

## **Quantitative und qualitative Zugänge in der Gesprächsforschung – Methodologische Betrachtungen am Beispiel einer Studie zu argumentativen Gruppendiskussionen**

**Vera Mundwiler / Judith Kreuz / Daniel Müller-Feldmeth  
Martin Luginbühl / Stefan Hauser**

### *Abstract*

Dieser Beitrag diskutiert grundlegende und angewandte Aspekte des Verhältnisses qualitativer und quantitativer Zugänge in der Gesprächsforschung. Ausgehend von der Beobachtung, dass auch zahlreiche Arbeiten der klassischen Konversationsanalyse informale Codierungen und informale Quantifizierungen vornehmen, werden Merkmale des formalen Quantifizierens erläutert.

Im Hauptteil des Beitrags werden solche formalen, aber immer auf gesprächsanalytischen Kategorien basierenden Quantifizierungen an einem Korpus von 180 Gesprächen Deutschschweizer Schulkinder vorgenommen, wobei die vorangehende Erarbeitung eines Codierschemas im Detail diskutiert wird. Es handelt sich bei den Daten um Einigungsdiskussionen von Grundschulkindern. Ziel der Quantifizierungen ist es, verlässliche Aussagen über den Zusammenhang von interaktiven, argumentationsspezifischen Praktiken (Modalisierung, Perspektivierung, Argumentationskomplexität, Interaktivität) und anderen Variablen (Alter, Gesprächssetting, Geschlecht) zu erarbeiten und so repräsentative Aussagen über Teilaspekte argumentativer Gesprächskompetenz zu ermöglichen. Zudem werden verschiedene Visualisierungen von Gesprächen diskutiert, welche auf den Quantifizierungen beruhen und individuelle Differenzen zwischen den Interagierenden besonders gut sichtbar machen.

*Keywords:* Konversationsanalyse, Integration von qualitativen und quantitativen Zugängen, argumentative Gesprächskompetenz, formale Quantifizierung, Codierung, Visualisierung von Gesprächsdaten, Modalisierung, Perspektivierung, Argumentationskomplexität, Interaktivität.

### *English Abstract*

This paper discusses fundamental and applied aspects of the relationship between qualitative and quantitative approaches in conversation analysis. Starting from the observation that informal coding and informal quantification can be found in numerous works of classical CA, features of formal quantification are discussed.

In the main part of the paper, such formal quantifications (based on conversation analytic categories) are carried out on a corpus of 180 conversations of Swiss German schoolchildren, whereby the preceding elaboration of a coding scheme is discussed in detail. The data consist of discussions between elementary schoolchildren. The aim of the quantifications is to develop reliable statements on the relationship between interactive, argumentative practices (modalization, perspectivation, argumentative complexity, interactivity) and other variables (age, setting, gender) and thus on crucial aspects of oral argumentative skills. In addition, we discuss various visualizations based on quantified data, which make individual differences between interactants particularly visible.

*Keywords:* Conversation Analysis (CA), integration of qualitative and quantitative approaches, oral argumentative skills, formal quantification, coding, visualization of conversation data, modalization, perspectivization, argumentative complexity, interactivity.

1. Einleitung: Zur Integration quantitativer Analysen in die Gesprächsanalyse
2. Daten
  - 2.1. Forschungsgegenstand: Argumentieren in Gruppengesprächen
  - 2.2. Studiendesign und Datenerhebung
  - 2.3. Korpus
  - 2.4. Computergestützte Verfahren der Datenaufbereitung
3. Methodisches Vorgehen bei der Datenanalyse
  - 3.1. Theoriebasierte und qualitative Annäherung an die Daten
  - 3.2. Quantitative Annäherung an die Daten: das Codieren
  - 3.3. Normalisierung mündlicher Daten
4. Quantitative Auswertungen und Ergebnisse
  - 4.1. Fragestellungen und Auswertungen über das gesamte Korpus
    - 4.1.1. Modalisierung und Perspektivierung
    - 4.1.2. Komplexität von Argumentationen
    - 4.1.3. Interaktivität
  - 4.2. Visualisierung und Auswertung einzelner Gespräche
    - 4.2.1. Argumentative und thematische Sprecheranteile
    - 4.2.2. Thematische Cluster
5. Diskussion zur Kombination qualitativer und quantitativer Methoden
6. Danksagung
7. Literaturverzeichnis

## **1. Einleitung: Zur Integration quantitativer Analysen in die Gesprächsanalyse**

Zu den Zielen der Gesprächsanalyse gehört gemäß Sidnell (2013:77) "the discovery of previously unknown regularities of human interaction" und "to identify and describe the recurrent practices of interaction". Methodisch steht dabei zunächst die beobachtende Turn-by-Turn-Analyse, die Mikroanalyse der Form von Äußerungen sowie deren sequenzielle Positionierung anhand transkribierter Gespräche im Zentrum, aufgrund derer in Gesprächen hervorgebrachte Strukturen (wiederkehrende Praktiken oder aber abweichende Fälle) identifiziert und in Kollektionen gesammelt werden; darauf basierend können Praktiken beschrieben werden, welche sich auf einzelne Turn-Komponenten, auf ganze Turns oder Turnsequenzen beziehen können (dazu ebd.; Sacks 1992; Garfinkel 1967; Schegloff 2007). Dieses Vorgehen der 'klassischen' Gesprächsanalyse kann insofern als qualitativ angesehen werden, als es explorativ (es werden keine vorher festgelegten Praktiken überprüft) und hypothesengenerierend (Praktiken identifizieren) ist. Der Fokus liegt hier auf der Frage,

wie in der Interaktion (und das kann auch nur ein einziger konkreter Fall sein) gemeinsam eine geordnete (und ordnende) Struktur hervorgebracht wird. Die jeweils wirksamen Prinzipien können dann u.U. generalisiert werden. Hierfür ist der Vergleich mit anderen Daten innerhalb einer Kollektion notwendig, wobei dieser Vergleich qualitativ bleiben kann und nicht quantifizierend erfolgen muss (Sacks 1992; ten Have 2007).

Für die Erstellung von Kollektionen (seien diese nun basierend auf einer bestimmten Form, einer bestimmten Handlung oder einer bestimmten sequenziellen Position) müssen die einzelnen Vorkommen allerdings immer als Realisierung einer bestimmten Praktik identifiziert werden, um in eine Kollektion aufgenommen zu werden. Dasselbe gilt dann innerhalb einer Kollektion für Sub-Praktiken, die ausdifferenziert werden (z. B. die handlungs- und sequenzbezogene Ausdifferenzierung einer bestimmten Form). Diese Identifizierung kommt einer Codierung gleich; Steensig/Heinemann (2015:22) sprechen in diesem Zusammenhang von einem "informal coding", das in der Gesprächsanalyse schon lange praktiziert werde. Stivers (2015:5) schreibt dazu:

While CA does not formalize this coding process, it is a standard component of CA research that all cases in the collection should be accounted for in terms of matching the analysis to the various subtypes of the practice, a feature that can easily be used as the basis for formal coding. Thus, CA's insistence on clear characterizations of the phenomena being studied creates a solid foundation from which to build formal coding schemes.

Viele Arbeiten der Konversationsanalyse in ihrer klassischen Ausprägung nehmen jedoch nicht nur informale Codierungen vor, sondern auch informale Quantifizierungen. Dies zeigt sich in Charakterisierungen einzelner Praktiken mit Begriffen wie *oft*, *in der Regel*, *häufig* oder *selten* (z. B.: "Die präferierte Form [...] erfolgt *in der Regel* kurz und einfach und enthält *häufig* vorgeformte Elemente.", Gülich/Mondada 2008:52, Hervorhebung von uns) bzw. *massively*, *ordinarily* oder *occasionally*. Die drei englischsprachigen Beispiele stammen von Schegloff (1993: 118),<sup>1</sup> der in diesem Zusammenhang darauf verweist, dass wir es hier mit einer gängigen "informal 'quantification'" (ebd.) zu tun haben, welche

reports an *experience* or *grasp* of frequency, not a count; an account of an investigator's sense of frequency over the range of a research experience, not in a specifically bounded body of data; a characterization of distribution fully though tacitly informed by the analytic import of what is being characterized. (ebd.:119, Hervorhebungen im Original)

Dem von Schegloff als *informalen* Quantifizieren bezeichneten Vorgehen stehen Verfahrensweisen des *formalen* Quantifizierens gegenüber. In diesem formalen quantitativen Vorgehen stehen mathematisch fundierte Zählprozesse und statistische Auswertungsverfahren im Vordergrund, wobei es sich nicht um bessere Versionen von *informal quantifications* handelt, sondern um eine andere *Art* von Quantifizierung. Dennoch stellt auch in qualitativen Studien die Distribution der Phänomene oft einen wichtigen Aspekt gesprächsanalytischer Fragestellungen dar. Zwar lässt die allgemeine Frequenz einer Praktik keine Aussagen darüber zu, ob sich die Gesprächsteilnehmenden im Einzelfall wirklich präferiert an dieser Praktik orien-

---

<sup>1</sup> Ähnliche Beispiele mit Referenzangaben in Stivers (2015:6).

tieren (vgl. Stivers 2015:8), aber sie ist ein Indiz dafür, dass die entsprechende Praktik einer sozialen Norm entspricht und die Kenntnis dieser Praktik Teil des pragmatischen Sprachwissens einer kommunikativen Gemeinschaft ist; zudem zeigt das Abweichen von dieser Praktik eine bestimmte Handlung an. Dabei handelt es sich um ein Phänomen, das vor dem Hintergrund der Existenz einer sozialen Norm funktioniert (vgl. ebd.:7).

Die in den folgenden Kapiteln dargelegten Überlegungen zum Verhältnis qualitativer und quantitativer Zugänge in der Gesprächsforschung sind im Kontext eines Forschungsprojekts zu mündlicher Argumentationskompetenz von Schulkindern im Alter von 7-12 Jahren entstanden.<sup>2</sup> Hintergrund der Studie war unter anderem, dass entsprechende Kompetenzen in der Schule zwar vermittelt und bewertet werden müssen – dies sehen die entsprechenden Lehrpläne vor –, gleichzeitig aber ein großes Forschungsdesiderat dahingehend besteht, was in diesem Bereich in den einzelnen Klassenstufen erwartet werden kann. Zu den Herausforderungen dieses Projekts gehörte es, qualitative (und das heißt: auch an Einzelfällen orientierte) mit quantitativen (etwa auf bestimmte Kohorten bezogenen) Fragestellungen zu verbinden. Denn: Aussagen zum Erwerb argumentativer Gesprächskompetenz haben (nur) dann einen heuristischen Wert, wenn die situational-pragmatischen Spezifika einzelner Interaktionsereignisse ebenso Berücksichtigung finden wie Beobachtungen, die von Besonderheiten des Einzelfalls absehen.

Beim erwähnten Projekt kommt noch eine weitere – damit zusammenhängende – methodologische Problematik hinzu: Kompetenzbezogene Untersuchungen müssen in einem gesprächsanalytischen Kontext immer auf die Beschreibung dessen abzielen, was Deppermann (2004:20) als "faktische Kompetenz" beschrieben hat: Das kognitiv verankerte, individuelle Potenzial einer Person (Kompetenz) kann gesprächsanalytisch nur an ihrem tatsächlichen Handeln (Performanz) untersucht werden; letzteres etabliert die faktische Kompetenz. Kompetente SprecherInnen können globale und lokale Anforderungen einer aktuellen Gesprächssituation bewältigen, d.h. diese Situation einschätzen, angemessen reagieren, diese Reaktion angemessen zum Ausdruck bringen etc. (zu verschiedenen Aspekten einer allgemeinen Gesprächskompetenz siehe Hartung 2004; Grundler 2008; Becker-Mrotzek 2009; Quasthoff 2009). Gesprächskompetenz kann dann verstanden werden als "Anteil eines der Beteiligten an der gemeinsamen Herstellung von Sinn im Rahmen der jeweiligen Gesprächsaktivität" (Quasthoff 2009:86).

Eine Verbindung von qualitativen und quantitativen Analysemethoden (erstmalig in der Gesprächsanalyse realisiert von Heritage/Greatbatch 1986) erwies sich im Rahmen des Projekts als notwendig, als es uns nicht um die Beschreibung von Kompetenzen einzelner Schülerinnen oder Schüler ging,<sup>3</sup> sondern darum, empirisch basierte Kompetenzniveaus in Bezug auf mündliches Argumentieren im Gespräch zu beschreiben. Es ging also darum zu erfassen, welche argumentativ-mündlichen, interaktiven Prozessierungen die Kinder in den einzelnen Klassenstufen im Kontext von schulischen Übungen zu Argumentationsgesprächen typischerweise

<sup>2</sup> Das vom Schweizer Nationalfonds finanzierte Projekt "Argumentative Gesprächskompetenz in der Schule: Kontexte, Anforderungen, Erwerbsverläufe" (Laufzeit: 2014-2018) stand unter der Leitung von Brigit Eriksson und Martin Luginbühl.

<sup>3</sup> Dazu müsse man u.E. dieselben Personen wiederholt und in grundsätzlich verschiedenen Situationen beobachten.

realisieren, wobei wir einen Fokus auf Begründungshandlungen gelegt haben, welche in unserem Verständnis den Kern von und eine Bedingung für Argumentationen darstellen (zum Begründen in mündlichen Argumentationen siehe Gohl 2006; Büker 2008; Grundler 2011:7-47; Bova/Arcidiacono 2013). Zudem interessierte die Frage, ob es neben dem Schulalter (also der Klassenstufe) noch andere, u.U. wichtigere Variablen gibt, die sich als relevant erweisen.<sup>4</sup> Die im Folgenden diskutierten quantifizierenden Herangehensweisen sind u.E. unumgänglich, weil es bei der großen, auch in den einzelnen Klassenstufen sehr heterogenen Datenmenge (180 Gespräche mit insgesamt 720 Kindern) unmöglich wird, einen verlässlichen 'Eindruck' ("*experience or grasp*", Schegloff 1993:119) zu erhalten und die Einzelfälle im Hinblick auf einzelne Variablen (*Klassenstufe, Alter, Geschlecht* etc.) zu generalisieren. Verlässliche Aussagen über den Zusammenhang von interaktiven Praktiken und anderen Variablen sowie differenzierte Aussagen über die Frequenzen solcher Praktiken sind u.E. in einer großen Datenmenge (welche wiederum Voraussetzung für die Beschreibung von empirisch basierten Kompetenzniveaus ist) nur möglich, wenn qualitative Analysen für die Identifizierung von relevanten Praktiken mit quantitativen Analysen verbunden werden.

Dabei ist es allerdings zentral, dass zunächst die relevanten Praktiken durch qualitative Analysen identifiziert werden (dazu auch Deppermann 2008:107), also gesprächsanalytisch basiert sind und nicht normativ bzw. theoriegeleitet gesetzt werden, wie es in einigen Arbeiten zur argumentativen Gesprächskompetenz der Fall ist. Dies bedeutet auch, dass diese Praktiken nicht nur Oberflächen-Formen berücksichtigen, sondern auch die durch Komposition, sequenzielle Anordnung etc. angezeigte kommunikative Funktion (dazu grundlegend Heritage 1984 oder Selting/Couper-Kuhlen 2001). Damit bleibt die Analyse den Grundsätzen der Gesprächsanalyse treu. Allerdings erfordert eine handhabbare Codierung breite, inhaltlich reduzierte Kategorien, um eindeutige Codierungen ohne großen zeitlichen Aufwand für den Einzelfall vornehmen zu können; dies bedeutet auch, dass das Problem der "hard boundaries" (Stivers 2015:13) bestehen bleibt. Dies muss bei der Interpretation der quantitativen Befunde immer mitbedacht werden. Dennoch ermöglicht die hier skizzierte Verbindung von qualitativer und quantitativer Analyse Erkenntnisse, die allein mit genuin gesprächsanalytischen Methoden nicht möglich wären. Im Folgenden stehen unsere methodischen Überlegungen im Zentrum, nicht die Beschreibung unterschiedlicher Kompetenzniveaus.

## 2. Daten

Im Folgenden werden wir die methodischen Überlegungen aufzeigen, die bei den verschiedenen Schritten vom Studiendesign über die Datenerhebung, die Komposition des Korpus sowie die Datenaufbereitung entscheidend waren. Bevor aber genauer auf die hier zentral diskutierten methodischen Fragen eingegangen wird, soll kurz unser Forschungsinteresse – das mündliche Argumentieren – expliziert werden.

---

<sup>4</sup> Stivers (2015:8) nennt die Analyse des Zusammenhangs von Praktiken der sozialen Interaktion und anderer Variablen als einen der Hauptgründe für die Integration formaler Codierungen in die Gesprächsanalyse.

## 2.1. Forschungsgegenstand: Argumentieren in Gruppengesprächen

Die Studie (vgl. Hauser/Luginbühl 2017) hatte zum Ziel, Aussagen über unterschiedliche Ausprägungen argumentativer Gesprächskompetenz von Grundschulkindern treffen zu können und hierfür wurde ein Setting gewählt, in welchem den untersuchten Kindern eine Aufgabe gestellt wurde, die eine argumentative Prozessierung hochgradig erwartbar machte, den Kindern bezüglich Gesprächshandlungen aber keine einschränkenden Anweisungen gab (siehe dazu genauer 2.2). Da unser Verständnis von mündlichem Argumentieren maßgeblich Einfluss hatte auf die Datenerhebung, die Datenaufbereitung, die Entwicklung des Codierschemas sowie die Auswahl der Analysefragen, soll zunächst kurz umrissen werden, was unter dem mündlichen Argumentieren im Gespräch verstanden wird.

Grundsätzlich gehen wir in Bezug auf das mündliche Argumentieren von folgenden Annahmen aus:

- **Situiertheit:** Mündliche Argumentationen sind komplexe Diskurseinheiten, die als übersatzmäßige und abgrenzbare Einheiten / kommunikative Praktik in "interaktive und pragmatische Kontexte des Miteinanderhandelns" eingebettet sind (Morek 2017:70; vgl. auch Deppermann 2006).
- **Interaktivität:** Bei der Herstellung sozialer Kontexte, hier bei der Generierung und Prozessierung argumentativer Handlungen und Positionen, sind in spontanen Gruppendiskussionen mindestens zwei Interagierende beteiligt, die sich mit ihren Äußerungen aufeinander beziehen und in Gemeinsamkeit Argumentationen entfalten (Ko-Konstruktion) (vgl. z. B. Grundler 2011, 2015).
- **"Kernjob" Begründen:** Hauptmerkmal von Argumentationen sind Begründungshandlungen (Heller 2012).

In Anlehnung an einschlägige empirische Arbeiten verstehen wir unter argumentativen Diskurseinheiten global dimensionierte, sequenziell strukturierte und interaktiv hervorgebrachte (primär) sprachliche Handlungen, in denen verbale Begründungsaktivitäten realisiert werden, um differierende Geltungsansprüche zu bearbeiten (vgl. Quasthoff/Krah 2012; Grundler 2011; Heller 2012; Arendt 2015:21; Morek 2015). Obgleich davon ausgegangen wird, dass die kontroverse, "geltungskritische [...] Bezugnahme auf eine These [...] Kern des Argumentierens" ist (Deppermann/Lucius-Hoene 2006:142), muss nicht immer ein strittiger Fall Auslöser einer argumentativen Handlung sein, sondern es können auch Problematisierungen eines Sachverhalts, "Perspektivendivergenzen" (Rühl 1999:11) oder offene Fragestellungen, wie im hier beschriebenen Setting (vgl. Kapitel 2.2.), vorliegen.

Wir rechnen deshalb mit der Möglichkeit, dass sich Begründungshandlungen "nicht allein in einer kontroversen Veränderung von Wissen, sondern auch im gemeinsamen, kooperativen Aufbau von Wissen verorten" lassen (Scarvaglieri 2017: 146). Kotthoff (2015:83) verwendet für derart gemeinsames Abwägen des "Für und Wider einer sich entwickelnden Einschätzung" (Kotthoff 2015:83) den Begriff *Räsonieren*. Mercer (2009:184) spricht von *exploratory talk* und fasst darunter Sprachhandlungen, in denen "partners engage critically but constructively with each other's ideas". Ähnlich finden wir bei Ehlich (2014:46) den Begriff *exploratives Argumentieren* und die Beobachtung, dass die Wissenssysteme der Aktanten nicht zwangsläufig im Konflikt miteinander stehen, sondern lediglich verallgemeinerbares Wissen gemeinsam hergestellt und prozessiert wird. Argumentationen bewegen

sich abhängig von den Bedingungen des kommunikativen Rahmens also innerhalb von unterschiedlichen Stärkegraden der kognitiven Divergenz der Teilnehmenden und damit ihres konfligierenden Potenzials auf einem Kontinuum von konsensuellem und adversativem Argumentieren (vgl. dazu Cohen/Stevens 2017; vgl. auch Bose/Hannken-Illjes 2019).

Auch bei der Entwicklung des Codierschemas, auf das in Kapitel 4.1. noch ausführlich eingegangen wird, orientierten wir uns in Bezug auf Begründungshandlungen an der Sequenzialität und der interaktiven Einbettung von argumentativen Handlungen, d.h., es werden zwar isolierte Einzelhandlungen codiert, aber durch Codes abgebildet, die dem argumentativen Prozess/der Interaktivität gerecht werden und verschiedene interaktive Bezüge argumentativer Handlungen (z. B. 'zustimmend', 'ablehnend') zueinander zeigen (siehe ausführlich in Kapitel 4.1.). Damit haben die Codes ihr Fundament in den situieren Gesprächshandlungen und deren sequenziellen Prozessierung.

## 2.2. Studiendesign und Datenerhebung

Für die Erhebung der Daten waren gesprächsanalytische Prämissen und Anforderungen an die Quantifizierung und Vergleichbarkeit maßgebend. So wurde beispielsweise bewusst darauf verzichtet, in den beteiligten Schulklassen vorbereitende Unterlagen (z. B. mit möglichen Positionen und Argumenten) zu verwenden und es gab auch weder von den Lehrpersonen noch von den Projektmitarbeitenden Interventionen zum mündlichen Argumentieren. Die SchülerInnen wurden kurz instruiert und konnten dann in der Gruppe mit der Bearbeitung einer für das Schulfeld typischen Aufgabe beginnen. Auf eine vorgängige Intervention zum Argumentieren wurde verzichtet, da es nicht unsere Absicht war, Argumentationsfähigkeiten zu schulen und dann das entsprechende Wissen zum Argumentieren abzurufen. Vielmehr richtete sich unser Interesse auf die Frage, wie Kinder in unterschiedlichen Altersgruppen eine Einigungsaufgabe lösen, die das mündliche Argumentieren zwar nicht voraussetzt, aber hochgradig erwartbar macht. Demnach war es unser Anspruch, eine passende Aufgabe zu finden, die die Kinder zum Diskutieren animiert, die gleichzeitig für Kinder unterschiedlichen Alters von der 2. bis 6. Klasse der Grundschule (der Schweizer Primarschule) ansprechend ist (bzgl. Schwierigkeit, Verständlichkeit und Attraktivität der Aufgabe, möglichst mit Bezug zur kindlichen Alltagswelt), die vergleichbare und auswertbare Daten generiert und Störfaktoren gering hält.

Zu diesem Zweck wurden zwei Aufgaben entwickelt, die jeweils mehrere Wahlmöglichkeiten beinhalteten und eine Einigung zum Ziel hatten: Die 'Robinsonaufgabe' und die 'Spendenaufgabe'. Bei der Robinsonaufgabe sollten sich die SchülerInnen vorstellen, sie seien auf einer einsamen Insel gestrandet und hätten die Möglichkeit, mit einer Tauchausrüstung einmal zum Meeresgrund zu tauchen und drei gesunkene Gegenstände heraufzuholen, um das Überleben auf der Insel sicherzustellen. Hierfür erhielten sie ein Aufgabenblatt mit 12 abgebildeten Gegenständen (Taschenmesser [im Schweizer Standard 'Sackmesser'], Buschmesser, Schlafsack, Wolldecke, Zelt, Moskitonetz, Streichhölzer, Feuerzeug, Kochtopf, Leuchtraketen, Handy, Verbandszeug), aus denen sie gemeinsam drei auswählen mussten. Implizit

wird bei der Aufgabe eine Prüfung der Gegenstände, bezogen auf deren Nützlichkeit, nahegelegt. Absichtlich wurden deshalb auch einige in Bezug auf den Nutzen ähnliche Objekte in die Liste aufgenommen.

Bei der Spendenaufgabe wurden den SchülerInnen vier real existierende Möglichkeiten aufgezeigt, für die man Geld spenden könnte: für ein Projekt zum Schutz der Braunbären in der Schweiz (WWF), für eines zum Schutz der Bienen (WWF), für Kinder aus finanziell benachteiligten Familien (Stiftung Feriengestaltung für Kinder) und für die eigene Klassenkasse. Diese vier Spendenmöglichkeiten sollten die Kinder nach ihrer Präferenz auf einem 4-stufigen 'Siegertreppchen' gewichten.<sup>5</sup> In der einen Variante dieser Aufgabenstellung (Spenden *ohne* Handlungskonsequenz) sollten sich die SchülerInnen bloß vorstellen, sie hätten 50 Schweizer Franken und sollten sich als Gruppe einigen, in welcher Rangfolge sie für welches Projekt spenden würden. In der zweiten Variante dieser Aufgabenstellung (Spenden *mit* Handlungskonsequenz) erhielten sie jeweils als Schulklasse 50 Schweizer Franken, die im Anschluss real gespendet wurden.<sup>6</sup> Hintergrund dieser beiden Varianten der gleichen Aufgabe war die Frage, ob es für die Art und Weise der argumentativen Bearbeitung eine Rolle spielt, wenn die kollektiv zu fällende Entscheidung tatsächliche Handlungskonsequenzen im realen Leben hat (vgl. dazu die didaktische Diskussion um die oft positiv bewertete "Authentizität der Kommunikationsanlässe"; Becker-Mrotzek 2008:71).

Bei beiden Aufgaben lautete die Instruktion so, dass das erklärte Ziel eine Einigung in der Gruppe sei. Dies braucht grundsätzlich nicht durch Argumentieren, sondern kann auch durch Verfahren wie Abstimmen o.ä. gelöst werden. Das Konzept *Argumentieren* – sowie entsprechendes Vorwissen – sollte bewusst nicht aktiviert werden, jedoch wurde durch eine gezielte Wortwahl (*diskutieren, miteinander darüber reden, in Ruhe besprechen*) sowie dem abschließenden Satz "Dann interessieren wir uns dafür, was ihr euch überlegt habt" sichergestellt, dass den SchülerInnen klar war, dass es weniger das Ziel war, möglichst schnell eine Einigung herbeizuführen, sondern dass sie in irgendeiner Weise (begründend) dazu Stellung nehmen sollten.<sup>7</sup> Es gab keine explizite Vorgabe einer gewissen Gesprächslänge.

Nach verschiedenen Pilotierungsphasen haben wir entschieden, uns auf drei Variablen zu konzentrieren, die wir systematisch variierten, und weitere mögliche Variablen stabil zu halten oder (überwiegend aus Gründen der Machbarkeit) unbeachtet zu lassen.<sup>8</sup> Die systematisch variierten Variablen sind *Klassenstufe, Thema* und

<sup>5</sup> Dies ist eine Variation der Aufgabe, wie sie nach der Pilotierung eingeführt wurde. In der ursprünglichen Version sollten die Kinder lediglich einen Platz (für das favorisierte Projekt) vergeben. Die argumentativen Aktivitäten waren damit allerdings sehr gering. Mit der Anpassung konnte eine größere argumentative Aktivität evoziert werden.

<sup>6</sup> Die Handlungskonsequenz wurde sowohl bei der Instruktion hervorgehoben als auch im Anschluss an die Aufnahmen durch das Signieren eines Bestätigungsbriefes an das jeweilige Projekt deutlich gemacht.

<sup>7</sup> Diese Formulierungen wurden im Zuge der Pilotierung angepasst, da aufgrund von Aussagen einzelner SchülerInnen sowie basierend auf der Analyse teils sehr kurzer Gespräche deutlich wurde, dass die Aufgabe bisher stärker als Einigungsaufgabe und weniger als Diskussionsaufgabe interpretiert wurde. Die angepassten Formulierungen in der Instruktion hatten sodann einen positiven Einfluss auf die Interpretation der Aufgabe als Anlass zum Diskutieren.

<sup>8</sup> Im Rahmen der Pilotierung wurden weitere Variablen diskutiert und teilweise getestet, etwa Gruppengröße (Diskussionen in 3er-, 4er-, 5er-Gruppen sowie in Halbklassen), kommunikative Ordnung (schülerzentrierte Peer-Gespräche vs. lehrerzentrierte Gespräche bzw. Erwachsenen-Kind-Interaktion), Gruppenzusammensetzung (systematische Zusammensetzung, bspw. bzgl.

*Handlungskonsequenz.* Bei der Klassenstufe wurden Schulklassen der 2., 4. und 6. Klasse der Primarschule in der Deutschschweiz einbezogen (das entsprechende Alter variierte im Spektrum zwischen 7 und 14 Jahren). Damit stellten wir sicher, dass wir Aspekte der Entwicklung argumentativer Gesprächsfähigkeiten nachzeichnen können. Das Thema wurde wie oben beschrieben variiert mit dem Robinson- und dem Spendensetting. Durch die damit realisierte Vergleichbarkeit der Daten konnten themen- und aufgabenbezogene Aspekte ausgeschlossen bzw. kontrolliert werden. Die Variation der Handlungskonsequenz diente einer gezielten Überprüfung der Annahme, dass authentische Kommunikationsanlässe im Vergleich zu fiktiven Anlässen grundsätzlich andere Gesprächsprozessierungen nach sich ziehen.

Weitere Aspekte wurden ebenfalls stabil gehalten:

- Pro Gruppengespräch nahmen jeweils vier SchülerInnen<sup>9</sup> teil, die per Losentscheid ausgewählt wurden,
- es waren keine Erwachsenen anwesend (Peer-Gespräche),
- die Gesprächsfunktion bestand in allen Settings aus dem Ziel der gemeinsamen Einigung.

Die Peer-Gespräche wurden mit zwei Standkameras videografiert, sodass von allen vier SchülerInnen auch umfassende multimodale Handlungen erkennbar und analysierbar waren, was insbesondere bei der häufigen Realisierung von Zeigegesten auf das Arbeitsblatt bei gleichzeitiger Verwendung von Lokaldeixis zwingend notwendig war, aber auch bei nonverbalen Praktiken des Anzeigens von Einverständnis oder Widerspruch sowie bei Begründungen in Form von Gesten.<sup>10</sup> Im Anschluss an die Gruppendiskussionen führten die VersuchsleiterInnen mit den vier SchülerInnen ein Nachgespräch mit metasprachlich ausgerichteten Fragen über die soeben geführte Diskussion. Sowohl Instruktion als auch Nachgespräche wurden mit aufgenommen. Zusätzlich zu den Gesprächsaufnahmen wurden ethnografische Daten zum genauen Alter, zu den Sprachkenntnissen der Kinder (L1, L2 etc., DaZ-Förderunterricht) sowie zu schulisch erworbenen Vorkenntnissen zum Thema *Argumentieren/Diskutieren* erhoben. Es handelte sich um Auskünfte der Lehrpersonen über die Familiensprachen sowie im Falle von DaZ-Kindern über die entsprechenden Deutschkenntnisse. Leider fehlen uns teilweise die Angaben zu Sprachkenntnissen und schulisch erworbenen Vorkenntnissen (sowie in Einzelfällen auch zum genauen Alter) und zudem sind sie nicht in allen Fällen gleich präzise. Daher wurden diese Angaben nicht quantifiziert, sondern eher bei Detailanalysen im Sinne von ergänzenden und punktuell klärenden Wissensressourcen genutzt.

---

Gender), Gesprächsregeln (verfahrenszentrierte Gesprächsregeln können beispielsweise Regelungen der Rederechtsübergabe qua Meldeverfahren o.ä. umfassen, haben jedoch einen großen Einfluss auf die Interaktivität der Gespräche), Gesprächsfunktion (gemeinsam zu erreichendes Ziel vs. kontroverse Diskussion).

<sup>9</sup> Die kontrollierte Gruppengröße hatte praktische Folgen bei der Datenerhebung, da je nach Anzahl der Kinder in einer Klasse nicht alle Kinder an den (gefilmten) Gruppendiskussionen teilnehmen konnten, obschon eine Einwilligung der Eltern vorlag.

<sup>10</sup> Eine systematisch multimodale Analyse der Daten war nicht Gegenstand des Projekts, wäre aber ein lohnenswertes Unterfangen, z. B. bezüglich Praktiken der Affiliierung und Disaffiliierung (dazu Jacquin 2015), der Allianzenbildung oder Ko-Konstruktivität von gemeinsamen Beiträgen.

Im folgenden Abschnitt wird nun genauer beschrieben, wie sich das endgültige Korpus zusammensetzt.

### 2.3. Korpus

Das vorliegende Korpus besteht aus 180 videografierten Gruppendiskussionen und es wurden dabei die drei Variablen *Klassenstufe*, *Thema* und *Handlungskonsequenz* so variiert, dass jeweils ein Sample von 20 Gesprächen pro Variablengruppe resultierte (siehe Tabelle 1):

Klassenstufe	Thema/Setting	Handlungskonsequenz	Kürzel
<b>2. Klasse</b>	20x Robinson	nein	Ro_K2
	20x Spenden	ja	Sm_K2
	20x Spenden	nein	So_K2
<b>4. Klasse</b>	20x Robinson	nein	Ro_K4
	20x Spenden	ja	Sm_K4
	20x Spenden	nein	So_K4
<b>6. Klasse</b>	20x Robinson	nein	Ro_K6
	20x Spenden	ja	Sm_K6
	20x Spenden	nein	So_K6

Tabelle 1: Übersicht Vergleichskorpora

Die 180 Gruppengespräche wurden in insgesamt 53 Schulklassen an 24 Schulen erhoben, die in sechs verschiedenen deutschschweizerischen Kantonen liegen (Aargau, Basel-Landschaft, Basel-Stadt, Solothurn, Zürich und Zug). Wenn man die Schulstandorte zudem aufgrund ihres städtischen oder ländlichen Charakters einteilt, lassen sich von total 180 erhobenen Gruppen 97 als urban und 83 als rural beschreiben. Durch die Berücksichtigung verschiedener Orte und Einzugsgebiete wurde gewährleistet, dass verschiedene Gesellschaftsstrukturen mit einbezogen wurden.

Die Gesamtdauer der 180 Gruppengespräche beträgt 873 Minuten. Dabei zeigt sich eine altersbedingte Veränderung: Über alle Settings hinweg dauern die Gespräche in der 2. Klasse durchschnittlich 3.5 Minuten und in der 4. sowie 6. Klasse durchschnittlich 5.5 Minuten. Die Klassen 4 und 6 liegen hierbei sehr nah beieinander – ein Resultat, welches sich nicht nur bzgl. Gesprächsdauer zeigt, sondern auch bei einzelnen Gesprächsaktivitäten wie z. B. der Anzahl von Begründungshandlungen (siehe Kapitel 3) deutlich wird. Bei der Gesprächsdauer ist allerdings zu beachten, dass es eine große Streuung gibt und Ausreißer-Gespräche von 30 Sekunden sowie 22 Minuten im Korpus vorhanden sind.

Da die Zusammensetzung der Gruppen per Los entschieden wurde, gibt es bzgl. Genderdurchmischung unterschiedliche Gruppen, die aber insgesamt ein gutes Mischverhältnis ergeben: In 74 von 180 Diskussionen nehmen jeweils zwei Mädchen plus zwei Jungen teil. In 42 Gesprächen sind es drei Jungen plus ein Mädchen und umgekehrt in 40 Gesprächen drei Mädchen plus ein Junge. Und schließlich umfasst das Korpus je genau 12 reine Mädchen- und 12 reine Jungengruppen. Diese Gruppenverhältnisse ermöglichen die Bearbeitung von Fragestellungen hinsichtlich Geschlechterunterschiede in gesondert zusammengestellten Subkorpora.

Den Kindern wurde freigestellt, welche sprachliche Varietät sie in der Diskussion wählten. Obwohl die Schulsprache Standarddeutsch ist, ist es typisch, in Gruppenarbeiten in einen schweizerdeutschen Dialekt zu wechseln. Dies zeigt sich auch sehr deutlich in unseren Daten: In 139 Gesprächen wird ausschließlich Dialekt gesprochen und nur in acht Gesprächen wird die Schulsprache Standarddeutsch gewählt. In den übrigen 33 Gesprächen werden beide Varietäten eingesetzt, was meist damit zusammenhängt, dass einzelne Kinder Deutsch als Erstsprache sprechen oder als Zweitsprache (DaZ) gelernt haben und deswegen nicht aktiv Dialekt sprechen (können).

## 2.4. Computergestützte Verfahren der Datenaufbereitung

Im Folgenden diskutieren wir die verschiedenen Softwares, die wir teils umfassend und teils nur punktuell genutzt haben. Damit mündliche Daten quantitativ ausgewertet werden können, müssen sie in einem ersten Schritt transkribiert werden. Da es in Gruppengesprächen unter Kindern vielfach zu simultanem und überlappendem Sprechen kommt, wurde der Partitur Editor von EXMARaLDA mit der entsprechenden Partiturschreibweise gewählt. Bei EXMARaLDA<sup>11</sup> handelt es sich um ein XML-basiertes Software-Paket, welches u.a. Werkzeuge für die Transkription (Partitur Editor), die Annotation (Partitur inkl. flexibel programmierbares Annotationspanel), die Korpusverwaltung (COMA) und die korpuslinguistische Auswertung (EXAKT) enthält (vgl. z. B. Schmidt 2004a/b; Schmidt/Wörner 2009). Unser Korpus wurde mit dem Partitur Editor unter Anwendung des gesprächsanalytischen Transkriptionssystems GAT 2 (Selting et al. 2009) komplett transkribiert und annotiert. Am Ende enthielt eine fertige Transkription pro SprecherIn jeweils folgende Spuren:

- [v]-Spur: v=verbal, für das Gesprochene (i.d.R. in Dialekt);
- [sd]-Spur: sd=Standarddeutsch, für die interlineare Übersetzung in normiertes Standarddeutsch;
- [nv]-Spur: nv=nonverbal, für Beschreibungen von multimodalen Aktivitäten, die in der Interaktion bedeutsam und für das Verstehen notwendig sind;
- [a]-Spur: a=Annotation, für argumentationsrelevante Codes (vgl. Codierschema in Kapitel 3.2.).

Zudem wurde eine weitere [nv]-Spur eingefügt, um Pausen oder Handlungen von mehreren SprecherInnen zu erfassen – diese besitzt jedoch keine Sprecherzuordnung und ist daher von der oben genannten sprecherbezogenen [nv]-Spur unterscheidbar. Abbildung 1 zeigt einen Ausschnitt aus einem Transkript mit Codierung (zu den einzelnen Codes vgl. Kapitel 3.2.).

---

<sup>11</sup> Siehe Dokumentationen und Publikationen unter: <http://exmaralda.org/de/>

	19 [03:34.2]	20 [03:34.2*]	21 [03:35.1]	22 [03:35.7]	23 [03:37.3]	24 [03:37.7]	25 [03:38]	26 [03:39.9]
LUC [v]	Also ich WÜRD-							Aber mer cha
LUC [sd]	also ich würde							aber man kann al
LUC [nv]	<i>((rückt auf dem Stuhl nach vorne, schaut NAT und ANN an))</i>							
LUC [a]								4-meta
IRA [v]				aso ich find (.) ZÄ:LT?				
IRA [sd]				also ich finde Zelt				
IRA [nv]				<i>((zeigt auf das Zelt))</i>				
IRA [a]				1-für 2a-zel				
NAT [v]					nÄI (.) IIR (.) iir tüend zwÄI isch guet und mir zwÄI;			
NAT [sd]					nein ihr tut zwei ist gut und mir zwei			
NAT [nv]					<i>((zeigt zuerst auf LUC und IRA, dann auf ANN und sich))</i>			
NAT [a]					4-meta 4-norm			
ANN [v]		ich WÜRD <sub>i</sub> ,	ich würdi (so ÄI::S);					
ANN [sd]		ich würde	ich würde (so eins)					
ANN [nv]		<i>((zeigt auf das Moskitonetz))</i>						
ANN [a]		1-für 2a-mos 4-nonverbal						
VL1 [v]								
VL1 [sd]								
VL1 [nv]								
VL2 [v]								
VL2 [sd]								
VL2 [nv]								
[nv]								6)
[k]								

Abbildung 1: Ausschnitt aus einem EXMARaLDA-Transkript

Links sind die (anonymisierten) Sprecherkürzel zu sehen, jede Person hat vier Spuren: zunächst wird der transkribierte gesprochene Text wiedergegeben (v; hier Dialekt, z. B. ANN: "ich WÜRD<sub>i</sub>, ich würdi (so ÄI::S);"), gefolgt von der standarddeutschen Übersetzung (sd, "ich würde ich würde (so eins)") und wichtigen Beschreibungen nonverbaler Aktivitäten (nv, "((zeigt auf das Moskitonetz))"). Die nächste Spur (a) gibt dann die verschiedenen Codes an, welche den jeweiligen Äußerungen zugeordnet wurden (z. B. "1-für", "2a-mos", "4-nonverbal" in der Spur "ANN [a]"). Obwohl in dieser Spur mehrere Codes mit Bezug auf verschiedene Aspekte der Interaktion gleichzeitig vergeben werden und somit eine einzelne Code-Kombination die Zusammenhänge zwischen den vier Codierungsebenen aufzeigt, sind diese Codes in der Folge einzeln analysierbar; es können also sowohl einzelne Codes wie auch Code-Kombinationen analysiert werden. Dies machte es unnötig, für unterschiedliche Code-Gruppen je einzelne Spuren anzulegen. Gleichzeitig wurde damit auf die Praktikabilität des (technischen) Codiervorgangs Rücksicht genommen. Ganz unten folgen die Spuren [nv] für Nonverbales, welches nicht einer einzelnen Person zugeschrieben werden kann (hier handelt es sich, wie auch in obigem Beispiel, i.d.R. um Pausen), und [k] für weitere Kommentare der Transkribierenden (hier leer).

Die Aufbereitung der Daten in der Partiturschreibweise bzw. als interlineare Textspuren (verbal, standarddeutsche Übersetzung, nonverbal) hat Vorteile in der Visualisierung der Gleichzeitigkeit gesprochener Sprache (vgl. Schmidt 2003), allerdings ist die Übertragung der Daten in andere Software-Umgebungen vielfach problematisch. Während nämlich die Partiturnotation im Partitur Editor von EXMARaLDA als fortlaufendes und mit dem Video aligniertes Transkript angezeigt wird, müssen die interlinear angeordneten Spuren beim Export in Formate für die weitere Textverarbeitung (z. B. RTF) in Partiturflächen 'geschnitten' werden, sodass sie auf eine Seitenbreite passen. Dies erschwert einerseits die Lesbarkeit, da Zusammengehöriges dergestalt umgebrochen werden kann, dass u.U. die interlinearen

Übersetzungen, die Beschreibungen und/oder Annotationen nicht mehr untereinander, sondern in unterschiedlichen Partiturflächen abgebildet/sichtbar werden. Und andererseits hat es Konsequenzen auf die weitere Verarbeitung der Daten. Wenn beispielsweise eine Software für die qualitative Datenanalyse verwendet wird, wie MAXQDA<sup>12</sup>, ATLAS.ti<sup>13</sup> o.ä., so ist eine Codierung und Auswertung aufgrund der ungünstigen Umbrüche nur bedingt und mit Einschränkungen möglich.<sup>14</sup> Dennoch sind i.d.R. einfache quantitative Auszählungen und Analysen der einzelnen Codes in den exportierten Daten möglich.

Ein weiteres Problem, welches beim Export aus EXMARaLDA und Import in Analyse-Tools beachtet werden muss, ist die Tatsache, dass einmal importierte Transkripte nicht ohne Weiteres nachkorrigiert werden können bzw. nur über Umwege und Verluste bei den gesetzten Codierungen. Das heißt, die exportierten Transkripte werden als statische Textdokumente importiert, die nicht mehr mit der Videodatei verknüpft sind und zum Zeitpunkt des Exports als 'fertige' Daten behandelt werden (müssen). Schütte (2007:71) nennt dies ein "Killer-Kriterium" für die Verwendung von Programmen wie ATLAS.ti (gilt genauso für MAXQDA) in der Gesprächsanalyse. Denn typischerweise handelt es sich bei der Transkription um einen dynamischen Prozess, sodass während der Analyse je nach Schwerpunkt häufig einzelne Sequenzen noch genauer transkribiert, weiterbearbeitet und ergänzt werden. Da es zu empfehlen ist, die Anpassungen immer am Primärdokument vorzunehmen – in diesem Falle im Partitur Editor von EXMARaLDA, müsste danach theoretisch die Datei erneut in MAXQDA o.Ä. importiert werden, womit aber die bereits gesetzten Codes verloren gingen. Aufgrund dieser Schwierigkeiten wurde die Codierung der transkribierten Daten schließlich mithilfe eines eigens in den Quellcode programmierten Annotationspanels in EXMARaLDA selbst vorgenommen, um möglichst lange mit der ursprünglichen Datei arbeiten zu können.

Die Auswertungen, welche einerseits auf den Codierungen und andererseits auf den einzelnen Äußerungen basieren, wurden schließlich teilweise mit dem Auswertungsprogramm EXAKT von EXMARaLDA (wofür segmentierte Transkripte notwendig sind, welche via COMA von EXMARaLDA aufbereitet werden müssen) vollzogen. Teilweise wurde MAXQDA ergänzend für einfache Auszählungen genutzt. Weiterhin haben wir für komplexere Berechnungen und Visualisierungen R<sup>15</sup> bzw. die entsprechende Software RStudio<sup>16</sup> genutzt. Dank der XML-basierten Dateiformate von EXMARaLDA ist die Interoperabilität grundsätzlich gewährleistet, allerdings darf der Aufwand für die Datenaufbereitung in RStudio gerade bei großen Datenmengen nicht unterschätzt werden. Damit automatisierte Auswertungen zuverlässig sind, muss sichergestellt werden, dass die verschiedenen interlinearen

---

<sup>12</sup> Siehe <https://www.maxqda.de>

<sup>13</sup> Siehe <http://atlasti.com/de/>

<sup>14</sup> Bei MAXQDA ist es beispielsweise so, dass ein Code jeweils an den entsprechenden Text gehängt werden muss, um die verschiedenen Auswertungsmöglichkeiten nutzen zu können. Denn nur dann erhält man in den Auswertungen die Codes und die dazugehörigen Datenausschnitte. Da es aber nach Ausgabe der Transkripte aus EXMARaLDA in der Mehrheit der Fälle Umbrüche gibt und Zusammengehöriges über mehrere Partiturflächen verläuft, kann ein Code häufig nur an einen Textteil gehängt werden, was sich demnach einschränkend auf die weitere qualitative Analyse auswirkt.

<sup>15</sup> Siehe <https://www.r-project.org>

<sup>16</sup> Siehe <https://www.rstudio.com>

Spuren und Codes in eine durchsuchbare und analysierbare logische Struktur gebracht werden. Ein Vorteil von R ist, dass nach der Datenaufbereitung weitere Exportfunktionen möglich sind, womit dann wiederum weitere Analysewerkzeuge nutzbar werden. Entsprechend konnten die gesamten Datensätze beispielsweise für einfachere Auswertungen und Visualisierungen in Excel ausgegeben werden. Zudem konnten auch TXT-Dateien auf Basis der standarddeutschen Spuren erstellt und entsprechend für das korpuslinguistische Tool AntConc<sup>17</sup> verwendet werden.

Insgesamt ist festzuhalten, dass es bei den unterschiedlichen Transkriptionsprogrammen, die auf dem Markt sind, einer genauen Prüfung für den eigenen Forschungsprozess bedarf. In jedem Fall ist der Arbeitsaufwand bei einer systematischen Aufbereitung für quantitative Analyse Zwecke sehr groß. Da gerade auch bei qualitativ ausgerichteten Studien meist noch nicht alle Auswertungsfragen vordefiniert sind, zahlt es sich aus, auf ein interoperables System zu setzen, welches die flexible Übertragung der Daten in weitere Softwareumgebungen ermöglicht.

### 3. Methodisches Vorgehen bei der Datenanalyse

Hauptanliegen des Projektes ist die umfassende Beschreibung mündlicher Argumentationsfertigkeiten in kindlichen Peer-Interaktionen in der Kleingruppe, insbesondere deren Vergleich über verschiedene Altersstufen (siehe Kapitel 1). Das Datenmaterial wurde schwerpunktmäßig mit den Methoden der Gesprächsanalyse (vgl. Deppermann 2008; Sidnell/Sivers 2013) detailanalytisch untersucht,<sup>18</sup> bietet aber aufgrund seiner Anlage gleichermaßen das Potenzial quantitativ beforscht zu werden, um so von den zahlreichen individuellen Fertigkeiten auf kompetenzbezogene Fähigkeiten der untersuchten Altersklassen zu schließen (siehe Kapitel 2.1.).

Hierfür wurden die Daten einerseits durch einfache Wort-Suchabfragen auf ihre oberflächennahen, sprachlichen Phänomene hin gesichtet, die bezüglich argumentativer Aktivitäten relevant sind (z. B. lexikalische Markierungen von Begründungen durch Konjunktionen, Modalpartikeln etc.). Andererseits wurde den einzelnen Sprachhandlungen anhand eines Codierschemas ein Code zugeordnet (siehe Kapitel 3.2.), um auch größere und komplexere argumentative Einheiten quantifizieren zu können. Die Entwicklung eines sinnvollen Codierschemas ist besonders anspruchsvoll, wenn man spontan geäußerte Sprechhandlungen in ein Kategoriensystem einordnen will, das weder zu tief in das qualitative Detail geht (vgl. gesprächsanalytischer Zugang) noch zu sehr an der lexikalischen Oberfläche bleibt (vgl. korpuslinguistischer Zugang). Auch müssen Festlegungen getroffen werden, hinsichtlich welcher – möglichst klar beschreibbarer – Phänomene das Korpus codiert werden soll. Codieren bedeutet daher immer eine reduktionistische Analyse von Sprachdaten. Da das vorliegende Forschungsinteresse darin besteht, *argumentationsrelevante* Aktivitäten der Kinder zu beschreiben, wurden für das Codierschema wenige, aber für das Argumentieren zentrale Codes entwickelt, die v.a. den "Kernjob" des Begründens (Heller 2012) fokussieren.

<sup>17</sup> Siehe <http://www.laurenceanthony.net/software/antconc/> und Einführung unter:

[http://www.bubenhofer.com/korpuslinguistik/kurs/index.php?id=eigenes\\_AntConc.html](http://www.bubenhofer.com/korpuslinguistik/kurs/index.php?id=eigenes_AntConc.html)

<sup>18</sup> Vgl. Hauser/Luginbühl 2015, 2016, 2017; Kreuz/Mundwiler 2016; Kreuz/Mundwiler/Luginbühl 2017; Mundwiler/Kreuz 2018, Mundwiler et al. 2017.

Im Folgenden wird zunächst dargestellt, auf welchen theoretischen und empirischen Überlegungen die Entwicklung des Codierschemas basiert und wie die Gesprächsdaten für eine quantitative Analyse aufbereitet wurden, um Aussagen über alters- bzw. klassenbezogene Unterschiede argumentativer Teilkompetenzen treffen zu können. Bei der Entwicklung des Codierschemas stand das Interesse im Zentrum, nicht nur argumentationslogische Aspekte, sondern auch ausgewählte interaktionsspezifische, gesprächsanalytisch basierte Aspekte zu erfassen. Dann fokussieren wir die Auswertung und die Darstellungsmöglichkeiten der codierten Daten. Und zuletzt wird die Frage der Normalisierung diskutiert, die uns bei der Auswertung immer wieder beschäftigt hat: Wie bzw. mit welcher Bezugsgröße lassen sich unterschiedlich lange Gespräche sinnvoll miteinander vergleichen?

### 3.1. Theoriebasierte und qualitative Annäherung an die Daten

Ogleich die erste Annäherung an die Daten zunächst mittels qualitativer, gesprächsanalytischer Verfahren erfolgte (siehe Abschnitt unten), wurden zur Entwicklung eines Codesystems auch Forschungsarbeiten gesichtet, die sich mit der Quantifizierung mündlicher Interaktionen beschäftigen. Zu nennen sind hier beispielsweise die Arbeiten von Canary, Ratledge und Seibold (1982) und Canary, Brossmann und Seibold (1987), die auf Grundlage sowohl theoretischer Annahmen (z. B. Perelman/Olbrechts-Tyteca 1958) als auch empirischer Beobachtungen von argumentativen Interaktionen das umfassende *Conversational Argument Coding Scheme* (CACS) entwickelten. Das CACS wurde Ende der 70er/Anfang der 80er Jahre erarbeitet "as part of structuration theory work in communication [...], especially within the structuration approach to argument in group deliberations" (Canary/Seibold 2010:7). Das Codierschema sollte also auf mündliche spontane Interaktionen angewendet werden und enthält daher nicht nur Codes, die der argumentativen Logik unterworfen sind, sondern auch solche, die Eigenschaften argumentativer Elemente in ihrem konkreten Gebrauch bzw. ihrer Sequenzialität sowie das argumentative Verhalten der Interagierenden abbilden. Das Codierschema integriert damit einerseits die Komposition eines Arguments bzw. die Argumententwicklung durch bestimmte (einfache oder komplexe) Strukturen von einem oder von mehreren Interagierenden und andererseits auch sequenzielle Abläufe (vgl. "systematic repetitions of act-to-act argument behaviors" wie z. B. *Converging* und *Diverging Sequences*, Canary/Seibold 2010:21). Das CACS wurde seither für die unterschiedlichsten Kontexte interaktiven Argumentierens genutzt<sup>19</sup> und ist mittlerweile eine weit verbreitete inhaltsanalytische Methode "for investigating argument in involving diverse communicative interactions" (Canary/Seibold 2010:7).

Adaptiv nutzbar für das eigene Codierschema und kompatibel mit dem eigenen Verständnis von Argumentieren waren daneben auch die Überlegungen von Kyratzis et al. (2010). Sie beobachteten in ihren Daten von Interaktionen in Kindergartenruppen (3-6 Jahre), dass Begründungen nicht nur aus (potenziellen) Mei-

<sup>19</sup> Vgl. z. B. Untersuchungen von Gruppenberatungen (Seibold/Meyers 2007), von Entscheidungsgesprächen von Juroren (Burnett/Badzinski, 2000), von Konflikten unter Freunden (z. B. Semic/Canary 1997, 2003), von Problemlösegesprächen zwischen Eheleuten (z. B. Canary et al. 1987), von politischen Diskussionen im Fernsehen (z. B. Brossmann/Canary 1990) und von ethnopolitischen Gruppenkonflikten (Ellis/Maoz 2002, 2007) (vgl. Angaben in Canary/Seibold 2010:7f.).

nungsverschiedenheiten hervorgehen müssen, sondern auch die Aussagen des Partners stützen können, um beispielsweise seinen Spielzug zu legitimieren. Aufbauend auf diesen Beobachtungen arbeiteten sie drei Motivationen ("motivations") für Begründungen heraus (Kyratzis et al. 2010:126ff.):

- Opposition (die Ablehnung einer Idee wird begründet)
- Validierung (die Idee des Partners wird begründet)
  - "the speaker followed a partner's proposal by providing a reason for it in their next turn"
  - "the speaker agreeing with a partner's suggestion, and using a justification to justify the agreement"
  - "the speaker elaborating on the partner's suggestion, and using a justification to justify the elaboration"
- Selbsterweiterung (eine eigene Idee wird begründet)

Ebenso wie im CACS werden auch hier einzelne (Begründungs-)Aktivitäten in ihrem sequenziellen Zusammenhang aufgefasst und hinsichtlich ihres interaktionalen Status codiert. Eines der wichtigsten Paradigmen der Gesprächsanalyse – die Sequenzialität – wird damit berücksichtigt. Ebenso beziehen sich sowohl das CACS als auch das Schema von Kyratzis et al. (2010) auf *spezifische* Daten spontaner Interaktionen und beschreiben Sprachhandlungen an authentischem Material. Dies macht es jedoch nicht möglich, die vorgeschlagenen Codes für die eigene Datenanalyse unverändert zu übernehmen. Einige dieser Codes konnten zwar *adaptiv* übernommen werden, viele mussten aber auf Grundlage der Eigenlogik der eigenen Daten in gemeinsamen Datensitzungen *neu erarbeitet* werden. Des Umstands, dass ihre Codes nicht ohne Weiteres unverändert übernommen werden können, sind sich die Entwickler des CACS durchaus bewusst (Canary/Seibold 2010:23):

Finally, we hope that in the future researchers interested in conversational argument – whether they examine group decision making, interpersonal interactions, or attempts at communicative influence in other contexts – continue to connect their theoretical orientations with empirical observation of actual interaction.

Will man die Eigenlogik der erhobenen Daten berücksichtigen, setzt die Entwicklung von Codes daher immer eine *qualitative Voranalyse* der eigenen Daten voraus. Ziel ist es, einen Überblick über typische Charakteristika der Interaktionen zu gewinnen und zu eruieren, welche Codes sich als relevant für das spezifische Forschungsinteresse – mündliche Argumentationsfertigkeiten und deren alters- bzw. klassenbezogenen Veränderungen – erweisen könnten. So wurden die Daten unter Einbezug der in der Einleitung dargestellten Annahmen zum Argumentieren mittels Gesprächsanalyse zunächst global qualitativ analysiert, z. B. hinsichtlich typischer Gesprächsphasen (Kreuz/Luginbühl/Mundwiler 2019) und konversationeller Muster, dann auch hinsichtlich meso- und mikroskopischer Phänomene, wie einzelner Sprechhandlungen (z. B. Positionierungen), der Prozessierung verschiedener argumentativer 'Jobs', lokaler Argumentstrukturen und lexikalischer Markierungen von Begründungen (dazu Hauser/Luginbühl 2015, 2017; Kreuz/Mundwiler/Luginbühl 2017; Hauser/Kreuz 2018; Mundwiler/Kreuz 2018; Mundwiler et al. 2017). Auf Grundlage des so gewonnenen detaillierten Überblicks kristallisierten sich relevante Codes heraus, um argumentative Aktivitäten auch quantitativ untersuchen zu

können (siehe Kapitel 4.1.). Die Entwicklung des definitiven Codierschemas unterlag einem mehrstufigen, rekursiven Prozess, bei dem die Codes mehrmals getestet und modifiziert werden mussten (z. B. hinsichtlich ihres Detaillierungsgrads).

### 3.2. Quantitative Annäherung an die Daten: das Codieren

Auf Grundlage der oben beschriebenen theoretischen Annahmen und eigener qualitativer Analysen wurde ein Codebuch entwickelt, das Codierungen sowohl auf der Ebene einzelner Sprachhandlungen als auch auf argumentationsbezogener lexikalischer Ebene vorsieht.

Einer der Codes auf der Sprachhandlungsebene ist die sogenannte potenzielle Auslösehandlung für Argumentationen (Spranz-Fogasy 2006) – im Codierhandbuch als 'Thematisieren', als 'Positionieren für' oder 'Positionieren gegen' erfasst (Ebene 1). Im Spendensetting kommt durch die Spezifik der Aufgabenstellung (Reihung der Projekte in einer Rangliste) noch das 'Platzieren' hinzu. Für alle im folgenden diskutierten Code-Ebenen gilt, dass wir keine Mehrfachkodierungen innerhalb derselben Ebene vorgenommen haben. Uneindeutige Fälle (z. B. infolge von Formulierungsabbrüchen etc.) wurden als 'Thematisieren' codiert.

#### Beispiel Codes Ebene 1

Äußerung <sup>20</sup>	Code	Bedeutung
S1: Aso BUSCHmesser,	1-th	Das Kind thematisiert/benennt einen Gegenstand, ohne persönlich Stellung zu beziehen.
S2: Ich bin für das ZELT. S2: verbAnd braucht man eigentlich NICHT.	1-für 1-geg	Das Kind äußert seine Meinung und spricht sich für bzw. gegen einen Gegenstand aus.
S3: BIENE auf platz Eins;	1-platz	Das Kind nominiert ein Projekt für den ersten Platz.

Um den thematischen Bezug des aktuellen zu codierenden Turns mit zu berücksichtigen, wurde ein Code entsprechend der zwölf ausdiskutierenden Gegenstände bzw. der vier Spendenprojekte vergeben (Code auf der Ebene 2a und entsprechend für das Spendensetting 2b). Dieser wurde nach Möglichkeit *immer* gesetzt – nicht nur, wenn der thematische Bezug explizit genannt wurde, sondern auch, wenn er aus dem gesprächslokalen Kontext plausibel rekonstruiert werden konnte.

<sup>20</sup> Bei den Datenbeispielen handelt es sich um prototypische Beispiele aus dem Korpus, transkribiert nach GAT 2 (Selting et al. 2009). Da es an dieser Stelle um die Darstellung der inhaltsanalytischen Codierung geht, wurden die Beispiele so vereinfacht und bereinigt, dass sie den illustrierenden Zweck erfüllen können. Daher wurden keine Einzelquellen angegeben.

**Beispiel Codes Ebene 1 und 2a/b<sup>21</sup>**

Äußerung	Code	Bedeutung
S1: Aso BUSCHmesser ,	1-th 2a-bus	Das Kind thematisiert/benennt das Buschmesser, ohne Stellung zu beziehen.
S2: Ich bin für das ZELT.	1-für 2a-zel	Das Kind äußert seine Meinung und spricht sich für das Zelt aus.
S3: BIENE auf platz Eins;	1-platz 2b-bie	Das Kind nominiert das Bienenprojekt für den ersten Platz.

Wie in der Einleitung und in Kapitel 3.1. aufgezeigt wurde, ist a) der Kern des Argumentierens die interaktive Begründungsaktivität, die sich b) in unterschiedlich starker Divergenz (bzw. Konvergenz) an der Vorgängerausäußerung orientiert. Um diesen beiden Tatsachen gerecht zu werden, wurde ein weiterer Code auf der Sprachhandlungsebene entwickelt, der die interaktive Einbettung von Begründungen im Gespräch mit ihren jeweiligen Ausrichtungen berücksichtigt. Begründungshandlungen werden als 'initiiierend' (3a-in), 'zustimmend' (3a-zu), 'ablehnend' (3a-ab) und 'mehrfach ausgerichtet' (3a-mehrfach)<sup>22</sup> kategorisiert (Ebene 3a), wobei unklare Fälle nicht zugeordnet wurden. Dieser Code 3a lehnt sich an die gesprächsanalytische Forderung einer sequenziellen Datenanalyse an und soll der Vorgehensweise einer Turn-by-Turn-Analyse Rechnung tragen. Außerdem soll er etwa dafür Hinweise geben, inwiefern die Kinder im Gespräch bereits etablierte Positionen aufgreifen oder tendenziell eher eigene Positionen mit initiiierenden Begründungen plausibilisieren. Ebenso geben entsprechende Auswertungen Hinweise darauf, ob eine Diskussion primär kontrovers oder konsensuell geführt wird.

<sup>21</sup> Bei den Codes wurden für die Gegenstände bzw. die Spendenprojekte die folgenden Abkürzungen verwendet: Im Robinsonsetting (Ebene 2a) waren dies: Sackmesser [Taschenmesser] = '2a-sac', Buschmesser = '2a-bus', Schlafsack = '2a-sch', Wolldecke = '2a-wol', Zelt = '2a-zel', Moskitonetz = '2a-mos', Streichhölzer = '2a-str', Feuerzeug = '2a-feu', Kochtopf = '2a-koc', Leuchtraketen = '2a-leu', Handy = '2a-han', Verbandszeug = '2a-ver'. Im Spendensetting (Ebene 2b) waren dies: Projekt zum Schutz der Bären = '2b-bär', Projekt zum Schutz der Bienen = '2b-bie', Stiftung Feriengestaltung für Kinder = '2b-kin', eigene Klassenkasse = '2b-kl'. In beiden Setting gab es eine zusätzliche Sammelkategorie ('2a-andere' bzw. '2b-andere') für Verweise auf weitere zusammengehörende semantische Felder bzw. andere kontextuelle (Begründungs-)Zusammenhänge oder auf unspezifische Allgemeinplätze oder ambige (Rück-)Bezüge.

<sup>22</sup> Hierzu zählten wir Begründungen, in denen von einem einzelnen Sprecher/einer einzelnen Sprecherin mehrere Begründungszusammenhänge aneinandergereiht bzw. gekoppelt oder wieder relativiert oder modalisiert werden; auch das Abwägen mehrerer Projekte gegeneinander zählten wir dazu.

**Beispiel Codes Ebene 3a**

<b>Äußerung mit Vorgängeräußerung</b>	<b>Code</b>	<b>Bedeutung</b>
01 S1: Ich bin für das ZELT.  02 S2: ja aber wir können auch auf einem BAUM schlafen?	1-für 2a-zel  2a-zelt 3a-ab	Das Kind äußert eine ablehnende Begründung in Bezug auf die Vorgängeräußerung.
01 S1: nEhmen wir die LEUCHTrakete;  02 S2: ja weil dann können wir NOTsignale senden.	1-platz 2a-leu  2a-leu 3a-zu	Das Kind äußert eine zustimmende Begründung in Bezug auf die Vorgängeräußerung.
01 S1: wir könnten die WOLLdecke nehmen?  02 S2: und das bÜschmesser brauchen wir zum JAGEN;	1-für 2a-wol  2a-bus 3a-in	Das Kind initiiert eine Begründung für das Buschmesser, d.h., es wird ein neues Thema eingeführt, ohne dabei argumentativ Bezug zur Vorgängeräußerung zu nehmen.

Darüber hinaus wurden die Begründungshandlungen nicht nur hinsichtlich ihrer interaktiven Einbettung, sondern auch hinsichtlich ihrer lexikalischen Markierung (Explizitheit) codiert (Code 3b). Meyer (1975:56f.) schlägt hierzu eine Unterteilung verschiedener Arten von Explikationen bei Begründungen vor: Nullexplikation (keine Markierung), vage Explikation (Partikeln, Relativsätze),<sup>23</sup> volle Explikation (Konjunktion). Im Sinne einer möglichst einfachen und eindeutigen Analyse von Begründungsmarkierungen, wurde für die vorliegenden Daten lediglich die volle Explikation als Code verwendet.

Markierungen – so unser Eindruck nach der ersten Sichtung des Materials – werden von den Interagierenden v.a. als kausal (3b-kaus), konditional (3b-kond) und final (3b-fin) markierte Begründungen hervorgebracht. Bei anderen Formen der Markierung wurde ein Sammel-Code vergeben (3b-rest), welchem auch unklare Fälle zugeordnet wurden. Der Code 3b bezieht sich ausschließlich auf an der Oberfläche beobachtbare, also rein lexikalische Phänomene.

Das Codierschema fokussiert die interessierenden argumentationsrelevanten Phänomene, wurde aber zusätzlich um eine Ebene 4 erweitert. Codes der Ebene 4 beschreiben weitere auffällige Sprachhandlungen, die v.a. zum Zwecke der schnellen Auffindbarkeit für künftige qualitative Analysen, jedoch nicht systematisch, gesetzt wurden. Codes der Ebene 4 beschreiben z. B. metasprachliche Äußerungen, Quaestioverschiebungen, Normaufrufe, Ko-Konstruktionen etc.

<sup>23</sup> Die semantisch vage Explikation entsteht durch Verwendung von Partikeln oder Relativsätzen, ist also nicht eindeutig markiert, jedoch auch nicht gänzlich unmarkiert (Beispiel für die Markierung anhand einer Partikel: "Ich bleibe zu Hause. Es regnet ja.").

Sprechhandlungen wie einfache Zustimmung- und Ablehnungspartikel (*ja/nein, stimmt, eben*) sind zwar relevant, z. B. für Sprecherpositionierungen (vgl. z. B. Felton et al. 2015 bei dyadischen Gesprächen), wurden aber aufgrund ihrer Ambiguität und ihrer vielfältigen Funktionen (Zustimmung, vage Positionierung etwa durch gedehntes *ja* oder durch *continuer* etc.) nicht codiert.

Zur Illustration wird ein Beispiel einer zusammenhängenden Sequenz mit ihren entsprechenden Codes aufgeführt:

### Beispiel Codes Ebenen 1, 2 und 3

Äußerung	Code	Bedeutung
01 S1: HANdy?	1-th 2a-han	Das Kind thematisiert/benennt das Handy, ohne persönlich Stellung zu beziehen.
02 S2: jA Ich nehms HANdy.	1-für 2a-han	Das Kind äußert seine Meinung und spricht sich für das Handy aus.
03 S3: ja zum ANrufen;	2a-han 3a-zu 3b-fin	Das Kind äußert eine zustimmende Begründung in Bezug auf die Vorgängerausäußerung und markiert sie final.
04 S1: aber wAs ist wenn wir gar keinen emPFANG haben,	2a-han 3a-ab 3b-kond	Das Kind äußert eine ablehnende Begründung in Bezug auf die Vorgängerausäußerung und markiert sie konditional.
05 S3: stImmt. 06 dann nEhmen wir die lEuchtraketen weil da können wir signALE senden;	- 1-für 2a-leu 3a-in 3b-kaus	- Das Kind initiiert eine kausal markierte Begründung für die Leuchtrakete, da es ein neues Thema einführt und eine Begründung für dieses neue Thema einführt, weshalb die Begründung ohne direkten argumentativen Bezug zur Vorgängerausäußerung bleibt.

Das gesamte Korpus wurde nach diesem Schema von vier vorgängig geschulten studentischen Hilfskräften in EXMARaLDA (siehe Kapitel 2.4.) codiert und von den Projektmitarbeitenden in einem zweiten Durchgang überprüft. Dennoch stieß das Codieren oft an seine Grenzen (siehe Diskussion in Kapitel 5), da sich gewisse Sprachhandlungen als zu komplex und kontextabhängig herausstellten, als dass sie durch das Setzen eines bestimmten Codes hinreichend beschrieben werden konnten (Doppelcodierungen sollten möglichst vermieden werden). Außerdem kam es je nachdem, ob die Videos der Diskussionen bei der Codierung einbezogen wurden, zu teilweise unterschiedlichen Annotationen. Grundsätzlich wurde das Korpus basierend auf den Transkripten codiert und nur bei Unklarheiten oder punktuellen Tests wurden die Videos hinzugezogen. Erst dadurch kamen Abweichungen zum Vorschein, die verdeutlichten, dass die Videos in jedem Fall das detailliertere Bild des Kommunikationsereignisses repräsentierten. Jedoch war es aus zeitlichen Gründen nicht möglich, die Videos für den gesamten Codierprozess als Basis zu nehmen. Auch wenn quantitative Auswertungen bei der kompetenzorientierten

Analyse umfangreicher Daten hilfreich und u.E. unerlässlich sind und entsprechende Resultate grundsätzlich auf Tendenzen hinweisen, darf – und will – die Auszählung und mathematische Verrechnung von Codes nie über die Heterogenität der Interaktionen hinwegtäuschen.

### 3.3. Normalisierung mündlicher Daten

Für die quantitative Auswertung werden typischerweise Daten und Codes gezählt und miteinander verglichen. Um eine Vergleichbarkeit herzustellen, sind relative Werte notwendig und entsprechend bedarf es einer Referenzgröße, die für alle Teilkorpora als Bezugsgröße genutzt werden kann, um die teilweise heterogenen Daten ins Verhältnis zueinander zu setzen. Man spricht dabei von der *Normalisierung* der Daten (vgl. z. B. Bubenhofer 2009:150). Im vorliegenden Korpus finden sich zwar gewissermaßen standardisierte und somit vergleichbare Daten, da ausgehend von einem Stimulus (Instruktion) und einer Zielvorgabe (Einigung in der Gruppe) Diskussionen elizitiert wurden, in denen gewisse Aspekte (z. B. Gruppengröße) über das gesamte Korpus und andere Aspekte (z. B. Klassenstufe, Setting mit bzw. ohne Handlungskonsequenz) innerhalb von Subkorpora stabil gehalten wurden. Dennoch stellen wir – wenig überraschend – große Heterogenität fest, beispielsweise bzgl. Gesprächslänge, Sprecheranteile oder Turnlänge.

Eine mögliche Referenzgröße ist die Anzahl der Wörter, wie es beispielsweise in der Korpuslinguistik gängig ist (vgl. z. B. Bubenhofer 2009:150; Scherer 2014: 40f.). Die Problematik bei Peer-Gesprächen ist aber, dass es zum Teil kürzere oder längere Passagen gibt, in denen sich die Kinder nicht mit dem Lösen der Aufgabe beschäftigen, sondern über private Angelegenheiten oder über die Kamera bzw. die Aufnahmesituation sprechen. Solche Nebensequenzen oder die genannten individuellen Unterschiede im Sprechverhalten können das Bild demnach stark verzerren. Dennoch gibt es gewisse Fragestellungen, für die eine Normalisierung auf Basis der Anzahl der Wörter sinnvoll ist und so wurde beispielsweise bei tendenziell korpusanalytischen, lexikalischen Abfragen diese Referenzgröße genutzt. Grundsätzlich erlaubt es aber unsere Codierung jedoch, Nebensequenzen wenn nötig auszublenken und damit nur argumentationsrelevante Sprecherbeiträge in die Auswertung einzubeziehen (siehe dazu auch weiter unten).

Problematisch verhält es sich auch mit der Gesprächsdauer als mögliche Referenzgröße, da gewisse Kinder schneller oder langsamer sprechen und sich auch bei den Sprechpausen große Unterschiede zeigen. Die Gesprächsdauer wäre zwar einfach zu ermitteln, aber aufgrund der Varianzen kann sie ebenfalls nur unter Vorbehalt zur Normalisierung genutzt werden.

Ein weiterer Zugang ist es, die Anzahl codierter Gesprächsbeiträge als Bezugsgröße zu verwenden, da wir damit eine vergleichbare Einheit aus gesprächsanalytisch fundierter, handlungs- und argumentationsbezogener Sicht schaffen. Konkret bedeutet dies, dass nicht die Länge eines Gesprächsbeitrags (z. B. Anzahl Wörter) als Bezug gilt, sondern die Art der Sprachhandlung (z. B. zustimmendes Begründen, metadiskursive Handlung o.ä.). Zu bedenken gibt es aber, dass aufgrund der Entscheidung, Äußerungen nach gesprächsbezogenen Handlungen zu codieren, die Länge der codierten Beiträge unbeachtet bleibt und entsprechend sehr unterschiedlich ausfallen kann. So kann eine Begründung beispielsweise aus kurzen, ko-konstruierten Teiläußerungen mehrerer SprecherInnen bestehen, sie kann aber auch

längere narrative Diskurseinheiten eines einzelnen Sprechers/einer einzelnen Sprecherin umfassen.<sup>24</sup> Zudem wurden für weiterführende Analysen teilweise auch (unterschiedlich lange) Nebensequenzen codiert, die zwar in Bezug auf Gesprächsstrukturierung und Normaufrufe o.ä. interessant sind, jedoch nicht direkt mit der (argumentativen) Diskussion in Verbindung stehen. Demnach sind nicht alle codierten Beiträge gleichermaßen relevant für die Auswertung argumentativer Gesprächskompetenz.

Aufgrund dieser Gefahr der Verzerrung wurde in den meisten Fällen mit der Anzahl aufgabenbezogener Gesprächsbeiträge gerechnet. Demnach wurden nur diejenigen codierten Gesprächsbeiträge in die Berechnung einbezogen, in denen ein Bezug zur Diskussionsaufgabe erkennbar ist (vgl. z. B. auch *game-related utterance* bei Domberg et al. 2018 oder *discourse moves* bei Felton et al. 2015). So kann sichergestellt werden, dass Nebensequenzen oder gesprächssteuernde Verfahren die Werte nicht verzerren.

Punktuell haben wir auch *dieselbe* Suchabfrage mit unterschiedlichen Bezugsgrößen ausgeführt, um zu testen, wie sehr sich das gewählte Verfahren der Normalisierung auf die relativen Werte auswirkt. Die Ergebnisse waren teilweise fast identisch. Da es sich in unserer Studie um vergleichsweise große Datenmengen handelt, sind offenbar die genannten Unterschiede bzgl. Sprechtempo oder (Nicht)Einbezug von Nebensequenzen mehrheitlich unproblematisch.

#### 4. Auswertungen und Ergebnisse

In der vorliegenden Studie kamen einfache statistische Verfahren zur Anwendung, die eine Deskription der Gesprächsdaten auf Basis der Transkripte und der codierten Transkripte ermöglichen. Es wurden verschiedene Verfahren genutzt, um ausgehend von Werten, Grafiken und komplexeren Visualisierungen primär Vergleiche anzustellen und Ergebnisse zu erzielen, aber auch um hypothesengenerierend vorzugehen und daraus ableitend neue qualitative Auswertungen vorzunehmen (siehe z. B. Kapitel 4.2.).

Wir präsentieren in der Folge ausgewählte Suchabfragen und diskutieren die Aussagekraft der Berechnungen. Ein wichtiger Aspekt bei Auswertungen allgemein, aber gerade auch bei der Kombination quantitativer und qualitativer Auswertungsmethoden, sind Visualisierungen. Diese können einerseits abbildenden, veranschaulichenden Charakter haben, oder aber eine ideengenerierende Funktion erfüllen.

---

<sup>24</sup> Eine ko-konstruierte Begründung fällt so mehr ins Gewicht als eine lange monologische, d.h., die Quantitäten können nicht nur auf die Anzahl der Begründungen bezogen werden, sondern sind auch vom Grad der Ko-Konstruktion beeinflusst; zudem sagen sie nichts aus über den quantitativen durchschnittlichen Redeanteil (die Äußerungen können immer sehr kurz oder sehr lang sein; hohe Zahlen bedeuten nicht, dass sich die Kinder im Durchschnitt länger äußern).

#### 4.1. Fragestellungen und Auswertungen über das gesamte Korpus

Vor dem Hintergrund der Fragestellung(en) nach der Ausprägung argumentativer Gesprächskompetenzen auf verschiedenen Altersstufen und auf Basis von Literaturrecherchen und eigenen qualitativen Fallanalysen wurden verschiedene Teilbereiche definiert, die einerseits relevant für das mündliche Argumentieren sind und bei denen wir andererseits zusätzliche Hinweise durch quantitative Zugänge erwarteten. Der Ausgangspunkt unserer Analysen sind also qualitative Analysen, mit denen wir für mündliche Argumentationskompetenz relevante Aspekte in unserem Korpus identifiziert haben, ohne aber für das ganze Korpus (oder bezüglich der Altersstufen einzelne Teile davon) zuverlässige Aussagen machen zu können. Im Folgenden werden Auswertungen zu den Themenschwerpunkten *Modalisierung und Perspektivierung*, *Komplexität von Argumentationen* und *Interaktivität* vorgestellt und dabei Grafiken genutzt, um zusammenfassende Ergebnisse und Tendenzen über das *gesamte* Datenkorpus abzubilden.

##### 4.1.1. Modalisierung und Perspektivierung

In der Forschungsliteratur zum Argumentieren kommt der Modalisierung und Perspektivierung eine besondere Bedeutung zu, insbesondere unter Berücksichtigung der personalen Kompetenzausprägung (vgl. v.a. Grundler 2011:293ff.). Modalisierungen können einerseits durch den Verbmodus im Konjunktiv ausgedrückt werden, wodurch der Grad der Faktizität unter Bezugnahme auf (mangelndes) Wissen, Normen oder Wünsche o.ä. eingeschränkt oder variiert werden kann (vgl. z. B. Diewald 1999; Grundler 2011:293f.; Nuyts 2001; Redder 2009:91; Schmitt 2002:91; Schwitalla 2012:168-172). Andererseits werden in (argumentativen) Gesprächen vielfach Modal- bzw. Abtönungspartikeln genutzt, um Meinungen oder Sachverhalte in abgeschwächter Form hervorzubringen, u.a. auch als *face*-schonende Strategie (vgl. Grundler 2011:295; Locher 2004:113). Dabei können Modalisierungen dafür genutzt werden, die eigene Perspektive als (nicht) verhandelbar zu markieren oder beispielsweise anzuzeigen, in welchem Maße andere Perspektiven in der eigenen Position berücksichtigt werden.

Wir geben im Folgenden Beispiele aus der qualitativen Analyse sowie unsere vorläufigen Schlüsse, die sich daraus ergeben haben. Im folgenden Beispiel des Robinsonsettings einer sechsten Klasse (siehe auch Kreuz/Luginbühl, eingereicht) werden Modalpartikeln nicht nur bei der initiativen Stellungnahme zum thematisierten Gegenstand ("Buschmesser") eingesetzt. Der Gebrauch von Modalisierungen bestimmt den weiteren Gesprächsverlauf nicht unwesentlich mit, da er sich auch in der folgenden Begründungshandlung eines anderen Sprechers fortsetzt und konvergentes gemeinsames Begründen ausweist:

##### Transkript 1: Ro\_K6\_HZ\_G2a\_F36-F38 (Alessio, Jolina, Laurin, Rico)

- 01 JOL: BUSCHmesser; ((zu Alessio))  
 → 02 RIC: BUSCHmesser find ich vielleicht noch wIchtig;  
 → 03 ALE: (-) villecht [kame mit dem auch HOLZ] hacke und so;  
 04 RIC: [um TIERE zu killen; ]

Das von Jolina thematisierte Buschmesser (01) wird von Rico aufgegriffen und um seine persönliche Meinung ergänzt ("Buschmesser find ich vielleicht noch wichtig", 02). Er greift dabei auf verschiedene sprachliche Formen zurück, die eine modalisierende Funktion beinhalten. So verwendet er zum Beispiel die Partikel "vielleicht", die statt eines Fakttes lediglich eine Möglichkeit oder auch Mutmaßung nachzeichnet, welche verhandelbar ist und eine noch nicht definitive eigene Meinungsfindung suggeriert (epistemischer Unschärfemarker, Bergmann 2017:176). Auch die für das Schweizerdeutsche typische Partikel "noch" drückt in diesem Zusammenhang eine Abtönung der Äußerung aus.

Der modalisierende Charakter dieser Sequenz setzt sich auch nach dem Sprecherwechsel fort. In der ko-konstruierten Ergänzung einer Begründung von Alessio ("vielleicht kann man mit dem auch Holz hacken und so", 03), zeigt sich ebenfalls ein Modalisierungsaufwand (markiert durch "vielleicht" und "oder so"). Alessio recycelt mit der Verwendung von "vielleicht" Strukturen seines Gesprächspartners und stellt seine Begründung damit nicht nur als ko-konstruktiven Teil von Ricos Statement dar, sondern er zeigt – durch die formal und funktional ähnliche Modalisierung – auch seine Kooperativität (siehe z. B. "behavioral alignment", Pickering/Garrod 2004). Dies macht seine Äußerung ebenfalls verhandel- und erweiterbar und stellt – trotz der Hervorbringung eines inhaltlich elaborierten Beitrags – ein *downgrading* des eigenen epistemischen Status dar, welches dem Partner sein face-schonendes Agieren zusichert (vgl. Grundler 2011:295; Locher 2004:113). Sowohl die ko-konstruierte Begründung als auch die Verwendung modalisierender Partikeln zeigt Alessios Bemühungen, eine assertive Gesprächsmodalität aufrechtzuerhalten und dessen "concern for striking the balance between independence and affiliative cooperation" (Rafal 1996:287).

Die Gespräche der Klasse 2 zeigen sich hingegen – so unser Eindruck aufgrund der qualitativen Analyse – häufiger durch "Absolutheitsmarkierungen" in den Propositionen, teilweise mit Markierung einer epistemischen Autorität, z. B.:

### Transkript 2: Ro\_K2\_WB\_G2a\_F1-F3 (Sven, Cecilia, Lara, Björn)

01 SVE: aso ich glaub ich WÄISS wa mer [drIngend müend usenÄÄ;],  
**also ich glaub ich weiß was wir dringend herausnehmen müssen**

sowie Aufschlagswechsel (Knoblauch 1995:122; Morek 2015), durch den sich Dissens verhärtet:

05 BJÖ: NEI: ((noch mit Bezug auf das Feuerwerk))  
**nein**

06 CEC: NÄ:- ((noch mit Bezug auf das Feuerwerk))  
**nein**

07 SVE: dOch s FÜRwerch. ((tippt mehrmals auf die Leuchtraketen))  
**doch das Feuerwerk**

08 BJÖ: lUag ich zEig dir mol; ((mit erhobenem Zeigefinger))  
**schau ich zeig dir mal**

09 es ZÄLT de hEmmer, ((tippt mehrmals auf das Zelt))  
**ein Zelt dann haben wir**

10 CEC: denn chÖmmer SCHLA (.) denn chönd- ((zuckt mit den Schultern))  
**dann können wir schla dann könnt**

11 SVE: Aber ES (.) me wäiss jo nId ob,  
**aber es man weiß ja nicht ob**

Bei der Auswertung der Gespräche der jüngeren Kinder wird wie hier deutlich, dass diese insgesamt weniger Modalisierungsaufwand betreiben und dass das Argumentieren stärker kompetitive, wenig abgeschwächte Züge aufweist.

So konnte vor allem für die Klassen 4 und 6 ein häufiger Gebrauch von Modalisierungen im Zusammenhang mit Begründungshandlungen beschrieben werden (siehe folgende Ausführungen zu quantitativen Ergebnissen). Die älteren Kinder markieren in ihren supportiven Begründungen durch den Gebrauch verschiedener Modalisierungen nicht nur explizit, dass es sich um einen Vorschlag zu einer Plausibilisierung handelt, sondern dass dieser auch in einem gemeinsamen Für und Wider verhandelbar ist. Laut Kotthoff ist dies "typisch für den Rasoniermodus und die Entwicklung einer gemeinsamen Haltung" (Kotthoff 2015:86). Der Gebrauch von Modalisierungen zeugt daher vom Kooperationsstreben der Akteure – in den Daten der höheren Klassen konnte zudem festgestellt werden, dass auf bereits modalisierende initiale Äußerungen in den meisten Fällen supportive Begründungen folgen, die ihrerseits ebenfalls nahezu mimetisch und mit *face*-schonender Funktion modalisiert werden (siehe auch Transkriptbeispiel 1).

Das Interesse bestand nun darin, zu überprüfen, inwiefern im Bereich der Modalisierung und Perspektivierung Altersunterschiede über das ganze Korpus hinweg festzustellen sind, ob möglicherweise die drei getesteten Settings einen Einfluss nehmen und ob es Unterschiede zwischen Jungen und Mädchen gibt.

Im folgenden Abschnitt werden zunächst die angewandten Verfahren erläutert, mit Hilfe derer diese Fragestellungen quantitativ untersucht wurden.

### ***Quantitative Verfahren***

Zur Untersuchung der Verwendung von Modalisierung und Perspektivierung wurde die Häufigkeit von oberflächennahen, lexikalischen Einheiten (Konjunktivformen und abschwächende Modalpartikeln) ermittelt<sup>25</sup> und auf Unterschiede zwischen Klassenstufen, Setting und Geschlecht getestet. Da im Zentrum immer das Argumentieren steht, wurden nur diejenigen Vorkommnisse einbezogen, die entweder in Positionierungshandlungen (Code 1) oder Begründungshandlungen (Code 3a) auftraten. Der Vergleich der Häufigkeiten wurde jeweils getrennt für Positionierungs- und Begründungshandlungen durchgeführt. Es handelte sich bei dieser Auswertung also um einen korpuslinguistischen Zugang, der aber zusätzlich auf spezifizierte Sprachhandlungen Bezug nahm, die durch Codes identifiziert wurden.<sup>26</sup> Für den Vergleich der Häufigkeiten wurden unterschiedliche statistische Verfahren angewandt, die im Folgenden kurz erläutert werden (siehe auch Lijffijt et al. 2014).

Für den Vergleich von Häufigkeiten bietet sich zunächst der Pearson Chi-Quadrat-Test an. Das Verfahren testet, ob ein signifikanter Zusammenhang zwischen

<sup>25</sup> Ausgewertet wurden die Konjunktivformen *würd(en)*, *könnt(en)*, *sollt(en)*, *hätt(en)*, *müsst(en)* und die abschwächenden Modalpartikeln *vielleicht*, *eher*, *manchmal*, *nicht so*, *eigentlich*. Für die Auswertung wurden jeweils die standarddeutschen Übersetzungen genutzt, um bei der Suchabfrage nicht auf dialektale Varianzen Rücksicht nehmen zu müssen.

<sup>26</sup> Ein vergleichbares Vorgehen kam auch bei der Frage nach der Explizitheit von Begründungshandlungen zur Anwendung. Dort interessierten uns Häufigkeiten von explizit sprachlich markierten Begründungszusammenhängen, dies differenziert nach kausalen (z. B. weil, deshalb, darum), finalen (z. B. damit, dass, um...zu) konditionalen (z. B. wenn...dann, falls, sofern) und weiteren (Restgruppe) Markierungen.

zwei nominalskalierten Variablen besteht. In unserem Fall ist das zum einen eine Variable, die für jedes Element (Wort oder Sequenz) angibt, ob es der jeweiligen Suchkategorie entspricht oder nicht (ist ein Wort eine Modalpartikel bzw. kommt in der Sequenz eine solche vor), zum anderen die Gruppe (z. B. Klassenstufe). Anhand dieser Variablen wird eine Kreuztabelle erstellt, welche die Häufigkeitsverteilung in den einzelnen Gruppen repräsentiert. Der Chi-Quadrat-Test vergleicht diese beobachteten Häufigkeiten mit denjenigen Häufigkeiten, die erwartet werden, wenn die Gruppe keinen Einfluss auf die Häufigkeitsverteilung hat. Weichen beobachtete und erwartete Häufigkeiten deutlich voneinander ab, kann davon ausgegangen werden, dass ein Zusammenhang besteht. Bei großen Fallzahlen wird der Chi-Quadrat-Test allerdings auch schon bei relativ kleinen Unterschieden signifikant. So treten in unseren Analysen mehrfach signifikante Effekte auf, während aber die Effektstärke, also das Maß für die Stärke des Zusammenhangs (in unserem Fall wurde Cramers V berechnet) auf einen nur schwachen Zusammenhang hinweist. Dies hängt auch damit zusammen, dass als Vergleichswert für die Häufigkeit der Vorkommnisse alle Wörter in den durchsuchten Sequenzen dienen.<sup>27</sup>

Um die Ergebnisse des Häufigkeitsvergleichs zu validieren, wurde daher zusätzlich ermittelt, in welchen der Gespräche überhaupt einer der Suchbegriffe auftritt. Auch diese Häufigkeiten können dann mithilfe des Chi-Quadrat-Tests im Hinblick auf den Einfluss der Gruppe verglichen werden. Dies setzt natürlich voraus, dass in einer hinreichend großen Anzahl der Gespräche ein Suchbegriff gar nicht auftritt, was in unserem Korpus aber der Fall ist. Da hier die Bezugsgröße der Häufigkeit die Anzahl der Gespräche ist, bleiben die Fallzahlen deutlich niedriger. Wenn nun beide Verfahren in die gleiche Richtung weisen, kann dies als zusätzliche Validierung der Resultate dienen.

Der Chi-Quadrat-Test ist ein sogenannter Omnibus-Test. Ein signifikantes Ergebnis deutet also nur darauf hin, *dass* ein Zusammenhang besteht, sagt aber nichts darüber aus, welche Gruppen sich voneinander unterscheiden (was insbesondere bei mehr als zwei Gruppen relevant ist). Problematisch ist unter Umständen auch, dass die sehr unterschiedliche Länge der Gespräche die Resultate verzerrt, etwa wenn in einer Gruppe ein besonders langes Gespräch vorliegt.

Eine Möglichkeit, der Heterogenität der Daten besser Rechnung zu tragen, besteht darin, für jedes Gespräch die normierte Häufigkeit des Suchbegriffs (z. B. Anzahl Modalpartikeln pro hundert Wörter) zu ermitteln. Die normierten Häufigkeiten bzw. die Gruppenmittelwerte können dann verglichen werden, wobei den Unterschieden in der Gesprächslänge durch die Normierung Rechnung getragen wird. Da sich in den meisten Fällen die Daten innerhalb der Gruppen nicht als normalverteilt betrachten ließen, wurden hierfür nicht-parametrische Tests angewandt, die keine spezifischen Verteilungseigenschaften der Werte voraussetzen. Im Falle von zwei Gruppen kam hier der Mann-Whitney-U-Test, bei mehr als zwei Gruppen der Kruskal-Wallis-Test zur Anwendung. Um insbesondere bei mehr als zwei Gruppen zu ermitteln, zwischen welchen Gruppen signifikante Unterschiede bestehen, wurden anschließend paarweise Vergleiche mithilfe des Dunn-Tests (korrigiert für multiple Vergleiche nach Holm) vorgenommen. In einigen Analysen zeigte

---

<sup>27</sup> Eine Möglichkeit, hier Abhilfe zu schaffen, bestünde prinzipiell darin, nur diejenigen Positionen zu berücksichtigen, an denen die Suchbegriffe sinnvollerweise vorkommen können. Dies ist allerdings aufgrund des Gesprochensprachlichen kaum objektiv realisierbar.

sich ein deutlicher Zusammenhang insgesamt, in den paarweisen Vergleichen jedoch nur ein Unterschied zwischen Klasse 2 und 4, jedoch kein nennenswerter Unterschied zwischen Klasse 4 und 6. In solchen Fällen wurden die Klassen 4 und 6 zu einer Gruppe zusammengefasst und mit Klasse 2 verglichen, um den 'Sprung' zwischen diesen Klassenstufen zu validieren. Im weiteren Verlauf werden, wenn nicht anders angegeben, die Ergebnisse des soeben beschriebenen Verfahrens angegeben.

Eine weitere Analysemöglichkeit ist die Anpassung eines Regressionsmodells an die Daten. Bei einer Regressionsanalyse werden die Parameter einer mathematischen Funktion ermittelt, die die Daten bestmöglich beschreibt. In unserem Fall interessiert, inwieweit das Auftreten eines Ereignisses (also etwa das Auftreten eines Konjunktivs oder einer Modalpartikel) von den jeweiligen Faktoren wie etwa der Klassenstufe abhängt. Wir interessieren uns also für die Auftretenswahrscheinlichkeit eines Ereignisses in Abhängigkeit von bestimmten Faktoren. Da die abhängige Variable damit binär ist (d.h. eine Eigenschaft, etwa das Vorliegen einer Modalpartikel in einer Äußerung, trifft entweder zu oder nicht), kommt hier eine logistische Regression zum Einsatz. Um der Heterogenität der Daten Rechnung zu tragen, wurde eine gemischte logistische Regression angewandt, die es erlaubt, Zufallsfaktoren im Modell zu berücksichtigen. Dies ermöglicht, neben den sogenannten festen Faktoren (also etwa Klassenstufe) auch diejenigen Faktoren zu berücksichtigen, die ebenfalls auf die Auftretenswahrscheinlichkeit Einfluss haben können, die sich jedoch aus der zufälligen Wahl der Stichprobe ergeben. In unserem Fall wurden sowohl das jeweilige Gespräch als auch die/der individuelle SprecherIn als Zufallsfaktoren ins Modell aufgenommen. Damit verringert sich die Gefahr, einen Einfluss der festen Faktoren anzunehmen, der tatsächlich in erster Linie durch Unterschiede innerhalb der Gruppen zustande kommt. In der Regel wird zunächst ein 'Basismodell' an die Daten angepasst, welches den interessierenden Faktor (z. B. Klassenstufe) nicht berücksichtigt. Dieses Basismodell wird dann mit einem Modell verglichen, in welchem der Faktor enthalten ist. So kann ermittelt werden, ob der interessierende Faktor dazu führt, dass die Struktur der Daten besser erklärt werden kann, wenn der interessierende Faktor in der Regressionsformel enthalten ist.

Im Folgenden werden die Ergebnisse für einige Aspekte, bei denen sich deutliche Unterschiede zeigten, berichtet. Es zeigte sich, dass die unterschiedlichen Verfahren in der Regel ähnliche Muster zeigten. Daher werden wir nicht die Gesamtheit der Resultate der unterschiedlichen Verfahren ausführen. Es kann jedoch davon ausgegangen werden, dass für die präsentierten Ergebnisse die anderen Verfahren ebenfalls in dieselbe Richtung weisen.

### ***Ergebnisse***

Die deutlichsten Ergebnisse zeigen sich beim Gebrauch von *Konjunktiven* im Altersvergleich. Sowohl in Positionierungs- als auch Begründungshandlungen bestätigen alle oben beschriebenen Verfahren einen signifikanten Zusammenhang von Klassenstufe und Verwendungshäufigkeit konjunktivierter Formen (alle  $p < 0.05$ ). Die paarweisen Vergleiche weisen auf Unterschiede zwischen Klasse 2 und 4 sowie Klasse 2 und 6 hin, während sich zwischen Klasse 4 und 6 kein signifikanter Unterschied ergab. Klasse 4 und 6 wurden daher zu einer Gruppe zusammengefasst

und mit Klasse 2 verglichen. Der Vergleich der über Gespräche gemittelten Häufigkeit pro 100 Wörter zeigt eine signifikante Zunahme von Konjunktivformen von Klasse 2 (Mittelwerte Positionierung: 2.31; Begründung 0.92) zu Klasse 4/6 (Mittelwerte Positionierung: 2.79; Begründung: 1.24; Mann-Whitney-U-Test: Positionierung:  $U = 2738.5$ ,  $p < 0.01$ , Begründung:  $U = 2597$ ,  $p < 0.01$ ).

Auch beim Gebrauch von *abschwächenden Modalpartikeln* in Positionierungshandlungen besteht ein signifikanter Zusammenhang von Häufigkeit und Klassenstufe (alle  $p < 0.05$ ). Die paarweisen Vergleiche zeigen, dass die Häufigkeit pro 100 Wörtern in Klasse 6 mit durchschnittlich 1.01 signifikant höher ist als in der Klasse 2 mit durchschnittlich 0.62 (Dunn-Test:  $Z = -2.61$ ,  $p < 0.05$ , korrigiert nach Holm). Die Unterschiede zwischen Klasse 2 und 4 sowie zwischen Klasse 4 und 6 sind nicht signifikant. Auffällig ist auch der hohe Anteil an Gesprächen, in denen gar keine abschwächenden Modalpartikeln in Positionierungshandlungen geäußert wurden: In der Klasse 2 sind dies 47% und in den Klassen 4 und 6 noch 21% resp. 20% (Pearson Chi-Quadrat-Test:  $\chi^2 = 12.89$ ,  $p < 0.01$ , Cramer's  $V = 0.27$ ).

Weiter wurden auch Suchabfragen durchgeführt, die Positionierungs- und Begründungshandlungen identifizieren, in denen sowohl (mindestens) eine Konjunktivform als auch (mindestens) eine abschwächende Modalpartikel vorkommt. Denn bei kombinierter Verwendung kann eindeutiger bestimmt werden, ob die Äußerung modalisierend bzw. perspektivierend ist. Bei den Klassenstufen zeigt sich ähnlich wie zuvor, dass sich die Klassen 4 und 6 nur unwesentlich unterscheiden. Sie werden daher gemeinsam betrachtet und mit Klassenstufe 2 verglichen. Als Normalisierungsgröße konnte bei dieser Berechnungsvariante nicht mehr die Anzahl Wörter fungieren, da wir nach kombinierten Vorkommnissen innerhalb einer annotierten Sequenz suchen (Vorkommen von Konjunktivform und abschwächender Modalpartikel in einer annotierten Sequenz mit Positionierung bzw. mit Begründung). Deshalb diente hier die Anzahl annotierter Sequenzen als Normalisierungsgrundlage. In den Abbildung 2-a und 2-b (auf der nächsten Seite) ist ersichtlich, dass in Positionierungs- und auch Begründungshandlungen der Anteil von Äußerungen, die sowohl Konjunktivformen als auch abschwächende Modalpartikeln enthalten, deutlich zunimmt.

In den Abbildung 2-a und 2-b repräsentieren die beiden mittleren Balken die gemittelte Häufigkeit von Sequenzen mit sowohl Konjunktivform(en) als auch Modalpartikel(n) pro 100 annotierte Sequenzen, einerseits in Sequenzen mit Positionierungshandlungen (Code 1, Abbildung 2-a) und andererseits in Sequenzen mit Begründungshandlungen (Code 3a, Abbildung 2-b). Die Mittelwerte wurden über alle Settings berechnet. Die Fehlerbalken bezeichnen den Standardfehler, der Aufschluss über die Streuung der Daten gibt. Zusätzlich ist die Verteilung der Daten jeweils als Histogramm dargestellt, das Aufschluss darüber gibt, wie hoch der Anteil der Gespräche ist, in welchen die entsprechende Anzahl von Sequenzen mit Konjunktivform(en) und Modalpartikel(n) auftritt. Die Grafik ist demnach so zu lesen, dass z. B. in der Klasse 2 pro 100 Positionierungssequenzen durchschnittlich 1.2 Sequenzen sowohl einen Konjunktiv als auch eine Modalisierung enthalten, in den Klassen 4 und 6 dagegen 2.1 Sequenzen pro 100 Positionierungssequenzen. Zudem wird anhand der Histogramme ersichtlich, dass in Klasse 2 jeweils deutlich mehr Gespräche vorkommen, in welchen die Kombination von Modalisierung und Konjunktiv gar nicht auftritt (Histogrammbalken beim Wert 0) (Pearson Chi Quadrat-Test:  $\chi^2 = 7.0$ ,  $p < 0.01$ , Cramer's  $V = 0.20$ ).

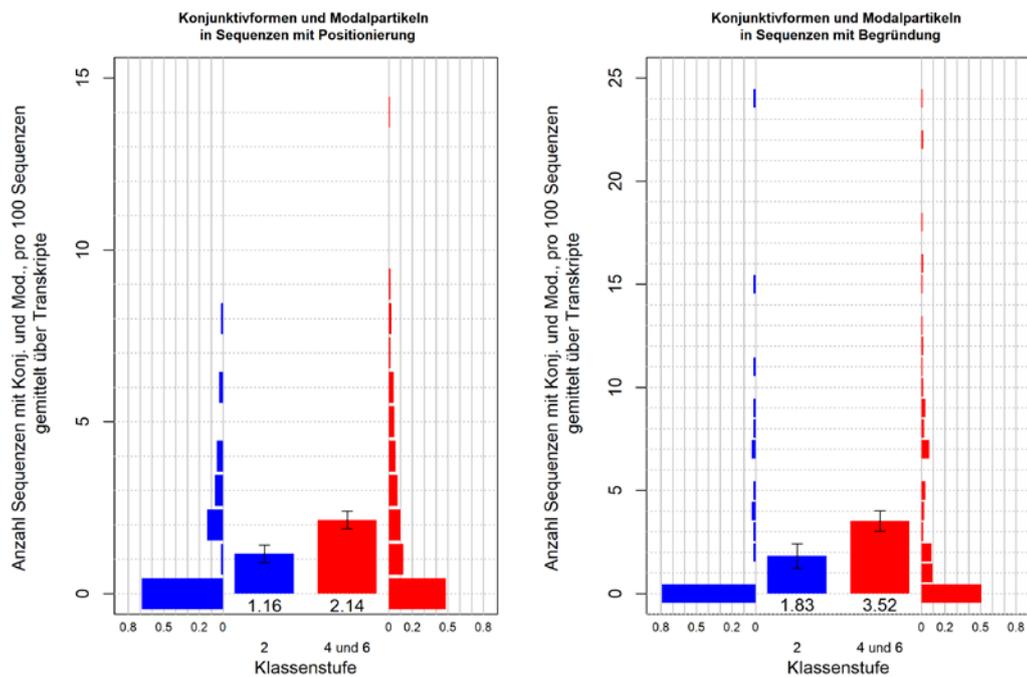


Abbildung 2-a (Positionierungshandlungen) und 2-b (Begründungshandlungen):  
Konjunktivformen und abschwächende Modalpartikeln in Klasse 2 versus Klasse 4 und 6

Die Verwendung von Äußerungen mit Konjunktivformen *und* abschwächenden Modalpartikeln verdoppelt sich von Klasse 2 zu Klasse 4/6 in beiden Handlungstypen und die Unterschiede sind signifikant (Mann-Whitney-U-Test: Positionierungshandlungen:  $U = 2840$ ,  $p < 0.05$ ; Begründungshandlungen:  $U = 2504.5$ ,  $p < 0.01$ ). Wenn wir diese kombinierte Auswertung mit den Einzelauswertungen oben vergleichen, fällt auf, dass sich unterschiedliche und teils widersprüchliche Interpretationsspielräume auftun: Gemessen an der Anzahl Wörter zeigt sich nur bei den Positionierungshandlungen ein signifikanter Anstieg des Gebrauchs abschwächender Modalpartikeln von Klasse 2 zu 6, nicht aber bei den Begründungshandlungen. Zudem sind die konjunktivierten Anteile pro 100 Wörter bei Positionierungshandlungen höher als bei Begründungshandlungen (vgl. weiter oben). Hingegen zeigt sich bei den kombinierten Suchabfragen (Konjunktivformen *und* abschwächende Modalpartikeln), gemessen an der Anzahl annotierter Sequenzen, dass sich bei den Begründungshandlungen durchaus ein Effekt zeigt und die Anteile konjunktivierter *und* modalisierter Äußerungen bei den Begründungshandlungen deutlich höher liegen als bei den Positionierungshandlungen (vgl. Abbildung 2-b versus 2-a).

Durch diese sich widersprechenden Ergebnisse aufgrund der unterschiedlichen Normalisierungen der Daten zeigen sich Grenzen eindeutiger statistischer Auswertungen für codierte Gesprächsdaten. Auch wenn die Verteilung von Modalpartikeln und Konjunktiven auf die einzelnen *Sprechhandlungen* widersprüchlich ist, so ist das Ergebnis hinsichtlich der Verteilung auf die einzelnen *Klassenstufen* eindeutig. Durch einen (erneuten) qualitativen Blick auf die Daten mit dem Fokus auf die tatsächliche, kontextabhängige Verwendung von Modalpartikeln und Konjunktiven könnten mögliche Gründe für ihren vermehrten Gebrauch in den höheren Klassen eruiert werden. Die qualitative Analyse (vgl. Transkriptbeispiele 1 und 2) legt nahe,

dass die älteren Kinder zunehmend ein Bewusstsein für doppelte und damit redundante Markierungen entwickeln – dies u.a. auch deshalb, da sie stark an die vorgängige Äußerung des Gesprächspartners anknüpfen und dessen Modalisierungen 'recyclen', gleichzeitig aber auch die selber eingebrachte Idee im Sinne des etablierten Gesprächscharakters hervorbringen. Sicher sind die Kinder auch durch den fortschreitenden allgemeinen Spracherwerb zunehmend in der Lage, Konjunktive zu verwenden – die Daten zeigen jedoch, dass sie dies auch schon im Alter von 8 Jahren tun. Das Interessante ist somit, dass der zunehmende Gebrauch offensichtlich nicht (nur) im isolierten Erwerb eines spezifischen Konstrukts aus Wortschatz und Grammatik begründbar ist, sondern im Kompetenzerwerb auf kontextuell-interaktionaler Ebene argumentativer Diskurseinheiten. Damit gemeint ist die situationsangemessene Verwendung von Konjunktiven und Modalisierungen zur funktionalen Lösung der Aufgabe (nämlich: Finden eines gemeinsamen Konsenses) im aktuellen Gesprächsvollzug – mit den verallgemeinerten Worten Hartungs (2004:50, formale Anpassung: die AutorInnen) zu Gesprächskompetenz:

Gesprächskompetenz ist die Fähigkeit, zu einem beliebigen Zeitpunkt in einem Gespräch zu einer angemessenen Einschätzung der aktuellen Situation und der lokalen Erwartungen der Gesprächspartner zu kommen, auf dem Hintergrund dieser Einschätzung eine den eigenen Interessen und den eigenen Ausdrucksmöglichkeiten angemessene Reaktion mit hoher Erfolgswahrscheinlichkeit zu finden und diese Reaktion der eigenen Absicht entsprechend körperlich, stimmlich und sprachlich zum Ausdruck zu bringen.

Analysen zur Untersuchung eben jener Gesprächskompetenzen können von quantitativen Analysen nur partiell geleistet werden und fallen somit wieder in den Hoheitsbereich der qualitativen Gesprächsanalyse. Die quantitative Verteilung von Konjunktiven und Modalisierungen gibt uns aber zum Beispiel Aufschluss darüber, welche Tendenzen zum Verständnis von Argumentieren (konvergent – divergent) in der untersuchten Altersspanne vorliegen, da durch den häufigen Gebrauch von Konjunktiven und Modalisierungen in den Klassen 4 und 6 u.a. deutlich wird, in welchem Maße eigene Sichtweisen als verhandelbar markiert, andere Perspektiven berücksichtigt und face-schonende Strategien verwendet werden.

Betrachten wir die Konjunktivformen und die abschwächenden Modalpartikeln nun im Settingvergleich, so zeigen sich unterschiedliche Effekte. Bei den Konjunktivformen gibt es bei den Positionierungshandlungen zwischen dem Robinsonsetting und dem Spendensetting (gruppiert Sm und So) keinen Unterschied, jedoch finden sich im Spendensetting ohne Handlungskonsequenz (So) signifikant mehr Konjunktivformen (Mittelwert: 3.0) als im Spendensetting mit Handlungskonsequenz (Sm, Mittelwert: 2.3) (Mann-Whitney-U-Test,  $U = 1366.5$ ,  $p < 0.05$ ). Da sich aber bei den abschwächenden Modalpartikeln im Settingvergleich Sm vs. So keine signifikanten Unterschiede zeigen und auch in den kombinierten Suchabfragen (Äußerungen mit sowohl Konjunktivformen als auch Modalpartikeln) keine Tests signifikant sind, muss das Resultat zu den Konjunktiven vorsichtig interpretiert werden. Denn es muss beachtet werden, dass eine Aufgabe mit fiktiver Problemstellung auch nahelegt, gewisse Entscheidungen im Konjunktivmodus auszuhandeln und insofern überrascht der erhöhte Gebrauch von Konjunktiven im So-Setting nicht. Ohne einen vergleichbaren Effekt bei den abschwächenden Modalpartikeln

oder bei den kombinierten Suchabfragen und ohne Betrachtung des Kontexts können wir daher nicht von einer intensiveren Perspektivierung und Modalisierung im Spendensetting ohne Handlungskonsequenz sprechen.

Im Gegensatz zu den Konjunktiven findet sich bei der Verwendung abschwächender Modalpartikeln ein signifikanter Unterschied zwischen dem Robinsonsetting und dem Spendensetting (siehe Abbildung 3):

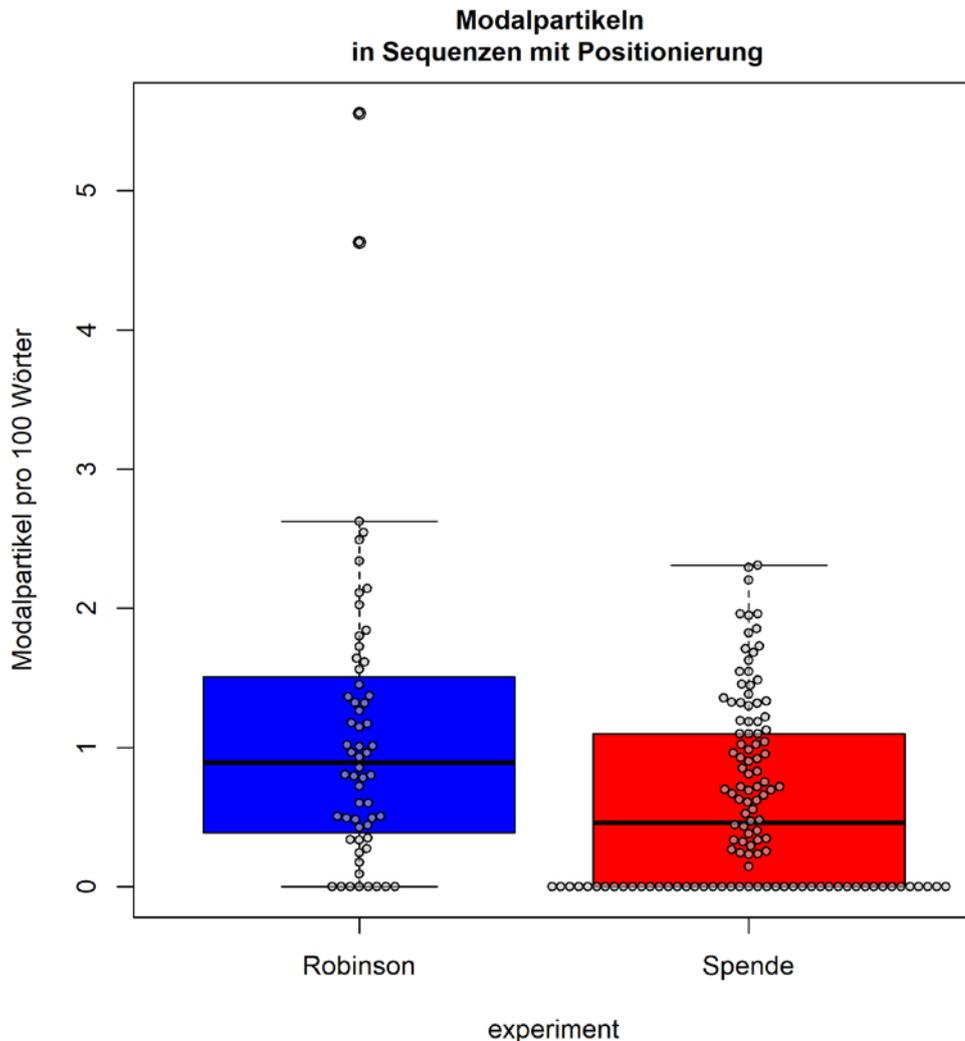


Abbildung 3: Modalpartikeln im Robinsonsetting und Spendensetting

In diesem Boxplot ist ersichtlich, dass im Robinsonsetting in Positionierungshandlungen mehr Gebrauch von abschwächenden Modalpartikeln gemacht wird als im Spendensetting (Sm und So) (Mann-Whitney-U-Test,  $U = 4660$ ,  $p < 0.01$ ). Die Linie innerhalb der Box markiert den Median.<sup>28</sup> Die ganze Box deckt jeweils die mittleren 50% der Daten ab, d.h. die Daten im 2. und 3. Quartil (25-75%). Die Antennen außerhalb der Box enden jeweils beim letzten Datenpunkt, der sich innerhalb des 1.5-fachen Interquartilsabstands (also der Länge der Box) von der Box

<sup>28</sup> Der Median gibt den Zentralwert eines Datensatzes, hier also der Wortanzahl der je 20 Gespräche, an: 50% der Daten sind größer oder gleich dem Median, 50% sind kleiner oder gleich groß. Der Median ist gegenüber Extremwerten (und damit gegenüber Ausreißern) unempfindlich.

befindet. Werte außerhalb der Antennen werden als Ausreißer bezeichnet. Je kleiner die Box ausfällt, desto näher liegen die Werte der mittleren 50% der Daten beim Median und desto kleiner ist also die Streuung.

Zusätzlich sind die Werte der einzelnen Transkripte als Punkte repräsentiert, um die Verteilung detailliert abzubilden. Hier ist auch klar sichtbar, dass im Spensetting deutlich mehr Gespräche gar keine Modalpartikeln beinhalten als im Robinsonsetting (Pearson Chi-Square-Test:  $\chi^2 = 11.25$ ,  $p < 0.001$ , Cramer's  $V = 0.25$ ).

Da sich dieser Effekt, dass im Robinsonsetting in Positionierungshandlungen mehr Gebrauch von abschwächenden Modalpartikeln gemacht wird als im Spensetting, (entgegen unseren Erwartungen) nicht beim Vergleich *Sm versus So*, sondern nur settingübergreifend zeigt, ist mit hoher Plausibilität ein Einfluss der Aufgabenstellung mit den unterschiedlichen Mengen an zu diskutierenden Gegenständen bzw. Objekten anzunehmen, die potenziell in das Gespräch auf perspektivierende Art und Weise eingebracht werden können.<sup>29</sup>

Bei Begründungshandlungen zeigt sich dann aber keinerlei Einfluss des Settings auf den Gebrauch von Konjunktivformen oder abschwächenden Modalpartikeln bzw. kombinierter Verwendung in einer Äußerung. Keiner der Tests wurde signifikant.

Und schließlich konnte keinerlei Einfluss des Geschlechts auf die Häufigkeit der Verwendung von Konjunktiven und abschwächenden Modalpartikeln bei Positionierungs- oder Begründungshandlungen nachgewiesen werden. Veranschaulichend sei dies in folgender Visualisierung dargelegt (Abbildung 4 nächste Seite).

In Abbildung 4 wird die Anzahl von Modalpartikeln pro hundert Wörter in Sequenzen mit Begründungshandlungen dargestellt und dies gruppiert nach Geschlecht. Jeder Datenpunkt stellt den Wert einer Sprecherin (links) bzw. eines Sprechers (rechts) dar. Wir sehen hier eine sehr ähnliche Verteilung bei den Mädchen und den Jungen. Bei beiden Gruppen sind die durchschnittlichen Werte in einem ähnlichen Bereich (graue Balken: weiblich 0.72, männlich 0.71) und bei beiden Gruppen gibt es Ausreißer nach oben. Auch der Anteil derjenigen SprecherInnen, die keine Modalpartikeln verwenden, unterscheidet sich kaum (siehe Balken beim Wert 0 in Abbildung 4).

Die quantitativen Analysen zeigen uns, ob und in welchem Maße ausgewählte Variablen einen Einfluss auf den Sprachgebrauch haben. Wir sehen beispielsweise, dass es in der 4. und 6. Klasse für die argumentationsrelevante Sprachhandlung des Positionierens wahrscheinlicher ist, Konjunktivformen und abschwächende Modalpartikeln zu verwenden als in der 2. Klasse, dass sich aber keine signifikanten Unterschiede zwischen der 4. und 6. Klassenstufe ergeben. Bezogen auf (faktische) argumentative Gesprächskompetenz bedeutet dies, dass viele Kinder zwischen der 2. und 4. Klasse (also zwischen 8 und 10 Jahren) lernen, sprachliche Mittel zur Markierung von Perspektivenmodalisierungen gezielt in Kombination einzusetzen. Wir sehen auch in aller Deutlichkeit, dass generell in Bezug auf Konjunktive und abschwächende Modalpartikeln keine geschlechtsspezifischen Unterschiede vorherrschen, wenn man auf das gesamte Korpus rechnet. Demnach können die Berechnungen auch als Korrektiv fungieren, da in unserem Fall die zuerst durchgeführten qualitativen Analysen einzelner Gespräche den Eindruck vermittelten, dass Mädchen stärker als Jungen Gebrauch von Modalpartikeln und Konjunktiven machen. Geht man mit dem quantitativen Befund, dass es keine Unterschiede zwischen

<sup>29</sup> So gibt es im Robinsonsetting mehr Gegenstände zu diskutieren.

Jungen und Mädchen gibt, noch einmal in die qualitative Analyse, zeigt sich allerdings, dass in Gruppen, die nur aus Jungen bestehen, häufig epistemische Autorität (Kirkham 2011) markiert wird, während in reinen Mädchengruppen oft konsensuell argumentiert wird. Unser ursprünglicher Eindruck in der qualitativen Analyse wird von den quantitativen Auszählungen widerlegt, es zeichnet sich aber in einer darauf basierenden nochmaligen qualitativen Analyse ein Effekt einer bezüglich des Geschlechts homogenen Gruppenzusammensetzung ab.

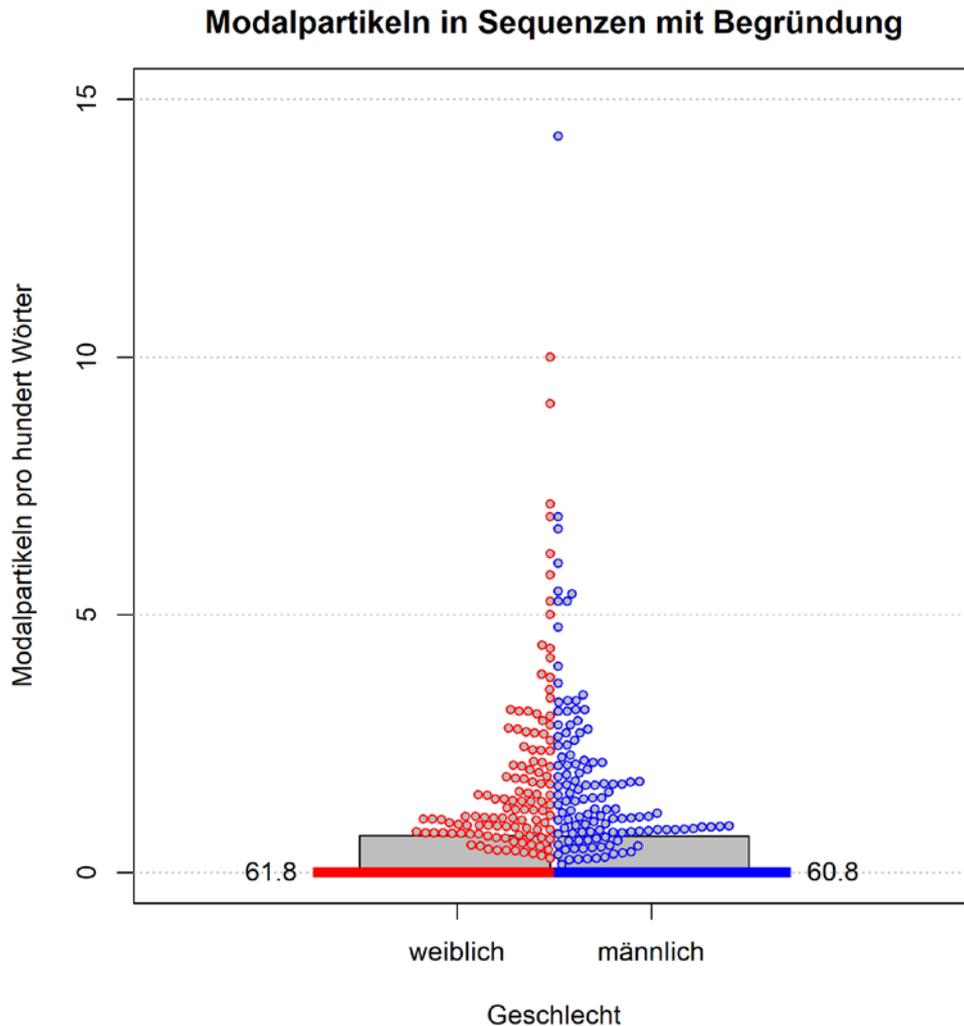


Abbildung 4: Abschwächende Modalpartikeln pro hundert Wörter in Sequenzen mit Begründungshandlungen bei Mädchen und Jungen. Jeder Datenpunkt repräsentiert eine/n SprecherIn; die grauen Balken repräsentieren die Gruppenmittelwerte; die farbigen Balken unten repräsentieren den prozentualen Anteil an SprecherInnen, die keine Modalpartikeln verwendeten.

Dies weist noch einmal darauf hin, dass die Auszählungen in diesem Fall losgelöst von der kontextuellen Einbettung durchgeführt wurden. Da Modalpartikeln kontextbedingt multifunktional sein können, müssen wir davon ausgehen, dass wir bei einer qualitativen, fallbasierten Analyse zu leicht anderen Ergebnissen kommen können. So finden wir beispielsweise in einem Gespräch auf Klassenstufe 4 die Äußerung: "ich würd villicht uf jede fall mal e SCHLAFsack;" ('ich würde vielleicht auf jeden Fall mal einen Schlafsack'; Ro\_K4\_HZ\_G3b, 02:38). Hanna kombiniert in dieser Positionierung den Konjunktiv *würde*, die abschwächende Modalpartikel

*vielleicht* sowie die verstärkende Modalpartikel *auf jeden Fall* (sie verwendet dasselbe Muster wiederholt bei ihren weiteren Vorschlägen). Da wir Letztere jedoch nicht in den quantitativen Berechnungen berücksichtigt haben, zählt dieses Beispiel zu den perspektivierten, modalisierten Äußerungen. Schaut man aber ins Gespräch, scheint die Verwendung widersprüchlich und die Zuordnung nicht mehr so deutlich. Entsprechend wichtig ist es, die Ergebnisse auf Basis von qualitativen Analysen wie dem obigen Beispiel von Hanna zu kontextualisieren, um sie umfassender zu verstehen.

#### 4.1.2. Komplexität von Argumentationen

Aus unseren bisherigen qualitativen Analysen geht hervor, dass in der 2. Klasse Themen nur selten über mehrere Redebeiträge vertieft diskutiert werden. Vielmehr treffen wir viele argumentativ isolierte Äußerungen an. Ab der 4. Klasse beteiligen sich die Kinder dann vermehrt gemeinsam an längeren argumentativen Einheiten. Auffällig ist zudem, dass das explorative (versus persuasive) Argumentieren (vgl. Ehlich 2014; Nonnon 1996) an Bedeutung gewinnt und die Kinder häufiger Begründungen oder auch längere Argumentationen ko-konstruieren (vgl. Jacoby/Ochs 1995). Wir illustrieren dies zunächst mit einem Beispiel aus unseren qualitativen Analysen.

Das folgende Beispiel zeigt eine argumentative Sequenz von vier Mädchen, die bereits beim Beginn der Diskussion Begründungshandlungen realisieren und – in vollkommener Übereinstimmung – im weiteren Verlauf fortfahren, gegenseitig Stützen zu ergänzen. Wir haben es hier mit einem gemeinsamen Begründen zu tun, das auch für die Beziehungsarbeit bzw. die Identitätsgestaltung relevant ist:

#### Transkript 3: Ro\_K6b\_PB\_G2b\_F29 (Tanja, Leonie, Juliana, Rana)

01 JUL: vielleicht das SACKmesser; ((zeigt auf das Sackmesser))  
 02 weisst du (.) [zum FRÜCHte ] ((imitiert das  
 Schneiden mit einem Taschenmesser))  
 03 RAN: [das SACKmesser das:-]  
 04 TAN: doch sAckmässer isch scho t find ich wichtig wäil das het  
 VIIL.  
 05 JUL: ja Eben [(.) für] ALles (komm),  
 06 LEO: [JÄ, ]  
 07 RAN: ((nickt)) und vIelleicht hats auch SO:- (.)  
 08 aso jetzt sag ich mal FLAschenöffner,  
 09 (weil) es hat so auch viele ANdere sachen; ((macht  
 Schraub- und Drehbewegungen mit den Händen))  
 10 TAN: JA;  
 11 JUL: JA;  
 12 LEO: und BUSCHmesser [kannst du Eins ] (.) ja;  
 13 TAN: [also Ich würd SACKmässer;]  
 14 LEO: Eigentlich MACHen;  
 15 JUL: jA denn kannst ja nur für DA:S; ((schwingt ein imaginäres  
 Buschmesser durch die Luft))  
 16 LEO: oder du kAnnst eins machen mit HOLZ,  
 17 RAN: JA;  
 18 JUL: Eben;

Die Sequenz wird durch Julia eröffnet, indem sie den thematischen Fokus durch ihren Vorschlag "vielleicht das Sackmesser" (01) auf ein neues Thema lenkt. Sie schließt selber eine Begründung an ("weißt du zum Früchte", 02), die sie zusätzlich gestisch illustriert. Hier wird die kommunikative Praktik des Argumentierens – als Explanat<sup>o</sup>n markiert – etabliert und auch von einer zweiten Sprecherin übernommen, indem diese ihre Äußerung nicht nur bei einer Zustimmung ("doch", 04) und einer eigenen Stellungnahme ("find ich wichtig", 04) belässt, sondern um eine sprachlich explizit markierte Begründung ergänzt ("weil das hat viel", 04). Mit der ko-konstruierten Begründung zu Julias Vorschlag etablieren die beiden Mädchen Konsens.

Dennoch werden beide thematischen Felder (Busch- und Sackmesser) im Folgenden von den anderen Kindern aufgegriffen und weiterbearbeitet. So bekräftigt Julia Tanias Argument zunächst durch eine Zustimmungspartikel und dann durch eine expandierende Stütze ("ja eben für alles", 05).

Bewegen sich beide Argumente zunächst noch auf einer recht abstrakten Ebene durch die unbestimmten Zahlwörter "viel" und "alles", werden sie in einem der nächsten Turns konkretisiert: eine vierte Sprecherin, Rana, führt eine weitere Stütze an, indem sie in den Zeilen 08 und 09 zwei Beispiele benennt und gestisch illustriert ("und vielleicht hats auch so also jetzt sag ich mal Flaschenöffner", 08 und "weil es hat so auch viele andere Sachen", 09). An dieser Stelle wäre diese argumentative Diskurseinheit abgeschlossen, da von allen Kindern Konsens signalisiert wird. Jedoch suchen die Mädchen weiterhin nach Argumenten. Diese beziehen sich zwar auf einen anderen, jedoch ähnlichen Gegenstand (Buschmesser) und werden herangezogen, um den eigentlichen Fokusgegenstand (Sackmesser) weiterhin zu plausibilisieren und neue Lösungsmöglichkeiten gemeinsam zu prüfen ("und Buschmesser kannst du eins ja eigentlich machen", 12, 14, 16).

Auffällig an dieser Sequenz ist, dass hier keine genretypischen Pro-Kontra-Abfolgen vorliegen, sondern sich eine über mehrere Turns hinziehende explorativ-argumentative Themenentwicklung durch die vier Kinder ergibt. Diese Sequenz zeichnet sich durch die Addition mehrerer stützender Begründungen für die initiale Proposition aus. In der analysierten Sequenz wird deutlich, wie die Kinder argumentativ – durch stützende Expansionen der Argumente der Gesprächspartnerinnen – ihre Zustimmungen bzw. ihre gemeinsame Einstellung explizieren und Konsens verstärken.

Bei den Kindern der zweiten Klasse ist im Gegensatz dazu augenscheinlich, dass diese ihre Positionen hervorbringen (i.S. eines Behauptens), jedoch keine weiteren Begründungen oder gar gegenseitige Stützen dafür anführen. Dazu folgendes Beispiel aus der Anfangssequenz eines Gesprächs:

#### Transkript 4: Ro\_K2\_SA\_G2a (Yara, Selvanila, Jonathan, Finn)

```

01 SEL      °h Oder (.) (de:nn) [dIese] BUSCHmesser,
02 FIN                                [hm- ]
03 SEL      (.) das kann AU: (.) vilI::[ch(t)],
04 FIN                                [busch],
05 YAR      jä °h strEichhÖ:lz (.) Oder f::FEUerzeug;
06 SEL      °h [Und noch] ein !ZELT! brAuch(en) wir;
07 JON      [aso ICH,]
08          jä (.) Und [s_m ]osKItonetz;
09 YAR      [j:ä-]

```

10 SEL j:ä;  
 11 YAR jä;  
 12 JON aso wÄmmer (mol) ZÄLT und moskItoneTz nee;  
 13 FIN !U:ND!?  
 14 YAR Und [no: (.) FEUer]zEug;  
 15 JON <<rufend> [WOLLdE:>cke; ]  
 16 SEL °h oder [noch (.) (und) verBANDzeug; ]  
 17 FIN [NÄ:I, DAS brUUch mer aber Au, ]  
 18 ver[BANDzeug, ]  
 19 SEL [verBAND ]zeug;

Selvanila schlägt das Buschmesser (01) vor. Dieser Vorschlag wird ohne Begründung geäußert und auch von den anderen Kindern nicht aufgegriffen oder argumentativ diskutiert; stattdessen werden weitere Vorschläge aneinandergereiht (04ff.) oder es wird die recht allgemeine und unspezifische Begründung angeführt, ein Objekt (17) zu "brauchen" (06, 17). Der Konsens wird in den Beispielen der jüngeren Kinder oft durch eine einfache Zusammenführung von Vorschlägen erreicht (Lindström/Sorjonen 2013), indem ein einmal vorgeschlagenes Objekt entweder wiederholt oder einfach von den anderen bestätigt wird. Begründungen werden meist erst nach verhärtetem Dissens und insbesondere nach Warum-Fragen, die explizit eine Begründung einfordern (Kreuz, eingereicht), gegeben.

Insgesamt fällt in den Daten der zweiten Klasse auf, dass nur selten unterschiedliche Perspektiven hierarchisiert werden. Zudem lassen sich nur vereinzelt Sprechhandlungen des begründeten Abwägens sowie des Nennens von Gegen Gründen, der Stützung mit Beispielen und des Zusammenführens von Begründungen auffinden. Daher ist die argumentative Komplexität (Klein 1980; Grundler 2011, 2015) der Diskussionen der ZweitklässlerInnen relativ gering, sodass von einer "flachen" argumentativen Auseinandersetzung gesprochen werden kann (vgl. auch Kreuz/Luginbühl, eingereicht).

Ausgehend von diesen Beobachtungen (vgl. auch Mundwiler/Kreuz 2018; Hauser/Kreuz 2018) war es nun unser Ziel, quantitative Instrumente zu nutzen, um diese Tendenzen zur Breite und Tiefe von Argumentationen über das gesamte Korpus zu überprüfen. In Anschluss an Grundler (2011:177) verstehen wir unter einer breiten Argumentation eine Argumentation, in der viele verschiedene Argumente zur Stützung einer Position realisiert werden. In einer tiefen Argumentation werden einzelne Argumente bezüglich ihrer Stützung ausdifferenziert (durch die Nennung von Regeln, Beispielen etc.). Unsere Annahme war, dass die gemeinsam hervorgebrachten Argumentationen bei den älteren Kindern breiter und tiefer sind und dass sie zunehmend ko-konstruiert, also unter Beteiligung von mehreren Kindern, hervorgebracht werden.

Ein erster quantitativer Zugang zur Beantwortung der Frage nach der Komplexität von Argumentationen, war eine Auszählung von Begründungshandlungen (vgl. Abbildung 5):

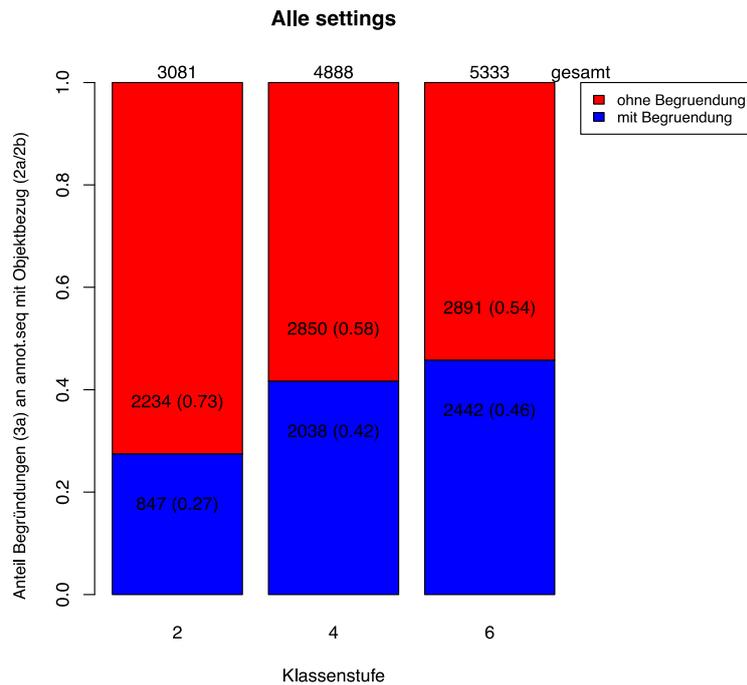


Abbildung 5: Anteil von Begründungshandlungen in annotierten Sequenzen mit Objektbezug, gruppiert nach Klassenstufe und über alle Settings hinweg

In Abbildung 5 geben die Werte innerhalb der Balkenabschnitte die absolute Anzahl der Sequenzen mit Begründung (unterer, blauer Abschnitt) bzw. ohne Begründung an (oberer, roter Abschnitt), sowie in Klammern die Anteile bezogen auf die jeweilige Klassenstufe. Die Werte oberhalb der Balken bezeichnen die Gesamtanzahl der themenbezogenen Handlungen (Sequenzen mit Objektbezug) in der jeweiligen Klassenstufe. Gerechnet über das gesamte Korpus zeigt sich bei dem Anteil an Begründungshandlungen am Total themenbezogener Handlungen ein wesentlicher Anstieg von der 2. zur 4. Klasse (27% zu 42%) und nur geringe Unterschiede zwischen der 4. und 6. Klasse (42% versus 46%). Diese Entwicklung, die ab der 4. Klasse einsetzt, zeigt sich auch in Bezug auf einige andere getestete Merkmale (siehe z. B. Modalisierungen/Verwendung von Konjunktiven, Abbildung 2-a und 2-b; Größe und Häufigkeit thematischer Cluster, Abbildung 9 und 10). In einem weiteren Schritt wurden die Begründungshandlungen gesondert nach der Einbettung in die Interaktion ausgewertet. In allen Settings werden Begründungshandlungen mehrheitlich in ablehnenden Kontexten realisiert (vgl. Abbildung 6):

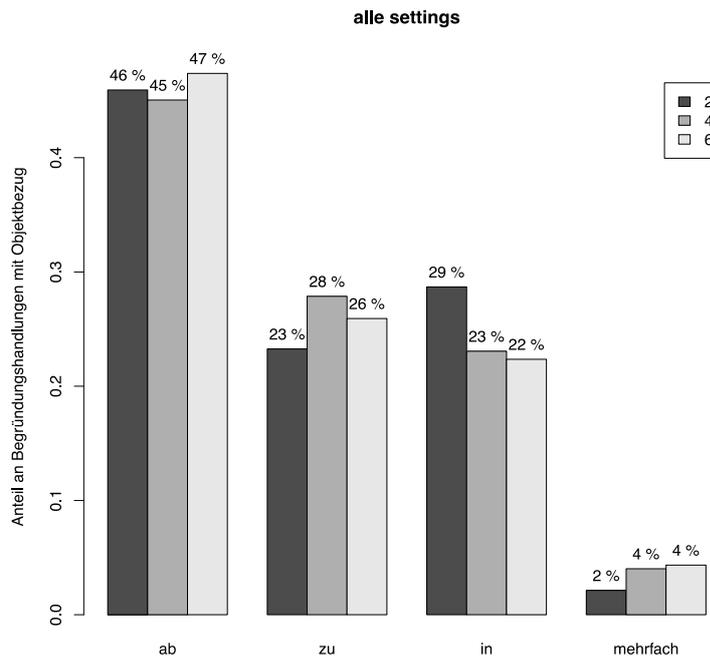


Abbildung 6: Interaktionale Einbettung von Begründungshandlungen (ganzes Korpus)

Hier ist ersichtlich, dass es sich bei 45-47% aller Begründungshandlungen um ablehnende Begründungen handelt, d.h. dass ein Widerspruch bzw. dass differierende Geltungsansprüche bearbeitet werden. Über die drei Klassenstufen hinweg sind 23-28% aller Begründungshandlungen zustimmende Begründungen, wobei es tendenziell eine Zunahme von der 2. zur 4. (und 6.) Klasse gibt. Ebenfalls eine schwache Zunahme ist bei den mehrfach ausgerichteten/abwägenden Begründungshandlungen (vgl. oben, Kapitel 3.2.) festzustellen. Die initiierenden Begründungen nehmen über die Altersstufen hinweg insgesamt ab, was bei dem Robinson-Korpus besonders deutlich wird (vgl. weiter unten).

Bei diesen Berechnungen über alle Settings hinweg fallen die Unterschiede bei den zustimmenden und ablehnenden Begründungen nicht sehr deutlich aus. Das Bild ändert sich aber, wenn die Settings gesondert analysiert werden. Hierbei ist es aber lediglich das Robinsonsetting, bei welchem sich ein klares Bild zeigt. Beim Spendensetting sind die Verläufe hingegen nicht ganz so schlüssig. Die Differenzen in den Einzelsettings zeigen sich teilweise geringer und entsprechend weniger aussagekräftig. Auch im Vergleich der beiden Spendensettings (mit/ohne Handlungskonsequenz) konnten wir keine signifikanten Unterschiede zeigen (möglicherweise aufgrund längerer 'Ausreißer'-Gespräche). Da sich das deutlichste und von der Interpretation her klarste Bild beim Robinsonsetting zeigt, wird dieses Ergebnis im Folgenden näher aufgegriffen (siehe Abbildung 7):

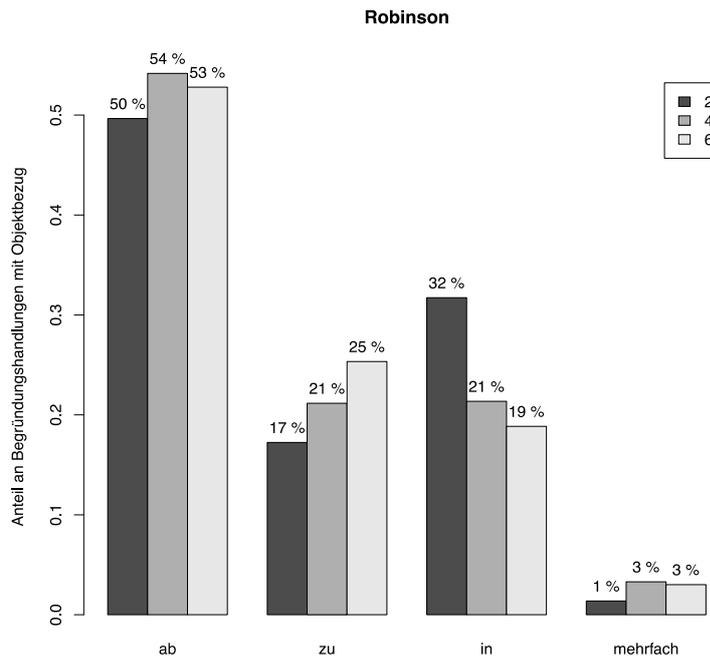


Abbildung 7: Interaktionale Einbettung von Begründungshandlungen (Subkorpus Robinson)

Während – wie schon in der Gesamtdarstellung – die ablehnenden Begründungen mit 50-53% etwa die Hälfte aller Begründungshandlungen ausmachen, zeigen sich nun auch deutliche Verläufe bei den zustimmenden und initiierenden Begründungen: Die zustimmenden Begründungen nehmen von der 2. zur 4. zur 6. Klasse jeweils um 4% zu. Und die initiierenden Begründungen sind mit 32% in der 2. Klasse deutlich stärker vertreten als in den 4. und 6. Klassen (21% resp. 19%). Hier bestätigt sich also das Fazit aus den qualitativen Analysen (vgl. Transkriptbeispiel 4), dass die jüngeren Kinder häufig einzelne Begründungen realisieren, die sich nicht auf eine unmittelbar zuvor von anderen formulierte Position beziehen (beinahe ein Drittel initiierende Begründungen) und dass mit zunehmender Klassenstufe häufiger auch in konsensuellen Kontexten argumentiert wird (Zunahme der zustimmenden Begründungen).

Insgesamt sind die Diskussionen aus dem Robinsonsetting die in Bezug auf die Anteile argumentativer Sequenzen ergiebigsten Daten, d.h., das Robinsonsetting ist übermäßig stark in den Durchschnittswerten vertreten und hat auch die Entwicklung der Codes für die qualitativen Analysen deutlich geprägt. Gleichzeitig sind bei dieser Aufgabe mehr Themen zu bearbeiten (zwölf Gegenstände beim Robinsonsetting vs. vier Projekte beim Spendensetting), was möglicherweise einen Einfluss auf die Verteilung von ablehnenden, zustimmenden oder isolierten/initiierenden Begründungen hat. Was wir aus den unterschiedlichen Berechnungen ableiten können, ist, dass die Art der Aufgabe einen Einfluss auf die argumentative Bearbeitung hat und demnach die Resultate nur bedingt mit anderen Studien und Settings vergleichbar sind. Ein direkter Vergleich ist nur dann problemlos möglich, wenn die Ergebnisse über verschiedene Settings hinweg konstant sind oder aber der spezifische Einfluss eines Settings plausibel identifiziert werden kann.

Im Kontext der Frage nach der Komplexität der Argumentationen, wurde auch ausgewertet, zu welchen Anteilen die verschiedenen Themen jeweils mit oder ohne Begründung verhandelt wurden. Wir sehen entsprechend in Abbildung 8, wie sich das Bild über die Klassenstufe hinweg verändert:

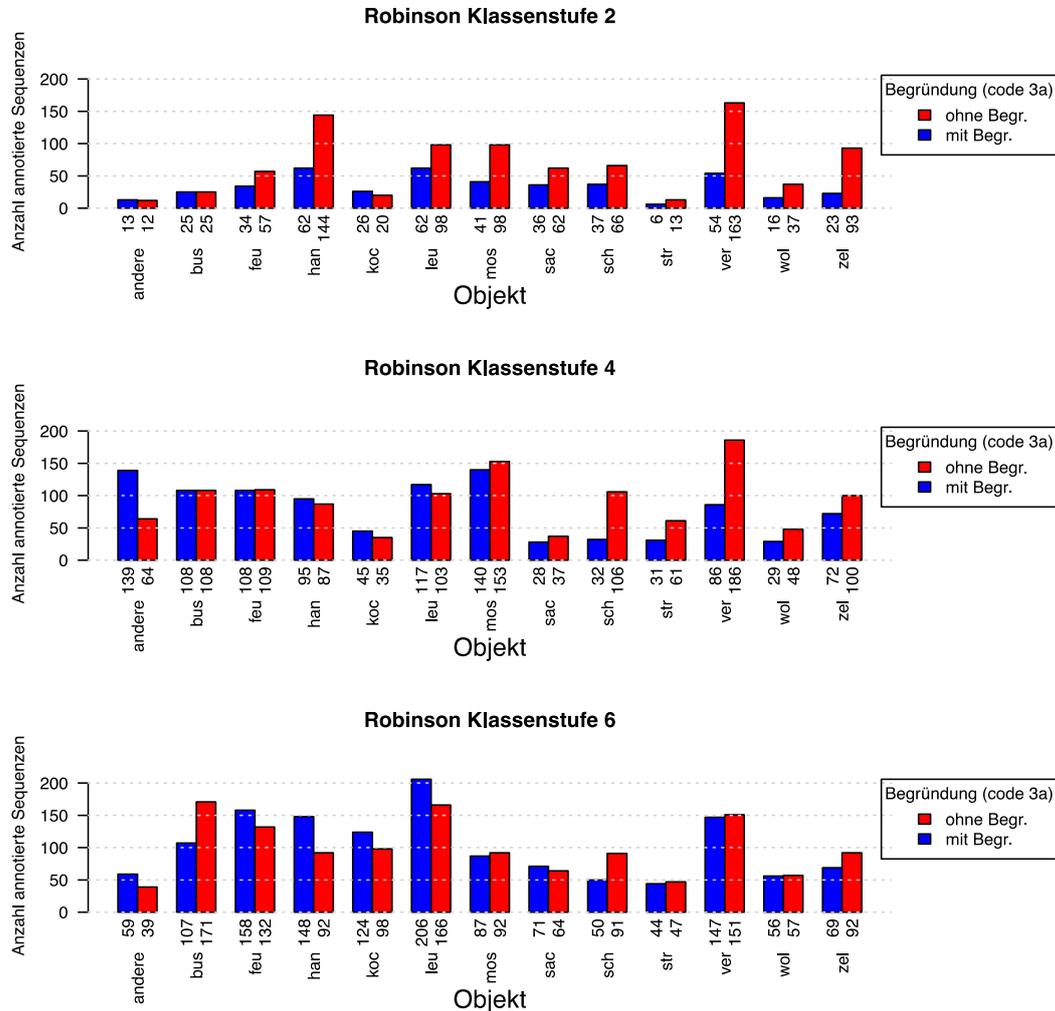


Abbildung 8: Anzahl von begründeten (links, blau) und unbegründeten (rechts, rot) Sequenzen pro Thema (Robinsonsetting)<sup>30</sup>

Abbildung 8 wird jeweils in einer eigenen Grafik pro Klassenstufe (2, 4, 6) gezeigt, wie viele Sequenzen pro Thema/Objekt eine Begründung enthalten (blau). Die Werte unter den Balken geben die absolute Anzahl Sequenzen für das jeweilige Thema/Objekt an. Einerseits ist ersichtlich, wie sich im Vergleich über die Klassenstufen hinweg die Verhältnisse ausgleichen oder umkehren (vgl. die Verhältnisse zwischen roten und blauen Balken, wobei die blauen Balken begründete Äußerungen abbilden) und bei den meisten Themen die Anteile an begründeten Sequenzen zunehmen (vgl. Zunahme der blauen Balken). Andererseits erhält man auch einen differenzierteren Blick auf kontroverse Themen. Dies kann insbesondere für didaktische Zwecke genutzt werden, wenn es um die Gestaltung von sinnvollen Übungsformaten und -themen geht.

<sup>30</sup> Die Abkürzungen für die Gegenstände werden in Kapitel 3.2. ausgeführt.

Um bezogen auf einzelne Gespräche, aber dann auch verallgemeinernd für Altersgruppen, noch besser eingrenzen zu können, ob und wann Kinder Themen vertiefend bearbeiten, wurde codiert, ob in einer Äußerung ein Thema (Objekt bzw. Projekt) behandelt wird, das auch in der vorangegangenen Äußerung im Fokus steht. Dadurch lassen sich *thematische Cluster* bestimmen und deren Größe (Anzahl aufeinanderfolgender Äußerungen zum gleichen Thema) darstellen und vergleichen. Die Clustergröße wurde auf Basis der Anzahl codierter Äußerungen bestimmt, d.h., die Clustergröße 1 entspricht einer einzelnen themenbezogenen Sprachhandlung, die Clustergrößen 2ff. entsprechen zwei (bzw. entsprechend mehr) aufeinanderfolgend codierten Sprachhandlungen zum selben Thema. Interessant sind also insbesondere Cluster ab der Größe 2. Damit können wir demnach eine Aussage darüber treffen, mit welcher Frequenz bzw. Anzahl themenbezogener Sprachhandlungen ein Thema gesprächslokal bearbeitet wird. In der folgenden Darstellung sehen wir die Entwicklung der durchschnittlichen Clustergröße am Beispiel der Robinsondaten:<sup>31</sup>

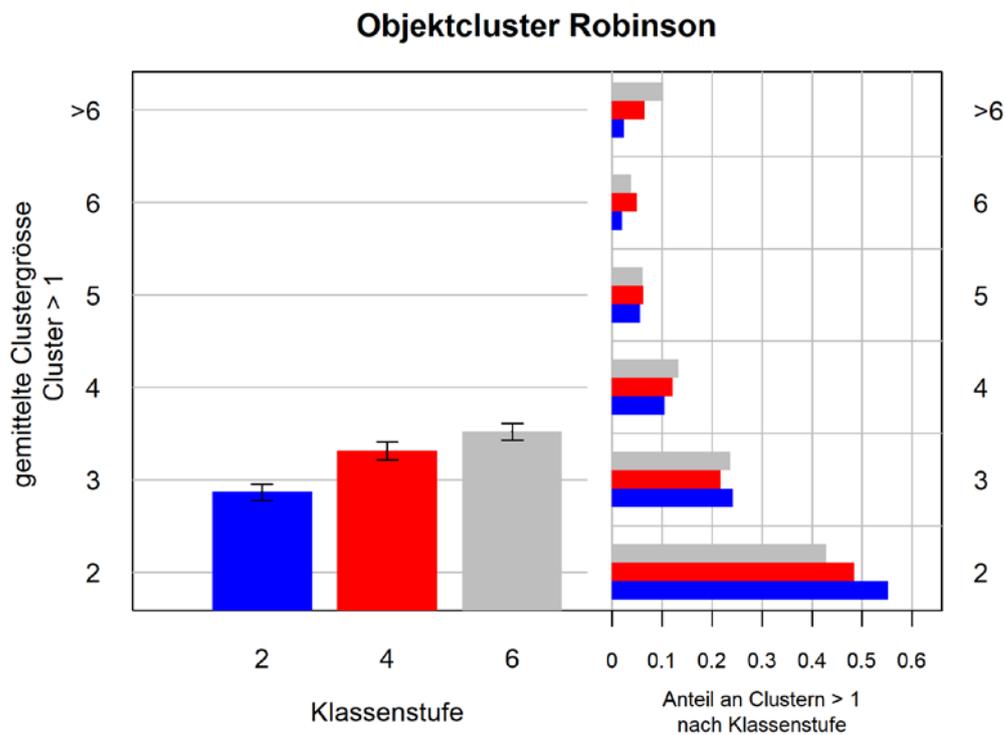


Abbildung 9: Durchschnittliche Größe thematischer Cluster, gruppiert nach Klassenstufe

Für die Berechnung thematischer Cluster in Abbildung 9 wurden alle themenbezogenen Sprachhandlungen einbezogen, die sich auf das gleiche Objekt wie die vorherige Sequenz beziehen, nicht dagegen Sequenzen, die sich auf ein anderes Objekt beziehen wie die vorhergehende Äußerung (nur Cluster > 1). Die Fehlerbalken repräsentieren den Standardfehler. Im rechten Teil der Abbildung ist für jede Clustergröße abgebildet, wie hoch der Anteil der entsprechenden Clustergröße an der Gesamtzahl der Cluster der jeweiligen Klassenstufe ist.

<sup>31</sup> Die Berechnungen beim Spendensetting ergaben keine klaren Entwicklungsverläufe oder Muster.

In den Robinsondaten nimmt die durchschnittliche Größe der thematischen Cluster von der 2. zur 4. zur 6. Klasse stetig zu, die einzelnen Themen werden also gesprächslokal umfangreicher kohärent diskutiert. Auch bei der Berechnung der *Häufigkeit* thematischer Cluster stellen wir im Robinsonsetting eine stetige Zunahme fest:

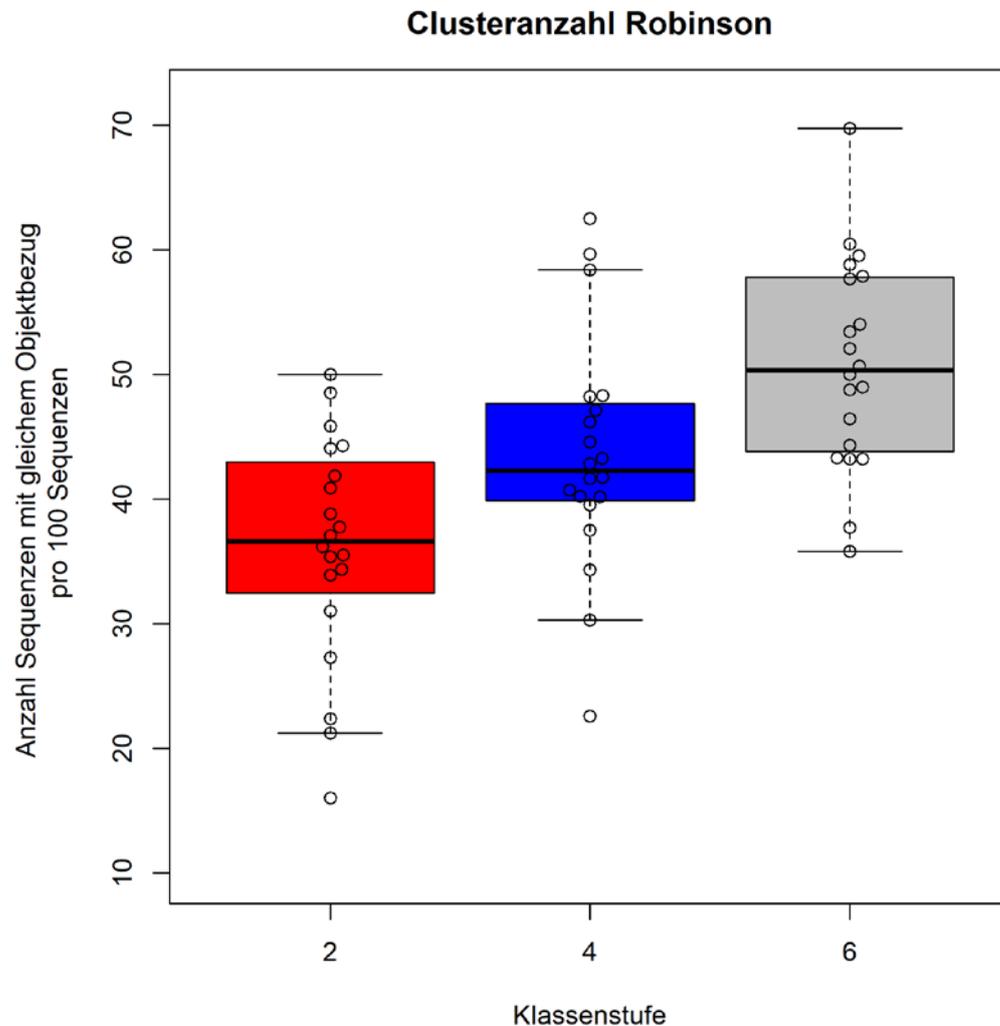


Abbildung 10: Häufigkeit thematischer Cluster im Robinsonsetting, gruppiert nach Klassenstufe

In den Boxplots (vgl. auch Erläuterungen zu Boxplots bei Abbildung 3 oben) in Abbildung 10 ist die Häufigkeit von Objektclustern dargestellt, also in wie vielen Sequenzen (bezogen auf 100 Sequenzen) ein Thema weitergeführt wird. Auch hier sind zusätzlich die Werte der einzelnen Gespräche als Punkte dargestellt. In der 2. Klasse wird in 35% der annotierten Handlungen ein Thema fortgeführt, in der 4. Klasse sind es 46% und in der 6. Klasse 50%. Eine einfaktorielle Varianzanalyse zeigte einen signifikanten Unterschied zwischen den Klassenstufen (ANOVA:  $F(2) = 13.34$ ,  $p < 0.001$ <sup>32</sup>). Die paarweisen Vergleiche (paarweise T-Tests, korrigiert nach Holm) sind alle signifikant (Klasse 2 vs. 4:  $p < 0.05$ ; Klasse 2 vs. 6:  $p < 0.001$ ; Klasse 4 vs. 6:  $p < 0.05$ ). Es zeigt sich also, dass die Schulkinder im Robinsonsetting

<sup>32</sup> Da hier die Werte innerhalb der Gruppen normalverteilt sind, konnte eine einfaktorielle Varianzanalyse (ANOVA) angewendet werden, sowie T-Tests für die paarweisen Vergleiche.

mit zunehmendem Alter häufiger in längeren Sequenzen ein Thema diskutieren. Weiter unten werden wir thematische Cluster in einzelnen Gesprächen betrachten (vgl. Kapitel 4.2.) und für zwei Einzelfälle visualisieren (siehe Abbildungen 15 und 16).

Unsere Analysen zeigen auch, dass der Anteil an zustimmenden Begründungen zunimmt, dass also zunehmend auch (partiell) konsensuell komplexer diskutiert wird (vgl. oben, Abbildung 7). Natürlich sind aber für die Komplexität von Argumentationen noch weitere Aspekte zentral, die wir nicht quantitativ erfassen können, so etwa die Frage, inwiefern bereits formulierte Begründungen wiederholt werden, wie komplex die einzelnen Begründungen sind etc. Hierfür sind qualitative Analysen notwendig, die nun aber vor dem Hintergrund der quantitativen Ergebnisse fokussierter erfolgen können.

#### 4.1.3. Interaktivität

In didaktischen Diskussionen wird vielfach die Wichtigkeit und Notwendigkeit betont, im Bereich des mündlichen Argumentierens Übungsgelegenheiten zu schaffen, die einen Bezug zur Realität der SchülerInnen schaffen und bei denen im besten Falle auch eine echte Handlungskonsequenz resultiert ("Ernstfalldidaktik", Wagner 2003:749, zitiert in Becker-Mrotzek 2008:71). Um diese Forderung empirisch zu überprüfen, haben wir die beiden Subkorpora 'Spendensetting mit Handlungskonsequenz' (Sm) und 'Spendensetting ohne Handlungskonsequenz' (So) erhoben und die Daten in Bezug auf verschiedene Kriterien verglichen. Dabei gingen wir u.a. der Frage nach der Interaktivität nach, und zwar basierend auf der Annahme, dass das reale Setting motivierender ist und mehr Involvierung der einzelnen SprecherInnen erzeugt (vgl. Grundler 2011:83; Spiegel 2006:39f.). Um einen entsprechenden möglichen Indikator zu prüfen, wurde die Häufigkeit von Sprecherwechseln<sup>33</sup> bestimmt und verglichen. Natürlich spielen bei der Beurteilung der Involvierung als möglicher Indikator für Motivation auch andere Aspekte eine Rolle, die quantifizierend untersucht wurden, so etwa die individuelle Redezeit (vgl. dazu exemplarisch die Auswertung einzelner Gespräche unter 4.2.), die Komplexität der Argumentationen (vgl. dazu Kapitel 4.1.2.), die Menge der Begründungen und deren Sequenzialisierung (konsensuell oder adversativ, wobei in letzterem Fall eher Pro und Kontra abgewogen wird, vgl. Abbildungen 7 und 8). Hinzu kommen Indikatoren für die Involvierung, die wohl kaum sinnvoll quantitativ analysiert werden können, die von nonverbalen über intonatorische zu verbalen Mitteln reichen. Es geht uns im Folgenden deshalb nur exemplarisch um die Analyse eines einzelnen Indikators.

Mit einem gemischten logistischen Modell wurde berechnet, ob das Auftreten eines Sprecherwechsels durch die Faktoren Klassenstufe und Setting bzw. deren Wechselwirkung beeinflusst wird. Als abhängige Variable wurde für jedes Gespräch codiert, ob eine Äußerung von einem anderen Sprecher bzw. einer anderen Sprecherin getätigt wurde als die vorangegangene Äußerung. Analysiert wurden

<sup>33</sup> Ein ähnliches Vorgehen findet sich in einer Studie von Lucas M. Bietti, die im Rahmen seines Vortrags "Interacting to Remember: Coordination and Distribution in Memory Collaboration" an der GAL Research School, Albert-Ludwigs-Universität Freiburg, 05.09.2017, vorgestellt wurde. Darin bestimmte er den Grad der Interaktivität als "distribution of speakers' turns" (PP-Folie 23).

zum einen die Haupteffekte von Klassenstufe und Setting. Ein signifikanter Haupteffekt etwa der Klassenstufe bedeutet, dass die Wahrscheinlichkeit eines Sprecherwechsel sich signifikant zwischen den Klassenstufen unterscheidet, die Zugehörigkeit zu einer Klassenstufe also eine Vorhersage der Wahrscheinlichkeit eines Sprecherwechsels ermöglicht (=signifikanter Prädiktor). Eine Wechselwirkung würde beispielsweise bedeuten, dass der Einfluss der Klassenstufe unterschiedlich ausfällt, je nachdem welches Setting bearbeitet wurde. In gemischten Modellen können neben den Prädiktoren oder *fixed factors* auch Zufallsvariablen berücksichtigt werden. Dies bedeutet, dass die Unterschiede innerhalb der Gruppen, die durch die zufällige Wahl der Stichprobe entstehen, in die Berechnung einbezogen werden. In unserem Fall wurden Gespräch (Transkript) und SprecherIn als Zufallsfaktoren (*random intercepts*) aufgenommen. Damit wird zum einen eine bessere Modellanpassung ermöglicht sowie gewährleistet, dass Effekte, die sich aus den Unterschieden innerhalb einer Gruppe ergeben, nicht fälschlicherweise als Einfluss der Prädiktoren bzw. *fixed factors* erscheinen. Als signifikanter Prädiktor zeigte sich allerdings lediglich die Klassenstufe (signifikanter Haupteffekt,  $p < 0.001$ ): In Klasse 6 ist die Auftretenswahrscheinlichkeit eines Sprecherwechsels höher als in den Klassenstufen 2 bzw. 4 (siehe Abbildung 11). Die Handlungskonsequenz hingegen sowie die Interaktion der beiden Faktoren (Handlungskonsequenz und Klassenstufe) sind nicht signifikant. In Abbildung 11 sind die geschätzten Wahrscheinlichkeiten eines Sprecherwechsels dargestellt, gruppiert nach Klassenstufe und Setting (Spende mit bzw. ohne Handlungskonsequenz):

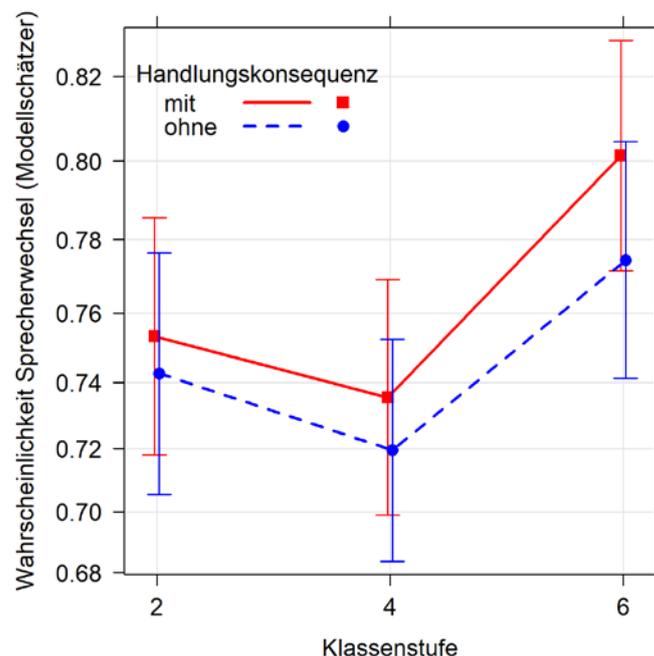


Abbildung 11: Geschätzte Wahrscheinlichkeit (gemischtes logistisches Modell) eines Sprecherwechsels in Abhängigkeit von Klassenstufe und Setting (Spende mit bzw. ohne Handlungskonsequenz). Abgetragen sind die Schätzwerte des gemischten logistischen Modells mit Interaktion der Variablen *Klassenstufe* und *Setting*; Gespräch und SprecherIn waren als Zufallsfaktoren spezifiziert (*random intercepts*), die Fehlerbalken repräsentieren Konfidenzintervalle (95%).

Die Grafik zeigt, dass in Klasse 6 die Tendenz besteht, dass Sprecherwechsel häufiger vorkommen, wenn eine Handlungsrelevanz vorliegt. Der Unterschied ist jedoch nicht signifikant. Auch ist ersichtlich, dass die Werte beim Spendensetting mit Handlungskonsequenz (Sm) in allen drei Klassenstufen über den Werten des fiktiven Vergleichssettings (So) liegen, die Unterschiede sind jedoch ebenfalls nicht signifikant. Insofern lässt sich die Erwartung einer erhöhten Involvierung aufgrund der Handlungskonsequenz anhand dieser Berechnung nicht bestätigen.

Allen bisher gezeigten Verfahren ist gemein, dass sie eine Aussage über das gesamte Korpus oder selektiv über Subkorpora zulassen. Gerade bei größeren Datenbeständen ist es oftmals anspruchsvoll oder unmöglich, basierend auf qualitativen Mikroanalysen den Blick von außen einzunehmen und globale Strukturen zu erkennen oder gar Zusammenhänge und Effekte zu berechnen, die erst durch kontrollierte Vergleichssettings möglich werden. Hier bieten die quantitativen Methoden Möglichkeiten, entweder basierend auf reinem Sprachmaterial (wobei die angefertigten Transkripte, d.h. mündliche 'Texte' durchsucht wurden) oder basierend auf vorgängig gesetzten Codes, einen Überblick zu den Daten zu bekommen. In unserem Fall wurden dabei nicht nur Eindrücke aus den qualitativen Studien bestätigt, sondern es konnten auch unerwartete Befunde in die Gesamtanalyse einbezogen werden. So hat sich beispielsweise unsere Erwartung, dass in Begründungen zunehmend modalisiert und perspektiviert wird, nur partiell bestätigt. Geschlechterbezogene Unterschiede im Bereich der Modalisierung und Perspektivierung konnten auch kaum gefunden werden. Ebenso zeigt sich nur in einem Setting (Robinson), dass sich der Anteil der zustimmenden bzw. initiiierenden Begründungen stabil in eine bestimmte Richtung entwickelt. Andererseits ist – unseren Erwartungen entsprechend – eine relative Zunahme an Begründungshandlungen über alle Daten hinweg zu beobachten.

#### 4.2. Visualisierung und Auswertung einzelner Gespräche

Im Zuge der Auswertungen haben wir verschiedentlich mit Visualisierungen<sup>34</sup> gearbeitet, die nicht nur zusammenfassende Ergebnisse veranschaulichend darstellen, sondern individuelle Differenzen besonders gut ersichtlich machen und weitere Analyseschritte anregen können. Beispielsweise haben wir die einzelnen Gesprächsverläufe hinsichtlich ausgewählter Aspekte farblich visualisiert. So konnten einerseits einfachere Aspekte wie reine Sprecheranteile dargestellt und ausgewertet werden, oder es konnten ausgewählte Codes in den Fokus genommen werden, so beispielsweise die thematischen Bezüge oder argumentative Handlungen wie Begründungen. Die Visualisierungen greifen demnach auf quantitativ erstellte Informationen zurück und bilden Häufigkeiten, Muster und Prozesse, aber auch individuelle Differenzen ab. Obschon diese Darstellungen bereits Auswertungen und Interpretationen zulassen können, werden sie u.E. jedoch v.a. dann ergiebig nutzbar, wenn sie mit weiteren qualitativen Analysen verknüpft werden.

Wir fokussieren in der Folge ausgewählte Auswertungen zu argumentativen und thematischen Sprecheranteilen sowie zu thematischen Clustern.

<sup>34</sup> Unter <https://argcomp.shinyapps.io/baselargVis/> sind verschiedene Visualisierungen aus dem Projekt für Interessierte zugänglich.

#### 4.2.1. Argumentative und thematische Sprecheranteile

In den folgenden beiden Abbildungen sind Visualisierungen der Sprecheranteile ersichtlich, wie sie für alle 180 Gespräche erstellt wurden:

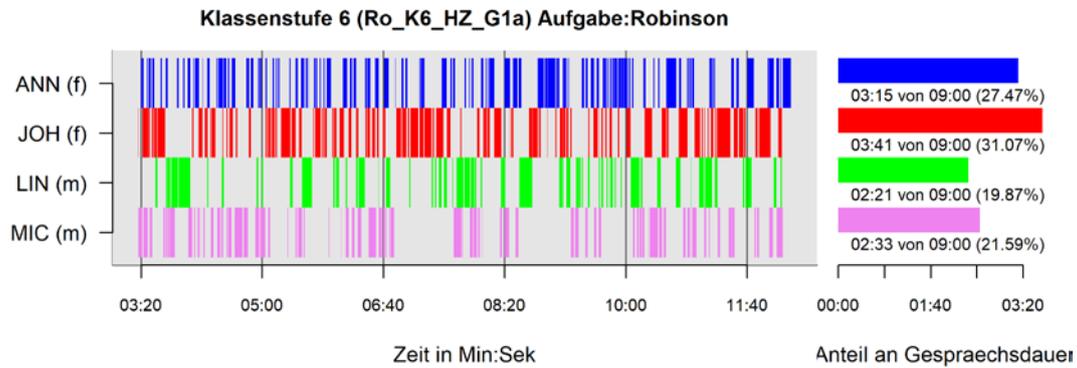


Abbildung 12: Sprecheranteile in einer 6. Klasse  
(Ro\_K6\_HZ\_G1a: Linus, Anna-Lena, Johanna, Michelle)

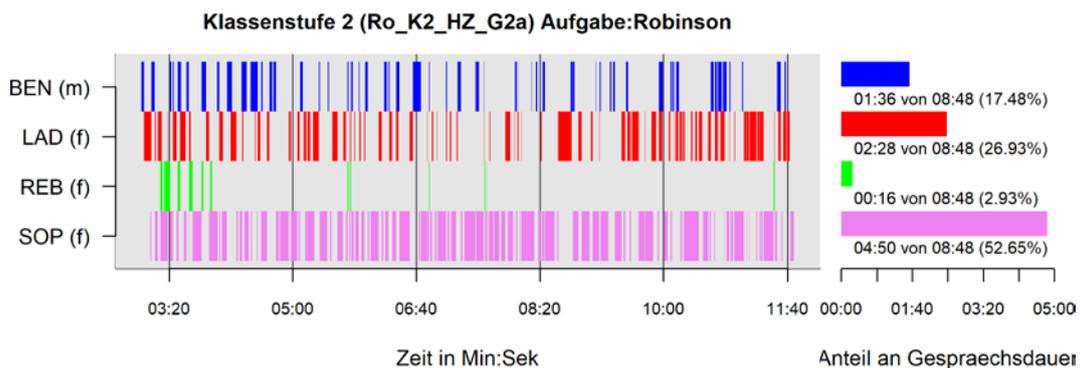


Abbildung 13: Sprecheranteile in einer 2. Klasse  
(Ro\_K2\_HZ\_G2a: Benjamin, Rebecca, Sophie, Ladina)

Bei der ersten Visualisierung (Abbildung 12) sehen wir ziemlich ausgewogene Sprecherbeteiligungen der vier SchülerInnen: Alle beteiligen sich – gemessen an der jeweiligen Sprechdauer – zu ca. 20-30%.<sup>35</sup> Wir sehen gleichzeitig aber auch, dass einzelne SchülerInnen phasenweise sehr dominant involviert sind (beispielsweise JOH um die 7. Minute), während andere sich zurücknehmen (in diesem Fall MIC).

Als Vergleich dazu sehen wir in Abbildung 13, dass eine derart ausgeglichene Beteiligung von allen vier SchülerInnen keineswegs selbstverständlich ist. Vielmehr finden wir auch Gespräche, in denen beispielsweise eine Person deutlich aktiver am Gespräch teilnimmt als andere Beteiligte und auch bei sämtlichen Phasen

<sup>35</sup> Gemessen wurde in diesem Fall die reine Sprechzeit, d.h. nur das verbal Geäußerte. Aus gesprächsanalytischer Sicht interessiert uns zwar grundsätzlich die umfassende, multimodal ausgestaltete Beteiligung der Kinder, jedoch ließen sich wegen des umfangreichen Korpus multimodale Aspekte nicht in demselben Maße standardisiert transkribieren und entsprechend auch nicht aussagekräftig quantifizieren.

stark involviert ist (hier dominiert SOP mit ca. 53%). Oder das Gespräch findet hauptsächlich zwischen zwei oder drei Kindern statt (hier beteiligen sich neben SOP die Kinder BEN und LAD mit je ca. 17% und ca. 27%). Was ebenfalls dank der Visualisierung erkennbar wird, ist das partielle oder komplette Ausbleiben von (verbaler) Beteiligung (hier spricht REB nur während ca. 3% der Gesamtsprechzeit). Während bei der gesprächsanalytischen Sequenzanalyse der Fokus oftmals auf kleineren Phänomenen und Kontexten liegt, ermöglichen die gezeigten Visualisierungen einen Gesamtblick über das Gespräch. Auch können so rasch Passagen (oder ganze Gespräche) identifiziert werden, bei denen ein qualitativer Blick besonders lohnenswert scheint.

Was wir bei diesen Visualisierungen noch nicht erfasst haben, ist, mit welchen sprachlichen Handlungen sich die Kinder beteiligen und ob die Sprecherbeiträge jeweils themenfokussiert sind, oder ob und wann es auch Phasen von Nebensequenzen gibt. In den folgenden Darstellungen (Abbildung 14) werden daher zwei Visualisierungen übereinandergelegt, sodass einerseits farblich ersichtlich wird, wann über welche Themen gesprochen wird, und andererseits, an welchen Stellen Begründungshandlungen realisiert werden. Es handelt sich um dasselbe Gespräch aus der 6. Klasse wie oben (vgl. Abbildung 12):

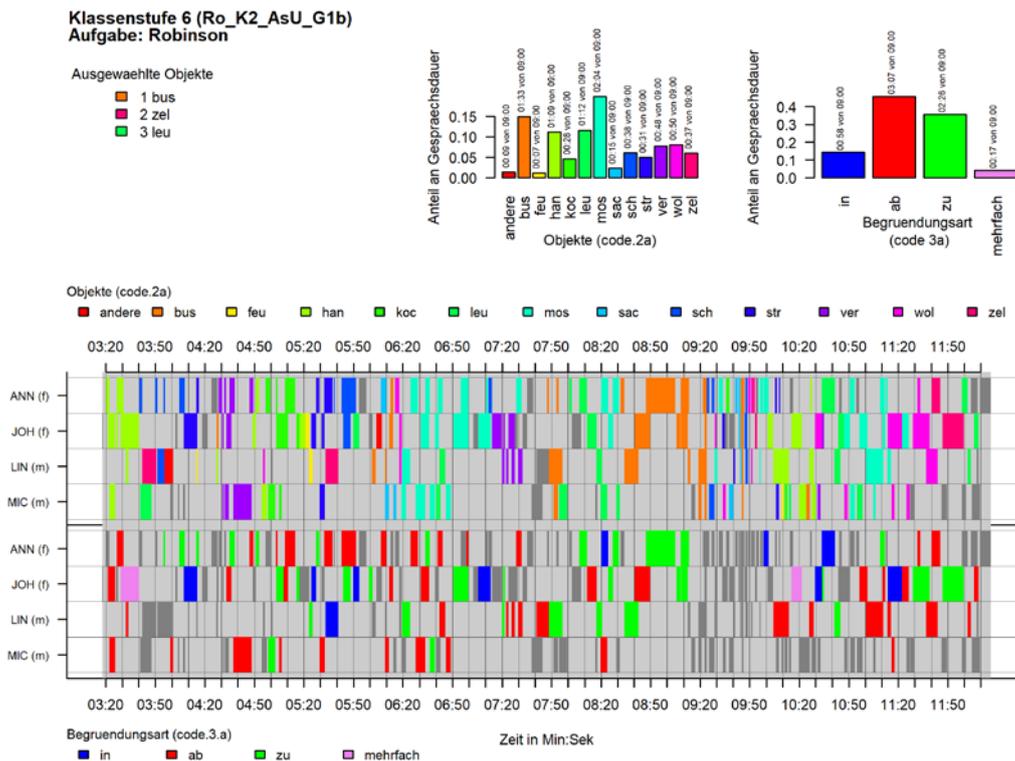


Abbildung 14: Gespräch einer 6. Klasse (Ro\_K6\_HZ\_G1a: Linus, Anna-Lena, Johanna, Michelle). Abgebildet sind oben die am Ende ausgewählten Objekte, der Anteil der Themenbezüge sowie der Begründungsarten an der Gesprächsdauer. Unten sieht man Themenbezug (obere vier Reihen) und Begründungshandlungen (untere vier Reihen) im Zeitverlauf (Gesprächsdauer in Minuten auf der unteren Achse ablesbar), getrennt jeweils nach den 4 SprecherInnen (auf der Achse links).

Oben links in Abbildung 14 wurde für das Gespräch festgehalten, auf welche Gegenstände bzw. Spendenprojekte sich die Gruppe am Ende geeinigt hat (im Beispiel: Buschmesser, Zelt, Leuchtrakete). Oben mittig und rechts finden sich die zusammengefassten Sprechanteile einmal aufgeschlüsselt nach Gesprächsthema (mittig, "Objekte (code.2a)") und einmal nach Art der Begründungshandlungen (rechts, "Begründungsart (code.3a)"). Diese Angaben korrespondieren mit den Visualisierungen der Gesprächsverläufe in den unteren zwei Dritteln der Abbildung. So ist im oberen Gesprächsverlauf farblich codiert, wer wann über welches Thema spricht. Beispielsweise können Cluster in den Blick genommen werden: Es sind drei längere Passagen zum Moskitonetz (türkis) ersichtlich (welches nicht zu den gewählten Gegenständen gehört). Das gewählte Buschmesser (orange) hingegen wurde nur einmal länger verhandelt, was offenbar bereits in Konsens mündete. In der unteren Visualisierung des Gesprächsverlaufs wurde farblich codiert, ob und welche Art der Begründungshandlung realisiert wird. Bei den dunkelgrau markierten Stellen handelt es sich nicht um Begründungshandlungen, sondern um codierte Handlungen der Ebene 1 (Auslösehandlungen) oder der Ebene 4 (weitere Beobachtungen, u.a. metadiskursive Handlungen).

Was beim Vergleich der beiden Visualisierungen auffällt, ist, dass sich MIC trotz ausgeglichenen Sprecheranteilen verhältnismäßig wenig begründend an der Diskussion beteiligt. D.h., sie leistet mehrheitlich andere Beiträge wie beispielsweise Positionen zu beziehen oder gesprächsleitende Aufgaben zu übernehmen. Außerdem ist ersichtlich, dass JOH verschiedentlich begründend interagiert, nämlich sequenziell so eingebettet, dass sie ablehnend, zustimmend, initiiierend und auch mehrfachbezogen/abwägend begründet. Solche personenbezogenen Beobachtungen können dafür genutzt werden, unter Einbezug weiterer Sequenzanalysen Sprecherprofile zu erstellen.

Zudem können die Visualisierungen dazu dienen, längere kontroverse oder konsensuelle Begründungssequenzen zu identifizieren. So ist in Abbildung 14 beispielsweise ersichtlich, dass das Buschmesser (orange in der oberen Visualisierung) nicht sehr kontrovers diskutiert wird, sondern mehrheitlich durch zustimmendes Begründen gefestigt wird (korrespondierender Bereich in der unteren Visualisierung in hellgrün für zustimmendes Begründen).

Insgesamt ermöglicht diese Art des Zugangs zu Einzelgesprächen, Hinweise zu individuellen Differenzen zu erhalten sowie auf potenziell interessante und analytisch relevante Stellen aufmerksam zu werden. Mit den hier präsentierten Visualisierungen lassen sich beispielsweise thematische Cluster, Häufungen von (zustimmenden) Begründungen, Sprecherprofile, Gesprächsprofile o.ä. erkennen und weiter vertiefen. Um die identifizierten Phänomene oder Stellen umfassend zu verstehen, ist dann wiederum ein Rückgriff auf qualitative Methoden und originale Videodaten sinnvoll, um die konkreten Ausprägungen im Kontext verorten und erklären zu können.

#### 4.2.2. Thematische Cluster

Die thematischen Cluster, welche weiter oben für zusammenfassende Ergebnisse genutzt wurden (vgl. Größe und Häufigkeit thematischer Cluster, Abbildung 9 und Abbildung 10), lassen sich auch für Einzelgespräche in ihrer Sequenzialität erstel-

len. Im Folgenden sei dies veranschaulicht durch Visualisierungen der beiden bereits besprochenen Gespräche aus der 2. und 6. Klasse. Die blauen Kreise bedeuten, dass unterschiedliche SprecherInnen am Cluster beteiligt sind, während die roten Markierungen gleichbleibende SprecherInnen anzeigen. Wir sehen in Abbildung 15, dass in diesem Gespräch z. B. zu Beginn (Zeit in Sekunden auf der Achse unten) ein Cluster von sieben aufeinanderfolgenden Äußerungen (Anzahl Äußerungen auf der Achse links) von mehreren SprecherInnen realisiert wird, gefolgt von thematisch isolierten Einzeläußerungen und einem Cluster von drei Äußerungen, wiederum von verschiedenen SprecherInnen realisiert etc.:

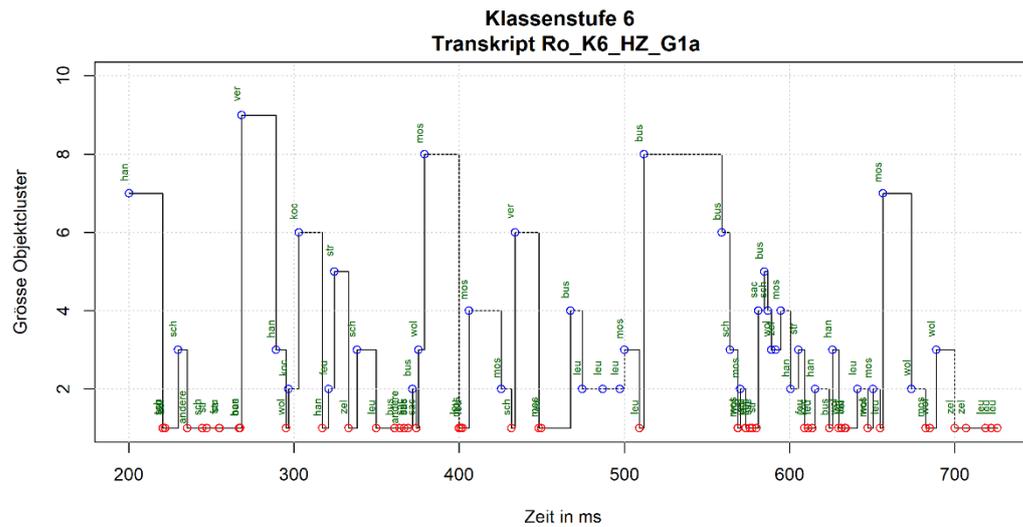


Abbildung 15: Thematische Cluster in einer 6. Klasse (Ro\_K6\_HZ\_G1a: Linus, Anna-Lena, Johanna, Michelle)

Anders präsentiert sich das Gespräch einer 2. Klasse (Abbildung 16):

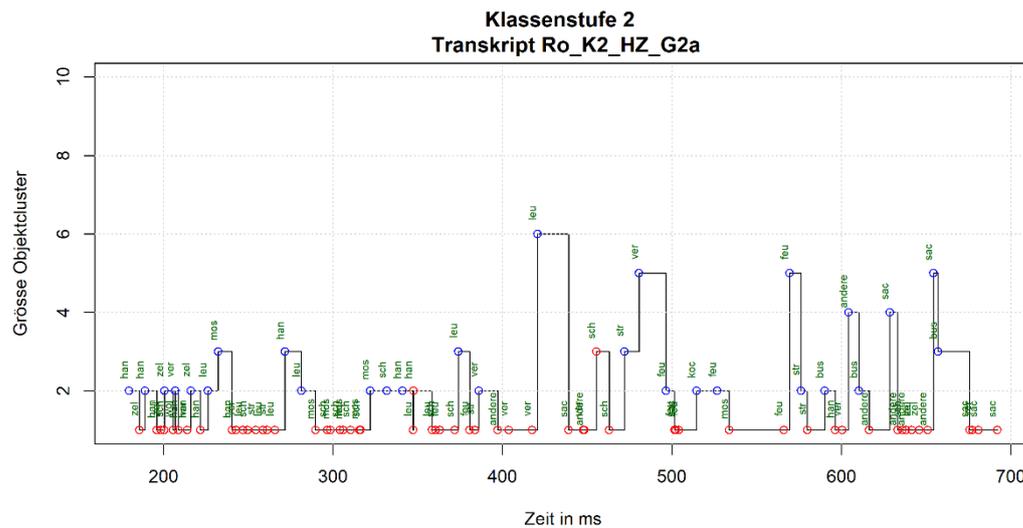


Abbildung 16: Thematische Cluster in einer 2. Klasse (Ro\_K2\_HZ\_G2a: Benjamin, Rebecca, Sophie, Ladina)

Im Beispiel aus der 6. Klasse (Abbildung 15) sind von Beginn weg längere Cluster von bis zu neun Beiträgen vertreten, was bedeutet, dass die SchülerInnen unmittelbar in eine vertiefende Diskussion zu den einzelnen Themen bzw. Objekten einsteigen. Hingegen zeigt sich im Beispiel aus der 2. Klasse (Abbildung 16), dass in der ersten Hälfte der Diskussion die Cluster von Einzelbeiträgen und zwei Beiträgen überwiegen und dass auch insgesamt die maximale Clustergröße sechs Beiträge nicht übersteigt. Auch diese Visualisierungen können nicht abschließend als Ergebnisse verstanden werden, sondern ermöglichen, einen anders eingestellten Blick auf die Gesprächsdaten zu werfen. So kann in einer qualitativen Analyse etwa nach Auslösehandlungen von größeren thematischen Clustern gefragt werden.

Zusammenfassend kann gesagt werden, dass die meisten der gezeigten Abfragen einerseits eine globale Gesamtsicht zulassen, aber andererseits wieder zurück zu den Daten führen, da die Details nicht abschließend abgebildet werden können. Zwar kann auch die Detailliertheit der Codes angepasst werden, was wiederum aussagekräftigere Quantifizierungen ermöglichen würde. Dies würde aber einen erneuten kompletten Codier-Durchgang verlangen, was sich als sehr aufwändig gestaltet. Eine weitere Möglichkeit läge in der Ausdifferenzierung von ausgewählten Codes in kleinere Subkorpora. Insgesamt scheint uns aber der Weg zurück zu den qualitativen Analysen in jedem Fall nötig zu sein, um entweder Interpretationen im Kontext zu verorten (siehe z. B. die kombinierte Verwendung von abschwächenden Modalpartikeln mit verstärkenden Modalpartikeln oder das bloße Wiederholen von Positionen) und Phänomene genauer zu erfassen (z. B. Unterschiede zwischen Mädchen- und Jungengruppen), oder um detailliertere Codes im Einzelfall anzuwenden (z. B. Propositionen mit Begründungsstatus, Gestik als Begründung oder Begründung durch prosodische Markierung). Die Kombination verschiedener Methoden und das rekursive Vorgehen scheint uns dabei vielversprechend zu sein. So zeigen die Ergebnisse, wie quantitative und qualitative Analysen gewinnbringend miteinander verknüpft werden können, um neue Erkenntnisse zu erzielen. Abschließend folgt nun eine Diskussion zu Chancen und Schwierigkeiten der in unserem Projekt angewendeten Methoden.

## **5. Diskussion zur Kombination qualitativer und quantitativer Methoden**

Die Daten wurden mit dem Ziel erhoben, nicht nur qualitative Einblicke in einzelne Gespräche zu bekommen, sondern auch repräsentative Generalisierungen hinsichtlich spezifischer Teilaspekte argumentativer Gesprächskompetenz vornehmen zu können. Deshalb wurde für das Korpus eine ausreichend große Menge an Daten angelegt (180 Diskussionen) und das Setting so stabil wie möglich gehalten. Zudem wurden schultypische Aufgaben gestellt und die Kinder führten die Gespräche ohne Leitung oder Anwesenheit von Erwachsenen, sodass dem gesprächsanalytischen Grundsatz der Erforschung authentischer Gespräche so weit wie möglich entsprochen werden konnte (vgl. Kapitel 2).

Will man Gesprächsdaten nun quantitativ auswerten, ist es zielführend, gewisse Interpretationsprämissen für die Beantwortung der Forschungsfragestellung zu veranschlagen und daraus aussagekräftige Codes zu entwickeln. Diese werden aber nicht durch rein hypothetische, theoretische Annahmen generiert (wie dies klassische quantitative Ansätze voraussetzen, vgl. dazu Einführungen wie z. B. Albert/

Marx 2010), sondern *bottom up* durch eine vorgängige hypothesengenerierende qualitative und in unserem Fall gesprächsanalytische Datenanalyse (Stivers 2015:9). Grundlegend bei qualitativen Forschungsmethoden – und ganz zentral auch bei quantitativen Zugängen – ist das Finden von Gemeinsamkeiten in vergleichbaren Daten (Witt 2001:Absatz 10; vgl. auch Kleining 1982):

Im Rahmen heuristischer qualitativer Forschung geht es bei der Analyse der Daten vorrangig um das Finden von Gemeinsamkeiten, d.h. um solche Aspekte in den Daten, die bei der größtmöglichen Heterogenität der Daten das Verbindende darstellen. Dieses Gemeinsame kann durch Gruppieren der Daten, durch 'Fragen' an die Daten, durch Kontrastierung mit dem Gegenteil, durch Negation usw. gefunden werden.

Die quantitative Erforschung von Gesprächsdaten steht damit – wie schon in Kapitel 1 kurz erläutert – zur Gesprächsanalyse nicht quer, sondern geht ebenso von einer qualitativen Analyse aus, welche genutzt wird als "a solid foundation from which to build formal coding schemes" (Stivers 2015:5). Demnach unterscheidet sich das Vorgehen beim Codieren nicht sehr von demjenigen in der Gesprächsanalyse<sup>36</sup> und ist generell gegenstandsbasiert. Durch das systematische Codieren werden nun aber neue Auswertungen ermöglicht, die beispielsweise Vergleiche über große Korpora zulassen (Stivers 2015:5):

Just as clear characterizations are necessary for formal coding and are already central to CA methods, distributional evidence is a key outcome of formal coding and is also a part of how CA findings are arrived at and represented.

Sorgfältige Überlegungen, welche konkreten Untersuchungsfragen an das Material herangetragen werden sollen und wie qualitative Beobachtungen in die Entwicklung von interessengeleiteten Codes einfließen können, sind Voraussetzung, um sinnvolle Codes für das Material generieren zu können. Allerdings weist Stivers (2015:13f.) zurecht darauf hin, dass "what we know at the time that we develop the coding scheme cannot be adjusted once the coding is done without redoing the coding entirely". Entsprechend ist die Weiterentwicklung der Theorie erschwert: "a coding scheme [...] represents the state of the art at the time and freezes it, making real-time CA advances difficult" (Stivers 2015:13f.). Zwar haben wir die Codes nicht rein theoretisch entwickelt, sondern in mehreren Testdurchläufen an kleinen Datenmengen erprobt und optimiert – so können Codes also auch später noch präzisiert oder weiter ausdifferenziert werden – aber insbesondere bei Auswertungen von großen Korpora ist es aus forschungspragmatischen Gründen meist nicht möglich, eine vollständige Nachcodierung zu leisten.

Daher ist die (rekursive) Phase der Ausarbeitung eines Codierschemas sehr zentral, da die Aussagekraft und die Möglichkeiten der Auswertung davon abhängen. Wie oben ausgeführt, können die Codes auf Basis von qualitativen Analysen entwickelt werden. Dabei gilt es zu beachten, dass die fokussierten Phänomene vor dem Codierprozess bereits hinreichend beschrieben wurden (vgl. Steensig/Heinemann 2015:21). Ist dieser Schritt gelungen, stellt der Prozess des darauffolgenden Codierens selber eine Art der Datenanalyse dar und erlaubt es, systematisch neue typische Fälle und Varianzen aufzudecken (Steensig/Heinemann 2015:21):

---

<sup>36</sup> Auch hier fragt man sich "Wie normativ muss oder darf Argumentationsanalyse sein und wie deskriptiv beziehungsweise rekonstruktiv muss oder kann sie sein?" (Deppermann 2006:17).

[W]e see two main ways that a coding process can open up to analytic possibilities: (1) It can test the 'maturity' of the phenomenon that is coded, with the result that it opens up to new qualitative studies, and (2) it can point to the process of coding as a topic for inquiry rather than as a resource in its own right.

Quantitative Analysen können somit neue Fragen aufwerfen und Anstoß geben, erneut qualitativ in die Daten vorzudringen und Subtypen detailanalytisch zu spezifizieren. Sie stellen also nicht das Ende der Untersuchung dar, sondern werden verstanden als "the prompt for new, qualitative, investigation" (Steensig/Heinemann 2015:21) und können einen anderen Zugang bieten, die soziale Bedeutung für die Wahl bestimmter Sprachhandlungen in der Interaktion besser zu verstehen (vgl. Holmes/Meyerhoff 2006:12f.)

Mit standardisierten Testverfahren lassen sich schlussendlich auf Basis der Codierungen oder auch auf Basis von Textsuchen Ergebnisse zu Häufigkeiten generieren, die verbindlichere Aussagen über umfangreiche Korpora zulassen, als dies in qualitativen Studien möglich ist. Auch erlaubt es die Quantifizierung, Sprachhandlungen mit soziodemografischen Daten der Interagierenden in Beziehung zu setzen und bspw. Zusammenhänge zwischen Aktivitäten wie Begründungen und Alter, Modalisierungen und Geschlecht usw. herzustellen. Ob die Ergebnisse tatsächlich repräsentativ und generalisierbar sind, hängt dann zudem von der Stichprobe und der Anzahl an Tokens des fokussierten Phänomens ab.

Natürlich bringt die nachträgliche quantitative Erfassung von qualitativ analysierten Daten eine Reduktion "from the intricate complexities of human behavior to broad and flattened categories" mit sich (Stivers 2015:2), da "auf einen Teil ihrer Bedeutungen verzichtet wird und sie in eine abstraktere Form überführt werden" (Witt 2001:Absatz 4). Bei Gesprächsdaten, die durch ihre Komplexität erstens sehr *vielen verschiedenen* potenziell relevante Phänomene beinhalten und zweitens in ihrem jeweiligen Kontext *stets anders ausgestaltet sind*, müssen daher besonders stark reduzierende Entscheidungen getroffen werden. Die zu entwickelnden Codes werden so weit generalisiert, dass sie auf die Mehrheit der Daten zutreffen und "hard boundaries" setzen, obwohl die Phänomene und Sprachhandlungen "may be better understood as continuous rather than categorical" (Stivers 2015:13). So können sich bei einer qualitativen Analyse beispielsweise auch solche Beiträge als Begründung zeigen, die nur durch die Einbettung in die syntaktische Struktur der Vorgängeräußerung als solche funktionieren (vgl. dazu Beispiel Codes Ebenen 1, 2 und 3, Kapitel 3.2.: S1: aber wAs ist wenn wir gar keinen emPFANG haben, → S3: stImmt. dann nEhmen wir die lEuchtrakten), ebensolche, die durch Prosodie Begründungsstatus erlangen (Mundwiler/Kreuz 2018) oder gar nur gestisch realisiert werden – Eingang in den Code "Begründungen" finden diese aber nicht. Die Ergebnisse werden in ihrem Detaillierungsgrad dann minimiert und folglich muss bei ihrer Interpretation berücksichtigt werden, dass es sich lediglich um eine eher oberflächliche Darstellung handelt, die erst mit Detailuntersuchungen ausdifferenziert werden kann. Auch wenn mit Vereinfachung und Generalisierungen dieser Art vorsichtig umgegangen werden muss, liegt aber auch hier der Vorteil darin, dass gewisse Phänomene über eine große Datenmenge beschrieben werden können und dadurch Verbindendes erst ersichtlich wird.

Aber selbst wenn der Blick auf die Daten reduktionistisch ist und nur wenige, klar bestimmbare Phänomene quantitativ analysiert werden, gestaltet sich der Pro-

zess des Codierens als sehr aufwendig: dies nicht nur aus dem Grund eines mengenmäßig großen Korpus', sondern auch, weil weiterhin ein Interpretationsspielraum bei der Vergabe der Codes bestehen bleibt – die Teilnehmenden in spontanen Gesprächen interagieren nicht nach einem Schema, sondern bringen fortlaufend situierte, ohne lokalen Gesprächskontext betrachtet aber auch ambige Sprachhandlungen hervor. Dadurch, dass es sich meist nicht um die Quantifizierung rein sprachlicher, oberflächennaher Phänomene handelt, sondern um eine Codierung anhand interpretativ gewonnener Kategorien, müssen während des Ratings fortlaufend Entscheidungen getroffen werden. Auch die Arbeitsweise während des Codierens – nur mit Transkript oder auch unter Einbezug der Videos – hat Einfluss auf die Vergabe bestimmter Codes.

So stellte sich auch immer wieder die Frage nach dem Detaillierungsgrad der Codes: Wie detailliert müssen und können die Codes sein, um die Daten hinreichend und gleichzeitig eindeutig zu erfassen? Die Gefahr bei zu stark ausdifferenzierten Codes ist zum einen der zeitliche Aufwand für die Bewältigung großer Sprachkorpora und zum anderen aber auch die schwindende Möglichkeit des Vergleichs (Stivers 2015:14):

[A]dditional rules can be introduced such as operationalizing the code of the answer to be the first TCU, for instance. Or, an additional code for multiple turn starts could be added. However, in practice, this will not solve the problem because these situations will be rare, and in the end either this will lead to an excluding of the cases or a folding of the cases with other categories of cases.

Daher müssen bei Zweifelsfällen entweder (immer wieder erneute) *Festlegungen* getroffen werden, die konsistent und konsequent auch bei anderen Fällen angewendet werden, oder aber es muss auf bestimmte Codes verzichtet werden. Kontinuierliche Datensitzungen unterstützen den Prozess des Eichens und auch ein doppeltes Rating ist unabdingbar zur Erhöhung der Objektivität sowie zur Herstellung von Interrater-Reliabilität.

Die Quantifizierung ist in mehrerlei Hinsicht ein zeitaufwendiges Unterfangen. Zunächst ist die Erhebung einer ausreichend großen Datenmenge von Nöten und ihre entsprechende Aufbereitung (Transkription). Dann erfolgt die Entwicklung der Codes durch qualitative Analysen, in teilweise rekursiven Prozessen mit mehreren Testdurchläufen und zahlreichen Datensitzungen, bis die Phänomene materialbasiert hinreichend charakterisiert und von mehreren Personen gleichermaßen beurteilt werden können. Anschließend wird die Annotation des gesamten Korpus vorgenommen, eine zweite, überprüfende Annotation und letztlich die Darstellung und Auswertung der Ergebnisse mit verschiedenen Tools (siehe Kapitel 3 und 4). Die Einarbeitung in die technischen Anforderungen und das Generieren von Ergebnissen erfordert eine umfassende und genaue Einarbeitung. Wenn zum ersten Mal mit quantitativen Methoden gearbeitet wird, empfiehlt es sich, ExpertInnen beizuziehen und auch bei der Interpretation der Ergebnisse (resp. generierter Zahlen und Diagramme) auf diese zurückzugreifen.

Sowohl qualitative als auch quantitative Zugänge zu Gesprächsdaten generieren u.E. wichtige Resultate und tragen zu einem umfassenderen Verständnis der Interaktion bei. Wir verstehen die unterschiedlichen Methoden und Visualisierungen dabei als komplementäre Zugänge, die ermöglichen, dass "der blinde Fleck einer Methode durch die jeweils andere ausgeglichen wird" (Krüger/Pfaff 2008:161). Ent-

sprechend wird bei der Methodenwahl der Forschungsgegenstand aus der Perspektive einer anderen Forschungslogik konzeptualisiert und es werden dabei andere Ergebnisse zu einem fokussierten Phänomen oder zur Gesamtstruktur der Gespräche erzielt. Während also möglicherweise bei der Quantifizierung die vielfältigen, kontextgebundenen Variationen punktuell vernachlässigt werden, können dafür aufschlussreiche Aussagen über globalere Phänomene, Sprecherprofile, Entwicklungstendenzen etc. getroffen werden, die bei einem Blick in die Transkripte allein nicht möglich sind. Wir hoffen durch die Darstellung verschiedener Zugänge gezeigt zu haben, welche Chancen, aber auch Grenzen bei quantitativen Auswertungen von Gesprächsdaten zu beachten sind. Gleichzeitig plädieren wir für eine gezielte Kombination gesprächsanalytischer und quantitativer Methoden, um sprachliche Interaktion aus verschiedenen Perspektiven analysieren und vergleichen zu können.

## 6. Danksagung

Das Projekt wurde zu verschiedenen Zeitpunkten von weiteren Personen unterstützt. Wir danken den studentischen Hilfskräften und PraktikantInnen für ihre Unterstützung bei der Transkription und Annotation der Gesprächsdaten: Raquel Coscoñas, Lea Bächli, Giulia D'Amico, Juliet Dorian, Yvonne Fuchs, Gabriela Giallombardo, Lisa Groelly, Silas Gusset, Mélanie Honegger, Melanie Horstmann, Hevin Karakurt, Johanna Meyer, Andrea Neyerlin, Raphaela Portmann, Fabian Schmid, Nicole Sutter, Chantal Wanderon.

Bei Manfred Jurgovsky bedanken wir uns für die Beratung bei der Auswahl von geeigneten Softwares, für die Unterstützung bei der Aufbereitung und Transcodierung von Videodaten sowie für die Entwicklung und Programmierung des Annotationspanels für EXMARaLDA und diesbezüglicher Beratung. Auch bedanken wir uns bei Thomas Schmidt, der als Software-Entwickler für EXMARaLDA zuständig war und uns bei technischen (Kompatibilitäts-)Problemen und Fragen stets beriet und unterstützte.

Und schließlich geht unser Dank auch an den Schweizerischen Nationalfonds (SNF), da unsere Forschung erst durch die finanzielle Förderung ermöglicht wurde (Projektförderung, Abt. I-III, Nr. 149382).

## 7. Literaturverzeichnis

- Albert, Ruth / Marx, Nicole (2010): Empirisches Arbeiten in Linguistik und Sprachlehrforschung. Anleitung zu quantitativen Studien von der Planungsphase bis zum Forschungsbericht. Tübingen: Narr Francke Attempto.
- Arendt, Birte (2015): Kindergartenkinder argumentieren. Peer-Gespräche als Erwerbskontext. In: *Mitteilungen des Deutschen Germanistenverbandes* 62(1), 21-33.
- Becker-Mrotzek, Michael (2008): Gesprächskompetenz vermitteln und ermitteln. Gute Aufgaben im Bereich "Sprechen und Zuhören". In: Bremerich-Vos, Albert / Granzer, Dietline / Köller, Olaf (Hg.): *Lernstandsbestimmung im Fach Deutsch. Gute Aufgaben für den Unterricht*. Weinheim: Beltz, 52-77.

- Becker-Mrotzek, Michael (2009): Mündliche Kommunikationskompetenz. In: Becker-Mrotzek, Michael (Hg.): Mündliche Kommunikation und Gesprächsdidaktik. Baltmannsweiler: Schneider Verlag Hohengehren, 66-83.
- Bergmann, Pia (2017): Gebrauchsprofile von *weiß nich* und *keine Ahnung* im Gespräch – ein Blick auf nicht-responsive Vorkommen. In: Blühdorn, Hardarik/Deppermann, Arnulf / Helmer, Henrike / Spranz-Fogasy, Thomas (Hg.), Diskursmarker im Deutschen. Reflexionen und Analysen. Göttingen: Verlag für Gesprächsforschung, 157-182.
- Bose, Ines / Hannken-Illjes, Kati (2019): Die Entwicklung von argumentativen Fähigkeiten bei Vorschulkindern. Zwischen Agonalität und Kooperativität. In: Bose, Ines / Hannken-Illjes, Kati / Kurtenbach, Stephanie (Hg.): Kinder im Gespräch – mit Kindern im Gespräch. Berlin: Frank & Timme, 11-31.
- Bova, Antonio / Arcidiacono, Francesco (2013): Investigating children's Why-questions. A study comparing argumentative and explanatory function. In: Discourse Studies 15(6), 710-731.
- Brossmann, Brent G. / Canary, Daniel J. (1990): An observational analysis of argument structures. The case of Nightline. In: Argumentation 4, 199-212.
- Bubenhofer, Noah (2009): Sprachgebrauchsmuster. Korpuslinguistik als Methode der Diskurs- und Kulturanalyse. Berlin: de Gruyter.
- Büker, Petra (2008): "Ich bin dafür, dass...". Kinder gewinnen Standpunkte. In: Praxis Deutsch 211 (Themenheft: Meinungen bilden), 14-18.
- Burnett, Ann / Badzinski, Diane M. (2000): An exploratory study of argument in the jury decision-making process. In: Communication Quarterly 48, 380-396.
- Canary, Daniel J. / Ratledge, Nancy T. / Seibold, David R. (1982): Argument and group decision making. Development of a coding scheme. Paper presented at the annual meeting of the Speech Communication Association, Louisville, KY.
- Canary, Daniel J. / Brossmann, Brent G. / Seibold, David R. (1987): Argument structures in decision-making groups. In: Southern Speech Communication Journal 53, 18-37.
- Canary, Daniel J. / Brossmann, Brent G. / Sillars, Alan L. / LoVette, Shannon (1987): Married couples' argument structures and sequences. A comparison of satisfied and dissatisfied dyads. In: Wenzel, Joseph W. (Hg.): Argument and social criticism. Proceedings of the fifth SCA/AFA conference on argumentation. Annandale, VA: Speech Communication Association.
- Canary, Daniel J. / Seibold, David R. (2010): Origins and development of the conversational argument coding scheme. In: Communication Methods and Measures 4(1-2), 7-26.
- Cohen, Daniel / Stevens, Katharina (2017): The Attraction of the Ideal Has No Traction on the Real. On Choices and Roles in Arguments (Presentation). 2nd European Conference on Argumentation: "Argumentation and inference", 20-23 June 2017, University of Fribourg, Switzerland.
- Deppermann, Arnulf (2004): 'Gesprächskompetenz'. Probleme und Herausforderungen eines möglichen Begriffs. In: Becker-Mrotzek, Michael / Brüner, Gisela (Hg.): Analyse und Vermittlung von Gesprächskompetenz. Radolfzell: Verlag für Gesprächsforschung, 15-28.
- Deppermann, Arnulf (2006): Desiderata einer gesprächsanalytischen Argumentationsforschung. In: Deppermann, Arnulf / Hartung, Martin (Hg.): Argumentieren

- in Gesprächen. Gesprächsanalytische Studien. 2. Aufl. Tübingen: Stauffenburg, 10-26.
- Deppermann, A. (2008): Gespräche analysieren. Eine Einführung in konversationsanalytische Methoden. 4. Aufl. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Deppermann, Arnulf /Lucius-Hoene, Gabriele (2006): Argumentatives Erzählen. In: Deppermann, Arnulf /Hartung, Martin (Hg.): Argumentieren in Gesprächen. Gesprächsanalytische Studien. 2. Aufl. Tübingen: Stauffenburg, 130-144.
- Diewald, Gabriele (1999): Die Modalverben im Deutschen. Grammatikalisierung und Polyfunktionalität. Tübingen: Niemeyer.
- Domberg, Andreas / Köymen, Bahar / Tomasello, Michael (2018): Children's reasoning with peers in cooperative and competitive contexts. In: *British Journal of Developmental Psychology* 36, 64-77.
- Ehlich, Konrad (2014): Argumentieren als sprachliche Ressource des diskursiven Lernens. In: Hornung, Antonie / Carobbio, Gabriella / Sorrentino, Daniela (Hg.): Diskursive und textuelle Strukturen in der Hochschuldidaktik. Deutsch und Italienisch im Vergleich. Münster: Waxmann, 41-54.
- Ellis, Donald G. / Maoz, Ifat (2002): Cross-cultural argument interactions between Israeli-Jews and Palestinians. In: *Journal of Applied Communication Research* 30, 181-194.
- Ellis, Donald G. / Maoz, Ifat (2007): Online argument between Israeli Jews and Palestinians. In: *Human Communication Research* 33, 291-309.
- Felton, Mark / Garcia-Mila, Merce / Villarroel, Constanza / Gilabert, Sandra (2015): Arguing collaboratively. Argumentative discourse types and their potential for knowledge building. In: *British Journal of Educational Psychology* 85(3), 372-386.
- Garfinkel, Harold (1967): *Studies in Ethnomethodology*. Cambridge: Cambridge Polity Press.
- Gohl, Christine (2006): Begründen im Gespräch. Eine Untersuchung sprachlicher Praktiken zur Realisierung von Begründungen im gesprochenen Deutsch. Tübingen: Niemeyer.
- Grundler, Elke (2008): Gesprächskompetenz. Ein Systematisierungsvorschlag im Horizont schulischer Bildungsstandards und Kompetenzen. In: *Didaktik Deutsch* 24, 48-69.
- Grundler, Elke (2011): Kompetent argumentieren. Ein gesprächsanalytisch fundiertes Modell. Tübingen: Stauffenburg.
- Grundler, Elke (2015): Interaktive Argumentationsentwicklung in schulischen Argumentationen. In: *Mitteilungen des Deutschen Germanistenverbandes* 62(1), 47-60.
- Gülich, Elisabeth / Mondada, Lorenza (2008): *Konversationsanalyse. Eine Einführung am Beispiel des Französischen*. Tübingen: Niemeyer.
- Hartung, Martin (2004): Wie lässt sich Gesprächskompetenz wirksam und nachhaltig vermitteln? Ein Erfahrungsbericht aus der Praxis. In: Becker-Mrotzek, Michael / Brüner, Gisela (Hg.): *Analyse und Vermittlung von Gesprächskompetenz*. Radolfzell: Verlag für Gesprächsforschung, 47-66.
- Hauser, Stefan / Kreuz, Judith (2018): Mündliches Argumentieren in der Schule zwischen pragmatischen Spielräumen und didaktischen Normsetzungen. In: Albert, Georg / Diao-Klaeger, Sabine (Hg.): *Mündlicher Sprachgebrauch zwischen*

- Normorientierung und pragmatischen Spielräumen. Tübingen: Stauffenburg, 179-199.
- Hauser, Stefan / Luginbühl, Martin (2015): Aushandlung von Angemessenheit in Entscheidungsdiskussionen von Schulkindern. In: *Aptum – Zeitschrift für Sprachkritik und Sprachkultur* 11(2), 180-189.
- Hauser, Stefan / Luginbühl, Martin (2017): Wenn Kinder argumentieren. Grundlagen und erste Befunde einer Studie zur mündlichen Argumentationskompetenz von Schulkindern. In: Meißner, Iris / Wyss, Eva Lia (Hg.): *Begründen – Erklären – Argumentieren. Konzepte und Modellierungen in der Angewandten Linguistik*. Tübingen: Stauffenburg, 89-105.
- Heller, Vivien (2012): Kommunikative Erfahrungen von Kindern in Familie und Unterricht. *Passungen und Divergenzen*. Tübingen: Stauffenburg.
- Heritage, John (1984): A change-of-state token and aspects of its sequential placement. In: Atkinson, J. Maxwell / Heritage, John (Hg.): *Structures of Social Action: Studies in Conversation Analysis*. Cambridge: Cambridge University Press, 299-345.
- Holmes, Janet / Meyerhoff, Miriam (2006): Different voices, different views. An introduction to current research in language and gender. In: Holmes, Janet / Meyerhoff, Miriam (Hg.): *The Handbook of Language and Gender*. Malden, MA: Blackwell, 1-17.
- Jacoby, Sally / Ochs, Elinor (1995): Co-construction. An introduction. In: *Research on Language and Social Interaction* 28(3), 171-183.
- Jacquin, Jérôme (2015): Multimodal counter-argumentation in the workplace. The contribution of gesture and gaze to the expression of disagreement. In: *Colloque Gesture and speech in interaction (GESPIN)* 4, 155-160.
- Kirkham, Sam (2011): Personal style and epistemic stance in classroom discussion. In: *Language and Literature* 20(3), 201-217.
- Klein, Wolfgang (1980): Argumentation und Argument. In: *Zeitschrift für Literaturwissenschaft und Linguistik* 38/39, 9-57.
- Kleining, Gerhard (1982): Umriß zu einer Methodologie qualitativer Sozialforschung. In: *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 34, 224-253.
- Knoblauch, Hubert (1995): *Kommunikationskultur: die kommunikative Konstruktion kultureller Kontexte*. Berlin, New York: de Gruyter.
- Kreuz, Judith (eingereicht): *Ko-Konstruiertes Begründen unter Kindern. Eine gesprächsanalytische Studie von Kleingruppeninteraktionen in der Primarschule*. Dissertation, Universität Basel.
- Kreuz, Judith / Luginbühl, Martin (eingereicht): From Flat Propositions to Deep Co-constructed and Modalised Argumentations: Oral Argumentative Skills among Elementary-school Children from Grades 2 to 6. In: Heller, Vivien / Luginbühl, Martin / Arendt, Birte (Hg.): *Becoming skilled at explaining and arguing: The role of co-construction and multimodality*. In: *Research on Children and Social Interaction*.
- Kreuz, Judith / Mundwiler, Vera / Luginbühl, Martin (2017): Mündliches Argumentieren im Spannungsfeld zwischen Kollaboration und Abgrenzung. Zu lokalen Gruppenidentitäten in schulischen Einigungsdiskussionen. *Bulletin VALS-ASLA (Vereinigung für angewandte Linguistik in der Schweiz)*, N° spécial 2017(2), 147-159.

- Kreuz, Judith / Luginbühl, Martin / Mundwiler, Vera (2019). Gesprächsorganisation in argumentativen Peer-Gesprächen von Schulkindern. In: Bose, Ines / Hancken-Iljes, Kati / Kurtenbach, Stephanie (Hg.): *Kinder im Gespräch – mit Kindern im Gespräch*. Berlin: Frank & Timme, 33-62.
- Krüger, Heinz-Hermann / Pfaff, Nicole (2008): Triangulation quantitativer und qualitativer Zugänge in der Schulforschung. In: Helsper, Werner / Böhme, Jeanette (Hg.): *Handbuch der Schulforschung*. 2., durchgesehene und erweiterte Auflage. Wiesbaden: VS Verlag für Sozialwissenschaften, 157-179.
- Kotthoff, Helga (2015): Konsensuelles Argumentieren in schulischen Sprechstunden. In: Hauser, Stefan / Mundwiler, Vera (Hg.): *Sprachliche Interaktion in schulischen Elterngesprächen*. Bern: hep, 72-94.
- Kyrtziz, Amy / Shuqum, Tamara / Koymen, S. Bahar (2010): Validating justifications in preschool girls' and boys' friendship group talk. Implications for linguistic and socio-cognitive development. In: *Journal of Child Language* 37, 115-144.
- Lijffijt, Jefrey / Nevalainen, Terttu / Säily, Tanja / Papapetrou, Panagiotis / Puolamäki, Kai / Mannila, Heikki (2014): Significance testing of word frequencies in corpora. In: *Digital Scholarship in the Humanities* 31(2), 374-397.
- Lindström, Anna / Marja-Leena Sorjonen (2013): Affiliation in Conversation. In: Sidnell, Jack / Tanya Stivers (Hg.): *The Handbook of Conversation Analysis*. Malden, MA: Blackwell, 350-369.
- Locher, Miriam A. (2004): *Power and Politeness in Action. Disagreements in Oral Communication*. Berlin: de Gruyter.
- Mercer, Neil (2009): Developing argumentation. Lessons learned in the primary school. In: Muller Mirza, Nathalie / Perret-Clermont, Anne-Nelly (Hg.): *Argumentation and Education. Theoretical Foundations and Practices*. Dordrecht: Springer, 177-194.
- Meyer, Paul Georg (1975): Satzverknüpfungsrelationen. Ein Interpretationsmodell für situationsunabhängige Texte. Tübingen: Narr.
- Morek, Miriam (2015): Dissensbearbeitung unter Gleichaltrigen. (K)ein Kontext für den Erwerb argumentativer Gesprächsfähigkeiten? In: *Mitteilungen des Deutschen Germanistenverbandes* 62(1), 34-46.
- Morek, Miriam (2017): Was lernt man bei einer Fishbowl-Diskussion im Unterricht? Zur Normativität und pragmatischen Situiertheit des Argumentierens. In: Hauser, Stefan / Luginbühl, Martin (Hg.): *Gesprächskompetenz in schulischer Interaktion. Normative Ansprüche und kommunikative Praktiken*. Bern: hep, 65-90.
- Mundwiler, Vera / Kreuz, Judith (2018): Collaborative decision-making in argumentative group discussions among primary school children. In: Oswald, Steve / Herman, Thierry / Jacquin, Jérôme (Hg.): *Argumentation and Language. Linguistic, Cognitive and Discursive Explorations*. Cham: Springer, 263-285.
- Mundwiler, Vera / Kreuz, Judith / Hauser, Stefan / Eriksson, Brigit / Luginbühl, Martin (2017): Mündliches Argumentieren als kommunikative Praktik. Schulbuchübungen und empirische Befunde im Vergleich. In: Hauser, Stefan / Luginbühl, Martin (Hg.): *Gesprächskompetenz in schulischer Interaktion. Normative Ansprüche und kommunikative Praktiken*. Bern: hep, 91-123.
- Nonnon, Elisabeth (1996): Activités argumentatives et élaboration de connaissances nouvelles. Le dialogue comme espace d'exploration. In: *Langue française* 112, 67-87.

- Nuyts, Jan (2001): *Epistemic Modality, Language and Conceptualization*. Amsterdam: John Benjamins.
- Perelman, Chaïm / Olbrechts-Tyteca, Lucie (1958): *Traité de l'argumentation. La nouvelle rhétorique*. Bruxelles: Edition de l'Université de Bruxelles.
- Pickering, Martin J./Garrod Simon (2004): *Toward a mechanistic psychology of dialogue*. In: *Behavioral and Brain Sciences*, 27, 169-190.
- Quasthoff, Uta M. (2009): *Entwicklung der mündlichen Kommunikationskompetenz*. In: Becker-Mrotzek, Michael (Hg.): *Mündliche Kommunikation und Gesprächsdidaktik*. Baltmannsweiler: Schneider Verlag Hohengehren, 84-100.
- Quasthoff, Uta / Krah, Antje (2012): *Familiale Kommunikation als Spracherwerbsressource. Das Beispiel argumentativer Kompetenzen*. In: Neuland, Eva (Hg.): *Sprache der Generationen*. Mannheim: Duden, 115-132.
- Rafal, Christine Theberge (1996): *From co-construction to takeovers: Science talk in a group of four girls*. In: *Journal of Learning Sciences*, 5, 279-293.
- Redder, Angelika (2009): *Modal sprachlich handeln*. In: *Der Deutschunterricht* 3, 88-95.
- Rühl, Marco (1999): *Interaktive Dynamik in argumentativen Gesprächen. Für eine kommunikationsprozess-orientierte Argumentationsanalyse*. In: *Zeitschrift für Sprachwissenschaft* 18(1), 3-38.
- Sacks, Harvey (1992): *Lectures on Conversation*. Hgg. von Gail Jefferson / Emanuel A. Schegloff. Oxford: Blackwell.
- Scarvaglieri, Claudio (2017): *"I a hundred per cent agree"*. *Konsensuelles Argumentieren in der interkulturellen Wissenschaftskommunikation*. In: Bühlig, Kristin / Schlickau, Stephan (Hg.): *Argumentieren und Diskutieren*. Frankfurt am Main: Lang, 145-175.
- Schegloff, Emanuel A. (1993) *Reflections on quantification in the study of conversation*. In: *Research on Language and Social Interaction* 26(1), 99-128.
- Schegloff, Emanuel A. (2007): *Sequence Organization in Interaction*. Cambridge: Cambridge University Press.
- Scherer, Carmen (2014): *Korpuslinguistik*. Heidelberg: Winter.
- Schmidt, Thomas (2003): *Visualising Linguistic Annotation as Interlinear Text*. *Working Papers in Multilingualism, Series B (46)*. Hamburg. [<http://www.exmaralda.org/files/Visualising-final.pdf>, 18.01.2018]
- Schmidt, Thomas (2004a): *EXMARaLDA. Ein System zur computergestützten Diskurstranskription*. In: Mehler, Alexander / Lobin, Henning (Hg.): *Automatische Textanalyse. Systeme und Methoden zur Annotation und Analyse natürlichsprachlicher Texte*. Wiesbaden: VS Verlag für Sozialwissenschaften, 203-218.
- Schmidt, Thomas (2004b): *Transcribing and Annotating Spoken Language with EXMARaLDA*. *Proceedings of the LREC-Workshop on XML based richly annotated corpora, Lisbon 2004*. Paris: ELRA. [[http://www.exmaralda.org/files/Paper\\_LREC.pdf](http://www.exmaralda.org/files/Paper_LREC.pdf), 18.01.2018]
- Schmidt, Thomas / Wörner, Kai (2009): *EXMARaLDA. Creating, analysing and sharing spoken language corpora for pragmatic research*. In: *Pragmatics* 19(4), 565-582.
- Schütte, Wilfried (2007): *ATLAS.ti 5. Ein Werkzeug zur qualitativen Datenanalyse*. In: *Gesprächsforschung – Online-Zeitschrift zur verbalen Interaktion* 8, 57-72.

- Schmitt, Reinhold (2002): Rollenspiele als authentische Gespräche Überlegungen zu deren Produktivität im Trainingszusammenhang. In: Brünner, Gisela / Fiehler, Reinhard / Kindt, Walther (Hg.): *Angewandte Diskursforschung. Band 2: Methoden und Anwendungsbereiche*. Radolfzell: Verlag für Gesprächsforschung, 81-99.
- Schwitalla, Johannes (2012): *Gesprochenes Deutsch. Eine Einführung*. 4., neu bearbeitete und erweiterte Aufl. Berlin: Erich Schmidt Verlag.
- Seibold, David R. / Meyers, Renee A. (2007): Group argument. A structuration perspective and research program. In: *Small Group Research* 38, 312-336.
- Selting, Margret / Auer, Peter / Barth-Weingarten, Dagmar / Bergmann, Jörg / Bergmann, Pia / Birkner, Karin / Couper-Kuhlen, Elizabeth / Deppermann, Arnulf / Gilles, Peter / Günthner, Susanne / Hartung, Martin / Kern, Friederike / Mertzlufft, Christine / Meyer, Christian / Morek, Miriam / Oberzaucher, Frank / Peters, Jörg / Quasthoff, Uta / Schütte, Georg / Stukenbrock, Anja / Uhmann, Susanne (2009): Gesprächsanalytisches Transkriptionssystem 2 (GAT 2). In: *Gesprächsforschung – Online-Zeitschrift zur verbalen Interaktion* 10, 353-402.
- Selting, Margret / Couper-Kuhlen, Elizabeth (2001): Forschungsprogramm 'Interaktionale Linguistik'. In: *Linguistische Berichte* 187, 257-287.
- Semic, Beth A. / Canary, Daniel J. (1997): Trait argumentativeness, verbal aggressiveness, and minimally rational argument. An observational analysis of friendship discussions. In: *Communication Quarterly* 45, 355-378.
- Semic, Beth A. / Canary, Daniel J. (2003): An investigation of the association among friends' argument sequences, perceptions of communicative competence, and communication satisfaction. Paper presented at the 2003 annual conference of the International Communication Association, San Diego, CA.
- Sidnell, Jack (2013): Basic Conversation Analytic Methods. In: Sidnell, Jack / Stivers, Tanya (Hg.): *The Handbook of Conversation Analysis*. Chichester: Wiley-Blackwell, 77-99.
- Sidnell, Jack / Stivers, Tanya (Hg.) (2013): *The Handbook of Conversation Analysis*. Chichester: Wiley-Blackwell.
- Spiegel, Carmen (2006): *Unterricht als Interaktion. Gesprächsanalytische Studien zum kommunikativen Spannungsfeld zwischen Lehrern, Schülern und Institution*. Radolfzell: Verlag für Gesprächsforschung.
- Spranz-Fogasy, Thomas (2006): Alles Argumentieren, oder was? Zur Konstitution von Argumentation in Gesprächen. In: Deppermann, Arnulf / Hartung, Martin (Hg.): *Argumentieren in Gesprächen. Gesprächsanalytische Studien*. 2. Aufl. Tübingen: Stauffenburg, 27-39.
- Steensig Jakob / Heinemann, Trine (2015): Opening up codings? In: *Research on Language and Social Interaction* 48(1), 20-25.
- Stivers, Tanya (2015): Coding social interaction. A heretical approach in conversation analysis? In: *Research on Language and Social Interaction* 48(1), 1-19.
- ten Have, Paul (2007): *Introducing qualitative methods. Doing conversation analysis: A practical guide* (2nd ed.). Thousand Oaks, CA: Sage Publications Ltd.
- Wagner, Roland W. (2003): Methoden des Unterrichts in mündlicher Kommunikation. In: Bredel, Ursula / Günther, Hartmut / Klotz, Peter / Ossner, Jakob / Siebert-Ott, Gesa (Hg.): *Didaktik der deutschen Sprache. Ein Handbuch*. Bd. 2. Paderborn: Schöningh, 747-759.

Witt, Harald (2001). Forschungsstrategien bei quantitativer und qualitativer Sozialforschung. In: Forum Qualitative Sozialforschung / Forum Qualitative Social Research 2(1), Art. 8, 36 Absätze.  
[<http://nbn-resolving.de/urn:nbn:de:0114-fqs010189>, 30.01.2018].

Prof. Dr. Stefan Hauser  
Dr. Vera Mundwiler  
Judith Kreuz, MA

Zentrum Mündlichkeit  
Pädagogische Hochschule Zug  
Zugerbergstrasse 3  
6300 Zug  
Schweiz

stefan.hauser@phzg.ch  
vera.mundwiler@phzg.ch  
judith.kreuz@phzg.ch

Prof. Dr. Martin Luginbühl  
Dr. Daniel Müller Feldmeth (c/o Prof. Dr. Martin Luginbühl)

Deutsches Seminar  
Universität Basel  
Nadelberg 4  
4051 Basel  
Schweiz

martin.luginbuehl@unibas.ch  
daniel.mueller@feldmeth.de

Veröffentlicht am 6.2.2020

© Copyright by GESPRÄCHSFORSCHUNG. Alle Rechte vorbehalten.