

Marcel Ebert ¹
Friederike Korneck ¹
Christian Wartena ²
Christian Schumburg ³
Andreas Nehring ³
David Schmitt ⁴
Marie Irmer ⁴
Birgit Neuhaus ⁴

¹ Universität Frankfurt
² Hochschule Hannover
³ Universität Hannover
⁴ Universität München

Talk Moves

Adaption eines Sprachmodells aus den USA

Hintergrund

Die Unterrichtsqualität stellt einen zentralen Gegenstand bildungswissenschaftlicher Forschung dar und spielt zugleich eine wichtige Rolle in der Ausbildung von Lehrkräften. Verschiedene empirische Studien haben gezeigt, dass insbesondere die Kommunikation im Klassenzimmer, insbesondere das Unterrichtsgespräch, einen entscheidenden Einfluss auf den Lernprozess und den Lernerfolg von Schülerinnen und Schülern hat (Walshaw & Anthony, 2017; Lipowsky & Rzejak, 2022).

Um qualitätsvolle Unterrichtsgespräche zu fördern haben Michaels et al. (2016) konkrete Teacher Moves im Rahmen von Accountable Talk, im Sinne eines verantwortlichen Gesprächs oder einer Diskussion, formuliert. Diese dienen den Lehrkräften als Gesprächsstrategien, um Unterrichtsgespräche nach den Accountable Talk Framework zu gestalten.

Jacobs et al. (2022) ergänzen die Teacher Moves mit Gesprächsstrategien für Schüler:innen die im Kategoriensystem der Talk Moves im Unterricht abgebildet werden. Diese Talk Moves und ein entsprechendes Kodierungsmanual („Talk Moves Coding Manual“, 2021) verwenden sie schließlich, um Interaktionen zwischen Lehrkräften und Schüler:innen in Unterrichtsgesprächen automatisiert durch Large Language Models (LLMs) zu analysieren.

Projektziele

Das vorliegende Projekt nutzt eigens kodierte Transkripte aus deutschen naturwissenschaftlichen Unterrichtsstunden und analysiert diese nach dem Vorbild von Jacobs et al. (2022). Es werden LLMs als Klassifikator eingesetzt und mit englischen Daten kategorisierter Talk Moves aus dem Mathematikunterricht trainiert. Diese Modelle werden schließlich auf die deutschen Daten angewendet. Dabei wird insbesondere untersucht inwieweit der Klassifikator eine Übertragung in verschiedenen Bereichen erreicht:

1. von englischen zu deutschen Daten.
2. vom Mathematikunterricht zum Unterricht in den naturwissenschaftlichen Fächern (Physik, Chemie und Biologie).
3. von der US-amerikanischen zur deutschen Unterrichtskultur.

Datengrundlage und Methodik

In einem ersten Schritt wurden Unterrichtsvideos aus den genannten naturwissenschaftlichen Fächern transkribiert. Mit Hilfe des englischen Kodierungsmanuals, das übersetzt und an den deutschen naturwissenschaftlichen Unterricht angepasst wurde, wurden einzelne Sätze der Transkripte in Bezug auf Talk Moves manuell kodiert. Für jedes Fach wurden je ca. 300 Sätze von Lehrkräften und Schüler:innen kodiert.

Abbildung 1 stellt die Häufigkeitsverteilung von Talk Moves in den Lehrkräfteaussagen dar. Auffällig ist die unausgewogene Verteilung sowohl in den englischen Test- und Trainingsdatensätzen als auch in dem deutschen Datensatz.

Dabei wird auch deutlich, dass im deutschen Datensatz mehr Talk Moves kodiert werden konnten und dementsprechend hier die Häufigkeit in der Klasse 0 (kein Talk Move in den Sätzen erkannt) geringer ist.

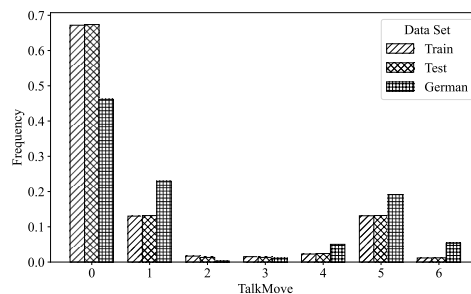


Abb. 1: Häufigkeiten der Teacher Talk Moves in den ursprünglichen Trainings- und Testdaten und im neuen deutschen Datensatz.

Mit dieser Datengrundlage wurden die Untersuchungen mit den Klassifikatoren durchgeführt. Verwendet werden zwei mehrsprachigen LLMs: mDeberta und XLM Roberta Base. Bei ausgewählten Modellen findet ein fine-tuning mit Low-rank Adaption (LoRA) nach Hu et al. (2021) oder einer Gewichtung der Kategorien (weighted) statt (Tab.1, II-V). Die Güte der verschiedenen LLMs werden für die deutschen und die englischen Testdaten von Suresh et al. (2021) verglichen. Die Güte der LLMs der englischen Testdaten werden zudem mit den von Suresh et al. (2021) berichteten Werten von deren Roberta Large Modells verglichen (Tab.1, I). Jedoch ist Roberta Large kein multilinguales Modell, wodurch eine genaue Übertragung des Modells auf den deutschen Datensatz nicht durchgeführt wird. Die Daten der jeweiligen Fächer werden getrennt voneinander durch die Klassifikatoren kodiert. Für die quantitative Analyse der Modellleistung wurden der F1-Score und der Matthew Correlation Coefficient (MCC) verwendet und die Klasse 0 exkludiert, da diese kein Talk Move im eigentlichen Sinne darstellt. Im Anschluss wurden die falsch kodierten Sätze qualitativ im Hinblick der drei beschriebenen Bereiche analysiert.

Ergebnisse

Tabelle 1 zeigt die Ergebnisse für die Talk Moves der Lehrkräfte. Bemerkenswert ist, dass mit den gleichen englischen Daten, jedoch mit anderen Modellen höhere Übereinstimmungswerte erreicht werden können als Suresh et al. (2021). Es zeigt sich auch, dass die Modelle mit den

englischen Daten immer besser abschneiden als mit den deutschen Daten. Die besten Ergebnisse für die englischen Daten erzielt für den F1-Score das Modell mDeberta (0.82) und die gewichtete Variante des gleichen Modells für den MCC (0.82). Bei den deutschen Daten erzielt das XLM Roberta Base Modell mit LoRA fine-tuning und gewichteten Klassen die besten Ergebnisse (F1: 0.62, MCC: 0.59).

In Tabelle 2 werden die Scores des fünften Modells für die deutschen Daten weiter für die drei naturwissenschaftlichen Fächer aufgeschlüsselt. Dabei zeigt sich, dass für die Fächer Chemie und Biologie die Modellleistung ähnlich ist. Für Physik erzielt dabei das Modell höhere Scores.

Des Weiteren zeigt eine vertiefte qualitative Analyse, dass Talk Moves nicht erkannt werden, wenn bestimmte sprachliche Besonderheiten auftauchen, wie z.B. plurale Imperative. Ein Beispiel dazu ist, dass „Ihr prüft bitte eure Vermutungen“ eher erkannt wird als „Prüft bitte eure Vermutungen“.

Modell	Englische Daten		Deutsche Daten	
	F1	MCC	F1	MCC
I. Roberta Large	0.79	0.78		
II. mDeberta	0.82	0.81	0.56	0.54
III. mDeberta (weighted)	0.82	0.82	0.56	0.54
IV. XLM Roberta Base / LoRA	0.75	0.76	0.60	0.57
V. XLM Roberta Base/ LoRA (weighted)	0.75	0.75	0.62	0.59

Tab. 1: Güte der Lehrkraftmodelle (gemittelter F1-Score und Mathews-Korrelationskoeffizient (MCC), I. von Suresh et al. (2021) berichtete Ergebnisse, II. – V. eigene Implementierung der beschriebenen Methode.

Daten	F1	MCC
Physik	0.69	0.70
Chemie	0.58	0.51
Biologie	0.58	0.56

Tab. 2: Ergebnisse zum Teacher Model für die drei deutschen Teilmengen unter Verwendung des XML Roberta Base Modell fine-tuned mit LoRA und Klassengewichtung.

Schlussfolgerungen

Die Ergebnisse zeigen, dass die mit englischen Daten trainierten Modelle in der Lage sind auch deutsche Äußerungen von Lehrkräften zu klassifizieren. Jedoch werden mit englischen Daten eine bessere Übereinstimmung der Modelle erreicht.

In Bezug auf die in den Forschungsfragen beschriebenen Bereich der unterschiedlichen Sprache legt das nicht Erkennen von Talk Moves durch sprachliche Besonderheiten nahe, dass für leistungsstärkere Modelle ein Training mit deutschen Daten ergänzt werden muss. Auch im Bereich der Disziplin, dem Unterschied zwischen der Mathematik und den Naturwissenschaften, zeigt sich, dass eine Übertragung möglich ist. Für die Physik gestaltet sich eine Übertragung besonders erfolgreich. Dieses Ergebnis ist möglicherweise durch die inhaltliche Nähe der beiden Fächer zu erklären. Bezüglich der unterschiedlichen Unterrichtskulturen lassen sich keine Indizien für einen Einfluss auf die Übertragung finden.

Literatur

- Hu, E. J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., Wang, L. & Chen, W. (2021). Lora: Low-rank adaptation of large language models.
- Lipowsky, F. & Rzejak, D. (2022). Unterrichtsgespräche erfolgreich führen. Eine zentrale Kernpraktik von Lehrkräften. *journal für lehrerInnenbildung*, 22(3), 58–73. <https://doi.org/https://doi.org/10.35468/jlb-03-2022-04>
- Michaels, S., O'Connor, M. C., Hall, M. W. & Resnick, L. B. (2016). *Accountable talk sourcebook: For classroom conversation that works*. Univeristy of Pittsburgh.
- Talk Moves Coding Manual. (2021). Verfügbar 5. Januar 2025 unter [https://github.com/SumnerLab/TalkMoves/blob/main/Coding% 20Manual.pdf](https://github.com/SumnerLab/TalkMoves/blob/main/Coding%20Manual.pdf)
- Jacobs, J., Scornavacco, K., Harty, C., Suresh, A., Lai, V. & Sumner, T. (2022). Promoting rich discussions in mathematics classrooms: Using personalized, automated feedback to support reflection and instructional change. *Teaching and Teacher Education*, 112. <https://doi.org/10.1016/j.tate.2022.103631>.
- Suresh, A., Jacobs, J., Lai, V., Tan, C., Ward, W., Martin, J. H. & Sumner, T. (2021). Using transformers to provide teachers with personalized feedback on their classroom discourse: The talkmoves application. <https://arxiv.org/abs/2105.07949>
- Walshaw, M. & Anthony, G. (2017). The teacher's role in classroom discourse: A review of recent research into mathematics classrooms. *Review of Educational Research*, 78(3), 516-551. <https://doi.org/10.3102/0034654308320292> (Original work published 2008)