

Approaching Structured Debate with Quantitative Ethnography in Mind

Jennifer Scianna, Rogers Kaliisa, Jamie J. Boisvenue, Szilvia Zörgő

This paper has been published in the proceedings of the International Conference on Quantitative Ethnography (2021)

Citation information:

2022 Scianna, J., Kaliisa, R., Boisvenue, J., Zörgő, S. Approaching Structured Debate with Quantitative Ethnography in Mind. In: Advances in Quantitative Ethnography. Communications in Computer and Information Science, Vol 1522. Eds. Wasson B and Zörgő S., pp 33–48. Cham, Switzerland: Springer Nature.

Approaching Structured Debate with Quantitative Ethnography in Mind

Jennifer Scianna, Rogers Kaliisa, Jamie J. Boisvenue, Szilvia Zörgő

This paper has been accepted to the International Conference on Quantitative Ethnography (2021) and will be published in its proceedings. Please check back here in December 2021 for precise citation.

Approaching Structured Debate with Quantitative Ethnography in Mind

Jennifer Scianna¹[0000-0003-1029-3452], Rogers Kaliisa²[0000-0001-6528-8517], Jamie J. Bois-venue³[0000-0003-3344-7952], Szilvia Zörgő⁴[0000-0002-6916-2097]

¹ University of Wisconsin, Madison, Wisconsin, USA

² University of Oslo, Oslo, Norway

³ University of Alberta, Edmonton, Alberta, Canada

⁴ Semmelweis University, Budapest, Hungary

jscianna@wisc.edu

Abstract. Structured Debate (SD) is a constrained discourse style that is popular in many different forums. The expansion of SD to online platforms leaves many questions about addressing this type of data during analysis. Quantitative Ethnography (QE) may provide a framework for the considerations that need to be made when analyzing SD datasets. In this paper, we review the ways in which QE methods are compatible with SD and the challenges associated with applying this method. Using data from an online, SD forum, we present a narrative of the decision-making process throughout the analysis process. We find that QE allows for myriad insights to be gained from this form of data depending on the approach one takes including insights into structures, content, and participation. This work intends to serve as a model for researchers hoping to utilize QE on SD and, more broadly, for approaching novel datasets.

Keywords: Structured Debate, Quantitative Ethnography, Data Preparation.

1 Introduction

1.1 Classroom and Online Structured Debates

Structured Debate (SD) is a form of discourse that occurs in various settings including politics, economics, law, and business [1], as well as education [2]. SD is unique in that participant contributions are heavily structured and are arranged in a dichotomous “pro” / “con” configuration, that is, arguments for or against an initial claim [2]. As a pedagogical tool, SD promotes critical and rational thinking; students are required to organize and convey their thoughts in a logical and structured manner and need to be able to scrutinize and respond to claims made by the opposing side. During classroom SDs students are usually divided into two teams, one group argues for and one against an initial claim, such as: “Information production on social media should be uncensored”. Arguments are disclosed in a structured manner, for example, a series of affirmative and opposing arguments are presented alternately, which is followed by a recess, and then rebuttals for the arguments. The teacher may impose strict time limits for each

section and may set requirements for logical and rhetorical devices students should employ. This is an established way of practicing critical thinking, building rational arguments, learning about logical fallacies, and exercising emotional self-regulation [3]. Online platforms are also beginning to offer a forum for such debating [4]; for example, DebateMap¹, eDeb8², and Kialo³ provide an interface for participants to join a discussion on a certain topic. Several of these online platforms have been intentionally designed for pedagogical usage [5], while others have been retrofitted to serve the needs of educators [6]. Each participant can add an argument by designating it as a pro or a con to the initial claim, or they may choose to comment on any subsequent pro/con claim also by designating their argument as pro/con. In any given debate, there is much to be gained by analyzing pro/con arguments in relation to each other and to the initial claim. Furthermore, the content of each claim can be examined individually, and quantifying these aspects of the data presents an opportunity for identifying patterns in the entire corpus.

At first glance, SD may seem like an ideal candidate for quantification: its highly structured nature offers opportunities for segmenting, coding, and modelling such a dataset, but several issues may arise precisely in these realms; these can be addressed with Quantitative Ethnography (QE). QE is a unified, quantitative – qualitative methodology that provides tools and techniques for quantifying discourse data and representing it together with other types of data in a unified dataset [7]. Although SD is a common form of discourse and denotes a valuable repository for examining several aspects of education and development, such as critical thinking and collaboration, and albeit QE is a methodology suited for curating and analyzing such data, literature is scarce in this area.

Quantitative Ethnography researchers are yet to endeavor into the domain of SD. Thus far, the conversation surrounding debate has labelled it as a feature of more varied discourse. Barany et al. [8] analyzed student discussions and included “debate” as a type of interaction, but the authors were not specifically centered on investigating ways in which students formally engage in debate, nor did they explicitly deal with the structure of debate in general. Similarly, Nachtigall and Sung [9] utilize debate as a code to describe student interactions during collaboration, but do not delve into the components of that debate. Hamilton & Hobbs [10] considered political debates which would tend to be included as SD, but their analysis strayed from the pro/con focus and towards a more epistemic view of arguments and positions. We were interested in utilizing QE methodology to map the structure of SDs, identify challenges and affordances of debate discourse, and explore some ways to model such data. In this paper, we consider the nuances of analyzing SD using quantitative ethnographic methods.

1.2 Key Facets of Structured Debate

Structured Debate, also referred to as constraint-based argumentation [11], is a specific form of discourse where participants support position statements with evidence to expose the different sides of an issue. Katzenstein [1] proposed that SD included several

¹ <https://debatemap.app/>

² <http://www.edeb8.com/>

³ www.kialo.com

features that should be considered when analyzing this style of discourse: argument structure, stimulus operationalization, subject training, sample, and debate method. While they were primarily concerned with SD in business settings, these considerations can help identify aspects of the discourse in alternative settings as well.

Argument structure is related primarily to the format of the debate. For example, in some SDs, participants may have clear roles that they perform throughout the event. Musselman [12] utilized roles to help engage all participants. Antagonists were responsible for arguing a set position, Questioners limited discussion once an argument had run its course, and Conciliators developed compromises between the extreme positions. In digital spaces, we see some of these same roles come to light; online forums often have different roles for those who are actively involved in the debate, those who are moderating, and those who are engaged from the sidelines. Participants in the debate practice evidence-based reasoning to construct arguments supporting their position and refuting their opponents. Participants in the audience may vote, deliberate, provide written feedback, or ask questions.

Other aspects of the formal structure include the number of team members, purpose of each participant's turn, and even the activities that occur outside of the debate [2]. While information about participant educational background or preparation is not readily available to researchers in many digital spaces, it may be critical to know whether participants in a debate accessed information on their own or if they were all given the same information prior to the debate. This relates to the concept of subject training because participants may not be familiar with the content of the debate or with the structure. Consideration should be given to whether the participants are experts or not when considering the data.

Finally, the pattern of discourse is variable between different forms of SD. Interactions typically occur in either rounds or branches. In rounds, participants commonly engage in three phases: introduction, refutation, and conclusion [2]. Conversely, a branching structure allows participants to add support or rebuttals at any stage of the debate. Virtual spaces that support branching debate, such as Kialo and DebateMap, do not limit participants to a certain number of rounds. There are also hybrid formats such as the Human Continuum where participants consider their stance towards a topic and position themselves along a continuum. They then debate with neighbors providing evidence for a set amount of time, move if they feel their position has changed, and then begin again [6].

1.3 Previous Approaches to Analyzing Structured Debate

Non-digital SD has typically been studied in terms of activity design [1], pedagogical practice [12], and impact on learning [2]. These studies primarily focus on the structure of the activity itself compared to the outcome, but they rarely focus on the intricacies of discourse happening within the debates themselves. QE could provide a toolkit for addressing this analysis gap in literature.

Research on digital SD platforms has been centered around natural language processing techniques. The ability to clarify a participant's overall stance has been of interest to language processing researchers [4]. For example, Bolton et al. used debate data from Kialo to train an AI agent that was capable of debating with a user [13].

Thus far, investigations into debate discourse have encountered common challenges. For example, understanding the stances expressed in these debates requires modeling both textual content and users' conversational interactions. Bolton et al. addressed this struggle by filtering down their data to a subset of the datasets where positions were predictable [13]. Currently, there is a gap in how researchers with a more computational or qualitative focus can model conversations which switch perspectives while still being interrelated. Challenges of determining the true meaning of sentiment or other discourse modelling strategies will likely also fail to deliver in this context. These issues can be addressed with QE, a unified methodology that enables the systematic coding and representation of discourse- and metadata (i.e., information characterizing or accompanying data). The following is a modular account and worked example of how the QE methodological framework can be employed for SD analyses, for data from online branching SD platforms. The materials produced during our process (e.g., operationalization, codebook, dataset) are openly available in our public repository: placeholder for blinding purposes.

2 A Quantitative Ethnographic Approach to Structured Debate: Challenges and Affordances

2.1 Identifying Potential Research Questions

General Considerations. As with all data, SDs are suitable for answering only certain types of research questions, which are based on the affordances and constraints of how the data was generated, its characteristics, content, and structure. Due to the aims and the nature of debate, collected data will lend itself well to comparing and contrasting two principal opinions, which are generally broken down into their respective supporting arguments. Thus, data is symbolically (or in the case of some virtual platforms, literally) colored according to this dichotomy. Apart from being inherently dichotomous, SD data can be considered threaded: each pro/con argument can be conceptualized as a parent claim, which can have one or more children, affirmative or negative claims of their own. Threading can take many forms in SD data based on the structure of the debate or any meaningful and relevant system designated by the researcher. Thus, SD data, among many other characteristics, exhibits two main qualities: it is usually both dichotomous and threaded. This provides an opportunity to gain an in-depth understanding of two opposing opinions, for example, examining its content, the kind of supporting evidence that was utilized, similarities between arguments on one or both sides of the debate, etc. It also enables analyses on the relationship among claims and the deeper structure of the entire corpus. Due to the ways in which participants interact in a SD, it would be more challenging to discern public opinion based on the resultant data. This is because participants may be participating in SD through the role of devil's advocate [1] or be assigned a viewpoint in conflict with their own [12].

Our Example. We used data from an online SD platform in the topic of COVID-19 and herd immunization; the initial claim was that "Herd immunization is achievable". We defined our research questions in three domains: 1) Content (RQ1: What is the content of the claims?), 2) Structure (RQ2: How are the claims related to each other and how

can the structure of the corpus be described?), and 3) Positionality (RQ3: What does the pro/con stance entail in terms of relation and meaning?). As explained in greater detail below, codes were developed for each of these domains separately.

2.2 Data Collection

General Considerations. While synchronous classroom debates would generally be audio-recorded and then transcribed, collecting data from online platforms usually entails manual or automatic scraping [14]. When data is scraped automatically, it may come in many forms with different types of content. For example, for each claim, Kialo provides a unique identifier, positionality, and the text itself, e.g., “1.1.1.2.1. Con: Text goes here [with hyperlink words here] (and actual link here).” The identifier denotes the position of a claim in the “generations” (claims belonging to different parents but on the same level of debate structure) succeeding the initial claim; here it is the fourth generation or level. The position of the line is noted as a “Con”. Furthermore, the use of brackets connotes areas of text that are hyperlinks with the address of the link designated by parentheses. DebateMap, another online SD forum, includes similar information (ID and positionality), but they disclose a list of parent claims and offspring, timestamp, author, and metadata that other participants vote on (truth, relevance, and impact). As a point of ethics and data protection, researchers need to be aware of who owns the data, even when it is public, and what kind of data they are processing and how [15].

Our example. We manually scraped data from an online platform by modelling it in a text file, mimicking the process of automatically scraping the data with the built-in downloading feature. We did not scrape the content of claims, rather, we performed discourse coding by using the interface; in our mock-up, we designated a unique identifier for each claim, noted their autogenic positionality specified by the interface (pro/con), and made note of whether hyperlinks were added to the claim as supporting evidence. We opted for creating a mock-up instead of scraping the data in order to protect user identity and avoid any intellectual property concerns. The online platform we chose allows participants to add any number of children to the initial claim and to each subsequent claim; thus, it can be classified as a forum for branching debate. Our publicly available dataset contained one initial claim and 65 subsequent arguments on 6 levels.

2.3 Data Curation

General Considerations. In the QE framework, discourse is first segmented into “utterances”, the smallest unit of analysis. Utterances are commonly operationalized as a sentence or a turn-of-talk from a participant. Each utterance receives a unique identifier. For utterances to be associated with discourse codes and metadata, they are eventually represented in a dataset called a qualitative data table. Such a dataset contains the raw data (e.g., individual claims) in rows, while columns contain variables, such as who made the claim, what positionality it had, when the claim was made, and where it is in a claim structure; these can all be considered metadata and can be represented with

categorical values. Raw data is also accompanied by discourse coding, codes developed for qualitative data describing any relevant aspects of the content, such as whether the claim contained a reference to a certain theme pertinent to the research question. This coding is represented in binary form, that is, every time a code is present in an utterance, it receives a 1, and if the code is not present, it is designated a zero. Metadata is used to group utterances, e.g., a team making a series of claims or a round in which a claim occurred, while discourse codes are employed to provide a quantified, more abstract version of the utterance’s content. Apart from the raw data, discourse codes, and metadata, the qualitative data table may also contain information on segmentation, i.e., ways of dividing discourse into meaningful parts. Segmentation may be performed based on metadata or other ways of grouping discourse, such as threading or theme-oriented delineation. The qualitative data table signifies a dataset where both qualitative and quantitative data can co-exist, be integrated, and be further processed in analysis and modelling.

Our example. We decided to use a QE tool, the Reproducible Open Coding Kit (ROCK) [16], to aid us in data curation and coding. The ROCK is a standard aimed at making qualitative and QE research more transparent and machine readable, and it is implemented in an R package that enables specifying metadata, aggregating information to create the qualitative data table, and performing various analyses. The ROCK is also implemented in a graphical user interface (iROCK) that eases manual coding of text files (see below). We placed scraped data into a plain text file; each line of data contained the following: UIDs indicating the location of the claim in the debate structure (e.g., 1.1.1.2.1.), its positionality (e.g., Con), and a label for any embedded hyperlinks. In the ROCK standard, data providers are called “cases”; in this instance we considered each utterance as belonging to a different data provider, even if a single participant could have added more than one claim. In other initiatives, where participants themselves are also units of analysis, each data provider could be given an ID and utterances could be assigned to these cases. Each utterance ended with a newline character; thus, every utterance was on a separate line. Our dataset contained 66 codable lines of data from 66 cases. See section 2.7 for a schematic version of our qualitative data table.

2.4 Designating Metadata

General Considerations. There are several variables that can be considered as potential ways of grouping utterances in SD. Metadata for online SD data could potentially include timestamps, participant IDs, number of comments added to claims, number of edits made to claims, number of views a claim receives, number of children a claim has, its positionality, and any other type of value the online platform or the researcher assigns to specific claims or debates.

Positionality. Pro/con positions hold a different meaning depending on platform and debate structure. For round-based platforms (such as eDeb8), the position of each participant is assigned prior to the discussion. It is therefore reasonable to assume that the content of each utterance by the participant will be consistent with the assigned position. Conversely, in the case of online branching SD (such as Kialo or DebateMap), a

student may take on several viewpoints, may be against the initial claim, but support several child claims, or vice-versa. Thus, the positionality of a claim to both its parent and to the initial claim is crucial in understanding claims and stances of individual participants, diverging/converging viewpoints, and in mapping discourse structure on a grander scale. Positionality can be represented as metadata; in this case, each claim may receive a categorical value “pro” or “con” and this may be used to group utterances in the corpus. If both the relationship to the initial claim, as well as the relationship to the parent (in instances when these two differ) is of interest, these can constitute separate variables in metadata.

Our example. In the ROCK standard, metadata that are characteristics of data providers are called attributes. We decided to designate the following attributes to each case: *case ID* (cid; label was almost identical to UID), claim’s *relation to its parent* (PC; pro/con), claim’s *relation to initial claim* (APC; pro/con), *alignment* of relation to parent and initial claim (align; “concordance” = if both pro or both con / “discordance” if mixed pro and con), number of *children* (child; claims directly stemming from claim), number of *descendants* (desc; all succeeding claims). Case IDs were placed into the discourse in their respective lines in a format that enabled the ROCK to process other attributes describing the same case. The chosen attributes were utilized to represent information relevant to our research questions: relation to parent and initial claim, as well as alignment all gave us information about positionality, whereas number of children and descendants supplied metadata on segmentation and threading (see section 2.6 below).

2.5 Discourse Code Development and Coding

General considerations. Discourse codes are typically used in qualitative research to capture elements of interest within discourse for later exploration and possible aggregation [17, 18]. Endless possibilities exist for defining discourse codes and they are ultimately determined by the research questions. Some codes may represent a more literal aspect of the data, such as the code “vaccine” in the case of a debate on herd immunization. In developing such codes (as with any kind of code), researchers should be mindful of the granularity they wish to represent. In SD, granularity may manifest in common themes associated with pro and con arguments; if these themes are captured in a label and definition that allows for coding regardless of positionality, it may act as a foundation upon which the two sides can be compared and contrasted regarding how a particular piece of content is being employed to make an argument. For example, instead of just indicating the general topic of “vaccine”, one might develop a code e.g., “vaccine efficacy” which could traverse positionality and be employed in arguments that are both for and against the initial claim. Codes in isolation will rarely describe claim content accurately [17, 18]; code co-occurrences, on the other hand, offer a way to capture what distinguishes a pro and con claim. Other potential discourse codes include those that encapsulate ways in which an argument is made in terms of rhetorical devices (e.g., logos, pathos, ethos), type of reasoning (e.g., inductive or deductive), logical fallacies (e.g., circular argument, false dilemma, ad hominem), and so on. Additionally, forms of evidence brought to support a claim (e.g., empirical, testimonial, anecdotal, analogical) may be of interest regarding SD as well.

Relevance. In branching debates, where an initial claim can have any number of descendants, the content of claims may have different relationships to the initial claim, depending on which generation they belong to. The content of the initial claim will most likely be closest to its children and grandchildren; as one inspects generations farther from the initial claim, their content may be vastly different. Using the example above, first generation claims may be making arguments about vaccine efficacy, while later generations are formulating opinions on vitamins and nutrition. On many online platforms, debates are allowed to branch out organically into related topics. Depending on the research question and the code system, this may mean that other discourse codes need to be developed, or this “distance” to the initial claim needs to be represented in some other way.

Our example. We developed codes that aimed to denote the literal content of claims ($n=3$) and types of evidence brought to support an argument ($n=3$). Content codes relied on the nature of the claim’s discourse whereas evidence codes were more related to the presence of supporting sources being linked to the claim. These were developed in several phases: 1) four researchers inspected the dataset and performed free, inductive coding on the corpus; 2) these tentative codes were triangulated; and 3) a final coding scheme was created. The final version of these codes was applied to the corpus deductively via synchronous social moderation with the aid of iROCK. Another code was developed to express “distance” from the content of the initial claim; the code Relevance was expressed on a scale of 1 to 3, one being the most relevant (based on the debate’s description available on the online platform and also based on the content of the initial claim), and 3 being the least relevant (content did not contain topics, themes expressed in the debate description and initial claim). This was a single code that could receive three different values, e.g.: [[Relevance|1]]. A full description of our codes can be found in our codebook available in our repository.

2.6 Discourse Segmentation

General considerations. Segmentation entails dividing discourse into meaningful parts [7]; this process aids interpretability and performing various analyses. One such analysis involves “coding-and-counting”, that is, looking at code frequencies in various segments of the data or the entire dataset, while other types of analyses may examine code co-occurrences (see below). In both these techniques, frequencies may be markedly affected by how the discourse is segmented. Segmentation delineates temporal context [19], which is defined based on several considerations. For example, in SD data, meaningful segments may be delineated by the parent – child relationship (one claim and all its children) or this may be further divided into a parent with its pro children in one segment and its con children in another. Yet another meaningful segment may be a parent claim and all its descendants, or all claims within a generation regardless of which parent they belong to. Again, positionality may play a crucial role in analysis, thus the latter two scenarios may only be meaningful if further subdivided according to positionality. Additionally, claims may retain content-based relationships with each other across parents and generations: claims are relational. A claim made further out on

one branch may actually come temporally before - and inform - a claim added to a relatively shorter branch. However, since each claim is a direct response to their parent, this relationship may be more consistent and therefore significant in analyses. These differing temporal contexts can be indicated with various types of segmentation. Decisions in segmentation are highly dependent on debate structure and aims of research. Figure 1 contains different forms of segmentation for “round-based” debates, which may all be meaningful in the context of e.g., a classroom SD.

POSSIBLE SEGMENTATION	POSITIONALITY AND ACTIVITY	DETAILS
}	Affirmative opening argument	Summary of pro arguments
	Opposing opening argument	
}	Affirmative argument 1	Claim backed by evidence
	Opposing argument 1	
	Recess	Confer and discuss with team
}	Opposing rebuttal	Defend own claim and rebut opposing team’s claim
	Affirmative rebuttal	
	Opposing closing argument	Summarize debate and make final arguments
	Affirmative closing argument	

Fig. 1. Example of Structured Classroom Debate

Any number of “rounds” can be added to the above structure, and that would have implications on segmentation as well. Although debate structure can differ significantly when taken to an online platform, decisions concerning segmentation are very similar in offline/online and round-based/branching debates. Figure 2 displays an example of a branching debate and some possible ways of creating meaningful segments, temporal relationships among claims. Red and green nodes represent claims that were categorized as con or pro claims, respectively – both were specified in relation to their parent.

components of the arguments. Alternatively, research may be more interested in the ways in which participants contribute to the debate in what ways which may be best demonstrated with a social networking or integrated approach. The key affordance of QE methodology is that both qualitative and quantitative data can be represented in a unified dataset, and such a qualitative data table can be employed to perform various analyses, for example, Epistemic Network Analysis (ENA). ENA is commonly used for analyzing the structure of connections in coded data by quantifying and modeling the co-occurrence of codes as dynamic, weighted node-link networks. Such networks can be employed to compare discourse visually and statistically from individuals or groups [20]. This method has been used, for example, to analyze the development of learners' epistemic frames during play [21] and measure the co-occurrence of concepts within the conversations, topics, or activities that take place during learning [20], yet it has not been employed for SD. ENA offers insight into how codes interact in a single claim or a group of claims to produce meaning; code co-occurrences can be aggregated to scrutinize patterns in claim content or debate structure.

Aggregation. To employ ENA as an analytical tool, discourse codes, metadata, and segmentation need to be aggregated in a systematic manner. Each utterance (claim) receives its own row in this data table, while metadata appears in separate columns in categorical form. Segmentation is represented in one or more columns, as a number indicating temporal context or threading, which allows for utterances to be grouped together. Discourse codes are represented in binary form in separate columns; if a code manifests in a particular claim, it receives a 1, if it does not, it is designated a zero. We employed the ROCK R package to aggregate the above information as illustrated in Table 1.

Table 1. The schematic version of our Qualitative data table.

UID	CID	PC	APC	ALIGN	LEVEL	CHILD	DESC	DISCOURSE CODE 1	DISCOURSE CODE 2	DISCOURSE CODE 3
1.1	1_1	Pro	Pro	Concord	2	3	10	0	0	1
1.1.1	1_1_1	Con	Pro	Discord	3	2	4	1	1	0
1.1.1.1	1_1_1_1	Pro	Con	Discord	4	1	1	0	1	0

Model parameterization. To produce networks with ENA, several parameters need to be specified, which depend on what the researcher intends to model and how. The parameter “unit” operationalizes who or what a network is produced for, e.g., individual or groups of data providers (cases). In our example, each claim originated from a separate case. However, if we had collected data from a classroom debate, for example, where data providers belong to distinct teams, several claims could originate from a single case. Alternatively, other forms of segmentation can be employed to model the interaction of discourse codes: a family (parent claim and its children) or a generation (level) in a debate structure can constitute a unit. Any form of segmentation or any piece of metadata can potentially be designated as a unit for which a network can be generated. Temporal context is operationalized by the model parameter “conversation”; this determines the segment(s) of discourse where code co-occurrences can take place, i.e., which lines of discourse are grouped together. Again, any form of segmentation or piece of metadata can constitute a conversation. Co-occurrences among discourse codes

are accumulated in ENA by a “stanza window”, which specifies how co-occurrences are computed. In our example, we utilized a “whole conversation” stanza window for our aggregation approach and applied it to individual levels; this aggregated code co-occurrences across all cases in each level (our conversations). We chose this because we were particularly interested in how the claims that were further from the central claim varied, not in their connection to their parent (which would lean more towards a threaded approach). If one was more interested in the development of discourse over time and believed that the conversation built on itself, an infinite or moving stanza might be more appropriate.

Model generation. We created several models to visualize our data regarding our research questions on content, structure, and positionality. Figure 3 shows the interaction of discourse codes related to the literal content of claims and types of evidence brought to support an argument (n=6). Levels of debate structure constituted our units as well as our conversations; co-occurrences were accumulated with a weighted whole conversation stanza window.

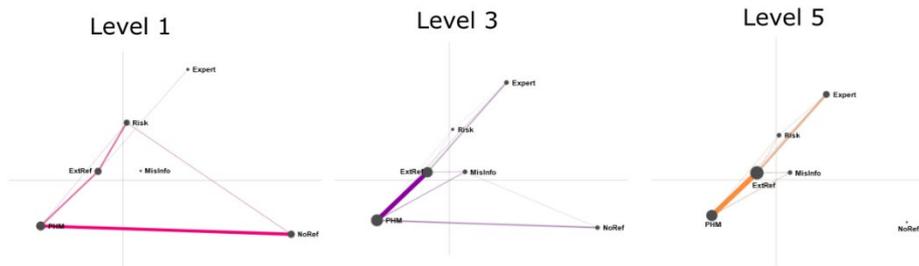


Fig. 3. ENA models of discourse across three labels of debate. Nodes reference codes whereas edges note the frequency of co-occurrence between the connected codes. Level 1 shows greater connections between the two bottommost codes whereas Level 5 shows codes concentrated to the leftmost x-axis and positive y-axis.

As depicted in these models, levels closer to the initial claim exhibited more connections among content codes, while later generations (more distant descendants of the initial claim) made less connections. Interestingly, the likelihood of a participant backing their claim with an external reference (as opposed to not employing such evidence) became higher in later generations.

We also modeled positionality on various levels of discourse. Figure 4 depicts how frequently autogenic pro and con claims aligned with their relation to the initial claim.

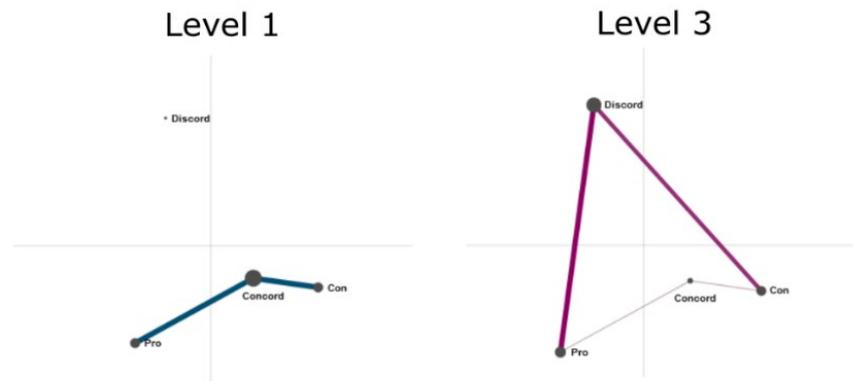


Fig. 4. A comparison of position and relationship to the initial claim on level 1 and level 3.

Figure 4 shows that on the first level (i.e., children of the initial claim), user designated pro and con claims aligned with their relation to the initial claim (“concord”), that is, claims which were e.g., against their parent, were also against the initial claim. By level three, more misalignment (“discord”) can be observed, as positionality with respect to the initial claim changes because of con arguments in the second level “re-positioning” their child claims.

3 Discussion

We aimed to map the affordances and challenges of taking a quantitative ethnographic approach to scrutinizing SD discourse. We discussed general considerations and worked examples in the domain of formulating feasible research questions, collecting and curating SD data, designating metadata, performing coding and segmentation, as well as analyzing and modelling the data. Our process description suggests that research questions answerable with SD discourse are primarily those comparing two viewpoints regarding the same initial claim, although online platforms may provide a venue for organic debate branching into other subject areas as well. Our research questions included mapping the content, structure, and positionality of SD discourse on an online platform enabling branching debate.

To begin, we curated a dataset containing 66 claims, including the one initial claim dictated by the debate creators; these were considered as originating from 66 different data providers (cases). We considered each claim as our smallest unit of segmentation (utterance), each claim received a unique identifier and was appended a newline character in a text file. We designated the following metadata (attributes): case ID (cid), positionality relative to parent (PC), positionality relative to initial claim (APC), alignment of the latter two attributes (concord/discord), number of children, and number of descendants. We developed discourse codes to capture the literal content and type of argumentation used in a claim, as well as indicate the relevance of utterances relative to the debate description and content of the initial claim. We segmented discourse by

indicating threading with a tilde; this linked parent claims to their children, but also made the scrutiny of generations possible.

Discourse coding was performed with iROCK by four researchers using synchronous social moderation. All information was parsed and aggregated using the ROCK R package, which also aided in creating a qualitative data table necessary for QE analyses and data modelling. Connections among discourse codes were modelled with ENA in network graphs showing both the structure and the content of our data. We visualized differences in content code co-occurrences according to level, which showed that although newer generations were making connections among fewer codes, they were also employing more external evidence to support their claims. In another model, we visualized how the positionality of a claim relative to its parent and to the initial claim becomes less aligned as we move farther from the initial claim.

Our process description aimed to highlight unique characteristics of SD discourse and how those can be addressed within a QE framework. Among these features are positionality, relevance, and threading. Positionality presents an issue analogous to sentiment analysis; in the latter technique, an algorithm tags pieces of content as “positive”, “negative”, or “neutral” [22]. Yet, in the case of branching SD on an online platform, a dataset can contain two manners of categorization: one pertaining to the claim’s relation to its parent and one to the initial claim, and these two may not be aligned. Continuing with the analogy, it is like retaining two sets of sentiments for each utterance, and in many cases, the utterance may be labeled as e.g., both a “positive” and a “negative” sentiment. Which is true? They are both accurate categorizations of positionality, and either one’s primacy will depend on how relational context is operationalized, which in turn is founded on the research question in focus.

Another consideration central to SD discourse is that of overall claim relevance. Although code development for most corpuses involves the same challenge - utilizing codes that capture discourse content well considering research aims - SD discourse does present a unique aspect. Since SD-related research questions will most likely involve comparing two key viewpoints on the same initial claim, branching debate structures where discussion can evolve organically into a plethora of topics poses a challenge for both inductive (bottom-up) and deductive (top-down) coding. It is difficult to develop a code structure that covers all pertinent components of branching SD discourse content, and most likely such code structures become large and unwieldy. Deductive coding (that is, using a limited, predetermined set of codes on the corpus) on the other hand, may lead to analyses that do not capture content to the desired extent.

Lastly, albeit SD is not the only source of threaded data, it does exhibit unique characteristics compared to e.g., more salient social media data (Twitter, Facebook, Reddit). Common to all these threaded data is the endeavor of determining temporal context, i.e., contextualizing an utterance: how far do we need to trace back in the data to understand an utterance and provide context for it? In Facebook data, for example, a comment can be contextualized within a thread (post and all comments), but a comment may have replies just pertaining to that comment, which may warrant a segmentation of its own. Yet branching SD data inspired the question of whether temporal context can span across a generation (children from multiple parents), which in Facebook data would mean contextualizing a comment-and-reply together with other comments-and-replies from under different posts. Thus, although segmentation frequently poses a

challenge, and threaded data is quite common, debates, especially online branching debates, introduce new questions in this domain.

4 Concluding Remarks

QE is a growing community of researchers interested in bridging a methodological gap between quantitative, qualitative, and mixed-methods researchers. The ubiquity of SD to varied settings supports the notion that it is not a matter of *if* QE researchers will encounter such discourse, but rather *when*. In this paper, we aimed to provide a framework of considerations for researchers as they undertake analysis of this type of discourse.

While these decisions were sensible for the dataset we sought to explore, there are more potential aggregation and display methods yet to be explored in the QE community. Each new development allows for expanded analytical opportunities for researchers, and it is here that we are excited to see the community grow. New digital environments will continue to develop in response to business, educational, and social demands, especially in a post-Covid era trying to navigate new-found freedoms of distance working, learning, and collaboration. As we continue to push on our own methods with novel data formats and sources, we will continue to refine our processes.

5 Acknowledgements

We would like to acknowledge the International Society for Quantitative Ethnography for their work in building a community which brought the authors together for the 2nd Covid Data Challenge.

References

1. Katzenstein, G.: The Debate on Structured Debate: Toward a Unified Theory. *Organizational Behavior and Human Decision Processes*. 66, 316–332 (1996). <https://doi.org/10.1006/obhd.1996.0059>.
2. Cariñanos-Ayala, S., Arrue, M., Zarandona, J., Labaka, A.: The use of structured debate as a teaching strategy among undergraduate nursing students: A systematic review. *Nurse Education Today*. 98, 104766 (2021). <https://doi.org/10.1016/j.nedt.2021.104766>.
3. Healey, R.L.: The Power of Debate: Reflections on the Potential of Debates for Engaging Students in Critical Thinking about Controversial Geographical Topics. *Journal of Geography in Higher Education*. 36, 239–257 (2012). <https://doi.org/10.1080/03098265.2011.619522>.
4. Sridhar, D., Getoor, L., Walker, M.: Collective Stance Classification of Posts in Online Debate Forums. In: *Proceedings of the Joint Workshop on Social Dynamics and Personal Attributes in Social Media*. pp. 109–117. Association for Computational Linguistics, Baltimore, Maryland (2014). <https://doi.org/10.3115/v1/W14-2715>.
5. McElfresh, J.: Spirited: A Web Application for Structured Debate, <http://nrs.harvard.edu/urn-3:HUL.InstRepos:33797385>.

6. McGreevy, P.D., Tzioumis, V., Degeling, C., Johnson, J., Brown, R., Sands, M., Starling, M.J., Phillips, C.J.C.: The Use of a Virtual Online Debating Platform to Facilitate Student Discussion of Potentially Polarising Topics. *Animals (Basel)*. 7, (2017). <https://doi.org/10.3390/ani7090068>.
7. Williamson Shaffer, D.: *Quantitative Ethnography*. Cathcart Press, Madison, WI (2017).
8. Barany, A., Shah, M., Foster, A.: Connecting Curricular Design and Student Identity Change: An Epistemic Network Analysis. In: Ruis, A.R. and Lee, S.B. (eds.) *Advances in Quantitative Ethnography*. pp. 155–169. Springer International Publishing, Cham (2021). https://doi.org/10.1007/978-3-030-67788-6_11.
9. Nachtigall, V., Sung, H.: Students' Collaboration Patterns in a Productive Failure Setting: An Epistemic Network Analysis of Contrasting Cases. In: Eagan, B., Misfeldt, M., and Siebert-Evenstone, A. (eds.) *Advances in Quantitative Ethnography*. pp. 165–176. Springer International Publishing, Cham (2019). https://doi.org/10.1007/978-3-030-33232-7_14.
10. Hamilton, E., Hobbs, W.: Epistemic Frames and Political Discourse Modeling. In: Ruis, A.R. and Lee, S.B. (eds.) *Advances in Quantitative Ethnography*. pp. 32–46. Springer International Publishing, Cham (2021). https://doi.org/10.1007/978-3-030-67788-6_3.
11. Cho, K.-L., Jonassen, D.H.: The effects of argumentation scaffolds on argumentation and problem solving. *ETR&D*. 50, 5–22 (2002). <https://doi.org/10.1007/BF02505022>.
12. Musselman, E.G.: Using Structured Debate to Achieve Autonomous Student Discussion. *The History Teacher*. 37, 335–349 (2004). <https://doi.org/10.2307/1555673>.
13. Bolton, E., Calderwood, A., Christensen, N., Kafrouni, J., Drori, I.: High Quality Real-Time Structured Debate Generation. arXiv:2012.00209 [cs]. (2020).
14. Mitchell, R.E.: Web scraping with Python: collecting data from the modern web. (2018).
15. Krotov, V., Redd, L., Silva, L.: Legality and Ethics of Web Scraping, *Communications of the Association for Information Systems (forthcoming)*. *Communications of the Association for Information Systems*. (2020).
16. Zörgő, S., Peters, G.-J.Y.: Epistemic Network Analysis for Semi-Structured Interviews and Other Continuous Narratives: Challenges and Insights, psyarxiv.com/j6n97.
17. Saldaña, J.: *The coding manual for qualitative researchers*. SAGE, Los Angeles (2013).
18. Williamson Shaffer, D., Ruis, A.: How We Code. In: *ICQE 2020: Advances in Quantitative Ethnography*. pp. 62–77. Springer, Virtual (2021). https://doi.org/10.1007/978-3-030-67788-6_5.
19. Siebert-Evenstone, A.L., Arastoopour, G., Collier, W., Swiecki, Z., Ruis, A., Williamson Shaffer, D.: In Search of Conversational Grain Size: Modelling Semantic Structure Using Moving Stanza Windows. *Journal of Learning Analytics*. 4, 123–139 (2017). <https://doi.org/10.18608/jla.2017.43.7>.
20. Williamson Shaffer, D., Collier, W., Ruis, A.: A Tutorial on Epistemic Network Analysis: Analyzing the Structure of Connections in Cognitive, Social, and Interaction Data. *Journal of Learning Analytics*. 3, 9–45 (2016). <https://doi.org/10.18608/jla.2016.33.3>.
21. Scianna, J., Gagnon, D., Knowles, B.: Counting the Game: Visualizing Changes in Play by Incorporating Game Events. In: *ICQE 2020: Advances in Quantitative Ethnography*. pp. 218–231. Springer, Virtual (2021). https://doi.org/10.1007/978-3-030-67788-6_15.
22. Misiejuk, K., Scianna, J., Kaliisa, R., Vachuska, K., & Williamson Shaffer, D.: Incorporating Sentiment Analysis with Epistemic Network Analysis to Enhance Discourse Analysis of Twitter Data. In: *ICQE 2020: Advances in Quantitative Ethnography*. Pp. 375–389. Springer, Virtual (2021). https://doi.org/10.1007/978-3-030-67788-6_26