Corpus Resources for Dispute Mediation Discourse

Mathilde Janier¹, Chris Reed¹
¹Centre for Argument Technology (ARG-tech)
University of Dundee, Scotland
m.janier@dundee.ac.uk, c.a.reed@dundee.ac.uk

Abstract
Dispute mediation is a growing activity in the resolution of conflicts, and more and more research emerge to enhance and better understand this (until recently) understudied practice. Corpus analyses are necessary to study discourse in this context; yet, little data is available, mainly because of its confidentiality principle. After proposing hints and avenues to acquire transcripts of mediation sessions, this paper presents the Dispute Mediation Corpus, which gathers annotated excerpts of mediation dialogues. Although developed as part of a project on argumentation, it is freely available and the text data can be used by anyone. This first-ever open corpus of mediation interactions can be of interest to scholars studying discourse, but also conflict resolution, argumentation, linguistics, communication, etc. We advocate for using and extending this resource that may be valuable to a large variety of domains of research, particularly those striving to enhance the study of the rapidly growing activity of dispute mediation.

Keywords: discourse, dispute mediation, open corpus

1. Introduction
Mediation is a rapidly growing practice among dispute resolution processes. The high costs and delays of traditional litigation lead people to prefer alternative dispute resolution (ADR) processes, and dispute mediation is becoming extremely popular, particularly in English-speaking countries¹. In different domains of research – such as sociology, linguistics or argumentation – an increasing number of academic publications focus on a better understanding of this growing practice and are therefore concerned with discourse in dispute mediation (see e.g. (Greco Morasso, 2011; Greatbatch and Dingwall, 1997; Tanaka et al., 2007; Stokoe, 2012; Hoffer, 1996), etc.). Academics, however, have difficulties in acquiring data to study discourse in mediation, in particular because of its confidentiality principle. This lack of resources is a challenge that leads us to advocate for an open corpus of mediation transcripts that would be valuable to research communities who strive to better understand this activity and try to make it more effective and more efficient. Such a corpus would be useful for diverse areas of research: conflict resolution, argumentation, linguistics, sociology, etc. It would allow for sharing transcripts of dialogues in this understudied context, and several different works of research would be made possible by building upon them. It will then be possible to compare, develop and expand previous studies. That, will ultimately lead to an extended knowledge of this growing domain.

Several corpora have been created to boost research in linguistics². They all are designed for different purposes and contain different data. For example, the Brown University Standard Corpus of Present-Day American English (or Brown corpus) (Kucera and Winthrop, 1967)³, one of the oldest corpora of natural language containing more than 100 million words from written and spoken texts, and the famous British National Corpus (BNC) (Leech, 1992)⁴ are intended for general use and present raw texts. The HCRC Map Task Corpus (Anderson et al., 1991)⁵, which comprises of 128 annotated dialogues, was built to support research in human communication, while other corpora such as the PennTree bank corpus, which presents linguistic trees (Marcus et al., 1993)⁶, or the AraucariaDB Corpus⁷ (Reed et al., 2008a; Lawrence et al., 2015), composed of argument analyses, contain already analyzed texts. These corpora, created for different uses, have supported a high number of works of research which were built upon them and have not been only useful to the persons who assembled them⁸.

We believe a corpus of mediation interactions would have the same impact on the community studying this practice. For this reason, we identify in Section 2. some sources of real and realistic data and present in Section 3. a newly created corpus of annotated mediation dialogues gathering data from many different sources, and most importantly, openly available for the purpose of supporting research in this growing activity.

2. Existing Sources
2.1. Academic Sources
Although understudied – compared to traditional litigation for example – a growing number of works of research has been concerned with discourse in mediation. Publications that rely on analyses of transcripts sometimes present extracts of dialogues – a transcript of an entire mediation

¹As an example, National Family Mediation, one of the largest mediation service in England and Wales conducted 16,000 mediations in 2012/2013
²Footnotes after the references that follow give the number of works that cite these publications, according to Google Scholar.
³cited by 7374
⁴The BNC Handbook and Users Reference guide have been cited over 1000 times
⁵cited by 905
⁶cited by 5509
⁷Accessed by over 3,000 unique users during 2015
⁸The numbers provided by Google Scholar may underestimate the total of works relying on the corpora but this gives an idea of their significance

1014
is, to our knowledge, never given. We list here some of the major publications where the authors used transcripts of mediations and mention their provenance. In (Greco Morasso, 2011; Greco Morasso, 2008; Greco Morasso, 2010), the corpus is constituted by transcripts of ‘exemplary interactions, from which mediators learn to mediate’ (Greco Morasso, 2011, p.148), and the publications show various passages of the transcripts. The transcripts come from video-recorded real mediation sessions that have been distributed worldwide to train mediators. The studies in (Stokoe, 2012) are based on transcripts of “200 intake calls to five different UK-based community mediation services” that were analyzed using conversation analysis. In (Jacobs and Aakhus, 2002) the authors base their study of mediators’ strategies on forty-one real mediation sessions, and present thirteen extracts. This small source of mediation data – due to the scarcity of the excerpts presented – has nevertheless the advantage of providing real and typical mediation dialogues. It can be used by researchers who may find the content of the excerpts valuable for their own project. As an example, in (Janier et al., 2014a), the authors analyze some passages taken from (Jacobs and Aakhus, 2002).

Using excerpts taken from academic works is the easiest way of obtaining mediation discourse data, and one can assume that the transcripts have been legally obtained, and have already proven to contain information suitable for analysis. The few works presented above have different goals so the exploitable information is different as well, but they can nevertheless be used for other research projects. Although the excerpts present authentic interactions, the absence of entire transcripts may be an issue for research that would focus on understanding the mediation process as a whole.

2.2. Online Sources

Another way of obtaining data concerning mediation discourse is to search resources online. Some websites present mediation scripts; they generally capture a small part of a mediation and are intended to training mediators or disputants willing to know how a ‘standard’ mediation unfolds, e.g. a guide for training mediators\(^9\), the script of the beginning of a mediation session\(^10\), or the typical introduction to a mediation\(^11\).

Even though still rare, another relevant source of data when searching for ‘mediation transcript’ or ‘transcripts mediation sessions’ are videos of mock mediation, ranging from small excerpts (of more or less ten minutes)\(^12\) to complete sessions\(^13\). Having such videos transcribed is a quick and easy way of getting data for the study of mediation interactions; however, role-plays may not suit all types of research: it is understandable that one may not rely on the authenticity of the dialogues in videos of mock mediations.

2.3. Professional Sources

Role-plays or mock mediations can also be acquired through mediation services, who are keen on sharing them than genuine mediation sessions. Although they do not present real disputes, we can suppose role-plays provide realistic data because they are generally used to train mediators. As an example, the research project led by the authors of this paper in the Centre for Argument Technology (ARG-tech)\(^14\) at the University of Dundee, primarily relied on a transcript of a mock mediation provided by the Early Dispute Resolution (edr) center in Dundee\(^15\). This corpus was used in several works (Janier and Reed, 2015; Janier et al., 2015) for the study of the argumentative activity in mediation.

Another interesting track to follow is to discuss with mediation professionals. As an example, ARG-tech organized a workshop\(^16\) with mediation professionals and researchers who have shown to be ready to share transcripts and videos in order to facilitate our research project; we have thus been provided with a real mediation transcript, some excerpts of which have been analyzed (see Section 3.). Depending on the type of research the data will be used for, it is sometimes important to emphasize that the transcripts can be anonymized.

This source of data has two advantages: one can be assured that the transcripts of real sessions provided by mediation services contain authentic interactions. As to role-plays, although they may seem less genuine, they represent typical and standard interactions. The transcripts, moreover, may capture the entire sessions, which rarely happens (see Sections 2.1. and 2.2.).

3. The Dispute Mediation Corpus to Support Research in Argumentation

The different sources to obtain data for mediation discourse presented in Section 2. all have their advantages and drawbacks and none can, alone, be fully satisfactory. For this reason, a corpus gathering data from the different sources would not only make it easier finding data but would also allow for making available other sources. The Centre for Argument Technology has worked in that direction and created the Dispute Mediation Corpus (DMC) as part of a research project on argumentation, available at arg.tech/DMC. It comprises of more than 100 annotated mediation excerpts. The annotations, carried out using the Inference Anchoring Theory (IAT) (Budzynska and Reed, 2011; Budzynska et al., 2014), elicit the dialogical and argumentative structures of the interactions. IAT is a theoretically grounded counter part of the Argument Interchange Format (AIF), a framework developed in response to the increasing number of argument theories and argument analyses that recommends a standardized representation of argument maps (Reed et al., 2008b). Analyses in the Argument Interchange Format are stored in the AIFdb database (Lawrence et al., 2012) to comply with what the AIF advocates, namely making argument analyses available and ex-

---

\(^9\) e.g. arg.tech/mediation-toolkit

\(^10\) e.g. arg.tech/mediationscript-beginning

\(^11\) e.g. arg.tech/mediation-intro

\(^12\) e.g. arg.tech/mediationvideo-summary

\(^13\) e.g. arg.tech/mockmediation

\(^14\) arg-tech.org

\(^15\) dundee.ac.uk/academic/edr/

\(^16\) mediationworkshopdundee.wordpress.com
changeable through a large variety of computational tools. The DMC has been analyzed using the Online Visualization of Arguments tool – OVA+ – (Janier et al., 2014b) and is stored in the AIFDb Corpora\textsuperscript{17} platform (Lawrence and Reed, 2014; Lawrence et al., 2015), an interface which further meets AIF recommendations by making publicly available and exchangeable argument maps.

### 3.1 Introduction to Inference Anchoring Theory

Although argument analyses that compose the DMC may not be of interest to everybody, it can be useful to understand the annotations. Here is a short introduction to Inference Anchoring Theory (IAT), the framework used to carry out the argument analyses.

IAT allows for eliciting the argument structure of texts in dialogical contexts. The argument structure is extracted from the representation of the dialogical and illocutionary structures, and allows for showing how dialogical moves create arguments in natural language thanks to a graphical representation. In the DMC, IAT analyses are carried out in OVA+. Let’s take a made up example to explain IAT.

\begin{enumerate}
\item a. Speaker 1: \textit{Don’t you think that an open corpus of mediation interactions is necessary?}
\item b. Speaker 2: \textit{Why would it be necessary?}
\item c. Speaker 1: \textit{Scholars of many disciplines could use it for their own research.}
\end{enumerate}

In Example 1 two speakers are talking about the usefulness of an open corpus of mediation interactions. In 1a, Speaker 1 poses a question, to which Speaker 2, in 1b, replies with another question, and in 1c, the first speaker replies with an assertion. In this example, we feel that Speaker 1 is arguing in favor of the open corpus although no linguistic cues (such as ‘because’ or ‘therefore’) allow us to prove it. Speaker 1’s argument can in fact be reconstructed by the sequence of locutions that form this dialogue. Let’s represent this in Figure 1 using the annotation scheme provided by IAT.

On the right-hand side the dialogical structure is represented with: (i) the sequence of locutions with the corresponding speaker’s identification and, (ii) the transitions between locutions. In IAT, transitions (represented by Default Transition nodes) correspond broadly to the rules of the dialogue. These are not logical relationships: they represent dialogical relevance rather than topical relevance. Here, there is a transition between the first two locutions, because Speaker 2’s locution is a reply to Speaker 1’s location; the second transition node also shows that the third locution is a reply to the second.

The left-hand side of the graph represents the argument structure with: (i) the propositional contents of the locutions and, (ii) the relations between these propositional contents. Relations between propositional contents in IAT can be of 3 types: inference, conflict or rephrase. Here, the relation between the third and the first propositional contents is of type inference, which means that the third propositional content acts as a support for the first\textsuperscript{18}. Indeed, in this dialogue, we can reconstruct Speaker 1’s position as: “An open corpus of mediation interactions is necessary because scholars of many disciplines could use it for their own research.” In IAT, conflict nodes are used to show that some propositional content is a sort of contradiction of another; and rephrase is used to represent propositional contents paraphrasing or reframing a previous propositional content. Inferences, conflicts and rephrases can be specified in IAT: for instance, an inference can be of type argument from position to know (Walton et al., 2008), however, in the DMC we only used generic nominations (i.e. Default Inference, Default Conflict and Default Rephrase).

In IAT, the argument structure can only be extracted via the representation of the illocutionary structure (i.e. the graphical elements between dialogical structures and argument structures). The illocutionary structure is composed of: (i) illocutionary connections between locutions and propositional contents and, (ii) illocutionary connections anchored in transitions. Illocutionary connections anchored in locutions represent the illocutionary forces (essentially inspired by (Searle and Vanderveken, 1985)) of the locutions. Here, the first locution is, according to IAT, an assertive question: although it takes the form of a question, the speaker gives it an assertive force. The second locution, in turn, is a pure challenge: Speaker 2 asks about the grounds for Speaker 1 believing 1a. The third locution, finally, bears the force of an assertion: Speaker 1 is stating something. The second type of illocutionary connections (the ones anchored in transitions) show the core argumentative functions of the dialogues: in Example 1, it is because Speaker 1 was challenged by Speaker 2 that she provided a premise to her first utterance (i.e. the assertive question), and therefore created an argument (shown by the arguing node). IAT graphical representations such as the one in Figure 1 therefore show how dialogues create arguments by eliciting relationships between locutions and propositional contents, and between dialogue rules and argument structures.

Let’s now take Example 2, a dialogue available in the DMC\textsuperscript{19}.

\begin{enumerate}
\item a. Eric: \ldots it’s just making my life a misery, actually and that’s the way things are at the moment.
\item b. Viv: I’m sorry.
\item c. Eric: I’m sorry if that’s the way it comes over, but you want me to be honest, so that’s the view and that’s kind of the way it is.
\item d. Mildred: What would you like out of today?
\item e. Eric: Well, I would like Viv to become part, a productive member of the team. Because we are a team and, you know, Viv was appointed to be my equal…
\end{enumerate}

In this excerpt, Mildred is the mediator, and Eric and Viv are the disputants. The IAT analysis of this example is shown in Figure 2. This analysis was presented in (Janier and Reed, 2015) to highlight the argument structure of a particular mediation strategy: redirection.

\textsuperscript{17}corpora.aifdb.org

\textsuperscript{18}Directed arrows show which propositional content is the premise and which is the conclusion.

\textsuperscript{19}Argument map # 6688, available at corpora.aifdb.org/mockmediation
The redirecting strategy is elicited here by the absence of Default Transition node between Eric and Viv’s discussion at the beginning of the example and Mildred’s intervention (between 2c and 2d). In IAT, the absence of such a node means that there is no relationship between two locutions. In this example, Mildred has detected that the parties’ discussion was leading nowhere (see from 2a to 2c that Viv is disagreeing, but Eric does not argue to justify his claim), so she poses a question (Pure Questioning node) that does not relate to the discussion between the parties. Afterwards, Eric answers to Mildred’s question and even argue without having been asked to. The analysis shows the impasse (the parties do not manage to have a reasonable discussion), the mediator’s strategy i.e. redirecting the discussion (shown by the absence of Default Transition node) along with the overall dialogical and argumentative structures.

3.2. The Dispute Mediation Corpus: Some Details

The DMC has been created as part of a recent project which aims at exploring argumentation in mediation. This corpus of analyses has been annotated by a unique analyst, however, IAT has been developed and shown stable in another project which focuses on dialogical interactions in the context of radio debates (see e.g. (Yaskorska and Janier, 2015)), and the results of the annotations gave an inter-annotator agreement reaching $\kappa = .68$ (Janier et al., under review). Example 2 and its analysis in Figure 2 give a flavor of what can be found in the DMC, but many other mediation features can be found in the corpus. We shall describe now some of its characteristics. Though still relatively small, this resource contains a large number of different data, summarized in the following table. Apart from the category ‘words’, the elements reported in the table have to be understood according to IAT definitions. A total of 2,805 locutions (with an average of 12.02 words) have been annotated, of which 1,545 are assertions and 248 are questions. Given that its current use is for the study of argumentation, we can also report more detailed and precise numbers: for instance, 590 schemes of inference and 202 schemes of conflict (roughly ‘arguments’ and ‘disagreements’, respectively) have been identified. These numbers show that the corpus and the annotation framework allow for extracting many different data on mediation discourse. For now, the corpus only contains texts in English, but it is conceivable to add excerpts of any language.

<table>
<thead>
<tr>
<th>Elements type</th>
<th>Occurrence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Words</td>
<td>18,628</td>
</tr>
<tr>
<td>Locutions</td>
<td>2,805</td>
</tr>
<tr>
<td>Assertions</td>
<td>1,545</td>
</tr>
<tr>
<td>Assertive Questions</td>
<td>76</td>
</tr>
<tr>
<td>Pure Questions</td>
<td>141</td>
</tr>
<tr>
<td>Rhetorical Questions</td>
<td>31</td>
</tr>
<tr>
<td>Assertive Challenges</td>
<td>5</td>
</tr>
<tr>
<td>Pure Challenges</td>
<td>11</td>
</tr>
<tr>
<td>Popular Concessions</td>
<td>18</td>
</tr>
<tr>
<td>Inferences</td>
<td>590</td>
</tr>
<tr>
<td>Conflicts</td>
<td>202</td>
</tr>
<tr>
<td>Rephrases</td>
<td>187</td>
</tr>
</tbody>
</table>

Table 1: Details of the DMC

The DMC is currently composed of 129 analyses of excerpts divided into six sub-corpora, according to the focus of the argument analyses:

- The sub-corpus Dispute mediation: excerpts taken from publications leaves the sub-corpora handle many different languages, such as

\[20\text{AlFdb} \text{corpora handle many different languages, such as Ukrainian, French or Hindi} \]

\[21\text{corpora.alfdb.org/mediationothers} \]
Figure 2: IAT analysis of Example 2, realized in OVA+

2002) (see Section 2.1.). It was mainly used as a preliminary step towards the development of a theory for the analysis of argumentation in mediation. The excerpts all come from real mediation sessions.

- The Mock mediation sub-corpus\(^{22}\) comprises 29 analyses from two role-plays, one provided by the edr Center (see Section 2.3.), the other transcribed from a video found online (see Section 2.2.). It has been mainly used to to support the findings in (Janier and Reed, 2015).

- The Critical discussion\(^{23}\), Bargaining\(^{24}\) and Therapeutic\(^{25}\) sub-corpora (14 analyses) were created for a project with Rutgers University and aims at comparing the dialogical and argumentative patterns of three types of discussions that can occur in mediation (Janier et al., 2014a).

- The Meta-talk in mediation sub-corpus\(^{26}\) (28 analyses) was created to explore meta-discourse elements in mediation interactions coming from all the various excerpts mentioned above.

3.3. Using the DMC

The DMC resources have been used in several works about argumentation in mediation. In (Janier and Reed, 2015), the Mock mediation corpus was used to present a method to analyze argumentative discourse; in (Janier et al., 2015), excerpts of this same corpus were used to analyze impasse and strategies; in (Janier et al., 2014a), we used the Critical discussion and Therapeutic corpora to show the argumentative and dialogical differences between two types of discussions, etc.

The DMC is openly available at arg.tech/DMC where both the original text of the dialogues and the annotations can be consulted, shared and exploited by everyone. Figure 3 shows the DMC webpage, where each already analyzed excerpts is stored under an ID number. On the left of the page, one can see extracts of the annotated texts. On the

\(^{22}\)corpora.aifdb.org/mockmediation
\(^{23}\)corpora.aifdb.org/critical
\(^{24}\)corpora.aifdb.org/bargain
\(^{25}\)corpora.aifdb.org/therapeutic
\(^{26}\)corpora.aifdb.org/metatalk
right, there are overviews of the argument analyses. Each argument analysis can be downloaded in several formats (e.g., .png, .json, .pl). To access a complete argument analysis, one can click on the OVA+ link: a window opens up with the IAT graphical analysis of the excerpt, as shown in Figure 4\textsuperscript{27}. A mere copying/pasting on the left of the OVA+ webpage allows for obtaining the original text of the analysis. A whole mock mediation transcript is also available by downloading the zip-file corresponding to the Mock mediation corpus on the AIFdb Corpora webpage.

4. Conclusion

The growing public interest in mediation has led to an increasing number of publications from different domains focusing on discourse in this context. To allow for developing research further, it is crucial to have reliable data to study. Although there is a lack of real transcripts (i.e., there is a large preponderance of role-plays), resources for mediation can be obtained in several ways, e.g., using previous research, looking for scripts online, discussing with practitioners, etc. A repository of mediation transcripts would be valuable for the research community to share and (re)use data for mediation discourse. For this reason, ARG-tech has created the publicly available DMC which currently comprises of 129 extracts from different sources (mock and real mediation transcripts, excerpts taken from academic publications, etc.). Although the corpus is composed of argument analyses, the original texts (i.e., raw texts) are stored in a database, therefore, its use is not limited to research in argumentation and can as well support works in other domains (e.g., sociology, linguistic, communication etc.), which will lead to a better understanding of this growing activity.

Extending and sharing this resource will facilitate access to mediation transcripts, and therefore allow for more studies to be carried out. Several areas of research can take advantage of such a corpus. For example, computational linguistics and machine learning techniques, in particular, provide opportunities where this corpus may be processed to support research or to implement a tool; argumentation theorists may find it in its resource for the study of arguments in dialogical contexts; some works in pragmatics could also use it for the various linguistic contexts it contains. These are only a few examples of possible uses, but it makes no doubt that research in natural language and conflict resolution in a broad sense will benefit from an expanded sharing of the DMC.

5. Acknowledgements

We gratefully acknowledge the support of the British Leverhulme Trust under grant RPG-2013-076.

6. Bibliographical References


\textsuperscript{27}arg.tech/5907


