

Einführungsvortrag: Korpora, Korpuslinguistik und datengeleitetes Lernen (DDL)

Prof. Dr. Christian Krekeler
Deutsch als Fremd- und Fachsprache
Hochschule Konstanz HTWG
krekeler@htwg-konstanz.de

Die Vermittlung von sprachlichen Mustern ist das zentrale Anliegen von DDL.

DDL und sprachliche Muster

1. Frage
2. Korpus
3. Suchmaschine



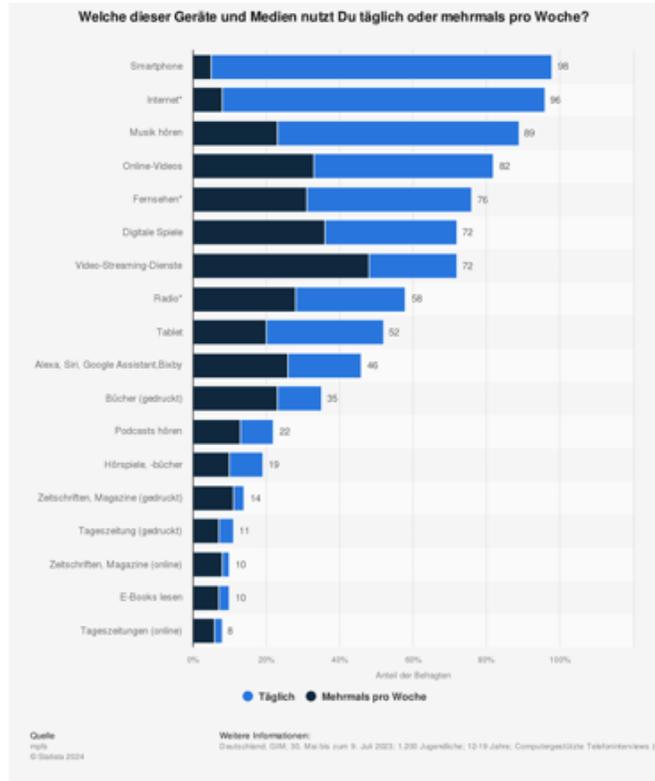
1. **Wie kann man Häufigkeiten angeben?**
2. **Texte: Mediennutzung**
3. **AntConc**

DDL und sprachliche Muster

Korpus „Mediennutzung“ – 10 Texte zu Umfragen

1

Mediennutzung durch Jugendliche in der Freizeit in Deutschland 2023



Basierend auf einer aktuellen Umfrage, die 2023 unter Jugendlichen in Deutschland durchgeführt wurde, gibt es wichtige Einblicke in die Mediennutzung in ihrer Freizeit. Die Daten verdeutlichen, wie digitale Medien im Alltagsleben junger Menschen fest verankert sind, wobei Smartphones und das Internet dominieren.

Der überwiegende Anteil der Jugendlichen nutzt täglich ein Smartphone; konkret gaben 93 Prozent der Befragten an, ihr Smartphone täglich zu verwenden, während weitere 5 Prozent es mehrmals pro Woche nutzen. Ähnlich verhält es sich mit dem Internet, das von 88 Prozent der Jugendlichen täglich und von weiteren 8 Prozent mehrmals pro Woche genutzt wird. Diese hohe Nutzungszahl deutet darauf hin, dass Smartphones und das Internet zentrale Werkzeuge für Kommunikation, Information und Unterhaltung darstellen.

Musik hören ist ebenfalls eine beliebte Freizeitaktivität, wobei 66 Prozent der Jugendlichen täglich Musik hören und 23 Prozent dies mehrmals pro Woche tun. Dies zeigt, dass Musik ein integraler Bestandteil des täglichen Lebens vieler Jugendlicher ist.

Online-Videos und Video-Streaming-Dienste erfreuen sich ebenso großer Beliebtheit. 49 Prozent der Jugendlichen schauen täglich Online-Videos, und 33 Prozent tun dies mehrmals pro Woche. Bei den Video-Streaming-Diensten nutzen 24 Prozent der Jugendlichen diese täglich und 48 Prozent mehrmals pro Woche. Diese Zahlen bestätigen den Trend weg vom traditionellen Fernsehen hin zu flexibleren Online-Angeboten, die es den Nutzern ermöglichen, Inhalte nach Bedarf zu ~~erschaffen~~.

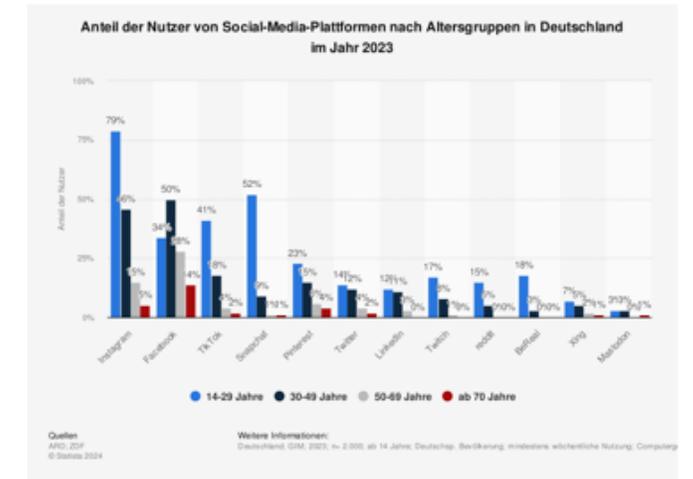
Das klassische Fernsehen wird von 45 Prozent der Jugendlichen täglich und von weiteren 31 Prozent mehrmals pro Woche genutzt, was darauf hindeutet, dass trotz der Popularität digitaler Medien das Fernsehen noch immer einen festen Platz in der Mediennutzung einnimmt.

Interessanterweise zeigt die Umfrage auch, dass traditionellere Medien wie gedruckte Bücher, Zeitschriften und Tageszeitungen weniger häufig genutzt werden, jedoch nicht vollständig obsolet sind. Gedruckte Bücher werden von 12 Prozent der Jugendlichen täglich und von 23 Prozent mehrmals pro Woche gelesen.

Diese Ergebnisse bieten einen umfassenden Überblick über die Mediengewohnheiten deutscher Jugendlicher im Jahr 2023. Sie belegen, dass junge Menschen eine breite Palette von Medien für Bildung, Unterhaltung und soziale Interaktion nutzen, wobei digitale Plattformen führend sind.

2

Anteil der Nutzer von Social-Media-Plattformen nach Alter in Deutschland 2023



Die Nutzung sozialer Medien ist zu einem integralen Bestandteil des täglichen Lebens geworden, und die Vorlieben bei der Wahl der Plattform variieren deutlich je nach Altersgruppe. Eine aktuelle Umfrage aus dem Jahr 2023 liefert aufschlussreiche Daten zur Nutzung von ~~Social~~-Medien-Plattformen nach Altersgruppen in Deutschland.

Instagram erweist sich als äußerst beliebt unter den jüngeren Nutzern, mit einer beeindruckenden Nutzerrate von 79 Prozent in der Altersgruppe der 14- bis 29-Jährigen. Die Beliebtheit nimmt mit steigendem Alter ab, wobei nur 46 Prozent der 30- bis 49-Jährigen, 15 Prozent der 50- bis 69-Jährigen und eine geringe Nutzung von 5 Prozent bei den über 70-Jährigen verzeichnet wird. ~~Instagram~~ visuell orientiertes Format scheint besonders ansprechend für jüngere Nutzer zu sein.

Facebook, einst der unangefochtene Riese unter den ~~Social~~-Medien-Plattformen, zeigt ein interessantes Muster: Es hat die höchste Nutzerrate unter den 30- bis 49-Jährigen mit 50 Prozent, gefolgt von den 14- bis 29-Jährigen mit 34 Prozent. Auch ältere Generationen sind hier noch relativ aktiv, was darauf hindeutet, dass Facebook aufgrund seiner ~~vielfältigen~~ und umfassenden Funktionen über verschiedene Altersgruppen hinweg eine breite Nutzerbasis hat.

DDL und sprachliche Muster

Suchmaschine für Korpusrecherchen: Mehrworteinheiten

Target Corpus
Name: temp
Files: 1
Tokens: 3754
bamberg_korpus_2.docx

KWIC Plot File Cluster **N-Gram** Collocate Word Keyword

N-Gram Types 3377 **N-Gram Tokens** 3752 **Page Size** 100 hits 1 to 100 of 3377 hits

	Type	Rank	Freq	Range
1	von prozent der	1	17	1
2	mehrmals pro woche	2	16	1
3	prozent der befragten	3	15	1
4	prozent der jugendlichen	4	13	1
5	der jugendlichen täglich	5	9	1
6	den bis jährigen	6	8	1
7	der bis jährigen	7	7	1
8	prozent mehrmals pro	7	7	1
9	bei den bis	9	6	1
10	prozent bei den	9	6	1
11	die von prozent	11	5	1
12	dies unterstreicht die	11	5	1
13	für den medienkonsum	11	5	1
14	jugendlichen täglich und	11	5	1
15	social media plattformen	11	5	1
16	werden von prozent	11	5	1
17	altersgruppen in deutschland	17	4	1

Search Query Words Case Regex **N-Gram Size** 3 **Open Slots** 0 **Min. Freq** 1 **Min. Range** 1

DDL und sprachliche Muster

Korpusrecherchen: Stichwortsuche, Konkordanzlinien

Target Corpus
Name: temp
Files: 1
Tokens: 3754
bamberg_korpus_KI.docx

KWIC Plot File Cluster N-Gram Collocate Word Keyword

Total Hits: 98 Page Size 100 hits 1 to 98 of 98 hits

	File	Left Context	Hit	Right Context
4	bam...	dies mehrmals pro Woche. Bei den Video-Streaming-Diensten nutzen 24	Prozent	der Jugendlichen diese täglich und 48 Prozent mehrmals pro Woche.
5	bam...	Inhalte nach Bedarf zu streamen. Das klassische Fernsehen wird von 45	Prozent	der Jugendlichen täglich und von weiteren 31 Prozent mehrmals pro
6	bam...	werden, jedoch nicht vollständig obsolet sind. Gedruckte Bücher werden von 12	Prozent	der Jugendlichen täglich und von 23 Prozent mehrmals pro Woche
7	bam...	Geräts im Alltag der jungen Generation unterstreicht. Genauer gesagt nutzen 93	Prozent	der Jugendlichen täglich ein Smartphone, und weitere 5 Prozent verwenden
8	bam...	ebenfalls ein integraler Bestandteil des jugendlichen Medienkonsums, wobei 88	Prozent	der Jugendlichen täglich und 8 Prozent mehrmals pro Woche online
9	bam...	en oder Streaming-Dienste, bleibt eine beliebte Freizeitbeschäftigung, wobei 66	Prozent	der Jugendlichen täglich Musik hören und 23 Prozent dies mehrmals
10	bam...	d Video-Streaming-Diensten zeigt sich eine interessante Dynamik: Während 49	Prozent	der Jugendlichen täglich Online-Videos anschauen, nutzen 33 Prozent diese
11	bam...	se Möglichkeit mehrmals pro Woche. Video-Streaming-Dienste werden von 24	Prozent	der Jugendlichen täglich und von 48 Prozent mehrmals pro Woche
12	bam...	und das Radio haben noch immer ihren Platz im Mediennutzungsverhalten. 45	Prozent	der Jugendlichen schauen täglich Fernsehen und weitere 31 Prozent mehrmals
13	bam...	gende Anteil der Jugendlichen nutzt täglich ein Smartphone; konkret gaben 93	Prozent	der Befragten an, ihr Smartphone täglich zu verwenden, während
14	bam...	oder Serien ist ebenfalls eine sehr beliebte Tätigkeit, die von 83	Prozent	der Befragten ausgeführt wird. Dies ist ein Hinweis darauf,
15	bam...	den Lernprozess unterstützt und erweitert. Online-Spiele spielen wird von 61	Prozent	der Befragten angegeben, was auf die große Beliebtheit von
16	bam...	An der Spitze der Beliebtheitsskala stehen Nachrichtensendungen, die von 33,5	Prozent	der Befragten favorisiert werden. Dies unterstreicht die kontinuierliche Nachfrag
17	bam...	ndungen ziehen ebenfalls eine große Zuschauerschaft an und werden von 24,2	Prozent	der Befragten regelmäßig gesehen. Die Faszination für Live-Sportereignisse

Search Query Words Case Regex Results Set All hits Context Size 10 token(s)

DDL und sprachliche Muster

in der Regel

mehrmals pro Woche

[Zahl] Prozent der [Gruppe]

sprachliche Muster

formelhafte Sequenzen

Chunks

usuelle Wortverbindungen

DDL und sprachliche Muster

Phraseologismus

Präpositionalphrasen
Präpositionalgruppen

Kookkurrenzen

N-Gramme

Satzbausteine

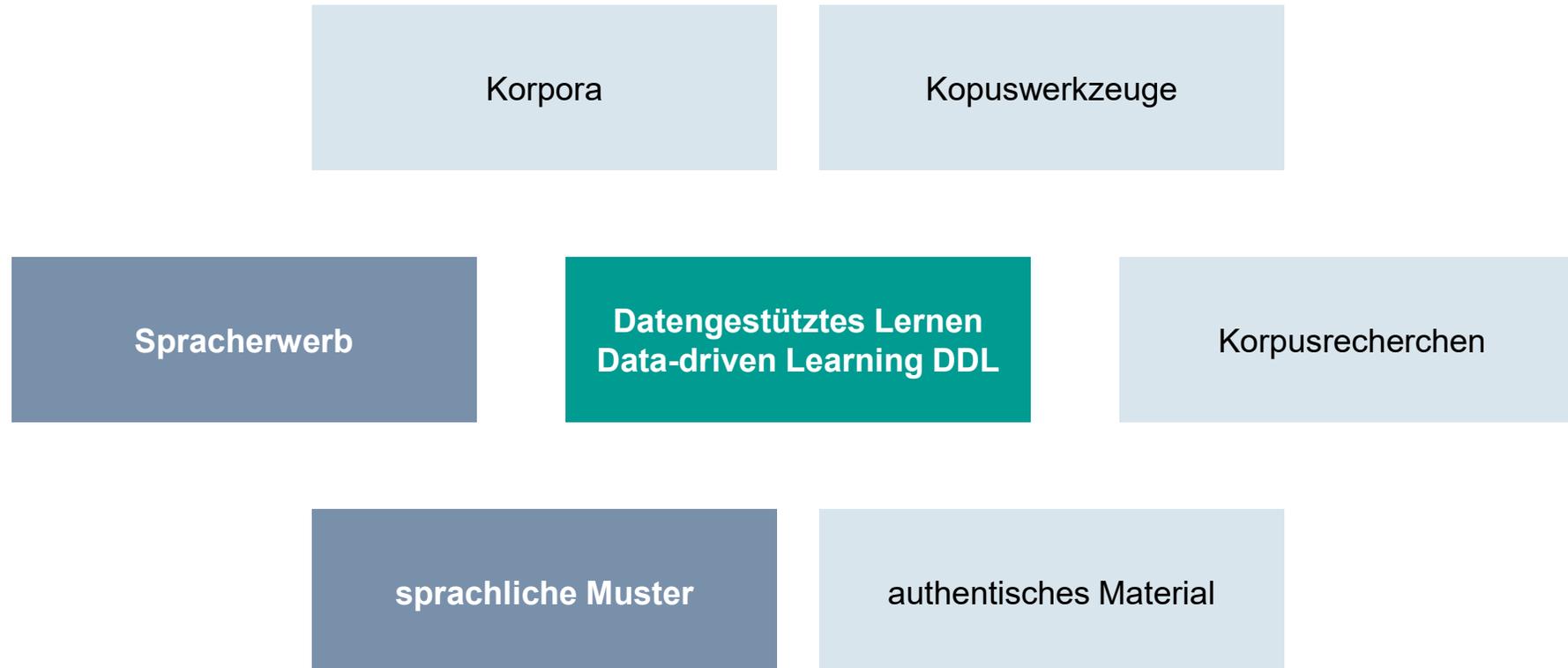
Textbausteine

Diskursmarker

Diskurspartikeln

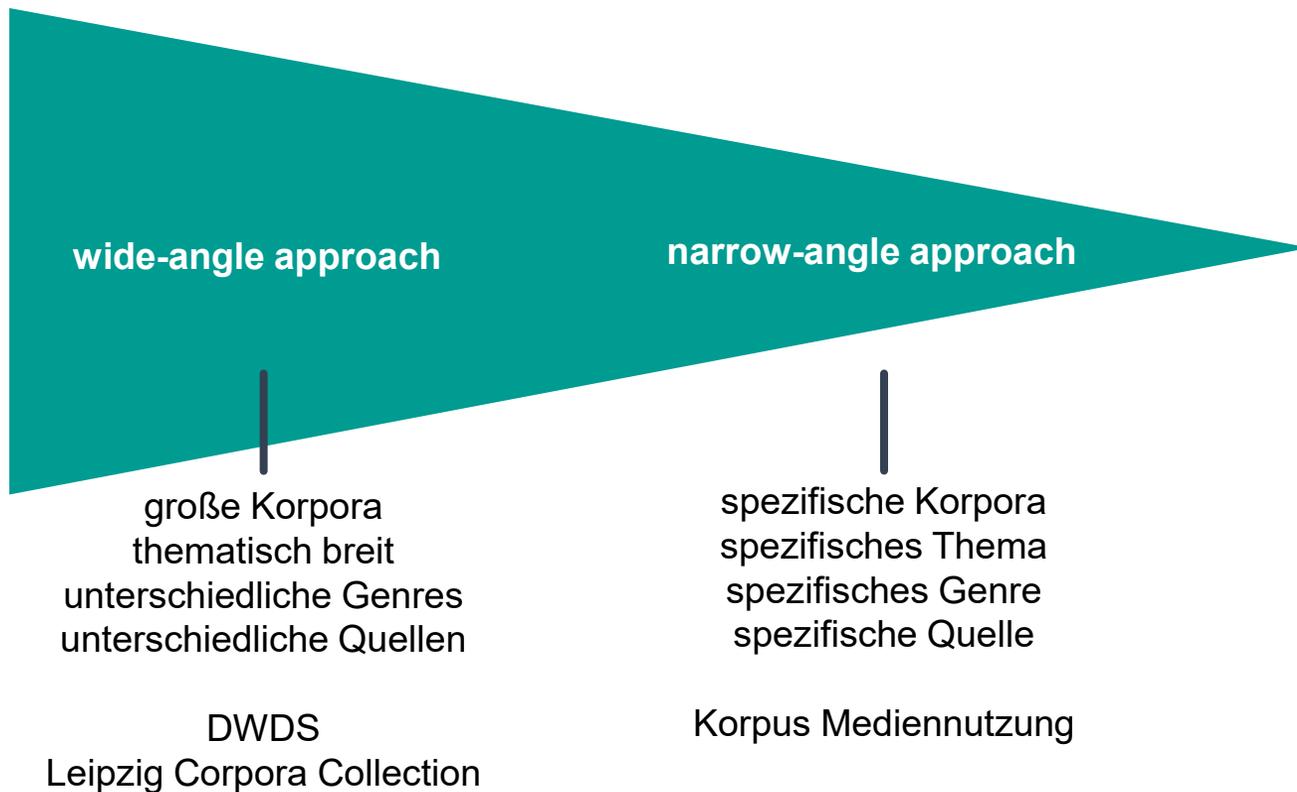
Redemittel

DDL und sprachliche Muster



DDL und sprachliche Muster

Auswahl der Korpora: Wie spezifisch soll das Korpus sein?



English for Specific Purposes 76 (2024) 1–13

Contents lists available at [ScienceDirect](#)

English for Specific Purposes

journal homepage: <http://ees.elsevier.com/esp/default.asp>

A critical review of corpus-based pedagogic perspectives on thesis writing: Specificity revisited

Lynne Flowerdew*, Bojana Petrić

Birkbeck, University of London, UK

ARTICLE INFO

Article history:
Available online 3 June 2024

Keywords:
Thesis and dissertation writing
Needs analysis
Specificity
Data-driven learning
Ethnographic approach
AI/ChatGPT

ABSTRACT

This writing (used here as an umbrella term to cover both master's and doctoral postgraduate-level writing) is a high-stakes genre for postgraduate students. This important student genre has been well-researched from a corpus-based perspective. Corpora of theses and also research articles have been used for data-driven learning (DDL) of this key genre. The purpose of this article is to critically examine key DDL initiatives, some of which take a 'research into practice' orientation. Importantly, the discussion is framed around the notion of 'specificity' in the context of needs analysis, and whether the initiatives take a wide-angle, narrow-angle or move from a wide-angle to a narrow angle approach. Accounts which focus on DIY (do-it-yourself) mini-corpus compilation and use by students are also reviewed. The final section of the article presents a critique of current pedagogic applications, taking a closer look at the issue of 'specificity' within the wider context of needs analysis and mapping out areas for future consideration. It is suggested that an ethnographic perspective may be particularly useful for conceptualising specificity relating to students' present situation needs. The article also considers the impact of AI/ChatGPT on future corpus-based pedagogy of thesis writing.

DDL und sprachliche Muster

vermeintliche Vorteile von DDL

Corpus linguistics and data-driven learning: a critical overview

Alex BOULTON

Université de Lorraine et Crapel, ATILF (UMR 7118)
44 avenue de la libération, 54000 Nancy, France
alex.boulton@univ-lorraine.fr

Henry TYNE

Université de Perpignan Via Domitia et VECT (EA 2983)
52 avenue Paul Alduy, 66000 Perpignan, France
henry.tyne@univ-perp.fr

L'utilisation de corpus s'est avérée intéressante dans de nombreux domaines depuis plus de vingt ans, notamment en didactique des langues. Un état de l'art des études publiées dans ce domaine nous permettra de constater que si les résultats sont en général assez positifs (l'apprentissage a lieu), l'exploitation de corpus reste une activité marginale. Si, en général, de telles recherches en didactique des langues ne donnent pas toujours des résultats fulgurants, on pourrait faire l'hypothèse que les approches sont plus ou moins adaptées à certains types d'apprenants, ou à certains types d'activité, et qu'elles contribuent au bagage général des outils et des techniques disponibles. Qui plus est, avec une approche sur corpus, on pourrait de même croire que les étudiants deviennent tout simplement des apprenants plus conscients de leur apprentissage grâce au travail effectué. Nous suggérons que l'apport des corpus pour l'apprentissage des langues offre ainsi de nombreuses possibilités.

DDL is alleged to

- enhance cognitive and metacognitive skills,
- increase sensitivity to authentic language use,
- provide an interactive approach to constructivist discovery learning,
- foster motivation especially through individualisation,
- promote reusable and transferable skills,
- favour autonomy for life-long learning,
- and correspond largely to current theories of second language acquisition.

Boulton, Alex; Tyne, Henry (2013): „Corpus linguistics and data-driven learning: A critical overview“. In: *Bulletin suisse de Linguistique appliquée* (97), 97–118.

EMPIRICAL STUDY

Corpus Use in Language Learning: A Meta-Analysis

Alex Boulton ✉, Tom Cobb

First published: 15 February 2017 | <https://doi.org/10.1111/lang.12224> | Citations: 218

Abstract

This study applied systematic meta-analytic procedures to summarize findings from experimental and quasi-experimental investigations into the effectiveness of using the tools and techniques of corpus linguistics for second language learning or use, here referred to as data-driven learning (DDL). Analysis of 64 separate studies representing 88 unique samples reporting sufficient data indicated that DDL approaches result in large overall effects for both control/experimental group comparisons ($d = 0.95$) and for pre/posttest designs ($d = 1.50$). Further investigation of moderator variables revealed that small effect sizes were generally tied to small sample sizes. Research has barely begun in some key areas, and durability/transfer of learning through delayed posttesting remains an area in need of further investigation. Although DDL research demonstrably improved over the period investigated, further changes in practice and reporting are recommended.

Boulton, Alex; Cobb, Tom (2017): „Corpus use in language learning: A meta-analysis“. In: *Language Learning* 67 (2), 348–393.



ELSEVIER



Effective corpus use in second language learning: A meta-analytic approach

Shotaro Ueno ^{a,*}, Osamu Takeuchi ^b^a Academic Success Center, Kanda University of International Studies, 1-4-1, Wakaba, Mishima-ku, Chiba-shi, 261-0014, Chiba, Japan^b Faculty of Foreign Language Studies, Kansai University, 3-3-35, Suita-shi, Osaka, Japan

ARTICLE INFO

Keywords:

Data-driven learning
Effect size
Meta-analysis
Moderator-analysis

ABSTRACT

Data-driven learning (DDL) refers to the use of corpora by second and foreign language (L2) learners to explore and inductively discover patterns of their target language use from authentic language data without interventions from others. Although previous meta-analyses have demonstrated the positive effects of DDL on L2 learning (Boulton and Cobb, 2017), the number of empirical studies has been increasing since then. Therefore, this study included more recent studies and used meta-analyses to examine the extent to which: (1) DDL exerts an effect on L2 learning; and (2) moderator variables affect DDL's influence on L2 learning. The results demonstrated small to medium effect sizes for experimental/control group comparisons and pre/post and pre/delayed designs. Moreover, the moderator analyses found that moderator variables, such as publication types, learners' factors, and research designs, influence the magnitude of DDL effectiveness in L2 learning.

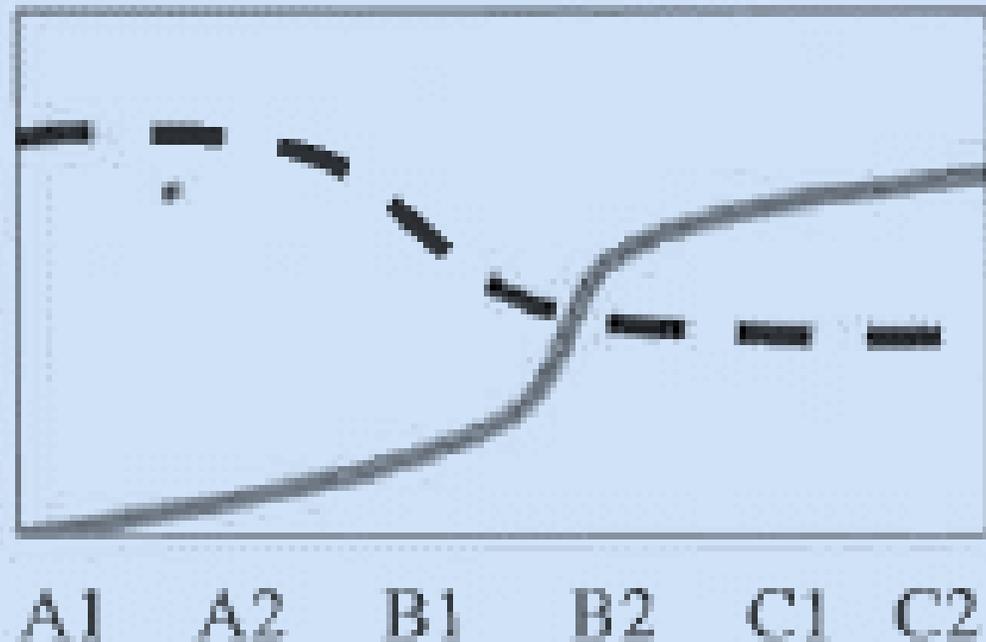
Die Beschäftigung mit sprachlichen Mustern hat einen Wert für den Spracherwerb, wenn es nicht beim Input bleibt.

Der Wert sprachlicher Muster für den Spracherwerb

Grafik 2: Die Schere zwischen *formulaic* und *creative speech*

Formulaic
speech

Creative
speech



Westhoff, Gerard (2007):
„Grammatische Regelkenntnisse
und der GER“. In: *Babylonia* (1),
12–21.

Der Wert sprachlicher Muster für den Spracherwerb



Langfristig gesehen können formelhafte Sequenzen eine wichtige erwerbsrelevante Funktion haben.

Wenn Lerner die ihnen im **Input** angebotenen formelhaften Sequenzen als solche **wahrnehmen** und es ihnen gelingt, mittels Segmentierung und Analyse die zugrundeliegenden **Strukturen und Regeln** zu extrahieren und das auf diese Weise erweiterte Lernerregelsystem anschließend **produktiv zu verwenden**, findet weiterer Spracherwerb statt.

Aguado, Karin (2002): „Formelhafte Sequenzen und ihre Funktionen für den L2-Erwerb“. In: *Zeitschrift für angewandte Linguistik* 37 (2), 27–49.

Entdeckendes Lernen passt zum DDL. Es muss aber etwas entdeckt und gelernt werden.

DDL und Sprachvermittlung

Corpus linguistics and data-driven learning: a critical overview

Alex BOULTON

Université de Lorraine et Crapel, ATILF (UMR 7118)
44 avenue de la libération, 54000 Nancy, France
alex.boulton@univ-lorraine.fr

Henry TYNE

Université de Perpignan Via Domitia et VECT (EA 2983)
52 avenue Paul Alduy, 66000 Perpignan, France
henry.tyne@univ-perp.fr

L'utilisation de corpus s'est avérée intéressante dans de nombreux domaines depuis plus de vingt ans, notamment en didactique des langues. Un état de l'art des études publiées dans ce domaine nous permettra de constater que si les résultats sont en général assez positifs (l'apprentissage a lieu), l'exploitation de corpus reste une activité marginale. Si, en général, de telles recherches en didactique des langues ne donnent pas toujours des résultats fulgurants, on pourrait faire l'hypothèse que les approches sont plus ou moins adaptées à certains types d'apprenants, ou à certains types d'activité, et qu'elles contribuent au bagage général des outils et des techniques disponibles. Qui plus est, avec une approche sur corpus, on pourrait de même croire que les étudiants deviennent tout simplement des apprenants plus conscients de leur apprentissage grâce au travail effectué. Nous suggérons que l'apport des corpus pour l'apprentissage des langues offre ainsi de nombreuses possibilités.

DDL is alleged to

- enhance cognitive and metacognitive skills,
- increase sensitivity to authentic language use,
- **provide an interactive approach to constructivist discovery learning,**
- foster motivation especially through individualisation,
- promote reusable and transferable skills,
- favour autonomy for life-long learning,
- and correspond largely to current theories of second language acquisition.

Boulton, Alex; Tyne, Henry (2013): „Corpus linguistics and data-driven learning: A critical overview“. In: *Bulletin suisse de Linguistique appliquée* (97), 97–118.

DDL und Sprachvermittlung

J Math Didakt (2017) 38:209–237
DOI 10.1007/s13138-017-0116-x



ORIGINALARBEIT/ORIGINAL ARTICLE

Entdeckendes Lernen: Eine Problematisierung.

David Kollosche

Eingegangen: 1. Oktober 2015 / Angenommen: 21. Februar 2017 / Online publiziert: 10. März 2017
© GDM 2017

Zusammenfassung Trotz der nachweislichen Popularität des Entdeckenden Lernens in der deutschsprachigen Mathematikdidaktik finden sich aktuell keine kritischen Beiträge, die dazu beitragen könnten, dieses grundlegende Unterrichtskonzept zu hinterfragen und auszuschärfen. In diesem Diskussionsbeitrag werden zunächst die Theorie und einige Umsetzungsbeispiele des Entdeckenden Lernens herausgearbeitet, um aufzuzeigen, dass das Entdeckende Lernen einem vagen Sammelbegriff gleicht, unter dem oft fragwürdige Unterrichtsumgebungen legitimiert werden. Anschließend werden an Hand erkenntnistheoretischer, lerntheoretischer, didaktischer und soziokultureller Betrachtungen Probleme des Entdeckenden Lernens im Mathematikunterricht und Möglichkeiten ihrer Überwindung thematisiert. Dabei zeigt sich, dass die Konzeption des Entdeckenden Lernens hinter dem aktuellen mathematikdidaktischen Erkenntnisstand zurückfällt und Lehrer sowie Schüler mit unmöglichen Forderungen konfrontiert, dass lerntheoretische Vorteile des Entdeckenden Lernens oft nicht nachweisbar sind, dass die Idee des Entdeckens auf einem problematischen platonistischen Verständnis von Erkenntnis beruht und dass Entdeckendes Lernen bildungsferne Schüler zu benachteiligen droht. Abschließend werden Forschungsdesiderata abgeleitet, deren Bearbeitung dazu beitragen könnte, die aufgezeigten Problemfelder zu überwinden.

EDUCATIONAL PSYCHOLOGIST, 41(2), 75–86
Copyright © 2006, Lawrence Erlbaum Associates, Inc.

Why Minimal Guidance During Instruction Does Not Work: An Analysis of the Failure of Constructivist, Discovery, Problem-Based, Experiential, and Inquiry-Based Teaching

Paul A. Kirschner

*Educational Technology Expertise Center
Open University of the Netherlands
Research Centre Learning in Interaction
Utrecht University, The Netherlands*

John Sweller

*School of Education
University of New South Wales*

Richard E. Clark

*Rossier School of Education
University of Southern California*

Evidence for the superiority of guided instruction is explained in the context of our knowledge of human cognitive architecture, expert–novice differences, and cognitive load. Although unguided or minimally guided instructional approaches are very popular and intuitively appealing, the point is made that these approaches ignore both the structures that constitute human cognitive architecture and evidence from empirical studies over the past half-century that consistently indicate that minimally guided instruction is less effective and less efficient than instructional approaches that place a strong emphasis on guidance of the student learning process. The advantage of guidance begins to recede only when learners have sufficiently high prior knowledge to provide “internal” guidance. Recent developments in instructional research and instructional design models that support guidance during instruction are briefly described.

DDL und Sprachvermittlung

DDL-Aufgabe „Suche in der Wildnis“

Schreiben Sie einen Text über statistische Informationen, die in Prozent angegeben sind.



Führen Sie eine Recherche in einem Korpus nach „Prozent“ durch und verwenden Sie sprachliche Muster mit „Prozent“, die häufig verwendet werden



Korpusbelege Gegenwartskorpora mit freiem Zugang

Prozent 🔍 ? ↺

Korpus: **Start:** **Ende:** **Teilkorpora:**

Anzeige: KWIC voll maximal **Sortierung:** **Treffer pro Seite:**

351–400 von 1078569 Treffern (1083588 insgesamt)

351:	2024	Kommunistisch	ihrem Spitzenkandidaten Kay-Michael Dankl in der Stadt Salzburg 3,8 Prozent , was den erstmaligen Wiedereinzug seit 1962 in den Salzburger Gem...
352:	2024	Stargard. In: Wil	Damit lag der jüdische Bevölkerungsanteil in Stargard bei rund vier Prozent und war somit durchschnittlich drei Mal höher als in anderen Städte...
353:	2024	Stargard. In: Wil	...en dem 1. und 3. März 1945 durch sowjetische Bombenangriffe zu 70 Prozent zerstört.
354:	2024	Bundesautobah	Aufgrund der starken Längsneigungen von mehr als sechs Prozent (statt der höchstens zulässigen vier Prozent) und den engen Kurvenr...
355:	2024	Bundesautobah	...en von mehr als sechs Prozent (statt der höchstens zulässigen vier Prozent) und den engen Kurvenradien legte Bundesverkehrsminister Hans-C...
356:	2024	Tschadsee. In: V	Zwischen 80 und 90 Prozent seines Wassers erhält der See aus den Zuflüssen des Schari (frz. Char...
357:	2024	Tschadsee. In: V	Weniger als zehn Prozent des Zulaufs stammen aus nigerianischen Flüssen und den lokalen Ni...
358:	2024	Ahmedabad. In:	Zwischen 2001 und 2011 wuchs die Einwohnerzahl um über 50 Prozent von 3,6 auf 5,6 Millionen.
359:	2024	Ahmedabad. In:	Bevölkerungswachstum von 4,5 Millionen auf 6,4 Millionen (plus 41 Prozent).
360:	2024	Ahmedabad. In:	Nach der Volkszählung 2011 sind 82 Prozent der Einwohner Ahmedabads Hindus und 14 Prozent Muslime.
361:	2024	Ahmedabad. In:	...ung 2011 sind 82 Prozent der Einwohner Ahmedabads Hindus und 14 Prozent Muslime.
362:	2024	Ahmedabad. In:	Knapp vier Prozent der Bevölkerung Ahmedabads sind Jainas, weniger als ein Prozent si

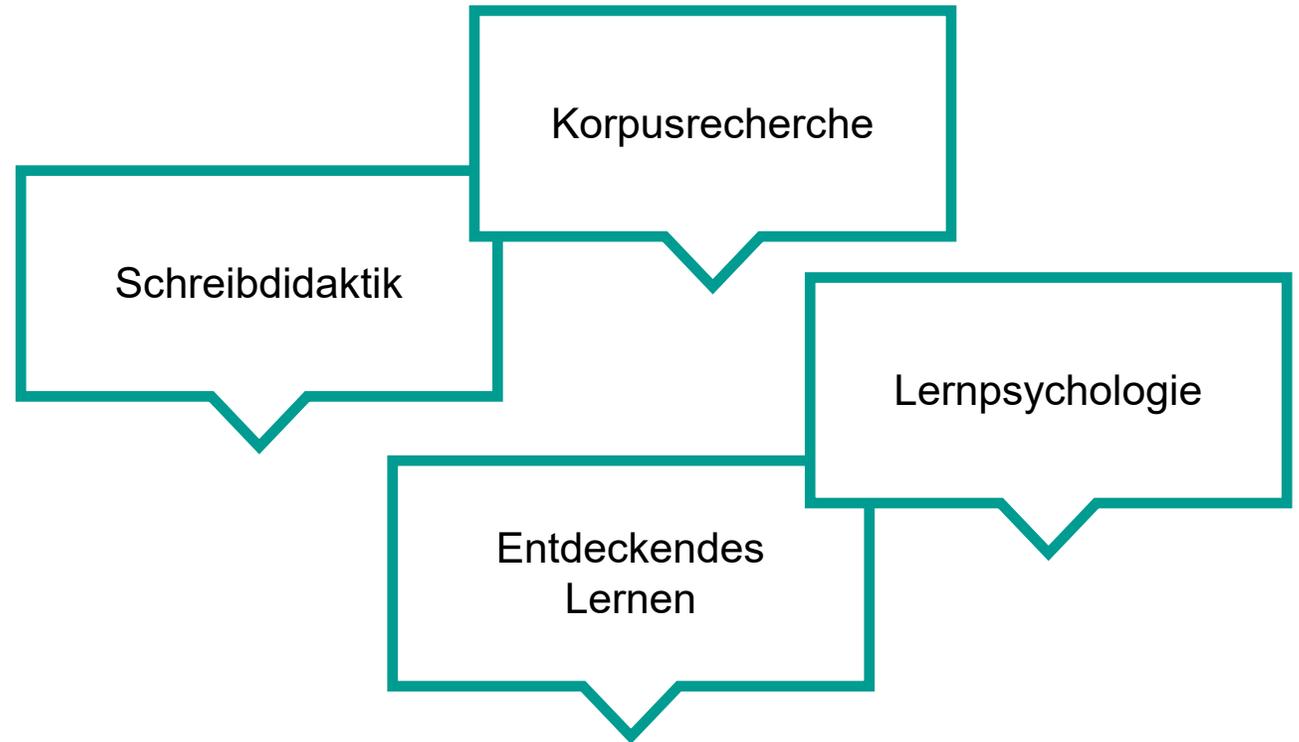
DDL und Sprachvermittlung

Kommentar zur DDL-Aufgabe „Suche in der Wildnis“

Aus Sicht der **Schreibdidaktik** ist die Aufgabe zu offen, es fehlen Kontext, Zielgruppe und Genre.

Die **Korpusrecherche** wurde nicht vorbereitet und führt nicht zur Entdeckung geeigneter sprachlicher Muster.

Aus Sicht der **Lernpsychologie** wurde kein verständlicher Input präsentiert. Es findet keine Entdeckung und kein Lernen statt.



DDL und Sprachvermittlung

DDL als Lernstrategie in einer Lektion

1. inhaltliches Thema



Mediennutzung

2. sprachliches Thema



**Ergebnisse einer Umfrage
präsentieren**

3. Unterrichtsprojekt



**Umfrage zur Mediennutzung
planen, durchführen, präsen-
tieren**

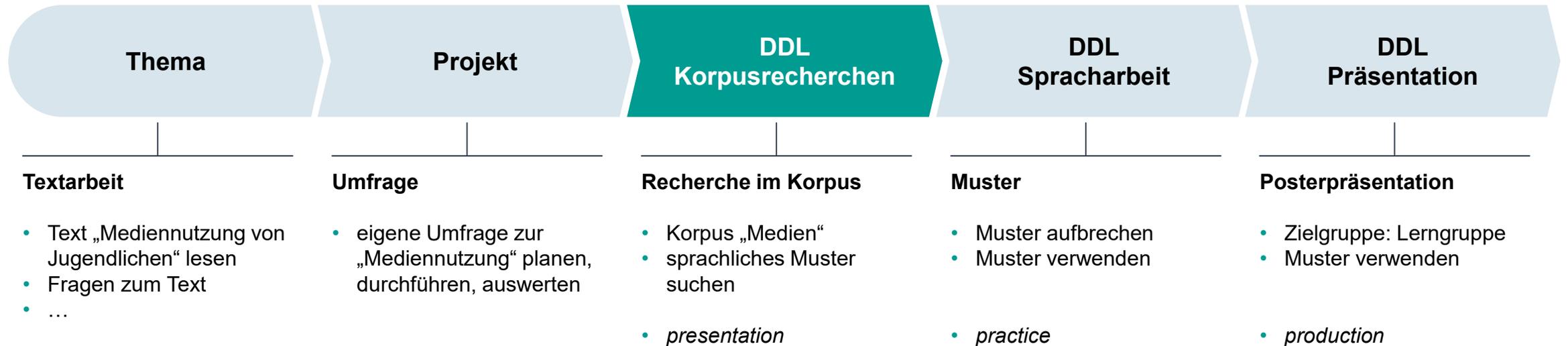
4. DDL als Lernstrategie



**sprachliche Muster zur Angabe
von Häufigkeiten mit Korpus-
recherchen entdecken**

DDL und Sprachvermittlung

DDL als Lernstrategie in einer Lektion



[Zahl]	Prozent
...	Prozent
...	Prozent
...	Prozent

DDL und Sprachvermittlung

DDL als Lernstrategie in einer Lektion

Target Corpus
Name: temp
Files: 1
Tokens: 3754
bamberg_korpus_KI.docx

KWIC Plot File Cluster N-Gram Collocate Word Keyword

Total Hits: 98 Page Size 100 hits 1 to 98 of 98 hits

	File	Left Context	Hit	Right Context
4	bam...	dies mehrmals pro Woche. Bei den Video-Streaming-Diensten nutzen 24	Prozent	der Jugendlichen diese täglich und 48 Prozent mehrmals pro Woche.
5	bam...	Inhalte nach Bedarf zu streamen. Das klassische Fernsehen wird von 45	Prozent	der Jugendlichen täglich und von weiteren 31 Prozent mehrmals pro
6	bam...	werden, jedoch nicht vollständig obsolet sind. Gedruckte Bücher werden von 12	Prozent	der Jugendlichen täglich und von 23 Prozent mehrmals pro Woche
7	bam...	Geräts im Alltag der jungen Generation unterstreicht. Genauer gesagt nutzen 93	Prozent	der Jugendlichen täglich ein Smartphone, und weitere 5 Prozent verwenden
8	bam...	ebenfalls ein integraler Bestandteil des jugendlichen Medienkonsums, wobei 88	Prozent	der Jugendlichen täglich und 8 Prozent mehrmals pro Woche online
9	bam...	en oder Streaming-Dienste, bleibt eine beliebte Freizeitbeschäftigung, wobei 66	Prozent	der Jugendlichen täglich Musik hören und 23 Prozent dies mehrmals
10	bam...	d Video-Streaming-Diensten zeigt sich eine interessante Dynamik: Während 49	Prozent	der Jugendlichen täglich Online-Videos anschauen, nutzen 33 Prozent diese
11	bam...	se Möglichkeit mehrmals pro Woche. Video-Streaming-Dienste werden von 24	Prozent	der Jugendlichen täglich und von 48 Prozent mehrmals pro Woche
12	bam...	und das Radio haben noch immer ihren Platz im Mediennutzungsverhalten. 45	Prozent	der Jugendlichen schauen täglich Fernsehen und weitere 31 Prozent mehrmals
13	bam...	gende Anteil der Jugendlichen nutzt täglich ein Smartphone; konkret gaben 93	Prozent	der Befragten an, ihr Smartphone täglich zu verwenden, während
14	bam...	oder Serien ist ebenfalls eine sehr beliebte Tätigkeit, die von 83	Prozent	der Befragten ausgeführt wird. Dies ist ein Hinweis darauf,
15	bam...	den Lernprozess unterstützt und erweitert. Online-Spiele spielen wird von 61	Prozent	der Befragten angegeben, was auf die große Beliebtheit von
16	bam...	An der Spitze der Beliebtheitskala stehen Nachrichtensendungen, die von 33,5	Prozent	der Befragten favorisiert werden. Dies unterstreicht die kontinuierliche Nachfrag
17	bam...	ndungen ziehen ebenfalls eine große Zuschauerschaft an und werden von 24,2	Prozent	der Befragten regelmäßig gesehen. Die Faszination für Live-Sportereignisse

Search Query Words Case Regex Results Set All hits Context Size 10 token(s)

Start Adv Search

Sort Options Sort center to right Sort 1 C Sort 2 1R Sort 3 2R Order by freq

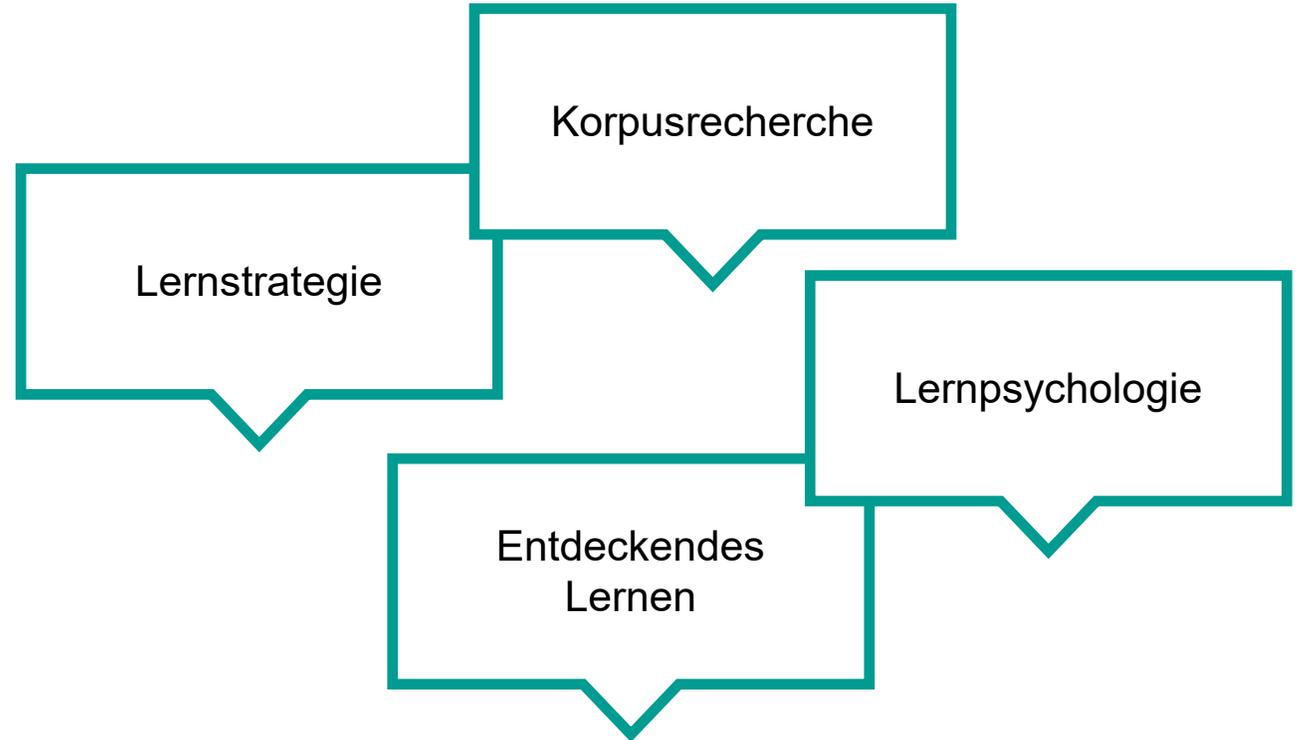
DDL und Sprachvermittlung

Kommentar: DDL als Lernstrategie in einer Lektion

DDL wurde als **Lernstrategie** vermittelt, um einen Sprachbedarf zu decken.

Die **Korpusrecherche** wurde vorbereitet und führt zur Entdeckung geeigneter sprachlicher Muster.

Es wurde ein verständlicher **Input** präsentiert. Es fand eine **Weiterverarbeitung** statt.

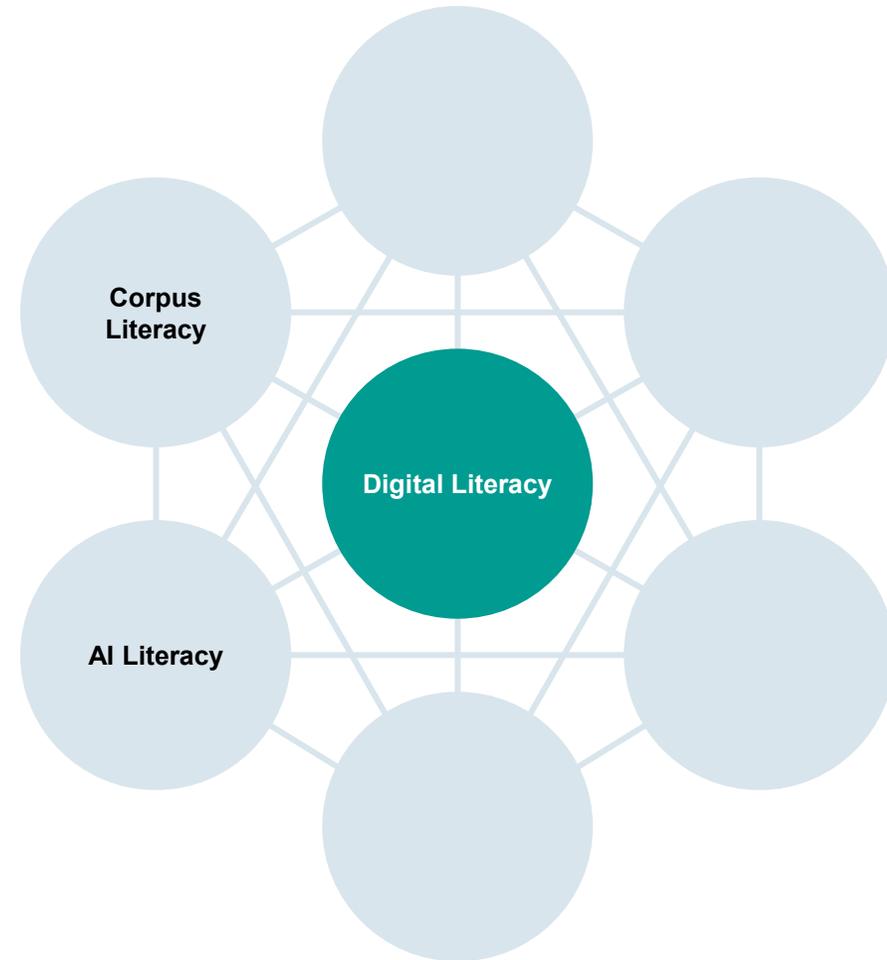


Die Vermittlung von Corpus Literacy ist ein Anliegen von DDL. Eine Herausforderung besteht darin, geeignete Untersuchungsgegenstände zu finden.

DDL und Corpus Literacy

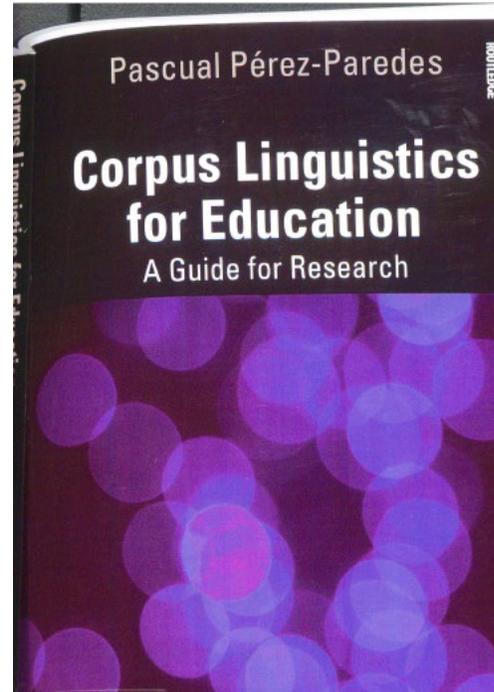
Corpus Literacy

- Korpora für den Spracherwerb nutzen können
- Vertrautheit mit Korpora, Korpusrecherchen, Konkordanzen
- Korpuslinguistik



DDL und Corpus Literacy

Pérez-Paredes, Pascual (2021): **Corpus linguistics for education. A guide for research.** London, New York: Routledge Taylor & Francis Group (Routledge corpus linguistics guides).



- 6 Examining lexis: analysing peace treaties and children's literature
 - 6.1 *Examining lexis* 117
 - 6.2 *Researching the lexicon: keywords* 117
 - 6.2.1 *Introducing keyword analysis* 120
 - 6.2.2 *Keyword analysis: a step-by-step guide* 121

DDL und Corpus Literacy

Lektion „Keywords“

- **Frage:** Welche lexikalischen Besonderheiten Friedensabkommen aufweisen?
- **Korpus:** 5 Friedensabkommen und Referenzkorpus
- **Suchmaschinen:** AntConc und SketchEngine

Korpuslinguistik

- Korpuslinguistische Bestimmung der „Keyness“
- Was ist ein Keyword?
- Wie ermittelt man ein Keyword?
- Beispiel: „education“ und „peace“ als Keywords

Table 6.3 Top 10 keywords using the British National Corpus as reference corpus

	Keywords	Score	Frequency in the target corpus	Frequency in the reference corpus
1	Tajikistan	2035.53	168	95
2	Tajik	1949.78	119	41
3	RUF	1252.98	57	2
4	Opposition	823.7	106	211
5	Leone	793.04	103	214
6	Autonomous	754.15	46	41
7	Reconciliation	718.8	53	73
8	MNLF	691.76	32	4
9	Signed	671.45	34	15
10	Nuri	661.97	38	32

DDL und Corpus Literacy

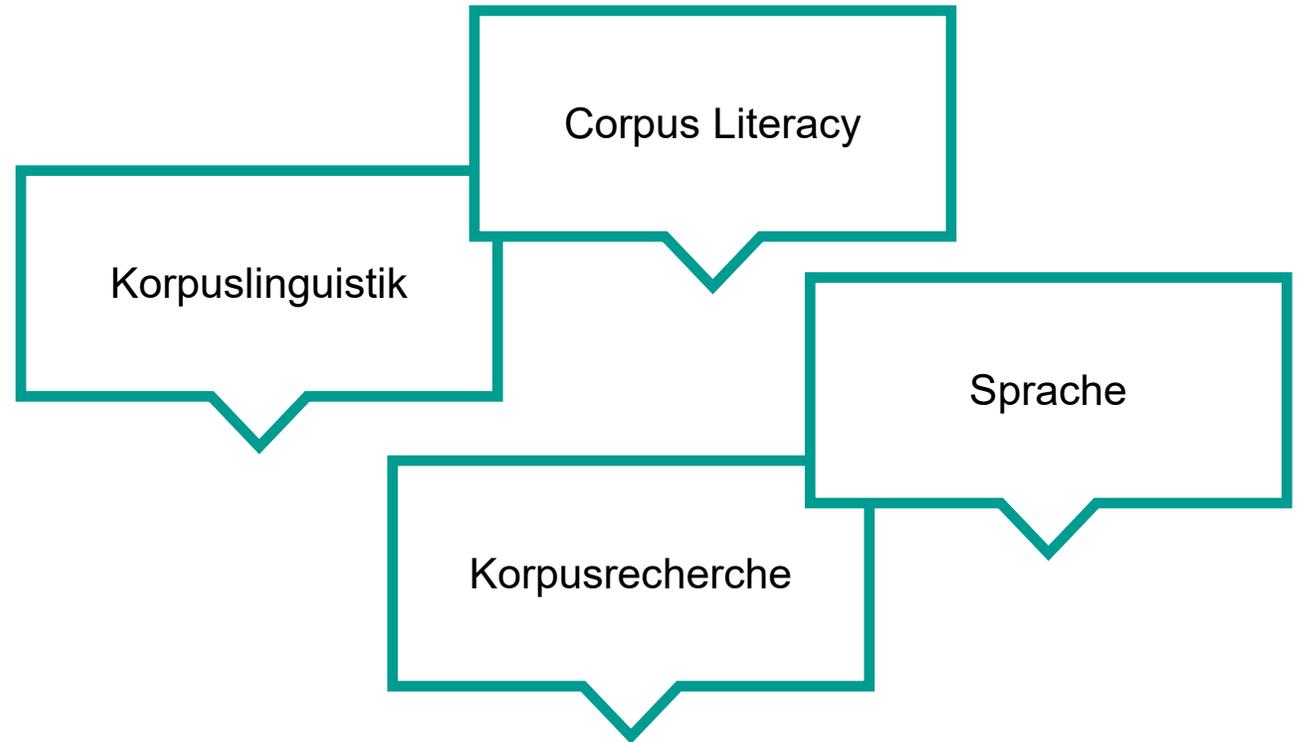
Kommentar zur Aufgabe von Pérez-Paredes

Corpus Literacy: methodische Kompetenzen

Korpuslinguistik: zentrale Konzepte

Sprache: Sprachvermittlung nicht im Zentrum

Korpusrecherche: Beispiel beliebig, Ergebnisse banal



**KI-Programme können
Korpusrecherchen nicht
ersetzen. Sie können DDL aber
ergänzen.**

DDL und KI-Programme



One of the main advantages of corpora is that we know exactly the **domain of texts** from which the corpus data is derived, something that we cannot track from current large language models underlying applications like ChatGPT.

Applied Corpus Linguistics 3 (2023) 100066

Contents lists available at [ScienceDirect](#)



Applied Corpus Linguistics

journal homepage: www.elsevier.com/locate/acorp



Generative AI and the end of corpus-assisted data-driven learning? Not so fast!

Peter Crosthwaite^{a,*}, Vit Baisa^b

^a School of Languages and Cultures, University of Queensland, Australia
^b Masaryk University, Faculty of Informatics, Natural Language Processing Centre, Czech Republic

ARTICLE INFO

Keywords:
Data-driven learning
generative AI
ChatGPT
DDL
Corpora

ABSTRACT

This article explores the potential advantages of corpora over generative artificial intelligence (GenAI) in understanding language patterns and usage, while also acknowledging the potential of GenAI to address some of the main shortcomings of corpus-based data-driven learning (DDL). One of the main advantages of corpora is that we know exactly the domain of texts from which the corpus data is derived, something that we cannot track from current large language models underlying applications like ChatGPT. We know the texts that make up large general corpora such as BNC2014 and BAWE, and can even extract full texts from these corpora if needed. Corpora also allow for more nuanced analysis of language patterns, including the statistics behind multi-word units and collocations, which can be difficult for GenAI to handle. However, it is important to note that GenAI has its own strengths in advancing our understanding of language-in-use that corpora, to date, have struggled with. We therefore argue that by combining corpus and GenAI approaches, language learners can gain a more comprehensive understanding of how language works in different contexts than is currently possible using only a single approach.

DDL und KI-Programme

Ich lerne Deutsch auf dem Niveau B2 und soll einen Text zum Thema Mediennutzung schreiben. In dem Text wird es häufig um "Prozent" gehen.

Gib mir 20 Beispielsätze aus dem Bereich Mediennutzung, in denen das Wort "Prozent" vorkommt.

Schreibe das Wort "Prozent" fett und schreibe zwei Wörter nach dem Wort „Prozent“ kursiv.



Natürlich! Hier sind 20 Sätze zum Thema Mediennutzung, in denen das Wort "Prozent" fett hervorgehoben ist und die zwei nachfolgenden Wörter in Kursivschreibweise:

1. Laut einer Studie nutzen 78 **Prozent** *der Jugendlichen* täglich soziale Medien.
2. Etwa 65 **Prozent** *der Erwachsenen* schauen täglich mehr als zwei Stunden Fernsehen.
3. In den letzten fünf Jahren ist die Nutzung von Streaming-Diensten um 40 **Prozent** *gestiegen*.
4. Rund 50 **Prozent** *der Menschen* lesen Nachrichten hauptsächlich auf digitalen Plattformen.
5. Eine Umfrage ergab, dass 30 **Prozent** *der Teilnehmer* mindestens einmal pro Woche einen Podcast hören.
6. Knapp 60 **Prozent** *der Eltern* kontrollieren die Mediennutzung ihrer Kinder.
7. Fast 70 **Prozent** *der Senioren* nutzen das Internet, um mit ihrer Familie in Kontakt zu bleiben.
11. Die Nutzung von Smartphones ist unter Teenagern um 80 **Prozent** *verbreitet*.
16. Die Anzahl der Menschen, die täglich E-Mails checken, liegt bei etwa 95 **Prozent**.

Ich hoffe, diese Beispiele helfen dir bei deinem Text!

DDL und KI-Programme

CK

Mit welchen Präpositionen wird das Wort "Prozent" verwendet, um Änderungen zu beschreiben?



KI

Das Wort "Prozent" wird oft in Verbindung mit bestimmten Präpositionen verwendet, um Änderungen in Statistiken oder Anteilen zu beschreiben. Hier sind einige der gebräuchlichsten Präpositionen, die in diesem Zusammenhang mit "Prozent" benutzt werden:

1. **um** – Verwendet, um die Größe einer Änderung zu beschreiben. Beispiel: "Die Nutzung des Internets stieg um 10 Prozent."
2. **auf** – Kann verwendet werden, um ein neues Niveau nach einer Änderung anzugeben. Beispiel: "Die Arbeitslosenquote fiel auf 5 Prozent."
3. **von... auf** – Dient dazu, einen Anfangs- und Endpunkt einer Änderung zu beschreiben. Beispiel: "Die Zinsen wurden von 5 Prozent auf 3 Prozent gesenkt."

Diese Präpositionen helfen dabei, die Art und Weise, wie sich Zahlen verändern, klar und präzise zu kommunizieren.



DDL und KI-Programme

CK

Almanca'da değişiklikleri tanımlamak için "yüzde" kelimesini tanımlamak için hangi edatlar kullanılır?



KI

Almanca'da "Prozent" kelimesi ile birlikte değişiklikleri tanımlamak için genellikle aşağıdaki edatlar kullanılır:

1. **um** – Bir değişiklik miktarını belirtmek için kullanılır. Örneğin: "Die Nutzung des Internets ist um 10 Prozent gestiegen." (İnternet kullanımı %10 arttı.)
2. **auf** – Bir değişiklikten sonra yeni bir seviyeyi belirtmek için kullanılır. Örneğin: "Die Arbeitslosenquote fiel auf 5 Prozent." (İşsizlik oranı %5'e düştü.)
3. **von... auf** – Bir değişikliğin başlangıç ve bitiş noktasını belirtmek için kullanılır. Örneğin: "Der Zinssatz wurde von 5 Prozent auf 3 Prozent gesenkt." (Faiz oranı %5'ten %3'e indirildi.)

Bu edatlar, oranlardaki veya yüzdelerdeki değişiklikleri açıklarken kullanılır ve değişikliğin boyutunu ya da yönünü belirtmede yardımcı olur.



DDL und KI-Programme



Although the AI-generated output might appear convincingly human, subjecting the texts to a rigorous linguistic analysis [...] reveals significant **disparities**. Profound **differences** emerge at the lexicogrammatical level of AI texts, underscoring their considerable distance from human-generated texts.

Applied Corpus Linguistics 4 (2024) 100083

Contents lists available at [ScienceDirect](#)



Applied Corpus Linguistics

journal homepage: www.elsevier.com/locate/acorp



Articles

AI-generated vs human-authored texts: A multidimensional comparison

Tony Berber Sardinha*

Graduate Program in Applied Linguistics and Language Studies, Pontifical Catholic University of Sao Paulo, Brazil

ARTICLE INFO

Keywords:
Artificial intelligence
Multidimensional analysis
Register

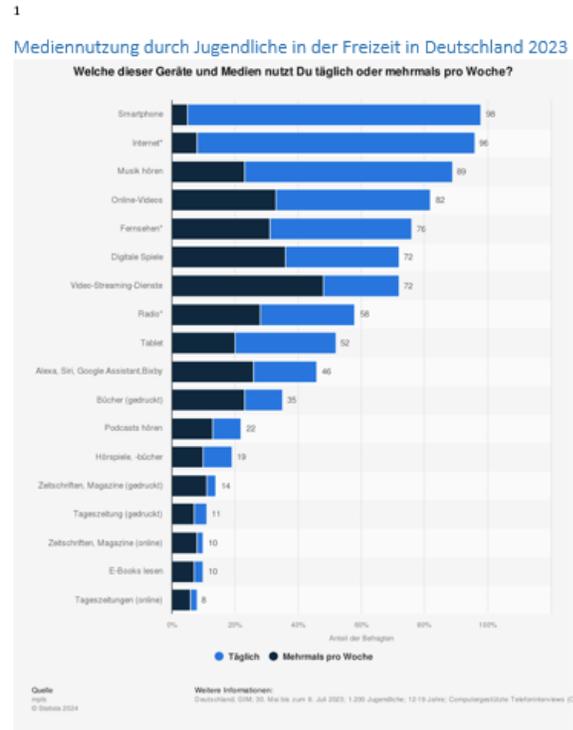
ABSTRACT

The goal of this study is to assess the degree of resemblance between texts generated by artificial intelligence (GPT) and (written and spoken) texts produced by human individuals in real-world settings. A comparative analysis was conducted along the five main dimensions of variation that Biber (1988) identified. The findings revealed significant disparities between AI-generated and human-authored texts, with the AI-generated texts generally failing to exhibit resemblance to their human counterparts. Furthermore, a linear discriminant analysis, performed to measure the predictive potential of dimension scores for identifying the authorship of texts, demonstrated that AI-generated texts could be identified with relative ease based on their multidimensional profile. Collectively, the results underscore the current limitations of AI text generation in emulating natural human communication. This finding counters popular fears that AI will replace humans in textual communication. Rather, our findings suggest that, at present, AI's ability to capture the intricate patterns of natural language remains limited.

DDL und KI-Programme

KI-Korpus

- 10 Texte zu Daten aus Statista
- Thema, Textsorte, Textlänge, Zielgruppe, Niveau können bestimmt werden
- narrow-angle approach
- Steuerung vs. Authentizität



Basierend auf einer aktuellen Umfrage, die 2023 unter Jugendlichen in Deutschland durchgeführt wurde, gibt es wichtige Einblicke in die Mediennutzung in ihrer Freizeit. Die Daten verdeutlichen, wie digitale Medien im Alltagsleben junger Menschen fest verankert sind, wobei Smartphones und das Internet dominieren.

Der überwiegende Anteil der Jugendlichen nutzt täglich ein Smartphone; konkret gaben 93 Prozent der Befragten an, ihr Smartphone täglich zu verwenden, während weitere 5 Prozent es mehrmals pro Woche nutzen. Ähnlich verhält es sich mit dem Internet, das von 88 Prozent der Jugendlichen täglich und von weiteren 8 Prozent mehrmals pro Woche genutzt wird. Diese hohe Nutzungszahl deutet darauf hin, dass Smartphones und das Internet zentrale Werkzeuge für Kommunikation, Information und Unterhaltung darstellen.

Musik hören ist ebenfalls eine beliebte Freizeitaktivität, wobei 66 Prozent der Jugendlichen täglich Musik hören und 23 Prozent dies mehrmals pro Woche tun. Dies zeigt, dass Musik ein integraler Bestandteil des täglichen Lebens vieler Jugendlicher ist.

Online-Videos und Video-Streaming-Dienste erfreuen sich ebenso großer Beliebtheit. 49 Prozent der Jugendlichen schauen täglich Online-Videos, und 33 Prozent tun dies mehrmals pro Woche. Bei den Video-Streaming-Diensten nutzen 24 Prozent der Jugendlichen diese täglich und 48 Prozent mehrmals pro Woche. Diese Zahlen bestätigen den Trend weg vom traditionellen Fernsehen hin zu flexibleren Online-Angeboten, die es den Nutzern ermöglichen, Inhalte nach Bedarf zu konsumieren.

Das klassische Fernsehen wird von 45 Prozent der Jugendlichen täglich und von weiteren 31 Prozent mehrmals pro Woche genutzt, was darauf hindeutet, dass trotz der Popularität digitaler Medien das Fernsehen noch immer einen festen Platz in der Mediennutzung einnimmt.

Interessanterweise zeigt die Umfrage auch, dass traditionellere Medien wie gedruckte Bücher, Zeitschriften und Tageszeitungen weniger häufig genutzt werden, jedoch nicht vollständig obsolet sind. Gedruckte Bücher werden von 12 Prozent der Jugendlichen täglich und von 23 Prozent mehrmals pro Woche gelesen.

Diese Ergebnisse bieten einen umfassenden Überblick über die Mediengewohnheiten deutscher Jugendlicher im Jahr 2023. Sie belegen, dass junge Menschen eine breite Palette von Medien für Bildung, Unterhaltung und soziale Interaktion nutzen, wobei digitale Plattformen führend sind.

Einführungsvortrag: Korpora, Korpuslinguistik und datengeleitetes Lernen (DDL)

Vielen Dank!

Prof. Dr. Christian Krekeler
Deutsch als Fremd- und Fachsprache
Hochschule Konstanz HTWG
krekeler@htwg-konstanz.de