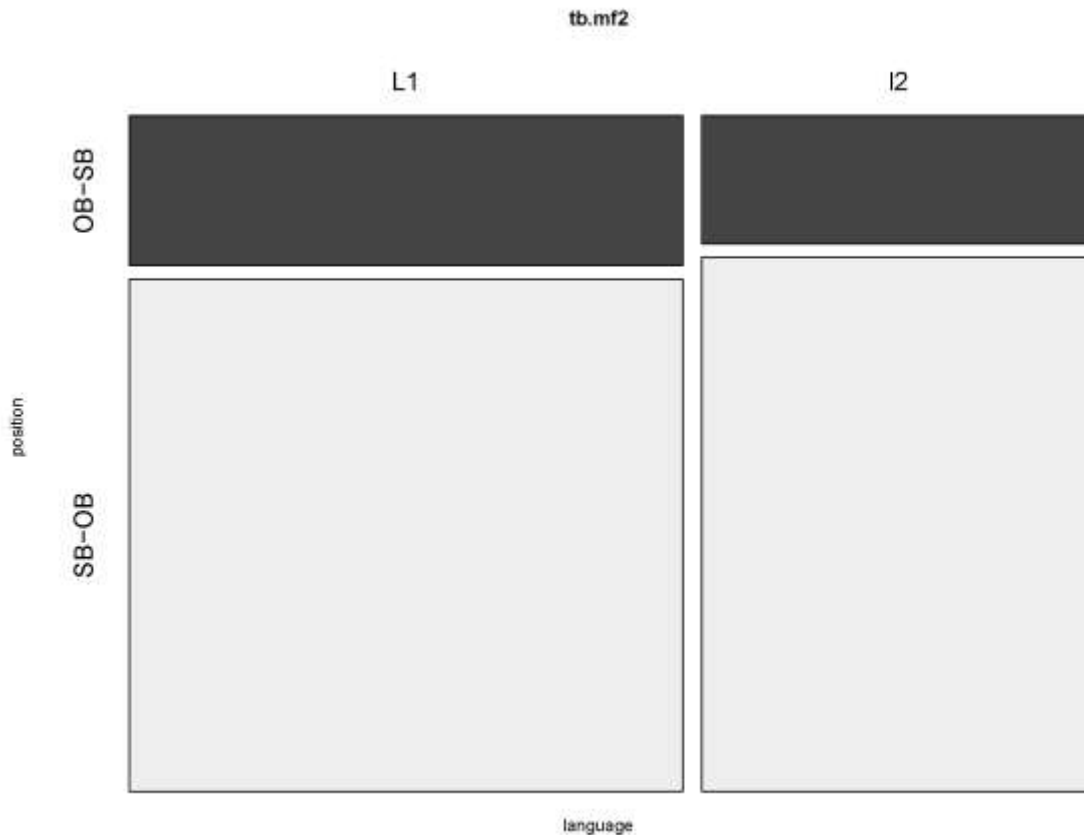
- Learners use significantly less object-subject middle fields in subordinate clauses



# results I: $\chi^2$

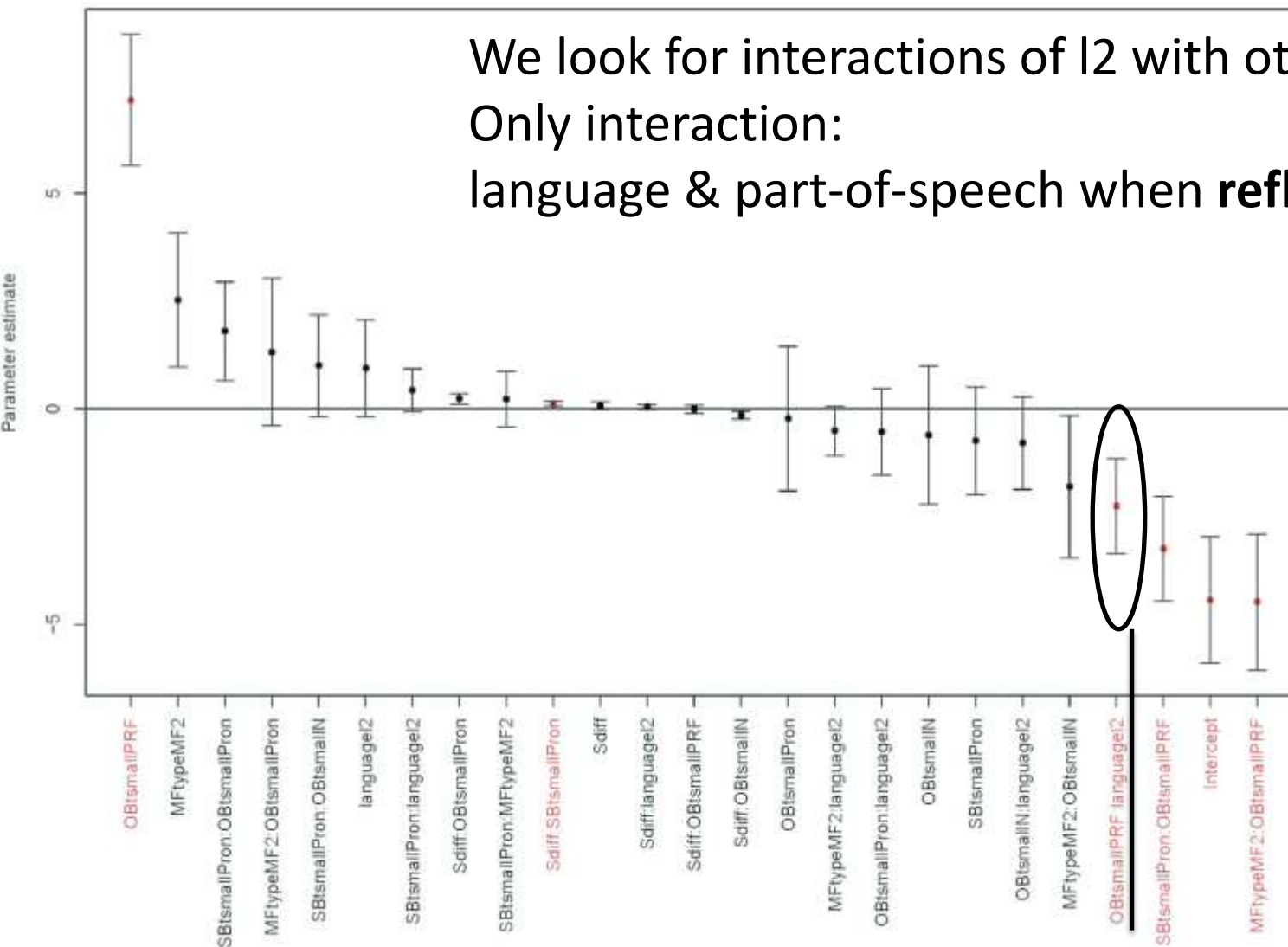
---

- Interestingly this is not the case in main clauses

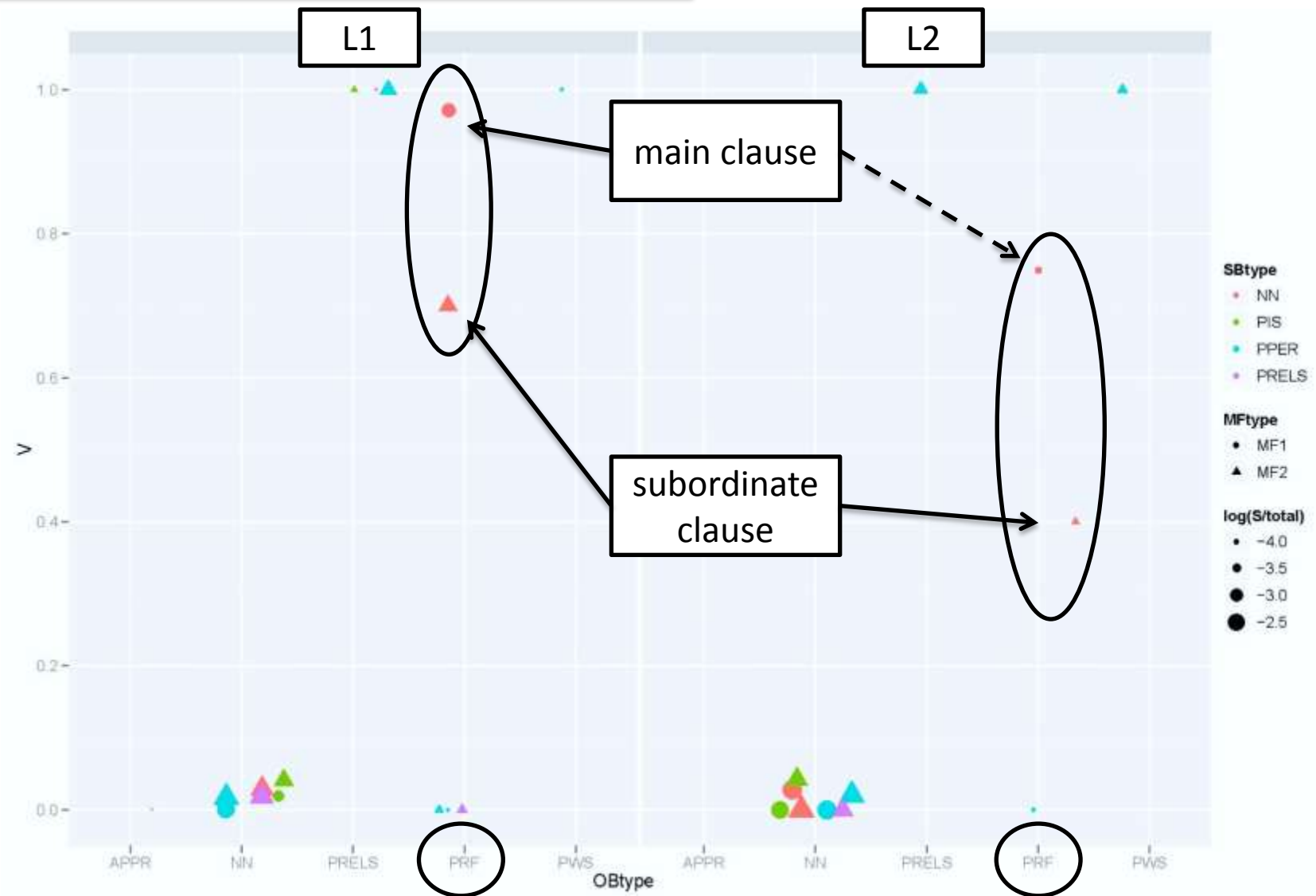


# results II: effects & interactions

We look for interactions of I2 with other factors  
Only interaction:  
language & part-of-speech when **reflexive pronoun**



# results II: effects & interactions



# discussion

---

- The learners in this study have shown less variation in the use of SB-OB-type subordinate clauses.
- This seems to mainly come from a significant bias of SB-OB-type clauses for reflexive pronouns.

# discussion

---

- The learners in this study have shown less variation in the use of SB-OB-type subordinate clauses.
- This seems to mainly come from a significant bias of SB-OB-type clauses for reflexive pronouns.
- **NO** effect found for case, weight.
- **case:** Too few datives in the data.
- **weight:** cognitive load → language independent



# discussion

---

- Quality of the parses have NOT been controlled.
  - Automatic edge-labeling quality is known to be problematic even for news paper text
- semi-automatic correction of parses will be necessary

# summary and outlook

---

- Advanced learners of German show different patterns of variation linked to the verb argument order in the German middle field
- This seems to be due to a non-native like weight of the factors 'sentence function' and 'part-of-speech' as influence of argument order

# summary and outlook

---

- Advanced learners of German show different patterns of variation linked to the verb argument order in the German middle field
- This seems to be due to a non-native like weight of the factors 'sentence function' and 'part-of-speech' as influence of argument order

Next step:

- more semantic and pragmatic factors

# Thanks to

---

Felix Golcher  
Berlin corpus linguistics team

# bibliography

---

- **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 S4 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.

# bibliography

---

- **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.

all sources checked on 09-05-2011.