Proceedings of the 13th Conference on Language Resources and Evaluation (LREC 2022), pages 3291–3300
Marseille, 20-25 June 2022

QT30: A Corpus of Argument and Conflict in Broadcast Debate

Annette Hautli-Janisz[1], Zlata Kikteva[1], Wassiliki Siskou[1,2], Kamila Gorska[3], Ray Becker[3], Chris Reed[3]
[1]University of Passau, [2]University of Konstanz, [3]University of Dundee
firstname.lastname@uni-passau.de, firstname.lastname@uni-konstanz.de,
{kamila, ray, chris}@arg.tech

Abstract

Broadcast political debate is a core pillar of democracy: it is the public's easiest access to opinions that shape policies and enables the general public to make informed choices. With QT30, we present the largest corpus of analysed dialogical argumentation ever created (19,842 utterances, 280,000 words) and also the largest corpus of analysed broadcast political debate to date, using 30 episodes of BBC's 'Question Time' from 2020 and 2021. Question Time is the prime institution in UK broadcast political debate and features questions from the public on current political issues, which are responded to by a weekly panel of five figures of UK politics and society. QT30 is highly argumentative and combines language of well-versed political rhetoric with direct, often combative, justification-seeking of the general public. QT30 is annotated with Inference Anchoring Theory, a framework well-known in argument mining, which encodes the way arguments and conflicts are created and reacted to in dialogical settings. The resource is freely available at `http://corpora.aifdb.org/qt30`.

Keywords: broadcast political debate, argumentation and conflict, Question Time, Inference Anchoring Theory

1. Introduction

Political discussions and interviews on TV & radio are the public's easiest access to opinions that shape policies and thereby impact on their lives. One of the most viewed political talk show in the UK, 'Question Time' (henceforth, QT), is a topical debate in which audience members request justifications from a panel of political and societal figures on current topics of broader national interest. The debate is moderated and features scheduled as well as spontaneous questions from the audience to the panel. The panel responds to those questions by presenting their standpoints (or those of the stakeholders they represent), arguing for them and often engaging in a discussion with the other panel members. QT has been broadcast since 1979, is aired on Thursdays at 10:45 pm on BBC One and is repeated twice later in the week. It reaches a public of around 1.4 million viewers every week[1] and is according to YouGov[2] the most viewed topical debate in the UK.

Despite the central role that broadcast political debate has on public opinion formation, there is a dire lack of resources that illustrate the ways in which the public interacts with political and societal figures in real time on issues that have an impact on their daily lives. Disagreeing with political decisions is one way of calling out individual issues, but the general tone in QT is one of asking for justifications of political decisions, sometimes diplomatically, but sometimes also outwardly hostile. In QT30, we use Inference Anchoring Theory (Budzynska et al., 2014b; Budzynska et al., 2016), a well-known framework in argument mining, to identify the argumentative structure of 30 Question Time debates: we identify all claims, attacks and supports, and the illocutionary force with which they were contributed in the debate. The episodes were broadcast in 2020 and 2021 and cover controversial debates on Brexit, the government's handling of the Covid-19 pandemic and national scandals like the Dominic Cummings affair. QT30 is the largest individual corpus of analysed dialogical argumentation, surpassing the second-biggest English-language corpus, US2016 (Visser et al., 2020), by a factor of three. The annotation is of high-quality: with an inter-annotator agreement of CASS = 0.56, we achieve state-of-the-art results in manual annotation of argumentation and conflict in dialogue. All QT30 data is publicly available at `http://corpora.aifdb.org/qt30`.

The paper proceeds as follows: Section §2 provides related work on argument and conflict annotation in dialogue, with Section §3 elaborating on the details of the annotation scheme and the methodology of creating the QT30 corpus. Section §5 gives an insight into the unique richness and detailedness of QT30 by eliciting some of the rhetorical patterns of conflict and argumentation that constitute dialogue between members of the public and the political and societal sphere. Section §6 summarises the paper.

2. Related work

This paper touches upon two strands of previous work: annotation of dialogical argumentation in natural communication and identification of rhetorical and conversational patterns in broadcast political debate. While the former has seen a sharp rise with the field of argument mining burgeoning, it is the intricate pragmatic

[1] `https://twitter.com/BBCNewsPR/status/1331526320795365378?ref_src=twsrc%5Etfw`
[2] `https://yougov.co.uk/ratings/media/popularity/current-tv-programmes/all`

structure of dialogue which has been dealt with to a lesser extent. Consequently, QT30 builds on one of the most prominent frameworks in manual argument analysis, Inference Anchoring Theory (IAT) (Budzynska et al., 2014b; Budzynska et al., 2016), to model and understand the ways in which arguments are created, referred to, supported and attacked in combative dialogical settings. IAT focuses on a specific subset of what is accounted for in general purpose annotation schemes for discourse such as the Penn Discourse Treebank (Prasad et al., 2008) or Segmented Discourse Representation Theory (Asher and Lascarides, 2003), namely relations of explanation, justification and opposition. IAT has been applied to a number of genres, from moral debates (Budzynska et al., 2014a), online deliberative democracy (Konat et al., 2016) to dispute mediation (Janier and Reed, 2016) and US presidential debates (Visser et al., 2018). It has also been the basis of state-of-the-art work in argument mining, particularly reserch focusing on domain-independent algorithms such as (Gemechu and Reed, 2019). With QT30 we do not only provide the largest corpus of broadcast debate, but also the largest corpus of dialogical argumentation.

With respect to broadcast political debate, the genre has attracted increasing attention given the societal interest in the process of political decision making. It is particularly election debates and TV interviews with political figures that have been resourced, for instance the 2016 US presidential debates (Haddadan et al., 2019; Visser et al., 2020) and the UK prime ministerial debates (Degano, 2016) have been annotated with argumentative structure. Smaller-scale studies investigated individual phenomena, e.g., (Luginbühl, 2007) investigated conversational violence in Swiss TV debate shows and (Hess-Lüttich, 2007) identified (pseudo)-argumentation in German and Swiss TV debates. With QT30 we make a significant contribution in this area: we provide an open-source dataset of significant size and state-of-the-art annotation, paving the way for a large-scale longitudinal analysis of rhetorical and argumentative maneuvering in broadcast political debate.

Debate as a whole has been attracting significant attention recently, not least because of the high-profile work conducted in the context of IBM's Project Debater (Slonim et al., 2021). Annotations in this work, however, are more shallow, aiming for broad coverage rather than capturing argumentative phenomena such as undercut, linkage and (dis-)agreement. Increasingly deeper analysis is now also becoming available languages other than English such as the 140,000-word VivesDebate corpus in Catalan (Ruiz-Dolz et al., 2021).

3. Annotation

3.1. Inference Anchoring Theory

Inference Anchoring Theory (IAT) (Budzynska et al., 2014b; Budzynska et al., 2016) provides a theoretical scaffolding to handle *dialogue and argument structures, and the relations between them* and has been applied to over 2.5 million words in fifteen languages (available online at `corpora.aifdb.org`). IAT has three types of relations: (i) relations between content (propositional content of locutions); (ii) illocutionary connections that link locutions with their content and (iii) relations between locutions in a dialogue, called transitions.[3]

Example (1), taken from the QT episode on 4 November 2021, illustrates a typical exchange between an audience member (AudienceMember 20211111QT10), the moderator (Fiona Bruce) and a panel member (Paul Polman).[4] The question at [00:30:22] by the audience member is the second out of four scheduled questions to the panel and refers to the upcoming UN Climate Change Conference (COP-26) in Glasgow:

(1) AudienceMember 20211111QT10 [0:30:22] *Without the likes of Russia, India and China on board, are this week's COP-26 breakthrough pledges such as the reduction of methane emissions a glimmer of hope or more blah blah blah?*

Fiona Bruce [0:30:38] *[...] Paul. [...]*

Paul Polman [0:30:52] *[...] The question is a pertinent one. We need the bigger countries like India, China, Russia, 100 percent on board. But the situation is actually a little bit better than the question would suggest. Before COVID started, we have all worried that nobody was serious to stick to the Paris Agreements [...]*

The full excerpt in the corpus[5] shows that the question by the audience member refers to Greta Thunberg, who used 'blah, blah, blah' to dimiss the outcomes of COP-26. Fiona Bruce redirects the question at panel member Paul Polman, who acknowledges the issue ('the question is a pertinent one') and provides support for why the issue pointed at in the question is correct ('we need the bigger countries like India, China, Russia, 100 percent on board'). But he also attacks the underlying message of the question that the situation with COP-26 breakthrough pledges is bad ('the situation is actually a little bit better [...]'). Polman supports this claim with the last locution ('before COVID started, we have all worried that nobody was serious to stick to the Paris Agreements').

[3]The IAT annotation guidelines are here: `http://www.arg.tech/f/IATannotationguidelines.pdf`.

[4]The QT panel on 4 November 2021 consisted of Caroline Lucas (Green Party MP), Emily Thornberry (Labour Party MP), Paul Polman (social entrepreneur), Paul Scully (Conservative Party MP) and Tim Stanley (journalist).

[5]The graph is in map 22357 of the corpus here: http://corpora.aifdb.org/qt11112021

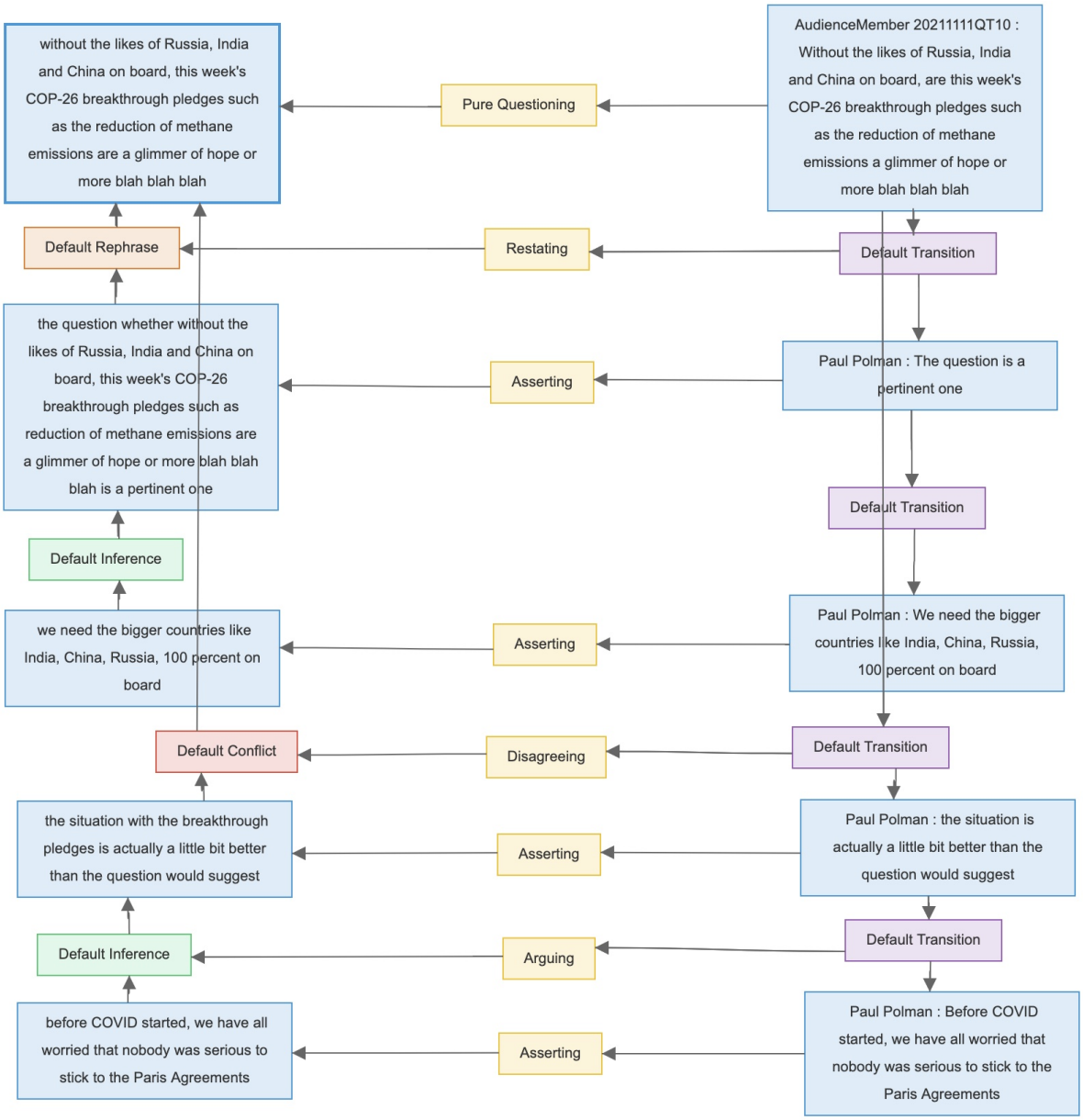

Figure 1: IAT diagram of example (1), featuring locations (blue nodes on the right-hand side), propositions (blue nodes on the left-hand side), illocutionary relations (yellow nodes in the middle), dialogical relations (purple nodes on the right) and propositional relations – 'Default Inference' (green), 'Default Rephrase' (orange) and 'Default Conflict' (red).

Figure 1 provides the IAT analysis for example (1), produced with OVA+ (Online Visualisation of Argument – `http://ova.arg.tech/`), an open-source online interface for the analysis of argumentation in dialogues (Janier et al., 2014). OVA+ allows for a representation of the argumentative structure of a text as a directed graph. The different components of IAT analysis for example 1, in particular locutions, propositions, propositional relations and illocutionary forces, are discussed in detail in the following.

3.1.1. Locutions

A locution (the 'right-hand side' nodes in the graph in Figure 1) is equivalent to an argumentative discourse unit (Peldszus and Stede, 2013), i.e., the minimal unit into which the transcribed text is segmented. It has discrete argumentative function and records the name of the speaker in the form of 'firstname lastname : *locution content*'. Content that is merely used for discourse management, for instance see Fiona Bruce's contribution at [00:30:38] ('Paul.'), is left out of the graph.

3.1.2. Propositions

Propositions are derived from locutions and are grammatical instantiations of the content of the locution. They have to be interpretable without context, i.e., they are standalone propositions that are intelligible without knowledge of surrounding propositional content.

As a consequence, propositions may have to be reconstructed, so for instance elliptical or anaphoric expressions contained in the locution are resolved in the proposition. An example of this is the proposition of the second locution in Figure 1, which is manually reconstructed from the locution 'the question is a pertinent one' (right-hand side) to 'the question whether without (...) is a pertinent one' in the proposition (left-hand side). The guideline for the annotators is to do minimal reconstruction in creating the proposition, though we do see some variability between annotators in balancing interpretability and reconstruction.
For questions, IAT makes very clear the separation of illocutionary force and propositional content (Hautli-Janisz et al., 2022), which is in line with speech act theory (Searle, 1969; Searle and Vanderveken, 1985): The first locution in Figure 1 is an information-seeking question, i.e., the audience member requests information from the panel members. The proposition of this alternative question ('a glimmer of hope or more blah blah blah?') encodes the content of the question but not the speaker intent – it is therefore recorded in declarative syntax ('the breakthrough pledges (...) are a glimmer of hope or more blah blah blah'). The intention as an information-seeking question is captured in the illocutionary connection of 'Pure Questioning' alone (more in §3.1.4).

3.1.3. Propositional relations

Argumentative structures are relations between propositions; core IAT assumes three different relations that are designed to capture argumentative structure in dialogue:

Inference (Support, 'Default Inference', RA, green node) Holds between propositions when one (or more) proposition is used to provide a reason to accept another proposition.

Conflict (Attack, 'Default Conflict', CA, red node) Holds between two propositions when one proposition is used to provide an incompatible alternative to another proposition.

Rephrase (Rephrase, 'Default Rephrase', MA, orange node) Holds between two propositions when one proposition is used to rephrase, restate or reformulate another proposition. Rephrases also hold between questions and answers.

In Example 1 we find instances of all three relations: The first locution by Paul Polman is annotated as a 'Default Rephrase' of the audience question – Paul Polman intends to refine the question, acknowledging the issue under discussion. He substantiates his rephrasing statement with the claim 'we need the bigger countries like India, China, Russia, 100 percent on board', which is encoded as a 'Default Inference', the IAT relation for supports in argumentative structures. He also provides justification of this claim ('before COVID started, we have all worried that nobody was serious to stick to the Paris Agreements'), again encoded as a 'Default Inference' to the previous statement.
These relations are 'Default' in the sense that they can be instantiated with more specific relation types, for instance with presumptive argument scheme types (Walton et al., 2008).

3.1.4. Illocutionary relations

Illocutionary relations are the glue between locutions and propositions and capture the communicative intent of the speaker in uttering a locution or a pair of locutions. We instantiate illocutionary relations with the concept of illocutionary forces of (Searle, 1969) and (Searle and Vanderveken, 1985), however we use a simplified version in which there are no commissives or expressives and just two simple types of directives (questions and challenges). For QT30, we use a set of nine illocutionary connections that either anchor propositions in locutions (as with 'Asserting') or propositional relations in the dialogical structure (as with 'Arguing', 'Disagreeing' and 'Restating' in the case of 'Default Inference', 'Default Conflict' and 'Default Rephrase', respectively). They are described as follows:

- **Asserting** Speaker S asserts information or communicates an opinion.
- **Agreeing** Speaker S declares that they share the opinion of the interlocutor.
- **Arguing** Speaker S provides justification to a claim. Anchors a 'Default Inference'.
- **Assertive Questioning** Speaker S communicates information and at the same time asks for confirmation/rejection from their interlocutor.
- **Challenging** Speaker S declares that they are seeking the grounds for the interlocutor's opinion.
- **Disagreeing** Speaker S attacks the interlocutor or declares not to share the interlocutor's opinion. Anchors a 'Default Inference' or disagreement when no propositional content is given, e.g., via 'No.'.
- **Pure Questioning** Speaker S is seeking information or asking for the opinion of their interlocutors.
- **Restating** Speaker S rephrases a previous claim. Anchors a 'Default Rephrase'.
- **Rhetorical Questioning** Speaker S is expressing an opinion in the form of an interrogative.
- **Default Illocuting** Used to anchor a a 'Default Rephrase' which captures an answer to a question, i.e., the answer instantiates (parts of) the question.

Figure 1 illustrates the 'glue' function of the illocutionary connections in IAT: For instance, the information-seeking function of the question in the first locution is captured by the illocutionary connection of 'Pure Questioning', i.e., the audience member intends to receive information from the panel. The intention of Paul Polman to reframe the question of the audience member in the second locution is captured with the illocutionary connection 'Restating', anchoring the 'Default Rephrase' in the dialogical (right-hand) side structure. All nodes on the left-hand side, i.e. the propositions as well as the relations between them, need to be anchored in the right-hand side (the dialogical structure), indicating the dialogical speech action that has created them.

3.2. Methodology

The procedure for creating QT30 across two years of broadcast roughly follows the general pattern for manual argument analysis in (Lawrence and Reed, 2020) of segmentation, argument/non-argument classification and relation identification. For QT30 this results in the following methodology for corpus creation:

Chunking Once an episode of Question Time is available online on BBC iPlayer, it is either transcribed offline, or else made available to a stenographer. In either case, the starting point for analysis is a text of around 10,000 words (+/- 20%) that is first chunked into (40-80) excerpts each comprising around 150-250 words – passages that are small enough to be considered in total by an analyst, but large enough to include substantial dialogical exchange. We exploit natural topical, thematic and turn-based breaks to guide this chunking process.

Segmentation In the first step, an analyst segments the text into argumentative discourse units (or 'locutions' in IAT), producing between five and thirty locutions per excerpt. In the same step, the analyst reconstructs information to be recorded in the proposition, for instance anaphora and ellipses (see §3.1.2).

Intra-map classification and structure identification In QT30, these two steps are combined: An analyst identifies whether a locution has argumentative function or not, and in the latter case immediately adds the type of propositional relation and the illocutionary connection between locutions (right-hand side) and propositions (left-hand side). The result is a map containing between a dozen to a hundred nodes in total, depending on the length and the content of the excerpt. Each analysis map then undergoes peer review by which a randomly chosen second analyst reviews and discusses annotation choices with the first.

Inter-map structure identification A separate process then sees connections *between* maps annotated – such connections are typically quite rare (over 90% of connections are proximate and contained within maps), but reference back to points made earlier is an important feature of the debate structure. In addition dialogical and content connections between consecutive maps (i.e. between the last turn in one map and the first turn in the next) are also annotated. These inter-map connections are then also reviewed.

The result for each QT episode is then a single map involving 500-1,000 propositions, with 200-500 relations between them, making up anything from 1,000 to 3,000 nodes in total.

4. The QT30 corpus

QT30 is a collection of 30 episodes of Question Time aired between June 2020 and November 2021. The annotation was conducted by 38 students of linguistics, philosophy, literature and computer science in Scotland, England, Germany and Poland. More than 60 students took part in one of three rounds of training in 2020 and 2021. Topic of the 15 hour course (taught in person once in 2020 and then virtually three times in 2020 and 2021) was a general introduction to argumentation theory and detailed instructions on applying Inference Anchoring Theory to dialogical argumentation across genres. Due to the strict quality restrictions for QT30, only the top 38 annotators were selected to contribute. The resource will be made publicly available upon final submission.

4.1. Corpus statistics

In total, QT30 consists of more than 29 hours of transcribed broadcast material and comprises of 19,842 locutions by more than 400 participants: one moderator, 125 panel members (7 of them appearing more than once) and 300+ audience members. Table 1 provides an overview of the size of the corpus in terms of propositional and illocutionary relations. In total, the corpus features 10,818 propositional relations, i.e. argumentative structures. Inference ('supports') and Rephrase have the highest frequency, 48% and 42.6%, respectively. Conflicts are significantly less frequent, making up only 9.4% of all relations between propositions. Section 5 provides a more detailed look into how these relations are distributed across roles in QT30.

For illocutionary connections, 'Asserting' is the most common illocutionary force, as is typical in corpora of dialogical argumentation (Budzynska et al., 2014c; Janier and Reed, 2016), meaning that information provision is the most important speaker intent across genres of argumentation. Next come the relations of 'Arguing', 'Restating' and 'Disagreeing', in the same order of frequency as with propositional relations. Explicit agreement to previous content in the form of 'Yes.', captured via 'Agreeing', is done very rarely in the dataset (0.8% in QT30). In the same way, direct challenges only constitute 0.3% of all illocutionary relations, which is surprising given that QT30 is intended as a forum of justification seeking. Similarly surprising, out of 2,353 instances of Questioning, only 899 are 'answered', i.e., the answer is connected to the question via 'Default Illocuting'.

	in total	in %
Propositional relations		
Default Inference	5,205	48%
Default Conflict	976	9.4%
Default Rephrase	4,637	42.6%
Total	10,818	100%
Illocutionary relations		
Asserting	18,576	57.5%
Agreeing	260	0.8%
Arguing	5,106	15,8%
Disagreeing	1,035	3.2%
Restating	4,074	12.61%
Questioning	2,353	7.28%
pure	1,177	3.64%
assertive	865	2.7%
rhetorical	215	0.64%
challenging	96	0.3%
Default Illocuting	899	2.8%
Total	32,303	100%

Table 1: Distribution of propositional and illocutionary relations in QT30.

Table 2 shows the distribution of locutions across the moderator, members of the panel and members of the audience. The biggest share (64%) is contributed by the panel, followed by the audience (24%) and the moderator (12%). This distribution is expected given the set-up of 'Question Time' as a broadcast format in which the public requests information which the panel provides, with the moderator guiding the exchange.

	# of locutions	in %
moderator	2,381	12%
panel	12,699	64%
audience	4,762	24%
total	19,842	100%

Table 2: Number of locutions per QT30 roles: moderator, panel members and audience members

4.2. Corpus structure

Table 3 provides the structure of the subcorpora in QT30 – each episode is a subcorpus which is identified by the date in the corpus name. The word counts show that the average number of words per locution (or segment) are relatively stable across episodes (14.15 words/locution, standard deviation of 0.95). The number of words per corpus ranges between 7,489 (14October2021) and 11,761 (29April2021). A manual investigation of the five episodes with more than 10,000 words per corpus shows that these episodes feature monologues by panel members or individuals with high speaking rate, which results in a higher number of words overall.

Table 3 also shows that the number of words per corpus is about 10% lower than the number of words in the transcript of the episode. This is due to the fact that discourse structuring material (as in Example (1) and mentioned in §3.1.1) is left unanalysed by the annotators and is therefore not contained in the IAT graph.

Corpus name	words/ locution	words/ corpus	words/ episode
28May2020	14,32	8,835	10,521
4June2020	14.96	9,425	10,292
18June2020	14.94	9,726	10,569
30July2020	13.77	8,840	9,833
2September2020	16.07	10,911	10,911
22October2020	16.09	8,350	9,032
5November2020	14.24	9,583	11,293
19November2020	14.87	7,970	7,979
10December2020	13.87	8,973	9,013
14January2021	14.16	8,708	8,722
28January2021	16.18	9,303	9,031
18February2021	14.49	8,824	8,980
4March2021	13.65	9,541	9,611
18March2021	14.65	8,218	8,750
15April2021	14.88	10,996	11,911
29April2021	14.12	11,761	13,310
20May2021	14.79	11,432	11,697
27May2021	13.54	9,870	10,527
10June2021	12.67	9,147	9,299
24June2021	13.65	9,541	9,548
8July2021	13.71	8,815	8,826
22July2021	13.53	8,997	10,679
5August2021	13.01	8,860	8,913
19August2021	14.02	8,960	8,999
2September2021	14.5	9,468	10,130
16September2021	14.19	8,769	10,199
30September2021	13.14	10,577	12,046
14October2021	12.63	7,489	8,691
28October2021	13.02	10,429	11,430
11November2021	13.38	7,948	9,026
QT30	**14.15**	**279,966**	**290,299**

Table 3: Number of words/locution of each subcorpus in QT30, the number of analysed words in the graph and and overall word count per transcript

4.3. Inter-annotator agreement

We use the Combined Argument Similarity Score (CASS) (Duthie et al., 2016) to measure inter-annotator agreement across all 30 episodes of QT30. CASS calculates separate scores for segmentation, argumentative structures and illocutionary forces and aggregates them into a single score for annotator agreement.

As the basis for the CASS score we randomly select four excerpts of each episode (about 8-10% of each transcript) and request a second annotation by a random other member of the annotation team. This second annotation is reviewed by another annotator, as in the standard annotation procedure detailed in §3.2, and are made available for evaluation purposes.

The CASS score of QT30 is 0.56, signaling moderate

agreement. In contrast to higher agreement scores in other corpora of dialogical argumentation – CASS of 0.752 in Visser et al. (2019) and $\kappa = 0.75$ in Budzynska et al. (2014b)) – inter-annotator agreement for QT30 is based on a very heterogeneous but realistic dataset for quantifying annotation reliability: it features annotations by all 38 annotators which are based on a variety of experience levels due to the incremental formation of the annotation team (in contrast, Budzynska et al. (2014b) only use two annotators for measuring inter-annotator agreement).

5. Rhetorical patterns of broadcast debate

QT30 presents an extensive and rich dataset in which to explore a range of different argumentative and dialogical dynamics. Here we focus specifically on argumentation and conflicts and the patterns they trigger. Although IAT only makes use of two types of illocutionary relation, namely 'Arguing' and 'Disagreeing', and two propositional relations, namely 'Default Inference' and 'Default Conflict', we can nevertheless build a more nuanced account of argumentation and conflict using the structural context in which these relations are found. The exploration is cast here as a series of questions that are put to the data.

5.1. Are conflicts well supported?

In the first investigation we extract those conflicts in which one proposition, either the source or the target of the conflict, is supported, i.e., targeted by a 'Default Inference'. Of all conflicts in QT30 (976, see Table 1), only 200 receive support: in 58 cases it is the target of the conflict that is supported, in 142 cases, the 'Default Inference' targets the source of the conflict. This suggests that conflicts are commonly not supported in QT30: instead of substantiating them in the subsequent discourse, the majority of them stand on their own.

5.2. Do patterns of conflict differ between roles?

A more nuanced view on conflict, in particular its distribution across roles in QT30, can be directly read off the IAT graphs, which record the speaker name in the locution (see §3.1.1). Compared to 976 conflicts overall in QT30 (see Table 1), 636 conflicts are instances where speakers present a conflicting statement with one of their own previous claims. This may seem counterintuitive at first, but covers instances as in (2) from QT on 2 September 2021: Mehdi Hasan, a British journalist and author, comments on the situation of Afghani citizens seeking asylum in the UK:

(2) Mehdi Hasan [00:24:34] *Up until Monday morning the Home Office's official guidance said they could be returned to Kabul safely, which I find absurd.*

The analysis in Figure 2 shows that the sentence is split into two argumentative units ('Up until Monday morning ... they could be returned to Kabul safely' and 'which I find absurd'), with the second proposition annotated as a source of conflict to the first proposition. These types of conflicts are therefore not instances of speaker incoherence, but signal inconsistencies that individual speakers point out.

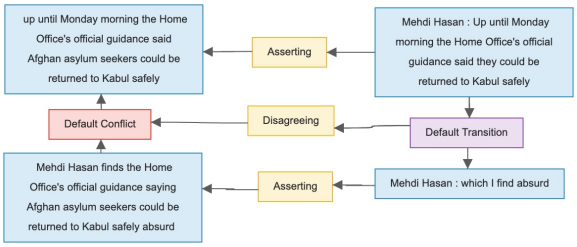

Figure 2: IAT graph for Example (2), illustrating the annotation of a speaker conflicting with their own previous claim.

It is this type of conflict that in fact constitutes the majority of conflicts in QT30, only one third of all conflicts are conflicts between different speakers. Table 4 gives an overview of the number of conflicts between roles and shows that the panel is the most combative of all roles: Their biggest share of attacks are directed at other panel members (117 out of 199, 62%) and the moderator (57 out of 199, 29%). Only 9% of attacks target the audience. Table 4 also shows that the moderator actively participates in the debate: 81 conflicts target the panel (82%), 18 (18%) target the audience. The QT30 audience, in absolute numbers, is significantly behind the panel and the moderator in the number of conflicts, in particular considering that they contribute about 25% of all locutions (see Table 2). In addition they attack each role uniformly, including speakers of their own group.

conflicts w/	moderator	panel	audience	total
moderator	–	81	18	99
panel	57	117	25	199
audience	13	16	13	42
total	70	214	56	340

Table 4: Overview of patterns of conflict across the different roles in QT30: moderator, panel members, audience members

5.3. Do patterns of support differ between roles?

In the same way as for conflicts we also shed light on the patterns of support in QT30. In the overwhelming majority, speakers support their own claims: Out of 5,025 'Default Inferences' in QT30, 4,637 support a claim by the same speaker (227 for the moderator, 3,271 for the panel and 1,139 for the audience). The panel is again the most active in substantiating

their own claims. Table 5 shows that only 121 out of 5,025 supports (2.4%) connect to a claim by a different speaker.[6] Again, members of the panel interact the most across speakers (66 out of 121, or 54%), the contributions of the audience in supporting different speakers across roles are very few (10%). Together with the results for conflicts, this suggests that speakers neither very often explicitly disagree with each other nor build alliances by way of explicitly substantiating each other's claims.

supports	moderator	panel	audience	total
moderator	–	30	12	42
panel	13	48	5	66
audience	7	6	–	13
total	20	84	17	121

Table 5: Overview of patterns of support across moderator, panel members and audience members in QT30.

However, support can also be signaled by rephrasing or restating previously mentioned content. In IAT this is encoded via the propositional relation of 'Default Rephrase', anchored via the illocutionary relation of 'Restating'. In QT30, out of 4,637 rephrases, 3,557 are used to rephrase a speaker's own material. Similar to conflicts and supports, it is only a small portion of rephrases that is uttered with the intention to relate to a different speaker's content, as shown in Table 6: again, the panel has the highest share of rephrases, mostly directed at the moderator (315 out of 650, 48%). The audience also uses rephrasing as a way to relate to the moderator, much more so than via direct conflict or support.[7]

rephrases	moderator	panel	audience	total
moderator	–	51	59	110
panel	315	49	52	416
audience	101	10	21	132
total	416	110	132	650

Table 6: Overview of patterns of rephrasing across moderator, panel members and audience members in QT30.

Overall, the number of supports and rephrases that are used to relate to material by different speakers is comparatively low and ranges between 2.4% for supports and 14% for rephrases. This is somewhat surprising, but can be explained by the core function and set-up of 'Question Time': It provides access to opinions that shape policies and enables the general public to make informed choices. Members of the panel are individuals speaking for the stakeholders they are representing, given that the panel is deliberately heterogeneous, we expect a significant amount of conflict and argumentation to be exchanged within that group. The audience is similarly heterogeneous, however, its role is rather one of pointing out individual issues than launching into extensive argumentation with the panel (or other audience members). The moderator is there to guide the exchange, shown by the comparatively high amount of rephrasing information by the panel and the audience.

[6]Table 5 shows the distribution of propositions which receive at least one support by a different speaker. Some of the propositions are supported by multiple speakers, including the same speaker, and are therefore counted as individual instances of 'Default Inference'.

[7]Again, as in Table 5, we count instances of propositions which are rephrased at least once by a speaker with a different role.

6. Summary and outlook

With QT30 we provide the largest individual corpus of analysed dialogical argumentation and the first resource of broadcast debate in which language of well-versed political rhetoric meets direct justification-seeking of the general public. Due to its size, time coverage and quality, QT30 paves the way for large-scale longitudinal analyses of rhetorical and argumentative maneuvering in broadcast political debate. The quantitative analysis in this paper is only the first step in investigating political discourse on a much larger and also more detailed scale: of particular interest lies in investigating the ways in which arguments evolve over time, information is reframed and critical questions are navigated by stakeholders in the public eye. We see significant merit in building on QT30 for identifying other pragmatic phenomena than argumentation and conflict, for instance automatic reply categorization, e.g., along the categories of 'Non-response', 'Nonanswer response', 'Answer' as in Stivers and Enfield (2010). Another avenue for further work is the automatic identification and extraction of rhetorical maneuvers such as the ABC strategy for answering critical questions ('Align, Bride, Categorise') – Paul Polman's reply in Example (1) shows an instance of this navigation strategy. Overall, the detailedness and size of QT30 offers a new data basis for a range of investigations into pragmatic phenomena and computational approaches, from topic-driven debate analysis to discourse processing and argument mining.

7. Acknowledgements

We thank the following people for the support in creating QT30 (in alphabetical order): Abbie Hamilton, Anna Mauz, Callum McGillivrey, Camilla Uhlbäck, Christopher Beckett, Clara Seyfried, Dominika Juszczak, Ella Schad, Fyfe Howard, Hye Won Yang, Joanne Frith, Jordan Robinson, Kamila Gorska, Karina Boriman, Katharina Hölzl, Kyle McWilliam, Louise Whyman, Mandela Shahto, Manos Apostolidis, Martyna Iwach, Marleen Kaesebier, Mascha Graupe, Matt Foulis, Nicole Orr, Mona Dörr, Pawel Knorps, Paweł Szczepański, Qi Yu, Rebecca Raddatz, Sahra Styger, Stephanie Gustedt, Stefanie Müller, Tuuli Jantti, Wassiliki Siskou, Weronika Wlodarek, Xiyuan

Gao, Zlata Kikteva, Zofia Trafas. We also thank Rory Duthie and John Lawrence for their technical support and Elaine McIntyre for covering the substantive administrative effort related to the endeavour.

8. Bibliography

Asher, N. and Lascarides, A. (2003). *Logics of Conversation*. Cambridge University Press.

Budzynska, K., Janier, M., Reed, C., Saint-Dizier, P., Stede, M., and Yaskorska, O. (2014a). A model for processing illocutionary structures and argumentation in debates. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, pages 917–924, Reykjavik, Iceland, May. European Language Resources Association (ELRA).

Budzynska, K., Janier, M., Kang, J., Reed, C., Saint-Dizier, P., Stede, M., and Yaskorska, O. (2014b). Towards argument mining from dialogue. In *Proceedings of the Fifth International Conference on Computational Models of Argument (COMMA 2014)*, pages 185–196. IOS Press.

Budzynska, K., Janier, M., Reed, C., Saint-Dizier, P., Stede, M., and Yakorska, O. (2014c). A model for processing illocutionary structures and argumentation in debates. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, may.

Budzynska, K., Janier, M., Reed, C., and Saint Dizier, P. (2016). Theoretical foundations for illocutionary structure parsing. *Argument & Computation*, 7(1):91–108.

Degano, C. (2016). Corpus linguistics and argumentation: Retrieving argumentative patterns in uk prime ministerial debates. *Journal of Argumentation in Context*, 5:113–138, 10.

Duthie, R., Lawrence, J., Budzynska, K., and Reed, C. (2016). The CASS technique for evaluating the performance of argument mining. In *Proceedings of the Third Workshop on Argument Mining (ArgMining2016)*, pages 40–49, Berlin, Germany, August. Association for Computational Linguistics.

Gemechu, D. and Reed, C. (2019). Decompositional argument mining: A general purpose approach for argument graph construction. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 516–526, Florence, Italy, July. Association for Computational Linguistics.

Haddadan, S., Cabrio, E., and Villata, S. (2019). Yes, we can! mining arguments in 50 years of US presidential campaign debates. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 4684–4690, Florence, Italy, July. Association for Computational Linguistics.

Hautli-Janisz, A., Budzynska, K., McKillop, C., Plüss, B., Gold, V., and Reed, C. (2022). Questions in argumentative dialogue. *Journal of Pragmatics*, 188:56–79.

Hess-Lüttich, E. W. (2007). (pseudo-)argumentation in tv-debates. *Journal of Pragmatics*, 39(8):1360–1370. Argumentation in Dialogic Media Genres—Talk Shows and Interviews.

Janier, M. and Reed, C. (2016). Corpus resources for dispute mediation discourse. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, pages 1014–1021, Portorož, Slovenia, May. European Language Resources Association (ELRA).

Janier, M., Lawrence, J., and Reed, C. (2014). Ova+: An argument analysis interface. In *Computational Models of Argument: Proceedings of COMMA*, volume 266, pages 463–464.

Konat, B., Lawrence, J., Park, J., Budzynska, K., and Reed, C. (2016). A corpus of argument networks: Using graph properties to analyse divisive issues. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, pages 3899–3906, Portorož, Slovenia, May. European Language Resources Association (ELRA).

Lawrence, J. and Reed, C. (2020). Argument mining: A survey. *Computational Linguistics*, 45(4):765–818.

Luginbühl, M. (2007). Conversational violence in political tv debates: Forms and functions. *Journal of Pragmatics*, 39(8):1371–1387. Argumentation in Dialogic Media Genres—Talk Shows and Interviews.

Peldszus, A. and Stede, M. (2013). From argument diagrams to argumentation mining in texts: A survey. *Int. J. Cogn. Inform. Nat. Intell.*, 7(1):1–31, jan.

Prasad, R., Dinesh, N., Lee, A., Miltsakaki, E., Robaldo, L., Joshi, A., and Webber, B. (2008). The Penn Discourse TreeBank 2.0. In *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08)*, Marrakech, Morocco, May. European Language Resources Association (ELRA).

Ruiz-Dolz, R., Nofre, M., Taulé, M., Heras, S., and García-Fornes, A. (2021). Vivesdebate: A new annotated multilingual corpus of argumentation in a debate tournament. *Applied Sciences*, 11(15).

Searle, J. and Vanderveken, D. (1985). *Foundations of Illocutionary Logic*. Cambridge: Cambridge University Press.

Searle, J. (1969). *Speech acts: An essay in the philosophy of language*. Cambridge: Cambridge University Press.

Slonim, N., Bilu, Y., Alzate, C., et al. (2021). An autonomous debating system. *Nature*, 591:397–384.

Stivers, T. and Enfield, N. (2010). A coding scheme for questionÄìresponse sequences in conversation. *Journal of Pragmatics*, 42(10):2620 – 2626. Question-Response Sequences in Conversation across Ten Languages.

Visser, J., Duthie, R., Lawrence, J., and Reed, C. (2018). Intertextual correspondence for integrating corpora. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki, Japan, May. European Language Resources Association (ELRA).

Visser, J., Konat, B., Duthie, R., Koszowy, M., Budzynska, K., and Reed, C. (2019). Argumentation in the 2016 us presidential elections: annotated corpora of television debates and social media reaction. *Language Resources and Evaluation*, Feb.

Visser, J., Konat, B., Duthie, R., Koszowy, M., Budzynska, K., and Reed, C. (2020). Argumentation in the 2016 us presidential elections: annotated corpora of television debates and social media reaction. *Language Resources and Evaluation*, 54:123–154.

Walton, D., Reed, C., and Macagno, F. (2008). *Argumentation Schemes*. Cambridge University Press.