QTMM2012c+: A Queryable Empirically-grounded Resource of Dialogue with Argumentation

Jacopo Amidei[1], Paul Piwek[2] and Svetlana Stoyanchev[3]

[1]The Open University, Milton Keynes, UK
[2]The Open University, Milton Keynes, UK
[3]Toshiba Cambridge Research Laboratory, Cambridge, UK

Abstract

This paper introduces QTMM2012c+, a resource which links relations between propositions (inference, conflict and rephrase) to dialogue act sequences. QTMM2012c+ builds on the MM2012c annotated corpus of BBC Moral Maze debates, extending it with new annotations – for speaker roles (chair, panellists and witnesses), speaker stances (neutral, pro and con) and locution chronological ordering – and making the information available in a queryable format. We show how the new resource allows for: i) automatic extraction of empirically-grounded dialogue rules which describe choice and frequency of dialogue acts with specific argumentative functions given the dialogue history, and ii) extraction of generation templates that reflect naturally-occurring argumentative locutions in empirically-grounded dialogue. QTMM2012c+ facilitates automatic analysis of argument transitions between speakers, extending previous manual analysis of the MM2012c corpus, enabling empirical tests of theories of argumentative dialogue.

Keywords

Dialogue based on argumentation, Argumentation in agent and multi-agent systems, Argument-based machine learning, Strategies in argumentation, Argumentation schemes, Dialogue rules, MM2012c dataset

1. Introduction

Most recent work on dialogue modelling and systems is empirical in nature, aiming to model natural task-oriented dialogue (e.g. [1], [2], [3] and [4]), chat (e.g. [5], [6] and [7]) or a mixture of the two (e.g. [8] and [9]). In contrast, research on argumentative dialogues has typically focused on the specification of normatively correct rules for dialogue, going back to the work of Hamblin [10], which aimed to characterise certain fallacious reasoning patterns as violations against such normative dialogue rules.

Hunter [11] argues that this strictly normative orientation is too restrictive for work on computational modelling of persuasive argumentative dialogue. Our interest is not so much in purely persuasive dialogue, but rather in dialogue that explores, via argument, different points of view. Nevertheless the point remains that a strictly normative perspective is inadequate and empirical modelling of natural argumentation in dialogue is called for, if the aim is to eventually build systems that meet human standards of naturalness and coherence.

5th Workshop on Advances In Argumentation In Artificial Intelligence (AI^3 2021)
jacopo.amidei1@open.ac.uk (J. Amidei); paul.piwek@open.ac.uk (P. Piwek);
svetlana.stoyanchev@crl.toshiba.co.uk (S. Stoyanchev)

A key challenge for this strand of work is to uncover how relations between propositions (e.g. inference, conflict and rephrase) can be translated into moves in argumentative dialogue. In this paper, we introduce QTMM2012c+, a resource that describes such translations based on a real argumentative conversation.

After an introduction and a first analysis of QTMM2012c+ (Section 3 and Section 4), we will present a description of two use cases scenario (Section 5). The examples will shows that QTMM2012c+ can be used to define argumentation strategies or as a training resource for multi-agent dialogue systems (for example statistical systems).

2. Related Work

Our work is based on the Moral Maze (MM2012) dataset [12, 13, 14]. This dataset has been used to define rule-based models for: i) automatic extraction of argument from dialogues [15, 16], and ii) automatic detection of discourse units and illocutionary structure from dialogues [17, 18]. QTMM2012c+, which specifically extends MM2012c, is in the form of a Queryable Table (hence the QT prefix) and has been enhanced with new annotations (indicated with the + post-fix).

Yaskorska-Shah [19] uses the Moral Maze dataset to define transition schemes that map propositional relations to dialogue acts, which in turn can be used to formulate dialogue rules. As a use case, we show how QTMM2012c+ allows us to automatically extract dialogue rules. As we will show in Section 5.1, our work extends that of Yaskorska-Shah in terms of (1) method (by using automatic rather than manual harvesting of schemes), (2) scale (significantly more data covered) and (3) content (we introduce new annotations that allow us to ground our schemes in dialogue speaker roles and speaker stances).

Stoyanchev and Piwek [20] developed a mapping between text segments and dialogue acts on a smaller corpus of human-authored dialogue, however they focused on discourse rather than propositional relations in text.

3. The argumentation dataset

For an in-depth description (including reliability study and dataset statistics) of the MM2012c dataset, which our work builds on, see [14].

MM2012c is composed of five episodes, all from 2012, of the Moral Maze BBC Radio 4 show. The Moral Maze format involves a chair/moderator, four panellists and four witnesses. They discuss the moral implications of current topics. In the discussion, two panellists and two witnesses take a position in favour of the topic under discussion and the others take positions against it. An episode proceeds in ‘rounds’ where two of the panellists interrogate a witness (whose opinion is opposite to that of the panellists).

The MM2012c dataset is annotated following the Inference Anchoring Theory (IAT) [21]. An example of an annotated fragment of dialogue is shown in Figure 1. On the right-hand side, we can see two *locutions*, ‘MT : Don’t you think there’s a streak that says, it was a pretty balanced account, there’s nothing wrong with colonialism.’ (with speaker MT) and ‘AM : I don’t think people do think that’ (with speaker AM). They are linked by a *transition* box, which signifies that the second locution is a reply or response to its predecessor.

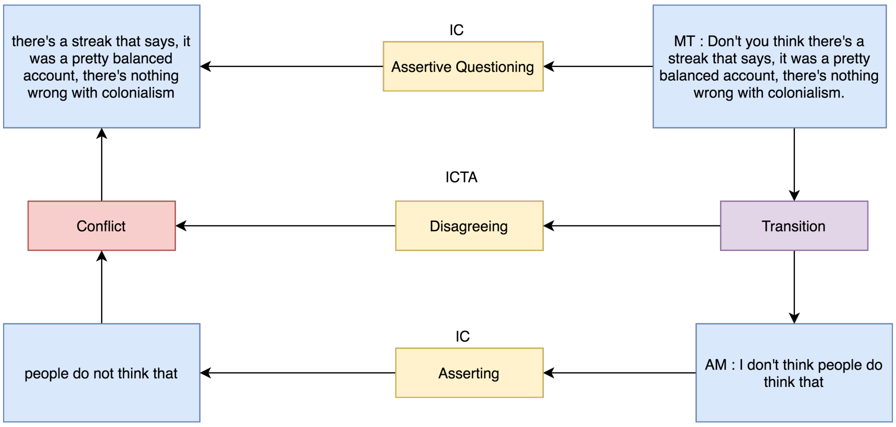

Figure 1: Example from Moral Maze British Empire, Map 6101 (http://ova.arg-tech.org/analyse.php?url=local&plus=true&aifdb=6101&akey=d06e4fa24e372e444586295ad192b085). IC refer to the Illocutionary connections associated to locutions and ICTA refer to the Illocutionary connections associated with the transitions.

Each locution is anchored to a *proposition* (shown in the two left-most blue boxes) via an *Illocutionary connection* (IC) (in the middle yellow boxes). These represent the propositional content and Illocutionary force from Speech Act theory [22].[1] In this case, the first locution's Illocutionary connection is 'Assertive Questioning' and the second 'Asserting'. Annotators were instructed to express the propositions as complete declarative sentences, removing ellipsis and reconstructing anaphoric references.

Finally, each transition between locutions is anchored, via an Illocutionary connection, to a *Propositional relation*. In this paper we refer to the Illocutionary connections associated with the transitions as *ICTA*. In this instance, Transition is anchored via the 'Disagreeing' Illocutionary connection to the Conflict Propositional relation. Apart from the *Conflict* relation between propositions, IAT singles out *Inference* (when one proposition is used to provide a reason to accept the other) and *Rephrase* (when one proposition is more or less a paraphrase of the other).[2]

This illustrates how IAT captures: 1) dialogue internal relations (via Transitions), 2) relations between the propositions expressed in the dialogue (via Propositional relations) and 3) the links between the two (i.e. between dialogue and propositions) via Illocutionary connections.

3.1. From MM2012c to QTMM2012c+: data cleaning and augmentation

The QTMM2012c+ Queryable Table[3] was constructed from the MM2012c dataset in two steps, during which all 1747 Transitions in MM2012c were processed.

(**1**) We started with a semi-automatic step. We developed an algorithm that extracted the transitions one map at a time following the dialogue flow (note that MM2012c is divided into maps which each covering part of the dialogue in an episode). We then manually checked

[1] The illocutionary force expresses the speaker's intention in producing an utterance.

[2] IAT has further subdivisions of these, but we ignore them for the purpose of this paper.

[3] QTMM2012c+ can be downloaded from https://github.com/jacopoamidei/Supplementary-material-A-Queryable-Empirically-grounded-Resource-of-Dialogue-with-Argumentation.

the correctness of the transitions the algorithm extracted against the original data and where needed corrected any errors introduced by the algorithm.
(**2**) In the second step we enriched the MM2012c with new/implicit information. (**2a**) Following Yaskorska-Shah [19], we enriched the MM2012c data set by adding *"Yes"* or *"No"* for those locutions not linked to any proposition and which disagreed or agreed with other locutions in the same transition. More precisely, in the MM2012c dataset, locutions that express disagreement or agreement elliptically (e.g., with locutions such as "That's true", "Yes", "I disagree", etc.) are not linked to propositional content. They are only indirectly linked to the proposition they respond to. To make these easier to process (without changing the actual information), we link them to "Yes" or "No". For example, Figure 2 shows a case where the second locution has been associated

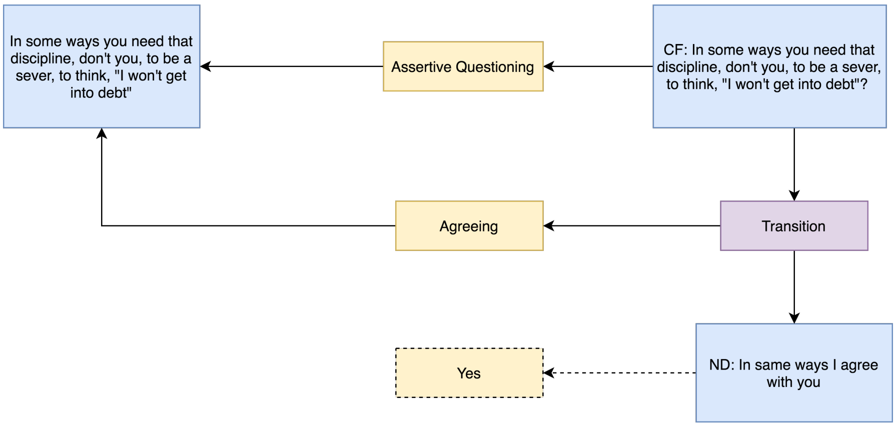

Figure 2: Example from Moral Maze Morality of Money, Map 6155 (http://ova.arg-tech.org/analyse.php?url=local&plus=true&aifdb=6155&akey=85209ec3fc0c83a0b637400bd1879fd7). The dashed box and the dashed arrow are added in QTMM2012c+.

with the IC "Yes" (dotted lines signal our addition to the original representation). Similarly,

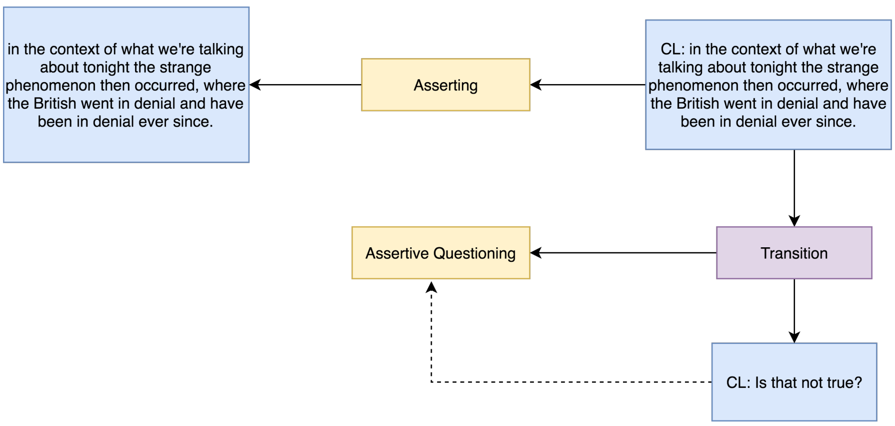

Figure 3: Example from Moral Maze British Empire, Map 6070 (http://ova.arg-tech.org/analyse.php?url=local&plus=true&aifdb=6070&akey=555352c8b250dfa1f6c57ba63d800869). The dashed arrow is added in QTMM2012c+.

for cases such as the one in Figure 3, the locution which is not linked to any proposition ('CL: Is that not true?'), has been linked to the illocutionary connection of its Transition.[4] (**2b**) We labelled the Chair as *Neutral* and the non-chair speakers as either *Pro* or *Con*, depending on whether they were for or against the main claim under discussion. The three authors of this paper independently annotated for *Pro/Con* and agreed 100% except for the episode on Banking System which had more than one claim under discussion. For this episode two claims can be identified: A) We need more rules to make the banks trustworthy and B) The banks' behaviour is immoral. After discussing which of the claims to designate as the main claim, resulting in selection of Claim A, we reached perfect agreement on *Pro/Con*. (**2c**) MM2012c only has a partial record of the order in which locutions occurred; we used the Moral Maze Transcripts [23] for associating with each locution a number (ID dialogue flow) representing its dialogue position. (**2d**) Using Moral Maze episode descriptions[5], we labelled each speaker with one of the following roles: Chair, Panellist or Witness.

4. Description and analysis of QTMM2012c+

Figure 4 shows the first three rows of QTMM2012c+.[6] The *ID* column stores a unique ID for each transition. The *ID map* column stores the ID of the map from which the transition comes. Similarly, the *Dataset* column stores the episode name from which a transition came. The *Reported Speech* column records the presence of reported speech in the transition (0 no reported speech; 1 for reported speech).[7] The *Speakers* column stores two values: Different and Same, which means that the locutions in the transition were uttered by different (same) speakers.[8] The *Degree* column stores the number of locutions involved in the transition. The *ICTA* column stores the Illocutionary connections associated with the transition. The *PropRel* column stores the propositional relations associated with the transition.

	ID	ID map	Dataset	Reported Speech	Speakers	Degree	ICTA	PropRel	L1	L1 IC	...	NOT PL2Reported Speech	NOT P1	Out of TA	To PL1Reported Speech	To PL2Reported Speech
0	1	6062	British Empire	0	Different	2	Asserting	N_A	MB : From your point of view, even if this hea...	Pure Questioning	...	0	0	0	0	0
1	2	6062	British Empire	0	Same	2	Arguing	Inference	ES : Most definitely	Yes	...	0	0	1	0	0
2	3	6062	British Empire	0	Same	3	Arguing	Inference	ES : In fact, some of us would argue the case ...	Popular Conceding	...	0	0	0	0	0

Figure 4: The first three columns of the QTMM2012c+ Queryable Table. Some columns start with 'NOT' for cases where a locution's proposition contains a negation

[4]Note that we did not remove the illocutionary connection linked to the transition itself. We just add that connection also to the locution. Thus, there is no loss of information.

[5]See https://www.bbc.co.uk/programmes/b006qk11/episodes/player.

[6]The columns name are listed in Appendix B.

[7]Reported Speech is equivalent to the Attribution discourse relation in Rhetorical Structure Theory (RST) [24].

[8]All transitions involve either one (Same) or two (Different) speakers.

The columns *Li* (for $1 \leq i \leq 7$) store up to 7 locutions from the transition (unused columns are populated with *N_A*). For each locution *Li*, QTMM2012c+ has a column about the locution IC (*Li IC*), the role of the speaker that uttered that locution (*Li Role*) – the possible values are: Chair, Witness, Panellist –, the stance of the speaker that uttered that locution (*Li Stance*) – the possible values are: Neutral, Pro, Con –, the order in which the locutions were uttered in the dialogue (*Li ID dialogue flow*) and the unique ID associated with that locution (*Li ID*).

The columns *LiReported Speech* (for $1 \leq i \leq 4$) store the reported speech used in the transition. If reported speech is not used in the transition, the column value is 0. For each reported speech *LiReported Speech*, QTMM2012c+ has a column about the reported speech IC (*LiReported Speech IC*), the role of the speaker that uttered that reported speech (*LiReported Speech Role*) – the possible values are: Chair, Witness, Panellist –, the stance of the speaker who uttered that reported speech (*LiReported Speech Stance*) – the possible values are: Neutral, Pro, Con –, and the unique ID associate to that reported speech (*LiReported Speech ID*).

The columns *Pi* (for $1 \leq i \leq 7$) store the propositions used in the transition. If a proposition is not used in the transition the column value is 0. For each proposition *Pi*, QTMM2012c+ has a column with a unique ID for that proposition (*Li ID).*

The columns *PLiReported Speech* (for $1 \leq i \leq 4$) store the proposition of the reported speech used in the transition. If a proposition of the reported speech is not used in the transition the column value is 0. For each reported speech *PLiReported Speech* QTMM2012c+ has a column for the unique ID associated with the proposition of the reported speech (*PLiReported Speech ID*).

The columns from *"L1Reported Speech to L2Reported Speech"* to *"TA is nonanchoring"* store information about the argument flow in the dataset. The argument flow describes the direction of the argumentation. For example the argument flow of Figure 1 goes from proposition 2 ('people do not think that') to proposition 1 ('there's a streak that says, it was a pretty balanced account, there's nothing wrong with colonialism'). In this case, QTMM2012c+ associates the value 1 with the column *P2 to P1* and the value 0 to all the columns representing alternative argument flows.

The value *N_A* means the information required for the column is not applicable to the transition. For example, in a transition of degree 2, all the columns *Li* (for $3 \leq i \leq 7$) get the value *N_A* .

4.1. QTMM2012c+ Analysis

Inference is the most used propositional relation in QTMM2012c+. It is used in 49% of transitions. *Conflict* is used in 12% of transitions, and *Rephrase* is used in 9% of transitions. In the remaining transitions (30%), the propositional relation is missing (see for example Figure 5). *Arguing* is the ICTA most frequently associated with *Inference* (99%), *Disagreeing* is the only ICTA associated with *Conflict*, and *Rephrase* is mostly associated with the ICTA *Default Illocuting* (95%). When a transition has no propositional relation, in the majority of the cases, the transition is not annotated with any illocutionary connection (74%). In some cases, the illocutionary connection of the transition is directly linked to a proposition (rather than via a propositional relation).[9] When this happens, 80% of the time the ICTA used is *Agreeing*. For more detail, see Appendix A Table 1.

[9]Figure 3 is an example of these cases.

In QTMM2012c+ most transitions have a low degree, i.e. 90% are degree 2 (that is, transitions that link together two locutions) and 8% are degree 3. *Conflict* is 94% of the time with degree 2 transitions, and *Rephrase* 92% of the time. Instances of *Inference*, although used mainly with transitions of degree 2, can also be found with higher degrees (17%). More detail can be found in Appendix A Table 2.

We found that the transitions with a single same speaker are used more than those with multiple different speakers: same speaker transitions make up 71.6% of the transitions (i.e. 1250 transitions), whereas the different speaker (multi-speaker) transitions make up 28.4% (496 transitions). More detail can be found in Appendix A Table 3.

Reported speech is used in 12% of the transitions.

Checking the number of ICTA per degree, we found that *Arguing* is the most used ICTA (49%) and it is the only one used with degrees 5, 6 and 7. Considering only the transitions that have an ICTA, the majority of ICTA are used in transitions of degree 2 (87%). More detail can be found in Appendix A Table 4. Regarding the use of Illocutionary connections (IC) that links locution to proposition, we found that *Asserting* is the most used IC (78% of the times). The other most used IC are questions (13% consisting of: 3% *Pure Questioning*, 6% *Assertive Questioning* and 4% *Rhetorical Questioning*). For more detail, see Appendix A Table 5.

Looking at the relation between IC and speaker roles, we found that the majority of the dialogue plays out between the *Panellists* and the *Witness*, with only a minor role for the *Chair*.

The *Chair* mainly poses questions. Indeed, the *Chair* uses IC of question type 50% of the time. The number of IC used between the *Panellists* and the *Witness* is balanced: the *Panellists* use the 47% of the annotated IC, whereas the *Witness* use the 48% of the annotated IC. Nevertheless, an interesting discrepancy between the use of IC among the *Panellists* and the *Witness* can be identified. Although both the *Panellists* and the *Witness* use mainly *Asserting* (71% for *Panellists* and 88% for *Witness*), the *Panellists* use more questions and challenging than the *Witness*. In particular, the *Panellists* use 22% question type ICs and 2% challenging type ICs, whereas the *Witness* use 3% question type ICs and 0.3% challenging type ICs. This is consistent with *Panellists* playing the role of inquisitive speakers who challenge the *Witness*. For more detail, see Appendix A Table 6.

Finally we checked the number of IC per speakers' stance. Although the *Pro* stance uses more IC than *Con* (*Pro* 51% of the time and *Con* 44% of the time), the types of IC used between the *Pro* and *Con* is reasonably balanced (for details, see Appendix A Table 7).

5. QTMM2012c+ use cases

In order to show the flexibility and utility of QTMM2012c+, in this section we present two use cases. As a first example, we show how QTMM2012c+ can be used to define a set of dialogue rules and we compare our rules with those reported in [19]. As a second example, we show how QTMM2012c+ can be used for extracting dialogue generation templates.

5.1. Dialogue Rules Extraction

QTMM2012c+ lends itself to extracting empirically-grounded rules for selecting the next dialogue act/proposition based on the dialogue history and propositional relation.

A dialogue rule is essentially an abstraction which collects together transitions that share certain properties.

We can count how many transitions fit a specific rule – below, we report this rule frequency in brackets at the end of each line. (For the sake of simplicity, the examples are extracted from transitions of degree 2 with no reported speech, and $\phi \rightarrow \psi$ stands for an (IAT) Inference from ϕ to ψ.)

Let us suppose we are interested in a rule that captures what follows after pure questioning for the case of same-speaker transition. We can query QTMM2021c+ to obtain the following rule:

Rule1: After Pure Questioning ϕ, a participant performs:

R1.1: Rephrase, via Pure Questioning ψ, where ($\psi \rightarrow \phi$) (6 times);
R1.2: Rephrase, via Asserting ψ, where ($\psi \rightarrow \phi$) (3 times);
R1.3: Rephrase, via Assertive Questioning ψ, where ($\psi \rightarrow \phi$) (2 times);
R1.4: Inference, via Asserting ψ, where ($\psi \rightarrow \phi$) (1 times);
R1.5: Introducing another statement via Pure Questioning ψ (11 times)
R1.6: Introducing another statement via Assertive Questioning ψ (1 times)
R1.7: Introducing another statement via Asserting ψ (1 times)

Yaskorska-Shah [19] defines a similar rule, but their rule for Pure Questioning only covers our R1.5 and R1.6. Also, as shown, from QTMM2012c+ we obtain rules together with their frequency. This is useful, for instance, when constructing a dialogue system, since it provides the raw material for choosing between dialogue rules using a statistical model.

Let us see a couple of further examples that illustrate the flexibility of QTMM2012c+. Suppose this time we are interested in a rule that involves the *Chair*, more specifically, a rule that describes the possible behaviour of a participant after the *Chair* performs a Pure Questioning.[10]

Rule2: After the *Chair*'s Pure Questioning ϕ, a participant performs:

R2.1: Agree, via Asserting Yes (2 times);
R2.2: Disagree, via Asserting Not-ϕ (3 times);
R2.3: Rephrase, via Asserting ψ, where ($\psi \rightarrow \phi$) (14 times);
R2.4: Rephrase, via Pure Questioning ψ, where ($\psi \rightarrow \phi$) (1 times);
R2.5: Inference, via Popular Conceding ψ, where ($\psi \rightarrow \phi$) (1 times);
R2.6: Introducing a second Pure Questioning (5 times).

With Rule 2, we capture the possible moves after the *Chair* performs a Pure Questioning. In this case, the rule does not specify properties of the participant who makes the response move. However, QTMM2012c+ allows for the extraction of more fine-grained rules:

Rule 3 describes a *Witness*'s behaviour after the *Chair* performs a Pure Question.

Rule3: After the *Chair* Pure Questioning ϕ, the *Witness* performs:

[10] For this example, this participant can be the Chair themselves, a Witness or a Panellist.

R3.1: Rephrase, via Asserting ψ, where ($\psi \rightarrow \phi$) (7 times);
R3.2: Inference, via Popular Conceding ψ, where ($\psi \rightarrow \phi$) (1 times);
R3.3: Agree, via Asserting Yes (2 times);
R3.4: Disagree, via Asserting Not-ϕ (2 times);

These examples illustrate the flexibility of QTMM2012c+: rules can be extracted based on a combination of roles and illocutionary connection, or a combination of stance and illocutionary connection or a combination of roles, stance and illocutionary connection. But they can be also more general, and be extracted using the illocutionary connection only.

Furthermore, frequency information can be used to build statistical models, for example, a model that aims to predict the next illocutionary connection. This kind of model can be then applied in a multi-party argumentative dialogue system.

Finally, because the rules are directly grounded in the rows that make up the Queryable Table QTMM2012c+ and which represent transitions, each rule can be traced back to the transitions that gave rise to it. For example, Figure 5 shows the transition instance underlying R1.6.[11]

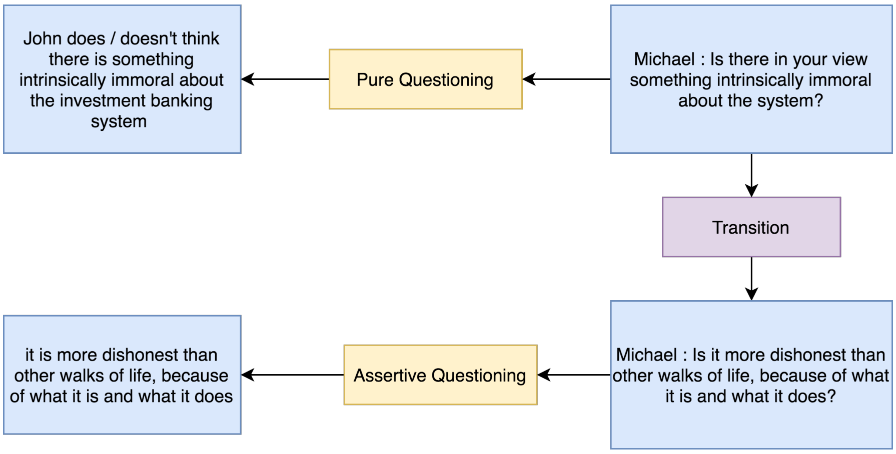

Figure 5: Example from Moral Maze Banking System, Map 5601 (http://ova.arg-tech.org/analyse.php?url=local&plus=true&aifdb=5601&akey=bd80882d6bd1c3c342b80d96b8d55f47).

Figure 5 shows a rule which was only observed a single time in the corpus (frequency equal to 1). This does not have to be interpreted as an aberration or an illegitimate response representing a breakdown in the dialogue. The corpus consists of real debates, where some patterns may occur very infrequently (e.g. because their felicity depends on a very specific dialogue context). Far from being considered as an error, they have to be considered as a possibly legitimate move in this kind of dialogue. That said, researchers that will use rules extracted from QTMM2012c+ can decide to use only the rules with a high frequency.

[11] As an illustrative example we have released the code for extracting Rule 1 at https://github.com/jacopoamidei/Supplementary-material-A-Queryable-Empirically-grounded-Resource-of-Dialogue-with-Argumentation. This code also prints the transition instances from which the rule arose.

5.2. Templates at sentence level

QTMM2012c+ can also be also used to extract generation templates at a sentence level [25].

As an example, suppose we are interested in defining templates for the *Chair* when (s)he performs a Pure Questioning. By querying QTMM2012c+, it is possible to extract all the locutions that where uttered by the *Chair* for which the Illocutionary connection is Pure Questioning.

In the dataset there are 28 locutions that satisfy this query. These locutions can be used to create generation templates. For example, among these locutions we can find the following:

- MB : What do you think?
- MB : Do you think we have got a lot to apologise for?
- Michael : do you think Cameron has a point?
- Michael : Or do you think it's all just demonising the poor?

All these examples have something in common, they query someone by using the formula: "(What) do you think".[12] Based on this observation we can define three templates: I) What do you think?, II) Do you think *[statement]* ? and, III) Do you think *[participant name]* has a point? These templates can be used in a multi-party argumentative dialogue system when the speaker in the turn is the *Chair* and the illocutionary connection to be used is Pure Questioning.

As the example shows, QTMM2012c+ can be used to isolate a set of locutions/propositions we are interested in (for example, locutions uttered by the *Chair* that are Pure Questioning). Once this set of naturally-occurring utterances from the multi-party argumentative dialogue is isolated, several strategies can be used to extract templates. For example, if the set of utterances is small, the template can be manually extracted. If the set of utterances is large, more complex models, for example statistical models, can be used.

6. Conclusion

In this paper we introduced QTMM2012c+, a queryable resource of annotated real-life debates which captures the links between propositional relations and dialogue act types, speaker roles and speaker stance. We present corpus analyses using the proposed resource and two use cases for the queryable resource. In particular, we suggest a way to use QTMM2012c+ to: i) define rules for multi-party argumentative dialogue and, ii) extract generation templates at a sentence level. The use cases are not fully developed here (this is ongoing work) and primarily aimed at providing two concrete instances of how the corpus can be used in the wider context of AI and argumentation research.

We are sharing QTMM2012c+ with the research community in the hope that its flexibility will allow others to explore natural argumentation in dialogue along the examples illustrated in this paper and beyond.

Acknowledgments

This work was supported by the UK Engineering and Physical Sciences Research Council (EPSRC) grant EP/T024666/1 'Opening Up Minds: Engaging Dialogue Generated from Argument Maps'.

[12]Both MB and Michael refer to Michael Buerk who is Chair of all Moral Maze episodes in the MM2012c dataset.

We would like to acknowledge discussions with and feedback from our project partners at Cambridge (Andreas Vlachos and Youmna Farag) and Sheffield (Tom Stafford and Lotty Brand) that informed the work reported in this paper.

References

[1] P. Budzianowski, T.-H. Wen, B.-H. Tseng, I. Casanueva, S. Ultes, O. Ramadan, M. Gašić, MultiWOZ - a large-scale multi-domain Wizard-of-Oz dataset for task-oriented dialogue modelling, in: Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, Brussels, Belgium, 2018, pp. 5016–5026. URL: https://www.aclweb.org/anthology/D18-1547. doi:10.18653/v1/D18-1547.

[2] A. Köhn, J. Wichlacz, C. Schäfer, A. Torralba, J. Hoffmann, A. Koller, Mc-saar-instruct: a platform for minecraft instruction giving agents, in: Proceedings of the 21th Annual Meeting of the Special Interest Group on Discourse and Dialogue, Association for Computational Linguistics, 1st virtual meeting, 2020, pp. 53–56. URL: https://www.aclweb.org/anthology/2020.sigdial-1.7.

[3] C. Zhu, Boosting naturalness of language in task-oriented dialogues via adversarial training, in: Proceedings of the 21th Annual Meeting of the Special Interest Group on Discourse and Dialogue, Association for Computational Linguistics, 1st virtual meeting, 2020, pp. 265–271. URL: https://www.aclweb.org/anthology/2020.sigdial-1.33.

[4] B. Fazzinga, A. Galassi, P. Torroni, An argumentative dialogue system for covid-19 vaccine information, in: International Conference on Logic and Argumentation, Springer, 2021, pp. 477–485.

[5] E. Dinan, S. Roller, K. Shuster, A. Fan, M. Auli, J. Weston, Wizard of wikipedia: Knowledge-powered conversational agents, in: 7th International Conference on Learning Representations, ICLR 2019, New Orleans, LA, USA, May 6-9, 2019, OpenReview.net, 2019. URL: https://openreview.net/forum?id=r1l73iRqKm.

[6] S. Roller, Y.-L. Boureau, J. Weston, A. Bordes, E. Dinan, A. Fan, D. Gunning, D. Ju, M. Li, S. Poff, P. Ringshia, K. Shuster, E. M. Smith, A. Szlam, J. Urbanek, M. Williamson, Open-domain conversational agents: Current progress, open problems, and future directions, 2020. arXiv:2006.12442.

[7] A. Cervone, G. Riccardi, Is this dialogue coherent? learning from dialogue acts and entities, in: Proceedings of the 21th Annual Meeting of the Special Interest Group on Discourse and Dialogue, Association for Computational Linguistics, 1st virtual meeting, 2020, pp. 162–174. URL: https://www.aclweb.org/anthology/2020.sigdial-1.21.

[8] W. Sieińska, C. Dondrup, N. Gunson, O. Lemon, Conversational agents for intelligent buildings, in: Proceedings of the 21th Annual Meeting of the Special Interest Group on Discourse and Dialogue, Association for Computational Linguistics, 1st virtual meeting, 2020, pp. 45–48. URL: https://www.aclweb.org/anthology/2020.sigdial-1.5.

[9] K. Sun, S. Moon, P. Crook, S. Roller, B. Silvert, B. Liu, Z. Wang, H. Liu, E. Cho, C. Cardie, Adding chit-chats to enhance task-oriented dialogues, 2020. arXiv:2010.12757.

[10] C. L. Hamblin, Fallacies, Reprint, Vale Press, 1986, 1970.

[11] A. Hunter, Towards a framework for computational persuasion with applications in behaviour change, Argument & Computation 9 (2018) 15–40.
[12] J. Lawrence, C. Reed, AIFdb corpora., in: Proceedings of the Computational Models of Argument (COMMA) Conference, 2014, pp. 465–466.
[13] J. Lawrence, M. Janier, C. Reed, Working with open argument corpora, in: European Conference on Argumentation (ECA), 2015.
[14] M. Janier, Dialogical dynamics and argumentative structures in dispute mediation discourse, Ph.D. thesis, University of Dundee, 2017.
[15] K. Budzynska, M. Janier, J. Kang, C. Reed, P. Saint-Dizier, M. Stede, O. Yaskorska, Towards argument mining from dialogue., in: Proceedings of the Computational Models of Argument (COMMA) Conference, 2014, pp. 185–196.
[16] K. Budzynska, M. Janier, J. Kang, B. Konat, C. Reed, P. Saint-Dizier, M. Stede, O. Yaskorska, Automatically identifying transitions between locutions in dialogue, in: Proceedings of 1st European Conference on Argumentation (ECA 2015), 2015.
[17] K. Budzynska, M. Janier, C. Reed, P. Saint-Dizier, M. Stede, O. Yaskorska, A model for processing illocutionary structures and argumentation in debates, in: LREC 2014: Ninth International Conference on Language Resources and Evaluation, European Language Resources Association, 2014, pp. 917–924.
[18] K. Budzynska, M. Janier, C. Reed, P. Saint-Dizier, Theoretical foundations for illocutionary structure parsing 1, Argument & Computation 7 (2016) 91–108.
[19] O. Yaskorska-Shah, Managing the complexity of dialogues in context: A data-driven discovery method for dialectical reply structures, Argumentation (2021) 1–30.
[20] S. Stoyanchev, P. Piwek, Harvesting re-usable high-level rules for expository dialogue generation, in: Proceedings of the 6th International Natural Language Generation Conference, Association for Computational Linguistics, Trim, Co. Meath, Ireland, 2010. URL: https://www.aclweb.org/anthology/W10-4215.
[21] C. Reed, K. Budzynska, How dialogues create arguments, in: Proceedings of the 7th conference on argumentation of the International Society for the Study of Argumentation, 2011, pp. 1633–1645.
[22] J. R. Searle, Speech acts: An essay in the philosophy of language, Cambridge University Press, 1969.
[23] S. Wells, Moral maze transcripts. figshare. dataset., 2013. URL: https://doi.org/10.6084/m9.figshare.741242.v1.
[24] W. C. Mann, S. A. Thompson, Rhetorical structure theory: Toward a functional theory of text organization, Text - Interdisciplinary Journal for the Study of Discourse 8 (1988) 243–281. URL: https://doi.org/10.1515/text.1.1988.8.3.243.
[25] K. v. Deemter, M. Theune, E. Krahmer, Real versus template-based natural language generation: A false opposition?, Computational linguistics 31 (2005) 15–24.

A. QTMM2012c+ detailed statistics

ICTA	Inference	Conflict	Rephrase	NA	Total
Disagreeing	0	215	0	3	**218**
Agreeing	1	0	1	107	**109**
Restating	0	0	7	4	**11**
Arguing	851	0	0	0	**851**
Asserting	10	0	0	7	**17**
Default Illocuting	0	0	153	1	**154**
Pure Questioning	0	0	0	3	**3**
Assertive Questioning	0	0	0	4	**4**
Pure Challenging	0	0	0	4	**4**
Rhetorical Challenging	0	0	0	1	**1**
TA is non-anchoring	0	0	0	374	**374**
Total	**862**	**215**	**161**	**508**	-

Table 1
Number of illutionary connections associated with the transitions (ICTA) per propositional relation. *NA* means that the propositional relation is missing from a transition. *TA is non-anchoring* means that the transition is not annotated with any illocutionary connection.

TA degree	Inference	Conflict	Rephrase	NA	Total
2	718	203	148	502	**1571**
3	116	10	12	4	**142**
4	18	2	1	2	**24**
5	5	0	0	0	**5**
6	3	0	0	0	**3**
7	2	0	0	0	**2**

Table 2
Number of propositional relations per degree.

TA degree	Same Speaker	Different Speaker
2	1099	472
3	125	17
4	18	5
5	4	1
6	3	6
7	1	1
Total	**1250**	**496**

Table 3
Same/Different speakers per transition (TA) degree

ICTA	2	3	4	5	6	7	Total
Disagreeing	206	10	2	0	0	0	**218**
Agreeing	108	0	1	0	0	0	**109**
Restating	11	0	0	0	0	0	**11**
Arguing	708	115	18	5	3	2	**851**
Asserting	16	1	0	0	0	0	**17**
Default Illocuting	141	12	1	0	0	0	**154**
Pure Questioning	3	0	0	0	0	0	**3**
Assertive Questioning	4	0	0	0	0	0	**4**
Pure Challenging	4	0	0	0	0	0	**4**
Rhetorical Challenging	1	0	0	0	0	0	**1**
TA is non-anchoring	369	4	1	0	0	0	**374**

Table 4
Number of ICTA per Degree.

IC	Number	IC	Number
Asserting	1799	**Popular Conceding**	29
Assertive Questioning	136	**Rhetorical Challenging**	14
Rhetorical Questioning	92	**Pure Challenging**	9
Pure Questioning	91	**Assertive Challenging**	8
Yes	68	**Ironic Asserting**	2
No	39		

Table 5
Number of illuctionary connection (IC).

IC	Chair	Panellist	Witness
Asserting	39	755	960
Yes	0	24	44
No	0	13	26
Pure Questioning	28	58	2
Assertive Questioning	14	116	6
Rhetorical Questioning	5	55	31
Pure Challenging	2	7	0
Assertive Challenging	1	7	0
Rhetorical Challenging	0	11	3
Popular Conceding	0	14	15
Ironic Asserting	0	2	0
Total	89	1062	1087

Table 6
Number of IC per speakers' roles.

B. QTMM2012c+ columns name

'ID', 'ID map', 'Dataset', 'Reported Speech', 'Speakers', 'Degree', 'ICTA', 'PropRel', 'L1', 'L1 IC', 'L1 Role', 'L1 Stance', 'L1 ID dialogue flow', 'L1 ID', 'L1Reported Speech', 'L1Reported Speech IC', 'L1Reported Speech Role', 'L1Reported Speech Stance', 'L1Reported Speech ID', 'L2', 'L2 IC',

IC	Chair	Pro	Con
Asserting	39	928	787
Yes	0	39	29
No	0	20	19
Pure Questioning	28	32	28
Assertive Questioning	14	65	57
Rhetorical Questioning	5	36	50
Pure Challenging	2	4	3
Assertive Challenging	1	3	4
Rhetorical Challenging	0	8	6
Popular Conceding	0	14	15
Ironic Asserting	0	2	0
Total	89	1151	998

Table 7
Number of IC per speakers' stance.

'L2 Role', 'L2 Stance', 'L2 ID dialogue flow', 'L2 ID', 'L2Reported Speech', 'L2Reported Speech IC', 'L2Reported Speech Role', 'L2Reported Speech Stance', 'L2Reported Speech ID', 'L3', 'L3 IC', 'L3 Role', 'L3 Stance', 'L3 ID dialogue flow', 'L3 ID', 'L3Reported Speech', 'L3Reported Speech IC', 'L3Reported Speech Role', 'L3Reported Speech Stance', 'L3Reported Speech ID', 'L4', 'L4 IC', 'L4 Role', 'L4 Stance', 'L4 ID dialogue flow', 'L4 ID', 'L4Reported Speech', 'L4Reported Speech IC', 'L4Reported Speech Role', 'L4Reported Speech Stance', 'L4Reported Speech ID', 'L5', 'L5 IC', 'L5 Role', 'L5 Stance', 'L5 ID dialogue flow', 'L5 ID', 'L6', 'L6 IC', 'L6 Role', 'L6 Stance', 'L6 ID dialogue flow', 'L6 ID', 'L7', 'L7 IC', 'L7 Role', 'L7 Stance', 'L7 ID dialogue flow', 'L7 ID', 'PL1Reported Speech', 'PL1Reported Speech ID', 'PL2Reported Speech', 'PL2Reported Speech ID', 'PL3Reported Speech', 'PL3Reported Speech ID', 'PL4Reported Speech', 'PL4Reported Speech ID', 'P1', 'P1 ID', 'P2', 'P2 ID', 'P3', 'P3 ID', 'P4', 'P4 ID', 'P5', 'P5 ID', 'P6', 'P6 ID', 'P7', 'P7 ID', 'L1Reported Speech to L2Reported Speech', 'L1Reported Speech to P2', 'L1Reported Speech to P3', 'L2Reported Speech to L1Reported Speech', 'L2Reported Speech to PL1Reported Speech', 'L2Reported Speech to P1', 'L2Reported Speech to P2', 'L3Reported Speech to P1', 'L4Reported Speech to P1', 'NOT L1Reported Speech to L1Reported Speech', 'NOT PL1Reported Speech to P1', 'NOT PL2Reported Speech to PL2Reported Speech', 'NOT P1 to P1', 'PL1Reported Speech to PL2Reported Speech', 'PL1Reported Speech to P2', 'PL1Reported Speech to P3', 'PL1Reported Speech to P4', 'PL2Reported Speech to NOT PL2Reported Speech', 'PL2Reported Speech to PL1Reported Speech', 'PL2Reported Speech to P1', 'PL2Reported Speech to P3', 'PL3Reported Speech to P1', 'P1 to L2Reported Speech', 'P1 to PL2Reported Speech', 'P1 to P2', 'P1 to P3', 'P1 to P4', 'P1 to P5', 'P2 to L1Reported Speech', 'P2 to Out of TA', 'P2 to PL1Reported Speech', 'P2 to P1', 'P2 to P3', 'P2 to P4', 'P2 to P5', 'P3 to PL1Reported Speech', 'P3 to P1', 'P3 to P2', 'P3 to P4', 'P3 to P5', 'P4 to PL1Reported Speech', 'P4 to P1', 'P4 to P2', 'P4 to P3', 'P4 to P5', 'P5 to PL1Reported Speech', 'P5 to P1', 'P6 to P1', 'P6 to P5', 'P7 to P1', 'P7 to P5', 'Rephrase P1 to P1', 'NOT L1Reported Speech', 'NOT PL1Reported Speech', 'NOT PL2Reported Speech', 'NOT P1', 'Out of TA', 'To PL1Reported Speech', 'To PL2Reported Speech', 'To PL3Reported Speech', 'To P1', 'To P2', 'Rephrase P1', 'TA is nonanchoring'.