Lernersprache und Lernerkorpora

BeMaTaC

Korpusdesign und gesprochene Sprache

Marc Reznicek

mit Folien des gesamten Korpuslinguistikteams

Berlin, 30. Mai 2013

Übersicht

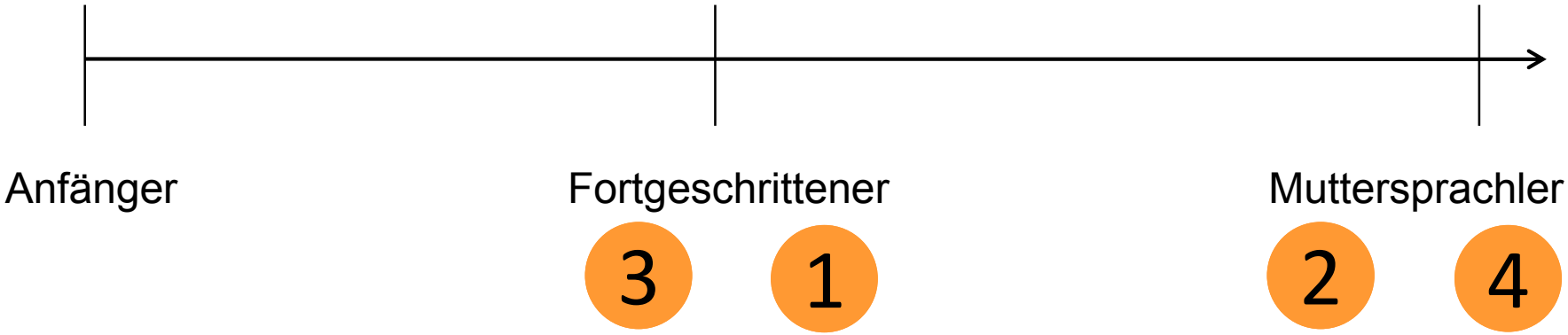
- Wie unterscheiden sich Lerner- von Muttersprachlertexten?
- Wie lernt man eine Fremdsprache?
- Wie kann man Lernersprache untersuchen?
- Was ist ein Fehler?
- Wie findet man Fehler?
- Meine Fehler, deine Fehler!

Wie unterscheiden sich Lerner- von Muttersprachlertexten?

Lernersprache

Vergleichen Sie die 4 Texte!

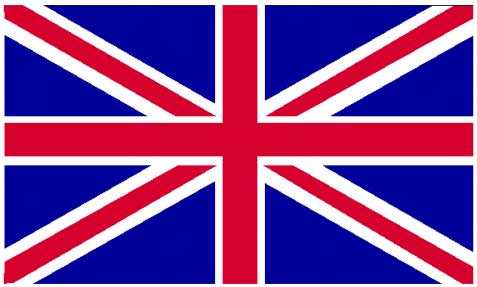
- Welche der Texte stammen von Muttersprachlern, welche von Lernern?
  - Woran haben Sie das erkannt?
- Wo würden Sie die Texte auf der Fortgeschrittenheitsskala eintragen?
  - Warum?



Wie lernt man eine Fremdsprache?

Kontrastivhypothese (Lado 1962)

Sprachsysteme werden verglichen



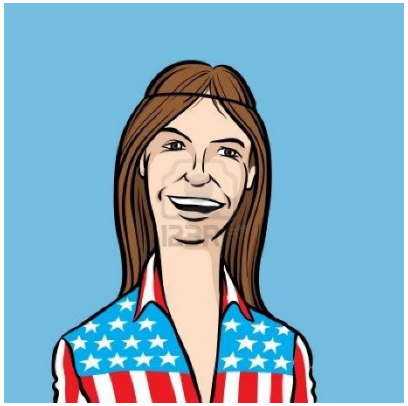
Vorhersage

Ähnlichkeiten
positiver Transfer

Unterschiede
negativer Transfer
Interferenzfehler

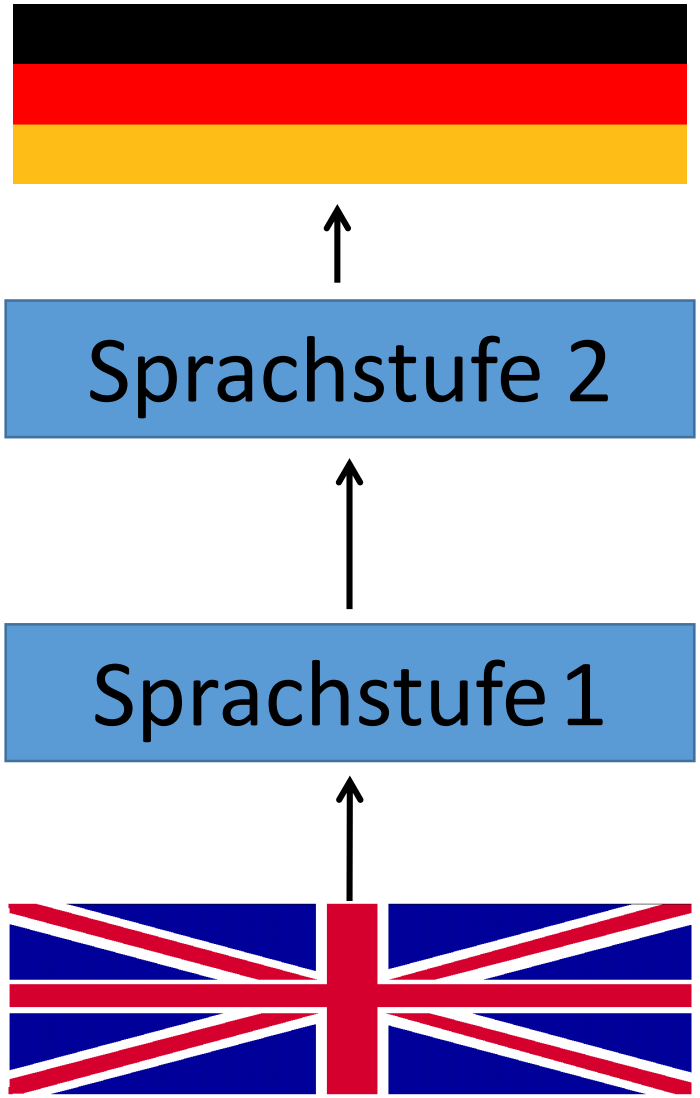
Identitätshypothese (Dulay & Burt 1974)

Erwerbsreihenfolgen werden verglichen



Universalgrammatik

Interlanguage-Hypothese (Selinker 1972)



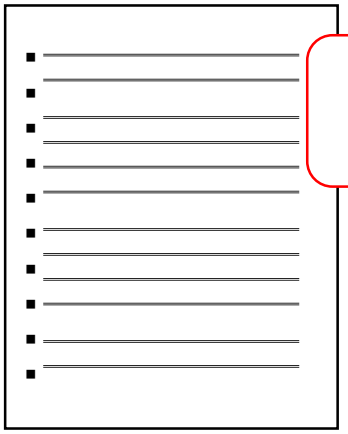
Interimssprachen sind **systematische**, echte Sprachen

Interimssprachen **nähern sich** in ihrer Struktur der **Zielsprache an**.

Wie kann man Lernersprache untersuchen?

Daten

Beobachtungen auflisten

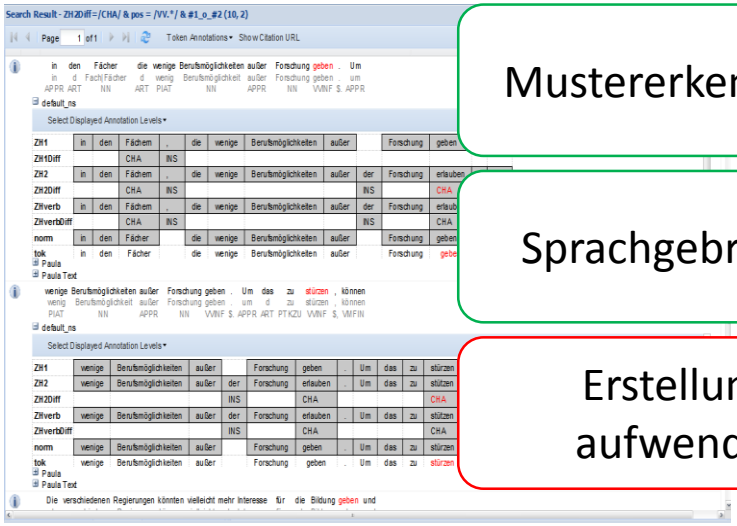


unsystematisch

Fragebögen zur Akzeptanz und Selbsteinschätzung

Sagen nichts über benutzte Sprache aus

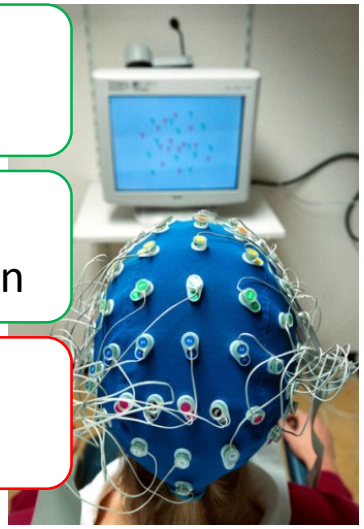
Korpora



- Mustererkennung
- Sprachgebrauch
- Erstellung aufwendig

Experimente

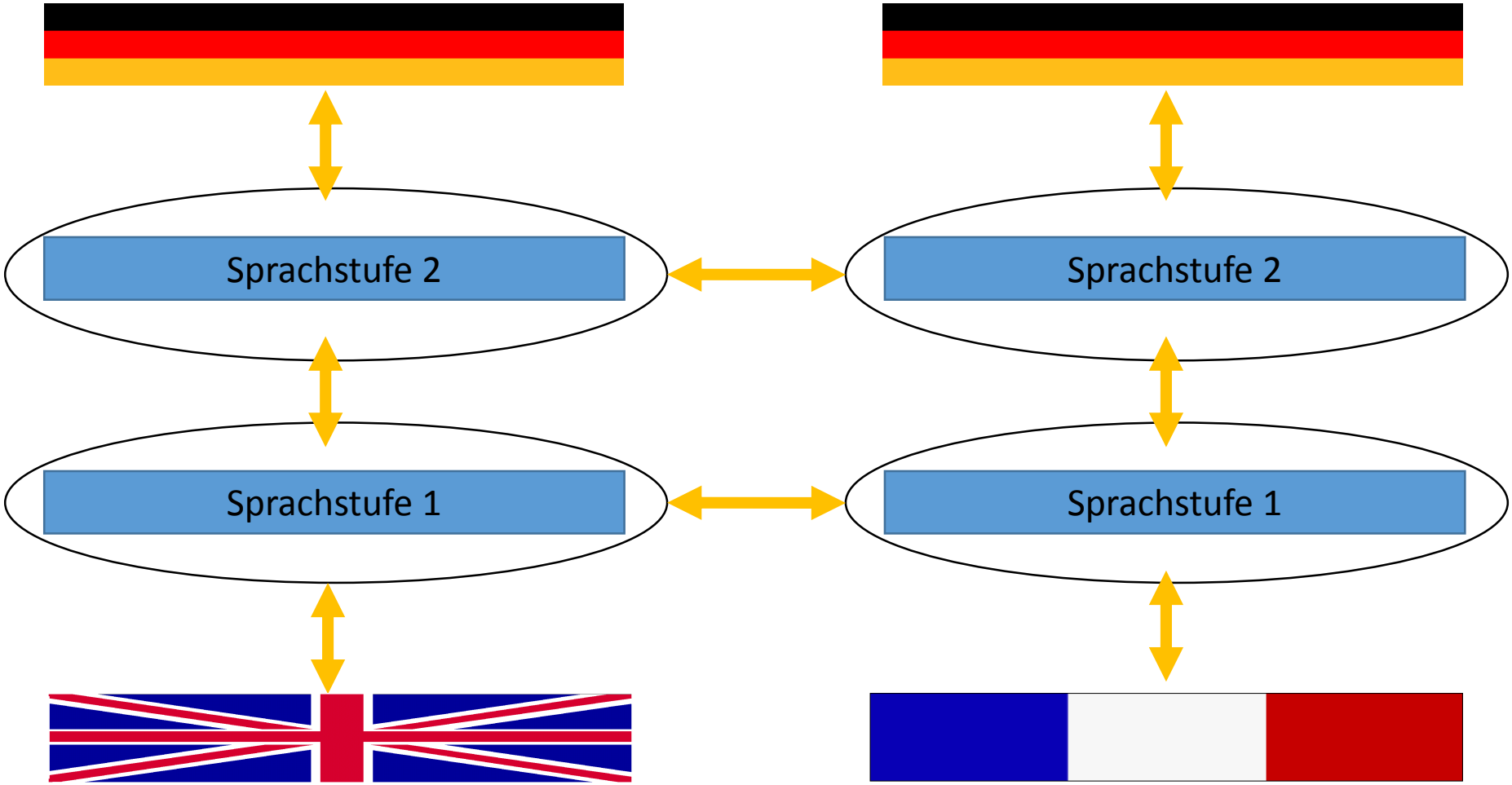
- Prozesse
- mentale Repräsentationen
- Durchführung aufwendig



Kontrastive Interlanguage-Analyse (CIA)

(Granger 2008)

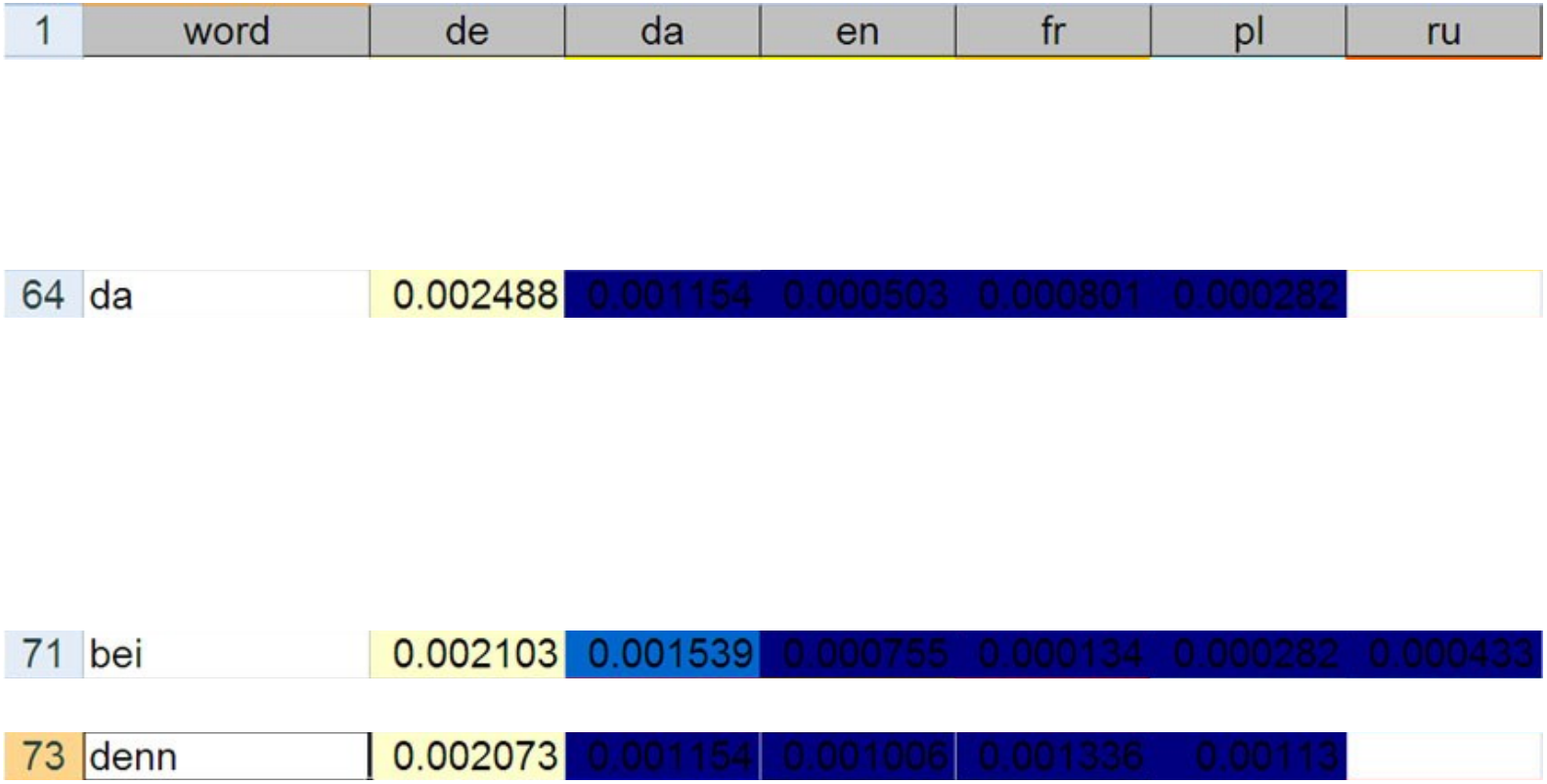
Vergleich der Frequenzen sprachlicher Strukturen

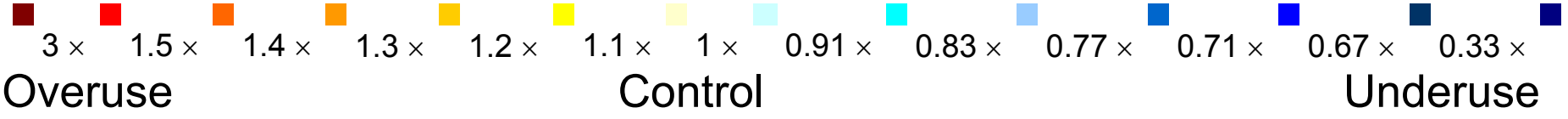


Overuse & Underuse

1	word	de	da	en	fr	pl	ru
2	,	0.055799	0.062897	0.064648	0.070923	0.050833	0.080987
61	dann	0.002754	0.002885	0.001459	0.001736		0.002165
62	einer	0.002666	0.001603	0.002465	0.002671	0.001412	0.003465
63	schon	0.002503	0.001154	0.000704	0.000801	0.001977	0.003032
64	da	0.002488	0.001154	0.000503	0.000801	0.000282	
65	Frau	0.002458	0.000769	0.001157	0.00187	0.001694	0.005197
66	muss	0.002384	0.002372	0.002515	0.001469	0.003389	0.007796
67	Geld	0.002369	0.000769	0.002515	0.000534		0.002599
68	viele	0.002281	0.003847	0.002214	0.00187	0.005083	0.002599
69	zum	0.002147	0.001282	0.002666	0.002938	0.000565	0.000866
70	des	0.002132	0.001731	0.003471	0.004408	0.001694	0.001299
71	bei	0.002103	0.001539	0.000755	0.000134	0.000282	0.000433
72	können	0.002088	0.003526	0.002616	0.003206	0.001977	0.001732
73	denn	0.002073	0.001154	0.001006	0.001336	0.00113	
74	Leben	0.002058	0.001411	0.001811	0.002003	0.000847	0.003465
75	sollte	0.00197	0.001026	0.000956	0.001469	0.000847	
76	einem	0.001925	0.001731	0.001459	0.000935	0.001694	0.000866

Overuse & Underuse

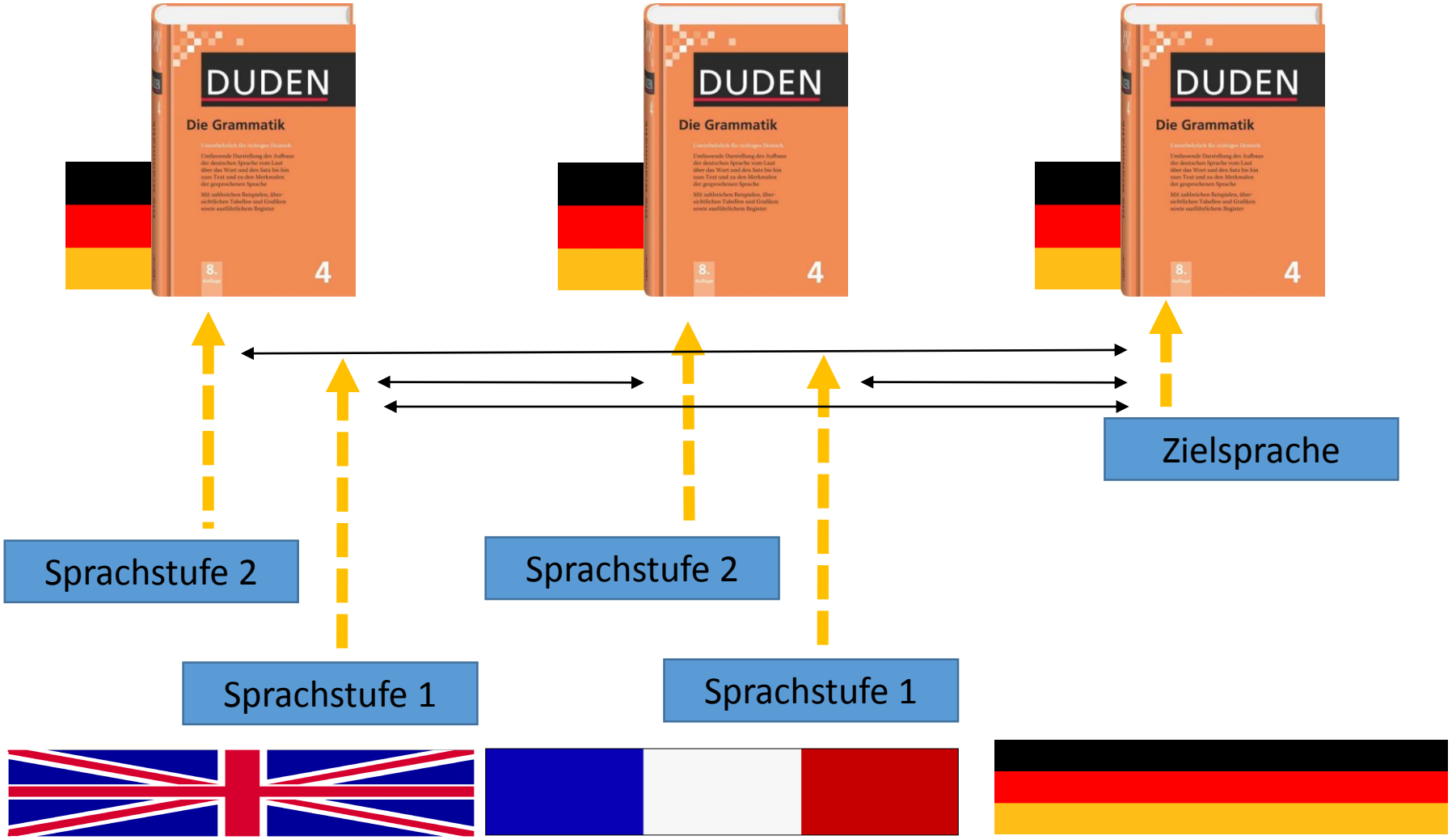




Kontrastive Fehleranalyse (EA)

(Granger 2008)

Vergleich der Frequenzen von Fehlern



Was ist ein Fehler?

Fehler definieren

"A linguistic form, ... which, in the same context would in all likelihood not be produced by the learner's native speaker counterparts." (Lennon 1991, 182)

„Eine sprachliche Form,… die im gleichen Kontext aller Wahrscheinlichkeit nach von einem Muttersprachler nicht produziert werden würde.“ (Lennon 1991, 182)

Fehler identifizieren

An die Arbeit!

Zählen Sie für diese beiden Sätze die Fehler.

Diese Frage schlägt vor, dass die wirkliche Welt nur der Beruf heißt, jedoch die wirkliche Welt heißt nicht nur die Karriere, sondern auch andere Bereiche des Lebens.

Die Studenten bleiben an der Universität, weil sie sich da viel besser vorbereiten sind.

Wie findet man Fehler?

Fehler identifizieren

"We identify errors by comparing original utterances with what I shall call *reconstructed utterances, that is, correct utterances having the meaning intended by the learner.*" (Corder 1986)

„Fehler kann man identifizieren, indem man die ursprüngliche Äußerung mit einer ‚wiederhergestellten' Äußerung vergleicht, d. h. mit einer grammatischen Äußerung, die der beabsichtigten Meinung des Originals entspricht." (Corder 1986)

- Fehler kann man nur vor dem Hintergrund einer Zielhypothese finden. (Lüdeling 2008)

Zielhypothesen explizieren

An die Arbeit!

Formulieren Sie zu den unterschiedlichen Sätzen eine Zielhypothese.

Diese Frage schlägt vor, dass die wirkliche Welt nur der Beruf heißt, jedoch die wirkliche Welt heißt nicht nur die Karriere, sondern auch andere Bereiche des Lebens.

Die Studenten bleiben an der Universität, weil sie sich da viel besser vorbereiten sind.

Zielhypothesen explizieren

Diese Frage schlägt vor, dass die wirkliche Welt nur ~~der~~ Beruf heißt, jedoch heißen nicht nur die Karriere die wirkliche Welt, sondern auch andere Bereiche des Lebens.

Die Studenten bleiben an der Universität, weil sie ~~sich~~ da viel besser vorbereitet sind.

nah an der sprachlichen Oberfläche

- der Zielsprache unähnlich
+ wenige Änderungen
+ weniger Unterschiede zwischen den Korrektoren

Zielhypothesen explizieren

Diese Frage impliziert, dass nur das Berufsleben als wirkliche Welt gilt, jedoch besteht die wirkliche Welt nicht nur aus ~~die~~ Karriere, sondern auch aus anderen Bereichen des Lebens.

Die Studenten bleiben an der Universität, weil sie ~~sich~~ da viel besser vorbereiten werden.

nah an der kontextuellen Adäquatheit

+ der Zielsprache sehr ähnlich
- viele Änderungen
- viele Unterschiede zwischen den Korrektoren

Zielhypothesen auswerten

- Zielhypothesen unterscheiden sich stark voneinander (Lüdeling 2008)
- Darauf basierende Fehlerzählungen weichen stark von einander ab.

Konsequenz:

- Zielhypothesen müssen ins Korpus **integriert** werden.
- Erstellung von Zielhypothesen muss **operationalisiert** werden.

Fehler klassifizieren

- Lexikalisch vs. grammatikalisch (traditionell)
- Morphologie, Syntax, Vokabular

(Politzer & Ramirez 1973)

- Fehlertaxonomien
  Oberflächenmerkmale (Dulay, Burt, Krashen 1982)

Omissions: *Wenn [DET] Professor etwas sagt*

Additions: *Sie geben wenig Lohn [zu] den Arbeiter*

Misformations: *Wenn das jeder Mann weißte, dann…*

Misorderings: [Früher] [die Menschen] [haben] geklaut

Fehlerannotation

Automatische Fehlerannotation in Falko

- Oberflächenabweichung von Lernertext und Zielhypothese

TAG	INS	DEL	CHA	MERGE	SPLIT	MOVS	MOVT

LT	Wegen	Bei	vielen	anderen	Studien	,	kommt	es	gar	nicht	**in**
ZH1	Wegen	Bei	vielen	anderen	Studien		kommt	es	gar	nicht	zur
ZH1Diff						DEL					CHA

LT	Frage		**in**	**wie**	**Weit**	ein	Praktikum	obligatorisch	oder	**un-obligatorisch**		ist
ZH1	Frage	,	inwieweit			ein	Praktikum	obligatorisch	oder	nicht	obligatorisch	ist
ZH1Diff		INS	MERGE							SPLIT		

Falko cbs012_2006_09

Tags beschreiben die Abweichung der Zielhypothese!!

Beispielanalyse

Fehleranalyse

Beispiel: Artikelverwendung

- **Hypothese:**

  Lerner aus Pro-Drop-Sprachen lassen den Artikel häufiger weg als Sprecher aus Nicht-Pro-Drop-Sprachen.

- **Untersuchung anhand eines Lernerkorpus**

  Welche Informationen muss das Korpus enthalten?

  1) Lernertexte
  2) Metadaten
  3) Annotationen

Wie sieht ein Lernerkorpus aus?

Korpusdesign

- **Fortgeschrittene Lerner des Deutschen als Fremdsprache**
  - C-Test > 60/100 Punkten
  - keine bilingualen Sprecher
  - geschriebene Texte
- **Zwei Aufgabenstellungen (Textsorten)**
  - Zusammenfassungen (FalkoSummary)
  - Argumentative Aufsätze (FalkoEssay)
- **Vier Themen**
  - Kriminalität, Entlohnung, Studium, Feminismus
- **Einheitliche Erhebungsbedingungen**
  - 90 Minuten
  - keine Hilfsmittel

Korpusdesign

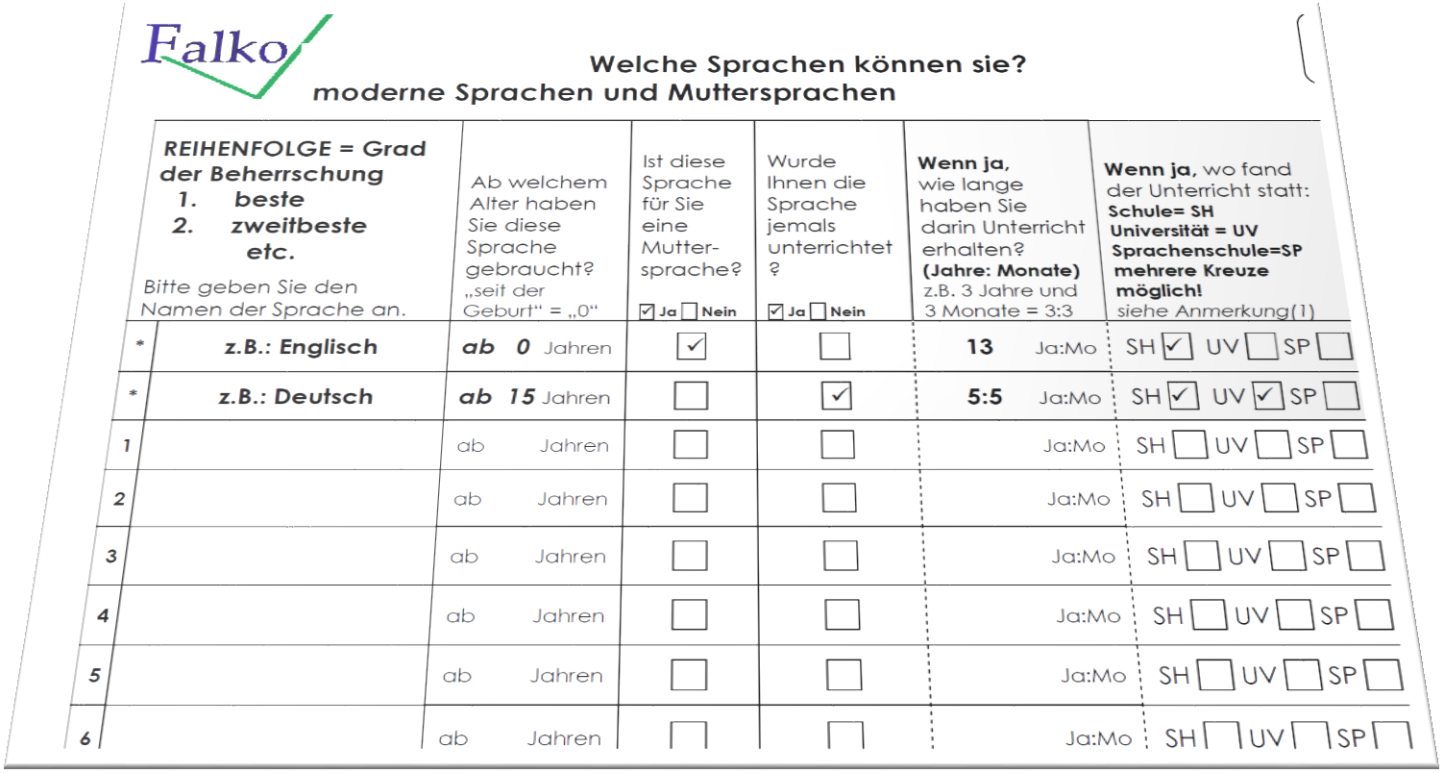
- **Verschiedene Muttersprachler**
  - englisch
  - französisch
  - russisch
  - polnisch
  - dänisch
  - ...
- **Deutsches Kontrollkorpus**
  - Abiturienten
  - Germanistik-Studierende

Metadaten

- **Wiederverwertbarkeit**
  Unterschiedliche Fragestellungen auf gleicher Datenbasis
  Satzverknüpfung, Wortschatzerwerb, Rechtschreibung
- **Vergleichbarkeit**
  Daten aus unterschiedlichen Erhebungen werde verglichen
  International Learner Corpus of English, Falko
- **Beschränkung**
  Nur kontrollierte Teile der Daten werden untersucht
  Fortgeschrittene Lerner, Männer

Metadaten

- **Erhebung**
  - nur relevante Daten (Dörnyei 2003)
  - maximale Ausfülldauer: 30 Min
  - maximal 2 Seiten



Falko

Welche Sprachen können sie?
moderne Sprachen und Muttersprachen

	REIHENFOLGE = Grad der Beherrschung ***1. beste*** ***2. zweitbeste*** ***etc.*** *Bitte geben Sie den Namen der Sprache an.*	Ab welchem Alter haben Sie diese Sprache gebraucht? „seit der Geburt" = „0"	Ist diese Sprache für Sie eine Mutter-sprache? ☑ Ja ☐ Nein	Wurde Ihnen die Sprache jemals unterrichtet? ☑ Ja ☐ Nein	**Wenn ja,** wie lange haben Sie darin Unterricht erhalten? **(Jahre: Monate)** z.B. 3 Jahre und 3 Monate = 3:3	**Wenn ja,** wo fand der Unterricht statt: **Schule= SH Universität = UV Sprachenschule=SP mehrere Kreuze möglich!** *siehe Anmerkung(1)*
*	***z.B.: Englisch***	***ab 0*** Jahren	☑	☐	**13** Ja:Mo	SH ☑ UV ☐ SP ☐
*	***z.B.: Deutsch***	***ab 15*** Jahren	☐	☑	**5:5** Ja:Mo	SH ☑ UV ☑ SP ☐
1		ab Jahren	☐	☐	Ja:Mo	SH ☐ UV ☐ SP ☐
2		ab Jahren	☐	☐	Ja:Mo	SH ☐ UV ☐ SP ☐
3		ab Jahren	☐	☐	Ja:Mo	SH ☐ UV ☐ SP ☐
4		ab Jahren	☐	☐	Ja:Mo	SH ☐ UV ☐ SP ☐
5		ab Jahren	☐	☐	Ja:Mo	SH ☐ UV ☐ SP ☐
6		ab Jahren	☐	☐	Ja:Mo	SH ☐ UV ☐ SP ☐

Metadaten

Lernerdaten

- Alter, Geschlecht, höchster Abschluss …
- Sprachbiografe: erste Fremdsprache, zweite …

Erhebungsdaten

- Erhebungsdatum, -ort,
- Modalität: geschrieben
- Thema: Studium, Entlohnung, Kriminalität, Feminismus

Fehleranalyse

Beispiel: Artikelverwendung

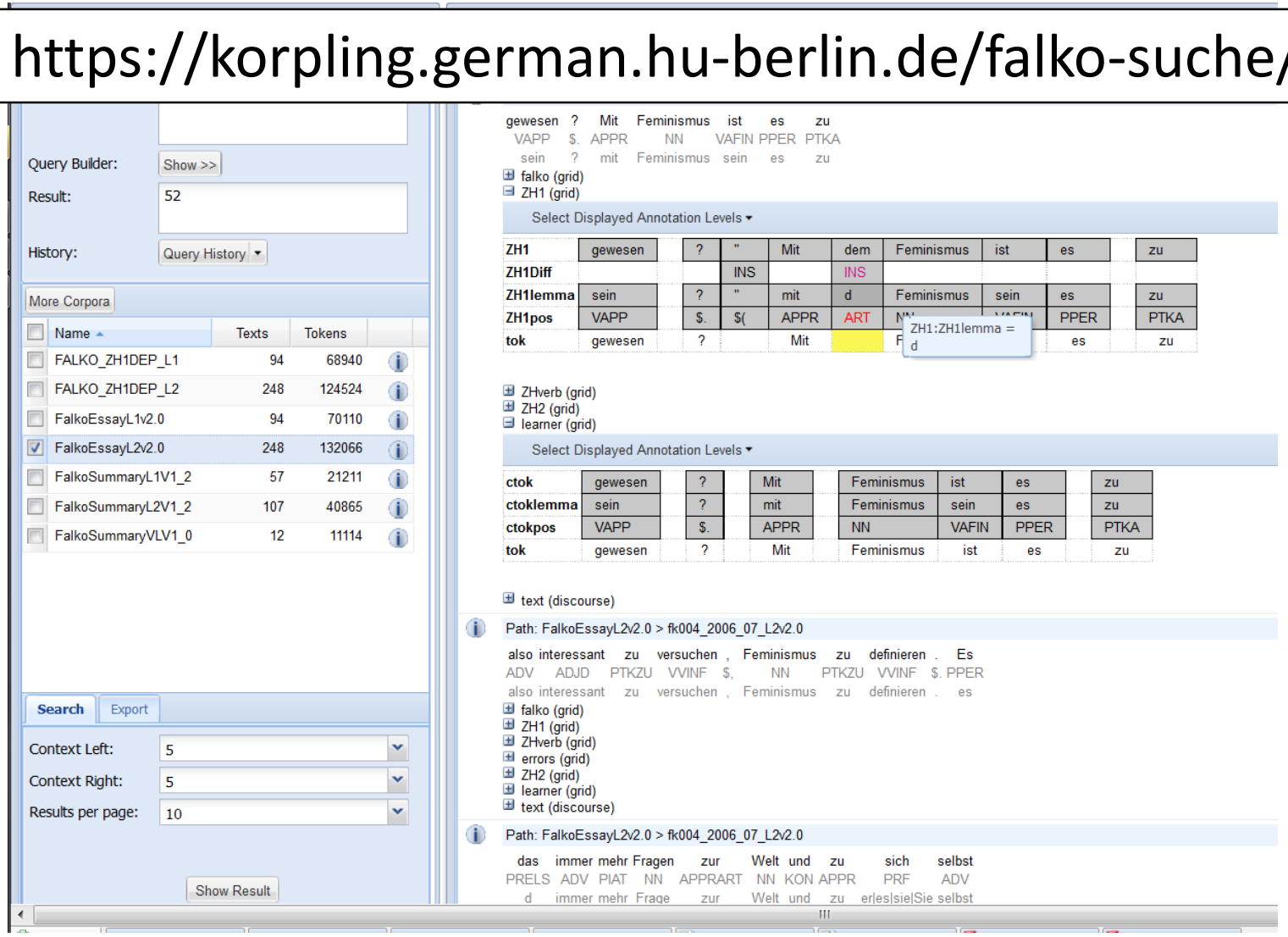
- **Hypothese:**

  Lerner aus Pro-Drop-Sprachen lassen den Artikel häufiger weg als Sprecher aus Nicht-Pro-Drop-Sprachen.

- **Vergleich zweier Sprachgruppen**
  - romanische L1 (Pro-Drop-Sprache / ohne Französisch)
  - germanische L1 (keine Pro-Drop-Sprache)
- **Suche**
  - Suche nach allen verwendeten Artikeln
  - Suche nach allen vergessenen Artikeln
  - Vergleich der relativen Häufigkeiten

Fehleranalyse

https://korpling.german.hu-berlin.de/falko-suche/



Fehleranalyse

Beispiel: Artikelverwendung

- **Vergleich der relativen Häufigkeiten**
  - Lerner aus Pro-Drop-Sprachen lassen den Artikel häufiger weg als Sprecher aus Nicht-Pro-Drop-Sprachen

p<0.01	**romanische Lerner SPA + ITA**	**germanische Lerner DAN**
Anzahl Texte	6 + 6 = 12	40
Artikel verwendet	35 + 371 = 406	1854
Artikel vergessen	18 + 16 = 34	54
Artikel gesamt	323 + 387 = 700	1908
% unterlassener Artikel	4,86 %	2,83 %

Zusammenfassung

- Es gibt **systematische Unterschiede** zwischen der sprachlichen Verwendung zwischen **Fremdsprachenlernern** und **Muttersprachlern**
- Diese können anhand von **Frequenzunterschieden** mithilfe von **Lernerkorpora** untersucht werden
- **EINE** interessante Betrachtung sind **Fehleranalysen**, für die eine **explizite und einheitliche Zielhypothese** nötig ist.
- In **kontrolliert erhobenen** und **gut dokumentierten** Korpora (**Metadaten**) lassen sich dann Gruppen kontrastieren
- Solche **Suchen** lassen sich u.a. in **ANNIS** durchführen.

Vielen Dank!

Literatur

Ellis, Rod (Hrsg.) (2009): The study of second language acquisition. Oxford [u.a.]: Oxford Univ. Press (= Oxford applied linguistics).

Granger, Sylviane (2008): Learner Corpora. In: Anke Lüdeling, Merja Kytö (Eds.): Corpus linguistics. Berlin, New York: Mouton de Gruyter (Handbücher zur Sprach- und Kommunikationswissenschaft / Handbooks of Linguistics and Communication Science, 29,1), 259–275.

Heringer, Hans-Jürgen (2001): Fehlerlexikon. Deutsch als Fremdsprache ; [aus Fehlern lernen: Beispiele und Diagnosen]. 1. Aufl., 1. Dr. Grammatikteil. Berlin: Cornelsen.

Lennon, Paul (1991): Error. Some Problems of Definition, Identification, and Distinction. In: Applied Linguistics 12 (2), S. 180–196.

Miller, George A. (1966): Language and Psychology. In: Lenneberg, Eric H. (Hrsg.): *New directions in the study of language.* Cambridge, Mass: The MIT Press.

Spillner, Bernd (1991): Error analysis. A comprehensive bibliography. Amsterdam: Benjamins (= Series 5, Library and information sources in linguistics; 12).