

EBEL, Louisa
Göttingen

Untersuchung der sprachlichen Komplexität von Abituraufgaben mithilfe von Natural Language Processing

Sprachliche Komplexität in Mathematikaufgaben

Zur Untersuchung der Schwierigkeit von Mathematikaufgaben wird Sprache zunehmend als Einflussfaktor in Betracht gezogen. Die Sprache der Mathematik im Allgemeinen charakterisiert sich zunächst durch ein Zusammenspiel der semiotischen Register natürliche Sprache, mathematische Symbole und Abbildungen (O'Halloran, 2015). Besonders den natürlichsprachlichen Elementen der mathematischen Kommunikation wird eine mathematikspezifische Struktur zugeschrieben, wengleich diese bisher noch nicht eindeutig charakterisiert werden konnte (Österholm & Bergqvist, 2013). Dies ist zum einen der Frage geschuldet, inwiefern mathematische Sprache mit etablierten linguistischen Methoden beschrieben werden kann; zum anderen hängt die sprachliche Struktur von ihrem Kontext, wie etwa der Textart, der Zielgruppe, dem Sachgebiet ab (Österholm & Bergqvist, 2013).

Für Aufgabenstellungen, als zentrale Textart des Mathematikunterrichts und in Prüfungen, stellt zunächst die COACTIV-Studie (Jordan et al., 2006) einen Vergleich zwischen der verwendeten Sprache und den mathematischen Bearbeitungsschritten an. Weitere Studien untersuchen einzelne Sprachebenen wie die Gebrauchshäufigkeit auftretenden Vokabulars (Dyrvold et al., 2015) und die syntaktische Struktur von Aufgaben (Österholm & Bergqvist, 2013). Zunehmend werden auch Analysen von Mathematikaufgaben durchgeführt, die mehrere Sprachebenen kombinieren. So verknüpfen beispielsweise Gürsoy et al. (2013) und Niederhaus et al. (2016) Strukturmerkmale wie Text- und Satzlänge, syntaktische Merkmale wie beziehungstragende Wortarten und lexikalische Komplexität, wie die Gebräuchlichkeit oder die Varianz des verwendeten Vokabulars, miteinander. In ihrer Untersuchung bildungssprachtypischer Sprachmittel machen Niederhaus et al. (2016) dabei Gebrauch einer computergestützten Sprachanalyse, die eine empirisch belastbare Betrachtung umfassenderer Aufgabenkorpora ermöglicht. Besonders die Nutzung von korpuslinguistischen Methoden und Natural Language Processing gilt als vielversprechend für eine empirisch gestützte Modellierung sprachlicher (Komplexitäts-)merkmale von Mathematikaufgaben.

Zielsetzung und Methodik

Dieses Projekt strebte eine erste empirische Charakterisierung der sprachlichen Komplexität von Mathematikaufgaben des Abiturs an. Hierzu wurden

In: L. Schick, M. Platz & A. Lambert (Hrsg.),
Beiträge zum Mathematikunterricht 2025.

58. Jahrestagung der Gesellschaft für Didaktik der Mathematik. WTM.

<https://doi.org/10.37626/GA9783959873307.0>

alle Aufgaben des Gemeinsamen Abituraufgabenpools der Länder der Jahre 2017 bis 2023 ($n = 562$) für die einzelnen Sachgebiete Analysis, Lineare Algebra, Stochastik und Vektorgeometrie betrachtet. Zielsetzung war hierbei auch eine Reflektion der verschiedenen Merkmale für sprachliche Komplexität für Mathematikaufgaben sowie eine Annäherung an das Natural Language Processing zur automatischen Verarbeitung größerer Datenmengen. Zur Bestimmung der sprachlichen Komplexitätsstruktur wurden an erster Stelle *Strukturmerkmale* wie die Anzahl Wörter und Sätze pro Aufgabe sowie die durchschnittliche Silbenzahl pro Wort erhoben. Ergänzt wurden *Wechsel semiotischer Register* von natürlicher Sprache zu symbolischen Sequenzen und Abbildungsbezügen, sowie zwei beispielhafte *Lesbarkeitsindices* (Flesch-Index und LIX) zur Überprüfung ihrer Anwendbarkeit für Mathematikaufgaben. Für die in den Aufgaben verwendete *Lexik* wurde die Nähe zum Alltagsregister, zum Sachtextregister und zu einem für Mathematikaufgaben typischen Register untersucht (vgl. Ebel et al., 2025). Zudem wurde die lexikalische Varianz bestimmt und die Konkretheit des verwendeten Vokabulars betrachtet. Die *syntaktische Komplexität* wurde zum einen durch die durchschnittliche Satzlänge, die mittlere Abhängigkeitsdistanz zwischen Wörtern sowie die Entropie operationalisiert. Weiterhin wurde das Maß an Beziehungsstrukturen über den Anteil der Wortarten Präposition, Pronomen, nebenordnende Konjunktion und unterordnende Konjunktion untersucht. Abschließend wurde die *Kohärenz* einer Teilaufgabe mit einleitenden Textbausteinen sowie dem Aufgabentext berechnet.

Die sprachlichen Kennwerte wurden mehrheitlich mithilfe des deutschen Sprachmodells der Python-Bibliothek spaCy (Honnibal et al., 2020) bestimmt. Ergänzend wurde die deutsche Formel des Flesch-Indexes sowie die Kohärenz, durch *sentence embedding* und Bestimmung des Kosinus zweier Textvektoren, berechnet. Die lexikalische Konkretheit beruht auf der *German list of extrapolated affective norms (GLEAN)* (Lüdtke und Hugentobler, 2023). Die Nähe zu Alltagsregister, Sachtextregister und Mathematikaufgabenregister wurde über die Häufigkeit im deutschen Untertitel-Korpus SUBTLEX (Brysbaert, 2024), im Wikipedia-Korpus (Zenodo, 2024) und in der Gesamtheit der Abituraufgaben operationalisiert. Merkmalsunterschiede zwischen Aufgaben verschiedener Sachgebiete wurden durch ANOVAs mit paarweisem t-Test (bzw. Kruskal-Wallis-Tests mit Dunn-Test) bestimmt.

Ergebnisse und Diskussion

Hinsichtlich ihrer Strukturmerkmale unterscheiden sich Aufgaben unterschiedlicher Sachgebiete vor allem in ihrer Wortlänge, hier über Silbenzahl pro Token ($F(3,558) = 51.91; p < .0001, \eta^2 = 0.22$). Ein paarweiser t-Test zeigt, dass Aufgaben der Stochastik die signifikant längsten Wörter

beinhalten, was komplexitätssteigernd wirken kann, in der Vektorgeometrie liegen die durchschnittlich kürzesten Wörter vor. Sowohl der Flesch-Index als auch der LIX als Lesbarkeitsskalen zeigen nach paarweisem t-Test dementsprechend eine signifikant verminderte Lesbarkeit von Abituraufgaben der Stochastik, während Gegenteiliges für die Vektorgeometrie gilt. Allgemein unterscheidet sich die Lesbarkeit der Abituraufgaben signifikant nach Sachgebiet, sowohl im Flesch-Index ($F(3,558) = 40.92; p < .0001, \eta^2 = 0.18$) als auch im LIX ($F(3,558) = 26.56; p < .0001, \eta^2 = 0.12$). Hinsichtlich der Nutzung symbolsprachlicher Ausdrücke, die eine Aufgabenstellung potenziell komplexer gestalten, bestehen ebenfalls deutliche Unterschiede zwischen den Sachgebieten ($F(3,558) = 83.89; p < .0001, \eta^2 = 0.31$). In Aufgaben der Stochastik sind nahezu keine symbolsprachlichen Ausdrücke enthalten, in der Analysis die höchste durchschnittliche Anzahl pro Aufgabe. Auch Bezüge zu Abbildungen ($F(3,558) = 37.15; p < .0001, \eta^2 = 0.17$) finden in der Stochastik am seltensten statt, in der Analysis am häufigsten. Die verwendete Lexik unterscheidet sich zwischen den Sachgebieten sowohl deutlich hinsichtlich ihrer Nähe zur Alltagssprache ($F(3,558) = 32.77; p < .0001, \eta^2 = 0.15$), sowie in geringem Maße zum Sachtextregister ($F(3,558) = 6.38; p < .001, \eta^2 = 0.03$). Zudem bestehen deutliche Unterschiede hinsichtlich der Konkretheit des verwendeten Vokabulars ($F(3,558) = 19.64; p < .0001, \eta^2 = 0.10$). Post-Hoc-Tests zeigen, dass Aufgaben der Analysis die geringste Konkretheit und die stärkste Distanz zum Sachtextregister aufweisen. Vektorgeometrie-Aufgaben zeigen sich als am wenigsten alltagssprachlich nach Untertitel-Korpus, jedoch als im Vergleich deutlich konkreter. Bezüglich der syntaktischen Komplexität zeigen sich vor allem starke Unterschiede hinsichtlich der Nutzung von Präpositionen ($F(3,558) = 59.84; p < .0001, \eta^2 = 0.24$), wobei diese in Aufgaben der Analysis den deutlich höchsten Anteil ausmachen, was auf eine syntaktische Verdichtung hinweisen könnte. Zwischen den Sachgebieten zeigen sich leichte Unterschiede in der Kohärenz von Teilaufgaben mit kontextualisierenden Textbausteinen ohne Instruktion ($F(3,558) = 6.28; p < .001, \eta^2 = 0.04$) und mit dem Gesamtaufgabentext ($F(3,558) = 3.51; p = .015, \eta^2 = 0.02$). Aufgaben der linearen Algebra und der Stochastik sind insgesamt am kohärentesten, Aufgaben der Analysis zeigen den stärksten Bezug von Teilaufgaben auf einleitende Textbausteine.

Die Ergebnisse zeigen, dass sprachliche Komplexitätsmerkmale unabhängig voneinander, jedoch abhängig vom Sachgebiet, dem die Aufgabe zugeordnet ist, häufig oder weniger häufig auftreten. Sprachliche Komplexität in Abituraufgaben wird demzufolge auf unterschiedlichen Ebenen gesteigert oder vermindert, sodass eine nur eindimensionale Betrachtung einzelner Parameter in jedem Fall zu kurz greifen würde. Unklar bleibt, ob bestimmte Sprachmerkmale gemeinsam verstärkt oder vermindert auftauchen, also zu,

eventuell kontextuell bedingten, Faktoren gehören, die bisher noch nicht eindeutig identifiziert wurden. So tauchen unterschiedliche beziehungstragende Wortarten nicht in den gleichen Aufgabengruppen gleichermaßen häufig auf und hinsichtlich der lexikalischen Gebräuchlichkeit bleibt offen, ob hier Wortfrequenzen in einem alltagssprachlichen Korpus oder die Konkretheit zielführender sind, da auch diese Kennwerte sich nicht parallel verhalten. Die Untersuchung bestätigt insgesamt Vorteile einer empirischen Analyse von Sprache in Mathematikaufgaben mithilfe von Natural Language Processing. Für eine belastbare Einschätzung der sprachlichen Komplexität sind jedoch weitere Einflussvariablen, konkrete Aufgabenabschnitte und Phrasen, sowie das Leseverhalten von Schüler*innen zu betrachten.

Literatur

- Brysbaert, M. (2024, Mai). *SUBTLEX-DE*. <https://osf.io/py9ba/>
- Dyrvold, A., Bergqvist, E., & Österholm, M. (2015). Uncommon vocabulary in mathematical tasks in relation to demand of reading ability and solution frequency. *Nordic Studies in Mathematics Education*, 20(1), 101–128.
- Ebel, L., Halverscheid, S. & Schroeder, S. (i.D.). On the Linguistic Complexity of Maths Leaving School Exams: Differences between content areas [zur Veröffentlichung eingereicht]. CERME14.
- Gürsoy, E., Benholz, C., Renk, N., Prediger, S., & Büchter, A. (2013). Erlös = Erlösung? - Sprachliche und konzeptuelle Hürden in Prüfungsaufgaben zur Mathematik. *Deutsch als Zweitsprache*, 1, 14–24.
- Honnibal, M., Montani, I., Van Landeghem, S., & Boyd, A. (2020). *spaCy: Industrial-strength Natural Language Processing in Python*. DOI: 10.5281/zenodo.1212303
- Jordan, A., Ross, N., Krauss, S., Baumert, J., Blum, W., Neubrand, M., Löwen, K., Brunner, M., & Kunter, M. (2006). *Klassifikationsschema für Mathematikaufgaben: Dokumentation der Aufgabenkategorisierung im COACTIV-Projekt*.
- Lüdtke, J., & Hugentobler, K. G. (2023, Dezember). *GLEAN*. <https://osf.io/a6w53/>
- Niederhaus, C., Pöhler, B., & Prediger, S. (2016). Relevante Sprachmittel für mathematische Textaufgaben - Korpuslinguistische Annäherung am Beispiel Prozentrechnung. *Deutsch als fremde Bildungssprache: Das Spannungsfeld von Fachwissen, sprachlicher Kompetenz, Diagnostik und Didaktik*, 135–162.
- O'Halloran, K. L. (2011). The semantic hyperspace: Accumulating mathematical knowledge across semiotic resources and modalities. In F. Christie & K. Maton (Hrsg.), *Disciplinarity: Functional linguistic and sociological perspectives* (S. 217–236). Continuum London.
- Österholm, M., & Bergqvist, E. (2013). What is so special about mathematical texts? Analyses of common claims in research literature and of properties of textbooks. *ZDM*, 45, 751–763. DOI: 10.1007/s11858-013-0522-6
- Zenodo. (2024). *Wikipedia-Korpus: Korpusquellen der deutschsprachigen Wikipedia im TEI-Format (Version 20240701)*. <https://doi.org/10.5281/zenodo.13120097>