Towards Automating the Measurement of Deliberative Communication

Valentin Gold*, Katharina Holzinger*, Christian Rohrdantz†
*Chair of International Relations and Conflict Management
†Chair for Data Analysis and Visualization
University of Konstanz, Germany
{firstname.lastname}@uni-konstanz.de

***Abstract*—The first step towards automating the measurement of deliberative communication is to analyze the content of a debate. In this paper we develop a novel visual method that allows to determine content-based episodes within discourses. Our framework reveals interesting subsequences of discourses that help to automate the measurement of deliberative communication.**

I. Introduction

The terms "deliberation" or "deliberative" are now widely used and may denote very different concepts. At the macro-level they refer to deliberative democracy, which is usually understood as a democracy that, in taking collective decisions, relies to a large degree on consensus-oriented discourse and argumentative communication instead of majority decision-making [1–4]. At the meso-level, they refer to specific deliberative fora, such as public discourses, policy dialogues, round tables, or citizens conferences, which are set up to deal with a given decision problem, be it the adoption of a collective rule or the resolution of a local conflict. At the micro-level, deliberation refers to deliberative ways of political communication, as opposed to, for example, strategic or rhetorical ways of communication. Deliberative political communication could be present in any setting, be it a specific deliberative forum, a parliamentary debate, speeches of politicians, or political negotiations at all levels. However, the concept of deliberation requires at all levels that certain (normative) attributes of the institutional setting, the communication and the individual behavior are present. As an ideal, the concept of deliberation requires not only for specific deliberative fora, in which political decision-making is both more discursive and more inclusive than in classical democracy, but for all political communication to be deliberative.

In this paper we focus on the content of deliberative communication. We are interested in detecting automatically the content of political discourses. Understanding the content of a discourse is the very first and basic step towards automating the measurement of deliberative communication. Only thereinafter, we can pose and answer questions such as: What makes communication deliberative? How can we distinguish deliberative communication from strategic or rhetorical speech? How can we measure the deliberative quality of political communication? To what exactly does the concept refer – to an individual utterance, or to the interaction of several speakers, or to a whole dialogue or discourse? These questions have partly been answered in 2005 by two special issues of *Acta Politica* in which attempts to empirically measure deliberation were collected. The most prominent concept for the measurement of deliberative communication is the Discourse Quality Index (DQI), developed by [5].

Our approach is distinct insofar as we develop methods for the *automated* measurement of deliberative communication. Even though we do not approach the automated measurement of deliberative communication in this paper, we start with a definition of deliberation. This is to illustrate our ambitions towards automating the measurement of deliberative communication. The definition is based on the most important properties of deliberative communication discussed in the literature. It encompasses ten dimensions, including institutional, communicative and individual properties. For the purpose of developing automatic methods it is sufficient to focus on those dimensions on which the actual verbal communication in a discourse can contribute information. This way, the paper aims to illustrate our ambitions as well as to demonstrate a novel framework of automatically determining the content of political discourses. Note that this paper presents work in progress.

The paper is organized as follows: In the next section, we present our definition of deliberative communication. Section III introduces our novel framework of lexical episodes. In section IV, we apply our framework to political discourses. Finally, in section V, we conclude with a short discussion of the findings and present an outlook of the next steps towards automating the measurement of deliberative communication.

II. Defining deliberative communication

The following definition is derived from a review of the literature on deliberative democracy and deliberative communication. The review included more than 70 books and articles on the subject which, regrettably, cannot all be cited here. We start from the ideas presented by Habermas in his works on communicative action [6] and discourse ethics [7, 8]. That is, conceptually, we rely on a Habermasian approach to deliberation. In order to put these ideas to an empirical test, they have to be made operational of which a first step is to define the concept. We do not know whether Jürgen Habermas would fully agree with our definition, however, based on our literature review, we think it captures most of his ideas and should be compatible.

The definition we propose (see table I) includes the goal of deliberative political communication, the institutional, and

TABLE I: Definition of deliberative communication

Goal	Deliberation is a communicative process that aims at taking a decision (or recommendation) on collectively binding rules or public projects. The substantive goal is to achieve the common good and universality of rules.
Procedure, institutional properties	As far as possible, the procedure • should take place in public (or at least be transparent to the public) • should guarantee inclusive participation • should guarantee equal rights to every participant
Individual behavior	Individual participants • should behave truthfully (authentically) • should communicate impartially • should behave respectful towards other persons and their positions, demands, proposals and arguments • should be open to become convinced by the better argument; that is, to change opinion and to sacrifice individual preferences in favor of the common good
Communication	The communication should be based on reason, that is, it • should be based on information as complete as possible about facts, norms, values and preferences of those participating or concerned by the decision • should be based on the argumentative justification of all positions and proposals • should lead to conclusions only if based on the power of the better argument; a good argument is considered to be empirically correct, logically consistent, and/or to refer to a universal/impartial and valid norm or value.

in particular procedural, requirements, requirements for communication, and requirements for individual behavior. It is important to note that the elements of our definition are normative requirements, as deliberative democracy and deliberative communication are normative concepts.

Obviously, as our goal is to analyze the micro-level of communication only, the institutional requirements are not what we want to measure, in the first place. Publicness and the right to participate is not part of the communication, although "inclusive participation" and "equal rights" can also be taken as the requirement that everyone can take part in the actual discussion to some extent. This we can measure by looking at each individuals share in the discussion, in terms of number of words or time spoken.

Similarly, the requirements for individual behavior cannot all be measured by looking at the language. For impartiality and respectfulness indicators may be found in the use of language. Whether someone speaks truthful or authentically, however, is a matter that is difficult to see from language used or any other observable characteristic. The openness to be convinced, to change opinion, or to sacrifice individual preferences is also difficult to be observed at the language level alone. Nevertheless, the language used might provide some indicators that can give hints on the attitudes and stances of speakers towards other speakers or the themes to be discussed. Those indicators may be sentiments and emotional speech, or the speakers intentions towards the others as revealed by speech act analysis, or the analysis of rhetorical means.

The requirements for communication are at the center of our research interest. The requirement that information on facts, norms, values and preferences should be as complete as possible is also something that cannot be evaluated by textual analysis of the discourse. The core of the concept of deliberation lies in the second and third requirement, that communication should be based on argumentative justification only, and conclusions be based on the better argument. A good argument is logically consistent, empirically correct, and refers to impartial and universal norms or values. Logical consistence of utterances can be evaluated within a text. Whether claims are empirically correct and norms are impartial and valid, can be seen from the text only insofar as the claims made are not contested by other participants. The extent of argumentative justification in a discourse can be measured in principle.

Many have tried to "turn Habermas empirical" and to test whether his ideas of consensual discourse as the foundation for democracy work in practice, how this can be achieved and under which conditions it is successful. All kinds of experimental deliberative fora have been set up [9], deliberative polling [10] and other experiments [11] have been carried out, trying to capture the institutional requirements Habermas demands as good as possible, and then measuring the effects in terms of, whether a consensus was reached [12, 13], whether participants were satisfied with the results, whether they had gained information [14], learned or changed opinions [4, 15]. Those studies have usually dealt with the institutional and sometimes the behavioral requirements and were interested in the effects of the setting or the conditions under which deliberation is successful [12, 16–18].

Not so many have dealt with deliberative communication at the micro-level of language and tried to measure the extent of deliberation. The most well-known approach is the Discourse Quality Index (DQI) that was introduced by Steiner, Bächtiger, Steenbergen and Spörndli [5, 19]. The DQI integrates indicators at seven levels: (1) participation in discussion, (2) depth of justification, (3) content of justification, (4) respect towards the other groups (5) towards its demands, (6) and towards counterarguments, (7) constructive behavior. It is clear that the DQI rests on those requirements of deliberation which can be coded on a transcript of a discourse, dialogue, discussion or debate, (whatever it may be called) – just as derived from the definition above. The DQI, however, is based on manual coding and seems difficult to automatize.

The second approach, which analyzes dialogues at the language level, was introduced and applied by Holzinger and Landwehr [20–22]. It is based on linguistic speech act analysis, and tries to get hold of the intentions of the speakers: are they engaged in strategic (bargaining) or in deliberative (arguing) communication? Although the academic debate on arguing versus bargaining showed that this juxtaposition is misleading, the instrument of speech act analysis might prove useful to uncover intentions of speakers and might thus contribute to the evaluation of the deliberative quality. So far, speech act analysis relied on manual coding, as well.

Based on this state of the art, two questions arise. The first question relates to whether the dimensions of participation in a debate, the extent of justification and argumentativeness, contents of justification, speaker intentions, impartiality and respectfulness, and more general constructive communicative behavior can be coded in an automatic manner using the modern tools of computer linguistics and information science. The second question is whether true deliberative communication has also other aspects that are not explicitly mentioned in Habermas concept and have not yet been addressed by the attempts to code it. For example, it seems intuitively plausible that successful communication also rests on the extent to which

the participants "talk to each other", i.e., to which they speak to the same subject and listen and respond to each other. This requirement does not directly follow from a Habermas based definition, but it is not incompatible with it. How good and intensively the participants in a dialogue interact might thus also be an element in the measurement of the quality of deliberation. Hence, in the next section, we introduce a novel visual framework to automatically detect the content of political discourses. Based on interdisciplinary experiences, the framework uses algorithms that are kept as simple as possible to also allow political science researchers with less knowledge of computer-based methods to understand the procedure. This is also why we strive for the visual method of discourse analysis.

III. Visual Analytics

Visual Analytics is "the science of analytical reasoning supported by interactive visual interfaces" [23]. The fundamental idea is to combine automatic and visual analysis methods, enabling human analysts to gain insight into the analysis and steer it.

From the political science perspective there are concrete analysis goals. For example, the degree of deliberation contained in a discourse shall be measured. While manual annotation schemes exist for this purpose (cf. [5]), there is no straightforward solution how to operationalize this in a way that it can be modeled computationally. Visual analytics research tackles the same problem from the opposed perspective starting from the questions: What potentially useful information can be measured and extracted from the data? Which meaningful components can be annotated automatically and presented to the human user? The aim is to first explore the range of opportunities for automatic annotation in order not to disregard any options that are technically possible, but have not been thoroughly investigated yet. Only then, in a second step it will be evaluated whether and how the extracted and measured information may contribute to beneficial automatic annotations, leaving room for innovation. Interactive visual interfaces may support such an evaluation and investigation. This proceeding follows the idea of designing so-called *quasi-semantic properties*, cf. [24]. A quasi-semantic property is a meaningful combination of low-level text features and statistics that support the analysis of certain semantic aspects contained in a text.

It is not to be expected that computational algorithms can faithfully imitate human annotation behavior. However, this is also not necessary and may even not be desirable in some cases.

While the annotation is done automatically, the interpretation is left up to the human analyst. Interactive visualization have been shown to help discover hidden underlying patterns, generate new hypotheses or verify or reject existing hypotheses. One example on how visual analytics methodology can complement traditional analyses and provide new perspectives is given in the Section III-A. While a human annotator can be much more accurate when an actual understanding of the text is required, computational methods for annotation also have their particular strengths that are valuable to be exploited. First, the algorithmic measurement always follows exactly the same process and in this sense is completely objective whereas human annotators are prone to be biased by subjective conceptions. This is shown by relatively low inter-rater agreement values in general text annotation tasks. Second, certain numerical measurements can be easily computed, but could not readily be performed by human annotators, i.e. they would require a too high cognitive load and be too time-consuming and error-prone to be practicable. One example for such an annotation that we designed will be detailed in the following.

A. Lexical Episodes

In this section we present a novel method for automatic content-based discourse annotation. If the whole discourse is seen as a sequence of words, the method discovers interesting subsequences based on word distributions. In particular, for each word appearing in the discourse, the algorithm searches for subsequences of the discourse in which the word is unexpectedly frequent. We denote such a subsequence as *lexical episode*. The method for the detection of lexial episodes was inspired by a time series analysis technique [25]. For each word of the discourse the algorithm performs the following analysis steps:

1) Retrieve the index positions (within the discourse sequence) of the word currently under investigation, which we denote as target word. For example, if the word occurs as the 4th, 27th, and 201st word within the discourse the list will be filled with the three indexes 4, 27, and 201.
2) Count the frequency of the target word in the whole discourse. In the simplified example this would be 3.
3) Calculate the expected gap (in words) between two occurrences of the target word assuming a uniform distribution of the target word over the discourse. If our example discourse contains 300 words, the expected gap between two occurrences of the target word in the example will be 100. The expected gaps vary from target word to target word, depending on the overall frequeny in the discourse.
4) Identify sequences of target word occurrences where the observed pair-wise gaps in the discourse are smaller than the expected gap. In the example, this criterion holds for the subsequence of the index 4 and index 27 ($(27-4) < 100$). According to our terminology the two index positions form a lexical episode. Yet, the word at index 201 does not belong to this episode, as the gap to the previous word is too large.

Lexical episodes mark parts of the discourse where the utterances refer to certain aspects, which in turn are underrepresented in the other parts of the discourse. Often, such aspects refer to concrete subjects of negotiation or parts of arguments. They indicate both content and argumentation patterns. Whereas individual lexical episodes have a rather limited expressiveness, the set of all lexical episodes as a whole may form insightful patterns. In the next section we suggest an innovative way of visualizing the discourse together with the lexical episodes for detailed inspection and hypothesis generation.

IV. Visualizing Lexical Episodes for Discourse Analysis

Before we present the visual analytics, we shortly describe the discourse data to which the lexical episodes are applied to.

A. Simulation-gaming experiments and "Stuttgart 21"

One of the main challenges in the analysis of deliberative discourses is the collection of data. This is due to several reasons: First, deliberative discourses are rarely recorded. Sometimes only the general discussions are taped, but not every discussion within the discourse. Moreover, in some cases where audio- or videotapes are available, the quality is rather bad which in turn increases the amount of money to spend for the text-transcription. Fortunately, some discourses are available as digital text.

One of these discourses is the public arbitration on "Stuttgart 21". Stuttgart 21 is a new railway and urban development project in Southern Germany. The project includes the restructuring of the central station in Stuttgart. Ever since the project was officially announced in the late 1980s, criticism against the project was expressed. It was not until the late 2000s, however, that large demonstrations and protests with over 100.000 participants took place. The main aim of the protests was directed against the demolition of the central railway station. On September 30, 2010, hundreds of protesters were injured when the police tried to secure the beginning of the construction work. In response to this event, the government finally agreed to establish a public arbitration procedure to discuss the facts of the project with both supporters and opposition. Between October 22 and November 27, 2010, the public arbitration took place. Within eight rounds of arbitration, supporters and opposition discussed the merits of the project. The discussions were live broadcasted and are all available online. The data we use to demonstrate the automated methods were voluntarily transcribed by several citizens that were interested in the public availability of the transcripts [26]. Each of the single rounds of arbitration were transcribed based on the videotapes of the dialogues and were split in several thematic episodes. For instance, one round was split in six episodes: three presentations and three discussions, each following one of the presentations.

The arbitration on Stuttgart 21 has one major disadvantage: As with any other real-world dialogue, one cannot compare the communication to other deliberative procedures, as the contexts are quite unique. Comparative data is, however, necessary to determine the validity of the single automated methods. Hence, in order to allow for a comparative perspective on deliberative dialogues, we also use laboratory data. The laboratory data stems from simulation-gaming experiments. In these experiments, experimental subjects were asked to discuss and design the political institutions of a fictitious country. In these role-play simulations, the subjects had to represent one of the main ethnic groups within the fictitious country. The subjects were confronted with opposing views on how political institutions have to be designed to satisfy the needs of all ethnic groups. To allow for a comparative analysis, the experimental subjects were provided with a predefined set of political institutions. On these institutions, they had to decide unanimously. Each subject was given a clearly defined political position which they had to fight for. For instance, the experimental subjects were asked to discuss the pros and cons of a parliamentary versus a presidential governmental system for the fictitious country. Whereas some subjects had to favor parliamentary systems, others were in favor for a presidential system. At the end, only one possibility was allowed to agree on. Overall, we run 15 experiments. The experiments lasted about 3.5 hours, with a total time of 2 hours of group discussion. In most experiments, the maximum of 2 hours of discussion was fully made use of by the subjects. This provides us with the necessary amount of comparative data to demonstrate the visual framework of lexical episodes.

B. Lexical episodes

Figure 1 and 2 each show one day of the Stuttgart 21 mediation together with automatically detected lexical episodes. On the right hand side the utterances are displayed as gray rectangles. The text becomes visible when zooming in, as for example in Figure 3. Further on the left, colored vertical bars represent lexical episodes. Each vertical bar contains a number of horizontal lines. The lines indicate the occurrence of the episode's target word in a sentence. Further to the left the target words are displayed in the same color as their episode's bar.

The episodes reveal which subjects are discussed at which time and breaks in content. On November 22nd, for example, there is a clear cut after the first topic discussed, see figure 2. Figure 3 is a zoom-in at the position of the break. The text reveals that the mediator Dr. Heiner Geißler explicitly requests a change in topic. Up to that point, the discussion addressed the modifications in the transport of goods (*Güterverkehr*) due to the new railway concept for Stuttgart and the surrounding region. The second topic addressed the advantages and disadvantages of dead-end stations (*Kopfbahnhof*). A similar pattern of topic distribution can be seen in figure 1: In the beginning of the discussion, obviously, some problems with the documents (*Unterlagen*) and the slides (*Folien*) were apparent. With regards to content, the first topic addressed the eco-balance of the new railway station (Co^2, *S21*) compared to road traffic in general and in specific to the existing dead-end station (*K21*). The third topic relates to the recreation area (*Bäume*) that will be created when most of the tracks run underground. The third and final topic in this discussion addressed the new urban district that was purchased by the city of Stuttgart already before the constructions began (*Gelände*, *Fläche*).

In figure 4, for four of our experiments it is shown how the discussions proceed over time. Since the experimental subjects had to make a decision on a predefined set of political institutions, the experiments can be compared with each other. For instance, in experiment 4, the discussion was much more structured than in the other three experiments: The first topic that the experimental subjects did address was the governmental system (*Regierungssystem*). They then proceeded with a discussion about the proportional electoral system (*Verhältniswahlrecht*). The third topic relates to the bicameral system (*Zweikammersystem*) and the potential inherited danger (*Gefahr*) for the fictitious ethnic group of the Komtoru. Finally, the subjects discussed the veto right (*Vetorecht*) and how the

government (*Regierung*) and the cabinet (*Kabinett*) should be organized.

In both experiments 1 and 2, at the very end of the discussion, the subjects tried to find a compromise (*Kompromiss*). Whereas in experiment 2, they were successful, in experiment 1 they failed because they could not agree on the electoral system. As a consequence, the subjects again addressed the topic of proportional electoral systems (*Verhältniswahlsystem*), before they finally gave up.

V. Conclusion

In this paper we have introduced a framework of how the content of discourses can be automatically identified. In general, the aim is not to replace human annotation and human analysis, but to complement and enhance it. The human analyst, of course, always has the option to design a measure based on the experiences gained through the interactive tool. For example, s/he could decide to up-rate dialogs with many episodes and test this variable's performance as a predictor for the success of a negotiation, i.e. whether an agreement is reached.

As with regards to our ambition to automating the measurement of deliberative communication, some caveats have to be expressed. First, the extent of deliberation and the deliberative quality of a given discourse or dialogue cannot only be assessed at the level of language. There are elements which require the test of other properties of the process, as well as elements which cannot be observed at all, such as the truthfulness of participants. Second, not all of the requirements for deliberative communication that can in principle be measured at the language or textual level can properly be captured by automatic procedures. Third, the automatic procedures rely on statistical measures; they are only approximations as natural language is so varied and complex that we can only grasp it to a certain percentage. Finally, from this follows that automatic coding and measurement only works with large amount of textual data, as only with great numbers statistical procedures become reliable. This is, however, exactly why we strive for the automated coding of discourses: we often have large bodies of texts which make manual coding too time consuming and expensive.

In order to complete the picture of deliberative communication, we currently consolidate a Java framework that unites single automated measures. For instance, we apply conceptual and topic recurrence plots (cf. [27]) to detect patterns of interactions across the discourse. We also make use of the Penn Discourse Tree Bank 2.0 Annotation scheme [28] in order to learn about the argumentative structure of discourses. The shallow analysis of the discourse is complemented by a deep and detailed linguistic analysis. The combination of different methodologies from information science and linguistics will allow us to close in on the factors that successfully determine deliberative discourses. In particular, the visualization of the discourse ensures that patterns can be quickly discovered across large corpora of texts without loosing relevant information on the deliberative quality of the discourse.

Funding

The joint research project "VisArgue – Why and when do arguments win?" is funded 09/2012 – 08/2015 under the research grant number 01UG1246 by the German Federal Ministry of Education and Research (Bundesministerium für Bildung und Forschung, BMBF).

References

[1] J. S. Dryzek, *Discursive Democracy: Politics, Policy, and Political Science*. Cambridge, MA: Cambridge University Press, 1990.

[2] ——, *Deliberative Democracy and Beyond: Liberals, Critics, Contestations*. Oxford: Oxford University Press, 2000.

[3] J. Bohman, *Public Deliberation: Pluralism, Complexity and Democracy*. Cambridge, MA: The MIT Press, 1996.

[4] A. Gutmann and D. F. Thompson, *Democracy and Disagreement. Why moral conflict cannot be avoided in politics, and what should be done about it*. Cambridge, MA: Harvard University Press, 1996.

[5] J. Steiner, A. Bächtiger, M. Spörndli, and M. R. Steenbergen, *Deliberative Politics in Action. Analyzing Parliamentary Discourse*. Cambridge, MA: Cambridge University Press, 2004.

[6] J. Habermas, *Theorie des kommunikativen Handelns*. Frankfurt am Main: Suhrkamp, 1981.

[7] ——, *Moralbewußtsein und kommunikatives Handeln*. Frankfurt am Main: Suhrkamp, 1983.

[8] ——, *Erläuterungen zur Diskursethik*. Frankfurt am Main: Suhrkamp, 1991, vol. 5.

[9] C. Landwehr, *Political Conflict and Political Preferences: Communicative Interaction between Facts, Norms and Interests*. Colchester: ECPR Press, 2009.

[10] J. S. Fishkin and R. C. Luskin, "Experimenting with a democratic ideal: Deliberative polling and public opinion," *Acta Politica*, vol. 40, no. 3, pp. 284–298, 2005.

[11] T. Sulkin and A. F. Simon, "Habermas in the lab: A study of deliberation in an experimental setting," *Political Psychology*, vol. 22, no. 4, pp. 809–826, 2001.

[12] A. Bächtiger, M. Spörndli, M. R. Steenbergen, and J. Steiner, "The deliberative dimensions of legislatures," *Acta Politica*, vol. 40, no. 2, pp. 225–238, 2005.

[13] A. Niemann, "Beyond problem-solving and bargaining: Genuine debate in eu external trade negotiations," *International Negotiation*, vol. 11, no. 3, pp. 467–497, 2006.

[14] B. Ackerman and J. S. Fishkin, "Deliberation day," *Journal of Political Philosophy*, vol. 10, no. 2, pp. 129–152, 2002.

[15] S. Chambers, *Reasonable Democracy: Jürgen Habermas and the Politics of Discourse*. Ithaca, NY: Cornell University Press, 1996.

[16] N. Deitelhoff and H. Müller, "Theoretical paradise: Empirically lost? arguing with habermas," *Review of International Studies*, vol. 31, no. 1, pp. 167–179, 2005.

[17] S. Chambers, "Measuring publicity's effect: Reconciling empirical research and normative theory," *Acta Politica*, vol. 40, no. 2, pp. 255–266, 2005.

[18] D. Naurin, "Most common when least important: Deliberation in the european union council of ministers," *British Journal of Political Science*, vol. 40, no. 1, pp. 31–50, 2010.

[19] M. R. Steenbergen, A. Bächtiger, M. Spörndli, and J. Steiner, "Measuring political deliberation: A discourse quality index," *Comparative European Politics*, vol. 1, no. 1, pp. 21–48, 2003.

[20] K. Holzinger, "Verhandeln statt argumentieren oder verhandeln durch argumentieren? eine empirische analyse auf der basis der sprechakttheorie," *Politische Vierteljahresschrift*, vol. 42, no. 3, pp. 414–446, 2001.

[21] ——, "Bargaining through arguing: An empirical analysis based on speech act theory," *Political Communication*, vol. 21, no. 2, pp. 195–222, 2004.

[22] C. Landwehr and K. Holzinger, "Institutional determinants of deliberative interaction," *European Political Science Review*, vol. 2, no. 3, pp. 373–400, 2010.

[23] J. J. Thomas and K. Cook, *Illuminating the Path: Research and Development Agenda for Visual Analytics*. IEEE-Press, 2005.

[24] D. Oelke, "Visual Document Analysis: Towards a Semantic Analysis of Large Document Collections," Ph.D. dissertation, Dissertation Thesis, University of Konstanz, 2010.

[25] C. Rohrdantz, M. C. Hao, U. Dayal, L.-E. Haug, and D. A. Keim, "Feature-based visual sentiment analysis of text document streams," *ACM TIST*, vol. 3, no. 2, p. 26, 2012.

[26] [Online]. Available: http://stuttgart21.wikiwam.de/index.php?title=Schlichtung\&oldid=5188

[27] D. Angus, B. Watson, A. Smith, C. Gallois, and J. Wiles, "Visualising Conversation Structure across Time: Insights into Effective Doctor-Patient Consultations," *PLoS ONE*, vol. 7, no. 6, 2012.

[28] R. Prasad, N. Dinesh, A. Lee, E. Miltsakaki, L. Robaldo, A. Joshi, and B. Webber, "The Penn Discourse TreeBank 2.0," in *Proceedings of LREC*, 2008.

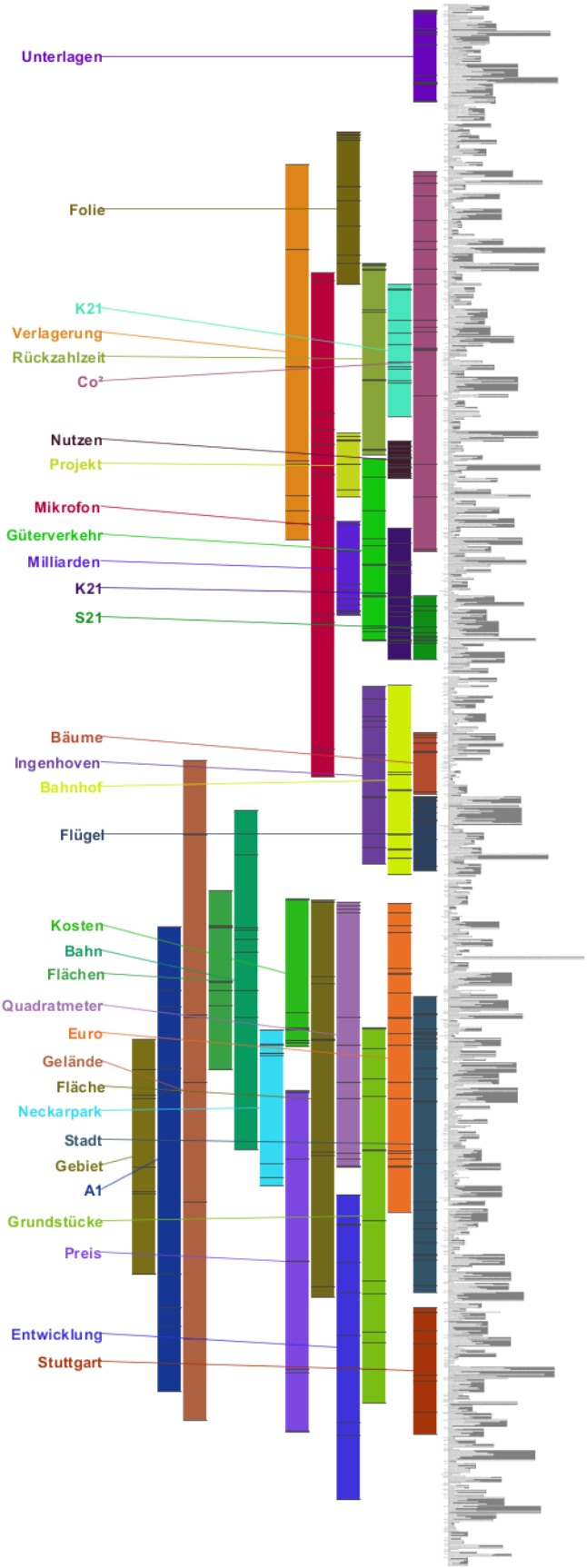


Fig. 1: Visualization of the lexical episodes for the Stuttgart 21 mediation on November 19th. It includes all lexical episodes having a length of at least 8 words which belong to nouns.

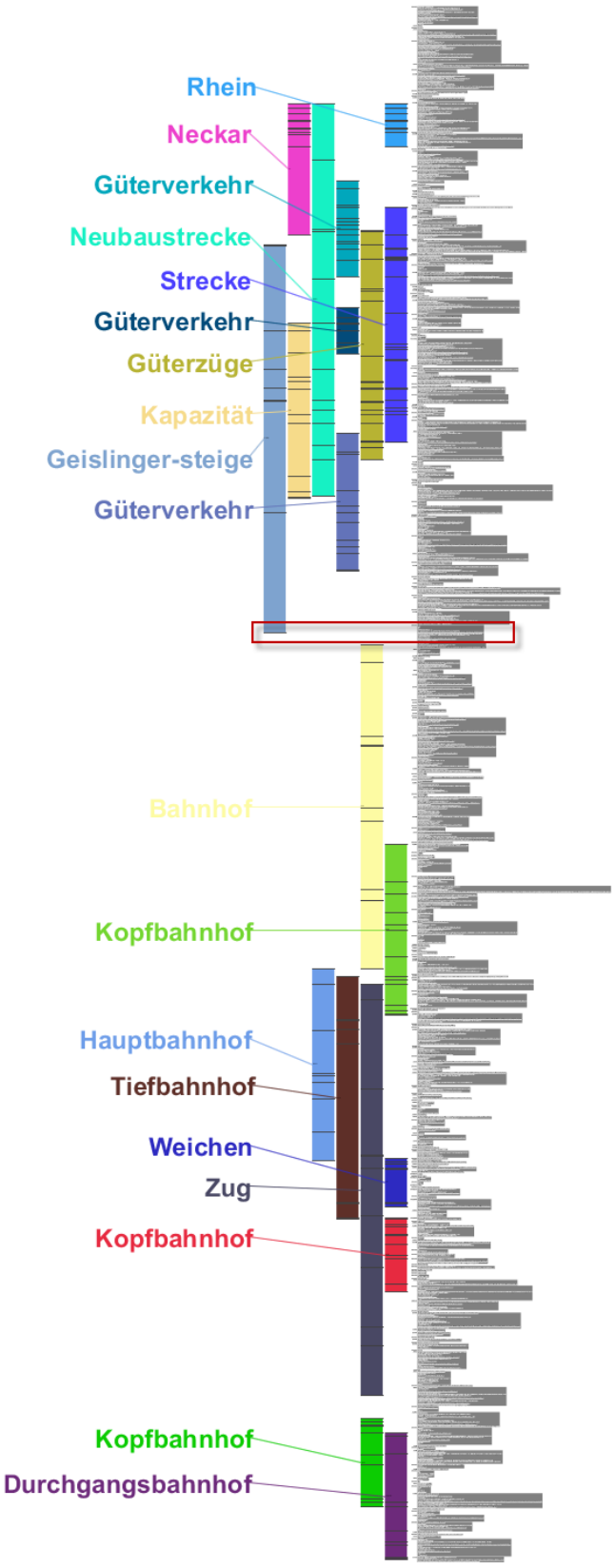


Fig. 2: Visualization of the lexical episodes for the Stuttgart 21 mediation on November 22nd. It includes all lexical episodes having a length of at least 8 words which belong to nouns.

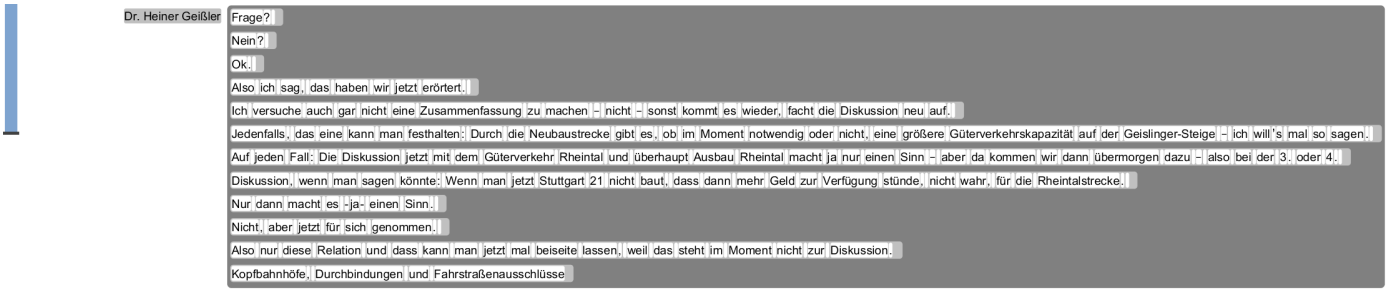
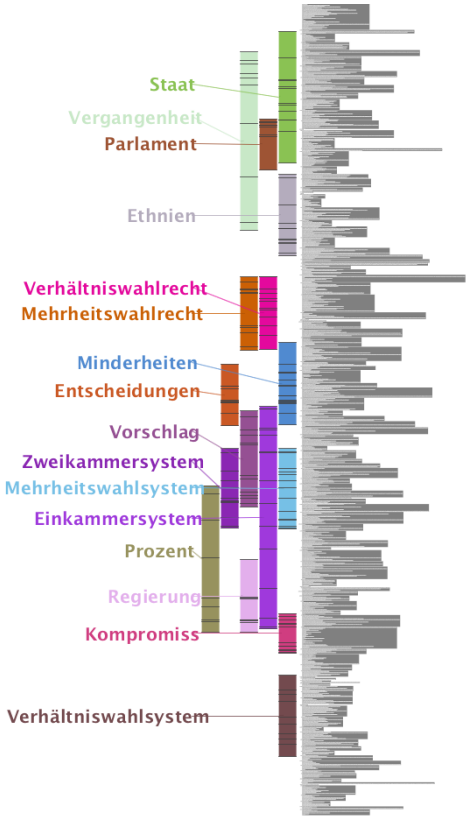
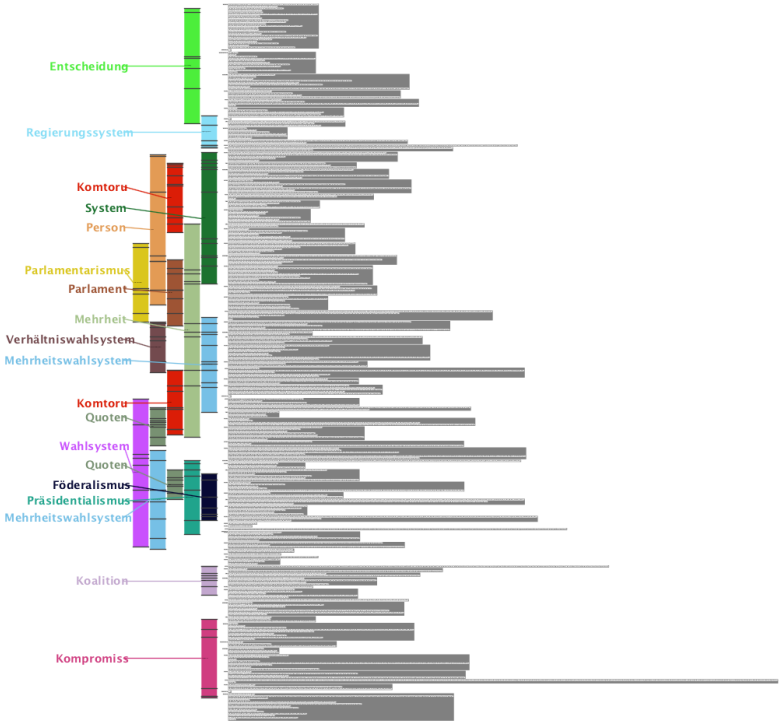


Fig. 3: Zoomed-in view on part of the visualization of the lexical episodes for the Stuttgart 21 mediation on November 22nd. It correponds to the part of the discourse presented in Figure 2, where the lexical episode of *Geislinger-Steige* ends and the lexical episode of *Bahnhof* begins.

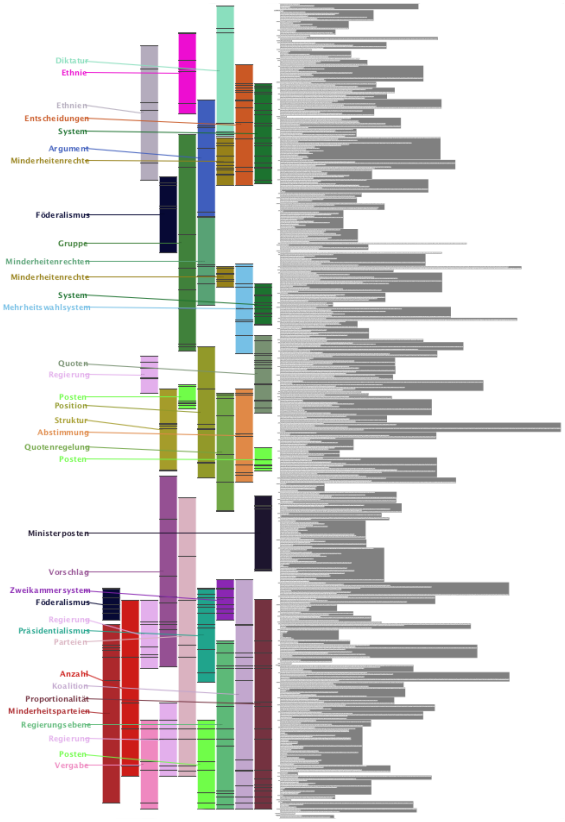
(a) Experiment 1



(b) Experiment 2



(c) Experiment 3



(d) Experiment 4

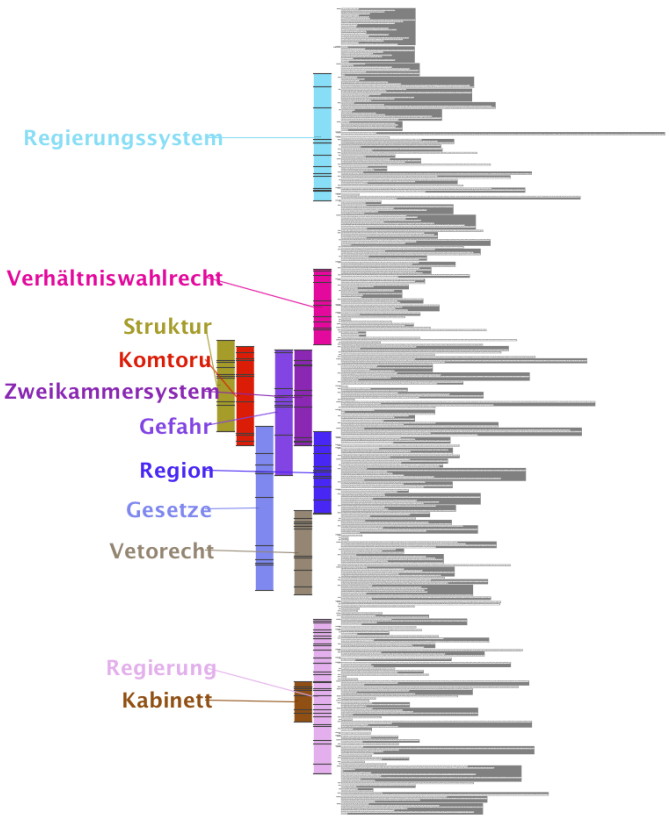


Fig. 4: Visualization of the lexical episodes for the simulation-gaming experiments. It includes all lexical episodes having a length of at least 8 words which belong to nouns.