

KOLAS – EIN LERNENDENKORPUS DER DEUTSCHEN WISSENSCHAFTSSPRACHE

- Was ist KoLaS?
- Wie wurde das Korpus erhoben?
- Wie sehen die Daten aus?
- Was wurde bisher mit dem Korpus gemacht?
- Welche Pläne gibt es für die Weiterentwicklung des Korpus?

ÜBERBLICK





WAS IST KOLAS?

WAS IST KOLAS?



WAS IST KOLAS?

- **K**ommentiertes **L**ernendenkorpus **a**kademisches **S**chreiben
- eine Sammlung studentischer Texte
- erhoben an der Schreibwerkstatt Mehrsprachigkeit der Universität Hamburg (Leitung: Dr. Dagmar Knorr)
- für die Verwendung in Forschung und Lehre frei verfügbar

WAS IST KOLAS?



ZIELE

- Schwierigkeiten beim wissenschaftlichen Schreiben auf Deutsch identifizieren
- anhand von Texten aus dem Studienalltag, die nicht experimentell erhoben werden
- Beratungsangebote darauf ausrichten
- Beratende entsprechend schulen

WAS IST KOLAS?



LERNERKORPORA ZUM WISSENSCHAFTLICHEN SCHREIBEN

- The British Academic Written English (BAWE, Warwick)
- The Corpus of Academic Learner English (CALE, Marcus Callies, Bremen)
- ...

WAS IST KOLAS?



AUSZUG

tion verfügt ([Gabka, 1978: 88](#)).

Nach Böttger (2008: 83) lassen sich im Russischen drei Typen von Stimmassimilationsregeln unterscheiden. Die erste Regel besagt, dass im Wortauslaut stehende stimmhafte Konsonanten /b/, /bʹ/, /v/, /vʹ/, /d/, /dʹ/ usw. stimmlos ausgesprochen werden.

Diesen phonologischen Prozess wird als „[Entstimmlichung im absoluten Auslaut](#)“ bezeichnet ([Gabka, 1978: 86](#)). Die Autoren ([Gabka, 1978: 86](#); Böttger, 2008: 83) haben

Autor
Kommentar [4]: Ich bin mir nicht ganz sicher ob ich diesen Absatz richtig verstehe. Wir sollten ihn am Montag besprechen.

Autor
Kommentar [5]: [Wieviele](#) gibt es denn noch?

WAS IST KOLAS?



WIE WURDE DAS KORPUS ERHOBEN?

DATENERHEBUNG



PEER TUTOR/INNEN SCHREIBBERATUNG

Schreibwerkstatt Mehrsprachigkeit

- Angebot der Schreibwerkstatt Mehrsprachigkeit (Knorr/Neumann 2014)
- Beratungsangebot für alle Studierenden der Universität Hamburg
- wöchentliche Beratung durch ausgebildete Peer Tutor/innen

DATENERHEBUNG



PEER TUTOR/INNEN SCHREIBBERATUNG

Schreibwerkstatt Gesprächigkeit

- Beratung zu Texten , die die Studierenden aktuell für ihr Studium schreiben
- Bei Einverständnis der Studierenden: anonymisierte Aufnahme der Texte ins Korpus KoLaS

DATENERHEBUNG



DATENERHEBUNG

- Datenerhebung wird nicht durch eine spezifische Fragestellung geleitet
- abhängig von den jeweiligen Ratsuchenden und ihren Texten
- „opportunistisches Korpus“ (Teubert & Cermáková 2004:120)
- dafür: besonders authentische Texte

DATENERHEBUNG



ANONYMISIERUNG

- personenbezogene Daten auf dem Titelblatt
 - direkte Ansprachen der Ratsuchenden in Kommentaren oder im Hauptdokument
 - Autornamen in Wordkommentaren
 - Dokument-Metadaten
 - Namensnennungen in Freitextfeldern in den Metadaten
- manuelle, durch ein Skript unterstützte
Bereinigung

DATENERHEBUNG



WIE SEHEN DIE DATEN AUS?

KORPUSBESCHREIBUNG



WIE SEHEN DIE DATEN AUS?

- 453 Texte (ca. 1 Mio. Wortformen)
- von 49 Studierenden der Universität Hamburg (L1 und L2)
- aus dem Zeitraum September 2011 bis Dezember 2013
- überwiegend im doc(x)-Format
- z. T. mit Kommentare von Peer Tutor/innen

KORPUSBESCHREIBUNG



TEXTSORTEN

- Hausarbeiten
- Bachelor- und Masterarbeiten
- Praktikumsberichte
- Protokolle
- ...

KORPUSBESCHREIBUNG



TEXTFORTSCHRITT

- Studierende können zu jedem Zeitpunkt im Schreibprozess in die Beratung kommen:
 - Ideensammlung, Themenfindung
 - Entwicklung der Argumentation
 - Darstellung von Forschungsliteratur und eigenen Ergebnissen
 - sprachliche Überarbeitung

KORPUSBESCHREIBUNG



TEXTFORTSCHRITT: VERSIONEN

- regelmäßiger Besuch der Schreibwerkstatt
 - mehrere Textversionen, die den Schreibprozess dokumentieren

KORPUSBESCHREIBUNG



TEXTVERSION 1 MIT KOMMENTAREN

Welche Funktionen sieht Bloch in der Geschichte? ¶

Bloch fängt **seinen Text** an mit der Frage, wozu eigentlich die Geschichte dient, und versucht anschließend sie zu beantworten. Zuerst erkennt er die Unterhaltungsfunktion. Teil davon ist der Charme, den sie auf die menschliche Vorstellungskraft ausübt. An dieser Stelle unterstreicht er die poetische Dimension der Geschichte. Danach **erwähnt** er die Funktion der Geschichte als Verständnismittel, **als Interpretation von Ereignissen** und nicht als die ohne Zusammenhang Aufzählung von Tatsachen. Denn die Geschichte als Gegenstand den Mensch und seine Taten **hat**, fällt sie ins Gewicht. Schließlich diskutiert er die Funktion der Geschichte als Wissenschaft, dass sie uns, die Menschen, helfen sollte, unsere Leben zu verbessern. Das hält er für nicht so wichtig (eigentlich habe ich diesen Punkt nicht genau **verstanden**). ¶

Außerdem betrachtet er die Geschichte als eine Wissenschaft, die noch jung ist, und er hindeutet auf die Appalos Schule und ihr Entfernung von die Politik- und Militärgeschichte und

Autor

Kommentar [1]: Ich würde von dem Text als Gegenstand abstrahieren. ¶

Autor

Kommentar [2]: erwähnen ist etwas eher oberflächliches. Meinst du das hier? ¶

Autor

Kommentar [3]: Gehört das noch zum Punkt "Verständnismittel" oder ist das schon die nächste Funktion? ¶

Autor

Kommentar [4]: In diesem Satz fehlt etwas. Vielleicht ein *weil* oder *da*? ¶

Autor

Kommentar [5]: ☺ ¶

TEXTVERSION 2, ÜBERARBEITET

Welche Funktionen sieht Bloch in der Geschichte? ¶

¶

Bloch stellt die Frage, wozu eigentlich die Geschichte dient, und versucht anschließend sie zu beantworten. Zuerst erkennt er die Unterhaltungsfunktion. Teil davon ist der Charme, den sie auf die menschliche Vorstellungskraft ausübt. An dieser Stelle unterstreicht er die poetische Dimension der Geschichte. Danach erläutert er die Funktion der Geschichte als Interpretation von Ereignissen und nicht als zusammenhangslose Aufzählung von Tatsachen. Die Geschichte fällt ins Gewicht, denn sie hat den Mensch und seine Taten als Gegenstand. Schließlich diskutiert er die Funktion der Geschichte als Wissenschaft, dass sie uns, den Menschen, helfen sollte, unsere Leben zu verbessern. Diese Funktion hält er für weniger zentral. ¶

¶

METADATEN ZU DEN STUDIERENDEN

- Studienfach (insb. Erziehungswissenschaft, Geisteswissenschaften)
- Semester (zwischen 1 und 30, Median: 6)
- Sprachbiografie: Welche Sprachen...
 - sprechen die Eltern?
 - wurden zuhause gesprochen?
 - wurden in der Schule gelernt?+ Freitextfeld

KORPUSBESCHREIBUNG



SPRACHLICHER HINTERGRUND: AUFGEWACHSEN MIT...

- 10- bis 11-mal genannt:

Deutsch, Russisch, Farsi/Persisch

- 2- bis 5-mal genannt:

Spanisch, Türkisch, Französisch, Englisch, Indonesisch, Japanisch,
Kasachisch, Tschechisch

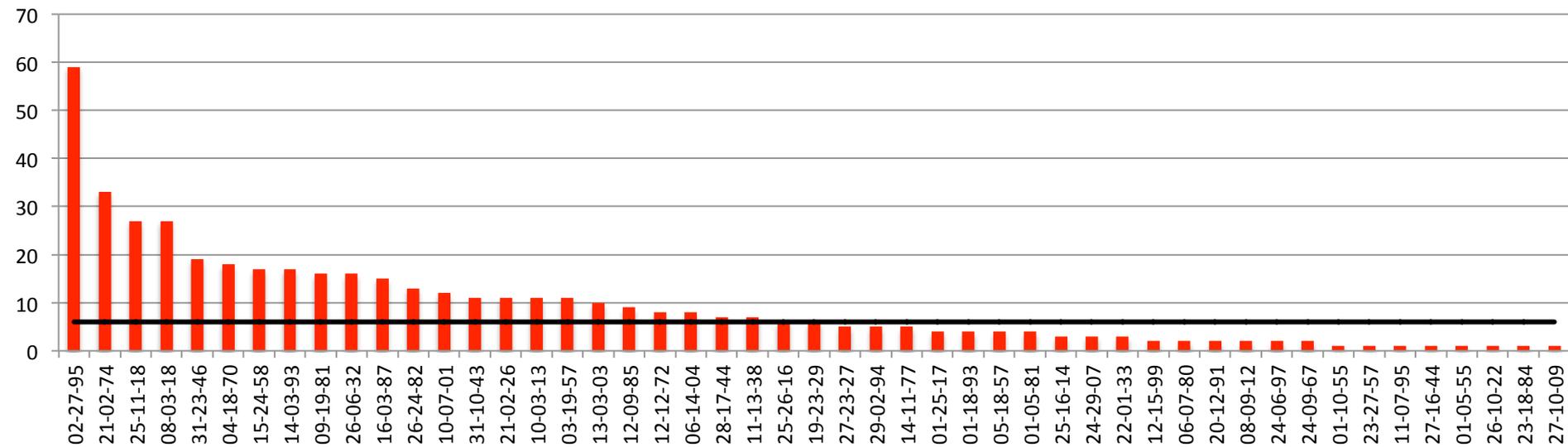
- einmal genannt:

Bassa, Chinesisch, Griechisch, Hindi, Kandhari, Katalan, Kirgisisch,
Kroatisch, Kurdisch, Multani, Slowakisch, Sundonesisch, Twi, Ukrainisch,
Yuruba

KORPUSBESCHREIBUNG



TEXTE PRO RATSUCHENDER/M



Median: 6 Texte

KORPUSBESCHREIBUNG



METADATEN ZU DEN BERATUNGSSITUATIONEN

- Datum und Dauer
- beteiligte Personen (i.d.R. ein/e Ratsuchende/r und ein/ Peer Tutor/in)
- besprochener Textabschnitt (Einleitung, Forschungsstand, ...)
- Freitextfeld für Notizen
- vereinzelt: Audioaufnahmen der Beratung und Transkriptionen

KORPUSBESCHREIBUNG



WAS WURDE BISHER MIT DEM KORPUS GEMACHT?

KORPUSNUTZUNG



FORSCHUNG ZUR WISSENSCHAFTSSPRACHE

- Wie unterscheidet sich die Verwendung des Verbs *diskutieren* in Presse- und Wissenschaftssprache?
- Wie verwenden Lernende das Verb? (Andresen 2016)

KORPUSNUTZUNG



FORSCHUNG ZUR WISSENSCHAFTSSPRACHE (Andresen 2016)

Fokus: Wie wird das Thema von *diskutieren* realisiert?

	Präpos.obj. mit <i>über</i>	direktes Objekt	Nebensatz	kein Thema	Summe
WissKo	9	109	16	–	134
PressKo	38	45	8	14	105
Summe	47	154	24	14	

KORPUSNUTZUNG



FORSCHUNG ZUR WISSENSCHAFTSSPRACHE (Andresen 2016)

Pressesprache:

- (1) In Fachkreisen werde seit längerem über das Handyverbot am Steuer im § 23 diskutiert. (BRZ, 21.08.2009)

Wissenschaftssprache:

- (2) In diesem Aufsatz werden zunächst die Besonderheiten beider Sprachen auf lexikalischer und syntaktisch-semantischer Ebene diskutiert. (InfoDaF_Bauer)

FORSCHUNG ZUR WISSENSCHAFTSSPRACHE (Andresen 2016)

Fokus: Wie wird das Thema von *diskutieren* realisiert?

	Präpos.obj. mit <i>über</i>	direktes Objekt	Nebensatz	kein Thema	Summe
WissKo	9	109	16	–	134
PressKo	38	45	8	14	105
Summe	47	154	24	14	

KORPUSNUTZUNG



FORSCHUNG ZUR WISSENSCHAFTSSPRACHE (Andresen 2016)

Wissenschaftssprache:

- (3) Nach wie vor stellen eventuelle Geschlechtsunterschiede in kognitiven Leistungen ein spannendes und interessantes Thema dar, über das auch öffentlich diskutiert wird.
(ZEP_Niklas)

FORSCHUNG ZUR WISSENSCHAFTSSPRACHE (Andresen 2016)

Lernensätze aus KoLaS:

(4) Im Theorie-Teil der Arbeit werden wir über Mehrsprachigkeit in Deutschland diskutieren und den Begriff definieren. (13-08-25_Einleitung_21-02-74_SB03)

→ In der Wissenschaftssprache eher unübliche Verwendung von *über*

KORPUSNUTZUNG



FORSCHUNG ZUR WISSENSCHAFTSSPRACHE (Andresen 2016)

Lernensätze aus KoLaS:

(5) Was ein „gutes Verhalten“ ausmacht und wie es anzustreben ist, ist bereits viel diskutiert worden. (12-11-26_Exposee_01-18-93)

→ Vorwissenschaftliche Relevanzbegründung

KORPUSNUTZUNG



STUDENTISCHE ARBEITEN

Bachelorarbeit: „Einleitungen: Exemplarische Analysen studentischer Arbeiten“ (Anna Tilmans, 2013)

Masterarbeit: „*Im Theorie-Teil der Arbeit werden wir über Mehrsprachigkeit diskutieren* – Verwendung von Sprechhandlungsverben in der deutschen Wissenschaftssprache“ (Melanie Andresen, 2014)

Sammelbandartikel: „Textkommentierung unter Berücksichtigung von Nicht-Direktivität“ (Özlem Alagöz-Bakan, 2015)

Hausarbeit: „Das Ich-Vorkommen in studentischen Hausarbeiten“ (Claudia Stern, 2015)

KORPUSNUTZUNG



ENTWICKLUNG EINES ANNOTATIONSSCHEMAS FÜR KOMMENTARE

- Wie kommentieren Studierende Texte von Mitstudierenden?

genauer:

- Auf welche Phänomene im Text reagieren sie?
- Wie wird die Rückmeldung sprachlich umgesetzt?
(Beyer/von Gunten 2016:10)

KORPUSNUTZUNG



ENTWICKLUNG EINES ANNOTATIONSSCHEMAS FÜR KOMMENTARE

- Wie kommentieren Studierende Texte von Mitstudierenden?
- Nutzung von KoLaS zur Weiterentwicklung eines an den primären Analysedaten entwickelten Annotationsschemas (Beyer 2016)

KORPUSNUTZUNG



NUTZUNG IN DER PEER TUTOR/INNEN-AUSBILDUNG

- Ziel:
Peer Tutor/innen in der Beratung zum Thema *ich*-Verwendung in der Wissenschaftssprache schulen
- Weg:
Data-Driven Learning (Johns 1991) anhand von Korpusdaten

KORPUSNUTZUNG



NUTZUNG IN DER PEER TUTOR/INNEN-AUSBILDUNG

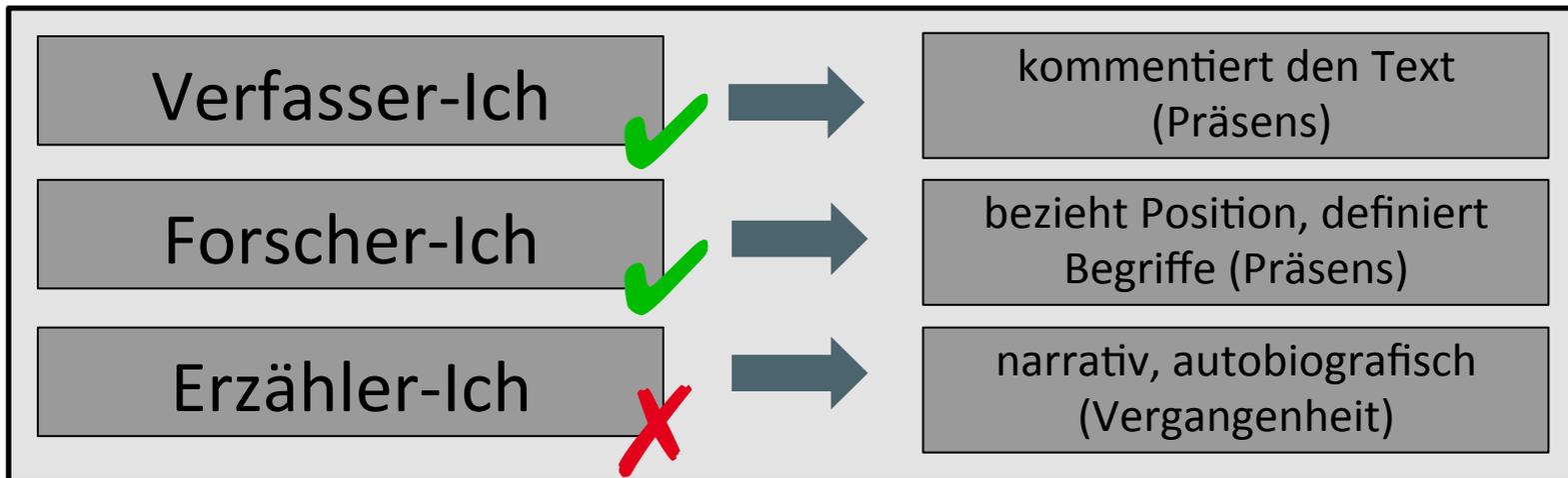
1. Modell zur *ich*-Nutzung kennenlernen
2. Expertensätze analysieren
3. Lernendensätze analysieren
4. Feedback schreiben

(Phasen orientiert an Thompson/Tribble 2001:101)

KORPUSNUTZUNG



EIN MODELL ZUR *ICH*-NUTZUNG KENNENLERNEN



basierend auf Steinhoff (2007)

KORPUSNUTZUNG

Ko **LaS**

ANALYSE VON EXPERTENSÄTZEN

(6) In diesem Beitrag möchte ich mich insbesondere mit Podcasts im Fremdsprachenunterricht beschäftigen.
(InfoDaF_Stork)

→ Verfasser- oder Forscher-Ich?

ANALYSE VON LERNENDENSÄTZEN

(7) Ich habe mich für diese Zwei entschieden, weil ich sie öfter bei anderen kommunikativen Herausforderungen benutzt habe und mich mit ihrer Anwendung sicher fühle.

(KoLaS, 11-12-13_HA-Analyse-einer-kommunikativen-Herausforderung_24-29-07)

Forscher-Ich → Erzähler-Ich

KORPUSNUTZUNG



LERNOUTPUT

- differenzierteres Wissen über die *ich*-Verwendung
- Bewusstsein für unterschiedliche Positionen dazu
- Bewusstsein für die Grenzen theoretischer Modelle
- Konfrontation mit authentischen Lernensätzen und Vorbereitung auf die Beratungssituation

KORPUSNUTZUNG





WELCHE PLÄNE GIBT ES FÜR DIE WEITERENTWICKLUNG DES KORPUS?

ZUKUNFTSPLÄNE



KORPUSERWEITERUNG

- Aufnahme weiterer Texte aus den Jahren 2014–2016 (342 Texte)
- aktuell: Anonymisierung der Texte

ZUKUNFTSPLÄNE



KORPUSERWEITERUNG: ESSAYKORPUS

- Erhebung im Rahmen der Lehrveranstaltung „Psychische Bedingungen von Bildung und Erziehung“ im Sommersemester 2016
- 126 BA-Studierende (Erziehungswissenschaft) im 2. Semester
- pro Person liegt ein Essay mit Kommentaren geschulter Fachtutoren vor
- 39 überarbeitete Fassungen

ZUKUNFTSPLÄNE



KORPUSZUGRIFF VERBESSERN

Ziele:

- Texte leichter durchsuchbar machen
- Suchabfragen über alle Texte ermöglichen (KWIC)
- Frequenzinformationen verfügbar machen
- Zugriff auf die Kommentare erleichtern
- einfache Filtermöglichkeit zur Erstellung von Teilkorpora schaffen
- Lösung vom softwarespezifischen, proprietären Format doc(x)

ZUKUNFTSPLÄNE



KORPUSZUGRIFF VERBESSERN

- Weg:
Konvertierung in cqp-Korpus („corpus query processor“),
abfragbar über die IMS Open Corpus Workbench
- Umsetzung in Kooperation mit der Universität
Hildesheim (Gertrud Faaß, Ulrich Heid)

ZUKUNFTSPLÄNE



ANREICHERUNG MIT ZUSÄTZLICHEN INFORMATIONEN

Textbezogene Metadaten:

- Welche Textteile enthält der Text (z. B. Einleitung, Hauptteil, Schluss)?
- Wie weit fortgeschritten ist der Schreibprozess?

Wortbezogene Annotationen:

- Lemmatisierung
- Wortarten-Tagging

ZUKUNFTSPLÄNE



KOLAS NUTZEN!

- Wir freuen uns, wenn auch andere mit KoLaS arbeiten!
- frei verfügbar für Forschung und Lehre
- einfache Beantragung über das HZSK
(Hamburger Zentrum für Sprachkorpora):
<https://corpora.uni-hamburg.de/drupal/>

ZUKUNFTSPLÄNE



**DANKE FÜR IHRE
AUFMERKSAMKEIT!**

Direkt zur KoLaS-
Webseite:



www.uhh.de/uk-kolas

FRAGEN UND KOMMENTARE?

Direkt zur KoLaS-
Webseite:



www.uhh.de/uk-kolas

LITERATUR

- Alagöz-Bakan, Özlem. 2016. Textkommentierungen unter Berücksichtigung der Nicht-Direktivität. In Özlem Alagöz-Bakan, Dagmar Knorr & Kerstin Krüsemann (eds.), *Akademisches Schreiben. Sprache zum Schreiben - zum Denken - zum Beraten*, 59–66. (Universitätskolleg-Schriften 14). Hamburg.
- Andresen, Melanie. 2016. Im Theorie-Teil der Arbeit werden wir über Mehrsprachigkeit diskutieren – Sprechhandlungsverben in Wissenschafts- und Pressesprache. *Zeitschrift für angewandte Linguistik* 64(1). 47–66. doi:10.1515/zfal-2016-0001.
- Andresen, Melanie & Dagmar Knorr (eingereicht). KoLaS – Ein Lernendenkorpus in der Schreibberatungsausbildung einsetzen.
- Beyer, Anke. 2016. *InliAnTe: Instrument für die linguistische Analyse von Textkommentierungen*. (Arbeitspapiere Projekt „Texte Kommentieren“ 1). Bern: PHBern. https://www.phbern.ch/fileadmin/user_upload/Dokumente-Microsites/Textkommentierungsverhalten/Arbeitspapier_1_InliAnTe_aktuell.pdf (5 October, 2016).
- Beyer, Anke & Anne von Gunten. 2016. *Wie Studierende Texte von Peers kommentieren. Exemplarische Kontrastierung von zwei Fällen und erste hochschuldidaktische Schlussfolgerungen*. (Arbeitspapiere Projekt „Texte kommentieren“ 3). Bern: PHBern.
- Johns, Tim. 1991. Should you be persuaded. Two samples of data-driven learning materials. *Classroom Concordancing: ELR Journal*(4). 1–16.
- Knorr, Dagmar & Ursula Neumann. 2014. Die Schreibwerkstatt Mehrsprachigkeit – (Lehramts-)Studierende mit Migrationshintergrund der Universität Hamburg schreiben. In Dagmar Knorr & Ursula Neumann (eds.), *Mehrsprachige Lehramtsstudierende schreiben: Schreibwerkstätten an deutschen Hochschulen*, 119–144. (FÖRMIG-Edition 10). Münster: Waxmann.
- Steinhoff, Torsten. 2007. Zum *ich*-Gebrauch in Wissenschaftstexten. *Zeitschrift für germanistische Linguistik* 35(1-2). 1–26.
- Teubert, Wolfgang & Anna Cermáková. 2004. Directions in corpus linguistics. In M. A. K. Halliday, Wolfgang Teubert, Colin Yallop & Anna Cermáková (eds.), *Lexicology and Corpus Linguistics. An Introduction*, 113–165. (Open Linguistics Series). London, New York: Continuum.
- Thompson, Paul & Chris Tribble. 2001. Looking at citations: Using Corpora in English for Academic Purposes. *Language Learning and Technology* 5(3). 91–105.