Modelling argumentative behaviour in parliamentary debates: data collection, analysis and test case

Volha Petukhova[1], Andrei Malchanau[1] and Harry Bunt[2]
[1] Spoken Language Systems Group, Saarland University, Saarbrücken, Germany
{v.petukhova,andrei.malchanau}@lsv.uni-saaarland.de
[2] Tilburg Centre for Communication and Cognition, Tilburg University, The Netherlands
harry.bunt@uvt.nl

June 13, 2017

Abstract

In this paper we apply the information state update (ISU) machinery to tracking and understanding the argumentative behaviour of participants in a parliamentary debate in order to predict its outcome. We propose to use the ISU approach to model the arguments of the debaters and the support/attack links between them as part of the formal representations of a participant's information state. We first consider the identification of claims and evidence relations to their premises as an argument mining task. It is not sufficient, however, to indicate what relations occur without establishing how these relations are created and verified during the interaction. For this purpose the model requires a detailed specification of the creation, maintenance and use of shared beliefs. The ISU model provides procedures for incorporating beliefs and expectations shared between speaker and hearers in the tracking model. To evaluate the content of the tracked information states, we compare them to those of the human 'concluder' who wraps up a debate, stating the claims which the majority of the debaters have agreed on.

1 Introduction

Argumentation constitutes a major component of human intelligence. Many domains, including philosophy, politics, journalism, law, and theology rely on the use of arguments. Argumentation is used to justify solutions to many problems. The problem of understanding argumentation has been addressed by many researchers in different fields including philosophy, logic and artificial intelligence. In natural language processing, a surge of interest in argumentation mining tasks has recently been observed. Much successful work has been done to extract arguments and analyse their structure. Argumentation mining methods are mostly focused on the identification and classification of argument components.

The argument detection task is generally defined as a binary task by separating argumentative and non-argumentative units. Based on a domain-independent theory of argumentation schemes [1, 2] an accuracy was obtained of 73.75% in identifying argumentative sentences in the Araucaria corpus [3], using features such as word pairs, verbs, and keywords indicative for argumentative discourse, e.g. discourse markers.

Argumentative structures have been well understood and modelled for argumentative texts and to a certain extent also for two-party argumentative dialogues, see e.g. [5, 12, 13, 14]. In order to identify arguments and relations between their constituents, discourse relations are often considered, inspired by Rhetorical Structure Theory, see [15]. Discourse relations help to identify to which other propositions a proposition serves as evidence and from which other propositions it receives support. One of the first studies on argument component classification is *Argumentative Zoning* [4]. In this study, sentences within one argument and texts as a whole are classified as having one of the rhetorical relations such as result, purpose, background, solution, and scope. When applied to scientific articles, this prediction method achieved an F-score of 0.46.

Although being very important, these methods are insufficient for many applications such as for example argumentative multi-agent multimodal interactions. For instance, it is not sufficient to indicate what claims and relations do occur without indicating how these relations are established and verified during a debate session. For a computer system to be engaged in the exchange of arguments, either as a direct participant or as side-participant like an observer, understanding the strength and sustainability of arguments along with the understanding of their structure is essential. The beliefs of a rational agent engaged in argumentation should be characterized not only by beliefs concerning his own supporting arguments but also by the beliefs concerning his partners' beliefs and relations between them. Cohen (1987), who provided an in-depth analysis of argument structures, emphasized that a tracking model of mutual beliefs between speakers and addressees is required, see [19]. The information state update (ISU) approach, see [20, 21], applied successfully in a variety of dialogue tasks, provides a computational model for the creation of shared beliefs and specifies mechanisms for their transfer.

In this paper we present an approach to modelling the interaction in parliamentary debates. We present a debate model based on an analysis of the tasks of the participants, of the structure of their contributions, and of the relations between them. In this analysis, an argument structure is defined in terms of claims and evidence and the connections between them.

This paper is structured as follows. In Section 2 we discuss the application domain, specifying participant roles and tasks, and highlighting important interactive phenomena to be modelled. Sections 3 and 4 describe and analyse our data and discuss ways to segment and annotate debates. Section 5 presents the semantic framework within which we model debate interactions; it describes the information state update process in debates, leading to the creation of mutual beliefs. Section 6 proposes an evaluation method to validate to what extent the system's understanding of debate arguments corresponds to that of human understanding and can be used to predict debate outcomes. Section 7 summarizes our conclusions and indicates perspectives for further research.

2 Application domain: the nature of parliamentary debates

A parliamentary debate is a communication process in which participants argue *for* or *against* a *motion*. A debate is thus a type of dialogue, but it differs from the well-studied task-oriented dialogue in the number and roles of its participants, their tasks and the form of interaction. A debate *session* is motivated by a motion which is concerned either with a general topic, e.g. health, or with a proposed law (legislation). The so-called 'closure motion' is a special motion which ends the debate and leads to voting. The motion is announced by a *Moderator* (or Speaker in the UK). The Moderator chairs the session, opens and closes it, and regulates the turn-taking. The actual debate starts by the *Proponent* presenting the motion and arguments in favor of it. An *Opponent* attacks the proponent's arguments. There is a number of *Proponent's Seconders* and *Opponent's Seconders* whose task is to counter-attack either the opponent's or the proponent's arguments, respectively, or those of their seconders.

There are different debate types depending on type of motion and status of participating parliamentarians: debate on legislation, general debate and short debate. In this study we considered general parliamentary debates - plenary sessions in the UK Youth Parliament (YP)[1]. Members of the Youth Parliament (MYPs) are elected to represent their constituency and do not belong to any political party. The YP does not decide on legislation, but simulates the environment of the actual Parliament plenary sessions discussing youth-specific current affairs issues. The results of YP debates are recorded in a publication called 'manifesto' which is available for the members of the actual Parliament. The general debate discussion is closed by the *Concluder* who wraps up the debate summarizing commonly agreed points and the most evident disagreements. Figure 1 shows a conceptual model of YP debates.

A YP debate is a formal interaction with certain rules, traditions and even rituals. Speakers present their positions by arguments that may take the form of quite lengthy verbal contributions with an articulate internal structure. Other MYPs listen to arguments and as a rule do not interrupt the current speaker.

[1] Youth Parliaments have been founded in many European countries and all over the world, e.g. in Greece, South Africa, Columbia. The European Youth Parliament is also active since 1987.

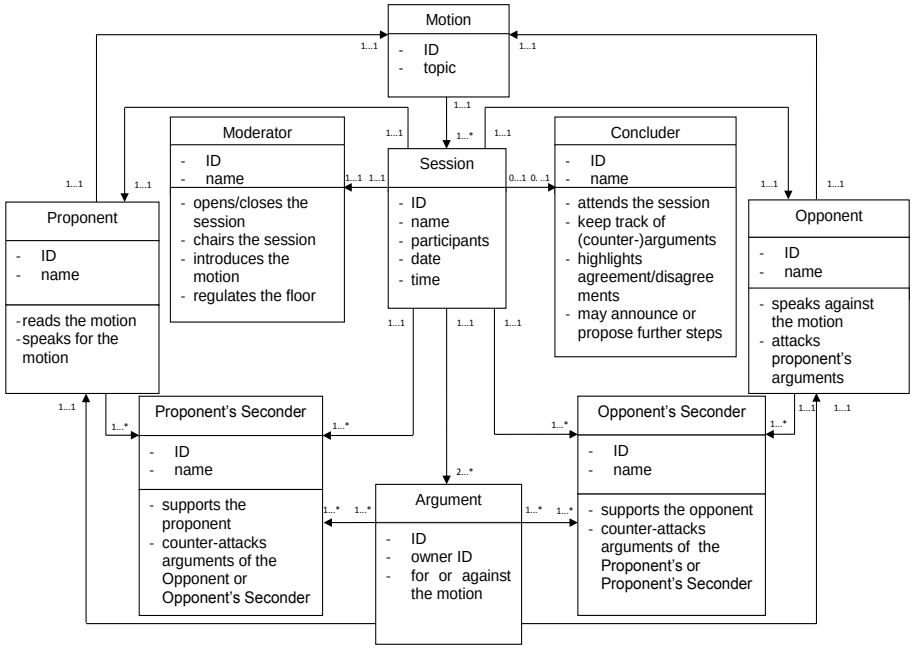

Figure 1: Conceptual model of a parliamentary debate session.

The right to have a turn is regulated by the Moderator. The Moderator nominates the next speaker. Participants who want to take the next turn, rise or half-rise from their seats in a bid to get the Moderator's attention.

3 Debate data: segmentation and annotations

The data that we analysed forms three UK YP[2] sessions. These sessions are video recorded and available on Youtube[3]. The YP members, aged 11-18, debate issues addressing three different motions: (1) sex and relationship education (SRE); (2) university tuition fees and (3) job opportunities for young people in the UK. The corpus is provided with automatically generated subtitles which are corrected manually.

First, segmentation has been performed together with dialogue act annotations into functional segments according to guidelines provided in ISO 24617-2 [22]. To each segment a communicative function has been assigned in one or more of the nine ISO dimensions (Task, Auto- and Allo-Feedback, Discourse Structuring, Time and Turn Management, Own and Partner Communication Management, and Social Obligation Management). The annotated corpus consists of 1388 functional segments from 35 different speakers. Table 1 provides in the last column an overview of the relative frequencies of functional tags per ISO-dimension.

3.1 Task acts

In our data, just over 41.4% of the dialogue acts performed by the debaters are Inform acts, which are often connected by discourse relations. For example,

(1) $D1_{21}$[4]: Sex education covers a wide range of issues affecting young people [*Inform*]
$D1_{22}$: These include safe sex practices, STIs and legal issues surrounding consent and abuse [*Inform Elaboration* $D1_{21}$]

We observed a small portion of set questions (3.4%) that are rhetorical in the sense they are not intended to get an answer but rather to emphasise an important issue. For example,

(2) $D3_{72}$: This is not a question of morality [*Inform*]
$D3_{73}$: But a question of equality [*Inform Contrast* $D3_{72}$]

[2] `http://www.ukyouthparliament.org.uk/`

[3] See as example `http://www.youtube.com/watch?v=g2Fg-LJHPA4`

[4] Here and henceforth D*k* stands for Debater *k*; the subscript is the index of the identified functional segment.

$D3_{74}$: Why should some children have sex education when others do not [*SetQuestion*]

About 1.7% of all task-related acts are explicit Agreements or Disagreements with previous speakers. For example,

(3) $D2_{37}$: it makes sense to teach a young person the basics of what a healthy relationship is before they want to have sex [*Inform*]
$D3_{62}$: we need a policy to reduce STDs [...] and address relationships aspects [*Inform*]
$D6_{153}$:I defend and I completely wholeheartedly agree that relationships are more important [*Agreement $D2_{37}$; $D3_{62}$*]

3.2 Dialogue control acts in debates

There are other utterances concerned mostly with Turn Management from the side of the Moderator; Time Management acts like Stallings; Own Communication Management acts like Self Corrections; Social Obligation Management acts like Thankings; and Discourse Structuring acts for signalling that the debater has finished his speech. Consider the following example:

(4) $D7_{20}$: The government needs to keep up with the media in speed and in terms of sexual imaging [*Inform*]
$D7_{21}$: Thank you [*Thanking; Closing*]
Audience: *applause* [*Positive Auto/Allo-Feedback $D7_1$ - $D7_{21}$; Thanking*]
$D8_1$: *stands up* [*Turn Take*]
M_{212}: Let's hear this young gentleman [*Turn Assign*]
M_{213}: Who I think is from London [*CheckQuestion*]
$D8_2$: *head nod* [*Confirmation M_{213}*]
M_{214}: Good [*Positive Auto/Allo-Feedback $D8_2$*]
$D8_3$: My name's Landry Adelard, MYP of London [*Turn Accept M_{212}; SelfIntroduction; Confirmation M_{213}*]

3.3 Automatic dialogue act recognition

For the automatic dialogue act recognition various machine learning techniques have been applied successfully. For example, Hidden Markov Models (HMM) have been tried for dialogue act classification in the spontaneous free two-party conversations (Stolcke et al., 2000), achieving a tagging accuracy of 71% on word transcripts. Another approach that has been applied to dialogue act recognition, by Samuel et al. (1998), uses transformation-based learning. They achieved an average tagging accuracy of 75.12% for the two-party phone negotiation data. Keizer (2003) used Bayesian Networks an average accuracy of 81% on the typed theater tickets booking system. Lendvai et al. (2004) adopted a memory-based approach, based on the k-nearest-neighbour algorithm, and report a tagging accuracy of 73.8% for the OVIS data, train information-seeking dialogues in Dutch.

Debate data, however, has certain properties that are different from the data used in classification experiments reported in the above mentioned studies. The differences become apparent if we compare dialogue act distributions in different collected dialogue data such as, for example, HCRC MapTask corpus[5] consisting of human-human two-party instructing dialogues where one participant plays the role of an instruction-giver and another participant, the instruction-follower, navigates through the map, and AMI[6] corpus containing human-human multi-party face-to-face meeting interactions of remote control design teams, with the debate data. Table 1 presents relative frequencies of dialogue act tags across ISO dimensions for 3 compared corpora. It can be observed that both AMI and HCRC MapTask contain more interactive phenomena related to explicit feedback providing indication of the speaker's and partner's processing state, as well as related to turn and time management interactive aspects. Debate participants, along with the task-related acts, were more often concerned with structuring their contributions, see also Sections 2 and 3.

We conducted series of machine-learning experiments using different features both automatically extracted from the corpora and computed using available for English linguistic parsers. In order to

[5] `http://groups.inf.ed.ac.uk/maptask/`
[6] `http://groups.inf.ed.ac.uk/ami/corpus/`

Dimension	Relative frequency (in %)		
	AMI meetings	HCRC MapTask	YP debates
Task	31.8	52.4	54.9
Auto Feedback	20.5	15.7	2.9
Allo Feedback	0.7	4.7	1.0
Turn Management	50.2	24.3	22.7
Time Management	26.7	13.4	21.1
Discourse Structuring	2.8	0.5	10.0
Own Communication Management	10.3	2.8	7.3
Partner Communication Management	0.3	0.3	0.0
Social Obligation Management	0.5	0.1	1.2

Table 1: Distribution of dialogue act tags across ISO-dimensions in terms of their relative frequency in the AMI, HCRC MapTask and YP debate corpora.)

Features set	unigrams	bi-grams	tri-grams
Chunks	0.45	0.71	0.41
Chunks, POS	0.63	0.75	0.55
Chunks, word tokens	0.66	0.68	0.60
Chunks, POS, word tokens	0.79	**0.84**	0.74
POS	0.62	0.58	0.64
POS, word tokens	**0.82**	0.79	0.76
word tokens	0.74	**0.81**	0.67

Table 2: Dialogue act classification results in terms of F-score on different feature set and with n-gram range computed for YP debate corpus.

train classifiers that are able to operate on data collected from various domains, along with frequently used n-grams and bag-of-words models we used part-of-speech (POS) information and shallow syntactic parsing features, and combinations of those. Linguistic features are expected to contribute to higher cross-domain portability of trained prediction models. For POS tagging Stanford CoreNLP[7] tagger was used and chunking was performed using Illinois shallow parser [11].

Support Vector Machine (SVM) classifier has been trained using scikit-learn implementation[8]. Training task has been defined as joined segmentation and classification task as proposed in [6]. To evaluate the classifiers' performance, the most commonly used performance metrics such accuracy, precision, recall and F-scores (harmonic mean) were computed. For the sake of simplicity, in this paper we report the best F-scores obtained in different classification experiments, see Table 2. As baseline the majority class, namely, `Task;Inform` of 0.41 has been used.

As for features, the best results were obtained on the complex features combining word bigrams, POS tags (unigrams) and chunking (bigrams) information. For indication, when trained on unigram language models only we observed the decrease in performance of about 10% compared to the performance on features combination; trained on POS tags - 20% on average, trained on chunking information - 40%. Thus, wording of an utterance is still very important, but when supplied with linguistic information the performance of the classifier improves. The general conclusion is that dialogue acts can be successfully learned from linguistically processed debate transcripts in data-oriented supervised way.

4 Detection of arguments and their structure

For our further analysis and modelling, arguments need to be identified. Toulmin (1958) proposed a scheme with six functional roles to describe the structure of an argument (see Figure 2). Based on evidence (*data*) and a generalization (*warrant*), which is possibly implicit and defeasible, a conclusion is derived. The conclusion can be *qualified*, e.g. by a modal operator indicating the strength of the inferential link between data and *conclusion*. A *rebuttal* specifies exceptional conditions that undermine this inference. A warrant can be supported by *backing*, e.g. reason, justification or motivation.

Toulmin's theory inspired many other argumentation schemes (see e.g. [12], [13]). A recently pro-

[7] `http://nlp.stanford.edu/software/corenlp.shtml`
[8] `http://scikit-learn.org/stable/`

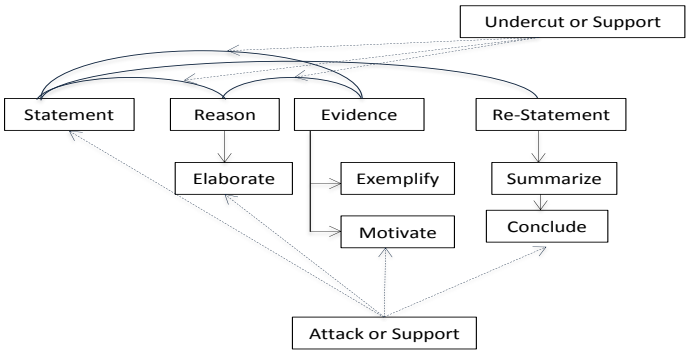

Figure 3: Analysed argument structure.

posed argumentation scheme is that of [14], where several previous approaches and theories are synthesized. It makes use of proponent and opponent moves as defined in [13]. The authors distinguish between *basic* elements of an argument which consists of a non-empty set of premises and a conclusion. Different patterns are observed linking premises and a conclusion. A premise supporting a conclusion form a basic argument. Several premises may either jointly (*linked support*) or independently (*multiple support*) support one conclusion. A premise may provide support for another premise, and indirectly support a conclusion (*serial support*). A special form of lending support to a claim is that of providing examples (*example support*).

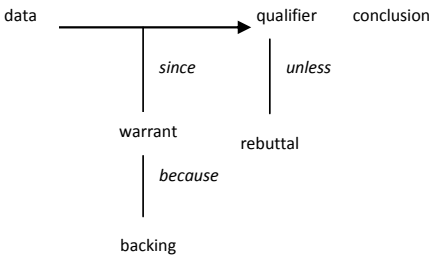

Figure 2: Argumentation diagram of Toulmin (1958).

Further, arguments can be either attacked by the opponent, anticipated by the opponent (temporal role with proponent vs opponent, e.g. express awareness of exceptions), or counter-attacked by the opponent. There are two possible ways to attack an argument: one is to present an argument against a conclusion or a premise (*rebutting*), the other is to diminish their supporting force (*undercutting*). When *counter-attacking*, it is possible to rebut the rebutter of a conclusion; to rebut an undercutter of a support link; to undercut an undercutter; and to undercut a rebutter.

Good debaters are distinguished by concise clear arguments and try to make their arguments understandable for others. In other words, if a debater wants to be successfully interpreted, he needs to signal his intentions as unambiguously as possible, e.g. by using markers or cues, unless he wants to be deliberately vague or deceptive. This is applicable not only to arguments, but also to the supporting or undermining links between them. To achieve this, debaters often use linguistic cues such as discourse markers and dialogue act announcement acts. For example, 'I will talk in favour of ... Because ... Since international research shows...'. Thus, *discourse* relations are often marked explicitly by means of discourse markers. Discourse relations can be of various types. For example, to signal linked support of two or more premises for a conclusion, two premises can be connected by Elaboration or Sequence relations. Supporting links between premises and conclusion can be of Justification, Motivation, Cause/Result, Background/Evaluation, Evidence and Circumstance type, and many others. Rebuttal or undercutting links are often enabled by presenting Contrast, Exception and other Comparatives. Discourse relations have been proposed as an explanation for the construction of coherence in discourse or at least as crucial modelling tools for capturing this coherence, see e.g. Hobbs (1985a); Mann and Thompson (1988); Sanders et al. (1992); Asher and Lascarides (2003). Discourse relations are learnable in a data-oriented way, using machine-learning techniques (see [23] and [24]). Figure 3 depicts the general analysed argument structure.

Based on discussed previous findings and defined argumentation schemes, we further segmented debates into Argumentative Discourse Units (ADUs), defined as a unit which consists of one or more

Discourse relation type	Relative frequency (in %)	Cohen's kappa scores
Elaboration**	28.1	0.67
Evidence**	21.4	0.72
Justify***	16.1	0.76
Condition***	0.7	0.34
Motivation**	1.4	0.48
Background**	0.3	0.18
Cause***	3.4	0.37
Result***	2.2	0.26
Reason*	10.6	0.65
Conclude**	5.7	0.71
Restatement***	10.1	0.76

Table 3: Distribution of Inform acts involved in a discourse relation in terms of their relative frequency in the corpus (* defined in DPTB; ** defined by Hovy and Maier, 1995; *** in both taxonomies) and the inter-annotator agreement in terms of Cohen's kappa.)

premises and one conclusion, possibly restated or paraphrased several times by the same speaker. To identify ADUs, we followed the approach proposed by Peldszus and Stede (2013)[14], who suggest to first segment into Elementary Discourse Units (EDUs) as minimal discourse building blocks, then establish relationships between two or more EDUs, and combine those into ADUs.

Segmentation into EDUs is well established for written discourse, where syntactic clauses are considered as such units. For spoken discourse prosodic units [25], speaking turns [26], and intentionally defined discourse segments [27] have been proposed. For debates, turns are obviously too coarse to be considered, as they are too lengthy and may contain more then one argument. Prosodic units like interpausal units, e.g. bounded by at least 100ms of silence [28], are too fine-grained since debaters often make pauses when emphasising a single word or phrase. EDUs in our data mostly coincide with intentionally defined units such as dialogue acts. The Task dialogue acts related to previous discourse by means of a discourse relation form the best-defined EDU for spoken discourse. In our corpus 1021 EDUs were identified meaning that about 73.6% of all dialogue acts constitute a part of an EDU.

Discourse relations were annotated using the annotation scheme designed for the Penn Discourse TreeBank (PDTB) corpus [29]), extended with discourse segment relations from the taxonomy proposed in [30]. Table 3 presents the types and frequencies of the relations along with the inter-annotator agreement reached annotating each relation type. It should be noted here that the inter-annotator agreement between three experienced annotators was moderate on this task (Cohen's kappa 0.54 on average), however on some relations like Elaboration, Evidence, Justification, Reason, Conclude and Restatement, which are important for our further processing, we obtained a substantial agreement (Cohen's kappa around 0.71).

Identifying ADUs, we observed a very frequent pattern[9]: an ADU will mostly start with a simple Inform act and end when an Inform Conclude or Restatement is identified, or before another Inform act which is not involved in any discourse relation. We assigned an index to each argument conclusion. Consider the following example:

(5) $D2_{30}$: Essentially we are experiencing a tragic loss of childhood [*Inform*]
$D2_{31}$: a walk down the high street reveals a depressing trend towards essentially adult's designs [*Inform Evidence* $D2_{30}$]
$D2_{32}$: children's pencil cases bearing playboy symbols [*Inform Evidence* $D2_{30}$*; Inform Motivate* $D2_{31}$]
$D2_{33}$: our children being sexualized too young [*Inform Result* $D2_{30}$*,* $D2_{31}$,$D2_{32}$;*Cause* $D2_{34}$]
$D2_{34}$: we must aim to protect this short-lived innocence [*Inform Result* $D2_{33}$]
$D2_{35}$;D2_2.1[10]: SRE is simply inappropriate within a primary curriculum [*Inform Conclude* $D2_{30}$ *-* $D2_{34}$*; Conclusion D2_2.1*]

In our data, 118 ADUs were identified in total, 37 to 40 per session.

[9]The inter-annotator agreement between three experienced annotators on this task was very high, 0.87 in terms of Cohen's kappa.

[10]Here and henceforth _x.y is the index assigned to the conclusion of an ADU, where x indicates the debater index and y stands for the index of an ADU conclusion.

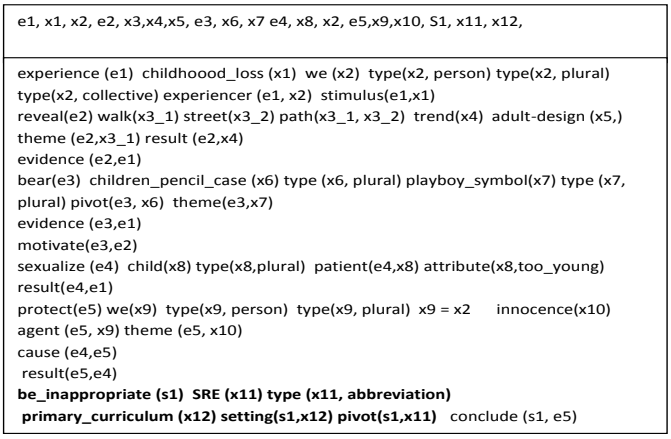

Figure 4: Example of DRS representation of the identified ADU presented in (5).

The semantic content of an argument is incrementally constructed from its premises and conclusion using the representation formalism of DRSs [31, 32]. An example of the DRS representation for the argument exemplified in (5) is shown in Fig. 4, where the conclusion is marked in bold face. Computing semantic content for each conclusion, we observed that for instance in the session on SRE nine main claims (henceforth also called 'propositions') are identified:

(6) p_1: SRE should be compulsory
p_2: SRE should be introduced in primary school
p_3: SRE should be valued by parents
p_4: SRE should be provided even in faith schools
p_5: SRE should be counter-part for media images
p_6: SRE should be about both sex and relationships education
p_7: SRE should be provided in an appropriate context
p_8: Government listens to the YP's campaign
p_9: SRE should not be provided at school but in peer-education

To incorporate support and attack links[11], we need the full specification of participants' information states. Only in this way can we establish beliefs concerning previously presented arguments that the current speaker either supports or attacks. We start with identified explicit and implicit agreement and disagreement dialogue acts signalling support or attack of arguments through the *functional dependence relations* defined in [22] between the detected argument conclusions. Consider the discussion on when SRE should be introduced at school.

(7) $D1_{47}$;$D1_{1.2}$: Sex education needs to start early to stop the damage before it's too late [*Inform*]
$D2_5$;$D2_{2.1}$: SRE is simply inappropriate within a primary curriculum [*Inform*& *Disagreement* $D1_{47}$] - Attack 1.2
$D7_2$;$D7_{7.1}$: I think involving sex education in primary school is perfectly sensible [*Inform*& *Agreement* $D1_{47}$& *Disagreement* $D2_5$]- Support 1.2/Attack 2.1

Debater 1 (Motion Proponent) states as his opinion that SRE needs to start early (read in primary school). Debater 2 thinks that SRE in primary school is inappropriate. Debater 7 supports SRE in primary school (argument 1.2) and thereby attacks the arguments 2.1.

The proposed complete argument identification and processing flow is illustrated in Figure 5. The process starts with segmenting a debater's turn into functional segments each of them having one or more communicative functions according to the ISO 24617-2 dialogue act annotation standard. Subsequently, we propose to identify discourse relations between dialogue acts that are mostly Informs and cluster them into EDU segments, and successively into ADUs as described in Section 4. The ADU's main statement can be then extracted which is either the opening Inform or the closing Conclusion or Re-statement. These propositions can be linguistically processed using the state-of-the-art parsers of various types, e.g.

[11]Note we do not distinguish between rebuttals and undercutters in this study.

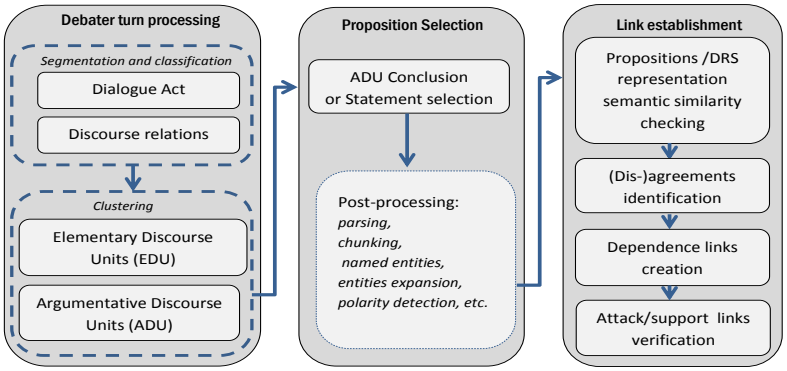

Figure 5: Argument identification and processing flow.

syntactic parser and (shallow) semantic parsers. One of the tools that incorporates many of the required existing up-to-date semantic analyzers is Boxer[12]. It takes as input CCG (Combinatory Categorial Grammar) derivations and produces DRSs (Discourse Representation Structures). Further, in many cases the identification of attack/support links requires an additional step, since our analysis showed most of them are expressed by implicit (dis-)agreements. We suggest to check the selected propositions for semantic similarity combined with polarity detection. Similarity checking operation can be performed on proposition's exact wording or using obtained semantic representations like DRSs. Additionally, to achieve wider coverage of possible linguistic expressions, entities expansion steps may be needed, e.g. expansion through lists of synonyms, homonyms and/or entities with some ontological relations using available resources like, for example, WordNet[13]. Semantically similar propositions produced by different speakers are selected and functional dependence links are established between them. Finally, after polarity detection, similar positive propositions are linked as having support link, and similar negative ones as having an attack link. To make the picture complete, arguments represented by their main propositions and support/attack links between them are semantically modelled as part of the debaters' information states (see Section 5).

5 Computing information states

Information state update approaches analyse dialogue utterances in terms of effects on the information states of the dialogue participants. An 'information state' (also called 'context') is the totality of a dialogue participant's beliefs, assumptions, expectations, goals, preferences and other attitudes that may influence the participant's interpretation and generation of communicative behaviour [22]. Dialogue acts are viewed as corresponding to update operations on the information states and consist of two main components: (1) the type of communicative act, expressed as its *communicative function*, e.g. Inform, Question, Request, etc., and (2) the *semantic content*, i.e. the objects, events, situations, relations, properties, etc. are addressed. Bunt (2014)[33] provides a detailed specification of the update semantics of dialogue acts.

5.1 Mutual belief creation and transfer in debates

To be successful in debate, the participants have to coordinate their activities on many levels. In the speaker role, a participant produces utterances with the aim to be understood by others. In dialogue act theory, understanding that a certain dialogue act is performed means creating the belief that the preconditions hold which are characteristic for that dialogue act. As the ultimate goal of a debater is to convince his audience of the rightness of his position, he wants the addressees to incorporate his beliefs as beliefs of their own (belief *adoption*).

[12] `http://svn.ask.it.usyd.edu.au/trac/candc/wiki/boxer`
[13] `http://wordnet.princeton.edu`

The coordination of the beliefs and assumptions of the participants is a central issue in any communication. A set of propositions that the dialogue participants mutually believe is called their *common ground*, and the process of establishing and updating the common ground is called *grounding*. The speaker expects under 'normal input-output' conditions [34] that what he is saying is perceived and understood as intended. These expectations may be strengthened when there is positive evidence from the audience, and if negative feedback occurs the expectations are canceled. Such evidence takes the form of explicit or implicit positive feedback; we observed instances of feedback on what was just said, such as laughter, applause, verbal 'yeah' and 'hear! hear!'. However, not all propositions are addressed immediately, and a debater may not get a chance to react to or correct misinterpretations or rejections of his contributions.

In parliamentary debates, where political confrontations and ideological convictions often play a significant role, the goals of a debater depend on the type of debate. In legislation debates the main goal is to gain the majority of supporters in terms of votes. A lot of preparatory work is done before the actual debate takes place, in committees and lobbies. To achieve their main goal parliamentarians may be ready to compromise on some points and negotiate on others. A governing party with a majority in the parliament has a bigger chance to get their beliefs adopted by the majority, therefore has stronger initial expectations. Parliamentarians also have certain knowledge about their opponents and their seconders, which should be modelled in the initial dialogue context together with knowledge about common and individual goals, and should be taken into consideration when computing the strength of expectations concerning the outcome of a debate. In HCI research it is common to incorporate user models where all available information about dialogue participants is specified [35]. This type of information is typically useful to design adaptive human-computer systems and can be profitably used when modelling interactive behaviour in dialogue, in particular related to grounding.

In general YP debates no strong political division is obvious a priori, and it is reasonable to assume that each debater expects that many of his partners will adopt his beliefs. At least, this is what he strives for, otherwise it would make little sense to participate in such a debate. With this goal in mind, a participant does his best to be convincing and persuasive, presenting his claims and evidence as convincingly as possible. Example (5) from our corpus can be used to illustrate this. Proponent D_1 presents arguments with the conclusion p_1: *'SRE should be introduced in the primary school curriculum'*. The debaters $D_2...D_n$ understand this proposition and make it part of their common ground. Following the computational model of grounding proposed by [36], beliefs are updated as follows:

(8) D1_1.2: Sex education needs to start in primary school to stop the damage before it's too late
preconditions: $Bel(D_1, p_2); Want(D_1, Bel(\{A1, ..., An\}, p_2))$
expected understanding:$Bel(D_1, MBel(\{D_1, A1, ..., An\}, WBel(D_1, Bel(Ai,$
$Bel(D_1, p_2))))))$ [for each addressee Ai];
$Bel(D_1, MBel(\{D_1, A1, ..., An\}, WBel(D_1, Bel(Ai, Want(D_1, Bel(Ai, p_2))))))$
expected adoption: $Bel(D_1, MBel(\{D_1, A1, ..., An\}, WBel(D_1, Bel(Ai, p_2))))$

D2_2.1: SRE is simply inappropriate within a primary curriculum
understanding: $MBel(\{D_1, D_2\}, Bel(D_1, p_2)); MBel(\{D_1, D_2\}, Want(D_1,$
$Bel(D_2, p_2)))$
cancelled adoption: ~~$Bel(D_1, MBel(\{D_1, D_2\}, WBel(D_1, Bel(D_2, p_2))))$~~
preconditions: $Bel(D_2, \neg p_2); Want(D_2, Bel(\{A1, ..., An\}, \neg p_2));$
expected understanding: $Bel(D_2, MBel(\{D_2, A1, ..., An\}, WBel(D_2, Bel(Ai,$
$Bel(D_2, \neg p_2))))); Bel(D_2, MBel(\{D_2, A1, ..., An\}, WBel(D_2,$
$Bel(Ai, Want(D_2, Bel(Ai, \neg p_2))))))$
expected adoption:$Bel(D_2, MBel(\{D_2, A1, ..., An\}, WBel(D_2, Bel(Ai, \neg p_2))))$

D7_7.1: I think involving sex education in primary school is perfectly sensible
understanding: $MBel(\{D_1, D_7\}, Bel(D_1, p_2)); MBel(\{D_1, D_7\}, Want(D_1,$
$Bel(D_7, p_2))); MBel(\{D_7, D_2\}, Bel(D_2, \neg p_2)); MBel(\{D_7, D_2\},$
$Want(D_3, Bel(D_2, \neg p_2)))$
adoption: $Bel(D_7, MBel(\{D_1, D_7\}, p_2))$

cancelled adoption: ~~$Bel(D_2, MBel(\{D_2, D_7\}, WBel(D_2, Bel(D_7, \neg p_2))$~~
preconditions: $Bel(D_7, p_2)$; $Want(D_7, Bel(\{A1, ..., An\}, p_2))$;
expected understanding: $Bel(D_7, MBel(\{D_7, A1, ..., An\}, WBel(D_7, Bel(Ai,$
$Bel(D_7, p_2)))))$; $Bel(D_7, MBel(\{D_7, A1, ..., An\}, WBel(D_7, Bel(Ai,$
$Want(D_7, Bel(Ai, p_2))))))$
expected adoption:$Bel(D_7, MBel(\{D_7, A1, ..., An\}, WBel(D_7, Bel(Ai, p_2))))$

We implemented a system that keeps track of all created and adopted beliefs on the part of each debater as the debate proceeds. We used the conclusions identified in Section 4 to update the information states of participants and that of the system. This leads to the system's creation and adoption of beliefs concerning these propositions. For example, with regard to the proposition p_1 in (6) the following system's beliefs are created: $Bel(S, MBel(\{S, D_1, D_3, D_4, D_{12}\}, Bel(\{D_1, D_3, D_4, D_{12}\}, (p_1)), Bel(S, MBel(\{S, D_1, D_3, D_4, D_{12}\},$ $Want(\{D_1, D_3, D_4, D_{12}\}, Bel(S, p_1))))$, where S stands for System. In the final state, the system may predict that the belief $Bel(S, MBel(\{S, D_1, D_3, D_4, D_{12}\}, p_1))$ will be adopted.

6 Concluder agent: evaluation

A system operating as described in the previous section can form the basis of an artificial agent that could play different roles in a debate. It could for instance play the role of one of the Debaters or their Seconders by supporting or attacking certain arguments. In this study we consider the system in the role of Concluder, whose task is to understand the arguments of all the debaters and to conclude the debate by stating the opinion of the majority. We call the system playing this role the C-Concluder (Computational Concluder).

In order to assess the quality of the C-Concluder final information state, we need to evaluate against some form of 'ground truth'. For this purpose we use the final state of a human concluder (H-Concluder). The human concluder is called by the Moderator at the end of the session to wrap up the debate.

The H-Concluder provides a general assessment of what was discussed by emphasizing all major arguments brought up by debaters. It is mostly a summary of the arguments that the majority is in favour of, and of points of strong disagreement. The summary exemplified in (9) is the basis of the H-Concluder' final state. The H-Concluder wraps up his summary by announcing further steps, e.g. the motion needs more discussion.

(9) $HC_{15.1}$: Compulsory sex and relationships education is something the UKYP strives for [*Support 1.1, 3.1, 4.3, 12.2/Attack 2.2*]
$HC_{15.2}$: Many believe teaching children about relationships from a young age is vitally important [*Support 1.2, 6.1, 7.1, 9.2, 10.1, 11.1, 14.2/ Attack 2.1, 5.1, 8.1*]
$HC_{15.3}$: Also it is highlighted that SRE is strongly valued by parents [*Support 1.3, 12.5, 14.3/Attack 2.3*]
$HC_{15.4}$: Many schools work successfully to provide effective SRE, even in faith organizations [*Support 1.4*]
$HC_{15.5}$: Our generation have a much disfigured view on sex from things such as peer pressure, and as many mentioned, sexualized media formats. [*Support 2.5, 7.2, 10.3, 14.1*]
$HC_{15.6}$: As it has already been mentioned by Poppie and many others before her that this is not just sex and sex education or the anatomy of it. This is sex and relationships education [*Support 2.4, 4.2, 6.2, 8.2, 9.1, 10.2, 11.2, 12.3, 13.1*]
$HC_{15.7}$: Children should understand the meanings of a relationship, trust and respect [*Support 3.2, 6.3, 7.3, 10.4, 11.3, 12.4*]
$HC_{15.8}$: I believe as a unified organization we can make the government sit up and listen to our campaign [*Support 1.5, 4.1, 7.4, 12.1*]

The evaluation method is depicted in Figure 6. Both C- and H-Concluders try to understand participants' arguments and links between them (strengthening, adoption and rejection effects). In the final state they have the beliefs of all participants resulting from their understanding of each other and adopting each others beliefs.

We compute the H-Concluder beliefs by applying the analysis exemplified in (8) to a summary given by a human concluder in (9). For the C-Concluder we compute the list of predicted beliefs resulting from understanding, grounding and the propositions supported by a 'winning' majority, as well as the

Table 4: Example of C-Concluder expected information state and H-Concluder actual information state. (pred.und = predicted understanding; und = understanding; pred.ad= predicted adoption; ad = adoption; pred.canc = predicted cancelling; canc = cancelling; Bel = believes; MBel = mutually believed; WBel = weakly believes)

source	C-Concluder (CC)	source	H-Concluder (HC)
pred.und	$Bel(CC, MBel(\{CC, D_1, D_3, D_4, D_{12}\}, Bel(\{D_1, D_3, D_4, D_{12}\}, p_1)))$ $Bel(CC, MBel(\{CC, D_1, D_3, D_4, D_{12}\}, Want(\{D_1, D_3, D_4, D_{12}\}, Bel(CC, p_1))))$ $Bel(CC, MBel(\{CC, D_2\}, Bel(D_2, \neg p_1)))$ $Bel(CC, MBel(\{CC, D_2\}, Want(D_2, Bel(CC, \neg p_1))))$ $Bel(CC, MBel(\{CC, D_1, D_6, D_7, D_9, D_{10}, D_{11}, D_{14}\}, Bel(\{D_1, D_6, D_7, D_9, D_{10}, D_{11}, D_{14}\}, p_2)))$ $Bel(CC, MBel(\{CC, D_1, D_6, D_7, D_9, D_{10}, D_{11}, D_{14}\}, Want(\{D_1, D_3, D_4, D_{12}D_1, D_6, D_7, D_9, D_{10}, D_{11}D_{14}\}, Bel(CC, p_2))))$ $Bel(CC, MBel(\{CC, D_2, D_5, D_8\}, Bel(\{D_2, D_5, D_8\}, \neg p_2)))$ $Bel(CC, MBel(\{CC, D_2, D_5, D_8\}, Want(\{D_2, D_5, D_8\}, Bel(CC, \neg p_2))))$ $Bel(CC, MBel(\{CC, D_1, D_{12}, D_{14}\}, Bel(\{D_1, D_{12}, D_{14}\}, p_3)))$ $Bel(CC, MBel(\{CC, D_1, D_{12}, D_{14}\}, Want(\{D_1, D_{12}, D_{14}\}, Bel(CC, p_3))))$ $Bel(CC, MBel(\{CC, D_2\}, Bel(D_2, \neg p_3)))$ $Bel(CC, MBel(\{CC, D_2\}, Want(D_2, Bel(CC, \neg p_3))))$ $Bel(CC, MBel(\{CC, D_1\}, Bel(D_1, p_4)))$ $Bel(CC, MBel(\{CC, D_1\}, Want(D_1, Bel(CC, p_4))))$ $Bel(CC, MBel(\{CC, D_2, D_7, D_{10}, D_{14}\}, Bel(\{D_2, D_7, D_{10}, D_{14}\}, p_5)))$ $Bel(CC, MBel(\{CC, D_2, D_7, D_{10}, D_{14}\}, Want(\{D_2, D_7, D_{10}, D_{14}\}, Bel(CC, p_5))))$ $Bel(CC, MBel(\{CC, D_2, D_4, D_6, D_8, D_9, D_{10}, D_{11}, D_{12}, D_{13}\}$ $Bel(\{D_2, D_4, D_6, D_8, D_9, D_{10}, D_{11}, D_{12}, D_{13}\}, p_6)))$ $Bel(CC, MBel(\{CC, D_2, D_4, D_6, D_8, D_9, D_{10}, D_{11}, D_{12}, D_{13}\}, Want(\{CC, D_2, D_4, D_6, D_8, D_9, D_{10}, D_{11}, D_{12}, D_{13}, \}, Bel(CC, p_6))))$ $Bel(CC, MBel(\{CC, D_3, D_6, D_7, D_{10}, D_{11}, D_{12}\}$ $Bel(\{D_3, D_6, D_7, D_{10}, D_{11}, D_{12}\}, p_7)))$ $Bel(CC, MBel(\{CC, D_3, D_6, D_7, D_{10}, D_{11}, D_{12}\}, Want(\{D_3, D_6, D_7, D_{10}, D_{11}, D_{12}\}, Bel(CC, p_7))))$ $Bel(CC, MBel(\{CC, D_2, D_4, D_7, D_{12}\}$ $Bel(\{D_2, D_4, D_7, D_{12}\}, p_8)))$ $Bel(CC, MBel(\{CC, D_2, D_4, D_7, D_{12}\}, Want(\{D_2, D_4, D_7, D_{12}\}, Bel(CC, p_8))))$ $Bel(CC, MBel(\{CC, D_{10}\}, Bel(D_{10}, p_9$ $Bel(CC, MBel(\{CC, D_{10}\}, Want(D_{10}, Bel(CC, p_9)))))))$	und	$Bel(HC, MBel(\{HC, D_1, D_3, D_4, D_{12}\}, Bel(\{D_1, D_3, D_4, D_{12}\}, p_1)))$ $Bel(HC, MBel(\{HC, D_1, D_3, D_4, D_{12}\} Want(\{D_1, D_3, D_4, D_{12}\}, Bel(HC, p_1))))$ $Bel(HC, MBel(\{HC, D_1, D_6, D_7, D_9, D_{10}, D_{11}, D_{14}\}, Bel(\{D_1, D_6, D_7, D_9, D_{10}, D_{11}, D_{14}\}, p_2)))$ $Bel(HC, MBel(\{HC, D_1, D_6, D_7, D_9, D_{10}, D_{11}, D_{14}\}, Want(\{D_1, D_3, D_4, D_{12}D_1, D_6, D_7, D_9, D_{10}, D_{11}, D_{14}\}, Bel(HC, p_2))))$ $Bel(HC, MBel(\{HC, D_1, D_{12}, D_{14}\}, Bel(\{D_1, D_{12}, D_{14}\}, p_3)))$ $Bel(HC, MBel(\{HC, D_1, D_{12}, D_{14}\}, Want(\{D_1, D_{12}, D_{14}\}, Bel(HC, p_3))))$ $Bel(HC, MBel(\{HC, D_1\}, Bel(D_1, p_4))$ $Bel(HC, MBel(\{HC, D_1\}, Want(D_1, Bel(HC, p_4))))$ $Bel(HC, MBel(\{HC, D_2, D_7, D_{10}, D_{14}\}, Bel(\{D_2, D_7, D_{10}, D_{14}\}, p_5)))$ $Bel(HC, MBel(\{HC, D_2, D_7, D_{10}, D_{14}\}, Want(\{D_2, D_7, D_{10}, D_{14}\}, Bel(HC, p_5))))$ $Bel(HC, MBel(\{HC, D_2, D_4, D_6, D_8, D_9, D_{10}, D_{11}, D_{12}, D_{13}\}$ $Bel(\{D_2, D_4, D_6, D_8, D_9, D_{10}, D_{11}, D_{12}, D_{13}\}, p_6)))$ $Bel(HC, MBel(\{HC, D_2, D_4, D_6, D_8, D_9, D_{10}, D_{11}, D_{12}, D_{13}\}, Want(\{HC, D_2, D_4, D_6, D_8, D_9, D_{10}, D_{11}, D_{12}, D_{13}\}, Bel(HC, p_6))))$ $Bel(HD, MBel(\{HC, D_3, D_6, D_7, D_{10}, D_{11}, D_{12}\}$ $Bel(\{D_3, D_6, D_7, D_{10}, D_{11}, D_{12}\}, p_7)))$ $Bel(HC, MBel(\{HC, D_3, D_6, D_7, D_{10}, D_{11}, D_{12}\}, Want(\{D_3, D_6, D_7, D_9, D_{10}, D_{11}, D_{12}\}, Bel(HC, p_7))))$ $Bel(HC, MBel(\{HC, D_2, D_4, D_7, D_{12}\}$ $Bel(\{D_2, D_4, D_7, D_{12}\}, p_8)))$ $Bel(HC, MBel(\{HC, D_2, D_4, D_7, D_{12}\}, Want(\{D_2, D_4, D_7, D_{12}\}, Bel(HC, p_8))))$
pred.ad	$Bel(CC, MBel(\{CC, D_1, D_3, D_4, D_{12}\}, p_1))$ $Bel(CC, MBel(\{CC, D_1, D_6, D_7, D_9, D_{10}, D_{11}, D_{14}\}, p_2))$ $Bel(CC, MBel(\{CC, D_1, D_{12}, D_{14}\}, p_3))$ $Bel(CC, MBel(\{CC, D_2, D_7, D_{10}, , D_{14}\}, p_5)$ $Bel(CC, MBel(\{CC, D_2, D_4, D_6, D_8, D_9, D_{10}, D_{11}, D_{12}, D_{13}\}, p_6)$ $Bel(CC, MBel(\{CC, D_3, D_6, D_7, D_{10}, D_{11}, D_{12}\}, p_7)$ $Bel(CC, MBel(\{CC, D_2, D_4, D_7, D_{12}\}, p_8)$	ad	$Bel(HC, MBel(\{HC, D_1, D_3, D_4, D_{12}\}, p_1))$ $Bel(HC, MBel(\{HC, D_1, D_6, D_7, D_9, D_{10}, D_{11}, D_{14}\}, p_2))$ $Bel(HC, MBel(\{HC, D_1, D_{12}, D_{14}\}, p_3))$ $Bel(HC, MBel(\{HC, D_1\}, p_4))$ $Bel(HC, MBel(\{HC, D_2, D_7, D_{10}, , D_{14}\}, p_5))$ $Bel(HC, MBel(\{HC, D_2, D_4, D_6, D_8, D_9, D_{10}, D_{11}, D_{12}, D_{13}\}, p_6)$ $Bel(HC, MBel(\{HC, D_3, D_6, D_7, D_{10}, D_{11}, D_{12}\}, p_7)$ $Bel(HC, MBel(\{HC, D_2, D_4, D_7, D_{12}\}, p_8)$
pred. canc	$Bel(CC, MBel(\{CC, D_2\}, WBel(CC, Bel(D_2, \neg p_1))))$ $Bel(CC, MBel(\{CC, D_2, D_5, D_8\}, WBel(CC, Bel(\{D_5, D_8\}, \neg p_2))))$ $Bel(CC, MBel(\{CC, D_2\}, WBel(CC, Bel(D_2, \neg p_3))))$ $Bel(CC, MBel(\{CC, D_1\}, WBel(CC, Bel(D_1, p_4))))$ $Bel(CC, MBel(\{CC, D_{10}\}, WBel(CC, Bel(D_{10}, p_9))))$	canc	$Bel(HC, MBel(\{HC, D_2\}, WBel(HC, Bel(D_2, \neg p_1))))$ $Bel(HC, MBel(\{HC, D_2, D_5, D_8\}, WBel(HC, Bel(\{D_5, D_8\}, \neg p_2))))$ $Bel(HC, MBel(\{HC, D_2\}, WBel(HC, Bel(D_2, \neg p_3))))$ $Bel(HC, MBel(\{HC, D_{10}\}, WBel(HC, Bel(D_{10}, p_9))))$

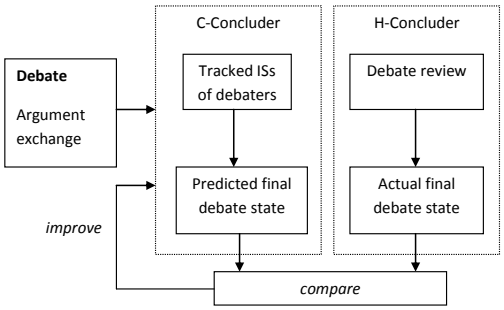

Figure 6: Evaluation model for the C-Concluder role.

negation of all rejected propositions that are not addressed by any of the debaters. The *predicted* final C-Concluder and computed *actual* H-Concluder states are compared.

Table 4 presents the predicted final information state of the C-Concluder and the actual final state of the H-Concluder. The representation of expected understanding effects has been omitted both for C- and H-Concluders, since they are identical. The proposition symbols p_1 to p_9 stand for conclusions.[14]

As we can observe, the predicted C-Concluder information state differs slightly from the actual H-Concluder state, but not significantly. The H-Concluder did not address the arguments concerning the propositions $\neg p_1$, $\neg p_2$, $\neg p_3$ and p_9, hence we do not find evidence in his final state for his understanding of the Inform and (Dis-)Agreement acts with that propositional content. As for the C-Concluder, we had taken the decision that in case of conflicting updates (e.g. $Bel(CC, p)$ and $Bel(CC, \neg p)$ we decide in favor of the majority, comparing the number of supporters. Thus, the adoption of beliefs concerning propositions $\neg p_1$, $\neg p_2$, $\neg p_3$ are cancelled for the C-Concluder state.

Closer inspection shows that of the two arguments that have not been supported or attacked, p_4 is addressed by the H-Concluder while p_9 is not. The H-Concluder considers p_4 as adopted and p_9 as cancelled. Our intuition says that human concluders may have personal considerations such as attitudes towards certain debaters or towards certain arguments, or maybe other factors play a role here. To model this computationally one would need to construct more sophisticated participant models which include their *a priori* beliefs and preferences.

7 Conclusions

In this study we showed how the ISU approach can be applied to modelling and managing argumentative multi-party discourse such as parliamentary debates. We argued that in order to model such complex interactions at least three models are needed. First, a domain model is required where the roles and tasks of the participants are specified. Second, we need a model for the analysis of their contributions. The identification of arguments and analysis of their internal structure (i.e. evidence relations from premises to conclusions) can be based on the identification and classification of discourse units and relations, and can be learned in a data-oriented way as shown by previous research [4, 2, 37]. Third, in order to identify support/attack links between arguments of different debaters, a computational model of belief creation and transfer is needed. An ISU model, where a dialogue is viewed as a sequential structure consisting of communicative acts that participants perform in order to change each other information states, is particularly suitable for this task. We showed how the participants' beliefs are created when a speaker's behaviour is understood and how it leads to the adoption or cancellation of beliefs when participants support or attack each other's arguments.

We evaluated the proposed approach against the debate review produced by a human who acts as a concluder. The system in the role of a concluder, having tracked the information states of the debaters, predicts which propositions will be adopted by the human concluder and which will be cancelled. The comparison shows that such predictions were fairly accurate.

[14] For the sake of simplicity we do not spell out the semantic content of the propositions and leave out evidence links here.

In conclusion, we believe that this paper has addressed a very challenging and exciting research topic, even though it is obviously still a long way to a fully automatic and robust system that is able to understand debate arguments with high accuracy and produce high-quality debate reviews, or even to replace one of the debaters.

Acknowledgements

The underlying research project is partly funded by the EU FP7 Metalogue project, under grant agreement number: 611073. We are also very thankful to anonymous reviewers for their valuable comments.

References

[1] Walton, D. N.: Argumentation schemes for presumptive reasoning. Routledge (1996)

[2] Moens, M., Boiy, E., Mochales Palau, R., Reed, C.: Automatic detection of arguments in legal texts. In: Proceedings of the ICAIL 07, pages 225230, Stanford, California (2007)

[3] Reed, C., Mochales-Palau, R., Rowe, G., Moens, M.: Language resources for studying argument. In: Proceedings of the LREC 08, pages 26132618, Marrakech, Morocco (2008)

[4] Teufel, S.: Argumentative Zoning: Information Extraction from Scientific Text. Ph.D. thesis, University of Edinburgh (1999)

[5] Toulmin, S.: The uses of arguments. Cambridge University Press, Cambridge, England (1958)

[6] Petukhova, V. and Bunt, H.: Incremental Recognition and Prediction of Dialogue Acts. Computing Meaning, Vol. 4, Harry Bunt, Johan Bos and Stephen Pulman, editors. Springer, Dordrecht, pp. 235-256 (2014)

[7] Keizer, S.: Reasoning under uncertainty in natural language dialogue using Bayesian Networks. PhD Thesis,Twente University Press,The Netherlands (2003)

[8] Lendvai, P., Bosch, van den A., Krahmer, E. and Canisius, S.: Memory-based Robust Interpretation of Recognised Speech. In Proceedings of the SPECOM '04, St. Petersburgh, Russia, pp. 415-422 (2004)

[9] Samuel, K., Carberry, S. Vijay-Shanker K.: Dialogue act tagging with transformation-based learning. In Proceedings of the ACL and CoLing, Montreal, pp. 1150 - 1156 (1998)

[10] Stolcke, A., Ries, K., Coccaro, K., Shriberg, E., Bates, R., Jurafsky, D., Taylor, P., Martin, R., Ess-Dykema van, C. and Meteer, M.: Dialogue act modeling for automatic tagging and recognition of conversational speech. In Computational Linguistics, 26(3), pp. 339-373 (2000)

[11] Punyakanok, V., and Roth, D.: The Use of Classifiers in Sequential Inference. NIPS, pp.995-1001 (2001)

[12] Klein, W.: Argumentation and argument. Zeitschrift für Literaturwissenschaft und Linguistik, 10(38/39), pp. 9-56 (1980)

[13] Freeman, J.B.: Argument structure: representation and theory. Argumentation Library, 18, Berlin: Springer (2011)

[14] Peldszus, A. and Stede, M.: From argument diagrams to argumentation mining in texts: a survey. International Journal of Cognitive Informatics and Natural Intelligence (IJCINI), 7(1), pp. 1-31 (2013)

[15] Mann, W. and Thompson, S.: Rhetorical structure theory: toward a functional theory of text organisation. Cambridge, MA, MIT Press (1988)

[16] Sanders, T., Spooren, W. and Noordman, L.: Toward a taxonomy of Coherence relations. Discourse Processes, Vol. 15, pp. 1-35 (1992)

[17] Hobbs, J.: On the coherence and structure of discourse. Research Report 85-37, CSLI, Stanford (1985)

[18] Asher, N. and Lascarides, A.: Logics of Conversation. Cambridge University Press (2003)

[19] Cohen, R.: A computational theory of the function of clue words in argument understanding. In: Proceedings of the Coling-ACL 1984, Standford, pp.251-258 (1984)

[20] Poesio, M. and Traum, D.: Towards an axiomatization of dialogue acts. In: Proceedings of the Twente Workshop on the Formal Semantics and Pragmatics of Dialogues, pp. 207-222 (1998)

[21] Bunt,H.: Information dialogues as communicative action in relation to partner modelling and information processing. In: The Structure of Multimodal Dialogue, Vol. 1, Taylor, M. and Neel, F. and Bouwhuis, D. (eds.), Elsevier, North Holland, pp. 47-73 (1989)

[22] ISO: Language resource management – Semantic annotation framework – Part 2: Dialogue acts. ISO 24617-2. Geneva, ISO Central Secretariat (2012)

[23] Sporleder, C. and Lascarides, A.: Using Automatically Labelled Examples to Classify Rhetorical Relations: An Assessment. In: Natural Language Engineering, vol. 14.03, pp. 369-416 (2008)

[24] Marcu, D.: The rhetorical parsing of natural language texts. In: Proceedings of of Association for Computational Linguistics Annual Conference (ACL), pp. 96-103 (1997)

[25] Hirschberg, J., Litman, D.: Empirical studies on the disambiguation of cue phrases. Computational Linguistics, 25(4), pp. 501-530 (1993)

[26] Sacks, H., Schegloff, E., Jefferson, G.: A simplest systematics for the organization of turn-taking for conversation. Language, 50(4), pp. 696-735 (1974)

[27] Grosz, B. J. and Sidner, C. L.: Attention, Intentions, and the Structure of Discourse. Computational Linguistics, 12, pp. 175-204 (1986)

[28] Ishimoto, Y., Tsuchiya, T., Koiso, H., Den, Y.: Towards Automatic Transformation Between Different Transcription Conventions: Prediction of Intonation Markers from Linguistic and Acoustic Features. In: Proceedings of the LREC 2014, Reykjavik, Iceland, pp. 311-315 (2014)

[29] Prasad, R., Dinesh, N., Lee, A., Miltsakaki, E., Robaldo,L., Joshi,A., Webber, B.: The Penn Discourse Treebank 2.0. In: Proceedings of the LREC 2008, Marrakech, Maroc, pp. 2961-2968 (2008)

[30] Hovy, E., Maier, E.: Parsimonious of profligate: how many and which discourse structure relations? Unpublished manuscript (1995)

[31] Bos, J.: Towards wide-coverage semantic interpretation. In: Proceedings of the 6th International conference on Computational Semantics (IWCS-6), pp. 42-53 (2005)

[32] Bunt, H.: Annotations that effectively contribute to semantic interpretation. Computing Meaning, 4, Harry Bunt, Johan Bos and Stephen Pulman (eds.), Springer, Dordrecht, pp. 49-69 (2014)

[33] Bunt, H.: A context-change semantics for dialogue acts. Computing Meaning, 4, Harry Bunt, Johan Bos and Stephen Pulman (eds.), Springer, Dordrecht, pp. 177-201 (2014)

[34] Searle, J.R.: Speech acts. Cambridge University Press (1969)

[35] Fischer, G.: User modeling in human-computer interaction. User Modeling and User-Adapted Interaction, 11:6568 (2001)

[36] Bunt, H., Keizer, S., Morante, R.: A computational model of grounding in dialogue. In: Proceedings of SIGdial 2007, Antwerp, Belgium, pp. 283290 (2007)

[37] Florou, E., Konstantopoulos, S., Koukourikos, A., Karampiperis, P.: Argument extraction for supporting public policy formulation. In: Proceedings of the LATECH Workshop, Sofia, Bulgaria (2013)