



Sprachliche Komplexität im Längsschnitt und im Querschnitt

Welche Merkmale charakterisieren die Entwicklung von Deutsch-als-Fremdsprache-Lernenden und welche sind indikativ für erfolgreiche Prüfungsleistungen?

Detmar Meurers

Leibniz-Institut für Wissensmedien (IWM) & TüCeDE, Universität Tübingen

basierend auf gemeinsamer Forschung mit

Yushan Li (Zhejiang University)

Tagung “Große Lernerkorpora – Möglichkeiten und Grenzen”

Universität Leipzig, 25.09.2025



Einführung

- Sprachtests bieten Momentaufnahmen von spezifischen Sprachkompetenzfacetten anhand von psychometrisch optimierten Testitems.
- Wie können wir sprachliche Kompetenz und ihre Entwicklung **longitudinal** und direkt auf Basis von **authentischen Sprachproduktionen** beobachten?
- (i) Profilanalysen zu ausgewählten sprachlichen Mitteln, die für bestimmte **Erwerbsstufen** charakteristisch sind, wie spezifische Wortstellungen, Morphosyntax, ...
 - Erstpracherwerb (Crystal et al. 1976; Fletcher et al. 2012), Deutsch (Clahsen & Hansen 2012)
→ Analysen unterstützen effektive Therapie (LARSP: Language Assessment, Remediation and Screening Procedure)
 - Zweitspracherwerb: Deutsch (Clahsen 1985; Grieshaber 2019), Rapid Profile basierend auf Processability Theory (Pienemann 1998), DAKODA
- (ii) Stufenlose Charakterisierung anhand der CAF-Triade (Skehan 1989; Bulté et al. 2025):
 - **Sprachliche Komplexität** (Complexity), Korrektheit (Accuracy), Flüssigkeit (Fluency)



Was ist sprachliche Komplexität?

- Complexity is “a matter of the **number and variety** of an item’s constituent elements and of the **elaborateness** of their interrelational structure.”
(Rescher 1998)
- Language complexity is “the extent to which the language produced in performing a task is **elaborate** and **varied**.”
(Ellis 2003)
- Wie lässt sich der Grad der **Elaboriertheit** und die **Vielfältigkeit** der von Lernenden verwendeten Sprache konkretisieren, und wie werden die Analysen interpretiert?



Analyse sprachlicher Komplexität aus zwei unterschiedlichen Perspektiven

- Sprachliche Komplexität als Maß sowohl für **sprachliche Entwicklung**, als auch für die Bewertung der **Qualität von Texten** (Crossley 2020)
 - Vgl. PISA Test, wo 30 Punkte im Test etwa dem Lernstoff eines Schuljahres entsprechen sollen.
- Sind die gleichen Merkmale indikativ für Entwicklung und Qualität? (Crossley & McNamara 2014)
- ⇒ Wir untersuchen: Welche sprachlichen Komplexitätsmaße sind charakteristisch für
 - die sprachliche Entwicklung von Fremdsprachlernenden
 - die Qualität von Fremdsprachaufsätzen
 basierend auf großen Korpora chinesischer Deutschlernender:
 - CDLK (Li & Wu 2023) und PGG (Dissertation von Yushan Li 2025)
- Komplexitätsanalyse traditionell reduktionistisch: wenige Merkmale (Bulté & Housen 2012)
 - ⇒ Wir untersuchen Komplexität empirisch breit in unterschiedlichen linguistischen Domänen (Morphologie, Lexis, Syntax, Diskurs) und der Sprachverwendung (Weiss & Meurers 2019a,b).



Breite Evidenz für sprachliche Komplexität

- Die **Elaboriertheit** und die **Vielfältigkeit** von Sprache hat viele unterschiedliche Facetten:
 - I. Welche sprachliche **Formen des linguistischen Systems** kommen vor?
 - z.B. Anzahl komplexer Nominalphrasen pro Satz, Anzahl subordinierter Sätze
→ theoretische Linguistik, Zweitspracherwerbsforschung
 - II. Welche **Verwendung** sprachlicher Formen lässt sich beobachten?
 - z.B. Wortfrequenz, Age-of-Acquisition (AoA) Normen
→ gebrauchsbasierte Linguistik, Korpuslinguistik, Psychologie



Evidenz für sprachliche Komplexität (cont.)

III. Welche Art & Menge von **Bedeutung** ist kodiert, und wie sind sie zu einem kohärenten **Diskurs** organisiert?

z.B. Konkretheit, Ideendichte, Konnektoren, referentielle Kohäsion
→ Linguistik, Psychologie (Textverstehen, z.B. Kintsch)

- Neben der Komplexität von Sprache sind auch manche Aspekte der Schwierigkeit der kognitiven Verarbeitung von Sprache allgemein modellierbar:

z.B. Gedächtnisanforderung (DLT, Gibson 2000), erwartete Fortsetzung (Surprisal, Boston et al. 2008)
→ Psycholinguistik (Satzverarbeitung)

⇒ CTAP (Chen & Meurers 2016b) berechnet 845 Komplexitätsmerkmale für das Englische und 500 Merkmale für das Deutsch (Weiss & Meurers 2019a,b).

- CTAP ist frei verfügbar unter <http://ctapweb.com>

Common Text Analysis Platform.

Automatic text feature extraction and visualization tools.

[Use the CTAP Tools](#)



Manage and analyze texts with state-of-the-art NLP tools.

The Common Text Analysis Platform is a set of tools that helps you manage your text corpus and automatically analyze them for various purposes. Potential uses of the system include text complexity assessment, plagiarism detection, authorship attrition, and native language detection, etc.

The CTAP Tools include modules for corpus management, feature selection, analysis generation, and result visualization. Combination of the modules makes your text analysis highly flexible. For more detailed introduction of each module, please refer to the Documentation.



Lexikon

- **Lexikalische Vielfalt** (*sagen* vs. *erzählen, behaupten, ...*)
 - Type-Token Ratio = *Typ/Tok*
 - textlängenunabhängig: Measure of Textual Lexical Diversity (MTLD, McCarthy 2005)
- **Lexikalische Frequenz** (*Schiff* häufiger genutzt als *Barke*)
 - Wortfrequenz in SUBTLEX-DE (Brysbaert et al. 2011)
 - Beim Aggregieren auch Varianz und Clustermethoden berücksichtigen (Chen & Meurers 2016a)
- **Lexikalische Semantik**
 - Synonymie: Anzahl Wörter mit gleicher Bedeutung (GermaNet Synsets, Hamp & Feldweg 1997)
 - Polysemie/Homonymie: Anzahl unterschiedlicher Wortbedeutungen eines Worts



Morphologie

- **Derivation**

- Nominalisierung (*Zerstör-ung, Heiter-keit, ...*)
- Komposita (*Stadt-entwicklung-s-potential, ...*)

- **Flexion**

- Kasus (Genitiv, ...)
- Verbformen (Konjunktiv, ...)



Syntax

- systematische Analyse von Sätzen, T-Units, Clauses:
 - mittlere **Länge** (durchschnittliche Satzlänge, ...)
 - **Anzahl** der Vorkommen (Anzahl Clauses pro Satz, ...)
 - Vorkommen von **Untertypen** (subordinierte Sätze, ...)
- **Strukturelle Vielfalt**
 - z.B., Parse Edit Distance zwischen adjazenten Sätzen oder global im Text
- Anzahl und Länge ausgewählter **elaborierter Konstruktionen**
 - z.B., komplexe Nominalphrasen pro Satz



Diskurs

- Konnektoren (temporal, kausal, . . .) z.B. nach (Breindl et al. 2014; Eisenberg et al. 2009)
- Kohäsion durch Wiederaufnahme sprachlichen Materials (lokal, global)



Aktuelle Untersuchung

- Welche sprachlichen Komplexitätsmaße sind charakteristisch für
 - (a) die sprachliche Entwicklung von Deutsch-als-Fremdsprache-Lernenden
 - (b) die Bewertung der Qualität von Fremdsprachaufsätzen
- Datenbasis: große Korpora chinesischer Deutschlerner:
 - (a) Longitudinales Teilkorpus des CDLK (Li & Wu 2023)
 - (b) Querschnittskorpus PGG (Dissertation von Yushan Li 2025)
- Methode:
 - computerlinguistische Analyse von 450 Komplexitätsmerkmalen (Weiss & Meurers 2019a,b) mit CTAP
 - datengetriebene Auswahl informativer Merkmale für Modelle mit Explainable Boosting Machines (GAM)
 - ermöglicht quantitative und qualitative Analyse auf unterschiedlichen Granularitätsebenen

Querschnittskorpus: Benotete Aufsätze (PGG)

- Datenquelle: Nationale Prüfung für das Germanistik-Grundstudium in China (PGG)
 - landesweit einheitliche Prüfung für Deutschstudierende in China zur standardisierten Erfassung des Lernerfolgs am Ende des Grundstudiums (Ende 4. Semester)
 - Daten aus allen Hochschultypen und Regionen
 - Bewertung des schriftlichen Teils der high stakes PGG-Prüfung von sehr hoher Qualität
- repräsentatives Sample und externe Validität (Moranski & Ziegler 2021)
- 31.395 Texte mit 5,4 Millionen Tokens
 - Bewertung mit 4 Noten: 1-fail: 7.600, 2-pass: 9.769, 3-good: 10.072, 4-excellent: 3.954



Longitudinales Korpus (CDLK Teilkorpus)

- Longitudinales Teilkorpus des Chinesischen Deutschlerner-Korpus (CDLK, Li & Wu 2023).
- Teilkorpus von 163 Deutsch-Studierenden (3 Universitäten aus verschiedene Regionen) für die longitudinale Daten für vier Semester vorliegen
 - chinesische Muttersprachler mit Englisch als erster Fremdsprache
 - Deutscherwerb an der Universität mit einheitlichem Curriculum und Lehrwerk
- Vier Datenerhebungen zum Ende jedes Semesters des zweijährigen Grundstudiums
 - letzte Erhebung zum Zeitpunkt der PGG Prüfung
 - transkribierte handschriftliche Texte
- 634 Texte mit 87.766 Tokens
 - Texte aus 4 Semestern: 1 (142), 2 (172), 3 (169), 4 (151)



Erster Schritt: Identifikation der informativen Komplexitätsmerkmale

- Welche Merkmale sind informativ hinsichtlich
 - Notenbewertung im Querschnittskorpus: 1-fail, 2-pass, 3-good, 4-excellent
 - Semesterzahl im Längsschnittkorpus: 1., 2., 3., 4. Semester
- Herausforderungen:
 - sehr viele Komplexitätsmerkmale (450) und vielfach hohe Kollinearität
 - Verhältnis Komplexitätsmerkmal zu Note/Semester nicht unbedingt linear
- Explainable Boosting Machines (EBM, Lou et al. 2012; Nori et al. 2019)
 - Erweiterung von Generalized Additive Models (GAM, Hastie & Tibshirani 1987): maschinelles Lernen (bagging, gradient boosting) zur effektiven, iterativen Merkmalsselektion zur Modelloptimierung



Merkmalsauswahl: sprachliche Komplexität aus zwei Perspektiven

- EBM wählt aus den 450 Komplexitätsmerkmalen
 - 23 für Entwicklung (CDLK): Lexik 15 (55%), Syntax 5 (22%), Diskurs 2 (9%), Morph. 1 (4%)
 - 39 für Benotung (PGG): Lexik 20 (51%), Syntax 13 (33%), Morph. 3 (8%), Diskurs (3%) 1, Länge 2 (5%)
- alle Dimensionen der linguistischen Modellierung sind relevant!
 - Die Benotungsperspektive fokussiert mehr auf Syntax und Morphologie.
- Entwicklungs- und Benotungsperspektive überlappen allerdings nur in 5 Merkmalen:
 - Syntaktische Vielfalt (Mean Local Edit Distance for POS)
 - Syntaktische Elaboriertheit (Dependent Clause Ratio)
 - Lexikalische Vielfalt (HDD excluding punctuation and numbers)
 - Lexikalische Frequenz (SD of Verb Word Frequency per Million SUBTLEX Token)
 - und einen potenziellen Aufgabeneffekt (Singular Proper Noun Density)

aber verwandte Merkmale spielen bei beiden eine Rolle (z.B. Textlängenaspekte)
- ⇒ Automatisierte Komplexitätsanalyse großer Lernerkorpora ermöglicht eine differenzierte, unterschiedlich feingranulare Sicht auf Sprachentwicklung und Textqualität.



Globale Charakterisierung: Klassifikation nach Semester vs. nach Note

- Longitudinal: Resultate für Klassifikation in 4 Semester für 634 Texte
 - Aufteilung: 80% Training und 20% Testing
 - Majority baseline: 27%
 - Accuracy: 65% mit 23 Komplexitätsmerkmalen
- Querschnitt: Resultate für Bewertung mit 4 Noten für 31.395 Texte
 - Aufteilung: 80% und Training 20% Testing
 - Majority baseline: 32%
 - Accuracy: 55% mit 39 Komplexitätsmerkmalen
- Schwächere Klassifikation für Bewertung, obwohl die Datenbasis weitaus größer ist.
- Benotung bildet neben sprachlicher Entwicklung anscheinend andere Aspekte ab:
 - Accuracy
 - Angemessenheit für Aufgabe
 - Qualität des Inhalts gegeben die Aufgabe (z.B. der Argumentation)



Analyse der Klassifikation nach Semestern

Semester	Precision	Recall	F1-Score	Support
1	93%	93%	93%	28
2	62%	74%	68%	35
3	55%	47%	51%	34
4	50%	47%	48%	30

- **Sehr gute Identifikation von Anfängern**
- **Schwächere Differenzierung von Fortgeschrittenen**
- **Insbesondere auch geringerer Recall**

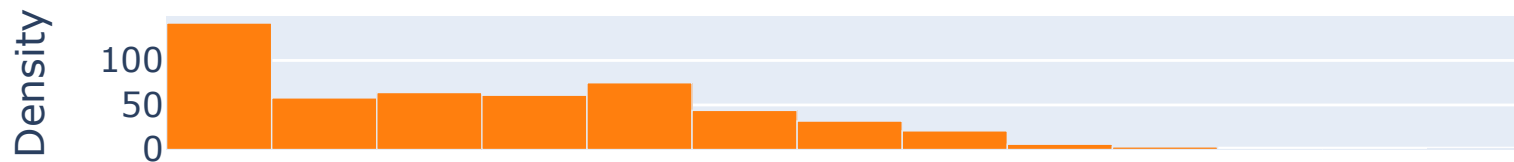
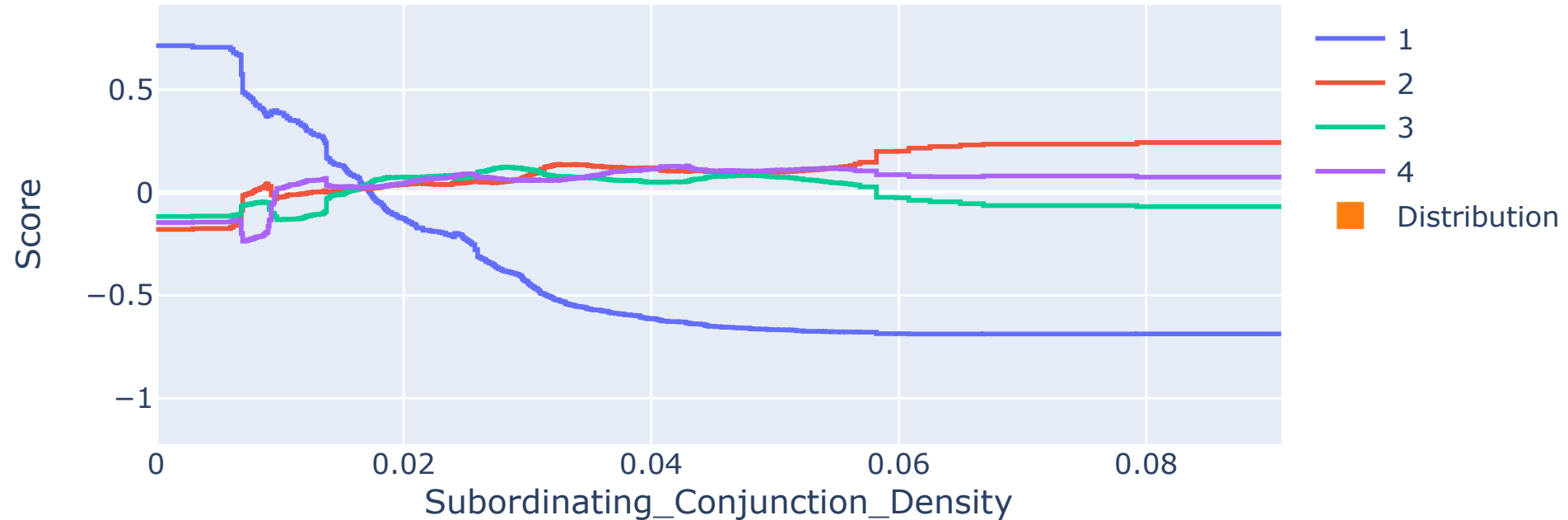


Analyse der Klassifikation nach Benotung

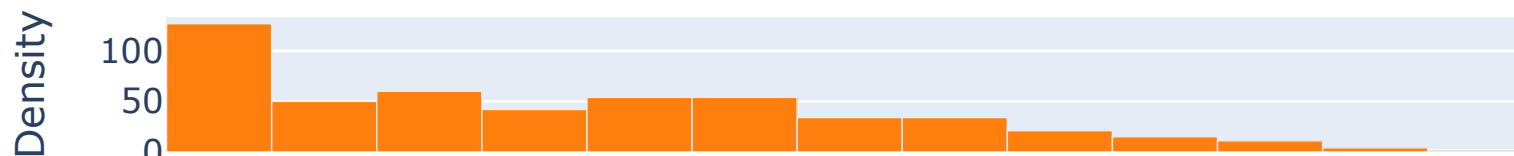
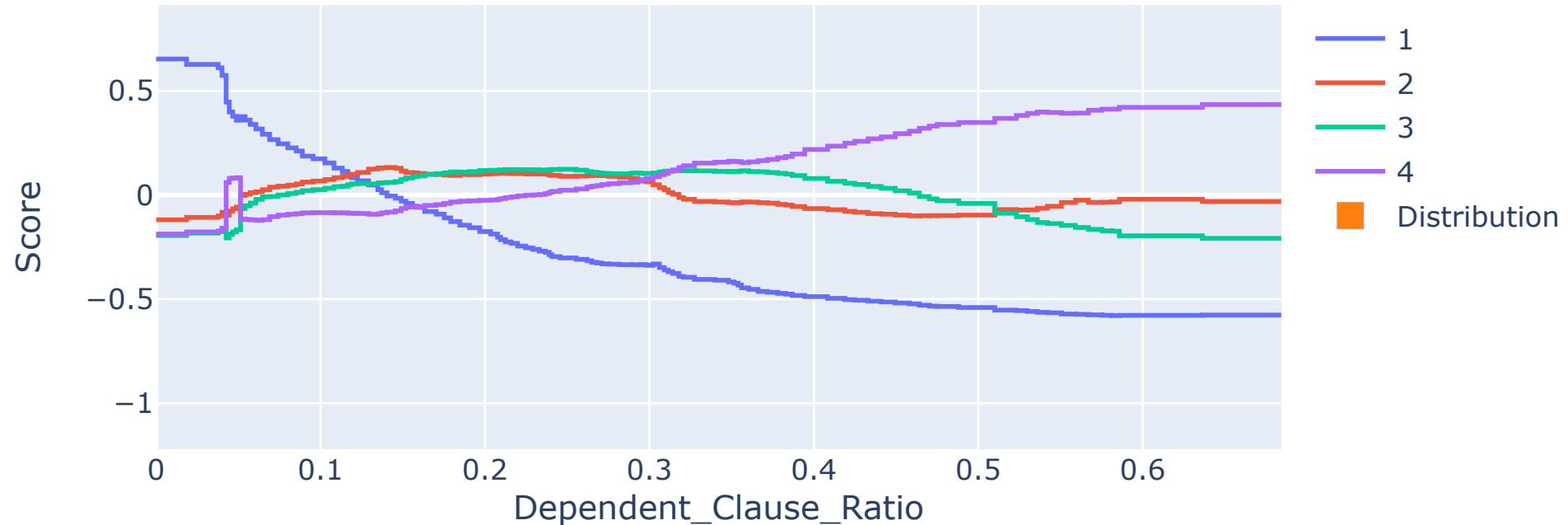
	Precision	Recall	F1-Score	Support
1-Fail	73%	63%	68%	1520
2-Pass	49%	49%	49%	1954
3-Good	50%	65%	56%	2014
4-Excellent	55%	26%	36%	791

- **Schwacher Recall bei exzellenten Aufsätzen**
 - unterstützt Hypothese, dass Benotung andere Aspekte als sprachliche Entwicklung berücksichtigt

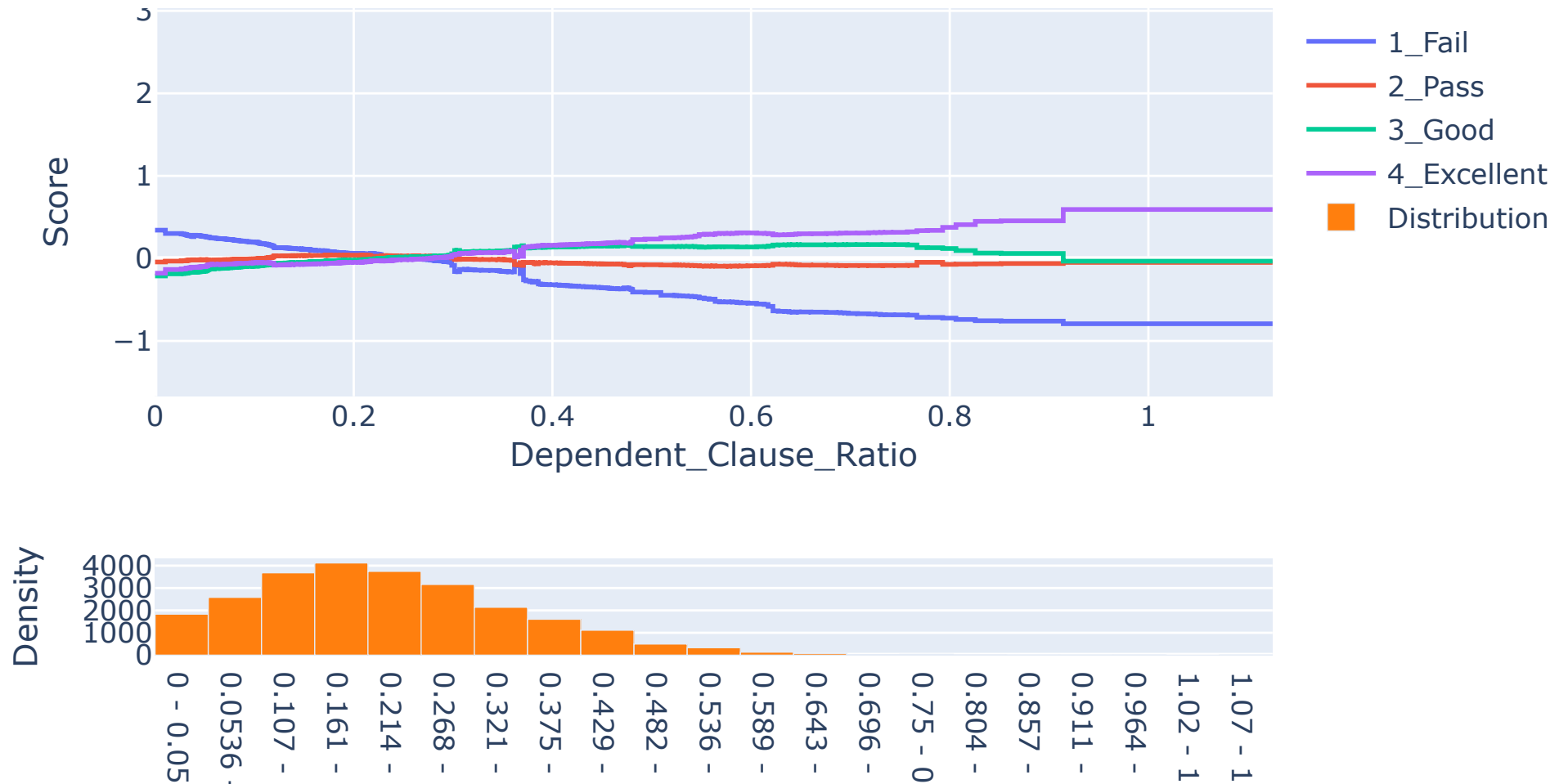
Feingranularer Blick auf Longitudinale Entwicklung: Syntaktische Elaboriertheit Subordinating Conjunction Density



Feingranularer Blick auf Longitudinale Entwicklung: Syntaktische Elaboriertheit Dependent Clause Ratio

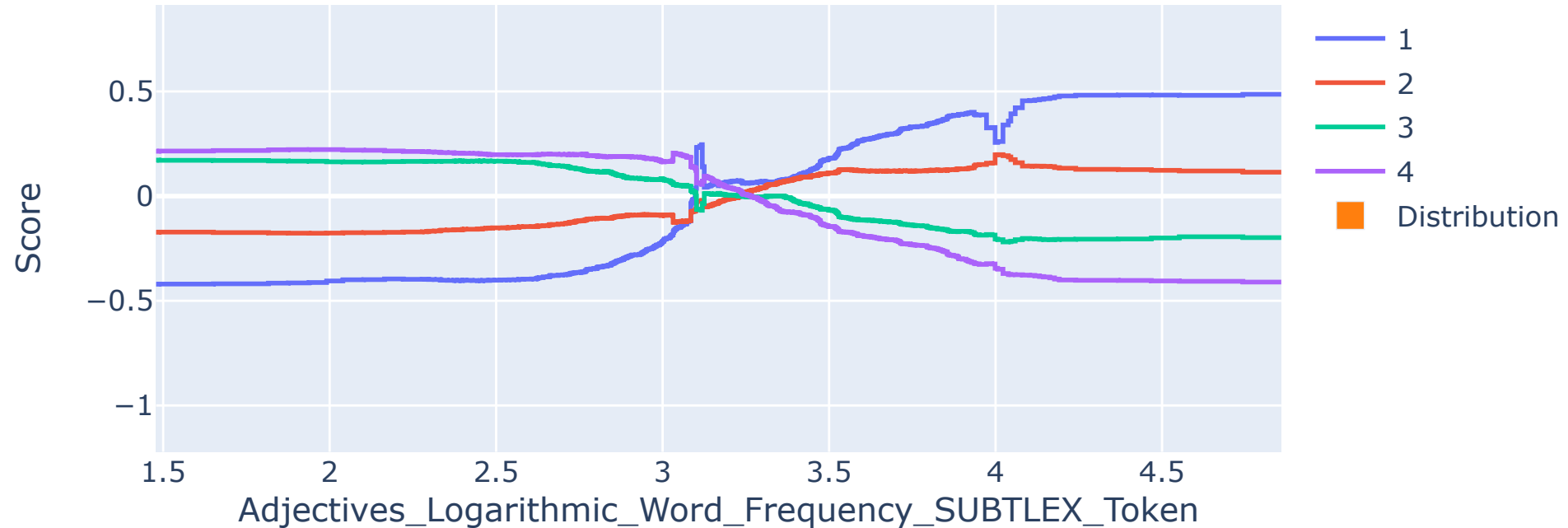


Feingranularer Blick auf Benotung: Syntaktische Elaboriertheit Dependent Clause Ratio



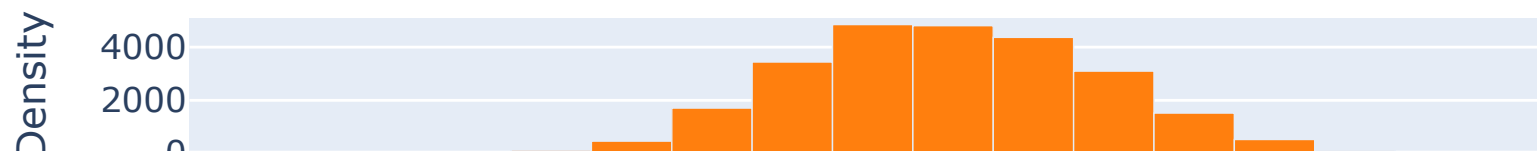
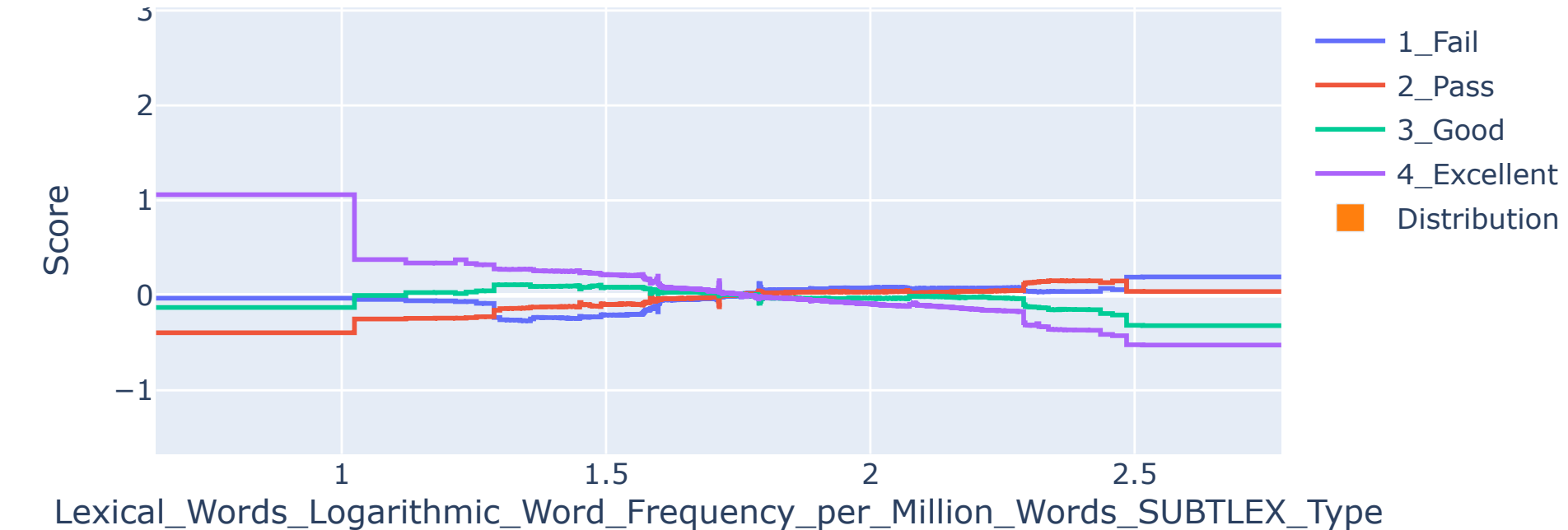
Feingranularer Blick auf Longitudinale Entwicklung: Lexikalische Frequenz

Adjectives Log. Word Frequency SUBTLEX Token

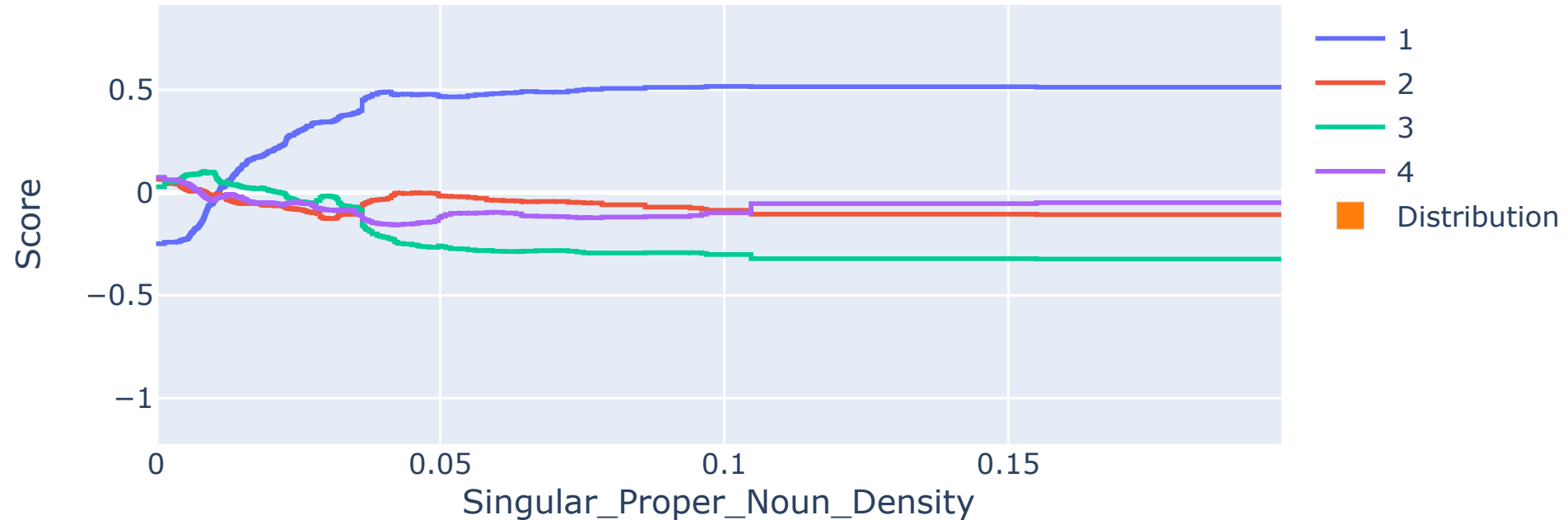


Feingranularer Blick auf Benotung: Lexikalische Frequenz

Lexical Words Log. Word Frequency per Million Words SUBTLEX Type

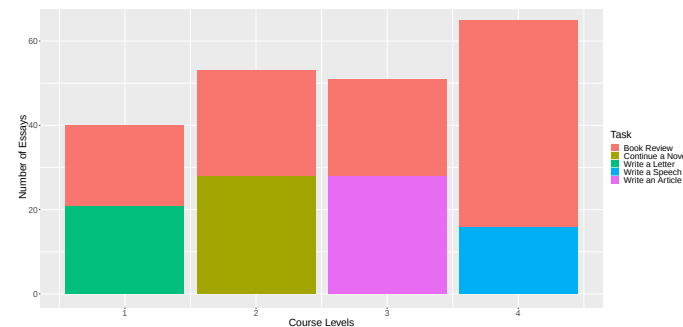


Feingranularer Blick auf Longitudinale Entwicklung: Idiosynkratische Lexikalische Eigenschaft Singular Proper Noun Density

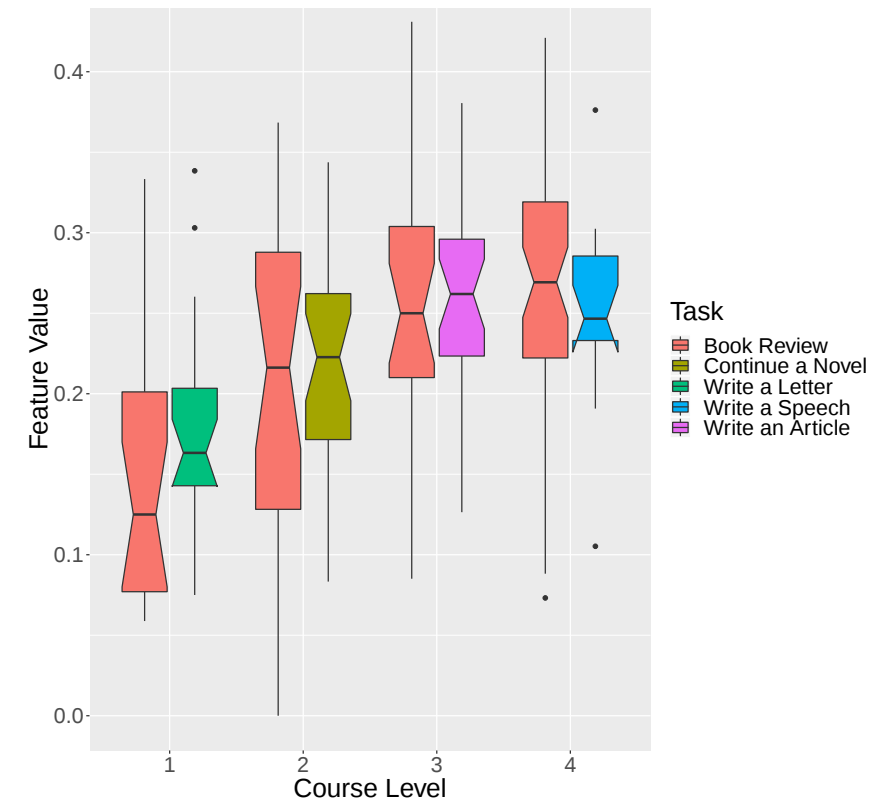
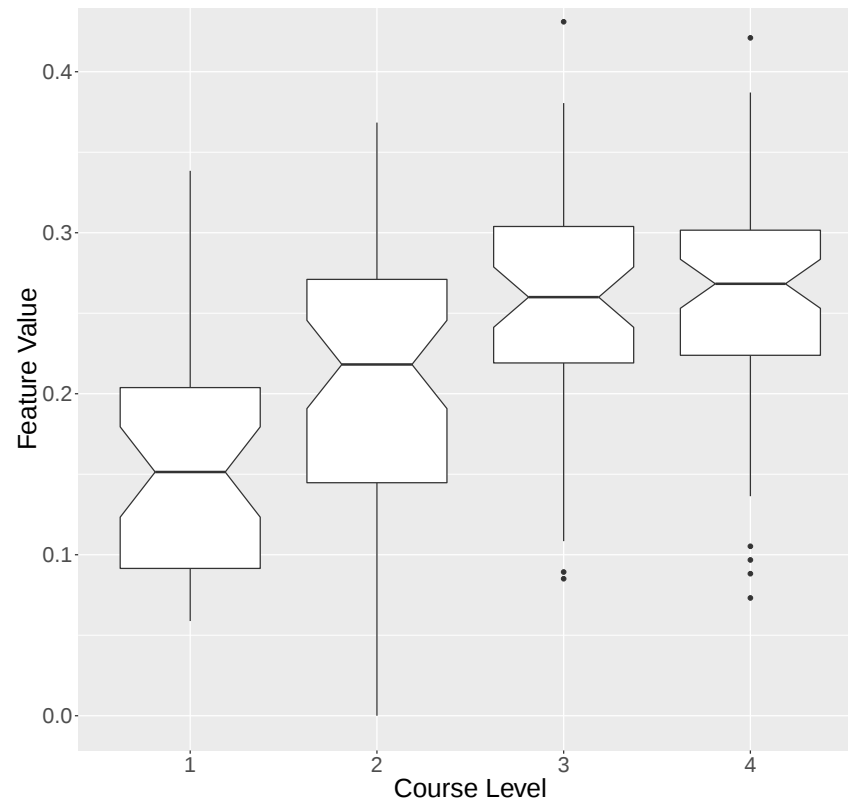


Einfluss der Aufgabe auf Komplexitätsanalyse

- Komplexität kann Indikator sein für sprachliche Kompetenzen
 - aber: Nutzung von Kompetenzen ist aufgabenabhängig
- ⇒ Analyse von Weiss (2017) basierend auf Falko Georgetown Korpus (Siemen et al. 2006)
- 209 Texte von 123 Deutsch L2 Lernern
- Take-home Klausuren aus vier Deutschkursen der Georgetown University
 - Kurslevel: mittleres (1) bis fortgeschrittenes (4) Niveau
 - Referenzaufgabe (book review) + kursabhängige Aufgabe (novel continuation, letter, speech, article)

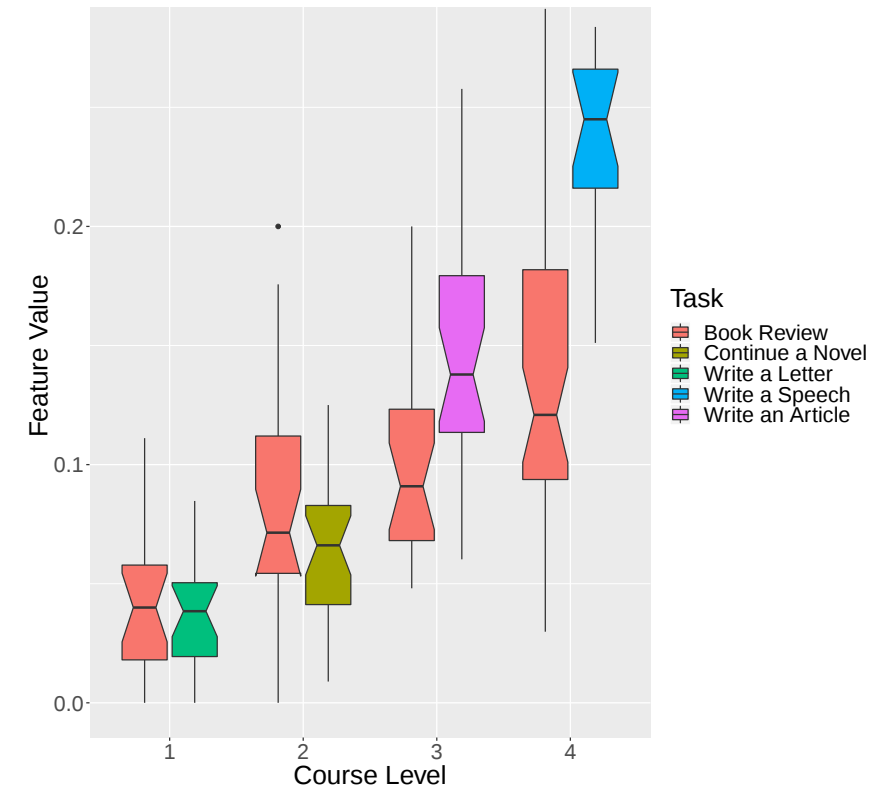
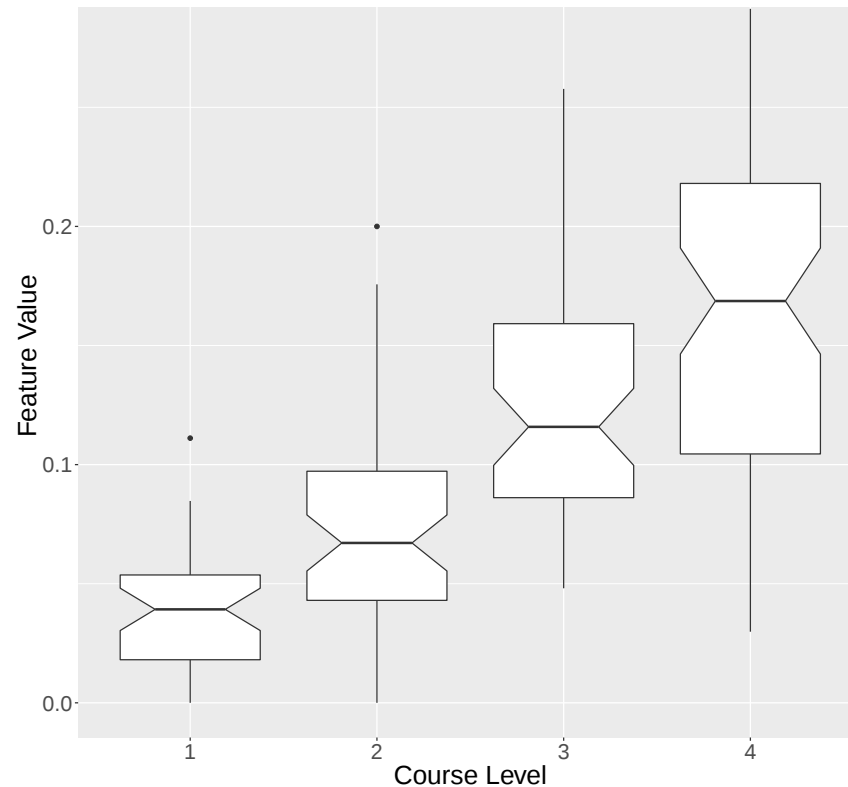


Klausale Komplexität (dependent clause per clause)



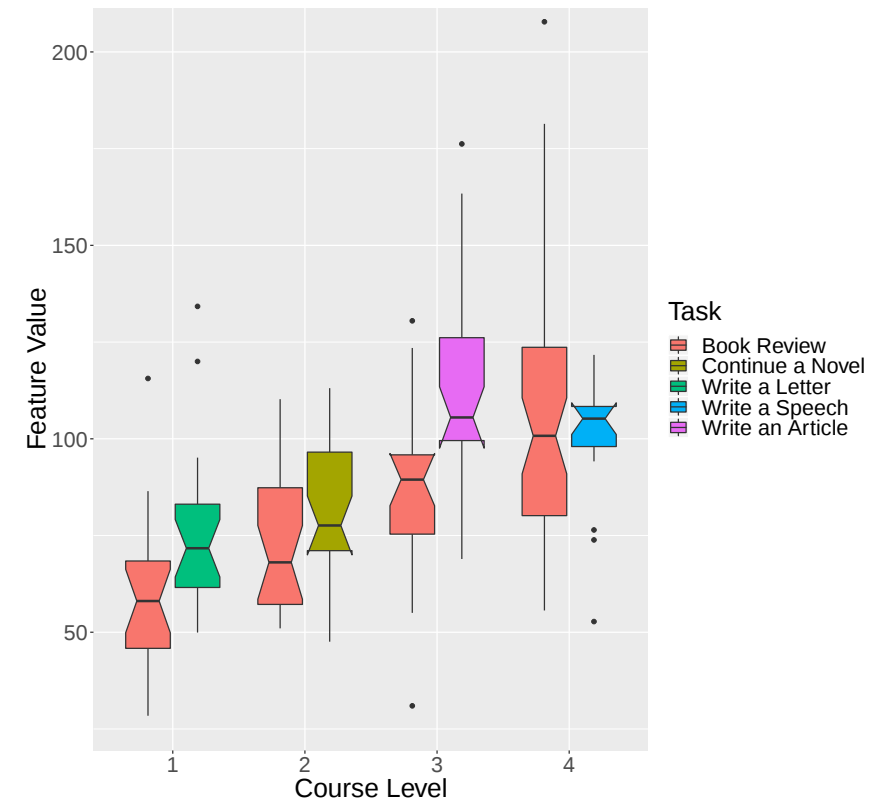
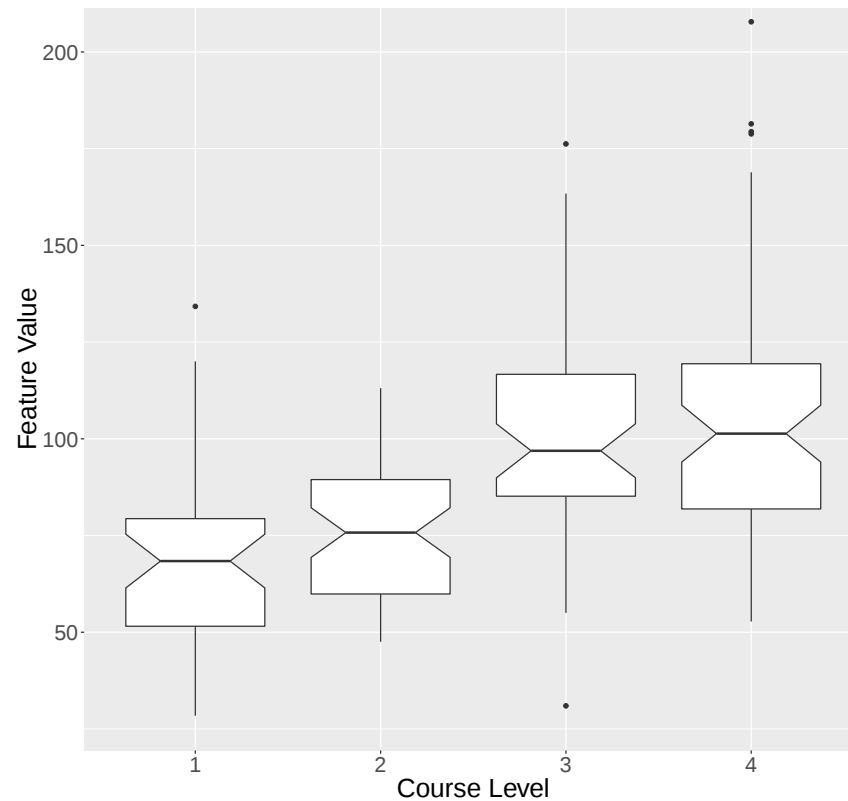
- systematischer Anstieg eingebetteter verbaler Projektionen
- wenig Aufgabenunterschiede

Morphologische Komplexität (derived nouns per noun)



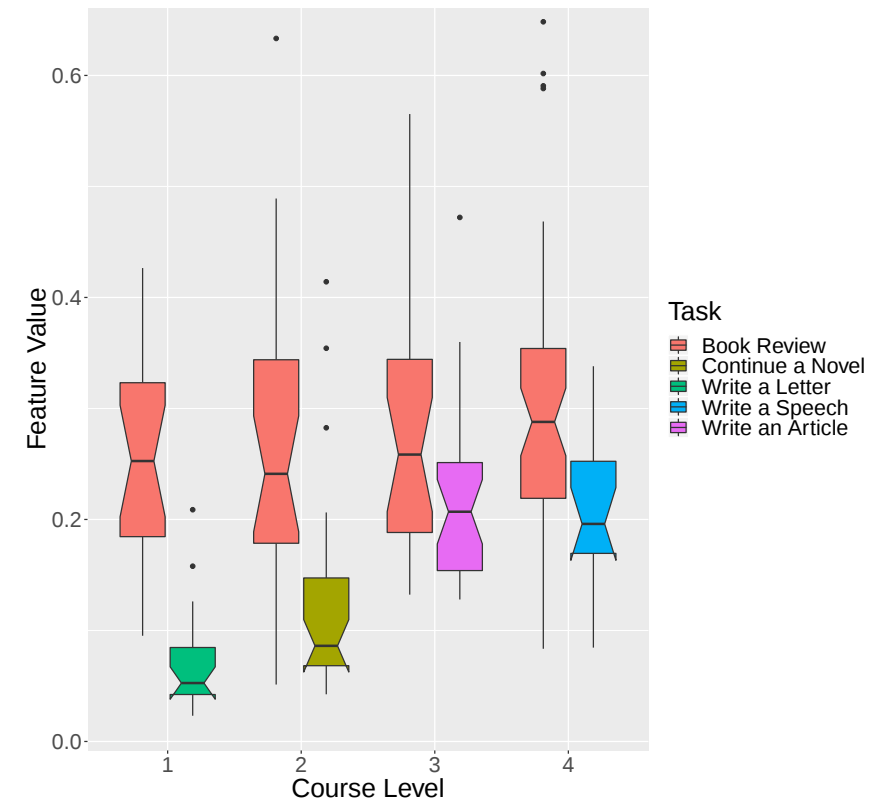
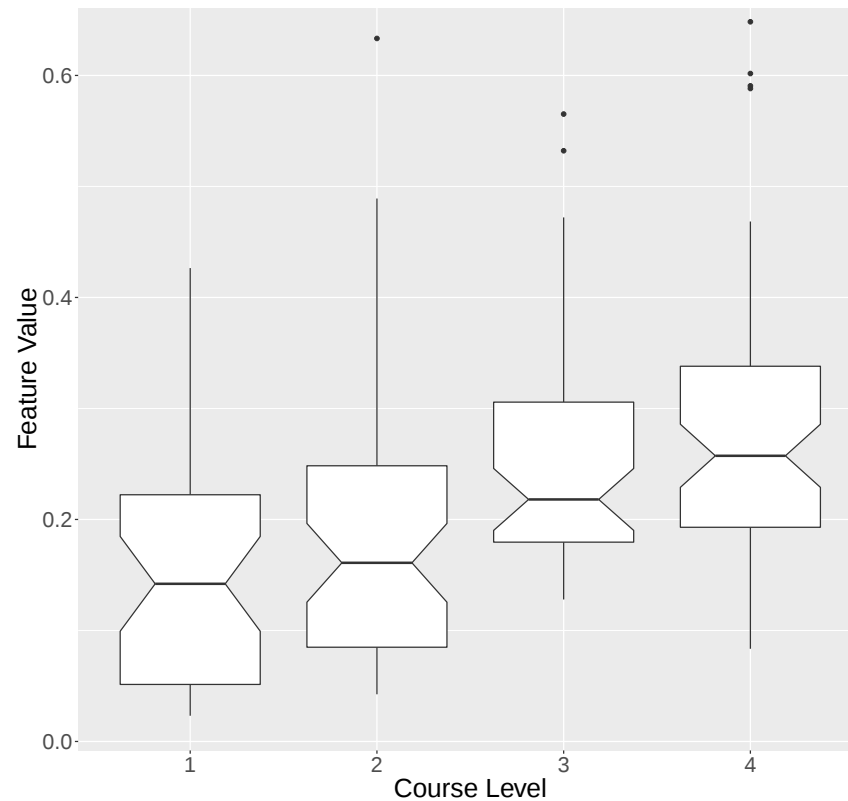
- systematischer Anstieg derivierter Nomen
- in höheren Kursen stärkere Aufgabenunterschiede

Lexikalische Komplexität (MTLD)



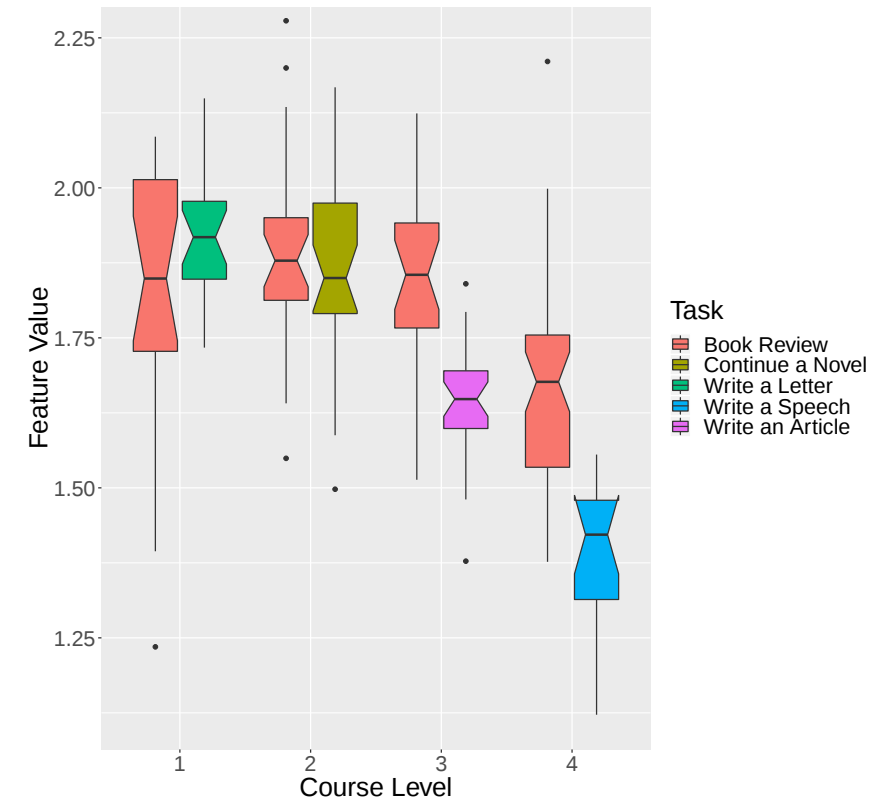
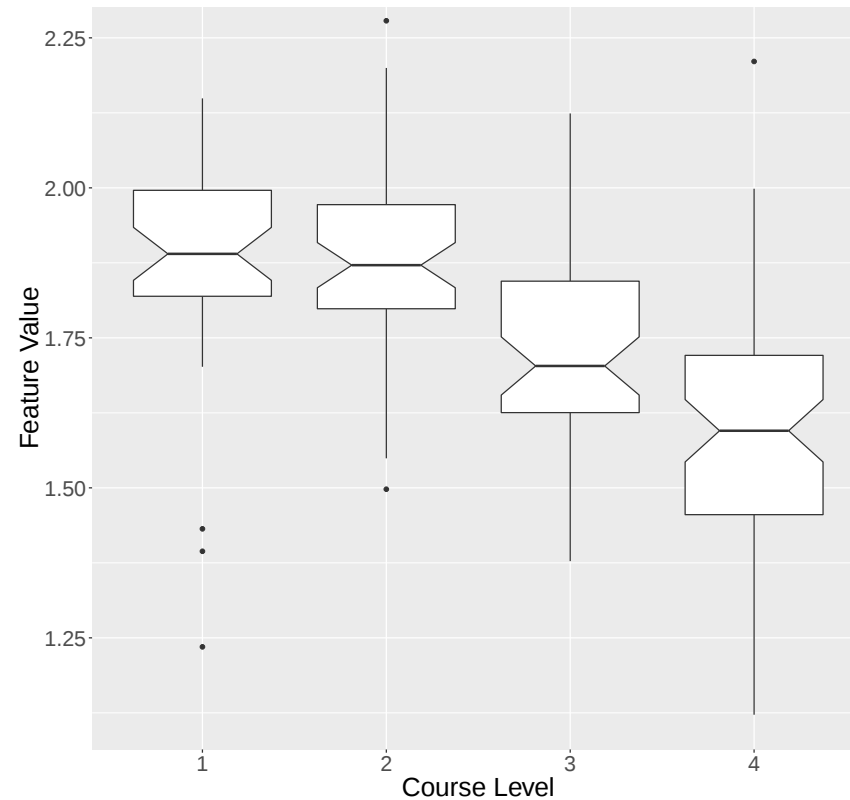
- Lexikalische Vielfältigkeit nimmt bis zum dritten Zeitpunkt zu
- leichte Aufgabenabhängigkeit, stärker bei Fortgeschrittenen

Diskurs-Komplexität (global stem overlap per sentence)



- Referenzaufgaben ohne Entwicklung
- spezifische Aufgaben zeigen systematischer Anstieg

Sprachverwendung (log type frequency per type in SUBTLEX-DE)



- zunahme seltenerer Wörter vor allem bei Fortgeschrittenen
- Aufgabenabhängigkeit bei Fortgeschrittenen



Diskussion zur Aufgabenabhängigkeit von sprachlicher Komplexität

- Aufgabenabhängigkeit der genutzten sprachlichen Komplexität vor allem bei fortgeschrittenen Lernenden beobachtet.
 - Fortgeschrittene verfügen über mehr unterschiedliches Sprachmaterial.
 - Fortgeschrittene Kompetenz liegt in der Verwendung von *appropriate complexity* (Pallotti 2023) entsprechend einer *functional adequacy* (Kuiken & Vedder 2017)
- Komplexitätsanalyse sollte zwischen individuell verfügbarem, genutzten Sprachmaterial und funktionaler Angemessenheit unterscheiden.
- Angelehnt an variationslinguistische Perspektiven könnten wir unterscheiden:
 - Welche Funktionen wird realisiert? (Variable)
 - Welche der sprachlichen Mittel, die diese Funktion realisieren können, werden genutzt? (Varianten)



Überlegungen zu einer variationslinguistischen Perspektive

- Komplexität (Variabilität, Elaboriertheit) lässt sich auf beiden Ebenen untersuchen:
 - Fortgeschrittene können vielfältigere und elaboriertere Funktionen sprachlich angemessen adressieren
 - und vielfältigere und elaboriertere Varianten für die Realisierung einer Funktion nutzen
- Beispiel: Wird für eine Bildbeschreibung funktional angemessen Modifikation genutzt?
 - Werden vielfältig unterschiedliche Varianten der Modifikation genutzt?
(prä-nominale/post-nominale/adverbiale, vielfältige lexikalische oder phrasale)
 - Wie elaboriert sind die dabei genutzten Optionen? (z.B. niederfrequente Adjektive, pränominale Partizipialkonstruktionen)



Von der Analyse des Sprachstandes zu möglichen Interventionen

- Profilanalysen wie die Language Assessment, Remediation and Screening Procedure (LARSP, Crystal et al. 1976) dienen systematisch als Grundlage einer Sprachförderung
 - Anders als Profilanalysen sind Komplexitätsanalysen nicht auf spezifische Erwerbsstufen bezogen, sondern liefern kontinuierliche Werte.
 - Komplexitätsanalysen lassen sich direkt für sprachförderliche Interventionen nutzen:
 - Komplexitätsanalyse sowohl von Lerner-Output als auch von Input möglich
- ⇒ Alignment von Lernenden mit adaptivem, entwicklungsproximalem Input in Complex Primed Input Continuation Writing Tasks (Chen & Meurers 2017, 2019; Chen et al. 2022)



Zusammenfassung

- Linguistisch breite Modellierung von Komplexität als Elaboriertheit und Vielfältigkeit der sprachlichen Mittel ist in großen Korpora umsetzbar
 - unterstützt Charakterisierung von Lernertexten, die unterschiedlich feingranular interpretierbar ist
- Sprachliche Komplexität ist hinsichtlich der longitudinalen Entwicklung und der Qualität von Schreibprodukten interpretierbar
 - aber manifestiert sich teilweise in unterschiedlichen Aspekten sprachlicher Komplexität
 - Bewertung von Schreibprodukten auch von andere Faktoren beeinflusst
- Interpretation von sprachlichen Komplexitätsmaßen ist abhängig von der Aufgabe.
 - Merkmale sind unterschiedlich anfällig für Aufgabeneffekte
 - Aufgabenabhängigkeit vor allem für fortgeschrittene Lernende beobachtbar
→ Funktion und Realisierung variationistisch berücksichtigen



Outlook: Verbindung von Profilanalysen und Komplexitätsanalysen?

- Profilanalysen mit mehr sprachlichen Aspekten erweitern, die sich empirisch als indikativ für bestimmte Stadien herausstellen (z.B. wie beim English Grammar Profile)?
- Komplexitätsanalysen mehr theoretisch motivieren statt rein datengetriebenen, dabei Aufgabe und ihre funktionalen sprachlichen Anforderungen explizit machen.



Zitierte Arbeiten

- Boston, M. F., J. T. Hale, U. Patil, R. Kliegl & S. Vasishth (2008). Parsing costs as predictors of reading difficulty: An evaluation using the Potsdam Sentence Corpus. *Journal of Eye Movement Research* 2(1), 1–12. URL <http://www.jemr.org/online/2/1/1>.
- Breindl, E., A. Volodina & U. H. Waßner (2014). *Handbuch der deutschen Konnektoren 2: Semantik der deutschen Satzverknüpfers*, vol. 13. Walter de Gruyter GmbH & Co KG.
- Brysbaert, M., M. Buchmeier, M. Conrad, A. M. Jacobs, J. Bölte & A. Böhl (2011). The word frequency effect: A review of recent developments and implications for the choice of frequency estimates in German. *Experimental Psychology* 58, 412–424.
- Bulté, B. & A. Housen (2012). Defining and operationalising L2 complexity. In A. Housen, F. Kuiken & I. Vedder (eds.), *Dimensions of L2 Performance and Proficiency*, John Benjamins, pp. 21–46. URL <https://doi.org/10.1075/llt.32.02bul>.
- Bulté, B., A. Housen & G. Pallotti (2025). Complexity and difficulty in second language acquisition: A theoretical and methodological overview. *Language Learning* 75(2), 533–574.
- Chen, X. & D. Meurers (2016a). Characterizing Text Difficulty with Word Frequencies. In *Proceedings of the 11th Workshop on Innovative Use of NLP for Building Educational Applications*. San Diego, CA: Association for Computational Linguistics, pp. 84–94.
- Chen, X. & D. Meurers (2016b). CTAP: A Web-Based Tool Supporting Automatic Complexity Analysis. In *Proceedings of the Workshop on Computational Linguistics for Linguistic Complexity*. Osaka, Japan: COLING, pp. 113–119.
- Chen, X. & D. Meurers (2017). Challenging Learners in Their Individual Zone of Proximal Development Using Pedagogic Developmental Benchmarks of Syntactic Complexity. In *Proceedings of the Joint 6th Workshop on NLP for Computer Assisted Language Learning and 2nd Workshop on NLP for Research on Language Acquisition at NoDaLiDa 2017*. Gothenburg, Sweden: ACL, Linköping Electronic Conference Proceedings 134, pp. 8–17. URL <http://aclweb.org/anthology/W17-0302.pdf>.
- Chen, X. & D. Meurers (2019). Linking text readability and learner proficiency using linguistic complexity feature vector distance.



Computer-Assisted Language Learning 32(4), 418–447. URL <https://doi.org/10.1080/09588221.2018.1527358>.
<https://doi.org/10.1080/09588221.2018.1527358>.

Chen, X., D. Meurers & P. Rebuschat (2022). ICALL offering individually adaptive input: Effects of complex input on L2 development. *Language Learning & Technology* 26(1), 1–21. URL <https://hdl.handle.net/10125/73496>.

Clahsen, H. (1985). Profiling second language development: A procedure for assessing L2 proficiency. In K. Hyltenstam & M. Pienemann (eds.), *Modelling and assessing second language acquisition*, Bristol, UK: Multilingual Matters, pp. 283–332.

Clahsen, H. & D. Hansen (2012). Profiling Linguistic Disability in German-Speaking Children. In M. J. Ball, D. Crystal & P. Fletcher (eds.), *Assessing Grammar: The Languages of LARSP*, Bristol, UK: Multilingual Matters, pp. 77–91.

Crossley, S. A. (2020). Linguistic features in writing quality and development: An overview. *Journal of Writing Research* 11(3), 415–443.

Crossley, S. A. & D. S. McNamara (2014). Does writing development equal writing quality? A computational investigation of syntactic complexity in L2 learners. *Journal of Second Language Writing* 26, 66–79.

Crystal, D., P. J. Fletcher & M. Garman (1976). *The grammatical analysis of language disability: A procedure for assessment and remediation*. London: Edward Arnold. Revised edition 1981.

Eisenberg, P., J. Peters, P. Gallmann, C. Fabricius-Hansen, D. Nübling, I. Barz & R. Fiehler (eds.) (2009). *Duden. Deutsche Grammatik*, vol. 4. Mannheim: Bibliographisches Institut & F. A. Brockhaus AG, 8. überarbeitete auflage ed.

Ellis, R. (2003). *Task-based Language Learning and Teaching*. Oxford, UK: Oxford University Press.

Fletcher, P., T. Klee & W. Gavin (2012). LARSP thirty years on. In M. J. Ball, D. Crystal & P. Fletcher (eds.), *Assessing Grammar: The Languages of LARSP*, Bristol, UK: Multilingual Matters, pp. 12–28.



- Gibson, E. (2000). The dependency locality theory: A distance-based theory of linguistic complexity. In A. Marantz, Y. Miyashita & W. O'Neil (eds.), *Image, language, brain: papers from the First Mind Articulation Project Symposium*, MIT, pp. 95–126.
- Grießhaber, W. (2019). 22. Profilanalysen. In *Sprachdiagnostik Deutsch als Zweitsprache*, De Gruyter Mouton, pp. 547–568.
- Hamp, B. & H. Feldweg (1997). GermaNet – a Lexical-Semantic Net for German. In *Proceedings of ACL workshop Automatic Information Extraction and Building of Lexical Semantic Resources for NLP Applications*. Madrid. URL <http://aclweb.org/anthology/W97-0802>.
- Hastie, T. & R. Tibshirani (1987). Generalized Additive Models: Some Applications. *Journal of the American Statistical Association* 82(398), 371–386.
- Housen, A. & F. Kuiken (2009). Complexity, Accuracy and Fluency in Second Language Acquisition. *Applied Linguistics* 30(4), 461–473.
- Keßler, J.-U. (2007). Assessing EFL-development online: A feasibility study of Rapid Profile. *Second language acquisition research: Theory-construction and testing* pp. 119–144.
- Kintsch, W. (2005). An Overview of Top-Down and Bottom-Up Effects in Comprehension: The CI Perspective. *Discourse Processes A Multidisciplinary Journal* 39(2 & 3), 125–128.
- Kuiken, F. & I. Vedder (2017). Functional adequacy in L2 writing: Towards a new rating scale. *Language Testing* 34(3), 321–336.
- Li, Y. & Z. Wu (2023). Chinesisches Deutschlerner-Korpus (CDLK). Ein umfangreiches Korpus mit Mehrebenen-Annotationen und multidimensionalen Metadaten. In M. Kupietz & T. Schmidt (eds.), *Neue Entwicklungen in der Korpuslandschaft der Germanistik*, Tübingen: Gunter Narr Verlag, pp. 223–236.
- Lou, Y., R. Caruana & J. Gehrke (2012). Intelligible Models for Classification and Regression. In *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, pp. 150–158.



- Mackey, A., M. Pienemann & I. Thornton (1991). Rapid Profile: A second language screening procedure. *Working Papers of the National Languages Institute of Australia* 1(1), 61–82.
- McCarthy, P. M. (2005). An assessment of the range and usefulness of lexical diversity measures and the potential of the measure of textual, lexical diversity (MTLD). Ph.D. thesis, University of Memphis. URL <https://umdrive.memphis.edu/pmmccrth/public/Papers/MTLD%20dissertation.doc>.
- McCarthy, P. M. & S. Jarvis (2010). MTLD, vocd-D, and HD-D: A validation study of sophisticated approaches to lexical diversity assessment. *Behavior Research Methods* 42(2), 381–392.
- Moranski, K. & N. Ziegler (2021). A case for multisite second language acquisition research: Challenges, risks, and rewards. *Language Learning* 71(1), 204–242.
- Nori, H., S. Jenkins, P. Koch & R. Caruana (2019). Interpretml: A unified framework for machine learning interpretability. *arXiv preprint arXiv:1909.09223* URL <https://doi.org/10.48550/arXiv.1909.09223>.
- Pallotti, G. (2023). Appropriate complexity. In C. Granget, I. Repiso & G. F. Sing (eds.), *Language, creoles, varieties: From emergence to transmission*, Berlin: Language Science Press, pp. 145–173.
- Pienemann, M. (1998). *Language Processing and Second Language Development: Processability Theory*. Amsterdam: John Benjamins.
- Rescher, N. (1998). *Complexity: A philosophical overview*. London: Transaction Publishers.
- Siemen, P., A. Lüdeling & F. H. Müller (2006). FALKO - ein fehlerannotiertes Lernerkorpus des Deutschen. In *Proceedings of Konvens*. Konstanz. URL <http://www2.informatik.hu-berlin.de/~siemen/daten/falko.pdf>.
- Skehan, P. (1989). *Individual Differences in Second Language Learning*. Edward Arnold.



- Weiss, Z. (2017). Using Measures of Linguistic Complexity to Assess German L2 Proficiency in Learner Corpora under Consideration of Task-Effects. Master's thesis, University of Tübingen, Germany. URL <http://www.sfs.uni-tuebingen.de/~zweiss/ma-thesis/weiss2017-distr.pdf>.
- Weiss, Z. & D. Meurers (2019a). Analyzing Linguistic Complexity and Accuracy in Academic Language Development of German across Elementary and Secondary School. In *Proceedings of the 14th Workshop on Innovative Use of NLP for Building Educational Applications (BEA)*. Florence, Italy: Association for Computational Linguistics.
- Weiss, Z. & D. Meurers (2019b). Broad Linguistic Modeling is Beneficial for German L2 Proficiency Assessment. In A. Abel, A. Glaznieks, V. Lyding & L. Nicolas (eds.), *Widening the Scope of Learner Corpus Research. Selected Papers from the Fourth Learner Corpus Research Conference*. Louvain-La-Neuve: Presses Universitaires de Louvain.



Themen

- Themen im longitudinalen CDLK Teilkorpus: *Meine Familie, Ein unvergessliches Erlebnis, Bildbeschreibung, Arbeit oder Masterstudium?, Sollten Handys im Klassenzimmer erlaubt sein?*

- Themen im PGG Querschnittskorpus Korpus:

Jahr	Themenstellung	# Texte	# Tokens
2012	<i>Praktikanten für den Chinesischunterricht gesucht</i>	2.783	494.525
2013	<i>Schreiben an die Eltern, um sie zu überzeugen, der Teilnahme ihres Kindes an einer ländlichen Mission zuzustimmen</i>	983	206.536
2014	<i>Vater und Sohn</i>	2.233	481.430
2015	<i>Eltern müssen Computerzeit der Kinder einschränken</i>	3.164	570.865
2016	<i>Auswanderung</i>	3.285	586.975
2017	<i>Masterstudium machen oder gleich arbeiten</i>	3.371	537.032
2018	<i>Die fünf beliebtesten Studienfächern</i>	3.432	550.801
2019	<i>Anteil der Männer mit Übergewicht und Adipositas in Deutschland in den Jahren 2005 bis 2017</i>	4.195	714.894
2020	<i>Verschuldung unter Jugendlichen</i>	3.968	666.475
2021	<i>Zeitmanagement</i>	3.981	616.631
		31.395	5.426.164



Longitudinales Korpus: CDLK Sample von 163 Lernenden

Gruppe	Semester 1		Semester 2		Semester 3		Semester 4		Summe	
	Text	Token	Text	Token	Text	Token	Text	Token	Text	Token
1	27	3.520	29	4.147	27	3.528	24	3.207	107	14.402
2	26	2.672	24	3.072	23	3.600	20	2.832	93	12.176
3	23	2.239	23	3.116	24	3.330	23	3.379	93	12.064
4	36	3.427	37	4.925	36	4.605	36	4.192	145	17.149
5	30	4.997	59	8.121	59	10.025	48	8.832	196	31.975
Summe	142	16.855	172	23.381	169	25.088	151	22.442	634	87.766