

Hagen Hirschmann
hirschhx@hu-berlin.de



**KORPUSGESTEUERTE SYNTAXANALYSEN
VON LERNERSPRACHE
-PRÄPOSITIONALOBJEKTE IM DAF-ERWERB-**

DGfS-Jahrestagung 2013
Universität Potsdam

Plan



- Präpositionalobjekte im DaF
- Daten und Methodische Vorüberlegungen
- Korpusanalysen
 - Fehleranalyse
 - CIA-Studie
- Fazit

Präpositionalobjekte - Spezifik



- Anzahl und Art der Ergänzungen ist mit Verblexemen assoziiert
- Einige Ergänzungen korrelieren mit einer spezifischen Semantik
(Fabricius-Hansen 2010, S. 221f)
- Aber: Präpositionalobjekte sind nicht antizipierbar und müssen als spezifische Ergänzungen gespeichert werden
- Präpositionalobjekte stellen eine besondere Herausforderung an den Lernprozess dar

Fragestellung



- *Wie zielsprachlich ist die Verwendung von Präpositionalobjekten bei fortgeschrittenen Lernenden des DaF?*

Untersuchungsmethode



- Methode: Korpusuntersuchungen
- 1. Fehlerstudie: *Wie häufig werden bei der Verwendung von P-Objekten Fehler produziert?*
- 2. Kontrastive Studie: *Wie viele grammatische P-Objekte werden von den Lernenden verwendet (quantitativ/qualitativ)?*

Rechtfertigung: Auswertung von Lernerkorpora



- EA (Error Analysis) vs. CIA (Contrastive Interlanguage Analysis), vgl. z. B. Granger 2008
 - EA:
 - Fehler=Abweichungen von der Zielsprache
 - Immer im Hinblick auf Zielstruktur (Zielhypothese)
 - 'Misuse'
 - CIA:
 - Vergleich **zielsprachlicher** Strukturen in L2 und L1
 - 'Overuse'/'Underuse'

Datengrundlage: Falko



- "Fehlerannotiertes Lernerkorpus"
- Handbuch: Reznicek et al. 2012
- frei nutzbar (ANNIS-Suchinterface) unter <https://korpling.german.hu-berlin.de/falko-suche/>
- Untersuchte Korpora: Essaykorpora L2 und L1
- alle Lerner fortgeschritten und aus gesteuertem Erwerb (144619 Token)
- Verschiedenste Muttersprachen → Metadaten; Trennung von Subkorpora möglich
- Vergleichskorpus mit L1-Sprechern (70615 Token)
- Annotationen: pos, Zielhypothesen, Dependenzparses

Hypothesen



- Gebrauch von P-Objekten bestimmt durch wenig Fehler (Präposition ist mit Lexem gespeichert; Lerner sind fortgeschritten)
- Gebrauch quantitativ vergleichbar mit L1
- Lexikalische Dichte abweichend:
Häufige P-Objekte übergebraucht, aber insgesamt weniger Lexeme (nicht erworbene Verbindungen können nicht ad hoc produziert werden)

Suchen nach Präpositionalobjekten in Korpusdaten



- pos-Tags → Präpositionen
(STTS; Teufel et al. 1999)
- Ohne syntaktische Analyse keine
Disambiguierung von P-Objekten und
P-Adjunkten

Suchen nach Präpositionalobjekten in Korpusdaten

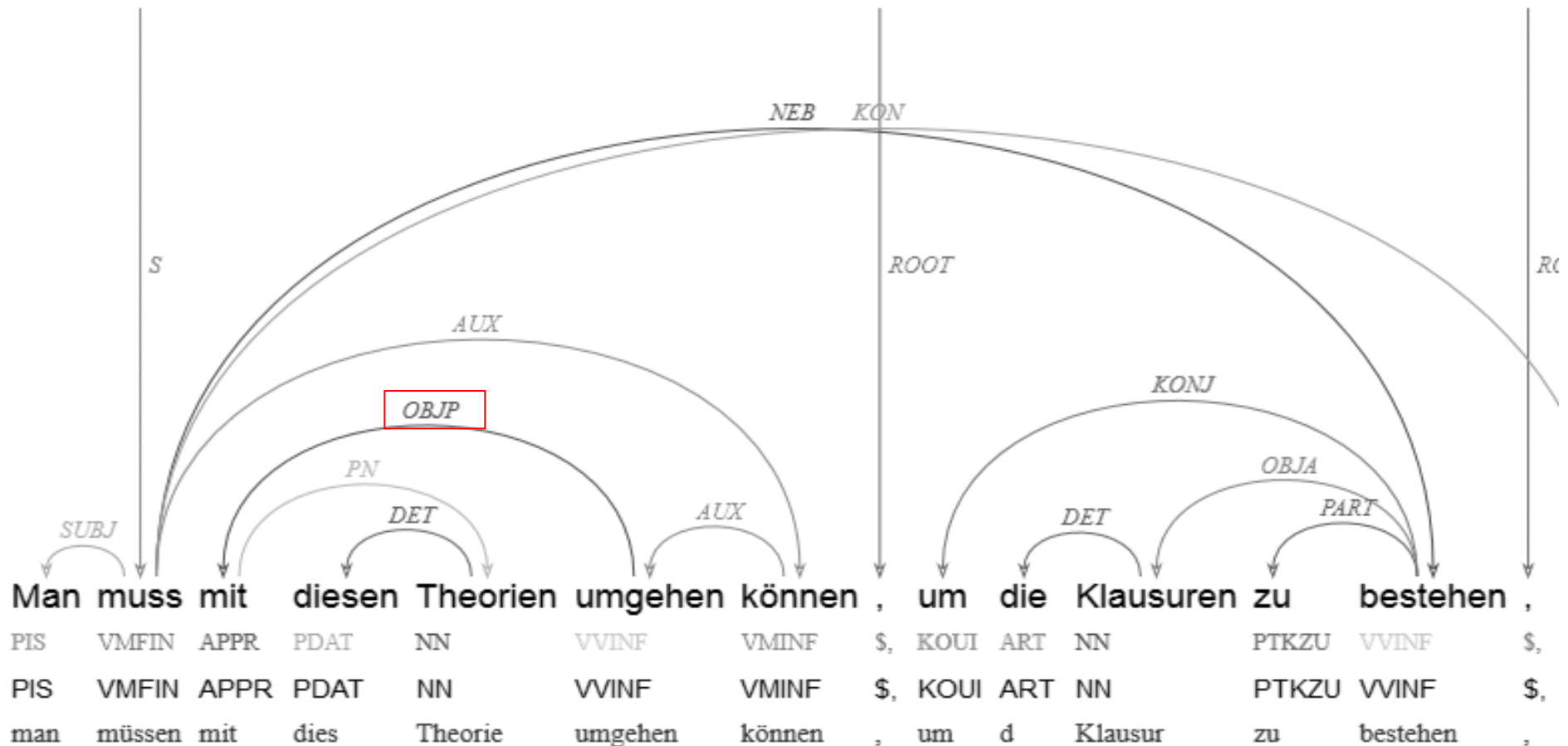


- Nutzung von Dependenzpareses
 - Schema: Foth 2006
 - Automatische Pareses mit Malt (<http://www.maltparser.org/>)
 - Evaluation, manuelle Korrekturen, Erstellung Goldstandard (Ines Rehbein)

Parsing: Beispiel



- Falko Essays L2, cbs001_2006_09



Lernerdaten: Konzeptionelle Probleme



- *Z. B. hat man oft über Greenpeace gehört*
(cbs001_2007_10)
- *Sie haben sich dazu gewöhnt (...)* (cbs014_2007_10)
- *Viel mehr achtet der Arbeitgeber , ob der Student ,
die relevante Arbeitserfahrung hat* (cbs006_2007_10)
- *Da die Studenten einen grossen Teil ihres Studiums
an die Theorien wittmen muss (...)* (cbs011_2006_09)
- *Man denke an den unterschiedlichen Gruppen (...)*
(cbs001_2007_10)

Lernerdaten: Konzeptionelle Probleme



- Grammatisch: Status 'P-Objekt' bei ungrammatischen Strukturen interpretationsabhängig
 - "Eigentlich wollte der Lerner..."
- Technisch: Ungrammatische Strukturen von Parser häufig nicht als P-Objekte erkannt

Zwischenfazit



- P-Objekte nur dort an Oberfläche messbar, wo Daten sich zielsprachlich verhalten
- Trennung: Zielsprachliche Äußerungen vs. abweichende Äußerungen
(→ Zielstruktur benötigt: „Zielhypothese“)
- In abweichenden Fällen zu überprüfen:
 - Verhältnis zwischen Lerneräußerung und Zielstruktur:
 - P-Objekt hinzugefügt
 - P-Objekt gelöscht
 - Präposition ausgetauscht

Falko-Methodik: Zielhypothesen



- Zielhypothesen als Vorverarbeitungsschritt: nicht zielsprachliche Strukturen werden in zielsprachliche Strukturen überführt
- Richtlinien: [Reznicek et al. 2012](#)
- Tokenbezogen
- Automatische Einfügung von Edit-Tags: „INS“, „DEL“, „CHA“, „MOVE“
- Zwei Ebenen/Zielhypothesen (Grammatik vs. Ausdruck/Stil)
- Relevant für P-Objekte: Zielhypothese 1

Zielhypothesen - Beispiele



1. Typ: Falsche Präposition

- Ergänzung inhaltlich korrekt, formal fehlerhaft

„CHA“

tok	Sie	haben	sich	dazu	gewöhnt
ZH1	Sie	haben	sich	daran	gewöhnt
ZH1Diff				CHA	
ZH1lemma	Sie sie	haben	er es sie	daran	gewöhnen
ZH1pos	PPER	VAFIN	PRF	PAV	VVPP

cbs014_2007_10

Zielhypothesen - Beispiele



2. Typ: Präposition fehlt

- Verb erfordert Ergänzung, die nicht realisiert wird

„INS“

tok	Beharrt	sie		,	bleibt	sie	Single
ZH1	Beharrt	sie	darauf	,	bleibt	sie	Single
ZH1Diff			INS				
ZH1lemma	beharren	sie	darauf	,	bleiben	sie	Single
ZH1pos	VVFIN	PPER	PAV	\$,	VVFIN	PPER	NN

cbs009_2006_09

Zielhypothesen - Beispiele



3. Typ: Präposition ist ungrammatisch

- Ergänzung inhaltlich korrekt, formal fehlerhaft

„DEL“

tok	Da	die	Studenten	einen	grossen	Teil	ihres	Studiums	an	die	Theorien	wittmen	muss
ZH1	Da	die	Studenten	einen	großen	Teil	ihres	Studiums		den	Theorien	widmen	müssen
ZH1Diff					CHA				DEL	CHA		CHA	CHA
ZH1lemma	da	d	Student	ein	groß	Teil	ihr	Studium		d	Theorie	widmen	müssen
ZH1pos	KOUS	ART	NN	ART	ADJA	NN	PPOSAT	NN		ART	NN	VVINF	VMINF

cbs011_2006_09

Zielhypothesen - Beispiele



4. Typ: Präpositionsergänzung im falschen Kasus
- Präpositionalobjekt korrekt, Subsystem der präpositionalen Rektion fehlerhaft

„CHA“
an Artikel/Adjektiv

tok	Man	denke	an	den	unterschiedlichen	Gruppen	
ZH1	Man	denke	an	die	unterschiedlichen	Gruppen	,
ZH1Diff				CHA			
ZH1lemma	man	denken	an	d	unterschiedlich	Gruppe	,
ZH1pos	PIS	VVFIN	APPR	ART	ADJA	NN	\$,

cbs001_2007_10

Fehleranalyse



- Fehleranalyse anhand Diskrepanz zwischen ZH1 und Lerneräußerung
- Fehlerkategorien oberflächenorientiert; Zuordnung aus Perspektive der Lerneräußerung:
 1. „Präp. falsch gewählt“ (ZH1Diff=„CHA“)
 2. „Präp. fehlt“ (ZH1Diff=„INS“);
 3. „Präp. fehlerhaft; Verb nimmt kein P-Objekt“ (ZH1Diff=„DEL“)
 4. „Präp. Ergänzung fehlerhaft“ (ZH1Diff=„CHA“)

Fehleranalyse - Ergebnisse



1. Falsche Präpositionen: 50
→ 11% pro P-Objekte insgesamt
2. Hinzugefügte P-Objekte: 33
→ 7% pro P-Objekte insgesamt
3. Getilgte P-Objekte: 33
→ 7% pro P-Objekte insgesamt
4. Falsche Kasus an der Nomen-Ergänzung: 32
→ 7% pro P-Objekte insgesamt

Fehleranalyse - Ergebnisse



- Häufigster Fehler: falsche Präposition
- Ca. jedes zehnte P-Objekt mit falscher Präposition (falsche Form)
- Alle Fehlertypen:
32% aller P-Objekte sind fehlerhaft
- Die Anzahl der fälschlich gesetzten P-Objekte ist gleich der Anzahl der fehlenden P-Objekte

Kontrastive Analyse



- CIA-Studie berücksichtigt diejenigen Stellen, an denen ZH1 und Lerneräußerung identisch sind
- Alle Präpositionalobjekte auf ZH1 abzüglich der Fälle mit Edit-Tags

Kontrastive Analyse - Ergebnisse



- Anzahl der auf ZH1 vorhandenen P-Objekte:
L2=3,52 / 100 VFIN
L1=3,70 / 100 VFIN
- Type-Token-Ratio:
L1=0,69 (184 Typen, 266 Token)
L2=0,50 (231 Typen, 462 Token)
L2 bei 266 Token=0,47 (124 Typen)

Präpositionalobjekte: Ergebnisse + Fazit



- Fehleranalyse: deutlicher Erwerbseffekt, trotz Fortgeschrittenheit
- Kontrastive Studie: Hypothesen bestätigt (gleichhäufiger Gebrauch; weniger Kandidaten)
- P-Objekte sind auch noch im fortgeschrittenen Bereich ein Erwerbsproblem
- Erklärung für hohe Fehlerquote: Präpositionalobjekte können nicht wie Adjunkte vermieden werden (im Adjunktbereich häufig Mindergebrauch als Reflex von Schwierigkeiten, bei P-Objekten sind Fehler der Reflex)

Methodik: Fazit



- Kopplung EA und CIA nötig
- Sowohl EA als auch CIA angewiesen auf Zielhypothesen (ZH)
- ZH oder ähnliche Konzepte immer dann erforderlich, wenn sprachl. Strukturen nicht mit gegebener Grammatik beschreibbar sind
- Auch andere Nichtstandardvarietäten benötigen ZH-Konzept, um auf Standard bezogen werden zu können



Herzlichen Dank

Kontakt: hirschhx@hu-berlin.de

Referenzen



- Fabricius-Hansen, Cathrine (2010): Syntax. In: Krumm, Hans-Jürgen; Fandrych, Christian; Hufeisen, Britta; Riemer, Claudia (Hg.): Deutsch als Fremd- und Zweitsprache. Ein internationales Handbuch. 1. Halbband. Berlin/New York; Mouton deGruyter.
- Foth, Kilian A. (2006): Eine umfassende Constraint-Dependenz-Grammatik des Deutschen. Tech. Rep. Universität Hamburg.
- Granger, Sylviane (2008): Learner corpora. In: Anke Lüdeling & Merja Kytö (eds): Corpus Linguistics. An International Handbook. Vol 1. Berlin: Mouton deGruyter, 259-275.
- Reznicek, Marc; Lüdeling, Anke; Krummes, Cedric; Schwantuschke, Franziska; Walter, Maik; Schmidt, Karin; Hirschmann, Hagen; Andreas, Torsten (2012): Das Falko-Handbuch. Korpusaufbau und Annotationen Version 2.01. Tech. Rep. Humboldt-Universität zu Berlin.
- Schiller, Anne; Teufel, Simone; Stöckert, Christine; Thielen, Christine (1999): Guidelines für das Tagging deutscher Textcorpora mit STTS. Tech. Rep. Institut für maschinelle Sprachverarbeitung, Stuttgart.

Suchanfragen



- Interface: <https://korpling.german.hu-berlin.de/falko-suche/>
- Korpora:
 - L2: FalkoEssayL2v2.4
 - L1: FalkoEssayL2v2.4
- Anfragen:
 - P-Objekte auf ZH1:
`ZH1 & ZH1 & #1 ->dep[func="OBJP"] #2`
 - P-Objekte, die im Originaltext durch eine falsche Präposition repräsentiert werden:
`ZH1 & ZH1 & #1 ->dep[func="OBJP"] #2 & ZH1Diff="CHA" & #2 _ _ #3`
 - Ersetze "CHA" mit "INS", um Belege für im Originaltext fehlende Pn zu erhalten
 - "Falsche" P-Objekte die auf ZH1 gelöscht sind:
`ZH1Diff="DEL" & pos="APPR" & #1 _ _ #2 & tok & #2 . #3 & ZH1 & ZH1 & #4`
`->dep[func=/O.*/] #5 & #3 .* #5`
 - Falsche Kasus bei den nominalen Ergänzungen der Präpositionen:
`ZH1 & ZH1 & ZH1 & ZH1 & ZH1Diff="CHA" & #1 ->dep[func="OBJP"] #2 & #2 ->dep[func="PN"]`
`#3 & #3 ->dep[func=/((DET|ATTR)/)] #4 & #4 _ _ #5`