Journal of Pragmatics 188 (2022) 56–79

Contents lists available at ScienceDirect

Journal of Pragmatics

journal homepage: www.elsevier.com/locate/pragma

Questions in argumentative dialogue☆

Annette Hautli-Janisz [a, *], Katarzyna Budzynska [b], Conor McKillop [c], Brian Plüss [c], Valentin Gold [d], Chris Reed [c]

[a] Department of Linguistics, University of Konstanz, Germany
[b] Warsaw University of Technology, Poland
[c] Centre for Argument Technology, University of Dundee, UK
[d] Methodenzentrum Sozialwissenschaften, University of Göttingen, Germany

ARTICLE INFO

Article history:
Received 25 February 2020
Received in revised form 19 October 2021
Accepted 29 October 2021
Available online 15 December 2021

Keywords:
Empirical pragmatics
Types of questioning
Natural language argumentation

ABSTRACT

Despite questions having a long-standing history in theoretical linguistics, the interface between empirically grounded corpus linguistics of questioning behaviour and analytically driven pragmatic theory of question structure and context has received significantly less attention. This paper aims to contribute to this field of research by showing that a four-way categorisation into question types, namely, pure questioning, assertive questioning, rhetorical questioning and challenge questioning, allows us to capture and represent questions in over two million words in natural language argumentative dialogue. In this type of dialogue questioning has been claimed to serve as the engine that drives the shape and development of a discourse. Our investigation covers three genres of argumentative discourse in which questions play a key role, namely political debates, moral dilemmas and sessions of participatory deliberative democracy. Through deep algorithmic analysis of the data, we test a variety of hypotheses from argumentation and linguistic theory, clearly demonstrating for the first time that (particular types of) questions directly catalyse argument structure and that the illocutionary consequences of non-canonical questions are much more varied than previously thought.

1. Introduction

Questions and their function in discourse are a complex phenomenon, with theoretical linguistics and in particular pragmatics trying to distil key functions of questions and their interaction with intonation and lexical or syntactic marking. Despite (or, partly due to) their interesting complexity, empirically-driven theories of questions are still under-represented in pragmatics, also due to the comparatively limited resources that explicitly deal with different categories of questioning and attempt to model the intent of the user in posing a question. This is even more so the case in specific types of discourse, such as natural argumentative discourse, where questions are known to be key in advancing a discussion.

☆ The work reported in this paper was supported by the VolkswagenStiftung under grant 92,182, by the Polish National Science Centre under grant 2015/18/M/HS1/00620, by the POB CyberDS of Warsaw University of Technology within the Excellence Initiative: Research University (IDUB) programme under grant 1820/1/Z01/POB3/2021 and in part by the Deutsche Forschungsgemeinschaft (DFG - German Research Foundation) under the Excellence Strategy of the German federal and state governments – EXC-2035/1–390681379 [Gefördert durch die Deutsche Forschungsgemeinschaft (DFG) im Rahmen der Exzellenzstrategie des Bundes und der Länder – EXC-2035/1–39068137a9].
* Corresponding author.
E-mail address: annette.hautli@uni.kn (A. Hautli-Janisz).

https://doi.org/10.1016/j.pragma.2021.10.029

In this paper we present a taxonomy for questioning that has been applied to corpora comprising over two million words in natural language argumentative dialogue. Based on this annotation, our aims are to (i) provide a question taxonomy that captures the function of question in argumentative dialogue and can be applied robustly; (ii) show the use and effect of questions in dialogue, and (iii) show what impact questions have on argumentation. Our question annotation is part of Inference Anchoring Theory (IAT) (Budzynska et al., 2014; 2016; Katarzyna et al., 2014), a theoretical framework whose primary focus lies on analysing dialogical argumentation. IAT has been at the core of a large body of work in philosophy (Reed, 2011; Reed and Budzynska, 2011, 2012) and argument technologies (Bex et al., 2017; Reed et al., 2017) and has so far been applied to more than 2.3 million words of dialogue across fifteen languages (English, German, Chinese, Polish etc.) – aall openly available at corpora.aifdb.org. This large-scale dataset allows us to shed new empirical insight into the functions of questions (and their responses) across genres of naturally occurring, dialogical argumentation in which questions play a very prominent role.

Based on loops of refining the annotation, we have now arrived at a stable categorisation of questioning which can be reliably applied across languages and genres, including, for instance the televised US 2016 presidential debates (Visser et al., 2020). It distinguishes four intentions: *Pure Questioning* (Speaker S seeks information from hearer H), *Challenge Questioning* (S asks H to justify her opinion), *Rhetorical Questioning* (S makes an assertion in the guise of a question) and *Assertive Questioning* (S not only seeks H's opinion, but also indirectly publicly declares their own opinion).

The data under investigation covers a wide range of argumentative settings, from political debates in the US2016 presidential election and moral dilemma discussions to sessions of real and experimental participatory deliberative democracy. By way of algorithmic methods we shed light on the use and effect of questions in these argumentative settings, yielding statistical profiles that show that questioning is not only used to update the information state of the participants, but is also vital in inviting other opinions on a standpoint and serves as a means of expressing a speaker's own standpoint by saving their own and the face of the interlocutor. We also demonstrate that a variety of questions do indeed catalyse argument structure, contrary to the prevailing view in argumentation theory whereby it is predominantly challenge questions that trigger argumentation (Freeman, 1991). Finally, the empirical analysis shows that the illocutionary consequences of non-canonical questioning, i.e. those questions that do not elicit information from an interlocutor, are much more frequent and varied than previously thought.

The paper is structured as follows. After presenting related work in §2 and §3 provides an overview of IAT, the different types of communicative structures it incorporates and the pragmatic 'glue' between dialogue structure and argument structure (including the logical relation of inference). In §4 we present the different categories of questioning and their analyses in detail. §5 describes the corpora underlying the investigation and provides statistics on the distribution of types of questioning across argumentative settings. In §6 we use deep algorithmic analyses in order to showcase the pragmatic effect of individual question types on the ongoing discourse, which are discussed in §7 in the light of several hypotheses from previous literature.

2. Related work

In argumentation theory, questions are considered crucial: for Aristotle what constitutes arguing is inevitably related to having doubts, and therefore asking questions (Aristotle, 2009). Walton (1988, p. 3) states that "no theory of argument could be adequate without dealing with the role that questions play in arguments" (Freeman, 1991), claims that argumentation depends upon critique, challenge, for instance realised by posing of 'Why questions'. Rhetorical questions are considered to be effective persuasion devices (see (Frank, 1990) and the references therein). (Johnson, 2014, p. 86) sees questions (and their answers) as the driving factors of the dynamics of argumentation, an assumption that forms a key part of the new computational field of argument mining (Lawrence and Reed, 2019), which is increasingly relying upon theoretical analysis and corpus techniques from linguistics and pragmatics in particular (Katzav and Reed, 2008). At the same time, (Hitchcock, 2020) shows that questions can be argued for, i.e. the argument provides a reason for asking the question and thus tries to establish that it needs to be answered.

In theoretical linguistics questions also have a long history and current research tries to distil those properties of grammar (from intonation to syntax and semantics) that determine the communicative functions of questions in discourse (see (Dayal, 2016)). A particularly complex case are non-canonical questions, often subsumed under the term 'rhetorical question' (RQ), a type of question which is generally considered not to elicit information from an interlocutor, but rather to make an assertion. Regarding the syntax-semantics-pragmatics interface, Sadock (1971) sets out to find explicit syntactic evidence for RQs, ("If an interrogative clause is introduced by 'after all' or followed by a 'yet'-clause, then it can only be a RQ"). Heritage (2002) shows that questions starting with 'Isn't it', 'Doesn't this … ', and 'Don't you … ' are used to assert an opinion rather than asking for information. Guerzoni (2003) shows that the presuppositions triggered by negative polarity items and the particle 'even' create a "negative bias" in a question (p. 61), signalling "the speaker's expectation for (bias towards) a negative answer". However, in sum, "it remains largely unclear what nonstandard question interpretations there are and where they have their sources." (Obenauer, 2004, p. 1).

Regarding the function of questions in discourse, a multitude of work in pragmatics has focused on non-canonical questioning: Truckenbrodt (2006) sees questioning not as means to request information but to establish mutual

knowledge. Rohde (2006) and Caponigro and Sprouse (2007) show that information-seeking questions and RQs only differ at the pragmatic level and are identical at the range of possible answers. Biezma and Rawlins (2017) employ the Question Under Discussion (QUD) model (Roberts, 2012) to show how highly context-dependent the interpretation of '…, or what?' questions is. RQs are also the focus of attention in conversation analysis: Quirk et al. (1985) describe rhetorical questions as having "the force of a strong assertion" (p. 825), whereas (Brown and Levinson, 1987) treat RQs as hedged assertions, directives or criticism. Koshik (2003) shows that in an environment of disagreement, accusation and complaint, wh-questions can be heard as challenges. Schaffer (2005) shows that RQs can be used as retorts to a preceding information-seeking question, similar to (Cerović, 2016)'s findings that RQs used as answers can signal opposition.

Regarding a taxonomy of questions, Freed (1994) proposes a continuum that ranges between purely information seeking and information conveying questions. The Switchboard corpus (Jurafsky et al., 1997; Calhoun et al., 2010) encodes information-seeking questions according to their syntactic property (e.g. yes-no-questions, wh-questions) rather than pragmatic function, whereas non-information seeking questions are summed up in the category 'rhetorical questions'. In the MapTask coding scheme (Carletta et al., 1997), rhetorical questions do not form a discrete category, but fall into two default categories of questions that are outside the spectrum of questions confirming information and checking speaker alignment, namely Query-W moves – these types of questions are "rare enough" (p. 19, Carletta et al. (1997) to require division into further subcategories Stivers and Enfield (2010) propose a cross-linguistic categorisation of questions into seven different 'social actions', among them 'Request for Information', 'Outloud' ("delivered to no one in particular", p. 2623) and 'Rhetorical Question' ("Questions that may seek a response but do not seek an answer", Stivers and Enfield, 2010), using multimodal input (audio and video) to refine their categorisation and disambiguation. As with any large-scale corpus linguistic task, the field is confronted with "a trade-off between usefulness and ease or consistency of coding" (Carletta et al., 1997, p. 15), in particular for a phenomenon as vague and complex as questions (and responses).

A number of taxonomies for question categorisation also take into account categories of responses: Stivers and Enfield (2010) use the four categories of 'Non-response', 'Nonanswer response', 'Answer', and 'Can't determine' (p. 2624) for spontaneous, naturally occurring conversation, showing some overlap with the six-way distinction of (Berninger and Garvey, 1981) with 'Possible answers', 'Indirect answers', 'Confessions of Ignorance', 'Clarification requests', 'Evasive replies' and 'Miscellaneous' (targeted at nursery school child conversation). Łupkowski and Ginzburg (2016) propose a taxonomy for query responses, i.e. responses that answer a query with a query, for instance 'Requests for underlying motivation' or 'I ignore your question' – the latter being identical to (Berninger and Garvey, 1981)'s 'Confession of ignorance' type. For the task-based interactions in MapTask, the BEE corpora and informal conversations in the BNC corpus, Ginzburg et al. (2019) propose a two-way distinctions of responses into 'Answers' and 'Non-answers', with more fine-grained categories subsumed under them, for instance 'Direct answers' versus 'Indirect answers' for types of answers and 'Clarification requests', 'Change-the-topic response' and 'Ignore' for types of non-answers.

Despite the variety of previous work, we find commonalities between the taxonomies of questions (and responses): For instance, Łupkowski and Ginzburg (2016)'s 'Requests for underlying motivation' (though these questions are responses to questions) correspond to what we call challenge questions: they request further reasoning from the hearer regarding an earlier point. However, instead of fine-grained distinctions such as (Stivers and Enfield, 2010)'s questions that suggest, offer or propose ('Did you want some?') or outloud questions (which are delivered to no one in particular, e.g. 'Now where did I put my keys?'), our taxonomy builds on the notion of *underspecification*: We propose four higher-level categories of questions (and three categories of responses) which can be input for more fine-grained distinctions, some of them possibly even automatically derivable. This also pertains to labelling the syntactic structure of the question: In contrast to the Switchboard corpus, we do not explicitly the syntactic configuration, e.g. *wh*-in-situ ('what' in 'Who bought what?') versus *wh*-fronted ('what' in 'What did John buy?') questions. This underspecification of question type and syntactic configuration is also guiding our treatment of inference structure and propositional content, allowing us to stay away from making "too many semantic distinctions [that] make coding difficult" (Carletta et al., 1997, p. 15). In addition, some of these distinctions, for instance in Stivers and Enfield (2010), are only possible with multimodal input which we do not take into account in IAT.

For the purpose of showing the use and effect of questioning in naturally occurring argumentation, we present three unique data sets of argumentative dialogue, comprising political debates in the US2016 presidential election through to moral dilemma discussions broadcast to nationwide public audiences and sessions of real and experimental participatory deliberative democracy. These constitute the largest datasets of closely annotated argumentative dialogue currently available. By annotating four different communicative uses of questions, we aim to bridge the gap between the abundance of theoretical linguistic work on questions (and responses), previous empirical investigations in the field and the constraints for reliably annotating a large-scale and cross-linguistic dataset of argumentative data across languages and genres. By tying the different strands of research together, we are able to support (and refute) a number of hypotheses in argumentation theory regarding the use of questions in such discourse.

3. Inference Anchoring Theory

The central motivation of Inference Anchoring Theory (IAT) (Budzynska et al., 2014; 2016) is to provide a theoretical framework for analysing dialogical argumentation. As such, IAT sets out to answer the question of where argumentation *comes from* in dialogical interaction and acts as a theory-neutral scaffolding that integrates different communicative structures, namely *dialogue structure*, *argument structure* (including inference understood in the logical manner) and *illocutionary forces* such as asserting, suggesting or promising, to anchor argument structure in dialogue structure. One of the primary loci of argumentation is precisely in the asking of, and responding to, questions of particular sorts. In general, IAT

IAT remains agnostic as to the specific formal analysis that is used, not only for questions and their context, but also regarding the analysis of propositional content (e.g. first-order predicate calculus) or dialogue procedure (e.g. logic programming). It also remains underspecified in terms of the model of argumentation and inference, not least because entire fields of academic endeavour have applied different strategies to the development of models of these connections.[1] Therefore, the communicative structures required to analyse dialogical argumentation are not instantiated with any theoretically or domain-driven assumptions of how communication unfolds. In particular, there is careful bracketing out of the ways in which the context of discourse impinges upon its structure (variously explored in the literature as activity-based communication analysis (Allwood, 2000) in pragmatics and dialogue games in philosophy (Walton and Krabbe, 1995) *inter alia*). This allows IAT analysis to adduce a variety of different theories of dialogical context in accounting for such discourse organisation.

The interplay between the communicative structures is illustrated by way of example (1), adapted from an excerpt from an experimentally controlled discussion on whether or not to allow fracking in Germany (Gold et al., 2017)[2]: In (1a), Luca puts forward the claim that fracking is worth destroying our nature for. This claim is challenged by Alex in (1b) who invites justification for that claim which Luca offers in his move in (1c): she thinks that there is no untouched nature anyway. While there is a clear inferential relation between the content of (1a) and (1c), there is also a dialogical dimension to which this inferential structure relates. The interplay of the communicative structures involved in creating an argument in dialogue, namely dialogue structure, argument structure and illocutionary structure, are detailed in the following.

(1) a. Luca: *Fracking is worth destroying our nature for.*
b. Alex: *Why?*
c. Luca: *We don't have untouched nature anyway.*

3.1. Dialogue structure

The right-hand side of the diagram in Fig. 1 represents the dialogue structure of (1), consisting of the speakers' moves, i.e. the actual utterances or *locutions*, and the connections between them. This structure captures which types of utterances can license or require the making of other utterances – the underlying motive for dialogue games, which specify a set of rules of the functional relations between moves ('dialogue protocol'). There is a rich variety of these dialogue games in, for example, philosophy (Mackenzie, 1990; Walton and Krabbe, 1995), jurisprudence (Prakken, 2005) and AI (Reed et al., 2017). However, these dialogue games are rarely exhaustive in specifying the relationships between moves. Therefore, we leave the exact nature of the transition between locutions underspecified: We are neither interested in the details of the dialogue protocols nor the characterisation of the rules from which they are composed, so we label all instances of these dialogue rules simply as *Default Transitions*. In (1), the Default Transition holds between the first and the second and the second and the third move as each move licenses the succeeding move. Such transitions are not always consecutive: it is perfectly common for interlocutors to return to earlier points of their own or of others, and in such cases, the transition captures the longer distance functional relationship. The longer the dependency, the more likely it is to be explicitly signalled on the linguistic surface ('returning to my earlier point,' 'And ninthly,' etc.).

[1] Three broad traditions can be distinguished (our goal here is not to provide a thorough review of the extensive literature, the subclassifications, and the overlaps, but to give a very broad-brush intuition of how fundamentally different accounts can be used interchangeably within the IAT framework). The first is the broadly mathematical, in which the formal-syntactic pattern is the driving force. In Modus Ponens, for example, one cares little of the internal structure of antecedent and consequent, only in the veracity of the major premise that connects them. The second tradition is the broadly linguistic, in which inference is intimately connected to the meaning of inferential constituents. In textual entailment, for example, it is a (typically first order predicate) account of the semantics of two canonical text spans that establish the entailment relation. The third tradition can be placed somewhere between the previous two, merging some aspects of propositional meaning with some aspects of informational structure. Argumentation theory encompasses many such accounts of inferential connection, but one that has been gaining particular momentum in philosophy, linguistics and AI is the model of presumptive argumentation schemes (Walton et al., 2008) which capture stereotypical patterns of reasoning that (arguably (Groarke, 1999) do not yield to a deductivist analysis). IAT allows any of these three accounts to be 'plugged in' as a specification of the inferential connection, and indeed can blend the different approaches ((Pease et al., 2017) for example, show how the informal, presumptive reasoning of natural dialogue between mathematicians can be connected to the formal, mathematical proofs that those dialogues yield).

[2] The original German example is given in example (7) and was simplified and translated for the sake of clarity. We will come back to it in §4.3.

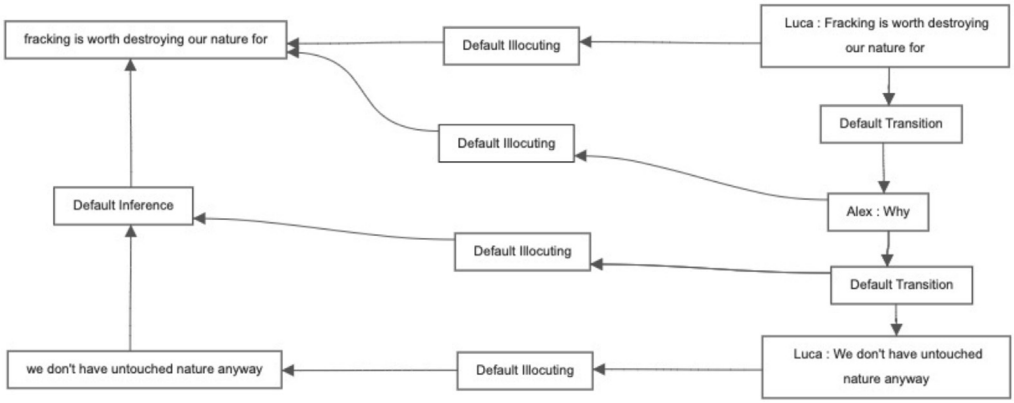

Fig. 1. Schematic IAT diagram for Example (1).

3.2. Argument structure

In IAT, three argument structures are considered: inference (support), conflict (attack) and rephrase. These hold between propositions, i.e. grammatically and semantically reconstructed contents of the utterances that people make. Propositions are, by convention, on the left-hand side of the diagram and are created during manual analysis based on the content of the propositional reports of discourse events that constitute IAT's locutions, recorded on the right. The goal for IAT is to have propositions that are understandable without context, and as a result, their formulation may include reconstruction such as the resolution of pronouns and unpacking of ellipsis.

This level of communicative structure also provides a linguistic connective fabric that remains underspecified with respect to theories of inference. Whilst argumentation theory has produced a wide range of models for representing inference relation – for instance, models of propositional logic, Bayesian reasoning and presumptive reasoning using argumentation schemes such as, e.g. Walton et al. (2008)'s scheme set in Lawrence and Reed, 2016 or Wagemans (2016)'s in (Visser et al., 2018), IAT avoids building in dependence upon any one such theory. At the same time, IAT also avoids making any assumptions about the formal representation of the propositional content, which is deliberately left in the form of natural language and not, for instance in first-order logic.[3]

In example (1), only the first and the last moves have propositional content – Alex's move in (1b) does not add propositional content but refers to a previous proposition, which is why the contribution of this locution is only captured via an illocutionary connection and not in its own proposition (more on this in §4.4). As shown in Fig. 1, the relation between the two propositions is recorded as one of *Default Inference*, i.e. an (undifferentiated, hence 'Default') inference between a premise ('we don't have untouched nature anyway') and a conclusion ('fracking is worth destroying our nature for'). Relations between propositions can also take the form of a *Default Conflict*, representing a (similarly undifferentiated) conflict between propositions (i.e. independent from any theory of conflict such as e.g. an abstract argumentation framework (Dung, 1995)), or a *Default Rephrase* where a proposition reformulates some earlier proposition. This latter connection is important in the treatment of questions in IAT and will be discussed in more detail in §4.

3.3. Illocutionary structure

The third type of connection serves as the glue between dialogical structure and inferential structure and captures the illocutionary force of an utterance. For this relation, there is a raft of work in the philosophy of language which could be imported. In the version of IAT employed in this paper, we instantiate illocutionary connections with the concept of illocutionary forces of (Searle, 1969) and (Searle and Vanderveken, 1985), such as asserting, suggesting or promising. The illocutionary scheme set adopted for the current work is a simplified one in which there are no commissives or expressives and just two simple types of directives (questions and challenges). Table 1 gives an overview of the illocutionary connections that are employed. As with the previous two types of relations, the connection between dialogical and inferential structure can be left underspecified – and marked simply as an undifferentiated type of illocution (*Default Illocuting*) or alternatively can be instantiated with one of a variety of theories capturing speaker intention.

[3] In this way IAT similarly avoids commitment to a particular set of advantages and disadvantages associated with a particular formal approach such as first order predicate calculus, Montague semantics or fuzzy set theory. Mirroring, however, the approach taken in (Ginzburg et al., 2019), we are "not really interested in the extension of the semantic entities (primarily propositions and questions)", we are rather "interested in the class each such entity is classified as since that is what determines the subsequent contextual evolution" (p. 321).

Table 1
Illocutionary structures used in IAT

Illocutionary connection		Description
Asserting		S communicates her opinion
Arguing		S defends a standpoint (a claim, a conclusion)
Agreeing		S expresses a positive reaction, i.e. S declares that S shares H's opinion
Disagreeing		S expresses a negative reaction, i.e. S declares not to share H's opinion
Restating		S expresses an opinion that rephrases what was uttered previously
Default		The connection is underspecified or does not match any of the other illocutionary structures
Questioning	Pure	S is directing hearer H to supply information
	Rhetorical	S is predominantly expressing an opinion that she expects H will find uncontentious
	Assertive	S may be directing H to supply information, but is also expressing an opinion
	Challenge	S declares that he is seeking (asking about) the grounds for H's opinion

Inferential structure can be *anchored* in dialogical structure in two different ways: If the locution on its own is enough for a full reconstruction of the illocution and its content – e.g. both of Luca's *Asserting* moves in (1) – the illocutionary connection is anchored in the locution. In cases where the locution requires content from previous material, as in Luca's argument structure between (1a) and (1c), we anchor the illocutionary force not in the locution but instead in the application of the rule of dialogue that has been followed, i.e. in the transition. Here, the propositional target of the illocutionary force is indexical; it is dependent in part upon the preceding locution. A good test regarding the anchoring is this one: Imagine you entered the room just before Luca's *We don't have untouched nature anyway* move. It would be impossible to identify the argumentative structure if the preceding context was unavailable.

Fig. 2 shows the IAT structure for example (1) with instantiated illocutionary connections according to the scheme in Table 1: Luca asserts that 'fracking is worth destroying our nature for' and that 'we don't have untouched nature anyway' (Asserting). Alex does not assert information, she challenges the content of Luca's first assertion, with 'Challenging' anchored in the locution of (1b) and pointing to the proposition of (1a) because that is the content she is challenging. The Default Inference between Luca's moves is anchored in the transition between move (1b) and (1c) with the illocutionary force of *Arguing*, i.e. only by taking into account the preceding context it becomes clear that the intention of Luca's second assertion is to substantiate his claim made in the first move.

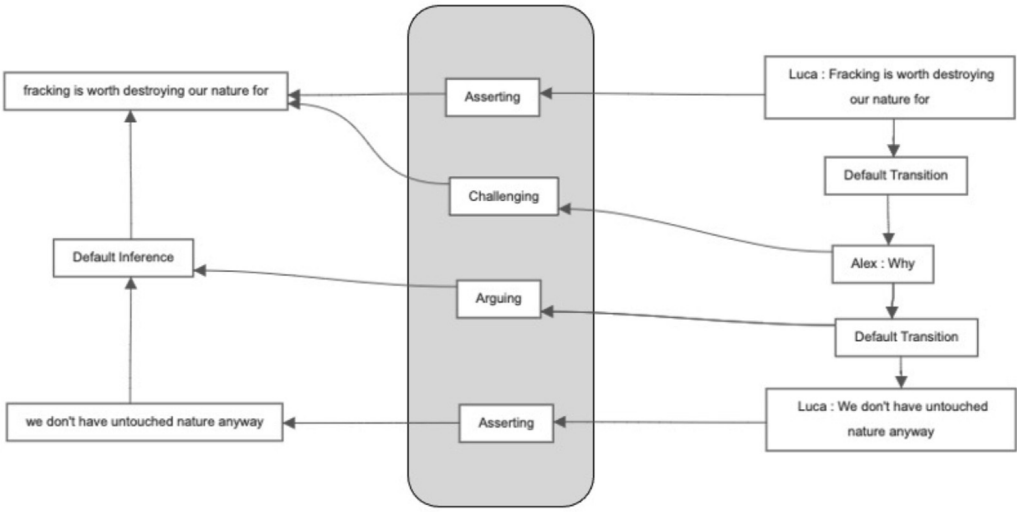

Fig. 2. IAT diagram with highlighted illocutionary structure for example (3).

3.4. Annotating IAT with OVA+

IAT analyses are produced with OVA+ (Online Visualisation of Argument – http://ova.arg.tech/), an open-source online interface for the analysis of argumentation in dialogues. OVA+ allows for a representation of the argumentative structure of a text as a directed graph (Janier et al., 2014). The user interface of OVA+ is shown in Fig. 3. Annotators work on a transcript of the dialogue, select relevant segments which automatically become locutions in the dialogue structure and propositions in the argument structure and connect them with the appropriate illocutionary connections (as specified in Table 1). The

propositional relations (Default Inference, Default Conflict, Default Rephrase) and dialogue relations (Default Transitions) are annotated in the next step.

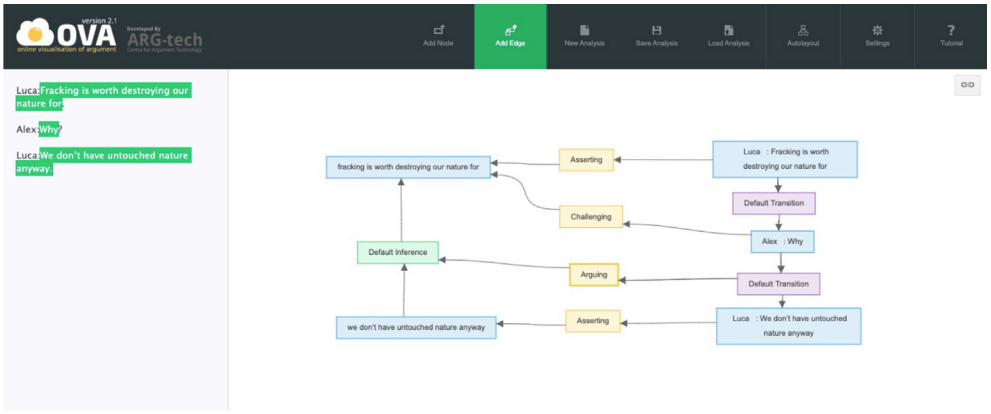

Fig. 3. The OVA annotation interface.

Manual annotation (Visser et al., 2019) and increasingly more sophisticated automatic argument mining (Stede and Schneider, 2018; Lawrence and Reed, 2020) has produced corpora in the Argument Interchange Format (AIF) (Chesnevar et al., 2006) which can be used in the Argument Web (Bex et al. 2017; Reed et al., 2017) and for argument evaluation and analytics (Lawrence et al., 2016). OVA+ can export analyses as images and files in the Argument Interchange Format (AIF), and also store these in AIFdb, the largest online database of openly available argument data, with almost 2.3 million words of dialogue across fifteen different languages (English, German, Chinese, Polish, Pashtu, Portuguese, etc. – see Table 2 for up-to-date statistics on AIFdb contents). Once in AIFdb, analyses can be grouped into corpora (corpora.aifdb.org) for distribution, evaluation and analytics, and broader access via web services (ws.arg.tech). Collectively, these technologies make up the Argument Web. For an overview of the Argument Web and its significance, see (Reed et al., 2017).

Table 2
AIFdb statistics at the time of writing – aifdb.org.

Node type	Count
Propositions	101,579
Locutions	105,342
Inferences	92,753
Conflicts	31,426
Rephrases	12,935
Illocutionary connections	171,821
Transitions	81,615
Total word count	2,292,040

So far, we have given an overview of the general architecture of IAT. In the following section, we illustrate the annotation of the different types of questions in IAT, both with respect to their syntactic structure and their discourse function.

4. Questions in IAT

The categorisation of questions in IAT is dictated primarily by the goal to uncover, surface and track – ultimately with computational means – argumentation in natural communication. In the genres analysed so far (see §5), we identified four types of questions with distinctive argumentative profiles. The most frequent type of questioning in our data are information-seeking questions ('Pure Questioning' – PQ) as they are used solely to seek and provide information without any intention of creating an argument structure of inference or conflict. If the speaker asks 'Is fracking worth destroying our nature for?' and the hearer responds with 'Yes' or 'No', then neither arguing nor disagreeing can be created between the question and the answer. Instead, the propositional content of the question and the propositional content of the response is connected in IAT through an *argument structure of rephrase* in order to capture the intended resemblance between the content of these two propositions. The criteria regarding the communicative intention which a speaker has in performing a given type of question;

the reaction to the performance of the question which the speaker expects from the hearer; and the argument structure which might be anchored between the speaker's question and the hearer's reaction are summarised in Table 3 and described in more detail in §4.1.

Table 3
Criteria for distinguishing question types.

Question type	Speaker's intention	Intended hearer's reaction	**Anchored argument structure**
Pure	Seeking information	Providing information	Rephrase
Rhetorical	Expressing opinion	(Tacit) agreement	None
Assertive	Seeking information & expressing opinion	Providing information	None or conflict
Challenge	Seeking justification	Providing justification	Inference

We further encounter many instances of another type of question studied in the literature, namely rhetorical questions (RQS). In this case, the speaker expresses her opinion using interrogative syntax without any expectation that the hearer will respond verbally. Instead, the speaker anticipates that the hearer will find this opinion uncontentious and will, consequently, tacitly agree with it. Of course, it is possible for the hearer to react to a rhetorical question against the speaker's expectation, and as a result create structures of inference, conflict or rephrase. Yet such a reaction and accompanying argument structure is not intended by the questioner, thus we do not include these reactions in the criteria for distinguishing this type of questions (see Table 3 and §4.2).

We then identify a type of question situated within (Freed, 1994)'s continuum between information-seeking and information-conveying questions (see §2). With assertive questions (AQS), e.g. 'Isn't fracking worth destroying our nature for?', a speaker seeks information from the hearer, but at the same time expresses her own opinion on the queried subject (i.e. that fracking is worth destroying our nature for). This type of question turns out to be frequently used in natural communication (see §5 for the empirical study). In case of a positive response, e.g. when the hearer says 'Yes' or 'Fracking indeed is worth destroying the nature', the speaker agrees with the assertion and no argument structure is created. In case of a negative response, e.g. when the hearer says 'No' or 'Fracking is definitely not worth destroying the nature', she disagrees with the speaker, and an *argument structure of conflict* is generated (see §4.3 for the details of specifying the category of assertive questions, and Figs. 8 and 9 for the representation of argument structures created by this type of questions in IAT). Some models of communicative function make a compositional assumption, *viz.* that the character of different types of questions can be assembled from simpler types – pragma-dialectics, for example, makes this assumption, and relies upon some compositional components being tacit to yield speech acts that capture different nuances of questioning (van Eemeren and Grootendorst, 1992, pp. 26–33). There is evidence, however, that not all questions are decompositional in this way, and that the felicity conditions that are associated with, for example, questions in which there is a suggested assertion, do not include the felicity conditions for assertion simpliciter: in Assertive Questioning, the conditions are diluted.

Finally, we find instances of Why-questions to which argumentation theory gives special credit, as it is assumed that so-called challenges are main triggers of argumentation in a dialogical interaction (*cf.* Freeman (1991)). In performing a challenge question such as, e.g. 'Why is fracking worth destroying our nature?', the speaker seeks justification from the hearer. If the hearer fulfils this expectation and responds by saying, e.g. 'Fracking is crucial to maintain a sustainable local economy', an *argument structure of inference* is created with the conclusion 'Fracking is worth destroying our nature' and the premise 'Fracking is crucial to maintain healthy local economy' (see §4.4).

From a syntactic point of view, questions come in different configurations: polar (yes/no) questions ('Did you finish the paper?'), alternative questions ('Did you finish the paper or not?'), wh-questions ('Why didn't you finish the paper?'), tag questions ('He finished the paper, hasn't he?') and embedded questions (e.g. 'I wonder why he didn't finish the paper.'). While polar, alternative and wh-questions are dealt with by specific configurations of the propositional content (see more details in the remainder of this section), tag questions rarely occur in our dataset and are treated as regular assertions. And in cases like "Speaker A: Banana's are yellow. Speaker B: Banana's are yellow. Speaker A: Yes, they are." we have a question without interrogative syntax/marking, but we would still annotate it as a question because speaker intention is unambiguous. In the case of embedded questions, we only deal with the content of the question and discard the epistemic verb in the front. This is due to IAT's main focus on argumentation: Only if main clause and subordinate clause have discrete argumentative function, i.e. serve as individual units in an argument (as a premise and a conclusion, for example), they form separate locutions. A frequent case are if-clauses (though not always), where in 'If it rains, I will take the bus home.', the subordinate if-clause is the premise and the main clause is the conclusion of the argument. In the example below, Hillary Clinton used a construction 'you've got to ask yourself, why won't he [Donald Trump] release his tax returns?' in one of the US 2016 TV election debates (US2016tv, map 10628). In this case, the annotator treated the clause 'You've got to ask yourself' as an epistemic statement (those are not included in the propositional content) and the embedded question as a rhetorical question which expresses the opinion that 'Donald Trump doesn't want to release his tax returns'.

(2) Hillary Clinton: *For 40 years, everyone running for president has released their tax returns. […] We know the IRS has made clear there is no prohibition on releasing it when you're under audit. So you 've got to ask yourself, why won't he release his tax returns? And I think there may be a couple of reasons. First, maybe he's not as rich as he says he is. Second, maybe he's not as charitable as he claims to be. Third, we don't know all of his business dealings, but we have been told through investigative reporting that he owes about $650 million to Wall Street and foreign banks.*

In the following we show how a combination of propositional content, illocutionary structure and anchoring captures the syntactic and communicative function of questions (and responses) in IAT.

4.1. Pure questioning

Pure questions are information-seeking. (3a) below is taken from the Citizen Dialogue corpus[4] (with (3b) added to the example from the corpus for illustrative purposes), a corpus that comprises Q&A sessions held during the public meetings organised by the US Department of Transportation. The meetings serve as part of a rulemaking process that allows for public consultation of proposed new policies. In (3a), after asserting that 'a lot of your steps are based on the availability of funding', Lee uses the polar question 'Has any money been approved?' to communicate his intention to seek information from Heather. Heather in turn provides the answer to the question: 'Yes, the money has been approved'.

(3) a. Lee: *I noticed a lot of your steps are based on the availability of funding.*
Has any money been approved? […]
b. Heather: *Yes, the money has been approved.*

Lee's first utterance ('I noticed a lot of steps') undergoes anaphora resolution for deriving the proposition, yielding 'Lee noticed a lot of steps'. The polarity of his question in (3a) with the possibility of an affirmative or negative response is encoded structurally in the proposition with the disjunction 'has or has not', a natural language instantiation of the choice space set up by the question. The resulting proposition 'any money has or has not been approved', as shown in Fig. 4, illustrates two things: First, IAT strongly assumes that illocutionary force and propositional content are encoded on two distinct levels, in line with speech act theory. Secondly, we see very clearly the tension between the IAT guidelines of (1) rendering grammatical propositions and (2) staying as close as possible to the NL form in reconstructing the propositional content. In this example, the annotator has given a higher priority to the latter guideline, namely that the NL form of the question is to be prioritised.

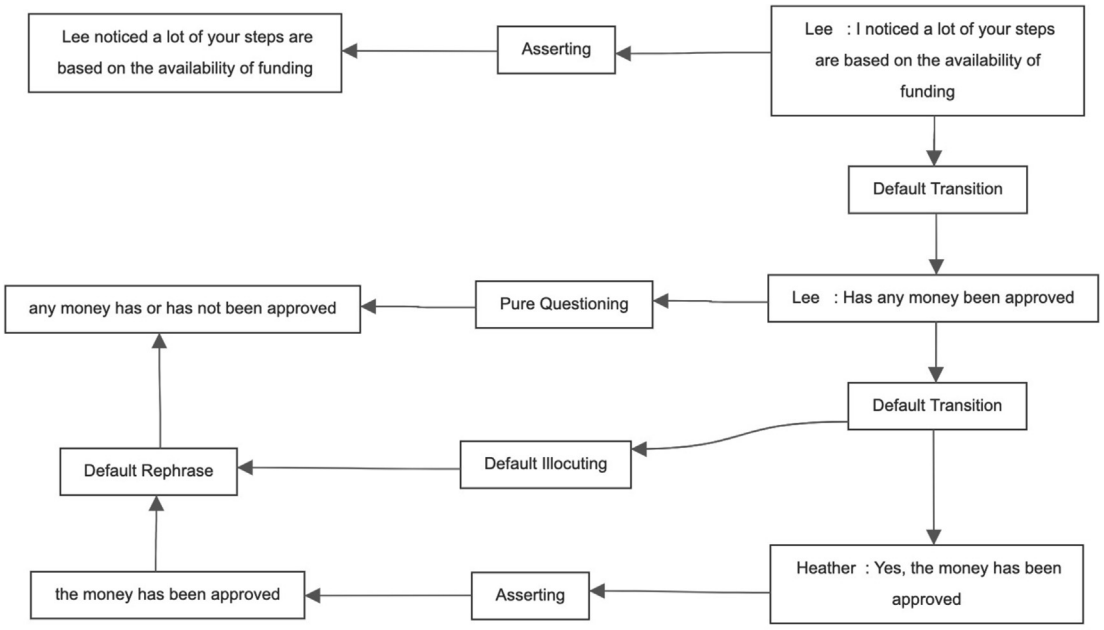

Fig. 4. IAT diagram for example (3).

[4] http://corpora.aifdb.org/cd.

Heather's response in (3b) constitutes what Berninger and Garvey (1981) and Stivers (2010) term an answer; it instantiates one of the conjuncts of the question. For answers in IAT, the proposition of the answer is linked to the proposition of the question via an argument structure of rephrase. The rationale is that answering typically involves rephrasing and reformulating the content of the question (or, in the case of (3b), instantiating it). Default Rephrase is anchored in the transition between Lee's and Heather's move, because for (3c) to be an answer, previous discourse material in the form of the question in (3b) is required – therefore the anchoring is in the transition and not in the locution.

Answerhood in IAT also follows the concept of underspecification (rather than focusing on the inferential relationship between propositions), in order to admit theories which do not rest upon foundations in the mathematical or linguistic traditions. A question's content might be represented as a lambda expression, as a Prolog clause, or as an underspecified Semantic Web graph, and in each case the relationship with the putative answer (an instantiation, a variable binding, or a graph fragment) can be related to that content through a common schematic relationship. What IAT does mandate is a clear separation between the interactional, dialogical facets of the question-answer relationship and the underlying propositional relations (and the relationships in turn between these two sets of relations). So regardless of the relationship between the content of a question and the content of a response, there will be other relationships between the dialogical performance of a question and of its response. IAT focuses on expressing the ways in which the relationships that are active in dialogue are connected to those that are active in the underlying content.

Another instance of a question-answer sequence is given in example (4), again taken from the Citizen Dialogue Corpus. Here a citizen uses a wh-question ('What has to happen to get the other funding?') to seek information from Heather. To capture the semantic content of the question, the proposition contains 'xxx' (a textually convenient shorthand for the lambda expression) as a way to account for the sets of items or individuals which can fill that proposition, following Hamblin's semantic treatment of wh-words (Hamblin, 1973) ('xxx has to happen to get other funding').

(4) a. Citizen: *What has to happen to get the other funding?* […]
b. Heather: That's something we continually work on.

The response that Heather provides is not as clear-cut an answer as her response in (3). However, it is still classified as an answer because IAT uses the speech act theoretical notion of 'intention': Heather makes an utterance with the communicative intention that her very intention of providing an answer to the question is recognised by the citizen. At first sight Heather's reply could also be what Berninger and Garvey's term an 'evasive reply', however evasive replies are only *discoursally* linked to the question, whereas Heather's response in example (4) is *grammatically* linked to the question, i.e., 'that's something we continually work on' filling the lambda position renders the proposition 'what has to happen to get other funding is something we continually work on'. This therefore makes it an answer also in Berninger and Garvey's (1981) terms.

Fig. 5 provides the diagram for example (4) which is parallel to the analysis in Fig. 4: The question proposition is anchored in the locution via Pure Questioning, the answer proposition is connected to the question proposition via Default Rephrase, which in turn is anchored in the Default Transition on the right-hand side via Default Illocuting. It is this (very broad) category of rephrase relations upon which the notion of answerhood turns.

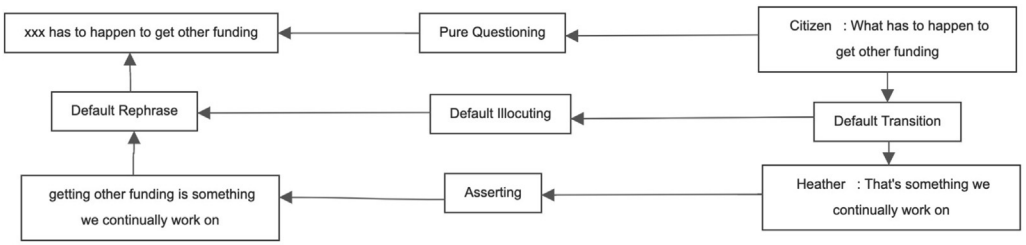

Fig. 5. IAT diagram for example (4).

4.2. Rhetorical questioning

Rhetorical questions (RQS) are those that are used to make an assertion and are not intended to receive an answer (though an interlocutor might provide one). RQS are tricky in the sense that they form a heterogeneous class of questions in theoretical

linguistic and empirical work on questions. With the criteria IAT proposes, we are in line with for instance Stivers and Enfield (2010) who characterise RQs as those questions "that may seek a response but do not seek an answer" (p. 2623, Stivers and Enfield, 2010). An example of the type of question we consider to be an RQ and their function in argumentative discourse is given in (5), taken from a BBC4 Moral Maze episode on the morality of money (MM2012, map 6151).[5] In this excerpt, Michael Portillo does not seek an answer from the interlocutor but uses the RQ to assert that 'borrowers shouldn't be doing so badly'. In the second move, he supports the claim in the RQ – borrowers should not be doing badly, because the interest rates are being kept low.

(5) Michael Portillo: *Why are borrowers doing so badly? I mean on the one hand interest rates are being kept really low, in order that borrowers won't bear the consequences of their foolish borrowing.*

Fig. 6 shows that the proposition of the RQ is an assertion which is anchored in the locution via the illocutionary connection Rhetorical Questioning. The IAT diagram also illustrates the Default Inference between the second proposition (the premise – 'interest rates are being kept really low […]') and the first (the conclusion – 'borrowers should not be doing badly'). The Default Inference is anchored in the transition via Arguing – the illocutionary force which signals argumentation in IAT. These topic-setting RQs with which a speaker does not invite a response are relatively frequent, as illustrated in the quantitative analysis of the data in §6.

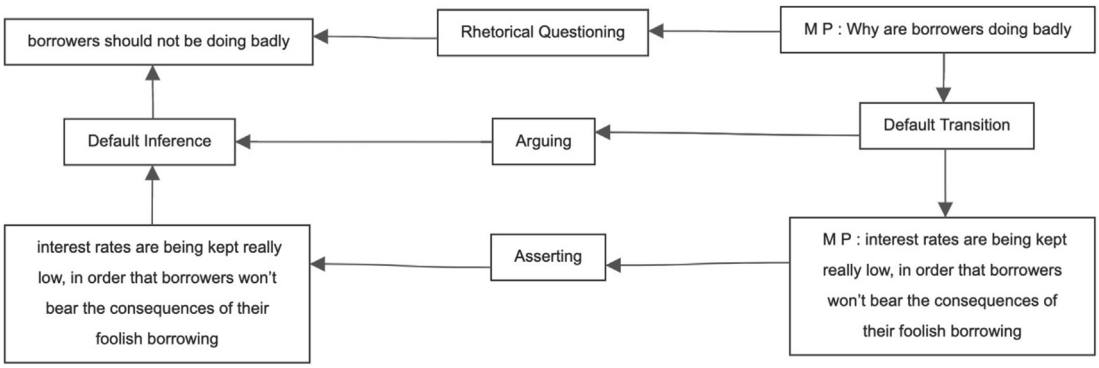

Fig. 6. IAT analysis for example (5).

Another type of RQ is shown in example (6), taken from one of the televised US2016 election debates: Donald Trump expresses his opinion on the US trade deficit and rounds up his point with two RQs, both intended as attacks on the capabilities of the previous governments: First, with 'You know what that is?', Trump intends to communicate a proposition along the lines of 'our trade deals are a disgrace', his second RQ 'who's negotiating these trade deals' seems to make an assertion similar to 'it is incompetent people who are negotiating US trade deals'.[6]

(6) Donald Trump: *We have a country that's doing so badly, that's being ripped off by every single country in the world […], Lester [Holt], we have a trade deficit with all of the countries that we do business with, of almost*
$800 billion a year. You know what that is? That means, who's negotiating these trade deals?

The diagram in Fig. 7 uses the wording above for creating the assertions of the RQs. The overall dynamics of example (6) are as follows: Trump uses a sequence of Default Rephrases to emphasise his initial point ('we have a country that's doing so badly'), namely 'we have a country that's being ripped off by every single country in the world' and 'we have a trad deficit […] of about $800 billion a year'. That last proposition is the premise in an argument with two conclusions, both of which are in the form of a rhetorical questions: 'our trade deals are a disgrace' and 'it is incompetent people who are negotiating the US trade deals'.

[5] The full corpus is available here: http://corpora.aifdb.org/MM2012

[6] In both cases, the variety of wording the propositional content is wide – an issue that we are aware of, but which requires more in-depth research on the presuppositions and implicatures triggered by questions in general.

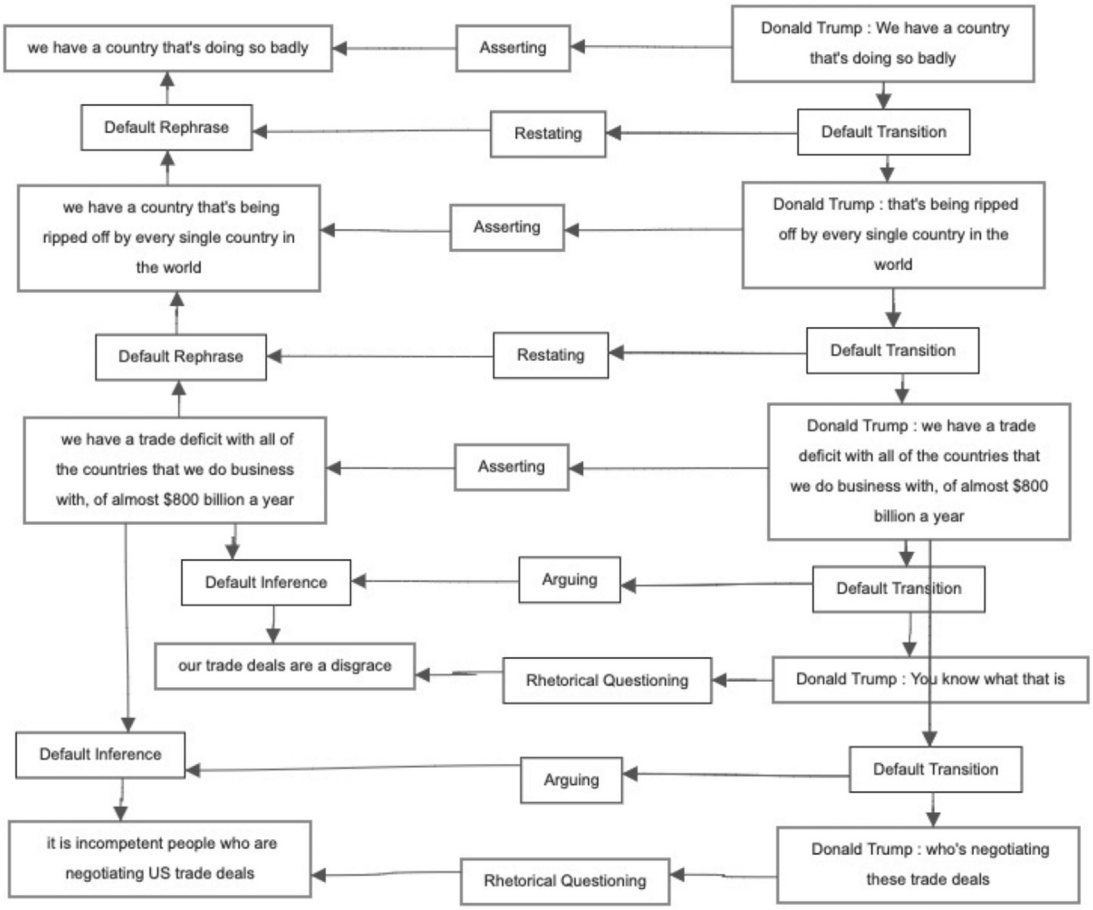

Fig. 7. IAT analysis for example (6).

4.3. Assertive questioning

Assertive questions (AQS) form an interesting category in the sense that they capture questions that fall between the categories of purely information-seeking and purely rhetorical. For a question to be assertive, IAT assumes that a speaker not only seeks the hearer's opinion on *p*, but also indirectly publicly declares her own opinion on *p* (Budzynska et al., 2014; 2016). The test for deciding between an AQ and a RQ in IAT is to check whether it is intended for the hearer to reply to a given question, e.g. whether his response 'Yes' to the question 'Are you an idiot or what?' would be treated as irrational (or humorous or naive) discursive behaviour (which would make the question a RQ).

AQS are not a simple composition of A and Q, because the constitutive rules of AQ are not a simple sum of rules for asserting and rules for questioning. In particular, AQ doesn't put on the speaker a burden of proof to such an extent as A does (*cf.* Gordon et al., 2007 for different levels of burden of proof in natural argumentation), since the speaker can withdraw from AQ easier than from A thanks to an interrogative form of its propositional content. For example, the burden of proof is stronger when the speaker says 'Fracking is not really worth destroying our environment for' than when she says 'Isn't is that fracking it not really worth destroying our environment for?'. In the first case, S can be challenged to deliver arguments supporting her opinion about fracking, while in the second case S can always respond 'I didn't claim that fracking is not worth destroying our environment, I was just asking about your opinion' which would not be eligible response in the first case. Still, AQS inherit some properties of the force of asserting, since the interlocutor can disagree (as in example (7)) or agree as a response to an AQ. This is not permitted in the case of a PQ, as its propositional content does not convey any claim or a standpoint to which the interlocutor can express her stance. The exact function of AQS is described differently across empirical investigations: Koshik (2003) illustrates their usage as challenges (Heritage, 2012), treats them as displaying the epistemic stance of the speaker and (Skilton, 2017) sees them as face-saving assertions.

For illustrating the IAT treatment of AQS in argumentation, we use the original version of the fracking example in (1), given here as example (7). Alex in (7a) uses an AQ to request Luca's opinion on whether she thinks fracking is 'worth damaging our

nature for', but she also states her opinion that it is not worth destroying our nature for. Luca in (7b) expresses her disagreement with Alex's assertion by responding to her question in (7a) ('Yes, we don't have untouched nature anyway').

(7) a. Alex: Lohnt sich das wirklich, wieder unsere Umwelt zu schädigen?
worth Refl that really again our nature to damage
'*Is it [fracking] really worth damaging our nature for?*'
b. Luca: Ja. Wir haben sowieso keine unberührte Natur.
yes we have anyway no untouched nature
'*Yes. We don't have untouched nature anyway.*'

IAT allows us to untangle this dynamic inferential structure in a way that both communicative intention and inferential structure are explicated: The diagram in Fig. 8 for the English translation of example (7) shows that the contribution of Alex's AQ is an assertion, i.e. the communicative intention of Alex is to assert *p* and not receive information whether Luca thinks *p* or not *p*. The proposition is anchored in the dialogical structure on the right via the illocutionary connection Assertive Questioning. Despite the affirmative 'Yes' from Luca, she in fact signals a contrasting opinion: she does think that fracking is worth damaging our nature for, captured with the Default Conflict and an illocutionary force of *Disagreeing* anchored in the transition between Alex and Luca's moves. Luca supports her point further by giving an argument why she thinks fracking is worth destroying nature for ('we don't have untouched nature anyway'). This support relation is represented with Default Inference, which is anchored in the dialogue structure via the illocutionary force of Arguing.

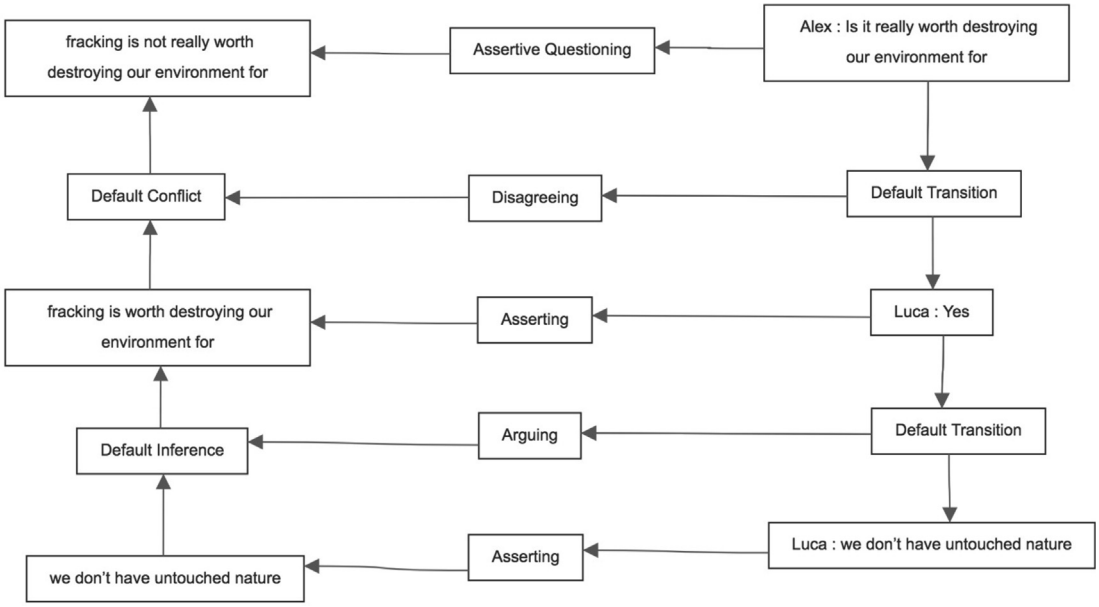

Fig. 8. IAT analysis for example (7).

Our data suggests that AQs are strongly signalled by cues such as 'Isn't it the case that … ', 'Can we agree that … ', 'Doesn't …', an example of which is illustrated in (8) from the BBC4 Moral Maze episode on the morality of money (MM2012, map 6143): Matthew Taylor invites a response on the question of the nature of the financial services industry, but also communicates his own standpoint on the issue:

(8) a. Matthew Taylor: *Doesn't the history of the financial services industry demonstrate that this is an industry with a deep commitment to exploiting […] weaknesses, in order to make a fast buck?*
b. John Lamiday: Well, I don't really agree with that at all.

As shown in Fig. 9, the propositional content of the AQ by Matthew Taylor captures the assertive nature of the question, namely that 'the history of the financial service industry demonstrates that this is an industry with a deep commitment to exploiting weaknesses in order to make a fast buck'. John Lamiday disagrees with this view ('I don't really agree with that at all'). This proposition does not add any additional content except for referring back to Matthew Taylor's proposition and disagreeing with it (Default Conflict). this is captured in IAT by anchoring the proposition of the response in the Default Transition between the two moves.

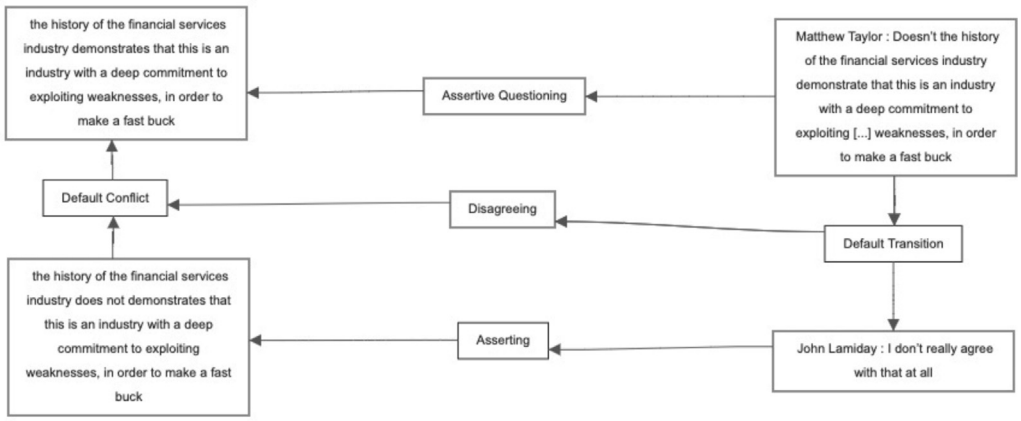

Fig. 9. IAT analysis for example (8).

In the category of AQs also fall those questions that (Koshik, 2002) terms 'reversed polarity questions', where the polarity of the question actually displays the reversed epistemic stance of the speaker. For instance, Koshik's example (p. 1855, Koshik (2002)) 'Mister President, in your zeal for funds during the last campaign, didn't you put the Vice President […] and all the others in your administration top side, in a very vulnerable position?', the assertion of the question is one of reversed polarity, namely 'In your zeal for funds during the last campaign you put the Vice President and all others in your administration top side in a very vulnerable position.' In IAT, this reversed polarity is accounted for in the configuration of the proposition on the left-hand side, but without an explicit label.

4.4. Challenge questioning

The fourth type of question is motivated by the fact that IAT comes from the tradition of formal dialogue systems where challenges, i.e. questions that challenge the interlocutor, typically play a critical role in the protocol that interlocutors follow. These challenges are stereotypically used to trigger argumentation because a speaker seeks the grounds for another speaker's point of view (see, e.g. Hamblin (1970), Mackenzie (1990), Walton and Krabbe (1995)). In IAT, this type of question falls outside the pure-assertive-rhetorical-questioning spectrum and is annotated with the illocutionary force 'Challenging'.

A large majority of challenge questions in our dataset (55 of 62, see Tables 4 and 5) are expressed with a why-interrogative, as illustrated in Example (1) in §3, repeated here as (9): Alex's question in (9b) creates a burden of proof on Luca to justify her claim in (9a), which she addresses in her reply in (9c).

(9) a. Luca: Fracking is worth destroying our nature for.
b. Alex: *Why?*
c. Luca: We don't have untouched nature anyway.

Table 4
Distribution of illocutionary structures across genres of argumentative discourse.

	Election debate	Moral debate	Public deliberation	Total
Total Non-questioning	6266 (95.37%)	5213 (94.99%)	3751 (95.42%)	15230 (95.25%)
Asserting	4218 (64.20%)	2953 (53.81%)	2245 (57.11%)	9416 (58.89%)
Arguing	1523 (23.18%)	1280 (23.32%)	879 (22.36%)	3682 (23.03%)
Agreeing	36 (0.55%)	180 (3.28%)	65 (1.65%)	281 (1.76%)
Disagreeing	153 (2.33%)	368 (6.71%)	151 (3.84%)	672 (4.20%)
Restating	243 (3.70%)	189 (3.44%)	376 (9.56%)	808 (5.05%)
Default	93 (1.42%)	243 (4.43%)	35 (0.89%)	371 (2.32%)
Total Questioning	304 (4.63%)	275 (5.01%)	180 (4.58%)	759 (4.75%)
Pure Questioning	166 (2.53%)	113 (2.06%)	52 (1.32%)	331 (2.07%)
Rhetorical Questioning	30 (0.46%)	39 (0.71%)	20 (0.51%)	89 (0.56%)
Assertive Questioning	85 (1.29%)	93 (1.69%)	99 (2.52%)	277 (1.73%)
Challenge Questioning	23 (0.35%)	30 (0.55%)	9 (0.23%)	62 (0.39%)
Total	6570	5488	3931	15989

Table 5
Distribution of question types across genres of argumentative discourse.

	Election debate	Moral debate	Public deliberation	Total
Pure questioning	166 (54.61%)	113 (41.09%)	52 (28.89%)	331 (43.61%)
Rhetorical questioning	30 (9.87%)	39 (14.18%)	20 (11.11%)	89 (11.73%)
Assertive questioning	85 (27.96%)	93 (33.82%)	99 (55.00%)	277 (36.50%)
Challenge questioning	23 (7.57%)	30 (10.91%)	9 (5.00%)	62 (8.17%)
Total	304	275	180	759

The IAT diagram in Fig. 10 replicates Fig. 2. Luca's move ('Why?') does not contribute its own proposition, instead the propositional target of the illocutionary force of 'Challenging' is the preceding locution, which is captured with 'Why?' not having its own propositional content but referring to the one by Luca.

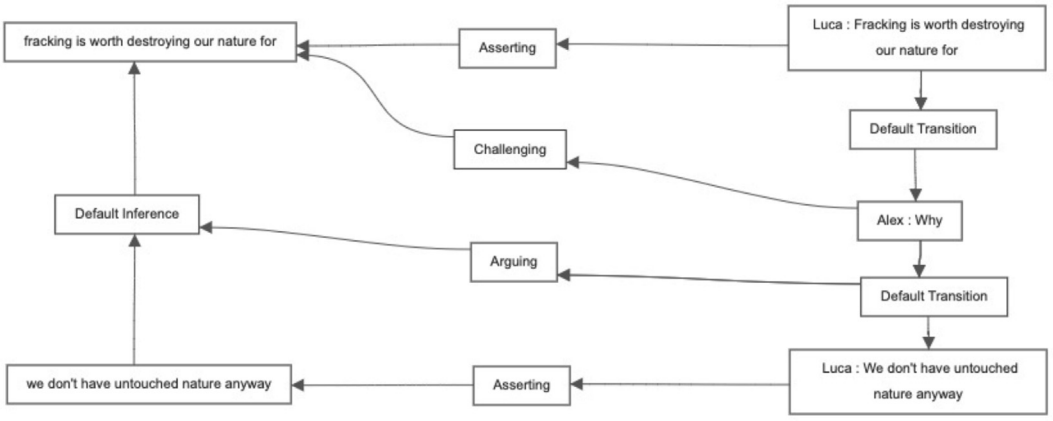

Fig. 10. IAT analysis of a challenge question.

In the following section we provide details of the corpora underlying the investigation (§5.1) and a first account regarding the distribution and relevance of question types in them (§5.2), followed by a detailed pragmatic account of questions in argumentative discourse in §6.

5. Empirical basis

5.1. Genres

The dataset underlying our investigation consists of six corpora in three argumentative genres: election debate, moral debate, and public deliberation. All corpora are annotated with IAT. In total, our dataset consists of 153,932 words (tokens), and 15,989 illocutionary structures.

For the 'Election Debate' genre, we use the US2016tv[7] corpus (Visser et al., 2019), which includes the transcripts of all televised debates in the US 2016 Presidential Election (i.e. the primary debates for Republicans and Democrats, and the three debates between Hilary Clinton and Donald Trump in the run-up to the general election). The 'Election Debate' dataset has 39,694 words, and 6570 illocutionary structures and an inter-annotator agreement of Cohen's $\kappa = 0.61$ (substantial agreement) (Visser et al., 2019). As described in more detail by Visser et al. [2019], the annotation for this dataset and for those that follow, was conducted by a small team of between two and ten annotators who were trained and working to a set of publicly available IAT annotation guidelines. Inter-annotator agreement is surprisingly high for such high-level pragmatic annotation with so many classes and segmentation variability.

For the genre 'Moral Debate', we include two sub-corpora. MM2012[8] is a collection of transcripts of BBC Radio 4's 'The Moral Maze', a series of debates about moral dilemmas in which witnesses are invited and questioned by a number of panellists. The second corpus, BBC2017[9], consists of a special issue of The Moral Maze and a related TV debate about the

[7] Available at http://corpora.aifdb.org/US2016tv.
[8] http://corpora.aifdb.org/mm2012
[9] http://corpora.aifdb.org/BBCaaaTV5RADIO.

morality of abortion, produced for the 50th Anniversary of the UK's Abortion Act. The 'Moral Debate' dataset has 60,273 words, and 5488 illocutionary structures.

For 'Public Deliberation', we include three sub-corpora. USCD2011[10] consists of several transcripts of public meetings and hearings from departments of transportation (DOTs) across the United States of America (Lawrence et al.; Konat et al., 2016). DEDD2019[11] is a set of two experimental 4-party deliberations on whether or not to allow fracking in Germany. The last corpus, UKDD2019[12] is a public deliberation organised by the Royal Society of Edinburgh in the context of an inquiry about the future of energy in Scotland. The 'Public Deliberation' dataset has 34,759 words and 3931 illocutionary structures.

Inter-annotator agreement in the US2016tv corpus is reported as substantial, but (Visser et al., 2019) do not provide individual scores for the different illocutionary structures. In order to elicit some of the disagreements in annotating questions in the corpus, we manually compare the analyses of questions in a subset of analyses.

The main point of disagreement lies in the correct reconstruction of the proposition, for instance, annotators forget to remove the epistemic markers, which sometimes leads to downstream analysis errors. For instance, in the debate between the Democrats Hillary Clinton and Bernie Sanders, the question to Sanders 'Do you think taxpayers should pick up the tab for wealthy children?', only one annotator (map 10817 in US2016tv) correctly removes the epistemic maker and renders the proposition as 'taxpayers should or should not pay up the tab for wealthy children'. The second annotator (map 10928 in US2016D1tv-iaa) keeps the epistemic marker and also applies the disjunction to it: 'Sanders does or does not think that taxpayers should pick up the tab for wealthy children'. So while both annotators picked the correct question type ('Pure Questioning'), only the former provides the correct propositional reconstruction.

In another case, propositional content is resolved and reconstructed slightly differently across annotations, with both analyses rendering a valid annotation. For example in (10), moderator Cooper questions Hillary Clinton as to her standpoint on instate college tuition for undocumented immigrants. Both annotators tag this as 'Pure Questioning'. However, one annotator (map 10441) interprets 'where' to signal an narrow alternative reading with a choice between either providing or not providing college tuition for undocumented immigrants. Consequently, the proposition is reconstructed as 'Clinton's stance is that instate college tuition should or should not be provided to undocumented immigrants'. The second annotator (map 10469) resolves the proposition with a wider interpretation in which 'where' hints at a variety of standpoints regarding the topic. In this case, the proposition of the question is reconstructed as 'Clinton's standpoint on providing instate college tuition to undocumented immigrants is xxx'. In addition, the surface realisation of the propositional content varies: 'stance' versus 'standpoint' and the inserted modality 'should' of annotator 1. Both instances still comply with the IAT guidelines, they simply represent the subtlety with which the annotators interpret the intention of the speaker in asking the question.

(10) Cooper: *Two of your rivals from your left, Governor O'Malley and Senator Sanders, want to provide instate college tuition to undocumented immigrants. Where do you stand on that?*

Despite this variation on the level of propositional reconstruction, the annotation of the question types in the sample that we manually analyse represents the overall substantial agreement in the corpus, meaning that the results derived in Section 6 are based on a well-motivated empirical basis.

5.2. Distribution and frequency of questions across genres

The statistics in Table 4 show that despite the varying sizes of the three argumentative genres (the 'Election Debate' corpus is the largest and almost twice the size of the 'Public Deliberation' data set), Asserting is consistently the most frequent illocutionary connection, making up almost 60% of all illocutionary connections, followed by Arguing with around 23%. Questioning makes up between 4.5% and 5% of all illocutionary connections, with around 2% of pure questioning and the rest divided between rhetorical, assertive and challenge questioning. Overall, the number of questions is slightly lower than in other corpora such as the Switchboard corpus (8%) (Jurafsky et al., 1997), however, rhetorical questions are just as frequent (aggregated 0.56% in our data sets v. 0.55% in the Switchboard corpus). In comparison, the Penn Discourse Treebank 3.0, questions make up only 0.2% of all explicit and implicit discourse relations (annotated with the HYPOPHORA relation (Webber et al., 2017)).

Table 5 provides the distribution of question types across argumentative genres: Pure, information-seeking questions are consistently the most frequent type of question, making up for 43.1% of all questions in the dataset. Except for 'Public Deliberation' they rank highest, followed by assertive questions, a type of questioning that is very prominent and statistically

[10] http://corpora.aifdb.org/cd.
[11] http://corpora.aifdb.org/DEDD2019
[12] http://corpora.aifdb.org/UKDD2019

significant (Pearson's Chi-squared test with Yates' continuity correction gives $\chi 2 = 33.823$, 1 degree of freedom, p-value $= 6.035e{-}09$). This indicates that assertive questions are a defining feature of argumentative discourse across genres. In possible contrast to the keystone role assumed for challenge questions in Mackenzie (1990) and Walton and Krabbe (1995) (amongst many approaches to formal dialectics), they rank least frequent in our data: Only 8.17% of all questions in the dataset are challenge questions. This is supported by the findings in §6 where we show that challenge questions are not as frequently used to trigger argumentation as previously stipulated.

In what follows we will look at the pragmatic import of questions in argumentative dialogue, i.e. the impact of questioning illocutionary structures on the illocutionary and propositional structures that follow, with a particular focus on different types of question-response and question-argument patterns.

6. The pragmatic import of questions

Given our substantial empirical basis, it becomes possible to start to explore the use and pragmatic import of questions in argumentative debate. Focusing on a question at a given point in the dialogue, it is possible (a) to tease apart the structures that are obligated, licensed and prohibited as a result; and (b) to look at the historical context of a given structure in the discourse and uncover the roles played by questions in getting to that point. The following two subsections explore these two directional orientations.

6.1. Question. What comes next?

The context succeeding a question is generally called 'the response space', with a number of coding schemes having worked on a partition or characterisation of that space. As mentioned before, IAT pursues an approach of underspecification – in the case of responses this means that we differentiate only four types of responses: First, we can distil those discourse moves that entirely ignore the question (*Ignored*) and relate to discourse material before the question. This category is parallel to (Ginzburg et al., 2019)'s 'Ignore' type of query response and (Stivers and Enfield, 2010) None-response category where the "person did nothing in response" (p. 2624).

Secondly, *Reactions* are those discourse moves that do not ignore a question but 'let it hang in mid air' and rather contribute to the ongoing discourse. Whether these reactions are queries (as in Łupkowski and Ginzburg (2016)) or declaratives is not analysed further, but could be derived by taking into account the illocutionary force of the succeeding move. Structurally, 'Reactions' are identified by transitions connecting the question to a succeeding utterance, i.e. the existence of an outgoing TA from the question locution.

The third category (*Responses*) subsumes those discourse moves that are directly related to a question (but do not answer it). Typical examples include, for instance, moves that evade a question or succeed a rhetorical question whose assertion and the assertion of the response together create an argumentative structure. Given our dataset, the category of 'Responses' is crucial given the frequency of questions that are not intended to elicit an answer, namely rhetorical and assertive questions: They can be responded to, but are typically not answered. 'Responses' is again an underspecified type of move, in line with what Stivers and Enfield (2010) call 'Non-answer responses' and Ginzburg et al. (2019) 'Non-answers', subsuming, for instance, Berninger and Garvey (1981)'s 'Evasive responses' and Ginzburg et al. (2019)'s 'Change-the-topic' responses. Those subtypes are not specified further in IAT, but can be derived by taking into account the propositional content in more depth. 'Responses' are captured by a transition that follows the question and anchors a YA, i.e. we count all transitions where the locution that follows is only interpretable given the content of the question.

The fourth category is for responses that do, in fact, constitute answers to a question (*Answers*); that is to say, the content of the response delivers propositional information that instantiates the lambda expression (in the case of a wh-question) or resolves the disjunction (in the case of a polar and alternative question). Answers in IAT are captured via a Default Rephrase between answer and question which is anchored via Default Illocuting in the transition between them (see the IAT diagram in Fig. 5 for an example). This category of answer is found with the exact same label in Stivers and Enfield (2010) and Ginzburg et al. (2019), Berninger and Garvey (1981) use the term 'Possible Answers'.[13]

Table 6 gives an overview of the types of responses that the different types of questions yield across genres of argumentative debate, broken down into each of the four categories of question as well as, in the final panel, in aggregate. Inherent in Table 6 is a notion of graded answerhood that originates in the constructionist approach to encoding questions-response structures in IAT: A 'Reaction' is encoded by way of a question having a simple outgoing TA (meaning that the discourse is somewhat related to the content of the question), the TA which, in the case of a Response, anchors an illocutionary force. For answers, this illocutionary force, in turn, anchors a Default Rephrase between answer and question.

[13] More formal analyses in the style of (Groenendijk and Stokhof, 1984) are in principle possible, for instance to what extent non-exhaustive answers invite requests for elaboration and whether an exhaustive answer triggers a shift to another topic, as one reviewer notes. However, these formal semantic analyses are beyond the account of questions and their role in argumentative discourse that are the focus of this paper.

Table 6
Analysis of the response space (aggregated).

	Election debate	Moral debate	Public deliberations	
Questioning (All Types)	**304**	**529**	**182**	**1015**
~**Ignored** (no outgoing TA)	**55**	**91**	**24**	**170**
~**with Reactions** (existence of outgoing TA)	**249**	**438**	**158**	**845**
~~**with Responses** (anchor a YA)	**129**	**369**	**100**	**598**
1 response/question	100	268	80	448
2 responses/question	22	77	17	116
>2 responses/question	7	23	3	33
~~**with Answers** (the content is MA)	**96**	**195**	**55**	**346**
Same speaker	23	62	26	111
Different speaker	73	112	29	214

Overall, as Table 6 shows, a very large majority (between 80 and 87%) of questions across argumentative genres and question types (Election Debate: 249 out of 304, Moral Debate: 438 out of 529, Public Deliberation: 158 out of 182) are reacted to and lead to subsequent, connected discourse material. We can also investigate to what extent a question is responded to, i.e. whether the question triggers an utterance that is directly related to it. To that end, Table 7 provides the frequencies for the individual question types. Overall, this shows that questions play an important role in driving a dialogue forward: between 50 and 60% of questions are responded to in the subsequent discourse and are therefore relevant to what follows. This number is stable across genres and question types.

In order to make the response patterns more insightful, Table 7 breaks down 'Responses (anchor a YA)' into the number of responses per question — one response per question, two responses per question and more than two responses per question. This allows us to derive to what extent individual questions trigger a large number of responses. Overall, 70–80% of questions are responded to once (Election Debate: 100 out of 129, Moral Debate: 268 out of 369, Public Deliberation: 80 out of 100), with around 5% responded to more than twice (Election Debate: 7 out of 129, Moral Debate: 23 out of 369, Public Deliberation: 3 out of 100). Across the complete dataset, the maximum number of responses to a single question is eight, found in the 2017 BBC documentary 'Abortion on Trial' (BBCaaaTV5RADIO, map 12385): Here, presenter Anne Robinson asks *'Who would like the law changed so you could take those drugs for a non-surgical abortion at home?'*. In the discourse that follows, all interview participants either signal their consent or dissent.

In some cases, a question-response structure does not involve two speakers but only one, as shown in example (11), taken from the US 2016 presidential election debate (US2016tv, map 10822): Here, Lincoln Chafee uses the rhetorical question to assert that he is the one who is best able to confront the challenges [that America is facing], backing up this claim with the statement that he is a proven peacemaker.

(11) Lincoln Chafee: *Who is best able to confront these challenges? I am a proven peacemaker.*
These questions have been previously termed 'header questions' or 'discourse-structuring questions'

Kalouli et al., 2018, as they are not used to elicit or assert information, but to structure the succeeding discourse by the speaker. This type of question is not explicitly catered for in IAT, but can be distilled from the data by looking at the speaker who provides the response: If it is the same speaker, it is a discourse-structuring question, if the response is provided by a different speaker, it constitutes a 'regular' question-response structure.

Finally, we can also see whether responses do, in fact, constitute answers to a question. In Table 7 we distinguish those answers that are given by different speaker ('By different speaker') from those that are given by the same speaker ('By same speaker'). In case of an answer by a different speaker, a typical example is the structure as in example (12), taken from the Moral Maze debate on the 'Morality of Money' (MM2012, map 6197), where the disjunction of Michael Portillo's 'it [stigmatising debt] is/is not an important thing to do' is resolved with the witness's answer that has the propositional content of 'it is not an important thing to do'.

(12) a. Michael Portillo: *I mean, first of all, do you think it's a very important thing to do?*
b. Witness: *No.*

Table 7 shows that around 50% of all responses are answers to a question (Moral Debate: 195 out of 369, Public Deliberation: 55 out of 100). An exception is the genre Election Debate in which 75% of responses are answers (96 out of 129). This genre is an exception in that the structure of the debate is rather fixed: a coin toss determines the candidate who gets to answer the first question, the answer can last up to 2 min and the subsequent questions alternate between the two candidates. Due to this setting, the proportion of answers compared to mere responses is higher than in the other, more unregulated argumentative subgenres. The impact of the particular setting is also mirrored in the fact that a large portion of all responses in Election Debate, around 76%, are indeed answers by a different speaker (73 out of 96). This is not the case in the other two subgenres, where only around 50% of questions are answered by a different speaker (Moral Debate: 112 out of 195, Public Deliberation: 29 out of 55). The highest proportion of answers by a different speaker are given for pure, information-seeking questions (Election Debate: 64 out of 68, Moral Debate: 104 out of 120, Public Deliberation: 18 out of 20). On the other hand, rhetorical questions are, despite their much lower frequency, generally 'answered' by the same speaker (Election Debate: 2 out of 2, Moral Debate: 20 out of 25). Investigating the underlying data shows that speakers often rephrase these types of questions (which are actual assertions in the guise of a question) in the immediately following discourse. This means

that the Default Rephrase in this type of question does not mean that content of the question is instantiated, but it is refined or elaborated. Assertive questions very seldomly yield an answer (Election Debate: 0 out of 85, Moral Debate: 20 out of 203, Public Deliberation: 9 out of 101) – the reason for this could be that they cannot be used as discourse-structuring assertion to which the same speaker can respond to, but AQs are more like assertions (in fact similar to what Koshik (2002) terms 'reversed polarity items') that communicate the epistemic stance of the speaker to which it is hard to answer to (though it is common to respond to them – as shown in Table 7). Challenge questions are very hardly answered (Election Debate: 0 out of 23, Moral Debate: 8 out of 46, Public Deliberation: 1 out of 9), a pattern that is expected given that challenges are used to elicit further explanation or justification from the hearer rather than to rephrase the content that was questioned. This is confirmed by the numbers of reactions and responses to challenge questions.

Table 7
Analysis of Question-Response patterns (individual question types).

	Election debate	Moral debate	Public deliberations	
Pure Questioning	**166**	**169**	**52**	**387**
~**Ignored** (no outgoing TA)	**21**	**17**	**8**	**46**
~**with Reactions** (existence of outgoing TA)	**145**	**152**	**44**	**341**
~~**with Responses** (anchor a YA)	**72**	**122**	**27**	**221**
1 response/question	61	97	24	182
2 responses/question	8	20	3	31
>2 responses/question	3	5	0	8
~~**with Answers** (the content is MA)	**68**	**120**	**20**	**208**
Same speaker	4	16	2	22
Different speaker	64	104	18	186
Assertive Questioning	**85**	**203**	**101**	**389**
~**Ignored** (no outgoing TA)	**18**	**32**	**10**	**60**
~**with Reactions** (existence of outgoing TA)	**67**	**171**	**91**	**329**
~~**with Responses** (anchor a YA)	**35**	**156**	**58**	**249**
1 response/question	20	105	45	170
2 responses/question	13	39	10	62
>2 responses/question	2	11	3	16
~~**with Answers** (the content is MA)	**0**	**20**	**9**	**29**
Same speaker	0	5	2	7
Different speaker	0	15	7	22
Rhetorical Questioning	**30**	**111**	**20**	**161**
~**Ignored** (no outgoing TA)	**7**	**29**	**3**	**39**
~**with Reactions** (existence of outgoing TA)	**23**	**82**	**17**	**122**
~~**with Responses** (anchor a YA)	**11**	**66**	**12**	**89**
1 response/question	10	46	9	65
2 responses/question	0	15	3	18
>2 responses/question	1	5	0	6
~~**with Answers** (the content is MA)	**2**	**25**	**0**	**27**
Same speaker	2	20	0	22
Different speaker	0	5	0	5
Challenge Questioning	**23**	**46**	**9**	**78**
~**Ignored** (no outgoing TA)	**9**	**13**	**3**	**25**
~**with Reactions** (existence of outgoing TA)	**14**	**33**	**6**	**53**
~~**with Responses** (anchor a YA)	**11**	**25**	**3**	**39**
1 response/question	9	20	2	31
2 responses/question	1	3	1	5
>2 responses/question	1	2	0	3
~~**with Answers** (the content is MA)	**0**	**8**	**1**	**9**
Same speaker	0	4	1	5
Different speaker	0	4	0	4

If we compare question types based on their response pattern, pure questions are distinct from all other question types in that they have the highest frequency of answers (Pure Questioning: 208 out of 387–53%). The other three question types (Assertive Questioning: 29 out of 289–16%, Rhetorical Questioning: 27 out of 161–7%, Challenge Questioning: 9 out of 78–11%) are on the other side of the spectrum regarding answers with much lower answer rates. However, regarding the rates of ignored, reacted and responded to questions, no clear picture emerges regarding a correlation of the type of question and a particular response type. About 50–65% of question are responded to, between 11% (Pure Questioning) and 30% (Challenge Questioning) are ignored. The latter is interesting in that one would expect challenges, i.e. question to obtain justification, would not go unnoticed in argumentative dialogue. A manual analysis shows that these questions are mostly interjections in ongoing discourse and are not reacted to by the interrupted speaker who continues his discourse (in this case there is no TA between the question and the following locution).

After having investigated the general function and impact of questions in different types of argumentative discourse, we will now shed light on the interplay of questions and argument structure.

6.2. How did we get to arguing?

There are many ways to explore the relationships between questions and argument structure. Our analysis in Table 8 is divided into the two sides of the argumentative coin: inference (support) and conflict (attack). For each, we look first at the overall frequencies of the antecedent and consequent (for inferences, the premise and conclusion) and the proportion of them that are realised by questions ('the premise/conclusion/antecedent/consequent is a question'). This allows us to answer the question whether (particular types of) questions are stereotypically associated with particular local roles in argumentative structures. In the next step, we look backwards from these argumentative structures to explore the role that questions may have had in their genesis. First, the focus is immediate: looking at cases where inferences and conflicts are direct responses to questions of one sort or another ('As immediate response to'). If we switch further back in the dialogical history by tracing a chain of relevant responses, from response-to-response back to response back to original locution (to any number of steps), we can identify how often inferences and conflicts occur in response to a question (of one type or another) at any number of (dialogically relevant) steps later ('As eventual response to').

Table 8
Analysis of Question-Argument patterns.

	Election debate	Moral debate	Public deliberations	Total
Inferences	**1551**	**1108**	**882**	**3541**
Premise is a question	**80**	**74**	**34**	**188**
Pure Questioning	2	1	2	5
Assertive Questioning	65	31	19	115
Rhetorical Questioning	6	30	12	48
Challenge Questioning	7	12	1	20
Conclusion is a question	**35**	**98**	**40**	**173**
Pure Questioning	0	6	1	7
Assertive Questioning	19	36	26	81
Rhetorical Questioning	11	45	12	68
Challenge Questioning	5	11	1	17
As immediate response to	**37**	**113**	**38**	**188**
Pure Questioning	0	4	38	42
Assertive Questioning	19	47	19	85
Rhetorical Questioning	13	50	17	80
Challenge Questioning	5	12	1	18
As eventual response to	**44**	**44**	**32**	**120**
Pure Questioning	0	7	0	7
Assertive Questioning	12	8	24	44
Rhetorical Questioning	30	23	8	61
Challenge Questioning	2	6	0	8
Conflicts	**194**	**262**	**178**	**634**
Antecedent is a question	**8**	**32**	**8**	**48**
Pure Questioning	0	0	0	0
Assertive Questioning	6	19	7	32
Rhetorical Questioning	2	10	0	12
Challenge Questioning	0	3	1	4
Consequent is a question	**25**	**89**	**13**	**127**
Pure Questioning	0	2	0	2
Assertive Questioning	14	68	10	92
Rhetorical Questioning	1	16	1	18
Challenge Questioning	10	3	2	15
As immediate response to	**33**	**104**	**21**	**158**
Pure Questioning	0	2	0	2
Assertive Questioning	20	77	17	114
Rhetorical Questioning	3	19	1	23
Challenge Questioning	10	6	3	19
As eventual response to	**28**	**4**	**0**	**32**
Pure Questioning	0	0	0	0
Assertive Questioning	16	4	0	20
Rhetorical Questioning	0	0	0	0
Challenge Questioning	12	0	0	12

The top half of each of the inferences and conflicts panels in Table 8 clearly demonstrates that speakers use questions to put forth argumentation: around 5% of antecedents and consequents of inferences and conflicts are formulated as questions, with assertive questions consistently being the most commonly used question type. An example of an argument component in the guise of a question is shown in the exchange in (13), taken from the US 2016 TV debate (US2016tv, map 10464): Chris Wallace, the moderator, inquires about Trump's business record. In (13a), the question is in fact an assertion by which he implicitly states that with Trump's record we should not trust him to run the nation's business. This assertive question is the conclusion of the sentence ultimately preceding the question, namely that the bankruptcies meant that lenders lost billions of dollars and people were laid off. In his response in (13b), Trump does not respond to Wallace's claim but diverts the question and shifts the focus to his company's lenders.

(13) a. *Wallace: Let's just talk about the latest example which is Trump Entertainment Resorts, which went bankrupt in 2009. In that case alone, lenders to your company lost over $1 billion and more than 1100 people were laid off. Is that the way that you'd run the country?*
b. *Trump: Let me just tell you about the lenders […]*

Rhetorical questions are used to almost the same extent, though the numbers are generally lower than for assertive questions. Determining the exact function of RQs in conversation, and in particular in argumentative, persuasive contexts, has yielded a variety of interpretations: Brown and Levinson (1987) consider them as part of a speaker's politeness strategy for face-threatening speech acts (e.g. criticism and sarcastic remarks). In contrast, Anzilotti (1982) sees RQs as a strategy to "affect persuasion by compelling listener's involvement" (p. 299), which leads (Frank, 1990) to the conclusion that using RQs enables a speaker "to make stronger statements, with greater implications, than would be possible if they had made a straightforward assertion" (p. 726). Going through the US televised presidential debate manually shows that an interpretation of the rhetorical and assertive questions in the corpus is in line with the assumption of Brown and Levinson (1987): They are used as a hedging device for strong claims made in a debate, i.e. they are a means to soften the blow of a purely asserted critique or accusation. We cannot provide more empirical support to the question whether they are an effective persuasion device, in the way that Anzilotti (1982) and Frank (1990) postulate, but we can inspect the reaction of the hearer, which does indeed show that the assertion underlying the RQ (i.e. the critique or accusation) is understood. This is also evident in example (13) above: The RQ packages the accusation that Trump would run the country in an irresponsible way, enabling Wallace to make a stronger point than with a simple assertion and thereby reducing the face threat to Trump (who in (13) exploits this room for manoeuvring the question with a response that is not in fact an answer).

When we look backwards from these argumentative structures in order to explore the role of questions, we see that assertive and rhetorical questions are amongst those that are ranked highest in generating immediate argumentation. From the point of view of argumentation theory, for instance Freeman (1991) who claims that argumentation depends upon critique, challenge, […] and is realised by posing 'Why questions', or Prakken (2006) who specifies that in formal dialogue systems for persuasion an argument can be given only in response to a '*why* φ' dialogical move but not to '*question* φ',[14] this adds empirical evidence that in fact other question types incite argumentation much more often. However, if we compare the overall number of challenge questions (see Table 5) with the challenge questions that immediately trigger an argument (sum over challenge questions triggering inferences and conflicts), we see that once they do occur, they trigger immediate argumentation in 50–60% of cases (Election Debate: 15 out of 23, Moral Debate: 18 out of 30, Public Deliberation: 4 out of 9). Pure, information-seeking questions are rarely followed by an argument. If we switch further back in the dialogical history, Table 8 shows that assertive and rhetorical questions are those with the highest effect on leading to an exchange of arguments.

7. Discussion

In our investigation of the function of questions in argumentative discourse we build on a large-scale and exceptionally diverse empirical basis, which allows us to relate to claims both in argumentation theory and pragmatic linguistic work on questions.

First and foremost, our study clearly shows that questions play an important role in argumentation, empirically supporting Walton (1988)'s claim that argumentation theory has to deal with the role of questions in argumentative discourse. In particular assertive and rhetorical questions, those whose communicative aim is rather to make a claim than request for information, are crucial in driving the argumentation forward. However, our analysis also shows that they serve a particular purpose, namely to hedge a statement that could otherwise threaten the hearer's face. As such they leave room for discursive manoeuvres such as evading a question, which empirically validates the claims in Brown and Levinson (1987) and Skilton (2017) that questions are used as hedged assertions, directives or offers.

Our analysis also shows that there is clear evidence for a particular type of question, namely discourse-structuring or header questions. We therefore contribute to efforts mentioned in (Obenauer, 2004) which are aimed at distilling different interpretations of questions in an empirically-driven way. We can also show very clearly that rhetorical questions can be answered in particular contexts, adding to the work of Caponigro and Sprouse (2007) and Rohde (2006) on the communicative function of rhetorical questions. This opens up the more general issue of how a theoretically as well as empirically well-motivated categorisation of questions could look like: for the purpose at hand, IAT and the four-way distinction into pure,

[14] In this framework, 'question φ' corresponds to an information-seeking question about the hearer's opinion whether φ is the case.

assertive, rhetorical and challenge questions serves as a good starting point, but there is clear evidence that this set is not exhaustive.

We cannot fully substantiate the hypothesis of Frank (1990) who states that "RQs [are found] in contexts where rhetoric is desired for accomplishing persuasion," for although our corpora are argumentative, they are not necessarily persuasive. For example in the televised US presidential debates, the intention of the participants is not to persuade the opponents or the hearers, but to present their own standpoint on certain issues. Similarly, public deliberations are not always conducted in order to persuade one part of the public, but in order to get the space of opinions on a public issue. Therefore, an empirical substantiation of Frank's claim would require a specific type of argumentative data as well as a measure regarding its persuasiveness – an interesting avenue for further work, but beyond the scope of the paper.

The framework we present also says nothing of the semantics (in the sense of Hamblin (1973) or Ginzburg (2012)) of the relationships between questions and their responses, nor of the semantics (in the sense of Baroni et al. (2011)), nor of the syntactic relationships, nor of the information theoretic relationships (Harrah, 1961), nor of the epistemological relationships (Hintikka, 1983), *inter alia*. Adducing such approaches would enable the framework to be refined both in its account of immediate, proximal discursive relationships between questions and their responses and also the longer-term discursive ramifications of questions. Bridging between these diverse perspectives on questions is a signal benefit of the approach we are advocating here.

8. Conclusions

In this paper, we combined empirically grounded corpus linguistics and analytically driven pragmatic theory in the study of questions. The dataset that we use is openly available at corpora.aifdb.org and consists of a variety of spontaneous, natural language argumentative dialogues across languages. Through deep algorithmic analysis of the data, we can test several hypotheses, clearly demonstrating for the first time that (particular types of) questions directly catalyse particular argument structure. We have also shown that the illocutionary consequences of non-canonical questions are much more varied than previously thought.

Looking at the actual linguistic surface in more detail is the natural extension of the more pragmatic investigation pursued here: by tying the type and the linguistic shape of the question, i.e. its syntactic shape or even its intonation, to the types of responses they get, we can contribute an empirical and cross-linguistic angle to current models of questions in dialogue, as for instance pursued in (Biezma and Rawlins, 2017). Our goal here is to lay the foundation upon which study of this interplay between linguistic surface, empirical generalisation and pragmatic theory can be built.

References

Allwood, Jens, 2000. An activity-based approach to pragmatics. In: Abduction, Belief and Context in Dialogue: Studies in Computational Pragmatics. John Benjamins, Amsterdam, pp. 47–80.

Anzilotti, Gloria Italiano, 1982. The rhetorical question as an indirect speech device in English and Italian. Can. Mod. Lang. Rev. 38 (2), 290–302.

Aristotle, 2009. Topics Books I and VIII, with Excerpts from Related Texts, Smith, R. (Translated with a Commentary). Oxford University Press, Oxford.

Baroni, Pietro, Caminada, Martin, Giacomin, Massimiliano, 2011. An introduction to argumentation semantics. Knowl. Eng. Rev. 26, 365–410, 12.

Berninger, Ginger, Garvey, Catherine, 1981. Relevant replies to questions: answers versus evasions. J. Psycholinguist. Res. 10, 403–420.

Bex, Floris, Grasso, Floriana, Green, Nancy, Paglieri, Fabio, Author, 2017. Argument Technologies: Theory, Analysis, and Applications. College Publications.

Bex, Floris, Lawrence, John, Snaith, Mark, Reed, Chris, Oct 2013. Implementing the argument web. Commun. ACM 56 (10), 66–73.

Biezma, María, Rawlins, Kyle, 2017. Or what? Semantics and Pragmatics 10.

Brown, Penelope, Levinson, Stepehn C., 1987. Politeness: Some Universals in Language Usage. Cambridge University Press, Cambridge.

Budzynska, Katarzyna, Janier, Mathilde, Author, Saint-Dizier, Patrick, 2016. Theoretical foundations for illocutionary structure parsing. Argument Comput. 7 (1), 91–108.

Budzynska, Katarzyna, Janier, Mathilde, Kang, Juyeon, Reed, Chris, Saint-Dizier, Patrick, Stede, Manfred, Yaskorska, Olena, 2014. Towards argument mining from dialogue. In: Parsons, et al. (Eds.), Proceedings of the Fifth International Conference on Computational Models of Argument (COMMA 2014). IOS Press, pp. 185–196.

Budzynska, Katarzyna, Janier, Mathilde, Reed, Chris, Saint-Dizier, Patrick, Stede, Manfred, Yakorska, Olena, 2014. A model for processing illocutionary structures and argumentation in debates. In: Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14).

Calhoun, Sasha, Carletta, Jean, Brenier, Jason M., Mayo, Neil, Jurafsky, Dan, Steedman, Mark, Beaver, David, 2010. The nxt-format switchboard corpus: a rich resource for investigating the syntax, semantics, pragmatics and prosody of dialogue. Comput. Humanit. 44 (4), 387–419.

Caponigro, Ivano, Sprouse, Jon, 2007. Rhetorical questions as questions. In: Puig-Waldmüller, E. (Ed.), Proceedings of Sinn und Bedeutung. Universitat Pompeu Fabra, Barcelona, pp. 121–133.

Carletta, Jean, Isard, Amy, Isard, Stephen, Kowtko, Jacqueline C., Doherty-Sneddon, Gwyneth, Anderson, Anne H., 1997. The reliability of a dialogue structure coding scheme. Comput. Ling. 23 (1), 13–31.

Cerović, Marijana, 2016. When suspects ask questions: rhetorical questions as a challenging device. J. Pragmat. 105, 18–38.

Chesnevar, Carlos, McGinnis, Jarred, Modgil, Sanjay, Rahwan, Iyad, Author, Simari, Guillermo, South, Matthew, Vreeswijk, Gerard, Willmott, Steven, 2006. Towards an argument Interchange format. Knowl. Eng. Rev. 21 (4), 293–316.

Dayal, Veneeta, 2016. Questions. Oxford University Press, Oxford.

Dung, Phanminh, 1995. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. Artif. Intell. 77 (2), 321–357.

Frank, Jane, 1990. You call that a rhetorical question?: forms and functions of rhetorical questions in conversation. J. Pragmat. 14, 723–738.

Freed, Alice F., 1994. The form and function of questions in informal dyadic conversation. J. Pragmat. 21 (6), 621–644. https://doi.org/10.1016/0378-2166(94)90101-5. ISSN 0378-2166.

Freeman, James B., 1991. Dialectics and the Macrostructure of Arguments: a Theory of Argument Structure. Foris Publications.

Ginzburg, Jonathan, 2012. The Interactive Stance: Meaning for Conversation. Oxford University Press.

Ginzburg, Jonathan, Yusupujiang, Zulipiye, Li, Chuyuan, Ren, Kexin, Lupkowski, Pawel, September 2019. Characterizing the response space of questions: a corpus study for English and Polish. In: Proceedings of the 20th Annual SIGdial Meeting on Discourse and Dialogue, pp. 320–330.

Gold, Valentin, El-Assady, Mennatallah, Hautli-Janisz, Annette, Bögel, Tina, Rohrdantz, Christian, Butt, Miriam, Holzinger, Katharina, Keim, Daniel, 2017. Visual linguistic analysis of political discussions: Measuring deliberative quality. Dig. Scholar. Hum. 32 (1), 141–158. https://doi.org/10.1093/llc/fqv033.
Gordon, Thomas, Prakken, Henry, Walton, Douglas, 2007. The Carneades model of argument and burden of proof. Artif. Intell. 171, 875–896.
Groarke, Leo, 1999. Deductivism within pragma-dialectics. Argumentation 13, 1–16.
Groenendijk, Jeroen, Stokhof, Martin, 1984. Studies on the Semantics of Questions and the Pragmatics of Answers. PhD thesis. Institute for Logic, Language and Computation (ILLC), Amsterdam.
Guerzoni, Elena, 2003. Why Even Ask?: on the Pragmatics of Questions and the Semantics of Answers. PhD thesis. Dept. of Linguistics and Philosophy.
Hamblin, Charles, 1970. Fallacies. Methuen, London.
Hamblin, Charles L., 1973. Questions in montague English. Found. Lang. 10, 41–53.
Harrah, David, 1961. A logic of questions and answers. Philos. Sci. 28 (1), 40–46.
Heritage, John, 2002. The limits of questioning: negative interrogatives and hostile question content. J. Pragmat. 34 (10), 1427–1446. https://doi.org/10.1016/S0378-2166(02)00072-3. ISSN 0378-2166, Negation and Disagreement.
Heritage, John, 2012. Epistemics in action: action formation and territories of knowledge. Res. Lang. Soc. Interact. 45 (1), 1–29.
Hintikka, Jaakko, 1983. New Foundations for a Theory of Questions and Answers. Springer Netherlands, Dordrecht, pp. 159–190.
Hitchcock, David, 2020. Arguing for questions. In: van Eemeren, Frans H., Garssen, Bart (Eds.), From Argument Schemes to Argumentative Relations in the Wild: A Variety of Contributions to Argumentation Theory. Springer International Publishing, pp. 167–184.
Janier, Mathilde, Lawrence, John, Reed, Chris, 2014. Ova+: an argument analysis interface. In: Parsons, et al. (Eds.), Proceedings of COMMA 2014, Computational Models of Argument: Proceedings of COMMA, vol. 266. IOS Press, pp. 463–464.
Johnson, Ralph H., 2014. The Rise of Informal Logic: Essays on Argumentation, Critical Thinking, Reasoning and Politics. University of Windsor.
Jurafsky, Daniel, Shriberg, Elizabeth, Biasca, Debra, 1997. Switchboard Swbd-Damsl Shallow-Discourse-Function Annotation – Coders Manual, Draft 13. Technical report. University of Colorado Institute of Cognitive Science.
Kalouli, Aikaterini-Lida, Kaiser, Katharina, Hautli-Janisz, Annette, Kaiser, Georg A., Butt, Miriam, 2018. A multilingual approach to question classification. In: Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018), Paris, France, May 2018. European Language Resources Association (ELRA). ISBN 979-10-95546-00-9.
Katzav, Joel, Reed, Chris, 2008. Modelling argument recognition and reconstruction. J. Pragmat. 40, 155–172.
Konat, Barbara, Lawrence, John, Park, Joonsuk, Author, Author, 2016. A corpus of argument networks: using graph properties to analyse divisive issues. In: Proceedings of the 10th Language Resources and Evaluation Conference (LREC 2016), pp. 3899–3906.
Koshik, Irene, 2002. A conversation analytic study of yes/no questions which convey reversed polarity assertions. J. Pragmat. 34 (12), 1851–1877. https://doi.org/10.1016/S0378-2166(02)00057-7. ISSN 0378-2166.
Koshik, Irene, 2003. Wh-questions used as challenges. Discourse Stud. 5 (1), 51–77.
Lawrence, John, Duthie, Rory, Budzysnka, Katarzyna, Author, 2016. Argument analytics. In: Baroni, P., Stede, M., Gordon, T. (Eds.), Proceedings of the Sixth International Conference on Computational Models of Argument (COMMA 2016). IOS Press, Berlin.
Lawrence, John, Reed, Chris, 2016. Argument mining using argumentation scheme structures. In: Baroni, P., Stede, M., Gordon, T. (Eds.), Proc. Of COMMA. IOS Press, Berlin.
Lawrence, John, Reed, Chris, 2019. Argument mining: a survey. Comput. Ling. 1–54. https://doi.org/10.1162/coli_a_00364. Early Access.
Lawrence, John, Reed, Chris, 2020. Argument mining: a survey. Comput. Ling. 45 (4), 765–818.
Lawrence, John, Snaith, Mark, Konat, Barbara, Budzynska, Katarzyna, Reed, Chris, June 2017. Debating technology for dialogical argument: sensemaking, engagement, and analytics. ACM Trans. Internet Technol. 17 (3), 24:1–24:23, ISSN 1533-5399.
Łupkowski, Paweł, Ginzburg, Jonathan, 2016. Query responses. Journal of Language Modeling 4, 245–292.
Mackenzie, Jim, Dec 1990. Four dialogue systems. Stud. Logica 49 (4), 567–583. ISSN 1572–8730.
Obenauer, Hans-Georg, 2004. Nonstandard wh-questions and alternative checkers in pagotto. In: The Syntax and Semantics of the Left Periphery. Mouton de Gruyter, Berlin, pp. 343–384.
Pease, Alison, Lawrence, John, Bduzynska, Katarzyna, Corneli, Joseph, Reed, Chris, 2017. Lakatos-style collaborative mathematics through dialectical, structured and abstract argumentation. Artif. Intell. 246, 181–219.
Prakken, Henry, 2005. Coherence and flexibility in dialogue games for argumentation. J. Logic Comput. 15 (6), 1009–1040.
Prakken, Henry, 2006. Formal systems for persuasion dialogue. Knowl. Eng. Rev. 21, 163–188.
Quirk, Randolf, Greenbaum, Sidney, Leech, Geoffrey, Svartvik, Jan, 1985. A Comprehensive Grammar of the English Language. Longman, New York.
Reed, Chris, 2011. Implicit speech acts are ubiquitous. why? they join the dots. In: Zenker (Ed.), Proceedings of the Conference on Argumentation: Cognition and Community (OSSA-2011). Windsor, ON.
Reed, Chris, Budzynska, Katarzyna, 2011. How dialogues create arguments. In: Proceedings of the 7th Conference of the International Society for the Study of Argumentation (ISSA 2010). SicSat. ISSA.
Reed, Chris, Budzynska, Katarzyna, 2012. The structure of ad hominem dialogues. In: Proceedings of the 4th International Conference on Computational Models of Argument (COMMA 2012). IOS Press, pp. 410–421.
Reed, Chris, Budzynska, Katarzyna, Duthie, Rory, Janier, Mathilde, Konat, Barbara, Lawrence, John, Pease, Alison, Snaith, Mark, 2017. he argument web: an online ecosystem of tools, systems and services for argumentation. Philos. Technol. 30 (2), 137–160.
Roberts, Craige, 2012. Information structure: Towards an integrated formal theory of pragmatics. Semantics and Pragmatics 5, 1–69.
Rohde, Hannah, 2006. Rhetorical questions as redundant interrogatives. San Diego Linguist. Pap. 2, 134–168.
Sadock, Jerrold, 1971. Queclaratives. In: Papers from the Seventh Regional Meeting of the Chicago Linguistics Society (CLS 7). Chicago Linguistics Society, Chicago, pp. 223–331.
Schaffer, Deborah, 2005. Can rhetorical questions function as retorts?: Is the Pope Catholic? Journal of Pragmatics 37 (4), 433–460. https://doi.org/10.1016/j.pragma.2003.12.007. ISSN 0378-2166.
Searle, John, 1969. Speech Acts: an Essay in the Philosophy of Language. Cambridge University Press, Cambridge.
Searle, John, Vanderveken, Daniel, 1985. Foundations of Illocutionary Logic. Cambridge University Press, Cambridge.
Skilton, Amalia, 2017. Assertive questions in Máíhōki. J. Pragmat. 109, 121–136.
Stede, Manfred, Schneider, Jodi, 2018. Argumentation Mining. Synthesis Lectures on Human Language Technologies. Morgan & Claypool Publishers. https://doi.org/10.2200/S00883ED1V01Y201811HLT040.
Stivers, Tanya, 2010. An overview of the question–response system in American English conversation. Journal of Pragmatics 42 (10), 2772–2781.
Stivers, Tanya, Enfield, N.J., 2010. A coding scheme for question-response sequences in conversation. Journal of Pragmatics 42 (10), 2620–2626. https://doi.org/10.1016/j.pragma.2010.04.002. ISSN 0378-2166, Question-Response Sequences in Conversation across Ten Languages.
Truckenbrodt, Hubert, 2006. On the semantic motivation of syntactic verb movement to c in German. Theor. Ling. 32 (3), 257–306.
van Eemeren, Frans, Grootendorst, Rob, 1992. Argumentation, Communication, and Fallacies. Routledge.
Visser, Jacky, Konat, Barbara, Duthie, Rory, Koszowy, Marcin, Budzynska, Katarzyna, Reed, Chris, 2019. Argumentation in the 2016 US presidential elections: annotated corpora of television debates and social media reaction. Comput. Humanit. https://doi.org/10.1007/s10579-019-09446-8. ISSN 1574-0218.
Visser, Jacky, Konat, Barbara, Duthie, Rory, Koszowy, Marcin, Budzynska, Katarzyna, Reed, Chris, 2020. Argumentation in the 2016 us presidential elections: annotated corpora of television debates and social media reaction. Comput. Humanit. 54, 123–154.
Visser, Jacky, Lawrence, John, Wagemans, Jean, Reed, Chris, 2018. Revisiting computational models of argument schemes: classification, annotation, comparison. In: Modgil, S., Budzynska, K., Lawrence, J. (Eds.), Proceedings of the Seventh International Conference on Computational Models of Argument (COMMA 2018). IOS Press, pp. 313–324.

Wagemans, Jean, 2016. Constructing a periodic table of arguments. In: Bondy, P., Benacquista, L. (Eds.), Proceedings of OSSA 2016: Argumentation, Objectivity, and Bias, pp. 1–12.
Walton, Douglas, 1988. Question-asking fallacies. In: Meyer, Michel (Ed.), Questions and Questioning. Walter de Gruyter, Berlin, pp. 195–221.
Walton, Douglas, Author, Macagno, Fabrizio, 2008. Argumentation Schemes. Cambridge University Press.
Walton, Douglas, Krabbe, Erik C.W., 1995. Commitment in Dialogue: Basic Concepts of Interpersonal Reasoning. State University of New York Press.
Webber, Bonnie, Prasad, Rashmi, Lee, Alan, Joshi, Aravind, 2017. The Penn Discourse Treebank 3.0 Annotation Manual. Technical report, LDC. https://catalog.ldc.upenn.edu/docs/LDC2019T05/PDTB3-Annotation-Manual.pdf.

Annette Hautli-Janisz is PostDoc in Linguistics at the University of Konstanz. Her research in computational linguistics covers a range of topics, with the most recent focus on the linguistically-motivated, computational analysis of multiparty discourse. She also has a track record of combining Visual Analytics and (computational) linguistics and developing resources using various machine learning techniques.

Katarzyna Budzynska is Associate Professor at Warsaw University of Technology and has a track record of formally analysing communicative processes of argumentation and persuasion linking philosophy of language, corpus linguistics and computational social science. She is spearheading computational ethos, a new research area in AI and heads the Laboratory of The New Ethos.

Conor McKillop is PhD student in the Centre for Argument Technology (ARG-tech) at the University of Dundee, working on computational argumentation. His research explores the application of argument mining and streaming data visualisation to monitor and improve large-scale deliberative processes.

Brian Plüss is PostDoc in the Centre for Argument Technology (ARG-tech) at the University of Dundee, working on applications of argument technologies and analytics to enhance everyday communication. His research combines computational linguistics, manual and automatic analyses of dialogue and argumentation, data analytics and interactive visualisations, with a special interest in applications in political communication.

Valentin Gold is PostDoc in political science at the University of Göttingen and has extensive knowledge in interdisciplinary, empirical work on deliberative communication. Recently, he has worked on operationalising the concept of deliberation in order to make it accessible to computational approaches.

Chris Reed is Professor of Computer Science and Philosophy and heads the Centre for Argument Technology (ARG-tech) at the University of Dundee, Scotland. Chris has a track record of working at the overlap between argumentation theory and artificial intelligence, developing foundational theory in philosophy, linguistics and cognitive science that facilitates the development of practical AI applications.