Zur Korpusarchitektur der Falko-Korpora und ihrer Auswertung

Hagen Hirschmann

20. 10. 2017

Universität Szeged

Plan

- Übergreifende Richtlinien und Tendenzen zur Korpuserstellung
- Die verschiedenen Falko-Korpora
  - Das **Falko_Essay**-Kernkorpus
    - Auswertungsbeispiel 1: Overuse- und Underuse grammatischer Kategorien
    - Auswertungsbeispiel 2: Grammatische und ungrammatische Strukturen bei der Analyse von Präpositionalobjekten
  - Fazit: Vor- und Nachteile von **Falko_Essay**
    - Lösungen für die Nachteile:
    - Homogene Lernerpopulation: Lernende des **WHIG-Korpus**
    - Homogene Aufgabenstellung: Lernende des **Kobalt-DaF-Korpus**
    - Longitudinale Lernerdaten: Lernende des **KanDeL-Korpus** (Analysebeispiel: Studie Vyatkina et al. 2015)
    - Erweiterung von **Falko_Essay** um spezifische Ebenen (Analysebeispiel: Studie Lüdeling et al. 2017)

Übergreifende Richtlinien und Tendenzen zur Korpuserstellung

- Für alle Korpora der Falko-Familie gilt:
  - Schriftliche Sprachdaten
  - Hauptsächlich argumentative Texte (Erörterungen)
  - Hauptsächlich fortgeschrittene Lernende (~ab B1-Niveau)
  - Texte Produziert unter kontrollierten Bedingungen
    - Klassenraumsituation
    - Textproduktion am Computer (notfalls handschriftlich)
    - Keine Hilfsmittel
    - Vorgegebene Bearbeitungszeit: bis zu 90min

Übergreifende Richtlinien und Tendenzen zur Korpuserstellung

- Alle Korpora der Falko-Familie erhalten …
  - Tokenisierung der Originaltexte

[txt]	Da die Studenten einen grossen Teil ihres Studiums an die Theorien wittmen muss, sollten sie lieber praxisorientiert sein.																		
[tok]	Da	die	Studenten	einen	grossen	Teil	ihres	Studiums	an	die	Theorien	wittmen	muss	,	sollten	sie	lieber	praxisorientiert	sein	.

  - Formulierung mindestens einer sog. Zielhypothese bei ungrammatischen Äußerungen, Äußerungsteilen oder Normverstößen bei Schreibungen

[txt]	Da die Studenten einen grossen Teil ihres Studiums an die Theorien wittmen muss, sollten sie lieber praxisorientiert sein.																		
[tok]	Da	die	Studenten	einen	grossen	Teil	ihres	Studiums	an	die	Theorien	wittmen	muss	,	sollten	sie	lieber	praxisorientiert	sein	.
[ZH]					großen				den			widmen	müssen							

(aus FalkoEssayL2v2.4, cbs011_2006_09)

Übergreifende Richtlinien und Tendenzen zur Korpuserstellung

- Alle Korpora der Falko-Familie erhalten …
  - Tokengenaue Alignierung von Originaläußerungen und Zielhypothesen
  - Wortartenanalyse und Lemmatisierung der Originaldaten und der Zielhypothesen (STTS-Tags und Lemmata vom Tretagger zugewiesen)

[tok]	Da	die	Studenten	einen	grossen	Teil	ihres	Studiums	an	die	Theorien	wittmen	muss	,	sollten	sie	lieber	praxisorientiert	sein	.
[pos]	KOUS	ART	NN	ART	ADJA	NN	PPOSAT	NN	APPR	ART	NN	VVINF	VMFIN	$,	VMFIN	PPER	ADJD	ADJD	VAINF	$.
[lemma]	da	d	Student	ein	groß	Teil	ihr	Studium	an	d	Theorie	wittmen	müssen	,	sollen	sie	lieb	praxisorientiert	sein	.
[ZH]	Da	die	Studenten	einen	großen	Teil	ihres	Studiums		den	Theorien	widmen	müssen	,	sollten	sie	lieber	praxisorientiert	sein	.
[ZHpos]	KOUS	ART	NN	ART	ADJA	NN	PPOSAT	NN		ART	NN	VVINF	VMINF	$,	VMFIN	PPER	ADJD	ADJD	VAINF	$.
[ZHlemma]	da	d	Student	ein	groß	Teil	ihr	Studium		d	Theorie	widmen	müssen	,	sollen	sie	lieb	praxisorientiert	sein	.

Übergreifende Richtlinien und Tendenzen zur Korpuserstellung

- Alle Korpora der Falko-Familie sind in ein und demselben Suchinterface, dem ANNIS-Suchprogramm, eingepflegt und können dort einzeln oder gemeinsam ausgewertet werden
- Dies erlaubt
  - Suchen nach Wortformen, Grundformen sowie Abfolgen derselben
  - Auflistung bestimmten Wörter und Grundformen nach deren Häufigkeit
  - Suchen nach grammatischen Mustern durch Wortartabfolgen usw.

Besonderheiten des Falko-Essay-Kernkorpus

- Vier kontroverse Themen
  (in Anlehnung an ICLE; Kriminalität, Entlohnung, Jugend, Studium)
- Derzeit 248 Lernertexte, 95 Muttersprachlertexte
  L2=144619 Token; L1=70615 Token
- Lernerdaten: diverse Muttersprachen; größte Gruppen: Englisch, Polnisch, Russisch, Französisch (Metadaten)
  - (Ungarisch nur ca. 2000 Token)
- Weitere Metadaten: L1, weitere L2, Erwerbszeiten, Alter, Geschlecht, …
- Zusätzlich L1-Vergleichsdaten

Besonderheiten des Falko-Essay-Kernkorpus

- Annotationen:
  - Formulierung zweier Zielhypothesen:
    ZH1 = "Korrektur" ungrammatischer Äußerungen bzw. Äußerungsteile
    ZH2 = "Korrektur" stilistischer Probleme
  - Tokengetreue Markierung der Unterschiede zwischen Originaläußerung und der ZH-Ebenen
  - Satzspannenannotation auf den ZH-Ebenen
  - Dependenzannotationen (Syntaxbäume) auf der ZH1-Ebene

Besonderheiten des Falko-Essay-Kernkorpus

- Beispiel für die ZH1-Bearbeitung mit Markierung der strukturellen Unterschiede:

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

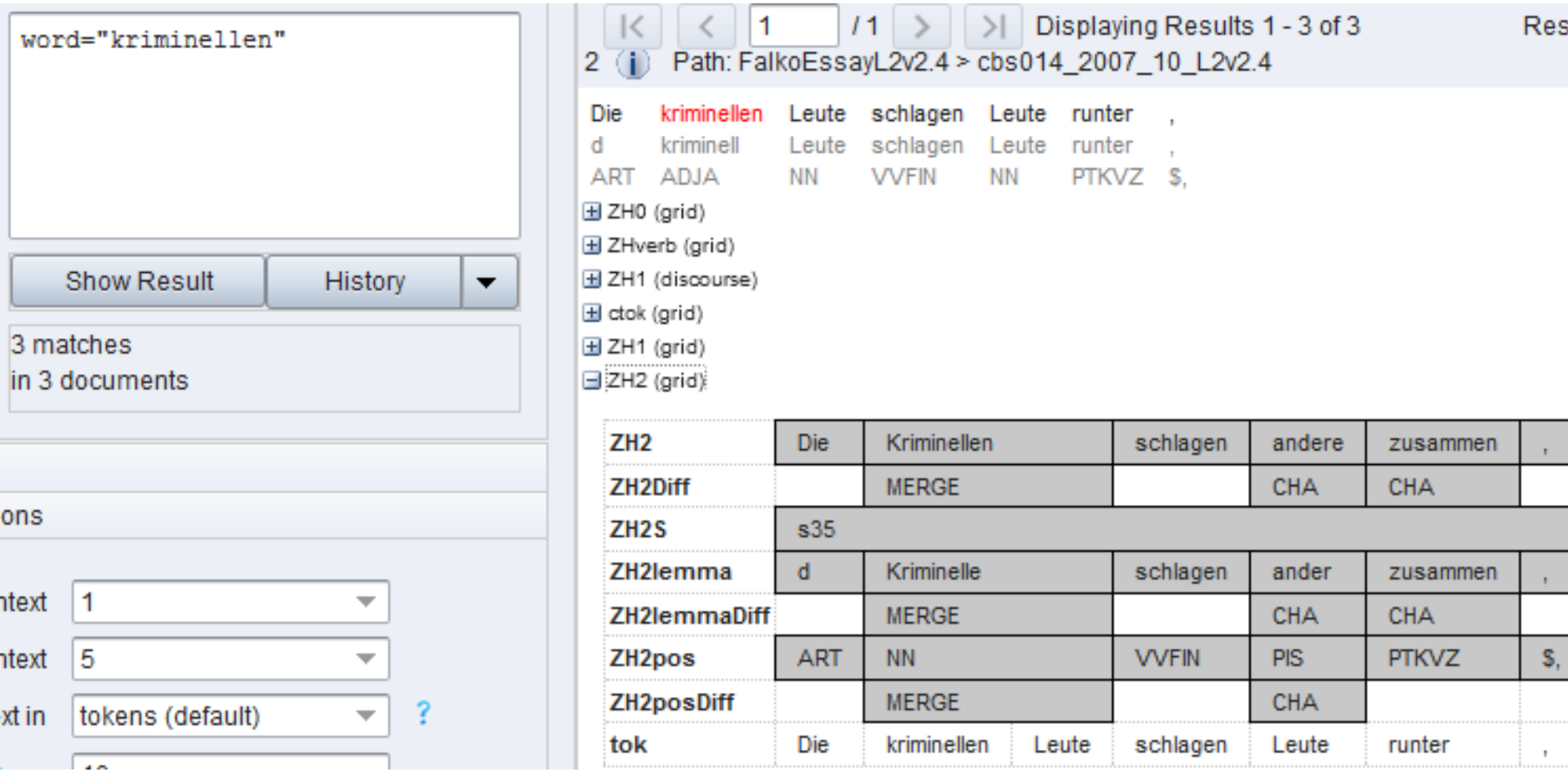
cbs011_2006_09

Besonderheiten des Falko-Essay-Kernkorpus: ZH1 vs. ZH2

- Zielhypothese I vs. 2

Besonderheiten des Falko-Essay-Kernkorpus: ZH1 vs. ZH2

- Zielhypothese 1 vs. 2



Vorteile des Zielhypothesen-Ansatzes

- Strukturen mit Grammatikalitätsproblemen suchbar
- Strukturen ohne Grammatikalitätsproblemen suchbar
- Fehler sind markiert und erhalten strukturelle Klassen (edit tags: INS, DEL, CHA, MERGE, SPLIT, MOVE)
- Zielhypothesen werden wie die originalen Strukturen behandelt (getaggt, geparst) und in die Analyse einbezogen
  - Mehrebenenarchitektur notwendig (MS Excel im Wesentlichen als Annotationsumgebung)
  - Unabhängigkeit der Ebenen erforderlich (standoff xml)
  - EXMARaLDA-xml als geeignetes Speicherformat Schmidt 2012; www.exmaralda.org

Vorteile des Zielhypothesen-Ansatzes

- Zwei grundlegende Wege zur Auswertung von Lernerkorpora (ähnlich wie Typ-A- vs. Typ-B-Studien nach Biber 2009 u.a.):
- EA (Error Analysis) vs. CIA (Contrastive Interlanguage Analysis) (vgl. z. B. Granger 2002 o. 2008)
  - EA:
    - Fehler=Abweichungen von der Zielsprache
    - 'Misuse'
  - CIA:
    - Vergleich zielsprachlicher (grammatischer) Strukturen in L2 mit L1-Strukturen
    - 'Overuse'/'Underuse'

Zusammenfassung: Annotationen in Falko und verwendete Tools

Annotation	Annotationswerkzeug
Tokenisierung, pos-Annotation, Lemmatisierung von Lerneräußerung sowie Zielhypothesen	Treetagger, manuelle Korrekturen in MS Excel
Erstellung Zielhypothesen, Diff-Ebenen, Satzspannenannotation	manuelle/automatische Annotation in MS Excel
Parsing von Zielhypothesen	Malt Parser, manuelle Korrekturen in Arborator

- Speicherung der Daten im EXMARaLDA-Format
- Zusammenführung der Annotationen mit Salt'n Pepper
- Importierung der Daten ins Suchsystem ANNIS

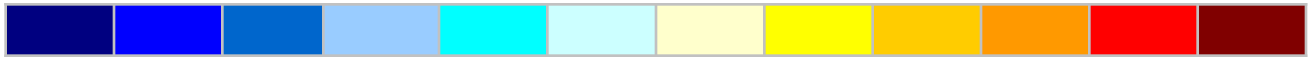
Fallstudie 1: Mindergebrauch von Modifikatoren (Hirschmann 2015, Hirschmann et al. 2013)

- Vorarbeit: Systematische Suche nach mindergebrauchten Wortkategorien
- Verschiedene L2-Verteilungen werden mit L1-Verteilungen verglichen
- Overuse und Underuse werden als (statistisch signifikante) Unterschiede zwischen den Populationen definiert
- Interpretation von L2-Underuse:
  - Lernende kennen die Struktur bzw. Form nicht
  - Lernende kennen die Struktur, aber vermeiden sie bewusst oder unbewusst

Visualisierung von Over- und Underuse

- Underuse → kühle Farben
- Overuse → warme Farben

Underuse **Overuse**



ExcelAdd-In und Doku von Amir Zeldes verfügbar unter
https://github.com/amir-zeldes/XLAddIns

Visualisierung von Over- und Underuse: Wortartenkategorien

bigram	tot_norm	de	da	en	fr	pl	ru
$.-PPER	0.042384	0.005297	0.009748	0.007963	0.006166	0.005801	0.007409
VVFIN-$,	0.042131	0.006457	0.00776	0.006343	0.006937	0.006243	0.008391
ADV-ADV	**0.041604**	**0.012858**	**0.010518**	**0.006111**	**0.006166**	**0.003094**	**0.002856**
PDAT-NN	0.03956	0.005409	0.004233	0.005509	0.007837	0.007735	0.008837
ADV-ART	0.037125	0.007629	0.006349	0.006898	0.005653	0.006133	0.004463

Adverbketten in allen L1-Gruppen mindergebraucht

Modifizierende Wörter

- Korpusbasierte Studien zu Adverbien in DaF
  - Typischerweise basiert auf Lexemen, selten auf Wortklassen, quasi nie auf grammatischen Funktionen
  - Typischerweise für ein Sprachpaar (Möllering 2004, Vyatkina 2007, Weinberger 2009 etc.)
- ADV-Mindergebrauch weist aber auf generelleres Problem: Modifikation

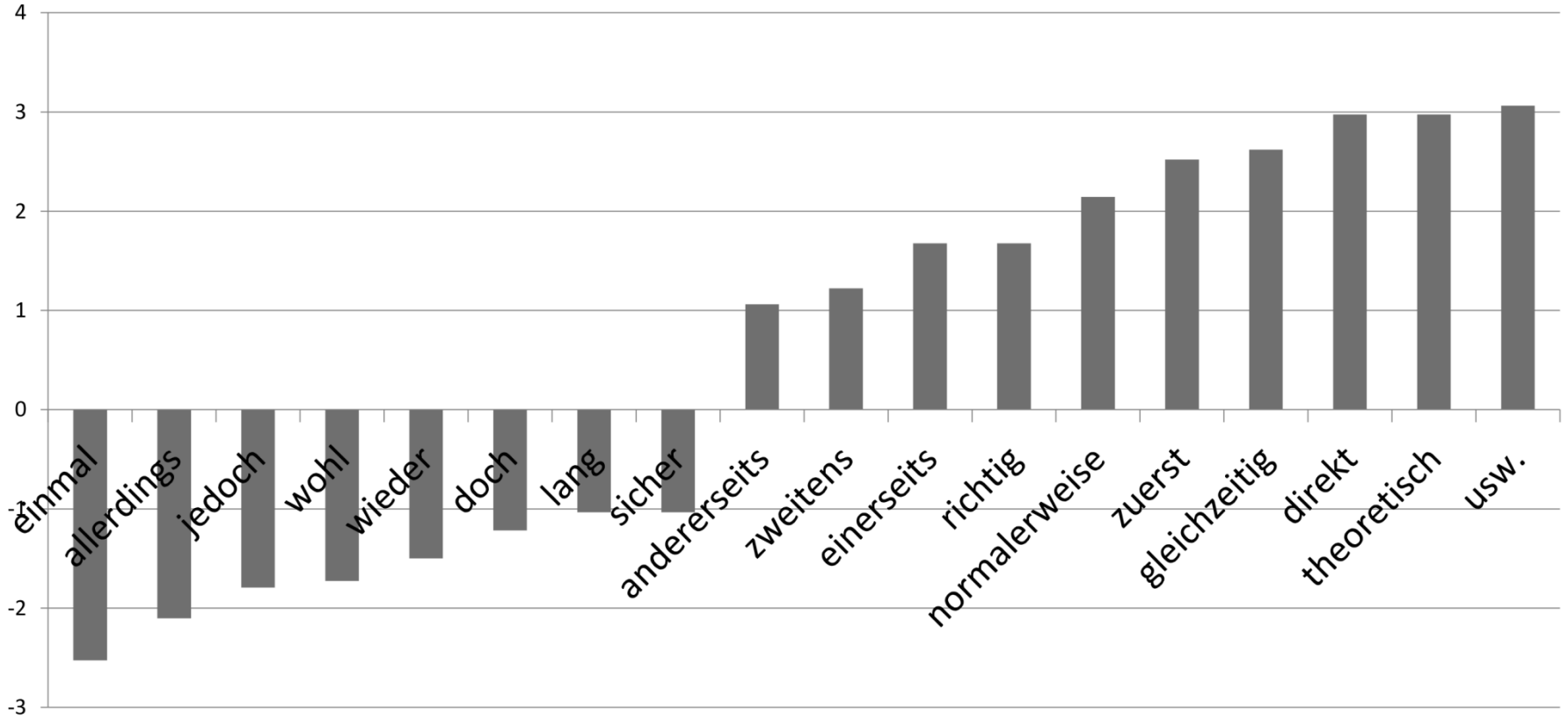
Modifikation

- Sind die beobachteten Effekte form- oder funktionsbasiert?
  - Werden Adverbien generell mindergebraucht?
  - Werden nur bestimmte Formen mindergebraucht?
  - Nur in bestimmten Funktionen (vgl. schon als Temporaladverb vs. Fokuspartikel vs. Modalpartikel)?
  - …

Modifikation

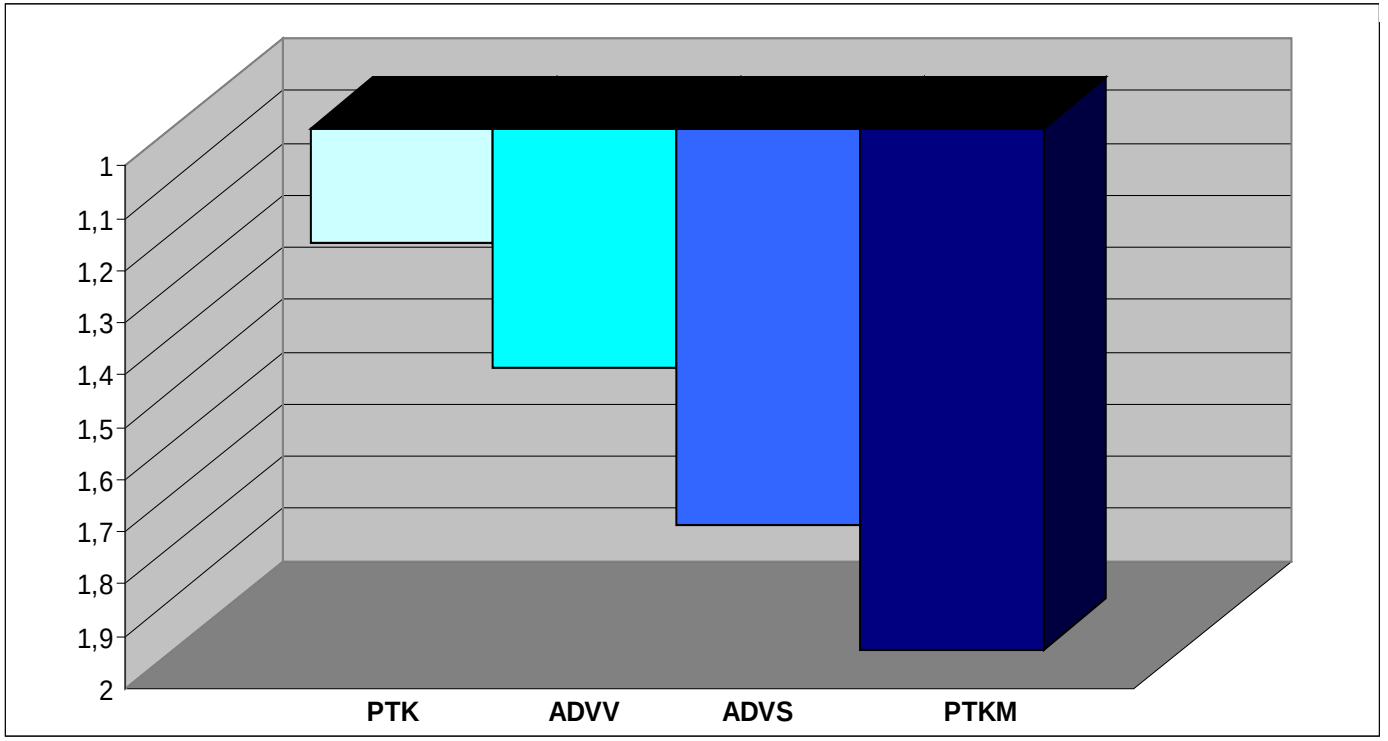
- Sind die beobachteten Effekte form- oder funktionsbasiert?
  - Werden Adverbien generell mindergebraucht?
    Nein: *einerseits*, *normalerweise* etc. werden übergebraucht
  - Werden nur bestimmte Formen mindergebraucht?
    Ja: *wohl*, *allerdings* etc. werden mindergebraucht
  - Nur in bestimmten Funktionen (vgl. *schon* als Temporaladverb vs. Fokuspartikel vs. Modalpartikel)?
    Ja: Adverbien als Satz- bzw. event-externe Modifikatoren werden viel stärker vermieden als Verb-Modifikatoren bzw. Eventmodifikatoren
  - (vgl. Auswertungen auf den kommenden Folien)

L2-L1-Differenzen im Gebrauch verschiedener ADV-Lexeme



(Hirschmann 2015, S. 272)

Mindergebrauch verschiedener ADV-Subklassen



PTK: Partikeln, v.a. Intensivierer/Gradpartikeln (***sehr*** *gut*)

ADVV: Verbmodifikatoren (*Das geht* ***schnell***)

ADVS: Event-externe Modifikatoren (***Bestimmt*** *schneit es bald*)

PTKM: Modalpartikeln (*Es schneit* ***wohl*** *gerade*)

Zusammenfassung: Modifikation in Falko

- Modifikation ist allgemein eine schwer zu erwerbende Funktionskategorie im DaF
  - Dies kann nicht nur lexikalisch, sondern auch anhand von syntaktisch-funktionalen Klassen nachgewiesen werden
    - Hier zeigt sich, dass Modifikatoren an höheren syntaktischen Positionen stärker betroffen sind als tiefer sitzende (VP-interne) Modifikatoren

Fallstudie 2

- Nutzung des Falko-Essay-Korpus zur Untersuchung von **Präpositionalobjekten**

Motivation/Fallbeispiel: Präpositional-objekte im fortgeschrittenen DaF

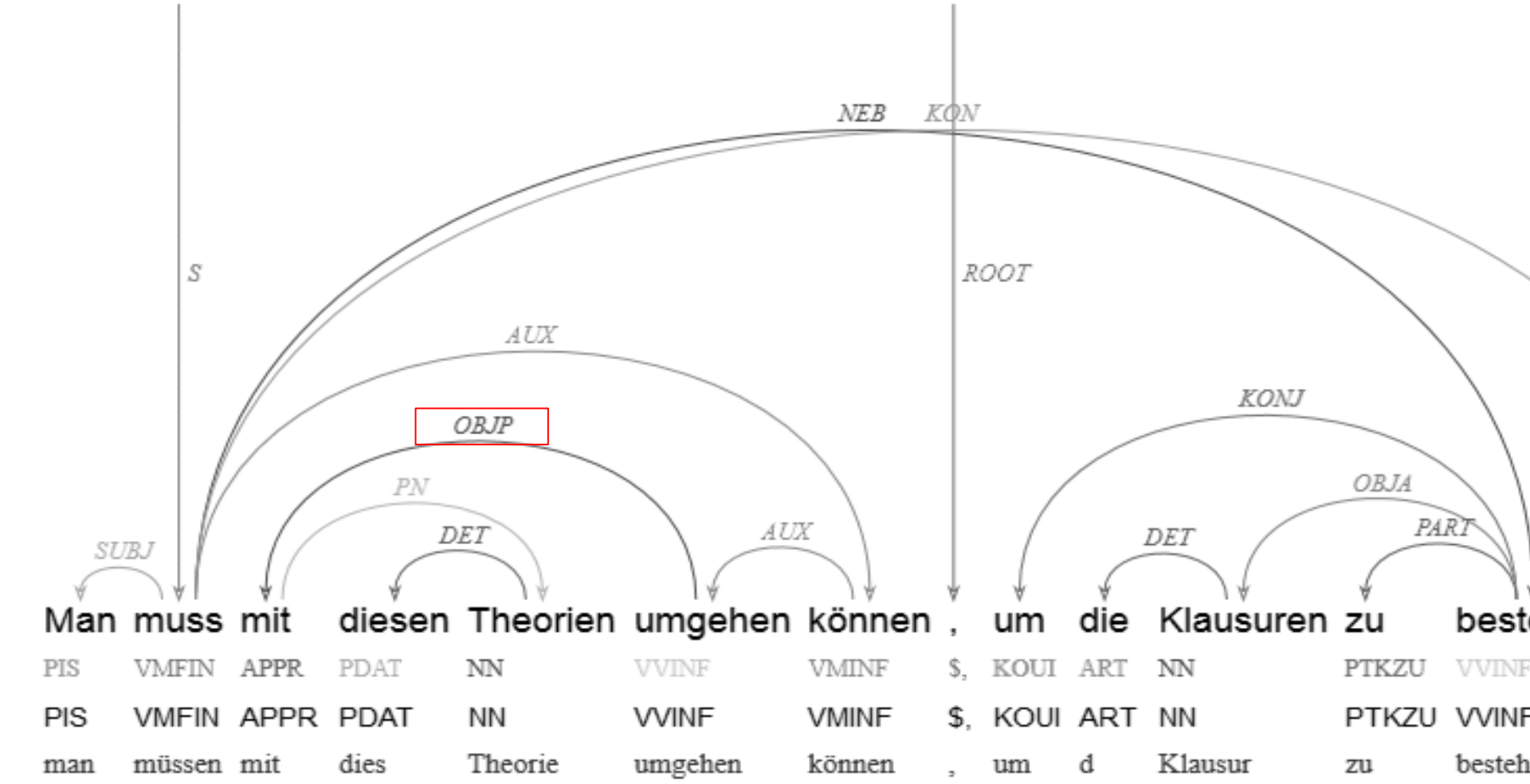
- *Studenten darum beklagen , dass ihr Studium sie nicht für die wirkliche Welt und ihre berufliche Zukunft vorbereitet .* (fk006_2006_08)
- Präpositionalobjekte stellen eine besondere Herausforderung für den Lernprozess dar (Präposition schlecht antizipierbar, semantisch keine homogene Objektklasse, …)
- Fragestellung: Wie zielsprachlich ist die Verwendung von Präpositionalobjekten bei den fortgeschrittenen Lernenden des DaF im FALKO-Korpus?

Fallstudie:
Welche Annotationen?

- 1. Teilfrage: Wie viele (grammatische!) Präpositionalobjekte werden von den Lernenden überhaupt verwendet? (CIA)
- 2. Teilfrage: Wie häufig treten bei der Verwendung von Präpositionalobjekten grammatische Regelverstöße auf? (Fehlerstudie)

Beispiel: Präpositionalobjekt in ANNIS-Dependenzdarstellung

(Falko Essays L2, cbs001_2006_09)



Zur Fallstudie: Kontrastive Analyse - Ergebnisse

- Vorgehen: Zählen aller analysierbaren Präpositionalobjekte in Lernerdaten und Muttersprachlerdaten
  - Normalisierung der absoluten Zahlen nach Vollverben
- Anzahl der in den grammatischen Strukturen vorhandenen Präpositionalobjekte:
  L2=3,52 / 100 VV
  L1=3,70 / 100 VV
- Type-Token-Ratio (Verb-PP-Kombinationen):
  L2=0,47
  L1=0,69

Zur Fallstudie: Kontrastive Analyse - Ergebnisse

- Nutzung von Präpositionalobjekten pro Vollverb erfolgt bei den Lernenden nur wenig geringer als bei den Muttersprachlern
  - Dieser Mindergebrauch verschwindet, wenn man die ungrammatischen Fälle (s. nachfolgend) hinzuzählt
- Die Lernenden verwenden in dem Bereich in etwa einen derart geringeren Wortschatz auf, wie sie es auch in anderen lexikalischen Bereichen (Verben im Allgemeinen) tun

Fallstudie:
Welche Annotationen?

- 2. Teilfrage: Wie häufig werden bei der Verwendung von Präpositionalobjekten Fehler produziert?
  - Markierung ungrammatischer Strukturen, deren Zielhypothese ein Präpositionalobjekt oder anstelle einer PP ein alternatives Objekt ist
    - Aufbereitung der Zielhypothesen analog zur Aufbereitung der grammatischen Lerneräußerungen

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)

Fehlertypen - 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

Fehlertypen - 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

Fehlertypen - 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

Fehlertypen - 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

Zur Fallstudie: 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

Zur Fallstudie: Fehleranalyse - Ergebnisse

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

Falko-Essay-Kernkorpus: Fazit: Vorteile

- Geeignet zur Untersuchung struktureller Erwerbsschwierigkeiten des DaF (heterogene Lernerpopulation, Fortgeschrittenheit der ProbandInnen)
- Geeignet für CIA-Studien (gute Vergleichbarkeit der L2- und L1-Daten)
- Beliebig erweiterbares, versionierbares (aktuell FalkoEssayL2v2.4) Korpus durch unabhängige Annotationsebenen, gespeichert in standoff xml

Falko-Essay-Kernkorpus: Nachteile

- Bedingt geeignet zur Untersuchung spezifischer Lernergruppen (sparse data in vielen L2-Gruppen; größte Lernergruppen: pl, ru, fr, da, en; hun nur ca. 2000 Token)
- Starke Themen-Effekte (task effects / topic effects) zu beobachten
- Kaum Möglichkeiten, einen Erwerbsverlauf zu beobachten
- Bedingt geeignet zu detaillierten Fehleranalysen (sehr grobe Kategorisierung, viel manuelle Nacharbeit nötig)

Falko-Essay-Kernkorpus: Nachteile – Lösungen

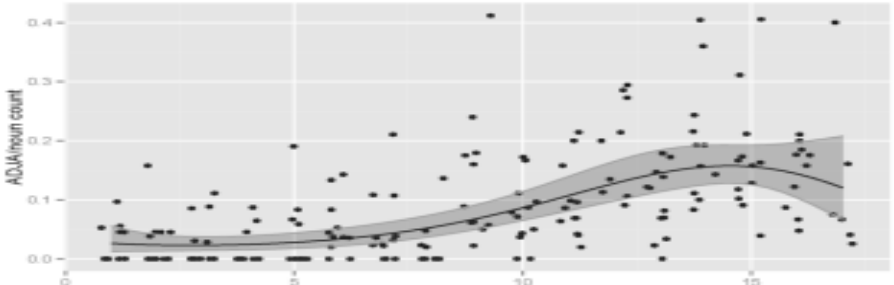
- Bedingt geeignet zur Untersuchung spezifischer Lernergruppen
  → WHIG-Korpus (130.949 Token, 196 Dokumente) beinhaltet ausschließlich englische Lernende
- Starke Themen-Effekte zu beobachten
  → KOBALT-DaF-Korpus (33.368 Token, 51 Dokumente) basiert auf einer Aufgabenstellung ("Jugend") und beinhaltet gleich große Lernergruppen (Schwedisch, Weißrussisch, Chinesisch)
- Kaum Möglichkeiten, einen Erwerbsverlauf zu beobachten
  → KanDeL-Korpus (121.878 Token, 688 Dokumente) beinhaltet dieselben Lernenden ab Anfängerniveau an 18 Erwerbszeitpunkten
- Bedingt geeignet zu detaillierten Fehleranalysen
  → Erweiterung des Kernkorpus um weitere, spezifische Annotationen: Fehlerkategorien bei komplexen Verben

Falko-Essay-Kernkorpus: Nachteile – Lösungen

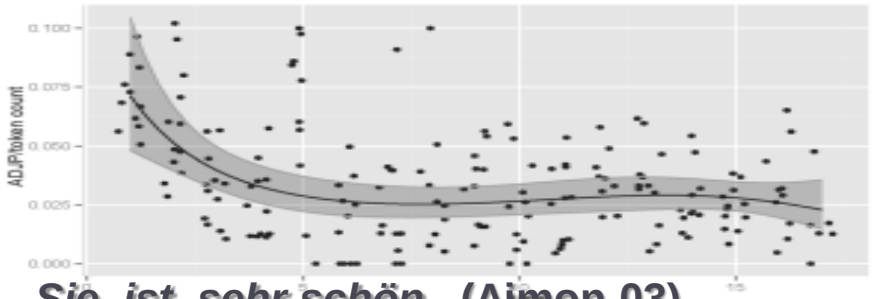
- Kaum Möglichkeiten, einen Erwerbsverlauf zu beobachten
  → KanDeL-Korpus (121.878 Token, 688 Dokumente) beinhaltet dieselben Lernenden ab Anfängerniveau an 18 Erwerbszeitpunkten

- Es folgt: Analysebeispiel für das KanDeL-Korpus in Vyatkina, Hirschmann, Golcher 2015

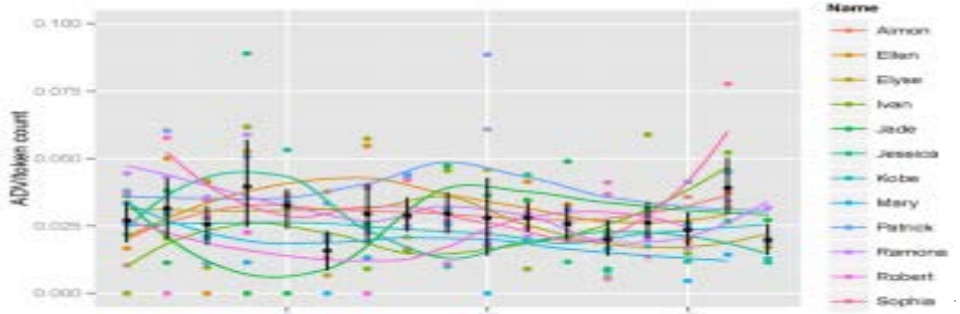
Erwerbsverläufe, gemessen anhand des KandeL-Korpus



Ich habe die beste Familie in der Welt. **(Aimon 03)**
I have the best family in the world.



Sie ist sehr schön. **(Aimon 03)**
She is very pretty.



Gestern kam Julchen zu mir. **(Patrick 15)**
Yesterday came Julchen to me.

- **Prenominal adjectives** significantly increasing over time (despite great variation)
- **Predicative adjectives** significantly decreasing over time (despite great variation)
- **Adverbs** show no significant trend

Falko-Essay-Kernkorpus: Nachteile – Lösungen

- Bedingt geeignet zu detaillierten Fehleranalysen (sehr grobe Kategorisierung, viel manuelle Nacharbeit nötig)
  → Erweiterung des Kernkorpus um weitere, spezifische Annotationen: Fehlerkategorien bei komplexen Verben

- Es folgt: Analysebeispiel für das Falko-Essay-Korpus in Lüdeling, Hirschmann, Shadrova 2017

Spezifizierte Fehlerannotation bei komplexen Verben

Table 7 Annotation of ungrammatical use of particle verb (particle superfluous/wrong verb lexeme) in learner text fkb041_2008_08

txt	Für	mich	scheint	es	aus	,	als	ob
	for	me	[form ungrammatical]	it	[particle]		as	if
lemma	für	mich	scheinen	es	aus	,	als	ob
pos	APPR	PRF	VVFIN	PPER	PTKVZ	$,	KOUS	KOUS
TH	Für	mich	scheint	es		,	als	ob
	for	me	seems	it			as	if
	To me it seems as if . . .							
THDiff					DEL			
THpos	APPR	PPER	VVFIN	PPER		$,	KOUS	KOUS
verb form			**finsep**					
verb category			**vpart**		**ppart**			
verb lemma			**ausscheinen**					
verb error type			**neo**					

Herzlichen Dank

- An und mit Falko arbeite(te)n auch:
  Anke Lüdeling, Seanna Dolittle, Marc Reznicek, Karin Schmidt, Maik Walter, Eva Breindl, Anna Shadrova u.v.m.

- Kontakt (Hilfe, Anmerkungen, …):
  hirschhx@hu-berlin.de

Online-Referenzen: Korpusdokumentation und -zugang

- Falko-, Kobalt-DaF-, KanDeL-Dokumentationshompages: https://www.linguistik.hu-berlin.de/de/institut/professuren/korpuslinguistik/forschung
- Falko-Handbuch: https://www.linguistik.hu-berlin.de/de/institut/professuren/korpuslinguistik/forschung/falko/FalkoHandbuchV2/
- Falko-Suchinterface (ANNIS): https://korpling.german.hu-berlin.de/falko-suche/

Referenzen 1

- Biber, Douglas (2009): Studying register and register variation. In: Lüdeling, Anke; Kytö, Merja (Hg.): Corpus Linguistics. An International Handbook. Vol 1. Berlin; de Gruyter, S. 823-855.
- Granger, Sylviane (2008): Learner corpora. In: Lüdeling, Anke; Kytö, Merja (Hg.): Corpus Linguistics. An International Handbook. Vol 1. Berlin; de Gruyter, S. 259-275.
- Granger, Sylviane. (2002): *A Bird's-eye View of Computer Learner Corpus Research*. In: Granger S., *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching* (Language Learning and Language Teaching; 6). Amsterdam & Philadelphia; John Benjamins, S. 3-33.
- Hirschmann, Hagen (2015) Modifikatoren im Deutschen. Ihre Klassifizierung und varietätenspezifische Verwendung. Tübingen; Stauffenburg.
- Hirschmann, Hagen; Lüdeling, Anke; Rehbein, Ines; Reznicek, Marc; Zeldes, Amir (2013) Underuse of Syntactic Categories in Falko. A Case Study on Modification. In: 20 years of learner corpus research. Looking back, Moving ahead (LCR2011). Louvain-la-Neuve, Belgium; Presses universitaires de Louvain.
- Lüdeling, Anke; Hirschmann, Hagen; Shadrova, Anna (2017) Linguistic Models, Acquisition Theories, and Learner Corpora: Morphological Productivity in SLA Research Exemplified by Complex Verbs in German. In: Language Learning, 1-34.

Referenzen 2

- Lüdeling, Anke; Doolittle, Seanna; Hirschmann, Hagen; Schmidt, Karin & Walter, Maik (2008): Das Lernerkorpus Falko. In: *Deutsch als Fremdsprache* 2(2008), S. 67-73.
- Maden-Weinberger, Ursula (2009): Modality in learner German: A corpus-based study investigating modal expressions in argumentative texts by British learners of German. Dissertation, Lancaster University.
- Möllering, Martina (2004): The Acquisition of German Modal Particles. A corpus-based approach. Bern; Peter Lang.
- 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
- Schmidt, Thomas (2012): EXMARaLDA and the FOLK tools. In: *Proceedings of LREC. ELRA*.
- Vyatkina, Nina; Hirschmann, Hagen; Golcher, Felix (2015) Syntactic modification at early stages of L2 German writing development: A longitudinal learner corpus study. In: Journal of Second Language Writing (JSLW).
- Vyatkina, Nina (2007): Development of second language pragmatic competence: The data-driven teaching of German modal particles based on a learner corpus. Dissertation, Pennsylvania State University.