



# „Dass wenn man etwas will, muss man dafür arbeiten“ - Zielhypothesen im Lernerkorpus Falko<sup>1</sup>.

Marc Reznicek+, Cedric Krummes\*, Hagen Hirschmann+, Anke Lüdeling+, Astrid Ensslin\*, Jia Wei Chan+, Amir Zeldes+, Thomas Krause+ und Florian Zipser+

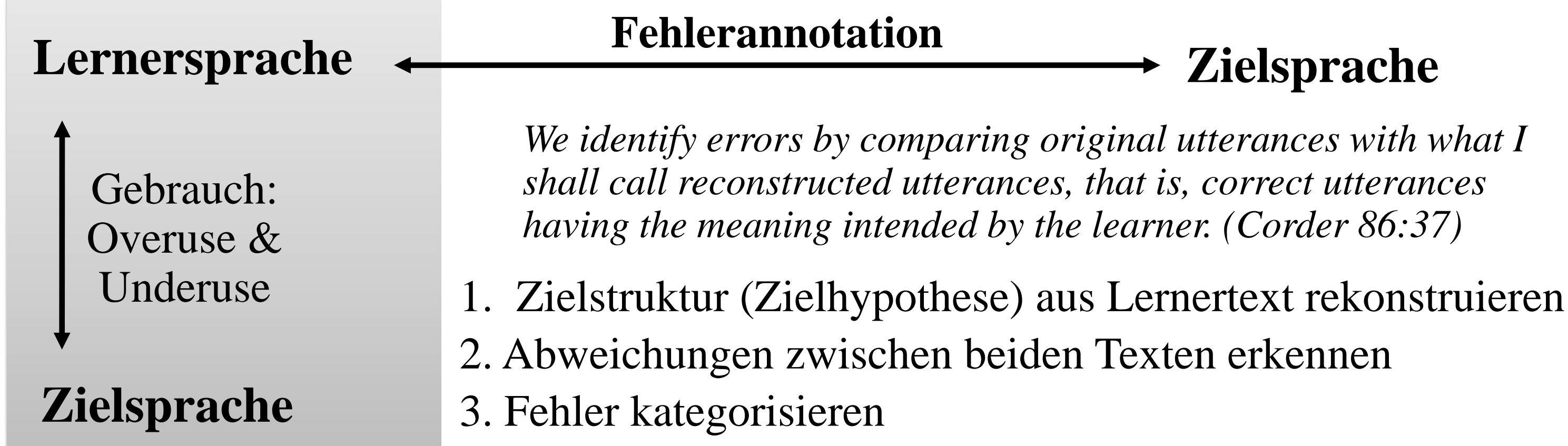
+ Humboldt-Universität zu Berlin, \* Bangor University



Arts & Humanities Research Council



## Fehlerannotation



## Zielhypothese

Fehlerannotationen in öffentlich zugänglichen Lernerkorpora (z. B. ICLE) basieren auf einer impliziten Zielhypothese.

Zu jeder Lerneräußerung können jedoch gleichberechtigte, konkurrierende Zielhypothesen formuliert werden. (vgl. Lüdeling 2008)

### Konsequenzen:

- a) Zielhypothesen müssen operationalisiert werden.
- b) Zielhypothesen müssen explizit gemacht werden.
- c) Weitere Zielhypothesen müssen zugelassen werden.

LT	Das	ist	vielleicht	das	Grund	,	aus	dem		mir	das	Wort	und	die	Ideologie	-	wenn	man	so	sagen	kann	-	,	die	sie	sich	hineinsteckt	,	nicht	echt	im	Begriff	habe	.
ZH2a	Das	ist	vielleicht	der	Grund	dafür	,			dass	ich	dieses	Wort	und	die	Ideologie	,	wenn	man	so	sagen	kann	-	die		darin	enthalten	ist	,	nicht	wirklich	begriffen	habe	.
ZH2b	Das	ist	vielleicht	der	Grund		,			weshalb	ich	für	das	Wort	und	die	Ideologie	-	wenn	man	so	sagen	kann	-	was	darin	steckt	,	keinen	Begriff	habe	.		
ZH2c	Das	ist	vielleicht	der	Grund		,			warum	ich	das	Wort	und	die	Ideologie	-	wenn	man	es	nennen	kann	-	die	darin	steckt	,	nicht	wirklich	begriffen	habe	.		
ZH2d	Das	ist	vielleicht	der	Grund		,	aus	dem		ich	das	Wort	und	die	Ideologie	-	wenn	man	so	sagen	kann	-	die	darin	steckt	,	nicht	wirklich	im	Griff	habe	.	

  Abweichung vom Lernertext (LT)

  Unterschiede zwischen den Zielhypothesen

Grafik 1: Unterschiedliche Annotatoren (a-d) formulieren konkurrierende Zielhypothesen (ZH2) als Grundlage für die Fehlerannotation  
Quelle: Falko L2 Essay Korpus fk016\_2006\_07.xml



Minimale Veränderungen stellen eine morphosyntaktisch korrekte, grammatische Zielstruktur her.

- + gut operationalisierbar
- + hohe Konsistenz der Annotation zwischen Texten und Annotatoren
- + große strukturelle Nähe zur Lerneräußerung

Einbeziehung lexikalischer, semantischer und pragmatischer Information stellt eine funktional äquivalente Zielstruktur her.

- + bestmögliche Rekonstruktion der Lernerintention

LT	Das	ist	vielleicht	das	Grund	,	aus	dem	mir	das	Wort	und	die	Ideologie	-	wenn	man	so	sagen	kann	-	,	die	sie	sich	hineinsteckt	,	nicht	echt	im	Begriff	habe	.
ZH1a	Das	ist	vielleicht	der	Grund	,	aus	dem	ich	das	Wort	und	die	Ideologie	,	wenn	man	so	sagen	kann	-	,	die	sie	sich	hineinsteckt	,	nicht	echt	im	Begriff	habe	.
ZH1b	Das	ist	vielleicht	der	Grund	,	aus	dem	ich	das	Wort	und	die	Ideologie	-	wenn	man	so	sagen	kann	-	,	die	sie	sich	hineinsteckt	,	nicht	echt	im	Begriff	habe	.
ZH1c	Das	ist	vielleicht	der	Grund	,	aus	dem	ich	das	Wort	und	die	Ideologie	(	wenn	man	so	sagen	kann	)	,	die	sie	sich	hineinsteckt	,	nicht	echt	im	Begriff	habe	.
ZH1d	Das	ist	vielleicht	der	Grund	,	aus	dem	ich	das	Wort	und	die	Ideologie	-	wenn	man	so	sagen	kann	-	,	die	sie	sich	hineinsteckt	,	nicht	echt	im	Begriff	habe	.

  Abweichung vom Lernertext (LT)

  Unterschiede zwischen den Zielhypothesen

Grafik 2: Die minimale Zielhypothese (ZH1) erlaubt eine hohe Übereinstimmung zwischen Annotatoren (a-d).

## Operationalisierung der ZH1

- 1) tokenbasierte Korrekturen (verbessern Suchergebnisse)
- 2) geringe Korrekturen
- 3) Bewegungen  
kurze Wege weniger und leichter Konstituenten
- 4) Erhaltung der Syntax (Konnektoren regieren)
- 5) Aufbau topologischer Felder (linke Satzklammer konstituieren)
- 6) Hierarchie im Verbalrahmen  
Vollverb>Objekte>Subjekt>Kopula
- 7) Hierarchie im Nominalrahmen  
Nomen>Artikel>Adjektive

LT	Heute		Männer	gehen	nicht	oft	einkaufen	.
ZH1	Heute		gehen	Männer	nicht	oft	einkaufen	.
*ZH1	Heute	gehen	Männer		nicht	oft	einkaufen	.

Grafik 3: Beispiel Regel 4) Finites Verb bleibt unbewegt

## Automatische Annotation der Abweichungen

Auf der Annotationsebene DIFF werden automatisch die Inhalte eines Tokens in LT und ZH1 verglichen und drei Tags vergeben.

LT gefüllt & ZH1 gefüllt & LT ≠ ZH1 → CHA (change)

LT gefüllt & ZH1 leer → DEL (deletion)

LT leer & ZH1 gefüllt → INS (insertion)

LT	Abend	-	er	schläft	
ZH1	Abends			schläft	- er
DIFF	CHA	-	DEL		INS

Abwandlung — Löschung — Einfügung

mögliche Bewegung

LT(x) = ZH(y) & DIFF(x)=„DEL“ & DIFF(y)=„INS“

→ Bewegung LT(x) nach ZH1(y)

## Anwendungsbeispiele:

**Fragstellung:** Wo unterlassen Lerner den expliziten Ausdruck von Definitheit?  
**Suche:** Liste aller in der Zielhypothese eingefügten definiten Artikel.

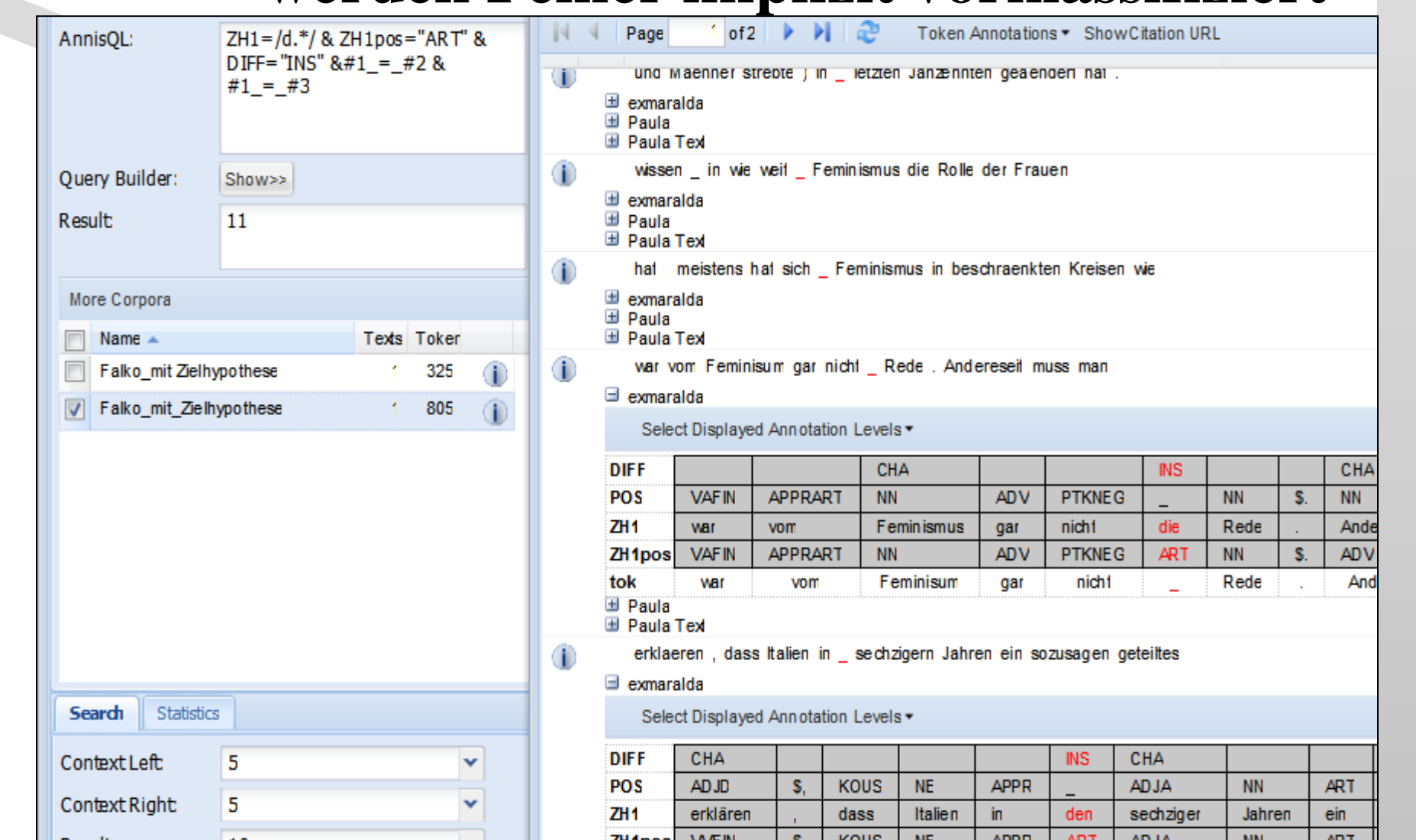
**Ausdruck in ANNIS Query Language:**  
ZH1=/d.\*/ & ZH1pos=„ART“ & DIFF=„INS“ & #1=#2 & #1=#3

**Fehlererkennung:** Fehlende Inversion  
**Fragstellung:** Wo unterlassen Lerner die Inversion von Verb und Subjekt in Nebensätzen?  
→ Die definitive Annotationszeit wird deutlich verkürzt  
→ Ohne zusätzliche, aufwendige Fehlerannotation werden Fehler implizit vorklassifiziert

**Ausdruck in ANNIS Query Language:**  
pos=„KOUS“ & pos=„/V.FIN/“ & pos=„/V.FIN/“  
& DIFF=„DEL“ & DIFF=„INS“  
& #1.\*#2 & #2.\*#3 & #2=#4 & #3=#5

Präzedensoperator — Überlappungsoperator

- findet alle Kandidaten für eine fehlende Inversion
- erlaubt eine deutlich kürzere Annotationszeit
- Über maximale Abstände zwischen den korrespondierenden Token lässt sich die Anzahl der positiven Funde optimieren.



Suche in Falko in ANNIS  
<http://www.sfb632.uni-potsdam.de/d1/annis/>

## Ausblick

- Aufbau syntaktischer Bäume über der Zielhypothese 1
- Erarbeitung von Richtlinien für Zielhypothese 2
- Semiautomatische Fehlerannotation
- Aufbau der Zielhypothesen als Parallelkorpora zum Lernertext
- Quantitative Fehlerbeschreibung

<sup>1</sup> Lüdeling, Anke, Doolittle, Seanna, Hirschmann, Hagen, Schmidt, Karin, & Walter, Maik (2008). Das Lernerkorpus Falko. Deutsch als Fremdsprache, (2), 67–73.  
Corder, Stephen Pit (1986): The role of interpretation in the study. In: Corder, Stephen P. (Hrsg.): Error analysis and interlanguage. 4. impr. Oxford: Oxford University Press, S. 35–44.  
Lüdeling, Anke (2008): Mehrdeutigkeiten und Kategorisierung. Probleme bei der Annotation von Lernerkorpora. In: Walter, Maik/Grommes, Patrick (Hrsg.): Fortgeschrittene Lernervarietäten. Korpuslinguistik und Zweitspracherwerbsforschung. Deutsche Gesellschaft für Sprachwissenschaft. Tübingen: Niemeyer (= Linguistische Arbeiten; 520), S. 119–140.