

MARCO MAGIRIUS

Klassenanalysen von Kontextpräferenzen Deutschstudierender beim Interpretieren literarischer Texte mittels poLCA und K-Means

Abstract

Beim Deuten literarischer Texte wird Bedeutung aus verschiedenen Kontexten vom Interpretieren an den literarischen Text herangetragen. In der vorliegenden Fragebogenstudie, die Teil eines Mixed-Methods-Projektes ist, wurden Kontextpräferenzen von 467 Studierenden des Lehramts Deutsch klassifiziert. Mit Hilfe von K-Means wurden 4 Typen von Präferenzmustern extrahiert und mit einer LCA ausgeschärft sowie (inferenz-)statistisch abgesichert. Der Beitrag, inhaltlich verortbar in der Lehrerprofessionalisierungsforschung, richtet sich insbesondere an (Fach-)Didaktiker/innen und Bildungswissenschaftler/innen, die eine Klassenanalyse selbst durchführen wollen, reflektiert das nötige Hintergrundwissen sowie Möglichkeiten und Grenzen der eingesetzten Verfahren.

When interpreting literature the interpreter utilizes different contexts to assign meaning to elements of the text or the text in a whole. The present questionnaire study is part of a mixed methods approach and categorizes context preferences of 467 German teacher students. In the first place 4 types of preferences were extracted using k-means. After that a latent class analysis was carried out to hone and statistically secure the found patterns. The article, to be located in the field of research on teacher education, is addressed to education researchers who want to implement a cluster analysis autonomously, reflects upon the needed background knowledge as well as possibilities and limitations of the employed methods.

1 | Einleitung

Empirieaffine Nachwuchswissenschaftler/innen im Bereich der Deutschdidaktik sehen sich nach der empirischen Wende innerhalb ihres Fachgebiets mit der Schwierigkeit konfrontiert, empirische Methoden für deutschdidaktische Fragestellungen fruchtbar zu machen, (meist) ohne während des Lehramtsstudiums mit dem dafür notwendigen fundierten Wissen ausgerüstet worden zu sein. Dies gilt insbesondere für quantitative Methoden, da das Unterrichtsfach Deutsch selten mit mathematischen Fächern oder Psychologie kombiniert wird. Die Einbindung empirischer Verfahren in Abschlussarbeiten oder gar Praktikumsberichte hilft im Falle quantitativer Empirie nur bedingt, da dort meist mit sehr kleinen Stichproben gearbeitet wird. Stand die Anwendung solcher Verfahren innerhalb der Deutschdidaktik auch in der Forschung anfangs noch unter „Welpenschutz“, wird nun ebenfalls methodisch state of the

art gefordert. Die Einarbeitung gestaltet sich dabei für Deutschdidaktiker/innen schon allein aufgrund der verfügbaren Literatur als überaus schwierig. Entweder bieten Lehrwerke eher Überblickswissen, das für die Anwendung in einem größer angelegten Projekt zu unspezifisch bleibt, oder die mathematischen Details wirken so einschüchternd, dass man sich unweigerlich die Frage stellt, was davon für die Deutschdidaktik relevant sein kann. Mit der Diskussion meiner Resultate möchte ich diese Frage für meine Untersuchung zu Kontextpräferenzen Studierender des Lehramts Deutsch bezüglich Klassenanalysen mit K-Means und einer LCA¹ exemplarisch beantworten. Ferner werde ich darlegen, für welche Funktion innerhalb eines Mixed-Methods-Ansatzes diese quantitativ ausgerichtete Teilstudie inwiefern geeignet ist. Dafür ist eine kurze Einführung in das Gesamtdesign nötig. Nach der darauffolgenden Skizzierung der Relevanz der Fragestellung wird das erhobene Konzept spezifiziert. Auf die Pilotierung folgt der Hauptteil zur Klassenanalyse mit K-Means und der Analyse latenter Klassen. Abschließend werden die Ergebnisse interpretiert, diskutiert sowie das Potential von Klassenanalysen umrissen.

2 | Das Mixed-Methods-Projekt

2.1 | Überblick

Die in diesem Beitrag diskutierte Fragebogenstudie ist Teil eines Mixed-Methods-Projekts, das im Bereich der empirischen Professionalisierungsforschung verortet werden kann. In jenem Projekt habe ich mit Fragebögen und Interviews die Überzeugungen Deutschstudierender zum Interpretieren literarischer Texte erhoben. Da die Fragebögen quantitativ und die Interviewtranskripte qualitativ ausgewertet werden, handelt es sich um einen Mixed-Methods-Ansatz im engeren Sinne, genauer gesagt um ein explanatory design oder Vertiefungsdesign (Kuckartz 2014), denn der quantitative Untersuchungsteil ist vorgelagert und dessen Ergebnisse werden durch den Interviewteil ausgeschärft und angereichert.

Ein Mixed-Methods-Ansatz kann im Falle geglückter Integration mehrerer Teilstudien durch reichhaltige Synergien überzeugen, bringt jedoch erheblichen Aufwand mit sich, weshalb die Entscheidung für dieses Vorgehen kurz zu begründen ist: In der vorliegenden Studie werden Konzepte Studierender erhoben, die sich im „[g]enerische[n] Strukturmodell professioneller Kompetenz“ (angehender) Lehrkräfte in die „Kompetenzfacette“ „Werthaltungen und Überzeugungen“ einordnen lassen, auch wenn trotz des „unterschiedlichen epistemologischen Status“ die Übergänge zum Bereich des Wissens „fließend sind“ (Kunter et al. 2011, 41).² Die Erhebung solcher „überdauernder[r] [...] Annahmen“ gestaltet sich mit Fragebögen oder Interviews allein als schwierig, insbesondere da es sich hier u.a. um abstrakte Metatheorien handelt, die zwar subjektiv für wahr gehalten werden, jedoch potentiell inkonsistent sein können.

Mit Fragebögen sind diese schwierig zu operationalisieren (vgl. Priemer 2006). Als Problem erscheint hierbei, dass „Themen der Erkenntnis- und Wissenschaftstheorie [...] einen sehr breiten inhaltlichen Umfang umfassen“ (ebd., 166) und forschers- wie probandenseitig mit erheblichem Interpretationsspielraum und daher Validitätsproblemen behaftet sind. Sie sind also schwierig durch Items als potentiell „extrem verkürzte Aussagen“ (ebd.) abbildbar. Mit Interviews allein ergibt sich wiederum das Problem der Bildung von Artefakten. Beeinflusst man die Proband/innen in der Gestalt, dass man sie auf ein metatheoretisches Reflexionsni-

¹ Im Deutschen wird dieses Verfahren meist als „latente Klassenanalyse“ bezeichnet, was m.E. eine unglückliche Übersetzung ist, da man keine latente Analyse betreibt, sondern latente Klassen analysiert.

² Der Bereich „Überzeugungen“ ist bisher für den Literaturunterricht domänenspezifisch nicht klar umrissen. Dies gilt für „beliefs“, verstanden als „misconceptions“, „doubt[s]“, „possessions“, „reduced prototypes“ oder „filters of mind“, ebenso in der Mathematikdidaktik als schwierig (Toerner 2005, 312). Das Ausschärfen und Abgrenzen der Begriffe wird wohl in einer ill-structured domain wie dem Literaturunterricht nicht einfacher fallen und kann in diesem Rahmen nicht geschehen.

veau hebt, das sie außerhalb der künstlichen Interviewsituation nicht erreichen würden, werden die erhobenen Konzepte wohl kaum in der späteren Praxis handlungswirksam sein.

Diese Probleme sind nicht vollends zu lösen, werden jedoch minimiert, indem man innerhalb eines Mixed-Methods-Ansatzes die Methoden so beansprucht, dass sie ihre Möglichkeiten entfalten und ihre Grenzen nicht überschritten werden. So versucht der erste quantitative Schritt, Zusammenhänge innerhalb eines breiten Panoramas von Überzeugungen aufzuspüren und Bereiche auszumachen, die weiterverfolgt werden sollten. Dieser Beitrag beleuchtet mit den Kontextpräferenzen Studierender nur einen Teil dieses Panoramas. Der zweite, qualitative Schritt fokussiert dann „individuelle[] Konstellationen“ von Überzeugungen (Wieser 2013, 175) sowie Motive Studierender für die Entscheidungsfindung. Bevor die Erhebung genauer vorgestellt und das Auswertungsverfahren erläutert wird, soll jedoch die deutschdidaktische Relevanz der Thematik kurz aufgezeigt und so die Studie motiviert werden.

2.2 | Zur deutschdidaktischen Relevanz

Ich erhebe Überzeugungen Deutschstudierender zum Interpretieren literarischer Texte. Es ist davon auszugehen, dass diese Überzeugungen von großen Unsicherheiten seitens der Studierenden (vgl. Wieser 2013) geprägt sind. Etwaige Übersimplifizierungen gründen dabei eventuell u.a. auf der unübersichtlichen Pluralität literaturtheoretischer Positionen, der genuin literarischen Ambiguität der Gegenstände und den damit einhergehenden Kontroversen in der Literaturwissenschaft. Da man mit Blick auf Resultate aus der Psychologie (vgl. Hofer 2004) und den Didaktiken der MINT-Fächer (vgl. Urhahne und Hopf 2004) von einem erheblichen Einfluss dieser Überzeugungen sowohl auf den Umgang mit Lehrinhalten im universitären Kontext als auch auf den späteren eigenen Unterricht ausgehen kann, stellt sich dringlich die Frage nach Förder- bzw. Interventionsbedarf. Bleibt die Elaboration jener Überzeugungen aus, werden in der Praxis des Deutschunterrichts möglicherweise defiziente Muster tradiert. Beispielweise ist denkbar, dass die angehenden Lehrkräfte im späteren eigenen Unterricht auf die Komplexität literarischen Lernens mit schematischem Abarbeiten von Algorithmen zur Literaturanalyse, mit der Vorgabe von Deutungen oder etwa mit der Öffnung des Unterrichts für beliebige Interpretationen reagieren. Folgt man literaturwissenschaftlichen Positionen wie der von Eco (1987), wäre ein solcher Umgang mit Literatur dem literarischen Gegenstand ästhetisch nicht angemessen und mit aktuellen literaturdidaktischen Zielvorstellungen von Deutschunterricht, beispielsweise jenen von Zabka (2015a), schwierig zu vereinbaren.

Bevor man die Auswirkungen in den Blick nimmt, müssen Struktur und Elaborationsgrad der Überzeugungen Deutschstudierender zum Interpretieren literarischer Texte empirisch erfasst und in Ansätzen typisiert werden. In Folgeprojekten ist dann zu untersuchen, ob diese Überzeugungen im Literaturunterricht handlungswirksam werden. Falls dem so ist, wäre die Entwicklung und Überprüfung der Wirksamkeit typspezifischer treatments anzuschließen, die zu einer Weiterentwicklung unvorteilhafter Überzeugungen in eine vorteilhaftere Richtung anregen. Die Erhebung der Überzeugungen Studierender ist also der erste Schritt. Im Folgenden gehe ich nur auf die Erhebung von Kontextpräferenzen innerhalb des quantitativen Untersuchungsteils dieses ersten Schritts ein.

3 | Erhebung von Kontextpräferenzen

3.1 | Konzeptspezifikation und Zielstellung

Unabhängig davon, wie man den „systematisch mehrdeutig[en]“ „Ausdruck“ *Interpretation* verwendet (Kindt und Köppe 2008, 10), sei hier postuliert, dass für eine geglückte Interpretation Wissen „zuerst ausgewählt und so reformuliert werden“ muss, „dass es zu dem interpretierten Text passt und als dessen Sinn behauptet werden kann“ (Zabka 2015b, 220). Diese „Kontextualisierung“ als Zuweisung von Bedeutung an den Text oder an Elemente des Textes

„ist [demnach] notwendige Bedingung“ jeder „Interpretation“ (Zabka 2007, vgl. Danneberg 2007). Die Kontexte lassen sich im Anschluss an Fricke (1992, 216) mit Hilfe dieses erweiterten Kommunikationsmodells klassifizieren.

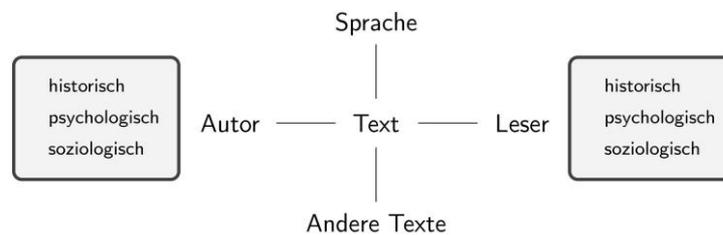


Abb. 1: Kommunikationsmodell analog zu Kindt und Köppe (2008)

Den Elementen des erweiterten Kommunikationsmodells lassen sich die bekannten literaturtheoretischen Ansätze/Positionen zuordnen – natürlich „nicht ohne Gewaltsamkeiten“ (ebd.), also eventuell von Vertretern dieser Positionen als unzulässig empfundene Vereinfachungen. Sicherlich häufig auftretende Mischformen lassen sich genauer beschreiben, wenn man dieses Prinzip der Zuordnung lediglich eines Elements zu einer Prioritätenliste von Elementen bzw. den damit assoziierten Kontexten verallgemeinert.

Statt einer literaturtheoretischen Beschreibung analysiere ich, welche Kontexte Studierende präferieren und wie hoch sie diese priorisieren. Dabei beschränke ich mich auf „biografisches“, psychologisches (jeweils bezüglich des Autors), „realhistorisches“ und ideengeschichtliches Wissen (Zabka 2015b, 221), also eine echte Teilmenge möglicher „extratextuelle[r] Kontexte“ (Danneberg 2007). Im Folgenden konzentriere ich mich nur auf die Analyse des Fragebogenteils auf Abbildung 2.

Bei einer Interpretation deutet man einen literarischen Text und versucht für diese Deutung gute Argumente zu finden. Wie wichtig sind folgende Arten von Argumenten für Ihre Interpretationen?
Argumente, die ...

	unwichtig	eher unwichtig	eher wichtig	wichtig
... sich auf die Psyche des Autors/der Autorin beziehen	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
... sich auf die Lebensgeschichte des Autors/der Autorin beziehen	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
... sich auf die gesellschaftlich-politischen Gegebenheiten der Entstehungszeit beziehen	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
... sich auf philosophisch-ideengeschichtlichen Aspekte der Entstehungszeit beziehen	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
... erläutern, wie Leser den Text (eventuell verschieden zur Autor/innen-Intention) verstehen können	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Abb. 2: Items für die Erhebung von Kontextpräferenzen in der Haupterhebung

Eine Klassenanalyse zielt darauf ab, die Proband/innen hinsichtlich der mit diesem Fragebogen erhobenen Kontextpräferenzen in Gruppen einzuteilen, sodass „das Antwortverhalten innerhalb einer Gruppe ein gemeinsames Muster aufweist“, während sich die „ermittelten“ Klassen paarweise „maximal“ voneinander „unterscheiden“ (Winkler 2011, 244).

Um dieses Ziel zu erreichen, wählte ich zwei Verfahren der Klassenanalyse, K-Means und die Analyse latenter Klassen. Beide sind in der Programmiersprache R implementiert und

im Gegensatz zu beispielsweise MPLUS frei verfügbar. Ein weiterer Vorteil von R besteht in der Möglichkeit, durch die Installation zusätzlicher Pakete den Funktionsumfang im Vergleich mit MPLUS erheblich zu vergrößern und somit äußerst flexibel auf die Notwendigkeit des Einsatzes weiterer Verfahren zu reagieren. Die R-community ist äußerst aktiv, weshalb neue statistische Verfahren schnell Umsetzung finden. Man kann deshalb vereinfachend konstatieren, dass sich für nahezu jedes statistische Verfahren Pakete finden lassen, weshalb ein Einstieg in R für (Nachwuchs-)Wissenschaftler/innen, die quantitativ arbeiten wollen, von Vorteil ist. Da die hier verwendeten Pakete, `poLCA` und `stats`, gut dokumentiert sind, eignet sich dieses Projekt durchaus als erste Erfahrung mit R. Für die Klassenanalysen hilfreiche Grundlagen, beispielsweise das Schreiben von Werten aus einer Tabelle in eine andere, findet man bei Kabacoff (2010).

Für Klassenanalysen, wie sie nach der Haupterhebung stattfanden, war die Stichprobe der Pilotierung mit $n=64$ zu klein. Dennoch möchte man die Plausibilität der Daten und die Möglichkeit, Klassen zu extrahieren, vorab ausleuchten. Hier ist eine Faktorenanalyse hilfreich. Statt wie bei der Klassenanalyse die Proband/innen zu klassifizieren, werden bei der Faktorenanalyse Items „gemäß ihrer korrelativen Beziehungen“ zueinander gruppiert (Bortz und Schuster 2010, 386). Jenen Gruppen von Items wird, vereinfacht gesagt, ein Faktor zugeordnet. Wenn alle Items einem einzigen Faktor zugeordnet werden, wäre eine Klassenanalyse aussichtslos, was bei meinen Daten nicht der Fall war.

Statt mit einzelnen Items zu arbeiten, wäre eine Erhebung durch mehrere Items pro Ansatz testtheoretisch natürlich vorzuziehen. Da den angestrebten Klassenanalysen jedoch vor allem eine heuristische Funktion innerhalb meines Gesamtdesigns zukommt, wurde eine exaktere Erfassung auf den qualitativen Teil verschoben. Dies ist gewiss suboptimal, jedoch mit der Schwierigkeit, metatheoretische Überzeugungen zu operationalisieren, begründbar. Auch wenn zudem eine Triangulation schon während der Pilotierungsphase sicherer gewesen wäre, wurde die Haupterhebung bezüglich dieses Fragebogenabschnitts also auf diese Weise vorbereitet. Die Haupterhebung fand mit 467 Studierenden des Lehramts Deutsch verschiedener Schulformen³ an den Universitäten Hildesheim (43%), FSU Jena (26%), Uni Hamburg (18%), TU Dresden (12%) und JLU Gießen (3%) statt.

3.2 Klassenanalyse mit K-Means

Die erste Klassenanalyse führte ich mit der R-Funktion `kmeans()`⁴ durch. Bei diesem Verfahren wird jedem Item eine Achse des in meinem Fall 4-dimensionalen Koordinatensystems⁵ zugeordnet. Die Abstufungen je Item, unwichtig – eher unwichtig – eher wichtig – wichtig, werden auf den jeweiligen Achsen abgetragen. Ein Proband bzw. eine Probandin, der oder die im ersten Item die erste Abstufung wählt, im zweiten die zweite, im dritten die dritte und im vierten die vierte Abstufung ankreuzt, wäre also durch den Punkt (1, 2, 3, 4) repräsentiert. Neben dem Ankreuzverhalten der Proband/innen muss die Zahl der zu extrahierenden Klassen dem Algorithmus mitgeteilt werden. Für jede Klasse wird nun im Koordinatensystem eine Punkthäufung, genannt Cluster, gesucht, weshalb man K-Means auch als Clusterzentrenanalyse bezeichnet. Entscheidet man sich für 4 Klassen, werden 4 Punkte als vorläufige Clusterzentren zufällig im Koordinatensystem gewählt.

Nun berechnet man den Abstand der „Fälle“ zu den Clusterzentren und weist jedem „Fall“ das Zentrum bzw. den Cluster mit dem geringsten Abstand zwischen ihm und dem jeweiligen Zentrum zu. Für alle so entstehenden Cluster berechnet man neue Clusterzentren als Mittelwerte aller Fälle des jeweiligen Clusters (4 „means“) und wiederholt die Schritte dieses Absatzes, bis sich die Clusterzentren nicht mehr oder nur noch sehr wenig ändern.

³ Gymnasium (34%), ausschließlich Sekundarstufe 1 (33%), Grundschule + Sekundarstufe 1 (25%), Berufsschule (6%), Sonderschulpäd. (2.2%)

⁴ Mit `kmeans()` und `poLCA()` referiere ich auf die konkreten Implementierungen/Befehle der verwendeten Pakete, während K-Means und LCA die Verfahren allgemein bezeichnen.

⁵ Ein 4-dimensionales Koordinatensystem kann man sich schlecht bildlich vorstellen. Im Falle von 2 oder 3 Items, also innerhalb eines 2- oder 3-dimensionalen Koordinatensystems, verläuft das Verfahren analog.

Die Clusterzentren repräsentieren zum Schluss die prototypischen Vertreter ihrer jeweiligen Klasse. Die simple Verfahrensweise bringt einige Probleme mit sich. Die Konstruktionsweise über Abstände erschwert beispielsweise das Erkennen konkaver Punkthäufungen als Klassen. Ferner ist das Ergebnis maßgeblich durch die Anfangswerte beeinflusst. Ähnlich problematisch ist die Wahl der Anzahl der Klassen, die dem Anwender oder der Anwenderin obliegt. Die beiden letztgenannten Probleme lassen sich meist wie folgt beheben. Mit der Standardimplementierung in R wird das Verfahren wie oben beschrieben angewandt, was dazu führt, dass für jeden Cluster die Summe S der Abstände zwischen den „Fällen“ und dem ihnen zugeordneten Clusterzentrum minimal wird. Dabei erlaubt es R über den Funktionsparameter `iter.max` der Funktion `kmeans()` das Verfahren mehrmals nacheinander anzuwenden, und nur das Resultat mit der kleinsten Summe aller S auszugeben. So lassen sich verschiedene Anzahlen von Klassen durchprobieren. Auf folgender Abbildung 3 sind auf der Abszisse die Anzahl der Klassen und auf der Ordinate die Summe aller Fehlersummen S abgetragen.

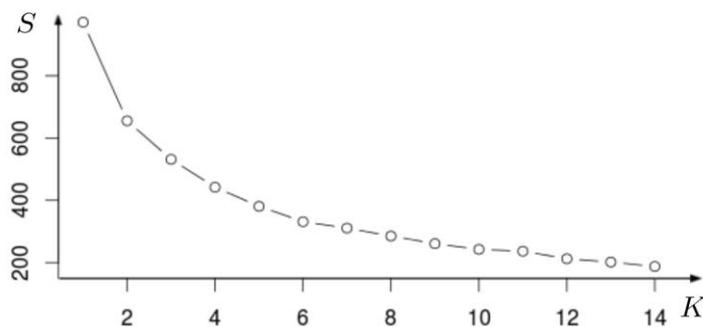


Abb. 3: Zusammenhang zwischen der Fehlersumme S und der Anzahl der Klassen K

Ähnlich zu Screeplots bei Faktorenanalysen wählt man als Zahl der Klassen den Abszissenwert des Punktes kurz vor dem „Scree“, engl. für Geröll, da man die Daten gewöhnlich mit so wenig wie möglich Klassen beschreiben und dennoch die Summe der S so klein wie möglich halten möchte.

In meinem Fall wird an der Glätte des Graphen deutlich, dass für keine Anzahl von Klassen die Zunahme einer weiteren Klasse die Summe aller Abstände zwischen den Fällen und ihren Clusterzentren signifikant verringert. Deshalb kann die Entscheidung für 4 Klassen nicht aus ihm abgeleitet werden. Obwohl der Graph am ehesten die Entscheidung für zwei Klassen nahelegt, lässt sich konstatieren, dass keine Anzahl dem Verlauf des Graphen widerspricht. Hier half der Screeplot also nur bedingt weiter. Man kann ihn jedoch auch verwenden, um die nötige Anzahl der Wiederholungen abzuschätzen, um somit die Abhängigkeit von den Anfangswerten zu eliminieren.

Ab 50 Wiederholungen pro Anzahl der Klassen veränderte sich der Graph nicht mehr, blieb also in der abgebildeten Gestalt, weshalb man mit guten Gründen annehmen kann, dass hinsichtlich der Größe S keine besseren Anfangswerte übersehen wurden. Die nachstehende Abbildung 4 ist wie folgt zu verstehen. Ein prototypischer Vertreter bzw. eine prototypische Vertreterin der Klasse C2 betrachtet für die Interpretation Argumente, die sich auf die Psyche des Autors beziehen als eher wichtig, auf die Lebensgeschichte und die gesellschaftlich-politischen Gegebenheiten bezogene als wichtig und die Kontexte Ideengeschichte/Philosophie als eher unwichtig. Analog dazu werden Prototypen der anderen Klassen abgebildet. Die Klassen unterscheiden sich nicht im Item „Leser“. Die prototypischen Vertreter/innen aller vier Klassen kreuzen bei diesem „wichtig“ an. Um die nachstehende Grafik mit der Klassenlösung so übersichtlich wie möglich zu gestalten, entfällt dieses Item, da vor allem Unterschiede zwischen den Klassen von Interesse sind.

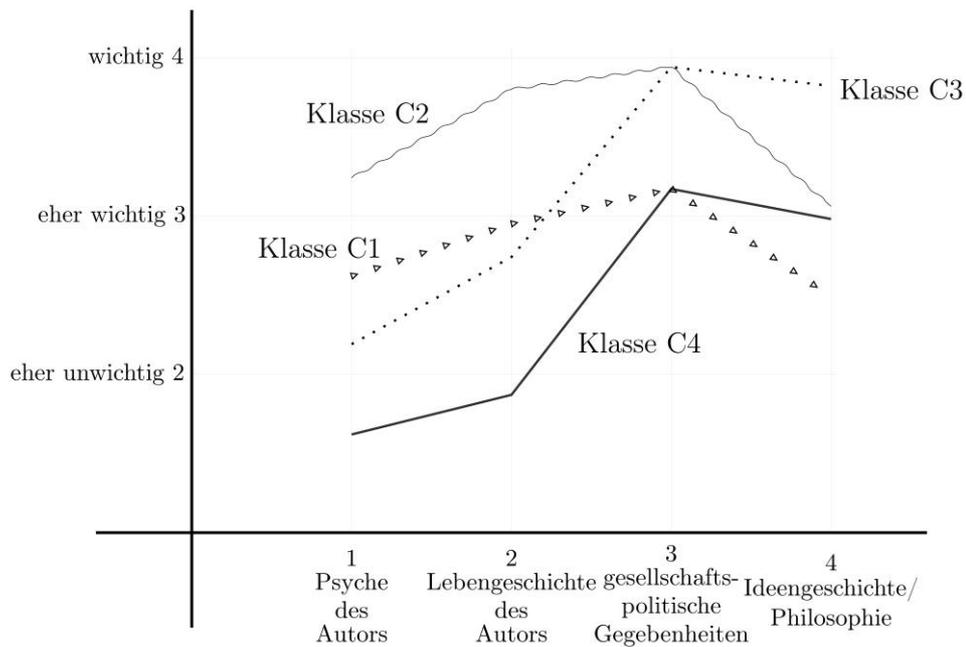


Abb. 4: Klassenlösung nach Anwendung von kmeans()

Betrachtet man die Abbildung genau, fällt auf, dass die Klassen C1 und C2 sowie die Klassen C3 und C4 durch ähnlich verlaufende, lediglich in der Höhe verschobene Graphen charakterisiert werden. Man sollte demnach entweder eine 2-Klassenlösung bevorzugen, was ebenso zu obigem Screeplot passt, oder ein aussagekräftigeres Verfahren bemühen, das uns weitere Hinweise zur zu bevorzugenden Anzahl der Klassen liefert und das die parallel verlaufenden Klassen ausdifferenziert. Letzteres geschieht nun mit Hilfe einer Analyse latenter Klassen. Diese hat gegenüber K-Means den weiteren Vorteil eines absoluten, also anwendungsübergreifenden Kriteriums für die Passung der gefundenen Lösung zu den Daten. Im Anschluss daran werden die Ergebnisse detaillierter interpretiert.

3.3 | Die Analyse latenter Klassen

Um die Schwächen und Probleme mit K-Means zu umgehen und dessen Lösung zu überprüfen bzw. auszuschärfen, verwende ich das kompliziertere Verfahren der Analyse latenter Klassen, kurz LCA. Der Algorithmus gehört ebenso wie K-Means und die Faktorenanalyse zur Gruppe der Data-Mining-Verfahren, welche durch das gemeinsame Ziel, Muster in großen Datenmengen zu finden, charakterisiert werden kann. Der LCA-Algorithmus bekommt die gleiche Eingabe wie kmeans() – das Antwortverhalten der Proband/innen für die ausgewählten Items sowie eine theoretisch hergeleitete bzw. vom Anwender oder der Anwenderin geratene Anzahl der Klassen. Er liefert als Ausgabe (vereinfacht gesagt) die Größen der extrahierten Klassen sowie für jede Klasse eine Tabelle, die im Aufbau dem Fragebogen gleicht, wobei statt der Kästchen die Wahrscheinlichkeit erscheint, mit der ein Vertreter oder eine Vertreterin der Klasse dort ein Kreuz setzen würde. Aus diesen Werten berechnet der Algorithmus für alle beobachteten Antwortmuster und jeweils alle Klassen, mit welcher Wahrscheinlichkeit ein Antwortmuster zu einer Klasse gehört. Den Prototypen jeder Klasse muss der Anwender oder die Anwenderin selbst berechnen, indem für alle Elemente jeder Klasse der Mittelwert für jedes Item gebildet wird.⁶

Das Vorgehen des Verfahrens zum Bestimmen der gesuchten Werte sei kurz umrissen, um im nächsten Abschnitt die Konstruktion der Güteindizes verstehen und einschätzen zu

⁶ Die Zuordnung der Zahlenwerte zu den Abstufungen ist wie bei kmeans(): unwichtig \approx 1, eher unwichtig \approx 2, eher wichtig \approx 3, wichtig \approx 4.

können. Der Algorithmus ermittelt die zu berechnenden Werte, indem er, vereinfacht gesagt, das Produkt aller Ankreuzwahrscheinlichkeiten maximiert. Dies geschieht unter der Bedingung, dass das Antwortverhalten innerhalb einer Klasse möglichst gleich sein soll. Dabei werden die Klassengrößen sowie die Ankreuzwahrscheinlichkeiten anfangs geraten und dann u.a. durch eine „Günstig-durch-möglich“-Rechnung aktualisiert, bis der zu maximierende Wert nicht mehr oder nur noch sehr wenig wächst. Man spricht hierbei von einer Maximum-Likelihood-Schätzung. Entgegen der default-Einstellung ist das Verfahren demnach analog zu `kmeans()` zu iterieren, um sich der Abhängigkeit von der zufälligen Wahl der Anfangswerte zu entledigen. Die geeignete Zahl der Iterationen muss ausprobiert werden. Da der Algorithmus nicht viel Rechenkapazität beansprucht, kann man hier großzügig sein. Der Algorithmus iterierte in meinem Fall 643 Mal. Bevor die Klassenlösung interpretiert wird, finden die für die LCA üblichen statistischen Indizes zur Bewertung der Lösung Anwendung. Die Beurteilung der Modellgüte kann sich dabei auf zwei Aspekte beziehen: (1) Vergleiche zwischen verschiedenen großen Modellen anhand der Informationskriterien und (2) Vergleich von Modell und Daten.

Zu (1): Die Anzahl der Klassen eines Modells wird mit Hilfe der sogenannten Informationskriterien BIC und AIC⁷ abgeschätzt. Das sind mathematische Funktionen, die die dem Schätzverfahren zugrundeliegende Erhöhung der Ankreuzwahrscheinlichkeiten belohnen und eine Vergrößerung der Anzahl der zu schätzenden Parameter bestrafen. Beim BIC fließt ferner die Stichprobengröße in den Strafterm ein (Moosbrugger und Kaleva 2008, 307). Die Werte der Informationskriterien werden von der Funktion `poLCA()` direkt ausgegeben.⁸ Der Vergleich der Werte für eine 2-,3-,4- und 5-Klassenlösung liefert für 4 Klassen die besten, weil niedrigsten Werte.

	2 Klassen	3 Klassen	4 Klassen	5 Klassen
<i>AIC</i>	3769.821	3695.638	3630.694	3626.943
<i>BIC</i>	3873.480	3853.199	3842.157	3892.308

Tabelle 1: Informationskriterien *AIC* und *BIC* für verschiedene Anzahlen von Klassen.

Eine 4-Klassen-Lösung ist insbesondere einer 5-Klassen-Lösung vorzuziehen, da $AIC(4) - AIC(5) \approx 5$, jedoch $BIC(5) - BIC(4) \approx 50$ gilt.

Zu (2): Die Passung zwischen Modell und Daten lässt sich inferenzstatistisch prüfen. Dazu vergleicht man die aus den mit Ankreuzwahrscheinlichkeiten (`poLCA.posterior`) berechneten geschätzten Häufigkeiten mit der Anzahl des tatsächlichen Auftretens der Antwortmuster (Linzer und Lewis 2011,7). Dies geschieht durch einen χ^2 -Test, dessen Teststatistik von `poLCA()` ausgegeben wird, und die man zusammen mit den Freiheitsgraden, die sich aus der Zahl der zu schätzenden Parameter ergeben und ebenfalls von der Funktion ausgegeben werden, in einen χ^2 -Rechner eingeben muss, um einen p-Wert zu erhalten. Dieser gibt darüber Aufschluss, ob die Nullhypothese, es gäbe keine Unterschiede zwischen Daten und Modell, oder die Alternativhypothese, Daten und Modell würden nicht zueinander passen, evident ist. Ein p-Wert kleiner als 0.05 würde darauf hindeuten, dass Daten und Modell nicht zueinander passen. Wir sind also an einem möglichst großen p-Wert interessiert. Damit ein Goodness-of-fit-Test durchgeführt werden kann, benötigt man mindestens so viele Proband/innen wie mögliche Antwortmuster, also „Anzahl der Abstufungen hoch Anzahl der Items“, in meinem Fall $4^4 = 256$ (vgl. Moosbrugger und Kaleva 2008, 307). Die benötigte Stichprobengröße würde bei 6 Abstufungen auf $6^4 = 1296$ anwachsen, weshalb man hier bes-

⁷ Bayesian Information Criterion $BIC = -2 \ln L + t \ln n$ und Akaike Information Criterion $AIC = -2 \ln L + 2 t$, für die Stichprobengröße n , Anzahl der zu schätzenden Parameter t , und Produkt über die berechneten Wahrscheinlichkeiten der möglichen Antwortmuster L .

⁸ Die Liste der Ausgabewerte findet sich bei Linzer und Lewis (2014) und auf <https://cran.r-project.org/web/packages/poLCA/poLCA.pdf> (letzter Zugriff am 25.04.2017). Die Werte mit Erklärungsbedarf werden im Laufe dieses Beitrags besprochen.

ser 4 als 6 wählt. Bei Stichproben unter 20, wenn mehr als ein Fünftel der (beobachteten oder berechneten) Häufigkeiten kleiner als 5 sind, oder berechnete Häufigkeiten kleiner als 1 auftreten, sind die χ^2 -Statistiken von `poLCA()` jedoch nutzlos, da ihre Werte einer χ^2 -Verteilung nicht genügen. Das war bei der Klassenlösung der Fall, weshalb ich auf eine Monte-Carlo-Variante des exakten Fisher-Tests zurückgriff. Dieser Test ist zwar nicht Teil des `poLCA`-Pakets, jedoch in den Standardpaketbibliotheken implementiert und ebenfalls gut dokumentiert.⁹ Zuvor muss man die tatsächlichen und die berechneten Häufigkeiten für alle Antwortmuster in eine Tabelle schreiben.¹⁰ Nullhypothese und Alternativhypothese bleiben unverändert. Die Funktion `fisher.test()`, angewandt auf diese Tabelle, lieferte einen p-Wert von .996, weshalb die Nullhypothese beibehalten wird. Das Modell passt also gut zu den Daten. Zum Schluss der statistischen Betrachtung seien noch die bereits erwähnten Zuordnungswahrscheinlichkeiten `poLCA.posterior` quantifiziert. Diese fallen für alle Klassen jeweils mit größer als .9 und im Schnitt mit .93 gut aus.

4 | Interpretation und Diskussion

Nach der Analyse latenter Klassen liegen 4 Klassen vor, die literaturdidaktisch nur dann von Interesse sind, wenn sie auch interpretiert werden können. Dies soll nun geschehen. Danach wird die Klassenlösung sowohl hinsichtlich ihrer Aussagekraft diskutiert als auch in mein Mixed-Methods-Projekt eingeordnet.

4.1 | Interpretation der Klassenlösung

Betrachtet man auf Abbildung 5 das Resultat der LCA, fällt auf, dass die von K-Means als parallel charakterisierten Klassen C1 und C2 sowie C3 und C4 insbesondere bezüglich des Items zur Lebensgeschichte ausdifferenziert wurden. Analog zum Vorgehen bei K-Means wird auf das Item „Leser“ verzichtet, da alle prototypischen Klassenvertreter hier „eher wichtig“ ankreuzen und es uns somit keinen Beitrag zur Differenzierung zwischen Klassen liefert.

⁹ <https://stat.ethz.ch/R-manual/R-devel/library/stats/html/fisher.test.html> (letzter Zugriff am 25.04.17)

¹⁰ Vor dem statistischen Test definiert man sich aus den absoluten Häufigkeiten und den berechneten Häufigkeiten eine zweispaltige Tabelle T. Die Grundidee des Tests besteht darin, dass für alle Tabellen mit gleichen Zeilen- und Spalten-Summen wie die eben definierte Tabelle T, für die die beiden Spalten sich mindestens so stark unterscheiden wie in T, die Wahrscheinlichkeit ihres Auftretens berechnet werden kann. Dies funktioniert nur unter der Annahme, die Spalten würden sich nicht systematisch voneinander unterscheiden. Wenn der erhaltene p-Wert als Summe der Auftretenswahrscheinlichkeiten dieser Tabellen inklusive T größer als .05 ist, kann man diese Annahme nicht verwerfen und hat keine systematischen Unterschiede zwischen beobachteten und berechneten Werten nachgewiesen. Da für große Tabellen wie T der Berechnungsaufwand eines exakten Fisher-Tests exorbitant ansteigt, muss man hier sogenannte Monte-Carlo-Methoden einsetzen, die T, statt mit konstruierten Tabellen, mit einer großen Anzahl von Zufallstabellen gleicher Spalten- und Zeilensumme vergleichen, um so den p-Wert abzuschätzen. Eine detaillierte Einführung anhand Fishers berühmter „tea tasting Lady“ findet sich in Kregel (2005,92).

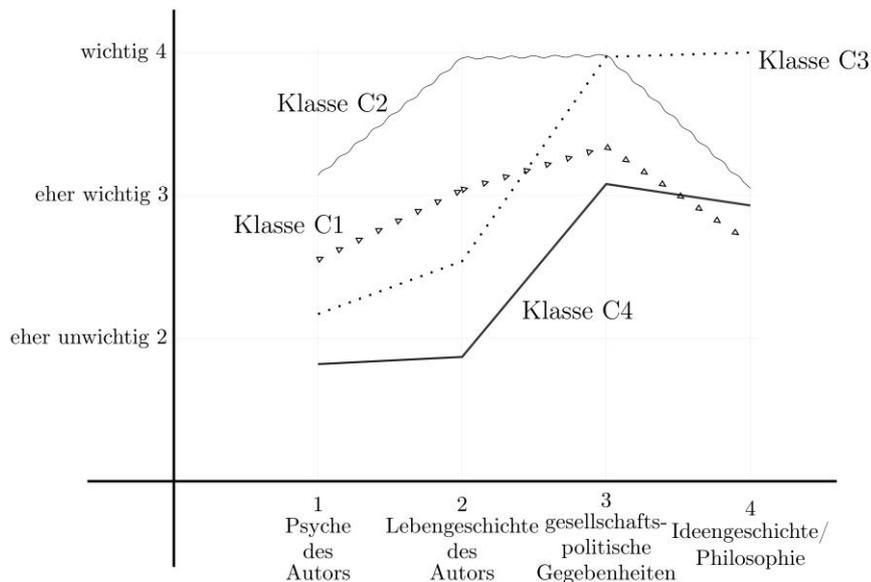


Abb. 5: Klassenlösung nach Anwendung von poLCA()

Diese Lösung lässt sich auf folgende Weise interpretieren: Die Klassen C1 und C2 stimmen allen Items tendenziell zu, während die Klassen C3 und C4 nur autorferne Kontexte favorisieren. Die Klasse C2 gewichtet konkrete, eventuell leichter handhabbare Kontexte wie die Lebensgeschichte des Autors oder realhistorische Gegebenheiten höher. Klasse C1 gibt ausschließlich realhistorischen Gegebenheiten leicht den Vorzug. Die Klasse C4 wertet die Bedeutung des Autors für das Interpretieren komplett ab. Für Klasse C3 gilt dies nur für die Psyche des Autors. Als deutschdidaktisches Resultat wären diese Erkenntnisse allein jedoch unbefriedigend. Beispielsweise stellt sich die Frage, an welchen Kriterien ein Vertreter oder eine Vertreterin der Klasse C1 festmacht, welche Kontexte zu wählen sind. Das können literaturdidaktisch plausible Gründe oder problematischere sein. Ähnliche Fragen stellen sich auch für die anderen Klassen. Dem ist qualitativ nachzugehen. Ich habe im qualitativen Untersuchungsteil in 22 Einzelinterviews Studierenden des Lehramts Deutsch diese Items mit der Bitte vorlegt, ihr Ankreuzverhalten zu begründen. Eine Probandin aus Klasse C1 sprach von Präferenzen „je nach Text und je nach Autor“, während eine andere alle Interpretationen und Kontexte für legitim hielt, die sich in „vernünftiger Sekundärliteratur“ finden lassen. Hohe Priorisierung der Bezugnahme auf den Autor inklusive seiner Psyche begründeten einige Proband/innen mit dem Verweis auf Autoren, für die solche Kontexte als fruchtbar erfahren wurden. Genannt wurden hier Kafka, Goethe, Büchner, Rilke und Rimbaud. Für Vertreter/innen der Klassen C3 und C4 waren das bloße Einzelfälle. Sie begründeten die Ablehnung jeglichen Autorbezugs mit der Unmöglichkeit, sich „in’s Hirn des Autors reinversetzen“ zu können, was auf ein verkürzendes Verständnis des intentional fallacy¹¹ hindeuten könnte. Bezüglich Vertreter/innen der Klasse C2 habe ich die Vermutung, dass von ihnen konkrete, leichter handhabbare Kontexte eher abstrakteren Konzepten vorgezogen werden. So sagte eine Probandin, dass sie beim Lesen nicht über „philosophische Sachen“ nachdenke, weil sie sich „da einfach nicht aus[kenne]“.¹² Dies soll nur einen Einblick in mögliche individuelle Begründungen geben. Weitere Details und Zusammenhänge zu anderen erhobenen Konzepten werden in meiner Dissertation ausführlich besprochen.

¹¹ Damit bezeichnet man in der Literaturtheorie des New Criticism die Vorstellung, die Autorintention wäre unzulässigerweise „available nor desirable as a standard for judging the success of a work of literary art“. Wenn der Autor seine Intention durch den literarischen Text, zum Beispiel ein Gedicht, vermittelt, „then the poem itself shows what he was trying to do. And if the poet did not succeed, then the poem is not adequate evidence“ (Wimsatt 1954, 3).

¹² Alle Zitate wurden von Studierenden geäußert, die kurz vor dem Examen stehen.

4.2 | Aussagekraft der Lösung

Die Möglichkeit der Weiterverwendung der nun interpretierten Klassenlösung ist davon abhängig, inwieweit man der gefundenen Klassenlösung trauen kann, was sich innerhalb und außerhalb des empirisch-quantitativen Paradigmas reflektieren lässt. Innerhalb dessen hatte ich Lösungen mit 2,3 und 5 Klassen ausgeschlossen und die Anpassungsgüte des Modells berechnet. Für Letzteres wurde gezeigt, dass man mit Hilfe der errechneten p-Werte die Nullhypothese, die besagt, dass die berechneten Häufigkeiten von den tatsächlich beobachteten Häufigkeiten nicht systematisch abweichen, nicht ablehnen konnte. Diese Argumentation liefert aber lediglich ein starkes Indiz und keinen Nachweis, da ich zwar den Fehler erster Art, also die Wahrscheinlichkeit der fälschlichen Zurückweisung der Nullhypothese, direkt kontrollierte, dies aber für den Fehler zweiter Art, also die fälschliche Annahme der Nullhypothese, so nur indirekt möglich ist (Bortz und Schuster 2010, 146). Dies geschieht, indem man das Signifikanzniveau, die forschenseitig „festgelegte Wahrscheinlichkeit, mit welcher die Ablehnung der Nullhypothese [...] zu einem Fehler 1. Art führt“ (ebd., 101), auf .25 erhöht (vgl. ebd., 146). Auch dem halten die berechneten p-Werte stand. Bei den berechneten Zuordnungswahrscheinlichkeiten ist jedoch Vorsicht geboten. Diese fallen zwar groß aus, mit Blick auf die Formeln, z.B. in Linzer und Lewis (2011), wird aber deutlich, dass sie nur aus Werten berechnet werden, die der Algorithmus asymptotisch ermittelt. Ein Rückbezug zu den tatsächlich vorliegenden Daten findet nicht statt. Darüber hinaus werden zwar nicht viele, aber dennoch einige Proband/innen mit geringer Wahrscheinlichkeit ihren Klassen zugewiesen, wie Abbildung 6 visualisiert. Wohl damit einher geht eine 80%-ige Übereinstimmung zwischen `kmeans()` und `poLCA()`. Es gibt also sehr wohl einige Proband/innen, die von beiden Algorithmen unterschiedlich zugeordnet wurden.

Es empfiehlt sich zum Schluss, noch einmal zwei Schritte zurückzutreten, um mit breiterer Perspektive einen Blick auf die Klassenanalyse zu werfen. Die mathematisch-deduktiven Herleitungen der Analyse latenter Klassen sind korrekt. Wie so oft bei formalen Argumentationen sind die Prämissen der Berechnung zu hinterfragen. So wird bei der Herleitung der Gleichungen die Wahrscheinlichkeit eines Antwortmusters unter der Bedingung der Annahme lokaler stochastischer Unabhängigkeit berechnet.

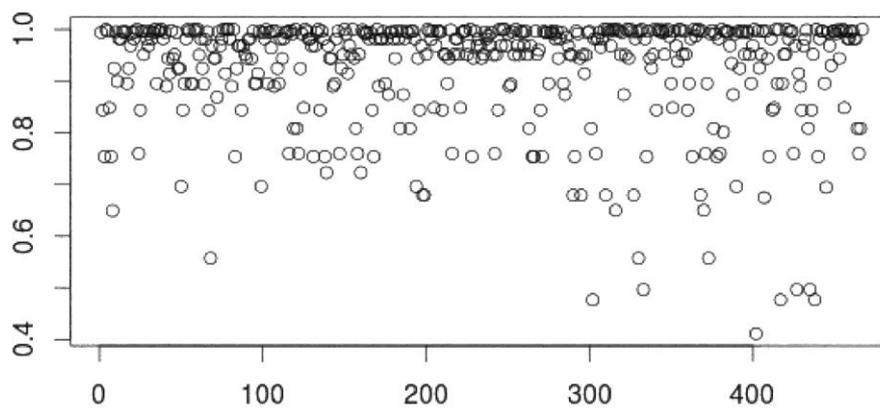


Abb. 6: Zuordnungswahrscheinlichkeiten aller 467 Proband/innen

Diese besagt, dass die Wahrscheinlichkeit, zwei Items auf eine bestimmte Weise anzukreuzen, nur von der Klassenzugehörigkeit abhängt und „nicht etwa von der Reihenfolge, in der die Items beantwortet wurden“ (Moosbrugger und Kaleva 2008, 301).¹³ In den Interviews

¹³ Die Definition lokaler stochastischer Unabhängigkeit entspricht strukturell der stochastischen Unabhängigkeit aus der Wahrscheinlichkeitstheorie (Krengel 2005, 25), wobei hier zusätzlich ein bedingendes Ereignis als latente Variable, im obigen Fall die Klassenzugehörigkeit, hinzukommt. Seien A und B die Ereignisse eines bestimmten Ankreuzverhaltens einer Person bezogen auf die Items a und b. Ferner bezeichne ξ das Ereignis einer bestimmten Klassenzugehörigkeit dieser Person. Lokale stochastische Unabhängigkeit bezüglich der Klassenzugehörigkeit und der Items liegt genau dann vor, wenn $P(A \cap B | \xi) = P(A | \xi)P(B | \xi)$ gilt (Moosbrugger und Kaleva 2008, 230), wobei $P(A | \xi)$ der Wahrscheinlichkeit

meinte ich jedoch zu beobachten, dass die Beantwortung der Items zur „Autorbiographie“ und „Psyche“ davon abhing, ob dazwischen zu den autorfernen Kontexten gesprungen wurde.¹⁴ Dies ist natürlich eine vage Hypothese, verdeutlicht aber die der Item-Response-Theorie notwendige Vereinfachung. Man würde im Fall stochastischer Abhängigkeit jene Interdependenzen zwischen den Items, die nicht durch die Klassenlösung erklärt werden, im Modell unberücksichtigt lassen müssen, weshalb ihre Nichtexistenz zur Prämisse wird. Noch bevor man die Klassenlösung interpretiert, ist sie also ein Artefakt, geht also nicht unabhängig vom bereits abstrahierenden Kalkül der LCA aus den Daten hervor. Die Verwendung zweier völlig verschieden verlaufender Verfahren kann hier etwas helfen. Die Ergebnisse bilden statt einer eindeutigen Partitionierung der Stichprobe dennoch lediglich Tendenzen ab.¹⁵ Ein Messen jener Merkmale, nach welchen klassifiziert wird, würde mit größeren Skalen pro Merkmal verlässlichere Ergebnisse liefern. Wenn große Skalen konstruierbar sind, ist ein solches Vorgehen natürlich zu bevorzugen. Dem prinzipiellen Charakter der Klassenlösung als bloßer Tendenz wirkt das wenig entgegen.

4.3 | Deutschdidaktische Funktionalisierung des Verfahrens

Im Rahmen meines Mixed-Methods-Projekts verwendete ich die Lösung entsprechend ihrer Eignung, also vor allem zur Generierung und Fokussierung von Hypothesen für die Interviewstudie. Dazu wurde mit einer großen Stichprobe eine Vielzahl von Hinweisen zur Möglichkeit literaturdidaktisch fruchtbarer Vertiefung untersucht. Durch die Größe der Stichprobe konnte zudem der Einfluss der besuchten Universität verringert werden, was gerade bei literaturtheoretischen Positionen wichtig ist. Im Falle der Kontextpräferenzen fragte ich mich also, ob hier diffuse, unscharfe oder klar konturierte und interpretierbare Klassen extrahiert werden können. Aus der in diesem Sinne erfolgreichen Klassenanalyse resultieren zwei naheliegende Anwendungsmöglichkeiten. Zum einen kann man die Klassenzugehörigkeit als Merkmal verstehen, dessen Zusammenhang mit anderen, ebenfalls mit Fragebögen erhobenen Merkmalen inferenzstatistisch überprüft werden kann. Dies ist jedoch nur bei Kontrolle der Anpassungsgüte zu empfehlen. Mit meinem Fragebogen konnte ich (schwache) signifikante Zusammenhänge zwischen den Kontextpräferenzen und Überzeugungen zur Wissenschaftlichkeit interpretatorischer Prozesse finden. Weitere deutschdidaktische Anwendungen dieses Vorgehens, also beispielsweise die Frage nach dem Zusammenhang verschiedener Orientierungen und Überzeugungen wie Aufgaben- oder Methodenpräferenzen, sind gerade innerhalb der Professionalisierungsforschung vorgelegt worden (Winkler 2011) oder denkbar. Darüber hinaus könnte man beispielsweise in der Aufgabenwirkungsforschung Fehler, Lernstände oder Verstehensniveaus von Schüler/innen klassifizieren und in ihrem Zusammenhang mit anderen personenbezogenen Merkmalen untersuchen. Zum anderen ist wie in meiner Studie die Verwendung der Klassenanalyse innerhalb eines Mixed-Methods-Ansatzes möglich. Bei vorgelagerten quantitativen Untersuchungsteilen könnte man zum Beispiel die Interviewproband/innen nach ihrer Klassenzugehörigkeit auswählen, um dann in Interviews Begründungszusammenhänge oder gar die biographisch eventuell erklärbare Genese der Klassenzugehörigkeit zu rekonstruieren. Ferner kann die Klassenlösung mit Hilfe der qualitativen Interviews kritisch geprüft werden. Beispielsweise war der Frage nachzugehen, ob nicht die Unterscheidung „Ablehnung des Autorbezugs“ versus „keine Ablehnung des Autorbezugs“ differenziert genug ist, was zum jetzigen Stand der Auswertung der Interviewdaten jedoch verneint werden kann. Wie bereits unter 4.1 dargelegt, passten in meinem Fall jene Begründungen der Proband/innen gut zur im Vorfeld der Interviews formulierten Interpretation der Klassenlösung. Abschließend bleibt die Frage zu klären, ob sich die Mühe der selbst durchge-

von A bei gegebenen ξ entspricht (vgl. Krengel 2005, 21). Abhängigkeit von der Reihenfolge würde jedoch auf eine von ξ nicht erklärte Abhängigkeit beider Items voneinander hindeuten.

¹⁴ Da ich in dieser Interviewphase keine Zwischenfragen stellte, schließe ich Beeinflussung durch mich an dieser Stelle aus.

¹⁵ Zur Sicherheit habe ich überprüft, ob häufig auftretende Antwortmuster durch diese Tendenzen erfasst werden, was der Fall ist.

fürten Klassenanalyse lohnte. Es zeigte sich, dass die Einarbeitung in die Verfahrensweise eine detaillierte Einschätzung der Aussagekraft der gefundenen Lösung ermöglichte. Da diese auch selbst sofort literaturdidaktisch interpretiert werden konnte, war man so hochgradig flexibel und konnte mit den Werten auf verschiedene Arten weiterrechnen und inferenzstatistische Tests anschließen. Dafür erwies sich gerade die Berechnung in R als vorteilhaft. Das Auffinden von Abhängigkeiten zwischen dem mathematisch-kombinatorischen Aufwand der Auswertungen und der Gestaltung des Fragebogens bewahrte vor bösen Überraschungen. So konnte vorab geklärt werden, wie viele Items und wie viele Abstufungen möglich sowie, davon abhängig, wie viele Proband/innen nötig waren. Mit Kenntnis der Verfahren behält man also von Anfang an das große Ganze im Blick, ohne den Rückbezug zur deutschdidaktischen Fragestellung zu vergessen. Die selbstständige Beschäftigung mit diesen recht elaborierten statistischen Algorithmen lieferte darüber hinaus wertvolle verallgemeinerbare Erfahrungswerte, die auch bei anderen statistischen Tests und Verfahren Anwendung finden, wie beispielsweise zum Umgang mit der Anpassungsgüte von Modellen oder der Maximum-Likelihood-Schätzung von Parametern. Diese Erfahrungswerte sind natürlich auch für das Verständnis und die Einschätzung fremder Publikationen, gerade aus dem Bereich der Bildungsforschung, hilfreich.

Literatur

- Bortz, J. / Schuster, C. (2010): Statistik für Human- und Sozialwissenschaftler. Lehrbuch mit Online-Materialien. 7. Aufl. Berlin: Springer.
- Danneberg, L. (2007): Kontext. In: G. Braungart, H. Fricke, K. Grubmüller, J. Müller, F. Vollhardt, K. Weimar (Hg.): Reallexikon der deutschen Literaturwissenschaft. Bd. II. Berlin: De Gruyter. S. 333–336.
- Eco, U. (2008): Streit der Interpretationen. In: T. Kindt, T. Köppe (Hg.): Moderne Interpretationstheorien. Ein Reader. Göttingen: Vandenhoeck & Ruprecht. S. 102–129. – 1. Auflage 1987
- Fricke, H. (1992): METHODEN? PRÄMISSEN? ARGUMENTATIONSWEISEN! Überlegungen zur Konkurrenz wissenschaftlicher Standards in der Literaturwissenschaft. In: L. Danneberg, F. Vollhardt, H. Böhme, J. Schönert (Hg.): Vom Umgang mit Literatur und Literaturgeschichte. Positionen und Perspektiven nach der Theorie-debatte. Stuttgart: Metzler. S. 211–227.
- Hemerén, G. (2008): Interpretation. Typen und Kriterien. In: T. Kindt, T. Köppe (Hg.): Moderne Interpretationstheorien. Göttingen: Vandenhoeck & Ruprecht, S. 248–276. – 1. Auflage 1983
- Hofer, B. K. (2004): Exploring the dimensions of the personal epistemology in differing classroom contexts: Student interpretations during the first year of college. In: Contemporary Educational Psychology, 29, S. 129–163.
- Kabacoff, Robert I. (2010): R in Action. Data Analysis and Graphics with R. New York: Manning.
- Kindt, T. / Köppe, T. (Hg.) (2008): Moderne Interpretationstheorien. Ein Reader. Göttingen: Vandenhoeck & Ruprecht.
- Krengel, U. (2005): Einführung in die Wahrscheinlichkeitstheorie und Statistik. Braunschweig: Vieweg und Sohn.
- Kuckartz, U. (2014): Mixed Methods: Methodologie, Forschungsdesigns und Analyseverfahren. Wiesbaden: Springer.
- Kunter, M. / Baumert, J. / Blum, W. (2011): Professionelle Kompetenz von Lehrkräften: Ergebnisse des Forschungsprogramms COACTIV. Münster: Waxmann.
- Linzer, D. A. / Lewis, J. B. (2011): poLCA. An R Package for Polytomous Variable Latent Class Analysis. In: Journal of Statistical Software, Vol. 42, 10, S. 1–29. – Online unter: <<http://www.jstatsoft.org/v42/i10/>> (letzter Zugriff: 25.04.2017)
- Linzer, D. A. / Lewis, J. B. (2014): Polytomous variable Latent Class Analysis. Version 1.4.1, Online unter: <<http://dlinzer.github.com/poLCA>> (letzter Zugriff: 25.04.2017)
- Moosbrugger, H. / Kelava, A. (2008): Testtheorie und Fragebogenkonstruktion. Berlin: Springer.
- Priemer, B. (2006): Deutschsprachige Verfahren zur Erfassung epistemologischer Überzeugungen. In: Zeitschrift für Didaktik der Naturwissenschaften, 12, S. 159-175.
- Törner, G. (2005): Epistemologische Beliefs. State-of-the-Art-Bemerkungen zu einem aktuellen mathematikdidaktischen Forschungsthema vor dem Hintergrund der Schraw-Olafson-Debatte. In: H.-W. Henn, G. Kaiser (Hg.): Mathematik im Spannungsfeld von Evolution und Evaluation. Festschrift für Werner Blum. Hildesheim: Div-Verlag Franzbecker. S. 308-323.
- Urhahne, D. / Hopf, M. (2004): Epistemologische Überzeugungen in den Naturwissenschaften und ihre Zusammenhänge mit Motivation, Selbstkonzept und Lernstrategien. In: Zeitschrift für Didaktik der Naturwissenschaften, 10, S. 71–87.
- W. K. Wimsatt / Monroe B. (1954): The Intentional Fallacy. In: Wimsatt, W. K. (Hrsg.): The Verbal Icon. Lexington, KY: University of Kentucky Press.– URL <http://faculty.smu.edu/nschwartz/seminar/fallacy.htm> (letzter Zugriff: 25.04.2017)
- Wieser, D. (2013) Was macht einen literarischen Experten aus? Epistemologische Konzepte von Schülerinnen und Schülern. In: C. Rieckmann, J. Gahn (Hg.): Poesie verstehen – Literatur unterrichten. Baltmannsweiler: Schneider Hohengehren. S. 159–178.
- Winkler, I. (2011): Aufgabenpräferenzen für den Literaturunterricht – Eine Erhebung unter Deutschlehrkräften. Berlin: Springer.
- Zabka, T. (2007): Diskursive und poetische Aufgaben zur Texterschließung. In: H. Willenberg (Hg.): Kompetenzhandbuch für den Deutschunterricht. Baltmannsweiler: Schneider Hohengehren. S. 199–209.
- Zabka, T. (2015a): Was ist Hochschulreife im Umgang mit Literatur? In: Didaktik Deutsch, 38, S. 136–150.
- Zabka, T. (2015b): Was sind Methoden der Literaturinterpretation im Unterricht und wie kann man sie untersuchen? In: H. Jonas, M. Kreisel (Hg.): Fachdidaktik Deutsch – Rückblicke und Ausblicke. Frankfurt a. M.: Lang. S. 215–236.

Marco Magirus

Wissenschaftlicher Mitarbeiter
Fakultät für Erziehungswissenschaft
Didaktik der sprachlichen und ästhetischen Fächer
Universität Hamburg
marco.magirus@uni-hamburg.de