Gesprächsforschung - Online-Zeitschrift zur verbalen Interaktion (ISSN 1617-1837)
Ausgabe 22 (2021), Seite 179-236 (www.gespraechsforschung-ozs.de)

Quantitative and Qualitative Approaches in Conversation Analysis: Methodological Reflections on a Study of Argumentative Group Discussions[1]

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

Abstract
This paper discusses fundamental and applied aspects of the relationship between qualitative and quantitative approaches in conversation analysis (CA). Starting from the observation that informal coding and informal quantification can be found in numerous works of classical CA, the paper discusses features of formal quantification. In the main section of the paper, such formal quantifications (based on conversation-analytic categories) are carried out on a corpus of 180 conversations of Swiss German schoolchildren, and 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 about the relationship between interactive argumentative practices (modalization, perspectivation, argumentative complexity, interactivity) and other variables (age, setting, gender) and thus about crucial aspects of oral argumentation competence. 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 argumentation competence – formal quantification – coding – visualization of conversational data – modalization – perspectivation – argumentative complexity – interactivity.

German 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.

---

[1] This article is a translation of: Mundwiler, Vera / Kreuz, Judith / Müller-Feldmeth, Daniel / Luginbühl, Martin / Hauser, Stefan (2019): Quantitative und qualitative Zugänge in der Gesprächsforschung. Methodologische Betrachtungen am Beispiel einer Studie zu argumentativen Gruppendiskussionen. In: Gesprächsforschung 20, 323-83.

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

1. Introduction: On the Integration of Quantitative Analyses into Conversation Analysis

According to Sidnell (2013:77) the aims of conversation analysis (CA) include discovering "previously unknown regularities of human interaction" and identifying and describing "the recurrent practices of interaction." Methodologically, the turn-by-turn-analysis, the microanalysis of the forms of utterances and their sequential positioning on the basis of transcribed conversations are most central. On the basis of these analyses, emerging structures (recurrent practices or deviating occurrences) in conversation relating to either components of single turns, whole turns or turn sequences, can be identified and gathered in collections (Sacks 1992; Garfinkel 1967; Schegloff 2007). This procedure of 'classical' CA can be regarded as qualitative in that it is explorative (no predetermined practices are examined) and hypothesis-generating (practices are identified). The focus lies on understanding how an ordered (and ordering) structure is established jointly in interaction (this can be a single concrete case). The identified practices may then be generalized by comparing them with other data within a collection, and this comparison can remain qualitative and does not need to be quantified (Sacks 1992; ten Have 2007).

When collections are formed (whether they are based on a certain form, a certain action or a certain sequential position), however, the individual occurrences must always be identified as the realization of a certain practice in order to be included in a collection. The same then applies within a collection of sub-practices that are

differentiated (e.g. the action- and sequence-related differentiation of a certain form). This identification is equivalent to coding; in this regard, Steensig and Heinemann (2015:22) refer to the idea of an "informal coding," which has long been used in CA. According to Stivers (2015:5):

> 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.

However, many works in classical CA carry out not only informal coding, but also informal quantifications. This can be seen in characterizations of individual practices with terms such as *often*, *usually*, *frequently* or *rarely* (e.g. "The preferred form [...] is *usually* short and simple and *often* contains preformed elements", Gülich/Mondada 2008:52, our translation, our emphasis), or *massively*, *ordinarily* or *occasionally*. The last three examples are taken from Schegloff (1993:118),[2] who points out that in this context we are dealing with a common "informal 'quantification'" (ibid.), which (ibid.:119, emphasis in original)

> 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.

The approach described by Schegloff as *informal* quantification is contrasted with methods of *formal* quantification. In the latter, mathematically founded counting processes and statistical analyses are foregrounded, and the result is not a better version of *informal quantifications,* but a different *type* of quantification. Nevertheless, the distribution of phenomena often represents an important aspect of CA in qualitative studies as well. Although the general frequency of a practice does not allow statements about whether the participants in a conversation really prefer to follow this practice in individual cases (see Stivers 2015:8), it does indicate that the practice corresponds to a social norm and that knowing this practice is part of the pragmatic language knowledge of a communicative community; moreover, deliberatively deviating from this practice indicates a certain action. This is a phenomenon that functions against the background of the existence of a social norm (see ibid.:7).

The considerations on the relationship between qualitative and quantitative approaches in CA presented here were developed in the context of a research project on oral argumentation competences of school children between the ages of seven and 12.[3] The background of the study is, among other things, the fact that although those competences must be taught and assessed at school – a requirement of the curricula – there is also a considerable desideratum of research on what can be expected in different grades. One of the challenges of this project was to combine

---

[2] Similar examples can be found in Stivers (2015:6).

[3] The project "Argumentative Gesprächskompetenz in der Schule: Kontexte, Anforderungen, Erwerbsverläufe" ("Oral Argumentation Skills at School: Contexts, Requirements, Acquisition Processes") (duration: 2014-2018), financed by the Swiss National Science Foundation, was headed by Brigit Eriksson and Martin Luginbühl.

qualitative (i.e. also case-oriented) with quantitative (e.g. cohort-related) research questions. Reliably capturing the acquisition of argumentation competence in conversation (only) has a heuristic value if the situational-pragmatic specifics of individual interaction events are taken into account just as much as observations that *disregard* special features of the individual case.

In this project, there was a further, and related, methodological problem: competence-related studies in the context of CA must always aim to describe what Deppermann (2004:20, our translation) described as "factual competence": a person's cognitively anchored potential (competence) can, by means of CA, only be examined on the basis of their actual actions (performance); the latter establishes the factual competence. Competent speakers can handle the global and local requirements of a current conversation, i.e. assess the situation, react appropriately, express their reaction appropriately and so on (for various aspects of general conversational competence, see Hartung 2004; Grundler 2008; Becker-Mrotzek 2009; Quasthoff 2009). Conversational competence can thus be understood as "the share of one of the participants in the joint construction of meaning within the framework of the given conversational activity" (Quasthoff 2009:86, our translation).

A combination of qualitative and quantitative methods of analysis (first implemented in CA by Heritage/Greatbatch 1986) proved necessary in the context of the project, as we were not concerned with describing the competences of individual students,[4] but rather empirically based competence *levels* of oral argumentation in conversation. Our goal was thus to capture which argumentative-oral interactive processes children in each grade typically apply in the context of school exercises for conversational argumentation, and we focused in detail on how justifications, i.e. reasons are given, which in our understanding represents the core of and a condition for argumentation (for such justifications in conversational argumentation, see Gohl 2006; Büker 2008; Grundler 2011:7-47; Bova/Arcidiacono 2013). In addition, we were also interested in whether there are other, possibly more important, relevant variables besides age (i.e. grade).[5] The quantifying approaches discussed in the following are unavoidable because it is impossible to obtain a reliable 'impression' ("*experience* or *grasp*," Schegloff 1993:119, emphasis in original) and generalize the individual cases with regard to single variables (grade, age, gender and so on) given the large amount of very heterogeneous data (180 conversations with a total of 720 children). Reliable statements on the relationship between interactive practices and other variables, and differentiated statements on the frequencies of such practices, are only possible with a large amount of data (which in turn is a prerequisite for the empirical description of competence levels) if qualitative analyses for the identification of relevant practices are combined with quantitative analyses.

It is crucial for our approach that the relevant practices are first identified through a qualitative analysis (Deppermann 2008:107), i.e. that they are based on principles of CA. Thus, our methodological approach is in contrast to various studies on argumentation competence in conversation that rely on normative or theory-based concepts. This also means that the practices we are looking at are not only defined

---

[4] For this purpose, in our opinion the same people must be observed repeatedly and in fundamentally different situations.

[5] Stivers (2015:8) identifies the ability to analyze the relationship between social interaction practices and other variables as one of the main reasons for integrating formal codes into CA.

by surface forms, but also the communicative functions indicated by composition, sequential arrangement and so on (see Heritage 1984 and Selting/Couper-Kuhlen 2001). Thus, the analysis remains compliant to the principles of CA. However, manageable coding requires broad, content-reduced categories in order to enable unambiguous coding without a great deal of time spent on individual cases; this also means that the problem of "hard boundaries" (Stivers 2015:13) remains. This must always be taken into account when interpreting quantitative findings. Nevertheless, the combination of qualitative and quantitative analysis outlined here provides insights that would not be possible with genuine CA methods alone. In the following, we focus on methodological considerations, and not on the description of different competence levels.

2. Data

In the following, we point out the methodological considerations that were crucial in the various stages, from study design to data collection, composition of the corpus and data preparation. But before dealing with the main methodological questions discussed in this paper in more detail, we briefly outline our research interest – oral argumentation.

2.1. Object of Research: Argumentation in Group Discussions

The study (see Hauser/Luginbühl 2017) aimed to make statements about the different characteristics of oral argumentation competence of primary-school children, and for this purpose we chose a setting in which the participating children were given a task that made an argumentative process highly expectable but did not give the children any restrictive instructions regarding conversational actions (see Section 2.2 for more detail). Since our understanding of conversational argumentation had a significant influence on the data collection, data preparation, development of the coding scheme and selection of the analysis questions, we first briefly outline what we mean by oral argumentation in conversation.

Generally, we make the following assumptions with regard to conversational argumentation:

- Situatedness: conversational argumentation consists of complex units of discourse that are embedded as transphrasal and delimitable units/communicative practices in "interactive and pragmatic contexts of interaction" (Morek 2017:70, our translation; see also Deppermann 2006).
- Interactivity: in the production of social contexts, in our context the generating and processing of argumentative actions and positions, at least two interactants who refer to each other with their statements and develop arguments conjointly (co-construction) are involved in spontaneous group discussions (see e.g. Grundler 2011, 2015).
- "Core job" justification: the key feature of argumentation consists of giving and negotiating justifications (Heller 2012:84, our translation).

Based on relevant empirical studies, we understand argumentative discourse units as globally dimensioned, sequentially structured and interactively produced (primarily) linguistic actions in which reasons are given in order to process differing claims of validity (see Quasthoff/Krah 2012; Grundler 2011; Heller 2012; Arendt 2015:21; Morek 2015). Although it is assumed that a controversial "validity-critical [...] reference to a thesis [...] is the core of argumentation" (Deppermann/Lucius-Hoene 2006:142, our translation), it is not always a controversial case that triggers an argumentation. There can also be problematizations of a fact, "divergences of perspective" (Rühl 1999:11, our translation) or open questions, as in the setting described below (see Section 2.2).

We therefore assume that justifications "cannot be located solely in a controversial modification of knowledge, but also in the joint, cooperative construction of knowledge" (Scarvaglieri 2017:146, our translation). This joint deliberation of "pros and cons of a developing assessment" (Kotthoff 2015:83, our translation) refers to *Räsonieren* (reasoning). Mercer (2009:184) calls this *exploratory talk* and defines it as practices in which "partners engage critically but constructively with each other's ideas." Similarly, in Ehlich (2014:46, our translation) we find the term *exploratory argumentation* and the observation that the knowledge systems of the interactants are not necessarily in conflict with each other, but that generalizable knowledge is jointly produced and processed. Depending on the conditions of the communicative framework, argumentations thus differ between various degrees of intensity of the participants' cognitive divergence and thus their conflicting potential moves along a continuum of consensual and adversarial argumentation (see Cohen/Stevens 2017; see also Bose/Hannken-Illjes 2019).

In the development of the coding scheme, which is discussed in detail in Section 3.2. we also focused on sequentiality and the interactive embedding of argumentative actions with regard to justifications, i.e. isolated individual actions are coded, but they are represented by codes that do justice to the argumentative process/interactivity and show various interactive references of argumentative actions (e.g. 'agreeing,' 'disagreeing') to one another. Thus, the codes have their foundation in situated action and its sequential processing.

2.2. Study Design and Data Collection

The data collection was based on the methodological premises of conversation analysis and the requirements for quantification and comparability. For example, we decided not to use preparatory documents (e.g. with possible positions and arguments) in the school classes, and there were no interventions concerning oral arguments by either the teachers or the investigators, since it was not our intention to train argumentation competence and then test the resulting knowledge of argumentation. The students were briefly instructed and could then start working in the group on a task typical for a school environment, as we were interested in how children of different age groups solve an agreement task that does not necessarily require argumentation but makes it highly expectable. Accordingly, we tried to find a suitable task that encouraged the children to engage in discussion and would appeal to children from grades 2 to 6 in Switzerland (in terms of difficulty, comprehensibility and attractiveness of the task, if possible with reference to the children's

everyday lives) and would generate comparable and evaluable data and minimize disruptive factors.

For this purpose, two tasks were developed, each containing several options and aiming at an agreement: the 'Robinson task' and the 'donation task.' In the Robinson task, the students were asked to imagine that they were stranded on a deserted island and were able to dive to the bottom of the sea once with scuba gear and retrieve three sunken objects to ensure their survival on the island. For this they received an illustrated task sheet with 12 objects (pocket knife, bush knife, sleeping bag, woolen blanket, tent, mosquito net, matches, lighter, cooking pot, flares, mobile phone, bandages), from which they had to choose three. Implicit in the task was an examination of the objects with regard to their usefulness. For this reason, a number of objects that were similar in terms of their usefulness were deliberately included in the list.

In the donation task, the students were shown four real projects to which they could donate money: the protection of brown bears in Switzerland (WWF), the protection of bees (WWF), a project for children from financially disadvantaged families (Foundation for Holiday Organization for Children) and their own class fund. The children had to rank these four donation options according to their preference on a four-step 'winners' podium.[6] In one variation of this task (donations without consequences), the students were only asked to imagine that they had 50 Swiss francs and should agree as a group in which order they would donate to which project. In the second variation (donations with consequences), they received 50 Swiss francs as a class, which were subsequently donated for real.[7] The background to these two variations of the same task was the question of whether the manner of argumentative processing changes if the collective decision has actual consequences in real life (see the didactic discussion about the often positively evaluated "authenticity of the occasion for communication"; Becker-Mrotzek 2008:71, our translation).

For both tasks, the children were instructed that the aim was for the group to reach an agreement. In principle, this aim does not need to be achieved by argumentation, but can also be achieved by procedures such as voting. The concept of *argumentation* – and corresponding prior knowledge – was deliberately not invoked, but a targeted choice of words (*discuss*, *talk to each other*, *deliberate in peace*) and the concluding sentence "We will be very interested to hear what you have thought about" ensured that it was clear to the students that the aim was not so much to reach an agreement as quickly as possible, but that they should in some (explanatory) way comment on it.[8] There was no explicit requirement that the conversation should last for any given length of time.

---

[6] This is a variation of the task as it was introduced after the pilot phase. In the original version, the children were only supposed to determine the favored project. However, the argumentative activities were very limited. With the adaptation, a broader argumentative activity could be evoked.

[7] The consequences were emphasized during the instruction and were also made clear after the recordings by signing a confirmation letter to the respective project.

[8] These formulations were adjusted in the course of the pilot project, since it became clear from statements made by some students and based on the analysis of very short conversations that the task had previously been interpreted more as an agreement task and less as a discussion task. The adapted wording in the instruction had a positive influence on the interpretation of the task as a cause for discussion.

After various pilot phases, we decided to concentrate on three variables, which we systematically varied, and we decided to keep other possible variables stable or (mainly for reasons of feasibility) leave them unconsidered.[9] The systematically varied variables were *grade*, *topic* and *consequence of action*. Classes from grades 2, 4 and 6 in German-speaking elementary schools in Switzerland were included (the ages ranged between seven and 14). In this way, we ensured that we could trace aspects of the development of oral argumentation competence. The theme was varied with the Robinson and donation setting as described above. Due to the thus achieved comparability of the data, topic- and task-related aspects could be excluded or controlled for. The variation in the consequences for future actions served to specifically test the assumption that authentic and fictitious communication occasions entail fundamentally different discussion processes. Other aspects were also kept stable:

- four students[10] participated in each group discussion; the groups were assembled by drawing lots;
- no adults were present (peer conversations);
- in all settings, the goal of the conversation was common agreement.

The peer conversations were videotaped with two cameras on tripods, so that the extensive multimodal actions of all four students could be identified and analyzed, which was particularly necessary to capture the frequent simultaneous use of pointing gestures toward the worksheet and local deixis, but also non-verbal practices of indicating agreement or disagreement and justifications in the form of gestures.[11] Following the group discussions, the investigators conducted a follow-up discussion with the four students, asking metalanguage-oriented questions about the discussion that had just taken place. Both instruction and follow-up talks were recorded. In addition to the recordings of the conversations, ethnographic data were collected on the children's exact age, language skills (L1, L2, etc., German as a second language classes) and previous knowledge acquired at school on the subject of *argumentation/discussion*. The teachers provided information about the children's family languages and, in the case of children with German as a foreign language, their German language level. Unfortunately, we sometimes do not have information on language skills or previous knowledge acquired at school (and in some cases on the children's exact age), and the information is not equally precise in all cases. For this reason, this information was not quantified, but rather used in detailed analyses as a supplementary and selectively clarifying knowledge resource.

---

[9] In the course of the pilot project, further variables were discussed and partly tested, such as group size (discussions in groups of 3, 4, 5 and in half classes), communicative order (student-centered peer conversations vs. teacher-centered discussions/adult-child interaction), group composition (systematic composition, e.g. with regard to gender), rules of conversation (procedural rules of conversation can include, for example, rules on the transfer of speaking rights e.g. by raising hands or similar, but have a great influence on the interactivity of the conversations) and function of conversation (goal to be achieved together vs. controversial discussion).

[10] The group size had practical consequences for the data collection, since, depending on the number of children in a class, not all children could participate in the (filmed) group discussions, although their parents had given their consent.

[11] A systematic multimodal analysis of the data was not the subject of the project, but would be worthwhile, e.g. with regard to practices of affiliation and disaffiliation (see Jacquin 2015), alliance building and the co-construction of joint contributions.

The following section describes the composition of the final corpus in more detail.

2.3. Corpus

The corpus consists of 180 videotaped group discussions, and the three variables *grade*, *topic* and *consequence of action* were varied in such a way that a sample of 20 discussions per variable group resulted (see Table 1):

Grade	Topic/Setting	Consequence of Action	Abbreviation
Grade 2	20x Robinson	No	Ro_K2
	20x Donation	Yes	Sm_K2
	20x Donation	No	So_K2
Grade 4	20x Robinson	No	Ro_K4
	20x Donation	Yes	Sm_K4
	20x Donation	No	So_K4
Grade 6	20x Robinson	No	Ro_K6
	20x Donation	Yes	Sm_K6
	20x Donation	No	So_K6

Table 1: Corpus overview

The 180 group discussions were conducted in a total of 53 school classes at 24 schools in six different Swiss-German cantons (Aargau, Basel-Landschaft, Basel-Stadt, Solothurn, Zurich and Zug). By classifying the school locations on the basis of their urban or rural character, 97 of the 180 assessed groups can be described as urban and 83 as rural. By considering different places and catchment areas, we ensured that different social structures were included.

The total duration of the 180 group discussions is 873 minutes. Here, an age-related change can be seen: across all settings, conversations last on average 3.5 minutes in grade 2 and 5.5 minutes in grades 4 and 6. Grades 4 and 6 are very close in this respect – a result that is evident in terms of not only discussion length, but also individual conversation activities such as the number of justifications (see Section 3). With regard to conversation duration, however, there is a big spread, and outlier conversations of 30 seconds as well as 22 minutes are part of the corpus.

Since the composition of the groups was decided by lot, individual groups differ with regard to gender balance, but the overall balance is good: in 74 of 180 discussions, two girls and two boys participated. There were three boys and one girl in 42 conversations, and three girls and one boy in 40. Finally, the corpus comprises exactly 12 groups of only girls and 12 groups of only boys. These group settings make it possible to deal with research questions regarding gender differences in separately compiled sub-corpora. The children were free to choose which linguistic variety they chose in the discussion. Although the teaching language at school is Standard German, it is common for Swiss German children to switch to a Swiss German

dialect during group work. This is also very clearly reflected in our data: in 139 conversations, dialect is spoken exclusively, and in only eight conversations is Standard German spoken exclusively. In the remaining 33 conversations, both varieties are used, which is usually related to the fact that individual children speak German as a first language or have learned German as a second language and therefore do not (or cannot) actively speak dialect.

2.4. Computer-mediated Methods of Data Processing

In the following, we discuss the various software products that we used, some extensively and some only selectively. In order to evaluate oral data quantitatively, it first must be transcribed. Since there is often simultaneous and overlapping speech in group discussions among children, EXMARaLDA's score editor was chosen with the appropriate partiture notation. EXMARaLDA[12] is an XML-based software package containing tools for transcription (score editor), annotation (score incl. flexibly programmable annotation panel), corpus management (COMA) and corpus-linguistic analysis (EXAKT) (see e.g. Schmidt 2004a/b; Schmidt/Wörner 2009). Our corpus was completely transcribed and annotated with the score editor using the CA transcription system GAT 2 (Selting et al. 2011). In the end, a complete transcription for each speaker contained the following tracks:

- [v]-track: v = verbal, indicating speech (usually in dialect);
- [sd]-track: sd = Standard German, for the interlinear translation into normalized Standard German;
- [nv]-track: nv = nonverbal, for descriptions of multimodal activities that are important in interaction and necessary for understanding;
- [a]-track: a = annotation, for argumentation-relevant codes (see coding scheme in Section 3.2).

In addition, a further [nv] track was inserted to record pauses or actions of several speakers – but this track has no speaker assignment and can therefore be distinguished from the speaker-related [nv] track mentioned above. Figure 1 shows a section of a transcript with coding (for the individual codes, see Section 3.2).

On the left, the (anonymous) speaker's abbreviations are shown, and each person has four tracks. First, the transcribed spoken text is reproduced (v; here dialect, e.g. ANN: "ich WÜRDi, ich würdi (so ÄI::S);" ("I would, I would (one of these)")), followed by the Standard German translation (sd, "ich würde ich würde (so eins)") and important descriptions of non-verbal activities (nv, "((points to the mosquito net))"). The next track (a) then shows the different codes that have been assigned to the respective utterances (e.g. "1-for," "2a-mos," "4-nonverbal" in the track "ANN [a]"). Although several codes relating to different aspects of the interaction are assigned simultaneously in this track, and thus a single code combination shows the relationships between the four levels of encoding, these codes can be analyzed individually; both individual codes and code combinations can be analyzed. This rendered it unnecessary to create individual tracks for different code groups. At the

[12] See documentations and publications at http://exmaralda.org/de/

same time, the practicability of the (technical) coding process was taken into account. At the very bottom follow the traces [nv] for non-verbal material that cannot be attributed to a single person (these are usually, as in the above example, pauses) and [k] for further comments by the transcriber (empty here).

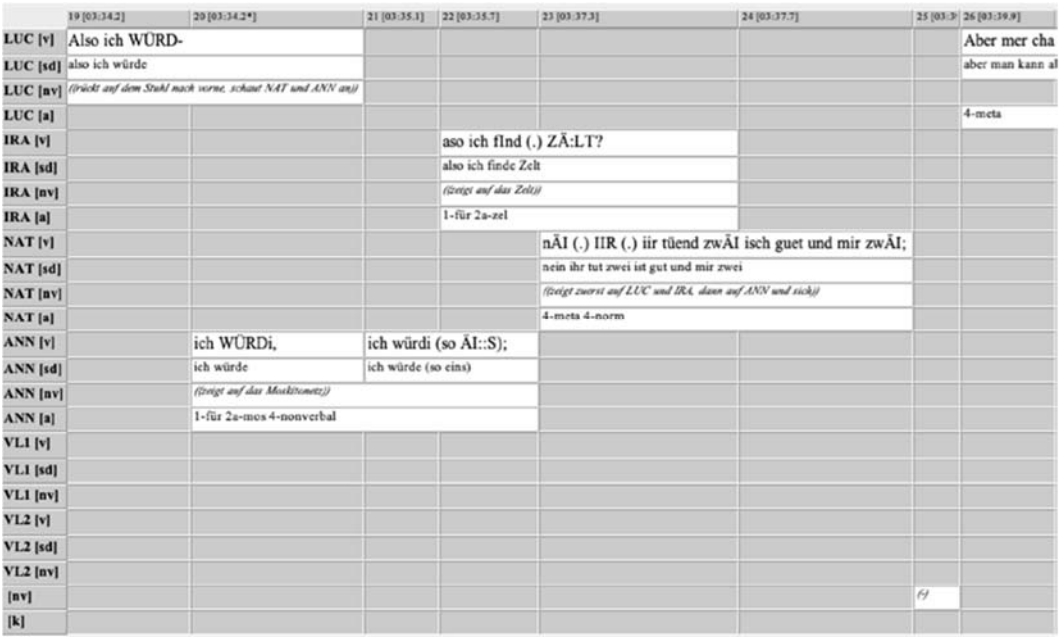


Figure 1: Extract from an EXMARaLDA transcript

The preparation of the data in partiture notation as interlinear text tracks (verbal, Standard German translation, non-verbal) has advantages in the visualization of the simultaneity of spoken language (see Schmidt 2003), but the transfer of the data to other software environments is often problematic. While the partiture notation in EXMARaLDA's score editor is displayed as a continuous transcript that is aligned with the video, the interlinearly arranged tracks must be 'cut' into score sections to fit the page width when exported to other formats for further word processing (e.g. RTF). On the one hand, this decreases readability, as the entire text can be portioned in such a way that, among other things, the interlinear translations, descriptions and/or annotations are no longer displayed one below the other, but in different areas of the score. On the other hand, it has consequences for the further processing of the data. If, for example, software such as MAXQDA[13] or ATLAS.ti[14] is used for qualitative data analysis, the unfavorable wrapping only makes coding and analysis possible to a limited extent and with restrictions.[15] Nevertheless, simple quantitative calculations and analyses of the individual codes in the exported data are usually possible.

[13] See https://www.maxqda.de

[14] See http://atlasti.com/de/

[15] With MAXQDA, for example, a code must be attached to the corresponding text in order to use the various analysis options. Only then can one obtain the codes and the corresponding data extracts in the analyses. However, since most transcripts from EXMARaLDA are accompanied by breaks and connections are spread over several parts of the score, a code can often only be attached to one part of the text, which restricts further qualitative analysis.

Another problem that has to be considered when exporting data from EXMARaLDA and importing it into analysis tools is that, once imported, transcripts cannot be corrected without further ado, or only via detours and losses in the set encodings. This means that the exported transcripts are imported as static text documents that are no longer linked to the video file and must be treated as complete data at the time of export. Schütte (2007:71, our translation) calls this a "killer criteria" for the use of programs such as ATLAS.ti (the same applies to MAXQDA) in speech analysis. This is because the process of transcription is typically dynamic, so that during the analysis individual sequences are often transcribed even more precisely, processed further and supplemented depending on the focus. Since it is advisable to always make the adjustments in the primary document – in this case in the score editor of EXMARaLDA – the file would theoretically have to be imported into MAXQDA (or a similar program) again, but the codes that have already been set would be lost as a result. Due to these difficulties, the encoding of the transcribed data was finally carried out in EXMARaLDA itself using an annotation panel that was specially programmed into the source code in order to work with the original file for as long as possible.

The analysis, which is based on the coding on the one hand and the individual statements on the other, was finally partly carried out with the EXAKT analysis program from EXMARaLDA (for which segmented transcripts are necessary, which must be prepared by EXMARaLDA via COMA). In some cases, MAXQDA was used as a supplement for simple counts. For more complex calculations and visualizations, we also used R[16] and RStudio.[17] Due to EXMARaLDA's XML-based data formats, interoperability is in principle guaranteed. However, the effort required for data preparation in R should not be underestimated, especially for large data sets. In order for the automated analyses to be reliable, the various interlinear tracks and codes must be brought into a searchable and analyzable logical structure. An advantage of R is that further export functions are possible after data preparation, which in turn makes further analysis tools available. Accordingly, the entire data sets could be exported to Excel, for example, for easier analysis and visualization. In addition, TXT files could be created on the basis of the Standard German tracks and used accordingly for the corpus-linguistic tool AntConc.[18]

Overall, it should be noted that the various transcription programs on the market require close examination for one's own research process. In any case, the amount of work involved in systematic preparation for the purposes of quantitative analysis is very high. Since usually not all analysis questions are predefined, it pays to use an interoperable system that enables the flexible transfer of data to other software environments, especially for qualitative studies.

---

16 See https://www.r-project.org

17 See https://www.rstudio.com

18 See http://www.laurenceanthony.net/software/antconc/ and, for an introduction to the software, http://www.bubenhofer.com/korpuslinguistik/kurs/index.php?id=eigenes_AntConc.html

3. Methodological Approach to Data Analysis

The principal aim of the project was to offer a thorough description of oral argumentation competence in children's peer interactions in small groups, especially a comparison of oral argumentation competence across different age groups (see Section 1). The data material was analyzed in detail using CA methods (see Deppermann 2008; Sidnell/Stivers 2013),[19] but due to its structure it can also be examined quantitatively in order to draw conclusions from the numerous individual skills (see Section 2.1) about the competence-related abilities of the age groups studied.

For the latter purpose, the data were, on the one hand, searched by simple word-search queries for their near-surface linguistic phenomena relevant to argumentative activities (e.g. lexical markings of justifications by subjunctives, modal particles and so on). On the other hand, a code was assigned to the individual conversational activities on the basis of a coding scheme (see Section 3.2) in order to be able to quantify larger and more complex argumentative units as well. The development of a suitable coding scheme is particularly challenging if one wants to classify spontaneously uttered turns in a category system that neither goes too deeply into qualitative detail (e.g. conversation-analytical approach) nor remains too much on the lexical surface (e.g. corpus-linguistic approach). It is also necessary to define the phenomena – which should be as clearly describable as possible – with regard to which the corpus should be coded. Coding therefore always entails a reductionist analysis of language data. Since this research is interested primarily in describing the *argumentation-relevant* activities of children, only a few basic codes were developed, which focus on the "core job" of justification (Heller 2012:84, our translation).

In the following, we first describe the theoretical and empirical considerations on which the development of the coding scheme is based and how the conversational data were prepared for a quantitative analysis in order to be able to make statements about age- and grade-related differences in partial argumentation competences. During the development of the coding scheme, we focused on capturing not only argumentation-logical aspects, but also interaction-specific conversation-analytically based aspects. We then evaluate and display the possibilities of the coded data. And finally, we discuss the question of normalization that preoccupied us throughout the analysis: how can conversations of different lengths be sensibly compared to each other, and what kind of reference value can serve as an appropriate basis for comparison?

3.1. Theory-based and Qualitative Approach to Data

Although we first approached the data by means of qualitative conversation-analytical procedures (see below in this section), research papers dealing with the quantification of oral interactions were also consulted in developing the coding system. These include Canary, Ratledge and Seibold (1982) and Canary, Brossmann and

[19] Cf. Hauser/Luginbühl (2015, 2016, 2017); Kreuz/Mundwiler (2016); Kreuz/Mundwiler/Luginbühl (2017); Mundwiler/Kreuz (2018); Mundwiler et al. (2017).

Seibold (1987), who developed the comprehensive *Conversational Argument Coding Scheme* (CACS) based on both theoretical assumptions (e.g. Perelman/Olbrechts-Tyteca 1958) and empirical observations of argumentative interactions. The CACS was developed in the late 1970s and early 1980s "as part of structuration theory work in communication [...], especially within the structuration approach to argument in group deliberations" (Canary/Seibold 2010:7). The coding scheme should therefore be applied to spontaneous oral interactions and thus contains not only codes that are subject to the argumentative logic, but also those that represent the characteristics of argumentative elements in their concrete use or sequentiality as well as the argumentative behavior of the interactors. The coding scheme thus integrates, on the one hand, the composition of an argument (i.e. the development of an argument through the specific (simple or complex) structures of one or more interlocutors) and, on the other, sequential sequences (see "systematic repetitions of act-to-act argument behaviors" such as *Converging and Diverging Sequences*, Canary/Seibold 2010:21). Since then, the CACS has been used in diverse contexts of interactive argumentation,[20] and today it is a widely used content-analytical method for investigating arguments "involving diverse communicative interactions" (Canary/Seibold 2010:7).

The considerations of Kyratzis et al. (2010) were also usable, with adaptations, for our coding scheme and compatible with our understanding of argumentation. In their data on interactions in kindergarten groups (ages three to six), Kyratzis et al. observed that justifications do not only emerge from (potential) differences of opinion, but can also support the statements of the partner, for example to legitimize their move in a game. Based on these observations, they developed three motivations for justifications (Kyratzis et al. 2010:126ff.):

- Opposition (the rejection of an idea is justified)
- Validation (the idea of the partner is justified)
    - "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 reasoning to justify the agreement"
    - "the speaker elaborating on the partner's suggestion, and using a reasoning to justify the elaboration"
- self-expansion (one justifies one's own idea)

As in the CACS, individual (explanatory) activities are understood in their sequential context and coded with regard to their interactional status. One of the most important paradigms of CA – sequentiality – is thus taken into account. Likewise, both the CACS and Kyratzis et al.'s (2010) scheme refer to *specific* data of spontaneous interactions and describe practices using authentic material. However, this does not

[20] Cf. e.g. investigations of group counselling (Seibold/Meyers 2007), decision-making discussions between jurors (Burnett/Badzinski 2000), conflicts between friends (e.g. Semic/Canary 1997, 2003), problem-solving discussions between married couples (e.g. Canary et al. 1987), political discussions on television (e.g. Brossmann/Canary 1990) and ethnopolitical group conflicts (Ellis/Maoz 2002, 2007) (cf. details in Canary/Seibold 2010:7f.).

make it possible to adopt the proposed codes unchanged for one's own data analysis. We could *adapt* some of these codes, but many had to be *newly developed* in joint data sessions based on the inherent logic of our own data. The developers of the CACS are well aware that their codes cannot simply be adopted unchanged (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.

If one wants to consider the inherent logic of the collected data, therefore, the development of codes always requires a *qualitative pre-analysis* of those data. The aim is to gain an overview of typical characteristics of the interactions and determine which codes could prove to be relevant to the specific research interest – oral argumentation competence and their age- and grade-related changes. Thus, we first analyzed the data qualitatively on a global scale by means of CA with regard to the assumptions about argumentation presented in the introduction, e.g. typical conversation phases (Kreuz/Luginbühl/Mundwiler 2019) and conversational patterns. We then analyzed data with regard to meso- and micro-level phenomena, such as individual turns (e.g. positioning), the processing of various argumentative 'jobs,' local argument structures and lexical markings of justifications (see Hauser/Luginbühl 2015, 2017; Kreuz/Mundwiler/Luginbühl 2017; Hauser/Kreuz 2018; Mundwiler/Kreuz 2018; Mundwiler et al. 2017). On the basis of the detailed overview thus obtained, relevant codes crystallized with which we were also able to investigate argumentative activities quantitatively (see Section 4.1). The development of the final coding scheme was subject to a multi-stage recursive process in which the codes had to be tested and modified multiple times (e.g. regarding their level of detail).

3.2. Quantitative Approach to the Data: Coding

On the basis of the theoretical assumptions described above and our own qualitative analyses, a codebook was developed that incorporates coding at both the level of individual conversational activities and the lexical level related to argumentation.

One of the codes at the level of conversational activities is the so-called *potential trigger for argumentation* (Spranz-Fogasy 2006) – recorded in the coding manual as 'thematizing,' (code 1-th) 'positioning for' (1-für) or 'positioning against' (1-geg) (level 1). In the donation setting, the specific nature of the task (ranking the projects in a ranking list) required the addition of 'placing' (1-platz). We have not done multiple coding within the same level for any of the levels of code discussed in the following. Ambiguous cases (e.g. as a result of discontinued formulations) were coded as 'thematizing.'

Example Codes Level 1[21]

Utterance[22]	Code	Meaning
S1: Aso BUSCHmesser, **(so, bush knife)**	1-th	The child thematizes/names an object without taking a personal stand.
S2: Ich bin für das ZELT. **(I am for the tent)** S2: verbAnd braucht man eigentlich NICHT. **(we don't really need the bandages)**	1-für 1-geg	The child expresses their opinion and argues for or against an object.
S3: BIEne auf platz Eins; **(bees in first place)**	1-platz	The child nominates a project for first place.

In order to take into account the thematic reference of a turn, a code was assigned according to the 12 items to be discussed or the four donation projects (code on level 2a and 2b, respectively). If possible, this was *always* assigned – not only when the thematic reference was explicitly mentioned, but also when it could be plausibly reconstructed from the local context.

Example Codes Levels 1 and 2a/b[23]

Utterance	Code	Meaning
S1: Aso BUSCHmesser, **(so, bush knife)**	1-th 2a-bus	The child thematizes/names an object without taking a personal stand.
S2: Ich bin für das ZELT. **(I am for the tent)**	1-für 2a-zel	The child expresses their opinion and argues for the tent.
S3: BIEne auf platz Eins; **(bees in first place)**	1-platz 2b-bie	The child nominates the bee project for first place.

[21] In this paper, the Swiss German original is given in examples, followed by the English translation. In order to ensure readability, the English translation is not word for word, but it nevertheless closely follows the original.

[22] The data examples are prototypical examples from the corpus, transcribed according to GAT 2 (Selting et al. 2011). Since the aim here is to illustrate the content-analytical coding, the examples have been simplified and adjusted so that they can fulfil the illustrative purpose. Therefore, no individual sources have been indicated.

[23] For the codes, the following abbreviations were used for the objects and the donation projects. In the Robinson setting (level 2a): pocket knife = '2a-sac,' bush knife = '2a-bus,' sleeping bag = '2a-sch,' wool blanket = '2a-wol,' tent = '2a-zel,' mosquito net = '2a-mos,' matches = '2a-str,' lighter = '2a-feu,' pan = '2a-koc,' flare = '2a-leu,' cell phone = '2a-han,' bandages = '2a-ver.' In the donation setting (level 2b): project for the protection of bears = '2b-bär,' project for the protection of bees = '2b-bie,' foundation for the organization of vacation for children = '2b-kin,' class funds = '2b-kla.' In both settings, there was an additional collection category ('2a-andere'or '2b-andere' = 'other') for references to further related semantic fields, other contextual (explanatory) connections, unspecific formulations or ambiguous (back) references.

As was shown in the introduction and Section 3.1, a) the core of argumentation is the interactive activity of justifying, which b) is oriented in varying degrees of divergence (or convergence) to the previous expression. In order to do justice to these two facts, a further code was developed at the language-action level that takes into account the interactive embedding of justifications in conversation with their respective orientations. Justifications were categorized as 'initiating' (3a-in), 'validating' (3a-zu), 'oppositional' (3a-ab) and 'multiple-aligned', i.e. oppositional-validating justifications (3a-mehrfach)[24] (level 3a), and unclear cases were not categorized. This code 3a is based on the conversation-analytical requirement of a sequential data analysis and should consider the procedure of a turn-by-turn analysis. In addition, it should also provide information on the extent to which the children take up positions already established in conversation or tend to plausibilize their own positions with initiating justifications. Similarly, corresponding analyses give indications as to whether a discussion is primarily controversial or consensual.

Example Codes Level 3a

Utterance with previous utterance	Code	Meaning
01 S1: Ich bin für das ZELT. **(I am for the tent)** 02 S2: ja aber wir kÖnnen auch auf einem BAUM schlafen? **(yes but we can sleep in a tree as well)**	1-für 2a-zel 2a-zelt 3a-ab	The child gives an oppositional justification in relation to the previous utterance.
01 S1: nEhmen wir die LEUCHTrakete; **(let's take the flare)** 02 S2: jA weil dann kÖnnen wir NOTsignale senden. **(yeah, because then we can send a distress signal)**	1-platz 2a- leu 2a-leu 3a-zu	The child expresses a validating justification in relation to the previous statement.
01 S1: wir kÖnnten die WOLLdecke nehmen? **(we could take the wool blanket)** 02 S2: und das bUschmesser brauchen wir zum JAgen; **(and we need the bush knife for hunting)**	1-für 2a-wol 2a-bus 3a-in	The child initiates a justification for the bush knife, i.e. a new theme is introduced without making any argumentative reference to the previous expression.

[24] This includes justifications in which a single speaker strings together – connects, relativizes or modalizes – several justification contexts; we also included the weighing of several projects against each other.

Furthermore, justifications were coded in terms of not only their interactive embedding, but also their lexical marking (explicitness) (code 3b). Meyer (1975:56f.) suggests a subdivision of different types of explications in justifications: zero explication (no marking), vague explication (particles, relative clauses)[25] and full explication (conjunction). In order to make the analysis of explanatory markers as simple and unambiguous as possible, only the full explication was used as code for the available data.

Markings – according to our impression after examination of the material – are produced by the interlocutors as marked justifications mainly as causal (3b-kaus), conditional (3b-kond) and final (3b-fin). For other forms of marking, a collective code was assigned (3b-rest), to which unclear cases were assigned as well. Code 3b refers exclusively to phenomena observable on the surface, i.e. purely lexical phenomena.

The coding scheme focuses on the argumentation-relevant phenomena of interest but has been extended by an additional level 4. Level 4 codes describe further noticeable conversational activities. These codes were set mainly to enable quick retrieval for future qualitative analyses but were not set systematically. Level 4 codes describe e.g. meta-linguistic utterances, shifts of the quaestio, normative references and co-constructions.

Turns consisting of simple consent and rejection particles (yes/no, true, exactly) are relevant, e.g. for speaker positioning (see e.g. Felton et al. 2015 for dyadic conversations), but they were not coded due to their ambiguity and their multiple functions (consent, vague positioning, for example by a prolonged yes or continuer, and so on).

For illustration purposes, an example of a continuous sequence with its corresponding codes is given:

Example Codes Levels 1, 2 and 3

Utterance	Code	Meaning
01 S1: HANdy? **(cell phone)**	1-th 2a-han	The child thematizes/names the mobile phone without taking a personal stand.
02 S2: ja Ich nehms HANdy. **(yes, I take the cell phone)**	1-für 2a-han	The child expresses his/her opinion and speaks in favor of the mobile phone.
03 S3: ja zum ANrufen; **(yes, to make calls)**	2a-han 3a-zu 3b-fin	The child expresses his/her agreement with the previous statement and marks it as final.
04 S1: aber wAs ist wenn wir gar keinen emPFANG haben, **(what if we don't have any reception)**	2a-han 3a-ab 3b-kond	The child expresses an oppositional justification in relation to the previous expression and marks it as conditional.

[25] The semantically vague explication results from the use of particles or relative clauses, i.e. it is not clearly marked, but also not completely unmarked (example of the marking by means of a particle: "I stay at home. It is raining, you know").

05 S3: stImmt. 06 dann nEhmen wir die lEuchtraketen weil da können wir sigNAle senden; **(that's right in that case we'll take the flares because we can send signals with them)**	1-für 2a-leu 3a-in 3b-kaus	The child initiates a marked causal justification for the flare because it introduces a new theme and introduces a justification for this new theme, which is why the justification remains without direct argumentative reference to the previous statement.

The entire corpus was coded according to this scheme by four student assistants previously trained in EXMARaLDA (see Section 2.4) and subsequently checked by the project staff. Nevertheless, the coding often reached its limits (see discussion in Section 5), as certain turns turned out to be too complex and context-dependent to be adequately described by setting a specific code (double coding should be avoided if possible). Furthermore, depending on whether the videos of the discussions were considered in the coding process, there were sometimes different annotations. In principle, the corpus was coded based on the transcripts and only in cases of ambiguity or selective tests were the videos consulted. Only then did deviations become apparent, making it clear that the videos always represented the more detailed version of the communication event. However, due to time constraints, it was not possible to use the videos as a basis for the entire encoding process. Even though quantitative analyses are helpful and, in our opinion, indispensable in the competence-oriented analysis of extensive data, and even though the results generally indicate trends, the counting and mathematical calculation of codes should never – and should never be intended to – conceal the heterogeneity of the interactions.

3.3. Normalization of Oral Data

For quantitative analysis, data and codes are typically counted and compared to each other. In order to establish comparability, relative values are necessary, and a reference value is therefore needed that can be used as a reference value for all subcorpora in order to place the partially heterogeneous data in relation to each other. This is referred to as the *normalization* of data (see e.g. Bubenhofer 2009:150). In the present corpus, there are to a certain extent standardized and thus comparable data, since starting from a stimulus (instruction) and a target (agreement within the group), discussions have been elicited in which certain aspects (e.g. group size) have been kept stable over the entire corpus and other aspects (e.g. grade, setting with or without consequences of action) have been kept stable within the sub-corpora. Nevertheless, we find – unsurprisingly – a great deal of heterogeneity, for example in terms of the length of the conversation, the speaker distribution and the length of the turns.

One possible reference value is the number of words, as is common in corpus linguistics, for example (see e.g. Bubenhofer 2009:150; Scherer 2014: 40f.). The problem with peer conversation, however, is that there are sometimes shorter or longer passages in which the children do not deal with solving the task, but talk about private matters or about the camera and the recording situation etc.. Such side sequences or the aforementioned individual differences in speaking behavior can therefore greatly distort the picture. Nevertheless, there are certain questions for which a normalization based on the number of words is useful, and this reference

value was used for corpus-analytical lexical queries, for example. In principle, however, our coding allows us to hide side sequences if necessary and thus include only argumentation-relevant speaker contributions in the analysis (see also below).

The duration of the conversation as a possible reference variable is also problematic, as some children speak more quickly or slowly and there are also big differences in the pauses between conversations. The duration of the conversation would be easy to determine, but due to the variances it can be used for normalization only with restrictions.

A further approach is to use the number of coded discussion contributions as a reference value, as this enables us to create a comparable unit from a CA-based action- and argumentation-related point of view. This means that it is not the length of a contribution to a conversation (e.g. number of words) that counts as a reference, but the type of language action (e.g. validating justification, meta-discursive action or similar). It should be noted, however, that due to the decision to encode statements according to conversation-related actions, the length of the encoded contributions remains unnoticed and can therefore vary considerably. For example, a justification may consist of short, co-constructed partial utterances by several speakers, but it may also include longer narrative discursive units by a single speaker.[26] In addition, for further analyses, side sequences (of varying lengths), which are interesting in terms of conversation structure and references to norms or the like but are not directly related to the (argumentative) discussion, were also partially coded. Accordingly, not all coded contributions are equally relevant for the analysis of oral argumentation competence.

Due to this risk of bias, the number of task-related conversation contributions was used in calculations in most cases. Accordingly, only those coded contributions to the conversation were included in the calculation in which a relation to the discussion task is recognizable (see also *game-related utterance* in Domberg et al. 2018 and *discourse moves* in Felton et al. 2015). This ensures that side sequences or conversation-organizing procedures do not distort the values.

We also performed the *same* search query with different reference values to test how much the chosen normalization procedure affects the relative values. The results were partly almost identical. Since our study deals with a comparatively large amount of data, the aforementioned differences in speech rate or (non-)inclusion of side sequences seem to be unproblematic for the most part.

4. Quantitative Analysis and Results

In the present study, simple statistical methods were used to describe the conversational data on the basis of the transcripts and the coded transcripts. Various methods were used, primarily to make comparisons and obtain results based on values, graphics and more complex visualizations, but also to generate hypotheses and carry out new qualitative analyses on the basis of these hypotheses (see e.g. Section 4.2).

[26] In this way, a co-constructed justification is more significant than a long monological one, i.e. the quantities not only can be related to the number of justifications, but are also influenced by the degree of co-construction; moreover, they do not tell us anything about the quantitative average speech length (the utterances can always be very short or very long; high numbers do not mean that the children express themselves longer on average).

We now present selected search queries and discuss the significance of the calculations. An important aspect for analysis in general, but especially for the combination of quantitative and qualitative methods, are visualizations. Visualizations can have a depictive, illustrative character, or they can fulfil an idea-generating function.

4.1. Research Questions and Analysis across the Entire Corpus

In the context of the research question(s) on the development of oral argumentation competence at different ages and on the basis of the literature and our own qualitative case analyses, we defined various sub-areas that, on the one hand, are relevant for oral argumentation and from which, on the other hand, we expected additional indications through quantitative approaches. The starting point of our analysis is therefore qualitative analysis, with which we identified aspects in our corpus that are relevant for oral argumentation competence, but we are not able to make reliable statements for the whole corpus (or with regard to each age group) based on those qualitative analyses. In the following, we present an analysis of the main topics of *modalization and perspectivation*, *complexity of argumentation* and *interactivity*. Graphs are used to illustrate results and trends across the *entire* data corpus.

4.1.1. Modalization and Perspectivation

Modalization and perspectivation are particularly important in the research literature on argumentation, especially when taking into account the personal level of competence (see above all Grundler 2011:293ff.). On the one hand, modalizations can be expressed by the verb mode in the subjunctive, whereby the degree of factuality can be limited or varied with reference to (a lack of) knowledge, norms, desires or the like (see e.g. Diewald 1999; Grundler 2011:293f.; Nuyts 2001; Redder 2009:91; Schmitt 2002:91; Schwitalla 2012:168-172). On the other hand, in (argumentative) discussions modal particles are often used to produce opinions or facts in a mitigated form, among other things also as a face-saving strategy (see Grundler 2011:295; Locher 2004:113). In this context, modalizations can be used to mark one's own perspective as (non) negotiable or, for example, to indicate the extent to which other perspectives are taken into account in one's own position.

In the following, we provide examples from the qualitative analysis as well as our tentative conclusions that have resulted from it. In the following example of a grade 6 Robinson setting (see also Kreuz/Luginbühl 2020), modal particles are not only used in the initial statement on the topic under discussion ("bush knife"). The use of modalities also has a significant influence on the further course of the conversation, since it increases the likelihood that the next speaker will also employ modalization, resulting in convergent joint reasoning (see transcript next page).

The bush knife (01) thematized by Jolina is taken up by Rico, who expands upon it with his personal opinion ("I think maybe a bush knife could be important," 02). In doing so, Rico draws on various linguistic forms that contain a modalizing function. For instance, he uses the particle "maybe" ("vielleicht"), which instead of a fact merely traces a possibility or even speculation, which is negotiable and sug-

gests a not yet definitive opinion of one's own (epistemic uncertainty marker, Bergmann 2017:176, our translation). Also, the particle "noch," which is typical for Swiss German, expresses in this context a mitigation of the statement.

Transcript 1: Ro_K6_HZ_G2a_F36-F38 (Alessio, Jolina, Laurin, Rico)

```
  01 JOL: BUSCHmesser;
          bush knife
          ((to Alessio))
→ 02 RIC: BUSCHmesser find ich vielleicht noch wIchtig;
          I think maybe a bush knife could be important
→ 03 ALE: (-) villecht [kame mit dem auch HOLZ] hacke und so;
          maybe one can chop wood with it or something like that
  04 RIC:              [um TIEre zu kIllen;   ]
                        to kill animals
```

The modalizing character of this sequence continues after the turn-taking. In the co-constructed addition of a justification by Alessio ("maybe one can chop wood with it or something like that," 03), there is also a modalization effort (marked by "maybe" and "or something like that"). With the use of "maybe," Alessio recycles structures of his interlocutor and thus not only presents his justification as a co-constructive part of Rico's statement, but also shows – through the formally and functionally similar modalization – his cooperativity (see, for example, "behavioral alignment," Pickering/Garrod 2004). This also renders his statement negotiable and expandable and represents – despite the production of a content-wise elaborated contribution – a *down-grading* of his own epistemic status, which assures the partner of his face-saving action (see Grundler 2011:295; Locher 2004:113). Both the co-constructed justification and the use of modal particles show Alessio's efforts to maintain an assertive conversation modality and his "concern for striking the balance between independence and affiliative cooperation" (Rafal 1996:287).

The conversations of grade 2 groups, in contrast, present themselves – according to our impression based on the qualitative analysis – more often by "markings of absoluteness" in the propositions, sometimes with the marking of an epistemic authority, e.g.:

Transcript 2: Ro_K2_WB_G2a_F1-F3 (Sven, Cecilia, Lara, Björn)

```
01 SVE: aso ich glaub ich WÄISS wa mer [drIngend müend usenÄÄ; ]
        so I believe I know what we urgently have to take out
```

as well as turn changes (Knoblauch 1995:122; Morek 2015), which harden dissent:

```
05 BJÖ: NEI:
        no
        ((still with reference to the fireworks))
06 CEC: NÄ:-
        nah
        ((still with reference to the fireworks))
07 SVE: dOch s FÜÜRwerch.
        yes the fireworks
        ((taps the signal flare repeatedly))
```

```
08 BJÖ: lUag ich zEig dir mol;
        look I'll show you
        ((with raised index finger))
09      es ZÄLT de hEmmer,
        a tent and then we have
        ((taps the tent multiple times))
10 CEC: denn chÖmmer SCHLA (.) denn chönd-
        then we can slee(p) then we can
        ((shrugs shoulders)
11 SVE: Aber ES (.) me wäiss jo nId öb,
        but it one does not know if
```

In the analysis of the younger children's conversations, it becomes clear, as in this case, that they generally invest less effort into modalization and that their argumentation has more competitive turns that are only slightly mitigated.

Thus, the children in grades 4 and 6 make frequent use of modalizations in connection with justifications (see the following elaborations on quantitative results). These children not only explicitly mark their validating justifications as a proposal to create plausibility through the use of various modalizations, but also show that a joint discussion for and against is negotiable. According to Kotthoff, this is "typical for the mode of reasoning and the development of a joint stance" (Kotthoff 2015:86, our translation). The use of modalizations therefore demonstrates the interlocutors' desire to cooperate – in the data of the higher grades it could also be observed that in most cases initial statements that have already been modalized are usually followed by validating justifications, which in turn are also modalized almost mimetically and with a face-saving function (see also Transcript Example 1).

The interest was now to examine the extent to which age differences can be determined across the whole corpus in the area of modalization and perspectivation, whether the three tested settings have an impact and whether there are differences between boys and girls.

The following section first describes the methods used to investigate these issues quantitatively.

Quantitative Methods

To investigate the use of modalization and perspectivation, the frequency of near-surface lexical units (subjunctive forms and mitigating modal particles) was determined[27] and tested for differences between grades, settings and gender. Since the focus is always on argumentation, only those incidents were included that occurred either in positionings (code 1) or justifications (code 3a). The comparison of frequencies was carried out separately for positionings and justifications. This analysis was thus a corpus-linguistic approach, which, however, additionally referred to

[27] Evaluated were the subjunctive forms *würd(en)* (would), *könnt(en)* (could), *sollt(en)* (should), *hätt(en)* (would) and *müsst(en)* (must), as well as the mitigating modal particles *vielleicht* (perhaps), *eher* (rather), *manchmal* (sometimes), *nicht so* (not really) and *eigentlich* (actually). The Standard German translations were used for the analysis in order to avoid having to take dialectal variances into account in the search query.

specified conversational activities that were identified by codes.[28] In order to compare the frequencies, different statistical methods were applied, which are briefly explained below (see also Lijffijt et al. 2014).

For the comparison of frequencies, the Pearson's chi-squared test is an obvious choice. The procedure tests whether there is a significant relation between two nominally scaled variables. In our case, the possible relation is between, on the one hand, a variable that indicates for each element (word or sequence) whether it corresponds to the respective search category or not (is a word a modal particle or does a modal particle occur in a sequence, respectively) and, on the other hand, the group (e.g. grade). These variables are used to create a contingency table that represents the frequency distribution in the individual groups. The chi-squared test compares these observed frequencies with the frequencies expected if the group has no influence on the frequency distribution. If observed and expected frequencies differ significantly, it can be assumed that the variables are interrelated. With high numbers of cases, however, the chi-squared test becomes significant even with relatively small differences. Thus, in our analysis, significant effects occur several times, while the effect strength, i.e. the measure for the degree of correlation (in our case Cramer's V was calculated) indicates that the relation is only weak. This is also due to the fact that all words in the searched sequences serve as a comparative value for the frequency of the occurrences.[29]

To validate the results of the frequency comparison, it was therefore additionally determined in which of the conversations one of the search terms occurs. These frequencies can then also be compared with regard to the influence of the group using the chi-squared test. This assumes, of course, that a search term does occur at all in a sufficiently large number of conversations, but this is the case in our corpus. Since the reference value for frequency here is the number of conversations, the case numbers remain significantly lower. If both methods now point in the same direction, this can serve as additional validation of the results.

The chi-squared test is a so-called omnibus test. A significant result therefore only indicates *that* a connection exists, but it does not say which groups differ from each other (which is particularly relevant when there are more than two groups). Another problem may be that the very different lengths of the conversations distort the results, for example if a group has a particularly long conversation.

One way to take better account of the heterogeneity of the data was to determine the normalized frequency of the search term for each conversation (e.g. number of modal particles per hundred words). The normalized frequencies – the group means – can then be compared, whereby the differences in conversation length are taken into account by the normalization. Since in most cases the data within the groups could not be regarded as having a normal distribution, non-parametric tests were used for this purpose, which do not require specific distributional properties of the values. In the case of two groups, the Mann-Whitney U-test was used, while in the

---

[28] A similar procedure was applied to the question of the explicitness of justifications. There, we were interested in the frequency of explicitly linguistically marked explanatory contexts differentiated according to causal (e.g. because, therefore, because), final (e.g. so that, in order to), conditional (e.g. if ... then, if, in case of), and further (rest group) markings.

[29] In principle, one way to improve this situation would be to consider only those positions where the search terms can occur in a logical manner. However, this is hardly objectively feasible given that it is spoken language.

case of more than two groups the Kruskal-Wallis test was used. Especially if there were more than two groups, in order to determine between which groups significant differences exist, pairwise comparisons were subsequently made using the Dunn test (corrected for multiple comparisons according to Holm). Some analyses showed a clear relationship overall, but in the pairwise comparisons only a difference between grades 2 and 4, but no significant difference between grades 4 and 6, was found. In such cases, grades 4 and 6 were combined into one group and compared with grade 2 to validate the 'gap' between these grades. In the following, unless otherwise stated, the results of the procedure just described are presented.

Another method of analysis is to fit a regression model to the data. In a regression analysis, the parameters of a mathematical function that best describes the data are determined. In our case, we are interested in the extent to which the occurrence of an event (e.g. the occurrence of a subjunctive or modal particle) depends on the respective factors such as the grade. We are therefore interested in the probability of an occurrence of an event depending on certain factors. Since the dependent variable is thus binary (i.e. a feature, such as the presence of a modal particle in an utterance, is either present or absent), a logistic regression is used here. To account for the heterogeneity of the data, a mixed logistic regression was applied, which allows random factors to be considered in the model. This makes it possible to take into account not only the so-called fixed factors (e.g. grade) but also those factors that may also have an influence on the probability of occurrence but result from the random selection of the sample. In our case, both the respective conversation and the individual speaker were included in the model as random factors. This reduces the risk of assuming an effect of fixed factors that is in fact primarily caused by differences within the groups. Usually a 'base model,' which does not consider the factor of interest (e.g. grade), is first fitted to the data. This base model is then compared with a model including the factor of interest. This allows to determine whether including the factor of interest in the regression formula provides a better explanation of the structure of the data, which would indicate that the factor contributes to the structure of the data.

In the following, the results are reported for some aspects where significant differences were found. It became apparent that the different processes generally revealed similar patterns. As a result, we do not elaborate on the totality of the results of the different procedures. However, it can be assumed that for the results presented, the other methods also point in the same direction.

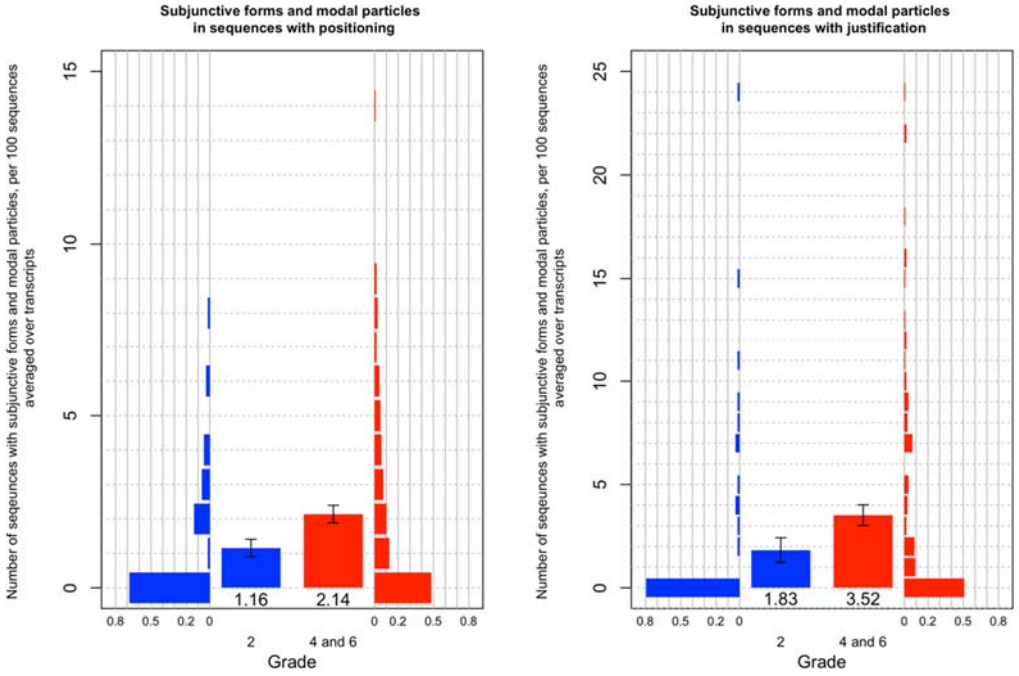
Results

The clearest results are seen in the relative use of *subjunctives* by age. In both positionings and justifications, all of the procedures described above confirm a significant correlation between grade and frequency of use of subjunctive forms (all $p < 0.05$). The pairwise comparisons indicate differences between grades 2 and 4 and between grades 2 and 6, but no significant differences between grades 4 and 6. Grades 4 and 6 were therefore combined into one group and compared with grade 2. The comparison of the frequency per 100 words averaged over conversations shows a significant increase in the subjunctive forms from grade 2 (mean values positioning: 2.31; justification: 0.92) to grades 4 and 6 (mean values positioning: 2.79; justification: 1.24; Mann-Whitney U-test positioning: $U = 2738.5$, $p < 0.01$; justification: $U = 2597$, $p < 0.01$).

The use of *mitigating modal* particles (such as *eher* (rather), *manchmal* (sometimes) or *nicht so* (not really)) in positionings also indicates a significant relationship between frequency and grade (all $p < 0.05$). The pairwise comparisons show that the frequency per 100 words in grade 6, with an average of 1.01, is significantly higher than in grade 2, with an average of 0.62 (Dunn test: $Z = -2.61$, $p < 0.05$, corrected according to Holm). The differences between grades 2 and 4 and between grades 4 and 6 are not significant. The high proportion of discussions in which no mitigating modal particles were used in positionings is also striking: in grade 2 it is 47%, and in grades 4 and 6 still 21% and 20%, respectively (Pearson's chi-squared test: $\chi 2 = 12.89$, $p < 0.01$, Cramer's $V = 0.27$).

Furthermore, search queries were also conducted to identify positionings and justifications in which both (at least) one subjunctive form and (at least) one mitigating modal particle occur. For when used in combination, it can be determined more clearly whether the utterance is modalizing or perspectivizing. With regard to the grades, it can be seen, as before, that grades 4 and 6 differ only marginally. They are therefore considered together and compared with grade 2. In this analysis, the number of words could no longer be used as the normalization variable, since we are looking for combined occurrences within an annotated sequence (occurrence of subjunctive form and mitigating modal particles in positionings and justifications). Therefore, the number of annotated sequences served as a basis for normalization. Figures 2-a and 2-b (on the next page) show that the proportion of utterances containing both subjunctive forms and mitigating modal particles increases significantly in positionings and justifications.

In Figures 2-a and 2-b, the two middle bars represent the average frequency of sequences with both subjunctive form(s) and modal particle(s) per 100 annotated sequences, on the one hand in sequences with positionings (code 1, Figure 2-a) and on the other in sequences with justifications (code 3a, Figure 2-b). The mean was calculated over all settings. The error bars indicate the standard error, which provides information about the variance of the data. In addition, the distribution of the data is shown in each case as a histogram, which provides information about the proportion of conversations in which the corresponding number of sequences with subjunctive form(s) and modal particle(s) occurs. The graph should therefore be read in such a way that, for example, in grade 2 an average of 1.2 sequences contain both a subjunctive and a modalization per 100 positioning sequences, whereas in grades 4 and 6, 2.1 sequences per 100 positioning sequences do so. In addition, the histograms show that in grade 2 there are significantly more conversations in which the combination of modalization and subjunctive does not occur at all (histogram bar at value 0) (Pearson's chi-squared test: $\chi 2 = 7.0$, $p < 0.01$, Cramer's $V = 0.20$).



Figures 2-a (positionings) and 2-b (justifications):
subjunctive forms and mitigating modal particles in grade 2 versus grades 4 and 6

The use of utterances with subjunctive forms *and* mitigating modal particles doubles from grade 2 to grades 4 and 6 in both types of action, and the differences are significant (Mann-Whitney U-Test: positionings: $U = 2840$, $p < 0.05$; justifications: $U = 2504.5$, $p < 0.01$). If we compare this combined analysis with the individual analyses above, it is noticeable that different and partly contradictory scopes for interpretation open up: measured by the number of words, a significant increase in the use of mitigating modal particles from grades 2 to 6 is only evident in the positionings, but not in the justifications. Moreover, the subjunctive proportions per 100 words are higher in positionings than in justifications (see above). Measured by the number of annotated sequences, however, combined search queries (subjunctive forms *and* mitigating modal particles) show that an effect is definitely evident in the justifications, and the shares of subjunctivized *and* modalized utterances are significantly higher in the justifications than in the positionings (see Figure 2-b versus 2-a).

These contradictory results that are due to the different normalizations of the data show the limits of conclusive statistical analysis of coded conversational data. Even though the distribution of modal particles and subjunctives in the individual *turns* is contradictory, the result is consistent with regard to the distribution in the individual *grades*. A (second) qualitative look at the data with a focus on the actual, context-dependent use of modal particles and subjunctives could reveal possible reasons for their increased use in the higher grades. The qualitative analysis (see Transcript Examples 1 and 2) suggests that older children are increasingly developing an awareness of double and thus redundant markings. One reason for this is that they link their utterance strongly to the previous utterance of their interlocutor and 'recycle' their modalizations, but at the same time they also present the idea they themselves introduce in the manner of the established conversation character.

Certainly, the children are increasingly able to use subjunctives due to their progressing general language acquisition – but the data show that they already do so in the second grade. Interestingly, the increasing use can obviously not (only) be justified by the isolated acquisition of a specific construct consisting of vocabulary and grammar, but (also) by the acquisition of competence at the contextual-interactional level of argumentative-discourse units – i.e. by the situation-appropriate use of subjunctives and modalizations for the functional solution of the task (namely, finding a common consensus) in the current execution of the conversation. As Hartung (2004:50, our translation) has argued regarding interactional competence,

> Interactional competence is the ability (1) to arrive at an appropriate assessment of the current situation and the local expectations of the interlocutor at any point in a conversation, (2) to find a reaction with a high probability of success that is appropriate to one's own interests and means of expression and (3) to express this reaction physically, vocally and linguistically in accordance with one's own intentions.

An analysis of precisely those interactional competences can only partially be performed quantitatively and must thus fall back under the jurisdiction of qualitative conversational analysis. The quantitative distribution of subjunctives and modalizations, however, provides us with information about, for example, tendencies regarding the understanding of argumentation (convergent – divergent) in the age range studied, since the frequent use of subjunctives and modalizations in grades 4 and 6 reveals, among other things, the extent to which one's own views are marked as negotiable, other perspectives are taken into account and face-saving strategies are used.

If we now look at the subjunctive forms and the mitigating modal particles in a setting comparison, different effects become apparent. In the subjunctive forms, there is no difference in the positionings between the Robinson setting and the donation setting (grouped Sm and So), but in the donation setting without consequences for future actions (So) there are significantly more subjunctive forms (mean value: 3.0) than in the donation setting with consequences for future actions (Sm, mean value: 2.3) (Mann-Whitney U-Test, $U = 1366.5$, $p < 0.05$). However, since no significant differences are found in the mitigating modal particles in the setting comparison Sm vs. So and no tests are significant in the combined search queries (statements with both subjunctive forms and modal particles), the result for the subjunctives must be interpreted with caution. It must be noted that a task with a fictional problem also suggests negotiating certain decisions in the subjunctive mode, and in this respect the increased use of subjunctives in the So setting is not surprising. Without a comparable effect in the mitigating modal particles or the combined searches and without considering the context, we therefore cannot speak of a more intensive perspectivation and modalization in the donation setting without consequence of action.

In contrast to the subjunctives, there is a significant difference between the Robinson setting and the donation setting in the use of mitigating modal particles (see Figure 3):

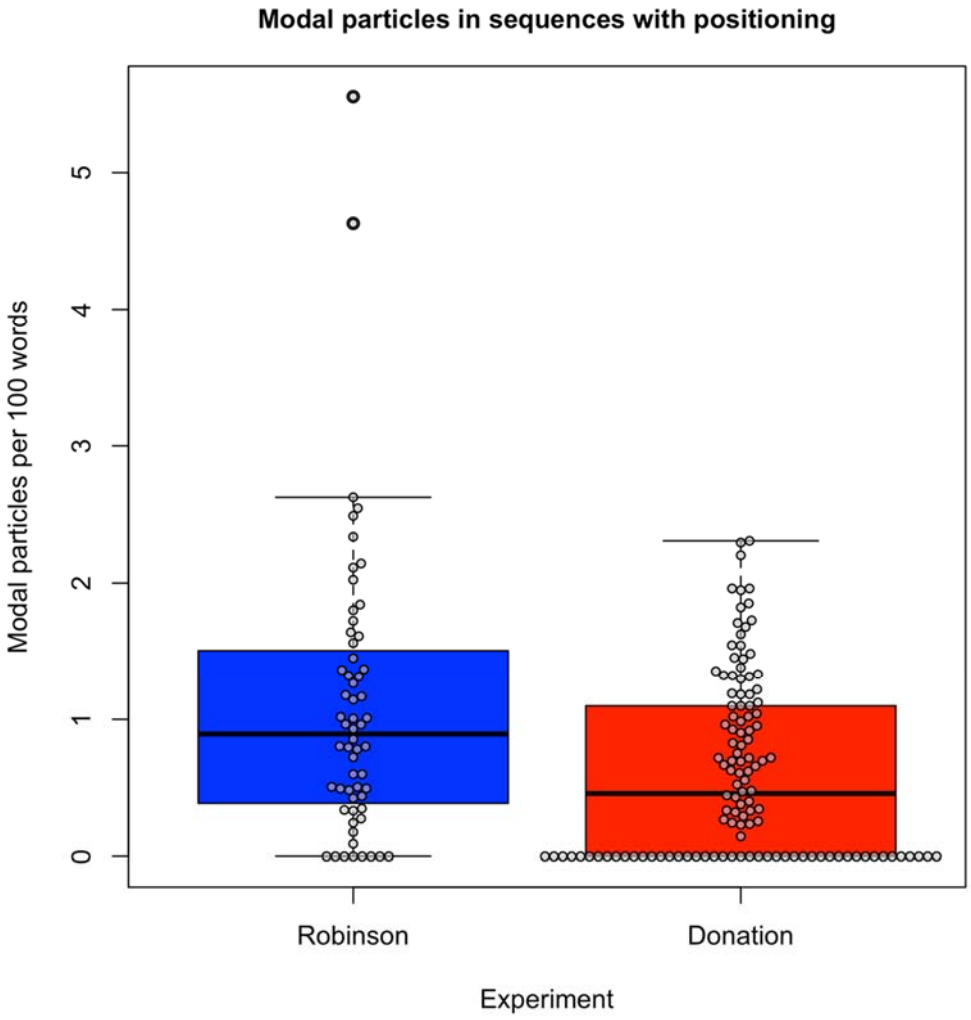


Figure 3: Modal particles in the Robinson setting and the donation setting

In this box plot, it can be seen that more use is made of mitigating modal particles in positionings in the Robinson setting than in the donation setting (Sm and So) (Mann-Whitney U-Test, $U = 4660$, $p < 0.01$). The line inside the box marks the median.[30] The whole box covers the middle 50% of the data, i.e. the data in the second and third quartiles (25-75%). The antennas outside the box terminate at the last data point within 1.5 times the interquartile distance (the length of the box) from the box. Values outside the antennas are regarded as outliers. The smaller the box, the closer the values of the middle 50% of the data to the median and the smaller the dispersion.

In addition, the values of the individual transcripts are represented as dots to show the distribution in detail. Here it is also clearly visible that considerably more conversations in the donation setting than in the Robinson setting contain no modal

[30] The median indicates the central value of a data set, in this case the word count of each of the 20 conversations. The median splits the dataset in two sets of equal size, whith values either not larger or not smaller than the median. In contrast to, for example, the arithmetic mean, the median is less sensitive to extreme values (and thus to outliers).

particles at all (Pearson's chi-squared test: $\chi 2 = 11.25$, $p < 0.001$, Cramer's $V = 0.25$).

Since this effect – more mitigating modal particles are used in positionings in the Robinson setting than in the donation setting – does not show up (contrary to our expectations) in the comparison Sm *versus* So, but only across all settings, it can be assumed with high plausibility that the type of task with the different number of objects to be discussed, which can potentially be brought into the conversation in a perspectivating way,[31] has an influence.

In the case of justifications, however, the setting has no influence whatsoever on the use in an utterance of either subjunctive forms or mitigating modal particles or the combination of the two. None of the tests turned out to be significant.

Finally, no influence of gender on the frequency of use of subjunctives or mitigating modal particles in positionings or justifications could be identified. This is illustrated in Figure 4.

Figure 4 shows the number of modal particles per hundred words in sequences with justifications, grouped by gender. Each data point represents the value of a female speaker (left) or male speaker (right). We can see a very similar distribution between girls and boys. In both groups, the average values are in a similar range (grey bars: female 0.72, male 0.71), and in both groups there are outliers toward the top. As well, the percentage of speakers who do not use modal particles hardly differs (see bar at value 0 in Figure 4).

The quantitative analyses show us whether and to what extent selected variables have an influence on language use. We see, for example, that children in grades 4 and 6 are more likely than those in grade 2 to use subjunctive forms and mitigating modal particles for the argumentation-relevant positionings, but also that there are no significant differences in this regard between children in grades 4 and 6. In terms of (factual) oral argumentation competence, this means that many children between grades 2 and 4 (i.e. between the ages of 8 and 10) learn to purposefully use linguistic means in combination to mark perspective modalizations. We also see very clearly that, in general, there are no gender-specific differences with regard to subjunctive and mitigating modal particles when considering the whole corpus. Accordingly, the calculations can also act as a corrective, since in our case the qualitative analyses of individual conversations, conducted first, gave us the impression that girls make more use of modal particles and subjunctives than boys. If one goes back to the qualitative analysis with the quantitative finding in mind that there are no differences between boys and girls, however, it becomes apparent that in groups consisting only of boys epistemic authority (Kirkham 2011) is often marked, whereas in groups consisting only of girls, arguments are often consensual. Our original impression in the qualitative analysis is refuted by the quantitative analyses, but in a further qualitative analysis based on the quantitative analyses, an effect of a homogeneous group composition in terms of gender becomes apparent.

[31] There are more items to discuss in the Robinson setting.

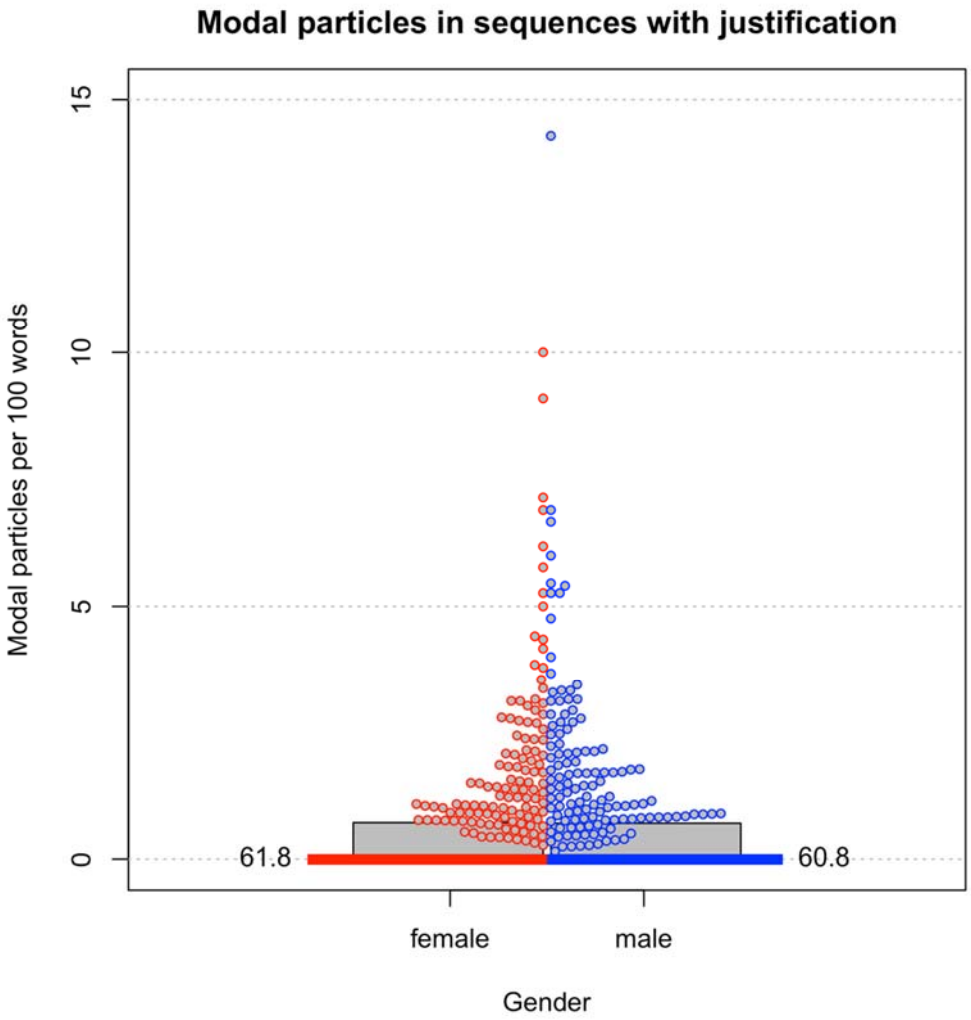


Figure 4: Mitigating modal particles per hundred words in sequences with justifications by girls and boys. Each data point represents a speaker; the grey bars represent the groups' mean values; the colored bars at the bottom represent the percentage of speakers who did not use modal particles.

This again shows that the calculations in this case were carried out detached from the contextual embedding. Since modal particles can be contextually multifunctional, we must assume that a qualitative, case-based analysis can easily lead to different results. For example, in a conversation in grade 4 we find the utterance: "ich würd villcht uf jede fall mal e SCHLAFsack;" ('maybe I would definitely a sleeping bag'; Ro_K4_HZ_G3b, 02:38). In this positioning Hanna combines the subjunctive *would*, the mitigating modal particle *perhaps* and the reinforcing modal particle *in any case* (she uses the same pattern repeatedly in her further suggestions). However, since we have not taken the latter into account in the quantitative calculations, this example is one of the perspectivized, modalized expressions. But if one looks into the conversation, the use seems to be contradictory and the categorization no longer so clear. Accordingly, it is important to contextualize the results on the basis of qualitative analyses, as in the above example of Hanna, in order to understand them more comprehensively.

4.1.2. Complexity of Argumentation

Our qualitative analyses so far show that in grade 2, topics are rarely discussed in depth over several turns. Rather, we encounter many argumentatively isolated statements. From grade 4 onwards, the children increasingly participate in longer jointly produced argumentative units. It is also noticeable that exploratory (as opposed to persuasive) argumentation (see Ehlich 2014; Nonnon 1996) increases in importance, and that children more often co-construct justifications or even longer arguments (see Jacoby/Ochs 1995). We illustrate this with an example from our qualitative analysis.

The following example shows an argumentative sequence of four girls who from the beginning of the discussion produce justifications and – in perfect agreement – continue to support each other subsequently. We are dealing with a joint reasoning here, which is also relevant for relational work and identity construction:

Transcript 3: Ro_K6b_PB_G2b_F29 (Tanja, Leonie, Juliana, Rana)

```
01 JUL: viellEicht das SACKmesser;
        maybe the pocket knife
        ((points at the pocket knife))
02      weisst dU (.) [zum FRÜCHte ]
        you know for fruit
        ((imitates cutting with a pocket knife))
03 RAN:               [dAs SACKmesser das:-]
                       the pocket knife it
04 TAN: doch sAckmässer isch scho t find ich wichtig wäil das het
        VIIL.
        well yes I reckon the pocket knife is important because it
        has many
05 JUL: ja Eben     [(.) für] ALles (komm),
        yes exactly for everything (come)
06 LEO:             [JÄ,    ]
                     yeah
07 RAN: ((nickt)) und vIelleicht hats auch SO:- (.)
        ((nods)) and maybe there also are such
08      aso jetz sag ich mal FLAschenöffner,
        I'll just call it bottle opener for now
09      (weil) es hat so auch viele ANdere sachen;
        (because) it has so many other things too
        ((imitates a screwing movement with her hands))
10 TAN: JA;
        yes
11 JUL: JA;
        yes
12 LEO: und BUSCHmesser [kannst du Eins ] (.) ja;
        and bush knife you can one well
13 TAN:                 [also Ich würd SACKmässer; ]
                         I would take pocket knife
14 LEO: Eigentlich MACHen;
        actually make one
15 JUL: jA denn kannst ja nur für DA:S;
        yes because you can only for this
        ((swings an imaginary bush knife through the air))
16 LEO: oder du kAnnst eins machen mit HOLZ,
        or you can make one out of wood
```

```
17 RAN: JA;
        yes
18 JUL: Eben;
        that's what I was saying
```

The sequence is opened by Julia, who directs the thematic focus to a new topic with her suggestion "maybe the pocket knife" (01). She herself adds a justification ("you know for fruit," 02), which she also illustrates gesturally. Here, the communicative practice of arguing – marked as explanation – is established and also taken over by a second speaker, in that she does not just limit her statement to an agreement ("well yes," 04) and her own statement ("I reckon the pocket knife is important," 04), but adds a linguistically explicitly marked justification ("because it has many," 04). With the co-constructed justification for Julia's proposal, the two girls establish a consensus.

Nevertheless, both thematic fields (bush knife and pocket knife) are taken up by the other children in the following and are developed further. Thus, Julia supports Tania's argument, first with a consent particle and then with an expanding support ("yes exactly for everything," 05).

At first, both arguments are initially rather abstract, with the use of the undefined numerical words "many" and "everything," but they are made concrete in one of the next turns: a fourth speaker, Rana, provides further support by naming two examples in lines 08 and 09 and illustrating them gesturally ("I'll just call it bottle opener for now," 08 and "because it has so many other things too," 09). At this point, this argumentative discourse unit would be completed, since consensus is signalled by all children. Nevertheless, the girls continue searching for arguments. These refer to a different but similar object (bush knife) and are used to further plausibilize the actual focus object (pocket knife) and examine new possible solutions together ("and bush knife, you can, well, actually make one," 12, 14, 16).

What is striking about this sequence is that there are no genre-typical pro-contra sequences, but rather an exploratory argumentative development of themes by the four children over several turns. This sequence is characterized by the addition of several validating justifications for the initial proposition: the children explicate their agreement and strengthen their consensus by expanding the arguments of their interlocutors.

The children in the second grade, in contrast, produce their positions (in the sense of an assertion) but do not provide any further justifications or even mutual support for them, as this example from the initial sequence of a conversation demonstrates:

Transcript 4 Ro_K2_SA_G2a (Yara, Selvanila, Jonathan, Finn)

```
01 SEL: °h Oder (.) (de:nn) [dIese] BUSCHmesser,
        or (in this case) this bush knife
02 FIN:                     [hm-  ]
                             hm
03 SEL: (.) das kann AU: (.) vilI::[ch(t),]
        it can as well maybe
04 FIN:                           [busch,]
                                   bush
05 YAR: jä °h strEichhÖ:lz (.) Oder f::FEUerzeug;
        yeah matches or lighter
```

```
06 SEL: °h [Und noch] ein !ZELT! brAuch(en) wir;
        and also a tent we need one
07 JON:    [aso ICH,]
            well I
08      jä (.) Und [s_m ]osKItonetz;
        yeah and a mosquito net
09 YAR:            [j:ä-]
                    yeah
10 SEL: j:ä;
        yeah
11 YAR: jä;
        yeah
12 JON: aso wÄmmer (mol) ZÄLT und moskItonetz nee;
        so should we take the tent and the mosquito net for now
13 FIN: !U:ND!?
        and
14 YAR:         Und [no: (.) FEUer]zEug;
                and also the lighter
15 JON: <<rufend>   [WOLLdE:>cke; ]
        wool blanket
16 SEL: °h oder [noch (.) (und) verBANDzeug;  ]
        or also (and) the bandages
17 FIN:         [NÄ:I, DAS brUUch mer aber Au,]
                 no, we also need
18      ver[BANDzeug,]
        bandages
19 SEL:    [verBAND  ]zeug;
            bandages
```

Selvanila suggests the bush knife (01). This suggestion is made without justification and is not taken up or argumentatively discussed by the other children; instead, further suggestions are strung together (04-06) or the quite general and unspecific justification is given that an object is "needed" (06, 17). In the examples of the younger children, consensus is often achieved by a simple combination of proposals (Lindström/Sorjonen 2013), in which an object once proposed is either repeated or simply confirmed by the others. Justifications are usually given only after hardened dissent and especially after why-questions, which explicitly demand a justification (Kreuz, in press).

Overall, it is noticeable in the data of the second grade that different perspectives are rarely hierarchized. In addition, there are only a few isolated turns that involve weighing justifications, naming oppositional and validating justifications with examples and merging justifications. Therefore, the argumentative complexity (Klein 1980; Grundler 2011, 2015) of the discussions among second graders is relatively low, so that one can speak of a "shallow" argumentative debate (see also Kreuz/ Luginbühl 2020).

Based on these observations (see also Mundwiler/Kreuz 2018; Hauser/Kreuz 2018), it was our goal to use quantitative instruments to verify these tendencies regarding the breadth and depth of argumentation across the entire corpus. Following Grundler (2011:177), we understand a broad argumentation as an argumentation in which many different arguments are realized to support a position. In an in-depth argumentation, single arguments are differentiated in regard to their support

(by naming rules, examples and so on). Our assumption was that the jointly produced argumentations by the older children are broader, deeper and increasingly co-constructed, that is, produced with the participation of several children.

A first, quantitative, approach to answering the question concerning the complexity of argumentation was counting the number of justifications (see Figure 5):

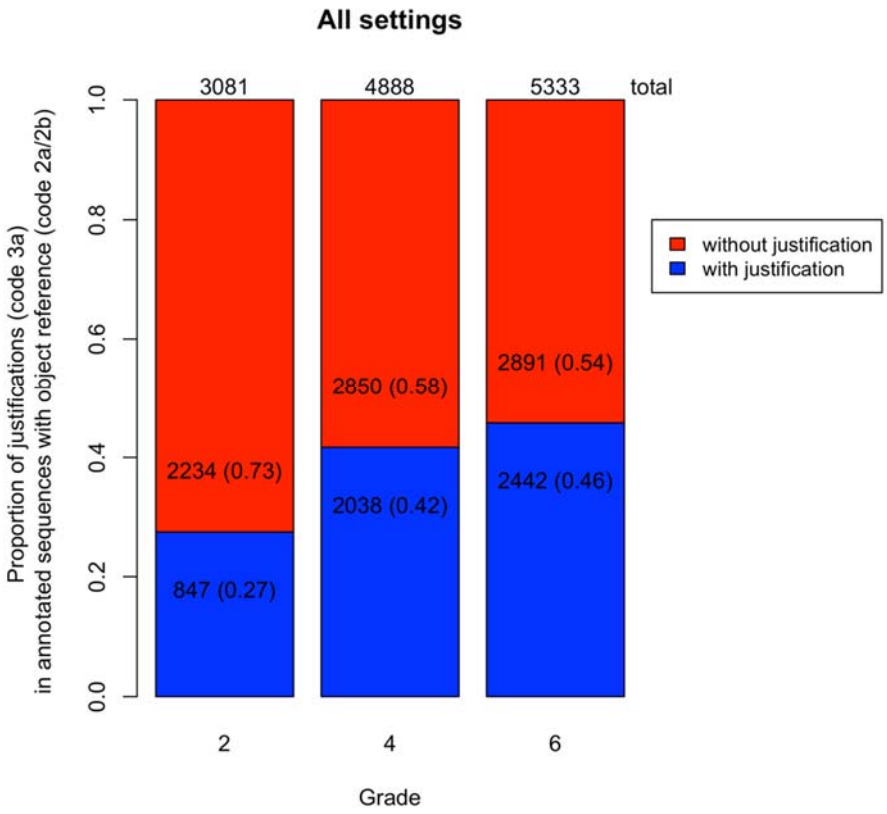


Figure 5: Proportion of justifications in annotated sequences with object reference, grouped by grade and across all settings

In Figure 5, the values within the bars indicate, for each grade, the absolute number of sequences with justification (blue) and without justification (red), and in parentheses the proportion of each. The values above the bars indicate the total number of topic-related actions (sequences with object reference) in the respective grade. Calculated over the entire corpus, the share of justifications in the total of topic-related turns increases significantly from grades 2 to 4 (27% vs. 42%) and only slightly from grades 4 to 6 (42% vs. 46%). This development, which begins in grade 4, is also evident in relation to some of the other characteristics tested (see, for example, modalizations/use of subjunctives, Figures 2-a and 2-b; size and frequency of thematic clusters, Figures 9 and 10). In a further step, the justifications were evaluated separately according to their embedding in the interaction. In all settings, justifications are mostly employed in oppositional contexts (see Figure 6):

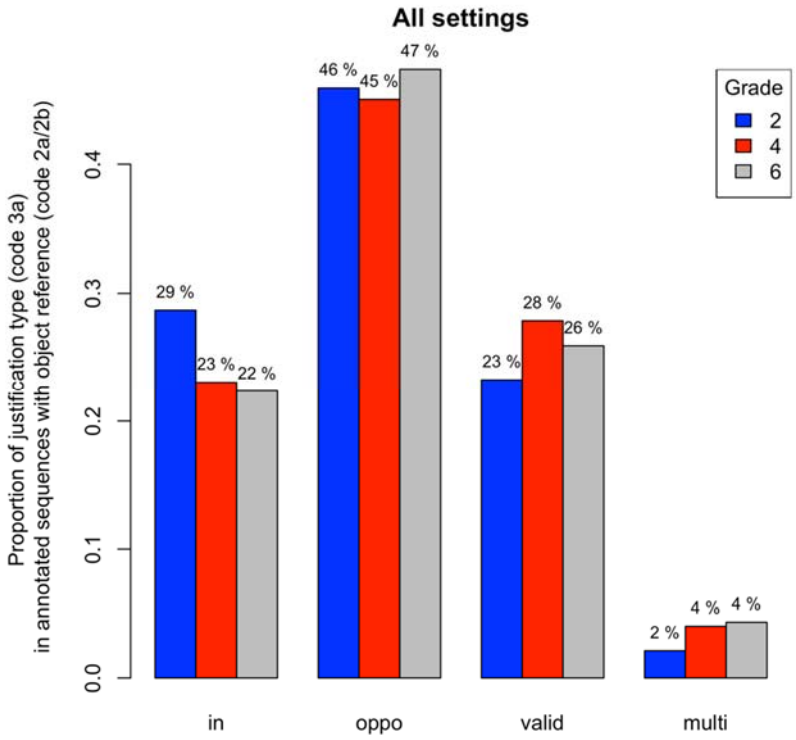


Figure 6: Interactional embedding of justifications (whole corpus)

Here it becomes apparent that 45-47% of all justifications are oppositional ("oppo"), i.e. that an objection or differing claims of validity are processed. Across the three grades, 23-28% of all justifications are validating ("valid"), with an increase from grades 2 to 4 (and 6). There is also a slight increase in the number of justifications that have multiple orientations/considerations ("multi", see above, Section 3.2). The initiating justifications ("in") decrease with age, as is particularly clear with the Robinson corpus (see below).

In these calculations across all settings, the differences within the validating and oppositional justifications are not very large. The situation changes, however, when the settings are analyzed separately. It is only in the Robinson setting that a clear picture emerges. In the case of the donation settings, the trends are not quite as conclusive. The differences in the individual settings are sometimes less pronounced and therefore less conclusive. Also, in the comparison of the two donation settings (with and without action consequence), none of the differences proved significant (possibly due to longer 'outlier' discussions). Since the most distinct and clearest picture shows up with the Robinson setting, this result is examined in the following (see Figure 7):

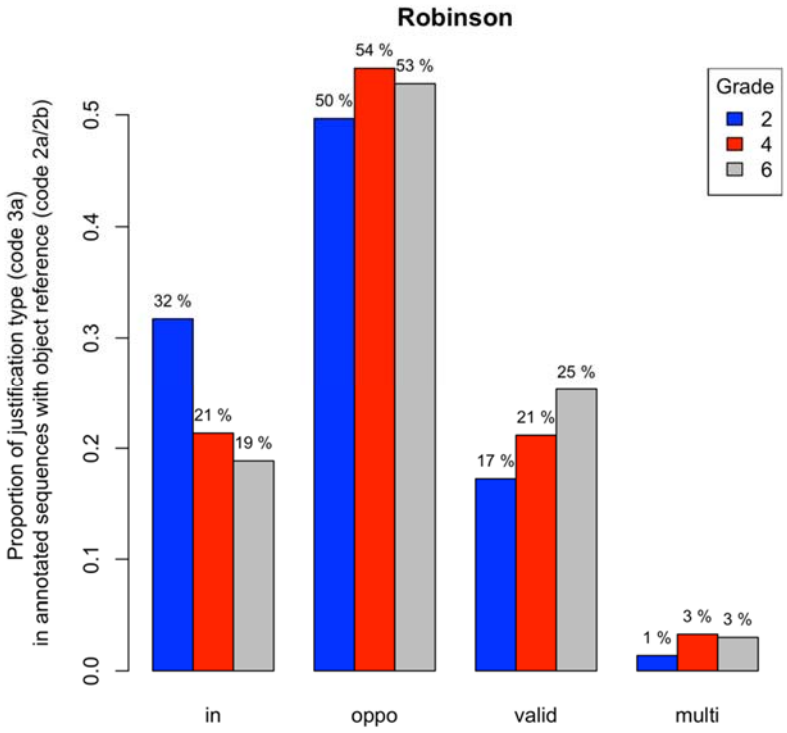


Figure 7: Interactive embedding of justifications (sub-corpus Robinson)

While – as is already clear in the overall illustration (figure 6) – oppositional justifications make up about half of all justifications with 50-54%, clear developments can now also be seen in the validating and initiating justifications: the validating justifications increase by 4% for each grade. And the initiating justifications are much more strongly represented in grade 2 (32%) than in grades 4 and 6 (21% and 19%, respectively). This confirms the conclusion from the qualitative analysis (see Transcript Example 4) that younger children often offer individual justifications that do not refer to a position formulated immediately before by others (almost one-third initiating justifications) and that, as the grade increases, arguments are more often advanced consensually (increase in the number of validating justifications).

All in all, the discussions from the Robinson setting are the most productive data with respect to the proportions of argumentative sequences, have also significantly influenced the development of the codes for the qualitative analysis. At the same time, this task requires more topics to be dealt with (12 items in the Robinson setting vs. four projects in the donation setting), which may have an impact on the distribution of oppositions, validations and isolated/initiating justifications. What we can deduce from the different calculations is that the type of task has an influence on the argumentative approach, and the results are therefore only partly comparable with other studies and settings. A direct comparison is only possible without difficulty if the results are consistent across different settings or if the specific influence of a setting can be plausibly identified.

Regarding the complexity of the argumentations, the extent to which the different topics were negotiated with or without justification was also evaluated. Accordingly, we can see in Figure 8 how the situation changes across the grades:

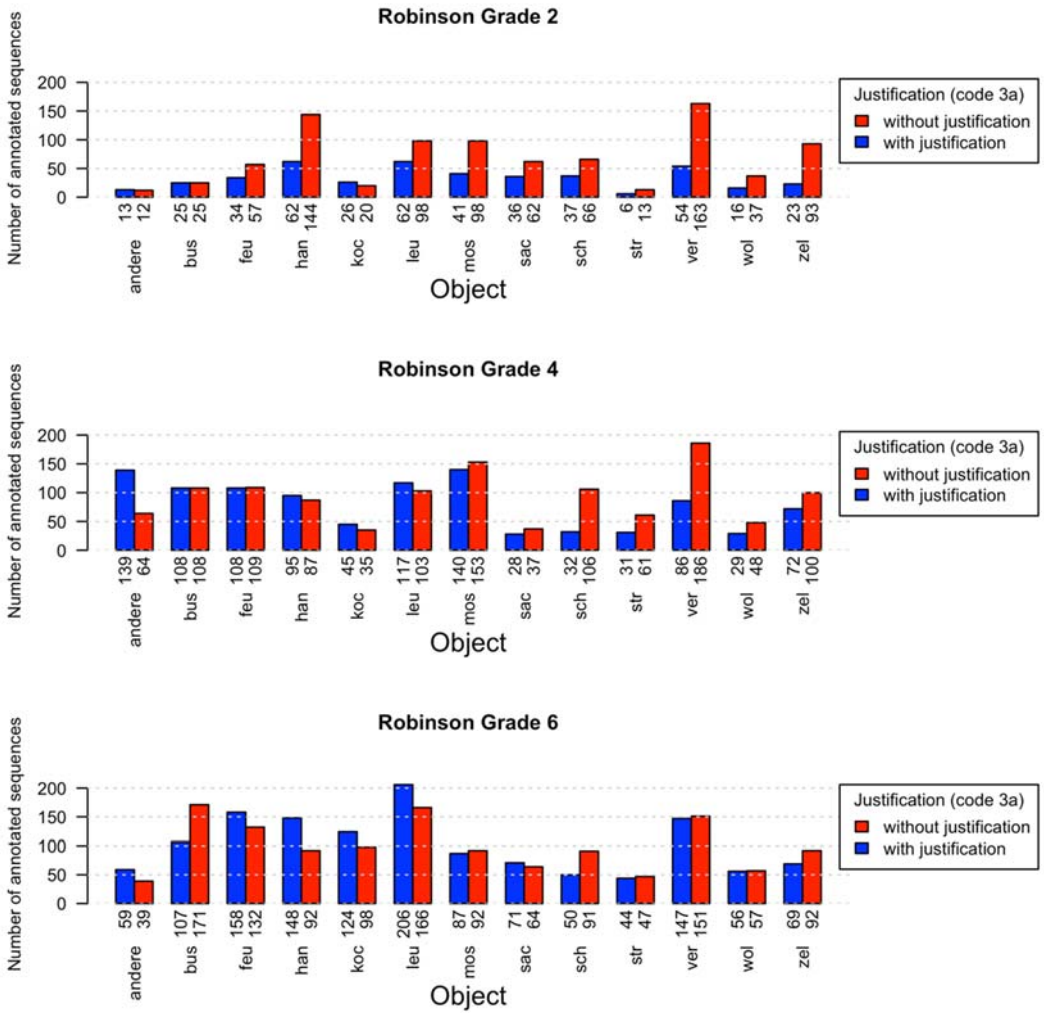


Figure 8: Number of justified (left, blue) and unjustified (right, red) sequences per topic (Robinson setting)[32]

In a separate graphic for each grade (2, 4, 6), Figure 8 shows how many sequences per topic/object contain a justification (blue). The values below the bars indicate the absolute number of sequences for the respective topic/object. On the one hand, it can be seen that the ratios balance out or reverse across the grades (see the ratios between red and blue bars, where the blue bars represent justified utterances), and that the proportions of justified sequences increase for most topics (see the increase in blue bars). On the other hand, one also gets a more differentiated view on controversial topics. This can be used in particular for didactic purposes when it comes to designing appropriate exercise formats and topics.

In order to better determine whether and when children are engaged in more in-depth discussions, in regard to both individual conversations and age groups, we

[32] The abbreviations for the objects are listed in Section 3.2.

coded whether a statement deals with a topic (object or project) that is also the focus of the previous statement. This allows *thematic clusters* to be determined and their size (number of consecutive statements on the same topic) to be displayed and compared. The cluster size was determined on the basis of the number of coded utterances, i.e. cluster size 1 corresponds to a single topic-related turn, while cluster size 2 (or more) corresponds to two (or more) successive coded turns on the same topic. Therefore, clusters from size 2 on are particularly interesting. This enables us to determine the frequency of the number of topic-related turns a topic is dealt with in a particular conversation. In the following figure, we see the development of the average cluster size using the example of the Robinson data:[33]

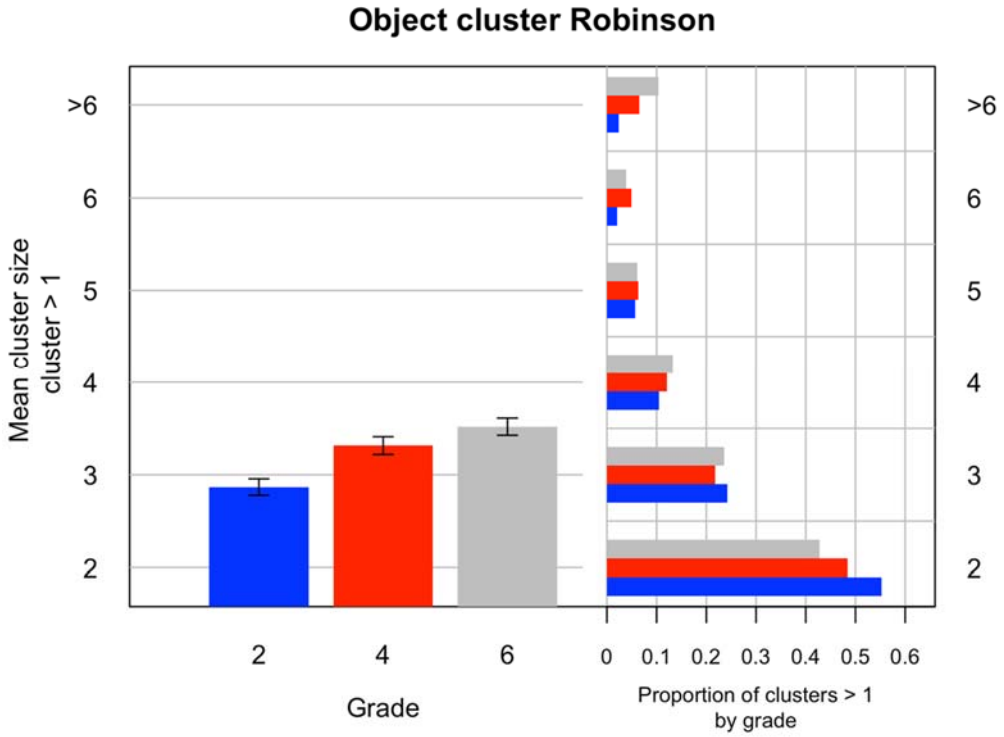


Figure 9: Average size of thematic clusters, grouped by grade

For the calculation of thematic clusters in Figure 9, all topic-related turns that refer to the same object as the previous sequence were included, but not sequences that refer to a different object than the previous utterance (only cluster > 1). The error bars represent the standard error. The right part of the figure shows, for each cluster size, the proportion of its corresponding cluster size relative to the total number of clusters of the respective grade.

In the Robinson data, the average size of the thematic clusters increases steadily from grades 2 to 4 to 6; that is, in each higher grade the individual topics are discussed more coherently and more extensively within the conversation. With the calculation of the *frequency* of thematic clusters, we notice a steady increase in the Robinson setting as well:

[33] The calculations for the donation setting did not result in clear trends or patterns.

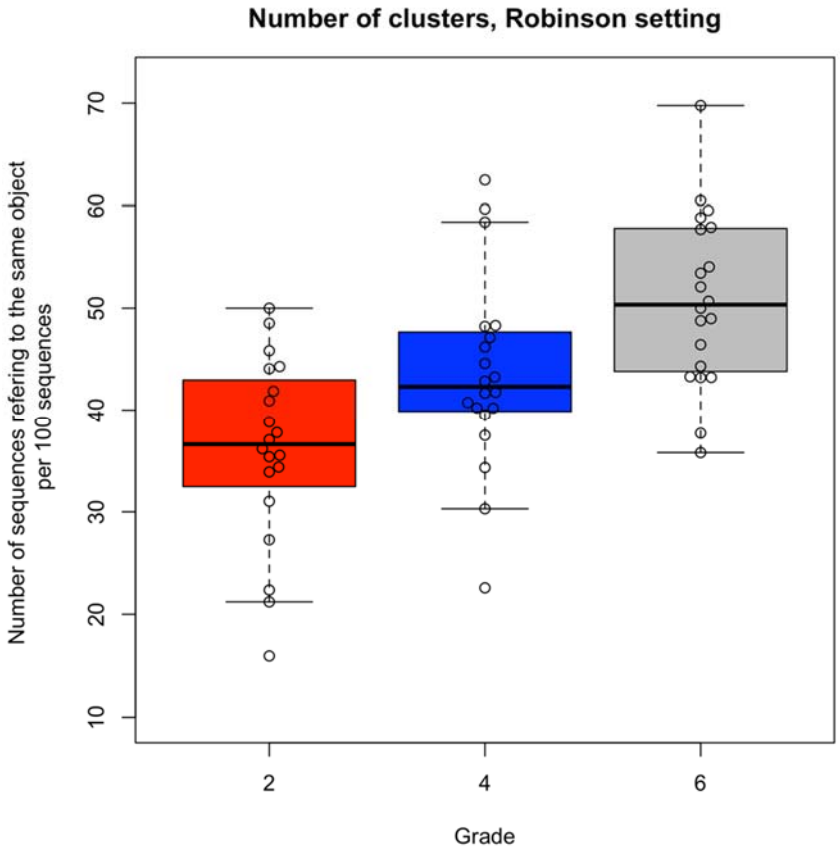


Figure 10: Frequency of thematic clusters in the Robinson setting, grouped by grade

In the boxplots (see also the explanation of boxplots in Figure 3 above) in Figure 10, the frequency of object clusters is shown, i.e. the percentage of sequences in which a topic is continued. Here, too, the values of the individual conversations are additionally displayed as dots. In grade 2, a topic is continued in 35% of the annotated actions, in grade 4 in 46% and grade 6 in 50%. A one-way analysis of variance shows a significant difference between the grades (ANOVA: $F(2) = 13.34$, $p < 0.001$[34]). The pairwise comparisons (pairwise T-tests, corrected according to Holm) are all significant (grades 2 vs. 4: $p < 0.05$; grades 2 vs. 6: $p < 0.001$; grades 4 vs. 6: $p < 0.05$). It thus becomes apparent that the school children in the Robinson setting discuss a topic in longer sequences more frequently with increasing age. In Section 4.2, we examine thematic clusters in individual discussions and visualize them for two particular cases (see Figures 15 and 16).

Our analysis also shows that the proportion of validating justifications increases, i.e. that (partially) consensus-based complex discussions also become increasingly common (see above, Figure 7). Of course, there are other aspects central to the complexity of argumentation that we cannot quantify, such as the extent to which already formulated justifications are repeated and the complexity of the individual justifications. This requires qualitative analysis, which can now be carried out in a more focused manner against the background of the quantitative results.

[34] Since the values are normally distributed within the groups, a one-way analysis of variance (ANOVA) could be applied, as well as T-tests for the pairwise comparisons.

4.1.3. Interactivity

Didactic discussions often emphasize the importance and necessity of creating practice opportunities in the area of conversational argumentation that allow a connection to the reality of the students and ideally have real consequences for future actions ("real world didactic," Wagner 2003:749, quoted in Becker-Mrotzek 2008:71, our translation). In order to test this requirement empirically, we collected data for the two sub-corpora 'donation setting with consequence of action' (Sm) and 'donation setting without consequence of action' (So) and compared the data with respect to different criteria. Among other things, we investigated the question of interactivity, based on the assumption that the real setting is more motivating and generates more involvement among the individual speakers (see Grundler 2011:83; Spiegel 2006:39f.). In order to test a possible indicator, the frequency of turn-taking[35] was determined and compared. Of course, when assessing involvement as a possible indicator of motivation, other aspects that have been quantitatively investigated also play a role, such as individual speaking time (see the analysis of individual conversations in Section 4.2 as an example), the complexity of the argumentation (see Section 4.1.2) and the quantity and sequencing of justifications (consensual or adversarial, whereby in the latter case the pros and cons are weighed up, see Figures 7 and 8). In addition, there are indicators of involvement that can hardly be meaningfully analyzed quantitatively, ranging from non-verbal to intonational to verbal means. In the following, we therefore only analyze a single indicator by way of example.

A mixed effects logistic regression model was used to calculate whether the occurrence of turn-taking is influenced by the factors grade and setting and their interaction. For the dependent variable it was coded in each conversation, whether an utterance was made by a different speaker than the previous utterance. On one hand, the main effects of grade and setting were analyzed. A significant main effect of the grade means that the probability of turn-taking differs significantly in the different grades, i.e. membership in a grade enables a prediction of the probability of a turn-taking (= significant predictor). An interaction would, for example, mean that the influence of the grade differs depending on which setting is processed. In mixed models, random variables can be considered in addition to predictors or *fixed factors*. This means that the differences within the groups resulting from the random selection of the sample are included in the calculation. In our case, conversation (transcript) and speaker were included as *random intercepts*. This enables a better adjustment of the model and ensures that effects resulting from differences within a group do not appear erroneously as the influence of predictors, or *fixed factors*. However, only the grade was found to be a significant predictor (significant main effect, $p < 0.001$): the probability of turn-taking is higher in grade 6 than in grades 2 or 4 (see Figure 11). The consequences for future actions and the interaction of the two factors (consequences for future actions and grade) are not significant. Figure 11 shows the estimated probabilities of turn-taking, grouped by grade and setting (donation with or without consequence of action):

[35] A similar approach is found in a study by Lucas M. Bietti, which he presented in "Interacting to Remember: Coordination and Distribution in Memory Collaboration" at the GAL Research School, Albert-Ludwigs-University Freiburg, September 5, 2017. In that study, he determined the degree of interactivity as "distribution of speakers' turns" (PP slide 23).

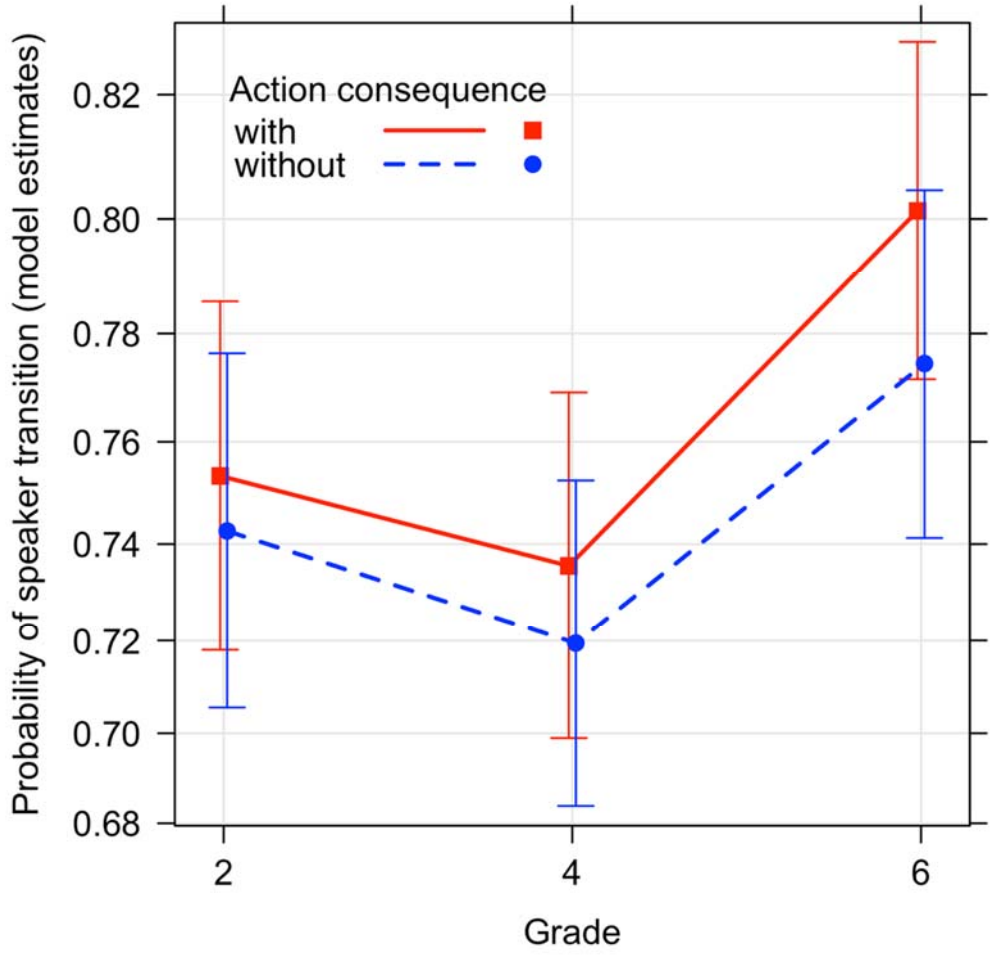


Figure 11: Estimated probability (mixed effects logistic regression model) of turn-taking depending on grade and setting (donation with or without consequence of action). The estimated values of the mixed logistic model with interaction of the variables grade and setting are shown; conversation and speaker were specified as *random intercepts*; the error bars represent confidence intervals (95%).

The graph shows that in grade 6 there is a tendency for turn-taking to occur more frequently when a relevance of action is present. However, the difference is not significant. It can also be seen that the values in the donation setting with consequences for future actions (Sm) in all three grades are above the values of the fictitious comparison setting (So), but the differences are also not significant. The expectation that consequence of action results in increased involvement cannot be confirmed by this analysis.

All the methods shown so far have in common that they allow statements about both the entire corpus and sub-corpora to be made. Especially with large data sets, it is often challenging or impossible to take an external view based on qualitative microanalyses and recognize global structures or even calculate correlations and effects, which only become visible in controlled comparison settings. Here, the quantitative methods make it possible to obtain an overview of the data based on either pure language material (where the transcripts, i.e. oral 'texts,' were searched)

or previously set codes. In our case, not only were the impressions from the qualitative studies confirmed, but unexpected findings could also be included in the overall analysis. For example, our expectation that justifications are increasingly modalized and perspectivized was only partially confirmed. Gender-related differences in the area of modalization and perspectivation were also hardly found. Similarly, only one setting (Robinson) shows that the proportion of validating or initiating justifications develops steadily in a certain direction. In line with our expectations, however, a relative increase in justifications can be observed across all data.

4.2. Visualization and Analysis of Individual Conversations

In the course of the analysis, we worked with visualizations[36] in various ways. The visualizations not only illustrate summarizing results, but also show individual differences particularly well and can stimulate further analysis. We have, for example, visualized the course of individual conversations with regard to selected aspects in color. Thus, on the one hand, simpler aspects such as pure share of contributions could be presented and evaluated, and, on the other, selected codes could be focused on, including thematic references or argumentative actions such as justifications. The visualizations therefore rely on quantitatively generated information and map frequencies, patterns and processes, but also individual differences. Although these representations may already allow analysis and interpretations, they are, in our opinion, particularly useful if they are linked to further qualitative analyses.

We subsequently focus on selected analyses of the share of argumentative and thematic contributions as well as thematic clusters.

4.2.1. Distribution of Argumentative and Thematic Contributions

We produced various visualizations for all 180 discussions. The following two figures show visualizations of the distribution of contributions in two conversations:

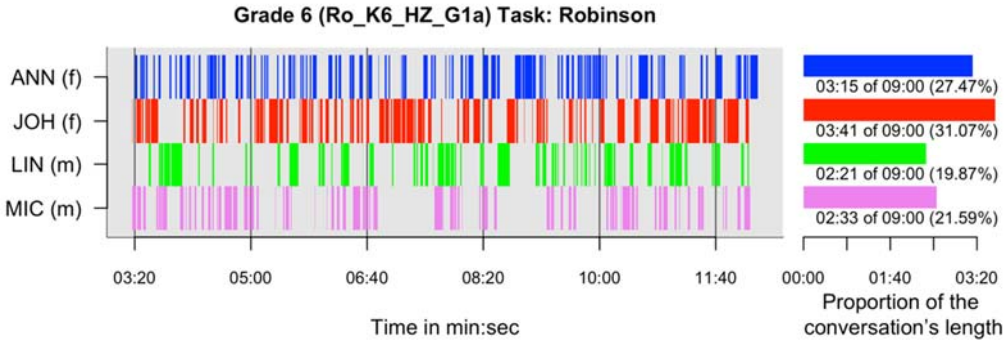


Figure 12: Distribution of contributions in a grade 6 group (Ro_K6_HZ_G1a: Linus, Anna-Lena, Johanna, Michelle)

[36] Various visualizations from the project are available at https://argcomp.shinyapps.io/baselargVis/

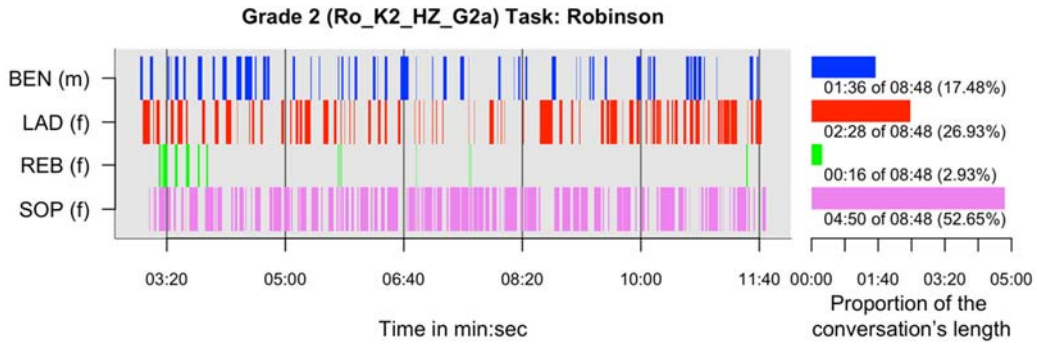


Figure 13: Distribution of contributions in a grade 2 group
(Ro_K2_HZ_G2a: Benjamin, Rebecca, Sophie, Ladina)

In the first visualization (Figure 12), we see a quite balanced speaker participation among the four students: all of them contribute about 20-30% of the total speaking time.[37] At the same time, however, we also see that some students are involved very dominantly at times (for example, JOH around the seventh minute), while others hold back (in this case MIC).

As a comparison, we see in Figure 13 that such a balanced participation among all four students is by no means given. Rather, we also find conversations in which, for example, one person participates much more actively than others and is also strongly involved in all phases (here SOP dominates with approx. 53%). As well, the conversation takes place mainly between two or three children (here, apart from SOP, the children BEN and LAD participate with approx. 17% and 27%, respectively). The visualization also allows us to see the partial or complete absence of (verbal) participation (here REB speaks only about 3% of the total speaking time). While conversation-analytical sequence analysis often focuses on smaller phenomena and contexts, the visualizations shown allow an overview of the entire conversation. They also make it possible to quickly identify passages (or entire conversations) in which a qualitative analysis seems particularly worthwhile.

What we have not yet captured in these visualizations are the linguistic actions with which the children participate, whether and when the speaker contributions are topic-focused and whether and when there are also phases of side sequences. In the following illustrations (Figure 14), therefore, two visualizations are superimposed on each other so that two things can be seen in color – which topics are discussed, and when justifications occur. This is the same conversation from grade 6 as above (see Figure 12):

[37] In this case, the pure speaking time was measured, i.e. only what was verbally expressed. From the point of view of CA, we are interested in the children's comprehensive, multimodal participation, but due to the extensive corpus multimodal aspects could not be transcribed to the same extent in a standardized manner and therefore could not be quantified meaningfully.

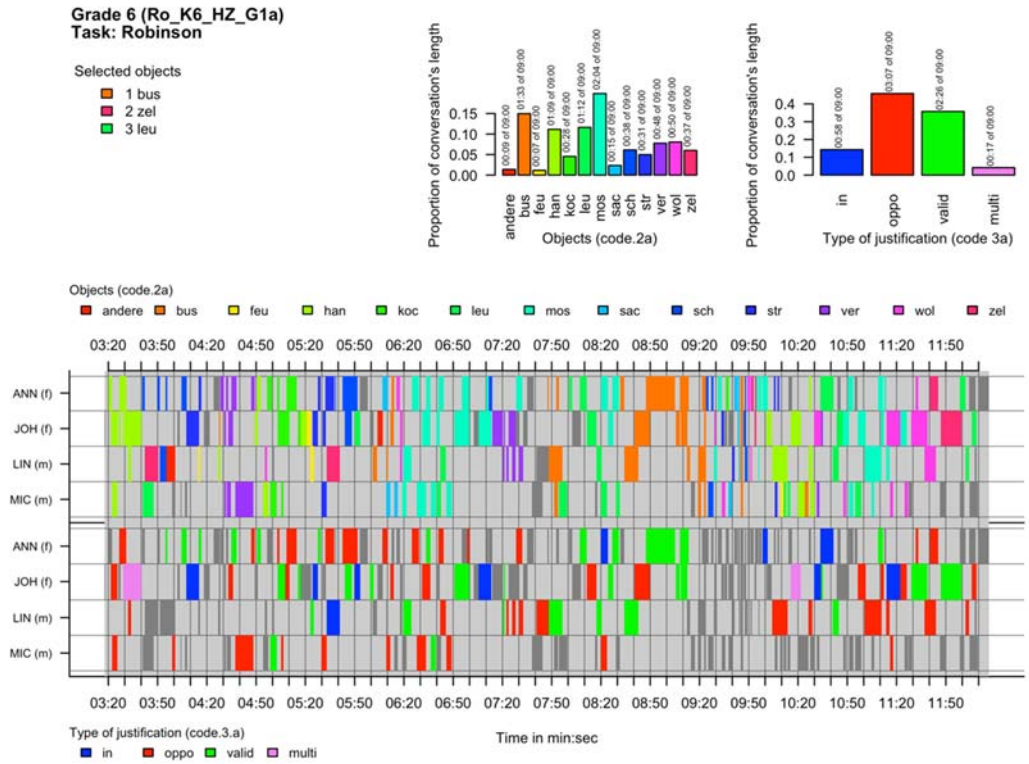


Figure 14: Conversation of a grade 6 group (Ro_K6_HZ_G1a: Linus, Anna-Lena, Johanna, Michelle). The objects selected at the end of the conversation, the proportion of topic references and the proportion of the duration of the different types of justification relative to the duration of the conversation as a whole are shown above. At the bottom are shown the topic reference (upper four rows) and justifications (lower four rows) over time (duration of the conversation in minutes, indicated on the lower axis), separated according to the four speakers (on the left axis).

In the upper left corner of Figure 14, the objects in the Robinson task the group agreed on at the end of the conversation are listed (in the example: bush knife, tent, flare). At the top in the center and on the right, the summarized speech components are listed by topic (center, "objects (code.2a)") and type of justification (right, "type of justification (code.3a)"). This information corresponds to the visualizations of the course of the conversation in the lower two-thirds of the figure. The upper course of the conversation is color-coded to show who speaks when about which topic. For example, clusters can be looked at: three longer passages concerning the mosquito net (which is not one of the selected objects) can be seen (turquoise). In contrast, the bush knife (orange), which was chosen, was only negotiated once for a longer period, apparently resulting in a consensus. In the lower visualization of the course of the conversation, whether and what kind of justification was realized is color-coded. The dark grey areas are not justifications, but coded level 1 (triggering actions) or level 4 (further observations, including meta-discursive actions) actions.

What is noticeable when comparing the two visualizations is that MIC takes part in the discussion with relatively few justifications, despite the balanced proportion of speakers; i.e. she makes mostly other contributions, such as taking up positions or taking on discussion-leading tasks. Furthermore, it is evident that JOH interacts by justifying in various ways, namely sequentially embedded in such a way that she reasons in an opposing and validating manner, initiates topics and relates to various

aspects and justifications. Such person-related observations can be used to create speaker profiles with the help of further sequence analyses.

In addition, the visualizations can serve to identify longer controversial or consensual justification sequences. For example, Figure 14 shows that the bush knife (orange in the upper visualization) is not discussed very controversially but is mostly strengthened by validating justifications (corresponding area in the lower visualization in light green for validating justifications).

Overall, this type of access to single conversations makes it possible to obtain information about individual differences and become aware of potentially interesting and analysis-relevant sections. With the visualizations presented here, thematic clusters, accumulations of (validating) justifications, speaker profiles, conversation profiles and so on can be recognized and further investigated. To comprehensively understand the identified phenomena or sections, it makes sense to then resort to qualitative methods and original video data in order to locate and explain the concrete characteristics in context.

4.2.2. Thematic Clusters

Thematic clusters, which were used above to summarize results (see size and frequency of thematic clusters, Figures 9 and 10), can also be created for individual conversations in their sequentiality. In the following, this is illustrated with visualizations of the two already discussed discussions from grades 2 and 6. The blue circles indicate that different speakers are involved in the cluster, while the red markings indicate unchanged speakers. In Figure 15, we see that in this conversation, for example, at the beginning (time in seconds on the bottom axis) a cluster of seven consecutive utterances (number of utterances on the left axis) is realized by several speakers, followed by thematically isolated single utterances and a cluster of three utterances, again realized by different speakers, and so on:

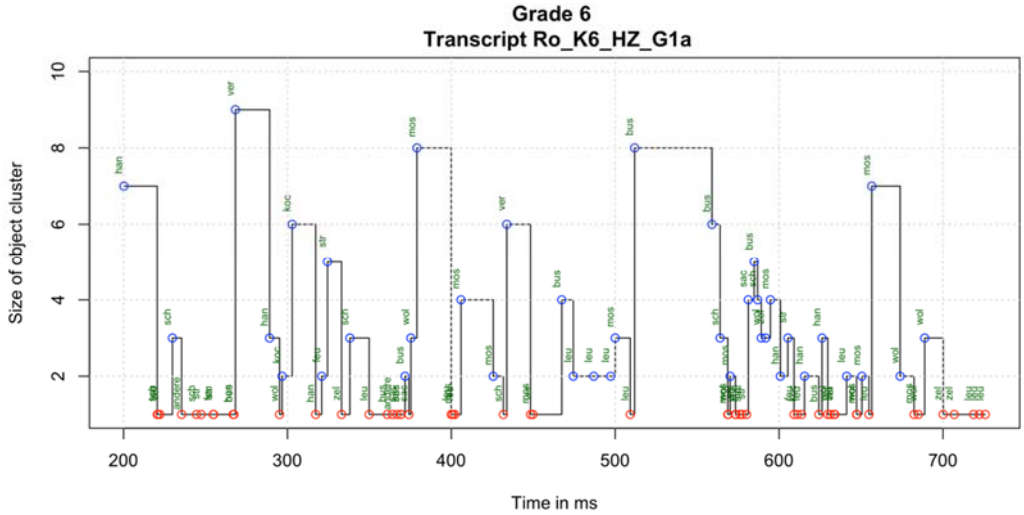


Figure 15: Thematic clusters in a grade 6 group
(Ro_K6_HZ_G1a: Linus, Anna-Lena, Johanna, Michelle)

The conversation of the grade 2 group presents itself differently (Figure 16):

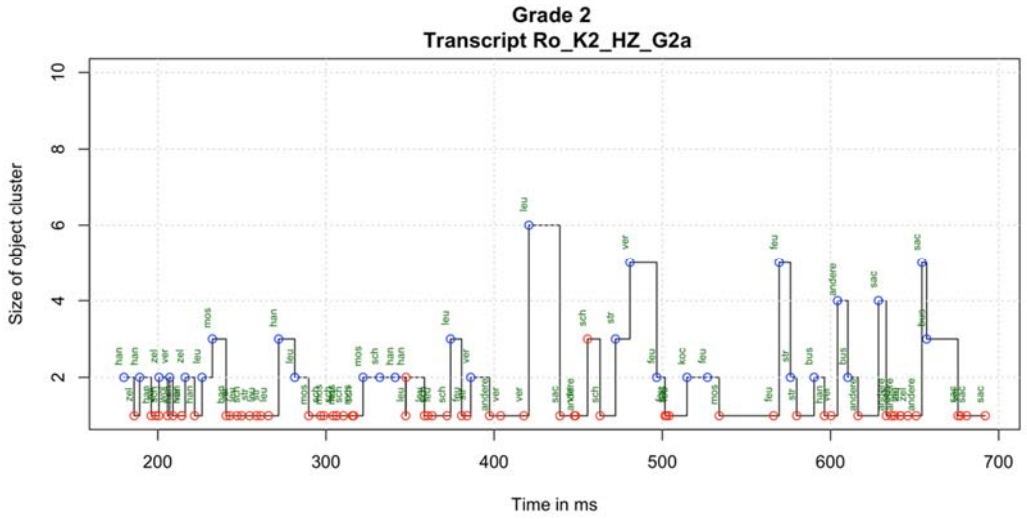


Figure 16: Thematic clusters in a grade 2 group
(Ro_K2_HZ_G2a: Benjamin, Rebecca, Sophie, Ladina)

In the example from grade 6 (Figure 15), longer clusters of up to nine contributions are present from the beginning on, which means that the students immediately enter an in-depth discussion on the individual objects. In contrast, the example from the grade 2 group (Figure 16) shows that in the first half of the discussion clusters of single contributions and two contributions predominate and that the maximum cluster size does not exceed six contributions. These visualizations cannot be understood conclusively as results, but they do allow a differently adjusted view on the conversational data. Thus, for example, a qualitative analysis could ask for triggering actions of larger thematic clusters.

In summary, most of the shown queries allow a global overall view, but they also lead back to the data, because the details cannot be mapped conclusively. It is true that the level of detail of the codes could be adjusted, which would make possible more reliable quantifications. This would require a complete re-encoding process, however, which would be extremely time-consuming. Another possibility would be to differentiate selected codes into smaller sub-corpora. On the whole, however, the return to a qualitative analysis seems to us to be necessary in any case, either to locate interpretations in context (see, for example, the combined use of mitigating modal particles with reinforcing modal particles and the mere repetition of positions) and to grasp phenomena more precisely (e.g. groups composed only of girls and groups composed only of boys), or to apply more detailed codes in individual cases (e.g. propositions with explanatory status, gestures as justifications or justifications through prosodic marking). The combination of different methods and a recursive approach seems to be promising to us. Thus, the results show how quantitative and qualitative analysis can be combined to achieve new insights. Finally, we discuss the advantages and difficulties of the methods applied in our project.

5. Discussion of the Combination of Qualitative and Quantitative Methods

The data were collected with the aim of not only gaining qualitative insights into individual conversations, but also making representative generalizations with regard to specific aspects of argumentation competence. For this end, we obtained a sufficiently large amount of data (180 discussions) and kept the setting as stable as possible. In addition, typical school tasks were assigned and the children conducted the conversations without the guidance or presence of adults, so that the conversational-analytical principle of researching authentic conversations could be followed as much as possible (see Section 2).

If one wants to evaluate conversational data quantitatively, it is helpful to estimate certain interpretation premises in order to answer the research question and develop meaningful codes from the data. However, these codes are not generated by purely hypothetical, theoretical assumptions (contrary to the presuppositions of classical quantitative approaches; see introductions such as Albert/Marx 2010), but *bottom-up* by prior hypothesis-generating qualitative and in our case conversation-analytical data analysis (Stivers 2015:9). Fundamental to qualitative research methods – and of significant importance to quantitative approaches as well – is the finding of shared characteristics in comparable data (Witt 2001: paragraph 10; see also Kleining 1982, our translation):

> Within the framework of heuristic qualitative research, the analysis of the data is primarily concerned with finding commonalities, i.e. those aspects in the data that represent the unifying factor in data with the greatest possible heterogeneity. These commonalities can be found by grouping the data, by 'asking' the data, by contrasting with the opposite, by negation.

As briefly discussed in Section 1, therefore, quantitative research of discussion data is not at odds with CA, but likewise proceeds from a qualitative analysis, which is used as "a solid foundation from which to build formal coding schemes" (Stivers 2015:5). Accordingly, the coding procedure is not very different from the one used in CA[38] and is generally object-based. However, systematic coding now enables new analyses that, for example, allow comparisons over large corpora (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.

Careful consideration of which specific research questions are to be applied to the material and how qualitative observations can be incorporated into the development of interest-driven codes is a prerequisite for generating meaningful codes for the material. However, Stivers (2015:13f.) rightly points out that "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." Accordingly, the further development of the theory is difficult: "[…] a coding scheme [...] represents the state of the art at the time and freezes it, making real-time CA advances difficult" (Stivers 2015:13). Although we did not develop the codes purely theoretically, but tested and optimized

[38] Here, too, one asks oneself "How normative must or may argumentative analysis be, and how descriptive or reconstructive must or can it be? (Deppermann 2006:17, our translation).

them in several test runs on small amounts of data – in this way, codes can be made more precise or further differentiated later on – it is usually not feasible to perform a complete post-coding, especially when evaluating large corpora.

Therefore, the (recursive) phase of the elaboration of a coding scheme is central, because the meaningfulness and possibilities of the analysis depend on it. As mentioned above, the codes can be developed based on qualitative analysis. Hereby, it is important to consider that the phenomena focused on have already been sufficiently described before the coding process (see Steensig/Heinemann 2015:21). If this step is successful, the process of subsequent coding itself represents a kind of data analysis and makes it possible to systematically uncover new typical cases and variances (Steensig/Heinemann 2015:21):

> [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 analyses can thus raise new questions and provide the impulse to qualitatively re-examine the data and specify subtypes analytically in more detail. Thus, they do not represent the end of the investigation, but are understood as "the prompt for new, qualitative, investigation" (Steensig/Heinemann 2015:21) and can offer another approach to better understand the social significance of the choice of certain linguistic actions in interaction (see Holmes/Meyerhoff 2006:12f.).

Finally, standardized test procedures can be used to generate results concerning frequencies on the basis of the codes or text searches, which allow more reliable statements about extensive corpora than are possible in qualitative studies. Quantification also makes it possible to relate speech behavior to socio-demographic data regarding the interlocutors and, for example, establish connections between activities such as justifications and age or modalizations and gender. Whether the results are actually representative and generalizable depends on the sample and the number of tokens of the phenomenon focused on.

Of course, the subsequent quantitative collection of qualitatively analyzed data entails a reduction "from the intricate complexities of human behavior to broad and flattened categories" (Stivers 2015:2), since "some of their meanings are dispensed with and transferred into a more abstract form" (Witt 2001: paragraph 4, our translation). In the case of conversational data, which due to its complexity contains *many different* potentially relevant phenomena and is *always differently structured* in its context, particularly strongly reducing decisions must be made. The codes to be developed will be generalized to the extent that they apply to the majority of the data and set "hard boundaries," although the phenomena and conversational activities "may be better understood as continuous rather than categorical" (Stivers 2015:13). In a qualitative analysis, for example, contributions can be shown to be justifications that only function as such because they are embedded in the syntactic structure of the previous utterance (see example codes levels 1, 2 and 3, Section 3.2. S1: aber wAs ist wenn wir gar keinen emPFANG haben, → S3: stImmt. dann nEhmen wir die lEuchtrakten (S1: but what if we do not have any reception, → S3: true. in that case we will take the flares)), as well as those that attain justification status through prosody (Mundwiler/Kreuz 2018) or are even realized through gestures – but these do not find their way into the code "justifications." The results are thus minimized in their level of detail, and any interpretation must recognize that it

offers only a rather superficial representation that can only be differentiated through a detailed investigation. Even though simplifications and generalizations of this kind have to be dealt with carefully, their advantage is that certain phenomena can be described over a large amount of data, which makes connecting elements visible.

Although the view of the data is reductionist and only a few, clearly identifiable phenomena are analyzed quantitatively, the process of coding is still very time-consuming, not only because of the large quantity of data, but also because there is still scope for interpretation in the assignment of the codes: the participants in spontaneous conversations do not interact according to a schema, but rather produce continuously situated linguistic actions that are ambiguous when considered without their local context. Since it is usually not a matter of quantifying purely verbal, near-surface phenomena, but rather of coding on the basis of interpretatively gained categories, decisions must continuously be made during the rating. Also, the method of working during the encoding process – whether one works only with a transcript or also considers the videos – has an influence on the assignment of certain codes.

Thus, the question of the level of detail of the codes arose again and again: how detailed must and can the codes be in order to render the data sufficiently detailed and at the same time unambiguous? The dangers of overly differentiated codes are, on the one hand, that it will take a long time to deal with large language corpora and, on the other, that it will become more difficult to compare the data (Stivers 2015:14):

> [A]dditional rules can be introduced such as operationalizing the code of the answer to be the first TCU [turn-constructional unit], 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.

Therefore, in cases of doubt, either (repeatedly new) *specifications* must be made that are consistently and consequently applied to other cases, or certain codes must be omitted. Continuous data sessions support the process of calibration, and a double rating is essential to increase objectivity and establish interrater reliability.

Quantification is a time-consuming task in several respects. First of all, it is necessary to collect a sufficiently large amount of data and prepare it accordingly (transcription). Then the codes have to be developed through qualitative analyses, in partly recursive processes with several test runs and numerous data sessions, until the phenomena can be sufficiently characterized based on the material and until they are evaluated in the same way by several people. Subsequently, the entire corpus has to be annotated, followed by a second, verifying annotation, and finally the results must be presented and evaluated with different tools (see Sections 3 and 4). Familiarization with the technical requirements and the generation of results requires comprehensive and precise training. If one is working with quantitative methods for the first time, it is also advisable to consult experts and resort to them when interpreting the results (or generating the figures and diagrams).

In our opinion, both qualitative and quantitative approaches to conversational data generate important results and contribute to a more comprehensive understanding of the interaction. We consider the different methods and visualizations complementary approaches that make it possible for "the blind spot of one method to

be compensated by the other" (Krüger/Pfaff 2008:161, our translation). Accordingly, the choice of methods conceptualizes the object of research from the perspective of a different research logic and produces different results on a given phenomenon or the overall structure of the conversations. While quantification may neglect the manifold, context-bound variations at certain points, it does enable insightful statements about more global phenomena, speaker profiles, developmental tendencies and so on that would not be possible by looking at the transcripts alone. We hope that by presenting different approaches, we have shown which opportunities, but also which limitations, need to be taken into account in quantitative analyses of conversational data. At the same time, we advocate a systematic combination of CA and quantitative methods in order to be able to analyze and compare linguistic interaction from different perspectives.

6. Acknowledgements

The project was supported by additional people at various times. We thank the student assistants and interns for their support in transcribing and annotating the conversational data: 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 and Chantal Wanderon.

We would like to thank Manfred Jurgovsky for his advice in the selection of suitable software, his support in preparing and transcoding video data and developing and programming the annotation panels for EXMARaLDA and his consulting in relation to them. We would also like to thank Thomas Schmidt, who, as one of the EXMARaLDA software developers, was responsible for the software and always advised and supported us with technical (compatibility) problems and questions.

Finally, we also thank the Swiss National Science Foundation (SNF), since our research was only possible because of its financial support (Project Funding, Dept. I-III, No. 149382).

7. References

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 (Eds.): Lernstandsbestimmung im Fach Deutsch. Gute Aufgaben für den Unterricht. Weinheim: Beltz, 52-77.

Becker-Mrotzek, Michael (2009): Mündliche Kommunikationskompetenz. In: Becker-Mrotzek, Michael (Ed.): 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 (Eds.), 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 (Eds.): 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. (Eds.): 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 / Brünner, Gisela (Eds.): Analyse und Vermittlung von Gesprächskompetenz. Radolfszell: Verlag für Gesprächsforschung, 15-28.

Deppermann, Arnulf (2006): Desiderata einer gesprächsanalytischen Argumentationsforschung. In: Deppermann, Arnulf / Hartung, Martin (Eds.): Argumentieren in Gesprächen. Gesprächsanalytische Studien. 2. Aufl. Tübingen: Stauffenburg, 10-26.

Deppermann, A. (2008): Gespräche analysieren. Eine Einführung in konversationsanalytische Methoden. 4th ed. Wiesbaden: VS Verlag für Sozialwissenschaften.

Deppermann, Arnulf / Lucius-Hoene, Gabriele (2006): Argumentatives Erzählen. In: Deppermann, Arnulf / Hartung, Martin (Eds.): Argumentieren in Gesprächen. Gesprächsanalytische Studien. 2nd ed. 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 (Eds.): 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 / Brünner, Gisela (Eds.): Analyse und Vermittlung von Gesprächskompetenz. Radolfszell: 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 (Eds.): 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 (Eds.): 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 (Eds.): 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 (Eds.): 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 38l/39, 9-57.

Kleining, Gerhard (1982): Umriß zu einer Methodologie qualitativer Sozialforschung. In: Kölner ZS für Soziologie und Sozialpsychologie 34, 224-253.

Knoblauch, Hubert (1995): Kommunikationskultur: die kommunikative Konstruktion kultureller Kontexte. Berlin, New York: de Gruyter.

Kreuz, Judith (in press): Ko-Konstruiertes Begründen unter Kindern. Eine gesprächsanalytische Studie von Kleingruppeninteraktionen in der Primarschule. Tübingen: Stauffenburg.

Kreuz, Judith / Luginbühl, Martin (2020): From Flat Propositions to Deep Co-constructed and Modalised Argumentations. Oral Argumentative Skills among Elementary-school Children from Grades 2 to 6. Research on Children and Social Interaction 4(1), 93-114.

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 / Hannken-Illjes, Kati / Kurtenbach, Stephanie (Eds.): 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 (Eds.): Handbuch der Schulforschung. 2nd rev. and expand. ed. Wiesbaden: VS Verlag für Sozialwissenschaften, 157-179.

Kotthoff, Helga (2015): Konsensuelles Argumentieren in schulischen Sprechstunden. In: Hauser, Stefan / Mundwiler, Vera (Eds.): Sprachliche Interaktion in schulischen Elterngesprächen. Bern: hep, 72-94.

Kyratzis, 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 (Eds.): 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 (Eds.): 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 (Eds.): 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 (Eds.): 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 (Eds.): 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 (Ed.): 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 (Ed.): 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. Ed. by Gail Jefferson / Emanuel A. Schegloff. Oxford: Blackwell.

Scarvaglieri, Claudio (2017): "I a hundred per cent agree". Konsensuelles Argumentieren in der interkulturellen Wissenschaftskommunikation. In: Bührig, Kristin / Schlickau, Stephan (Eds.): 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 (Eds.): 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, 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 8, 57-72.

Schmitt, Reinhold (2002): Rollenspiele als authentische Gespräche Überlegungen zu deren Produktivität im Trainingszusammenhang. In: Brünner, Gisela / Fiehler, Reinhard / Kindt, Walther (Eds.): Angewandte Diskursforschung. Band 2: Methoden und Anwendungsbereiche. Radolfzell: Verlag für Gesprächsforschung, 81-99.

Schwitalla, Johannes (2012): Gesprochenes Deutsch. Eine Einführung. 4th. rev. and expand. ed. 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 (2011): A system for transcribing talk-in-interaction. GAT 2 (translated and adapted for English by Elizabeth Couper-Kuhlen and Dagmar Barth-Weingarten). Gesprächsforschung 12, 1-51.

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 (Eds.): The Handbook of Conversation Analysis. Chichester: Wiley-Blackwell, 77-99.

Sidnell, Jack / Stivers, Tanya (Eds.) (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 (Eds.): Argumentieren in Gesprächen. Gesprächsanalytische Studien. 2nd ed. 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 (Eds.): Didaktik der deutschen Sprache. Ein Handbuch. Part 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. Martin Luginbühl
Dr. Daniel Müller-Feldmeth

German Department
University of Basel
Nadelberg 4
4051 Basel
Switzerland

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

Prof. Dr. Stefan Hauser
Dr. Vera Mundwiler
Dr. des. Judith Kreuz

Centre for Oral Communication
University of Teacher Education
Zugerbergstrasse 3
6300 Zug
Switzerland

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

Veröffentlicht am 14.5.2021