The Metalogue Debate Trainee Corpus: Data Collection and Annotations

Volha Petukhova[1], Andrei Malchanau[1], Youssef Oualil[1], Dietrich Klakow[1],
Saturnino Luz[2], Fasih Haider[3], Nick Campbell[3], Dimitris Koryzis[4],
Dimitris Spiliotopoulos[5], Pierre Albert[2], Nicklas Linz[6], Jan Alexandersson[6]

[1]Saarland University, Spoken Language Systems, Saarbrücken, Germany
[2]The University of Edinburgh, UK; [3]Trinity College Dublin, Ireland
[4]Hellenic Parliament, Greece; [5]University of Houston, TX, USA
[6]German Research Center for Artificial Intelligence (DFKI), Germany

[1]{v.petukhova,andrei.malchanau,youssef.oualil,dietrich.klakow}@lsv.uni-saarland.de;
[2]{s.luz,pierre.albert}@ed.ac.uk;[3]{haiderf,nick}@tcd.ie; [4]dkoryzis@parliament.gr
[5]dspiliotopoulos@uh.edu; [6]{nicklas.linz,jan.alexandersson}@dfki.de

Abstract

This paper describes the Metalogue Debate Trainee Corpus (DTC). DTC has been collected and annotated in order to facilitate the design of instructional and interactive models for Virtual Debate Coach application - an intelligent tutoring system used by young parliamentarians to train their debate skills. The training is concerned with the use of appropriate multimodal rhetorical devices in order to improve (1) the organization of arguments, (2) arguments' content selection, and (3) argument delivery techniques. DTC contains tracking data from motion and speech capturing devices and semantic annotations - dialogue acts - as defined in ISO 24617-2 and discourse relations as defined in ISO 24617-8. The corpus comes with a manual describing the data collection process, annotation activities including an overview of basic concepts and their definitions including annotation schemes and guidelines on how to apply them, tools and other resources. DTC will be released in the ELRA catalogue in second half of 2018.

Keywords: debate argumentation, multimodal data collection, ISO standard annotations

1. Introduction

We currently observe a steadily growing interest of researchers and practitioners in natural argumentation modelling and in developing argumentation technologies. There are systems developed and deployed for legal domains to assist the lawyer in his search for similar past cases, (Teufel, 1999; Brüninghaus and Ashley, 2005); for mining arguments in social media with the goal to predict consumers sentiment (Bai, 2011), to analyse opinions in public discussions (Murakami and Raymond, 2010), to study citizen engagement (Purpura et al., 2008) and to recognize stance in political online debates (Somasundaran and Wiebe, 2010; Walker et al., 2012a).

Argumentation constitutes an important component of human intelligence. Educational studies have shown that constructing arguments and engaging in argumentative discussion enhance conceptual understanding of the subject matter (Wiley and Voss, 1999; Zohar and Nemet, 2002). Argumentation training systems are designed for the legal domains, e.g. to training hypothetical reasoning (Ashley et al., 2007). The TruthMapping[1] web application facilitates collaborative learning through argumentation. DebateGraph[2] used to train how to prevent opinion manipulation marking inconsistent arguments.

These and other developments were supported by corpora collected for various genres, domains and modalities. For example, the *AIFdb* corpora collection (Lawrence and Reed, 2014) of the Centre for Argument Technology, University of Dundee includes data harvested and analysed from ArguBlogging[3], BBC Radio programmes (e.g.

MM2012), Araucaria argument database (Reed, 2006). There is the *Internet Argument Corpus* (IAC) (Walker et al., 2012b) of political debates on internet forums, consisting of about 11,000 discussions and 390,000 posts. Subsets of the data have been annotated for topic, stance, agreement, sarcasm, and nastiness among others. The *Yahoo News Annotated Comments Corpus* (Napoles et al., 2017) is one of the largest annotated corpora of online human argumentative dialogues, with the most detailed set of annotations to identify argumentative, respectful exchanges containing persuasive, informative, and/or sympathetic comments.

Larger projects have been used successfully as resources to study written and spoken argumentative discourse, e.g. *Online Debate Forum*[4], *CE-EMNLP-2015*, also known as IBM corpus, a selection of annotated arguments from Wikipedia articles (Rinott et al., 2015), documents of the *European Court of Human Rights*[5], *UK Youth Parliament* (UKYP)[6] debates (Petukhova et al., 2016), the *American Presidency Project* (APP)[7], and many more.

For the application designed within the Metalogue project[8] - Virtual Debate Coach - an interactive system used to train young parliamentarians to debate efficiently (Petukhova et al., 2017b), the Debate Trainees Corpus (DTC) of 'natural' multimodal arguments was collected. Trainees were trained to make choices from a wide range of rhetorical, lexical, syntactic, pragmatic and prosodic devices to deliver strong persuasive speeches (Petukhova et al., 2017c).

---

[1]https://www.truthmapping.com/
[2]http://debategraph.org/
[3]http://www.argublogging.com/

[4]http://www.debate.org/
[5]http://echr.coe.int
[6]http://www.ukyouthparliament.org.uk/
[7]http://www.presidency.ucsb.edu/index.php
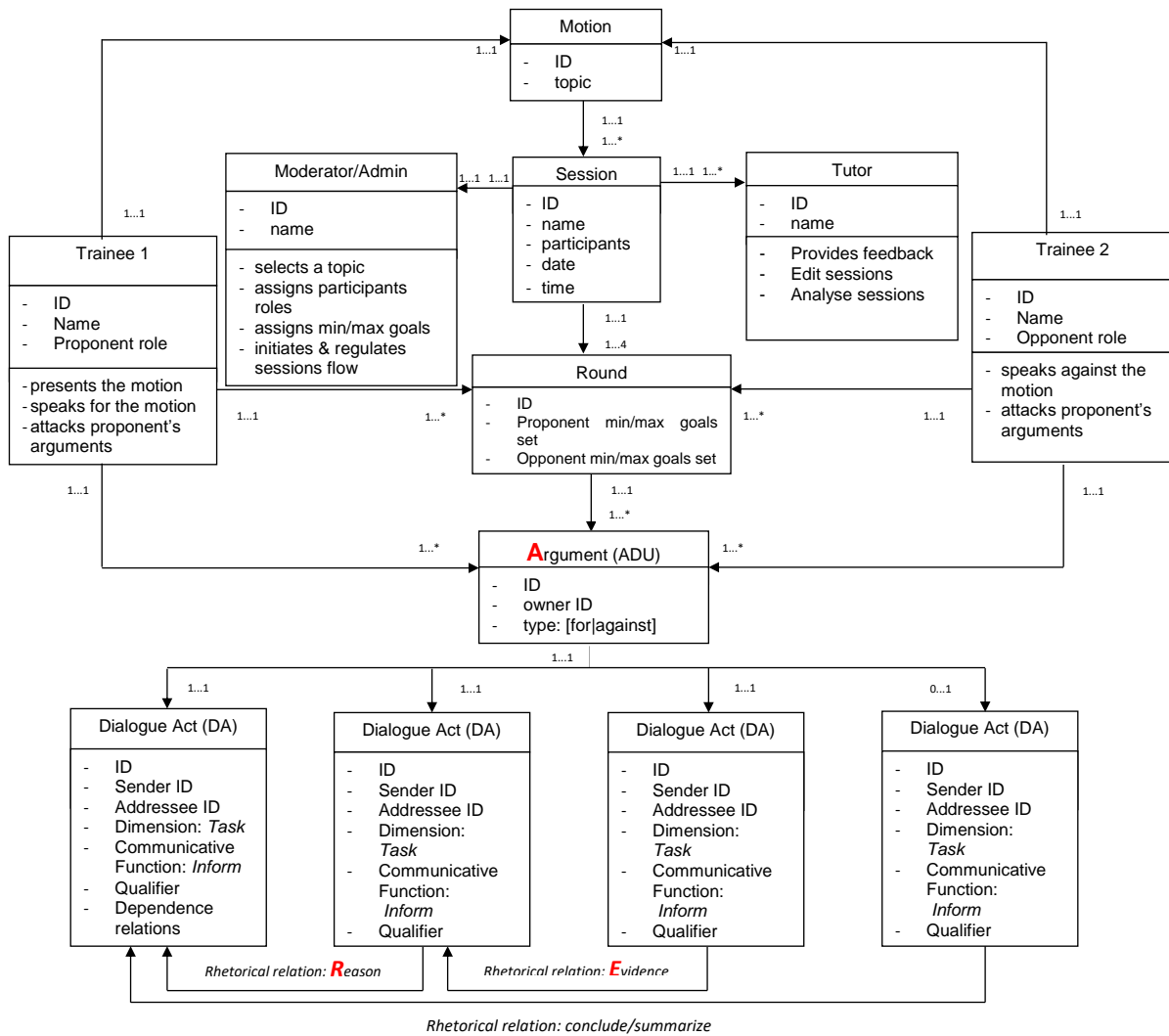[8]http://cordis.europa.eu/project/rcn/110655_en.html

Figure 1: Conceptual class diagram for a Metalogue debate training session.

This paper describes the semantically annotated debate data collection undertaken within the project. The corpus will be delivered with audio and video recordings, Kinect tracking data, automatic and manual transcriptions, ISO compliant semantic annotations. Annotations will be provided in Anvil[9]-specific XML format, and converted to ISO 24617-2 Dialogue Act MarkUp Language (DiAML) (Bunt et al., 2012).

The corpus guidelines, reports, annotation schemes as well as data collection instructional material will be also provided to enable the replication of the carried out experiments.

2. Training Debate Argumentation

Debates, in particular political debates, constitute a large portion of public speeches. Skilled professional debaters give the impression that they truly believe what they say, know how to catch and keep the attention of the audience, and express authority, confidence, respect and friendliness. People generally associate certain speech, personality and

interaction features with what they think is a 'good public speaker', see e.g. (Strangert and Deschamps, 2006).

The training of debate skills typically involves ad-hoc face-to-face classroom debates. The debater's skills proficiency level is often judged on three criteria: (1) argument organization, (2) argument content, and (3) argument delivery.

A debate is a communication process in which participants argue *for* or *against* a certain position proposed for the dispute. In a parliament setting, such initial position is called *motion*. In our training scenario, each debate *session* is motivated by a motion - new law proposal or changes to an existing law. A session consists of one or multiple training *rounds*, e.g. our session comprised four debate rounds, featuring different goals assigned to trainees by a *Moderator* (or Admin). Moderators initiate and further formally regulate the session(-s). One or more *Tutors* attend the session and provide feedback to *Trainees*. Tutoring interventions are expected to inform trainees of mistakes, propose corrections, provide instructions, initiate 'try again' rounds, or highlight trainees' successes. This involves immediate real-time 'in-action' and summative 'about-action' feedback (Schön, 1983) on the three debate aspects mentioned

Issue under debate	Trainees minimum goals to achieve		
	Proponent	Opponent (conservative)	Opponent (liberal)
Smoking ban scope	Not all public places should be affected, allow smoking in bar and restaurants and open air areas like outside buildings, parks and beaches	Forbid smoking inside all public spaces, special smoking areas outside buildings	Allow smoking in special areas in bars and restaurants, open air places also need smoking areas
Tobacco prices	Tobacco price already high, increase no more than 2% a year	Tobacco prices are low, increase by 10% a year	Tobacco prices are still too attractive, increase by 5% a year
Access to tobacco	Tobacco sold in supermarkets, specialized licensed tobacco shops, in bars and restaurants, and vending machines on street with secured buyer's age control	Tobacco should be sold only in special licensed tobacco shops	Tobacco sold in supermarkets but hidden in special containers, prohibited to sell around schools (5km distance) and not available in bar or street vending machines
State control	No police control but municipal and administrative control, no penalties but warnings for the 1st time, repeated disobedience may be punished with penalties	Strong police presence in public places and penalties without warnings	No police control, municipal and administrative control, 1st time disobedience gets warning; second time penalties
Anti-smoking campaign	on TV (state channels 20 min broadcasting time a week); posters in every public place; 'educated' slogans on cigarettes; big newspapers 5 lines a week on the first 2-3 pages	on TV (all channels 30 min broadcasting time a week + one documentary a month); posters in every public place; slogans and scaring images on cigarettes; big newspapers 10 lines a week on the bottom of the front page	on TV (state channels 20 min broadcasting time a week); posters in every public place; 'educated' slogans on cigarettes; big newspapers 10 lines a week on the first 2-3 pages

Table 1: Example of participants' minimal goals in own debate round.

above. The actual debate training session starts by the *Proponent* presenting the motion and an argument in favor of it. An argument is defined as consisting of a statement that can be supported by evidence. A statement (*claim*) is an assertion that deserves attention. There may be a conclusion which presents a result, which can be derived from certain evidence (*premises*). An argument has certain internal structure and correspond to a discourse unit, often called Argumentative Discourse Unit (ADU, see e.g. (Peldszus and Stede, 2013). Claim and premises are dialogue acts related to each other by means of discourse relations, see Section 5.2. The task of the *Opponent* is to attack the proponent's argument by rebutting parts of the argument (premises) or its conclusion, or by undercutting their supporting inference(-s). Both trainees can be in the role of either an proponent or opponent. Moderator(-s) may terminate debating and collect trainees feedback about the debate process and their progress in achieving goals assigned to them. Figure 1 presents the conceptual class diagram for a Metalogue debate training session.

3. Scenario and Data Collection

The specific setting considered for the data collection involves a debate scenario about anti-smoking legislation in Greece. The initial proposal for a smoking ban is supported by the proposing (governmental) party. The goal of the proposer is to get a majority vote while agreeing on as few amendments as possible.

Our core data collection activity involved debate *trainees*, school children aged 14-15 years who have been exposed to little debate training. A session involved a pair of participants: one assigned the role of proposer, the other the role of either liberal or conservative opponent. Each participant was given a set of minimal goals concerning: (1) the total ban on smoking in public spaces; (1) limiting youth access to tobacco products; (3) improving the effectiveness of anti-smoking campaign; (4) state control and reinforcement policy; (5) and raising prices on tobacco products. Participants were not allowed to disclose their goals to the other parties prior to the interaction. Three human tutors evaluated debate performance. Table 1 provides an example of minimal

goals that trainees playing different roles should achieve in one debate round.

The collected data consists of 12 sessions with a duration of 2.5 hours, comprising 400 arguments (Argumentative Discourse Units, ADUs[10]) from 6 different bilingual English/Greek speakers.

4. Multimodal Recording and Synchronisation

Training sessions were recorded in a quiet room under special lighting conditions, ensuring that there were no windows behind the participants and that the participants' faces were not in shadow. Two Kinect $V1$ sensors, each facing one participant as much as possible, were placed at a distance of 1.5-2m to the participants. A Kinect $V2$ sensor was also used to track both participants. Body and face tracking data were stored in an XML format containing elements for frames, faces, joint orientation and bone rotation with respect to the camera's coordinates.

Participants faced each other, and markers were placed on the floor to constrain the participants to a limited area. In addition to Kinect's videos, the recordings included two separate video streams, recorded by conventional video cameras. Figure 2 depicts the technical set up for Metalogue debate sessions.

Speech was captured by two audio Tascam Dr-40 recorders and saved in MS WAV format[11]. Speech files are of two types: (1) full dialogue session recorded per speaker, and (2) cut audio files per speaker and roughly per turn (after speaker diarization). Speaker diarization has been partly carried out manually using the Audacity tool[12] and partly automatically using LIUM tool (Rouvier et al., 2013). The speech signal files contain timestamps - start and end time

---

[10]For more details on segmentation and annotation performed, we refer to (Petukhova et al., 2016).

[11]The recordings were performed in the following setting: sample rate (48KHz), sample size (16-bit), sample format (linear PCM) with stereo channel which was later converted to mono .
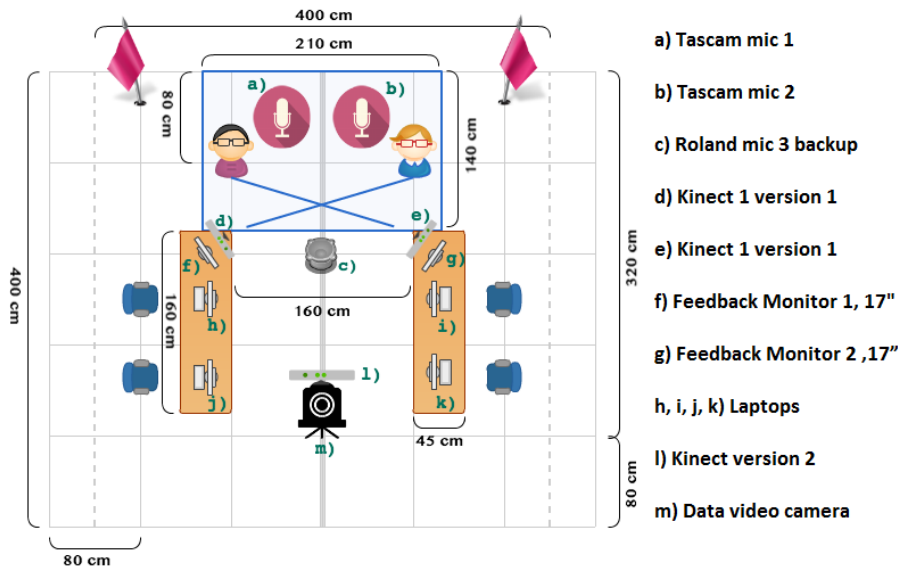
[12]http://www.audacityteam.org/

Figure 2: Recording set up for Metalogue debate sessions, adapted from Haider et al. (2017).
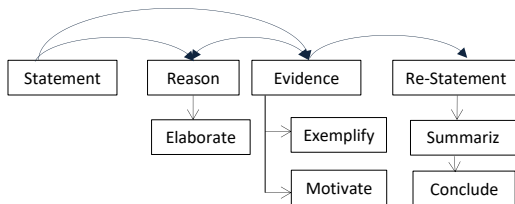

Figure 3: Argument structure observed in Metalogue debate data.

- and additional comments on acoustic and temporal conditions (noise, long silences, etc.) in the file name. For example, `08.22-08.30.n.wav` is the segment which started at 8 minutes and 22 seconds and finished at 8 minutes and 30 seconds during the recording session; and it contains some noise indicated by "n".

The Kinect and video streams were synchronised with audio using the Final Cut Pro X software. The resulting media were converted to a ANVIL compatible format.[13]

Participants' speech has been transcribed semi-automatically by (1) running the Automatic Speech Recognizer (ASR) Kaldi (Povey, 2011) and (2) correcting ASR output manually. Corrected transcriptions are used to re-train/improve language models.[14] All types of transcriptions are stored in plain text and converted to TEI compliant format (ISO, 2006).

5. Annotation Design

Organization of arguments is the planning and preparation involving **A**rgument as a general claim , **R**eason(-s) and **E**vidence. This structure is often called **ARE**[15].

Good debaters are distinguished by concise clear arguments and try to make their arguments understandable for their addressees. For this purpose, debaters often use linguistic cues such as discourse markers and meta-discoursive acts[16] For example, 'I will talk in favour of ... Because ... Since according to the international research shows...'. Thus, *discourse relations* between two or more *dialogue acts* (argument's premises or conclusions) are often marked explicitly by means of discourse markers to support Justification, Motivation, Cause/Result, Background/Evaluation, Evidence and Circumstance links. Figure 3 depicts the most frequently observed of the Metalogue arguments, about 80% of data accounts for this pattern. The main claim, i.e. Statement, is supported by either a Reason or Evidence, and is wrapped up by a Re-Statement in the form of a Summary or Conclusion. For example:

(1) $D1_{21}$[17]: Past anti-smoking campaigns were useless [*Inform*]
$D1_{22}$: I haven't actually seen any of those implemented [*Inform Motivate* $D1_{21}$]
$D1_{23}$: I have personally walked into a store and seen a fourteen years old buying a pack of cigarettes [*Inform Evidence* $D1_{21}$]
$D1_{24}$: Many cases of civil disobedience make this campaign look nice only on paper [*Inform Re-Statement* $D1_{21}$]

5.1. Dialogue Acts

In Metalogue, we mostly consider annotations of semantic and pragmatic multimodal phenomena. For this purposes, *dialogue acts* play an important role. The ISO 24617-2 dialogue act annotation standard is used which allows the analysis of dialogue behaviour as having communicative func-

---

[13]For more details on the data collection and synchronisation process we refer to (Haider et al., 2017).

[14]It should be noticed that the corpus contains a significant proportion of non-native English speakers, varying in fluency from nearly-native to challenging-to-transcribe.

[15]See http://www.slideshare.net/Cherye/advanced-debating-techniques and (Petukhova et al.,

2016)

[16](Crismore et al., 1993) define metadiscourse as "linguistic material in texts, written or spoken, which does not add anything to the propositional content but that is intended to help the listener or reader organize, interpret and evaluate the information given", e.g. Shifting Topic, Marking Asides, etc.

[17]Here and henceforth D*k* stands for Debater *k*; the subscript is the index of the identified dialogue act.

Discourse relation	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	0.73
Conclude**	5.7	0.71
Restatement***	10.1	0.76

Table 2: Distribution of Inform acts connected by a discourse relation in the corpus (* defined in DPTB; ** defined by Hovy and Maier, 1995; *** in both taxonomies).

tions in several dimensions. The ISO 24617-2 taxonomy (ISO, 2012) distinguishes 9 dimensions, addressing information about a certain (*Task*); the processing of utterances by the speaker (*Auto-feedback*) or by the addressee (*Allo-feedback*); the management of difficulties in the speaker's contributions (*Own-Communication Management*) or that of the addressee (*Partner Communication Management*); the speaker's need for time to continue the dialogue (*Time Management*); the allocation of the speaker role (*Turn Management*); the structuring of the dialogue (*Dialogue Structuring*); and the management of social obligations (*Social Obligations Management*). For Metalogue purposes, we used 3 additional dimension-specific Discourse Structuring functions that are not included in ISO 26417-2, however, defined in DIT++[18]: Topic Introduction, Topic Shift and Topic Shift Announcement.

In the data, more than 41.4% of the dialogue acts performed by the debaters are Inform acts, which are often connected by discourse relations forming an argument. Small portions of *Set Questions* (3.4%) and *Agreements* or *Disagreements* (1.7%) are observed. Other dialogue acts are concerned with *Turn Management* (22.7%); *Time Management* (21.1%); *Own Communication Management* (7.3%); *Social Obligation Management* (1.2%); and *Discourse Structuring* acts (10%).

5.2. Discourse Relations

Discourse relations were annotated using the annotation scheme designed for the Penn Discourse TreeBank (DPTB) corpus (Prasad et al., 2008)), extended with discourse segment relations from the taxonomy proposed in (Hovy and Maier, 1995). Table 2 presents the types and frequencies of the relations along with the inter-annotator agreement reached annotating each relation type. For relations like Elaboration, Evidence, Justification, Reason, Conclude and Restatement, which are important for the debate argument identification and processing, a substantial agreement has been achieved. The annotated discourse relations were mapped to those defined in ISO 24617-8 standard, which was published after all DTC sessions were annotated.

5.3. Argumentative Discourse Units

We segmented debates into Argumentative Discourse Units (ADUs), defined as a unit which consists of one or more

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), who suggest to first segment into Elementary Discourse Units (EDUs)[19] as minimal discourse building blocks, then establish relationships between two or more EDUs, and combine those into ADUs.

Identifying ADUs, we observed a very frequent pattern[20]: 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 is performed by the same speaker which is not involved in any discourse relation, see Figure 3 and example in (1), or another speaker claimed the turn. Finally, to capture support and attack links between arguments produced by different speakers, we identified explicit and implicit agreement and disagreement dialogue acts signalling support or attack of arguments through the *functional dependence relations* defined in (ISO, 2012) between the detected argument conclusions. For example:

(2) $D1_{47};D1_{1.2}$: The government should launch effective anti-smoking campaign before it's too late [*Inform*]
$D2_5;D2_{2.1}$: Decision to smoke or not is the personal issue and state shouldn't interfere [*Inform* & *Disagreement* $D1_{47}$] - Attack $D1_{1.2}$
$D7_2;D7_{7.1}$: I think public health is one of the most important tasks that the government should perform [*Inform* & *Agreement* $D1_{47}$ & *Disagreement* $D2_5$] - Support $D1_{1.2}$/Attack $D2_{2.1}$s

Debater 1 states that an anti-smoking campaign is needed and it is the government responsibility. Debater 2 thinks that smoking is the personal responsibility and government should not interfere. Debater 7 supports argument 1.2 and thereby attacks the arguments 2.1. These links are modelled as part of the debaters' information states, see (Petukhova et al., 2016).

5.4. Dialogue Act Markup Language

ISO standard 24617-2 includes the definition of the Dialogue Act Markup Language (DiAML). The representation of a dialogue act annotation makes use of the XML element `<dialogueAct>` with attributes such as speaker, addressee, communicative function, dimension, qualifiers, dependence relations. Additionally, rhetorical (discourse) relations among dialogue acts are represented by means of `<rhetoLink>` elements. All these types are defined in `diaml` namespace in the defined `DiAML_Types.xsd` scheme. In `DiAML_Containers.xsd` elements such as primary data tokens and sounds, and functional segments are specified without a namespace. The last allows to reuse DiAML_Containers in other (not-`diaml`) schema that may also specify domain-dependent semantics. Such semantics is usually a description of all possible elements of the semantic content of dialogue acts. Semantics can be specified to represent predicate-argument structures, named entities, semantic roles or other semantic relations, etc. Similarly to Metalogue Multi-issue Bargaining (MIB) corpus

---

[18] http://dit.uvt.nl/

[19] EDUs in our data mostly coincide with intentionally defined segments such as dialogue acts - functional segments as defined in ISO 24617-2 (ISO, 2012).

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

Type	Content	Format	Comment
Debate minimal goals cards	4 rounds	pdf	defined for Proponent and Opponent
Metadata	participants (id, native language sex, age at collection)	xml, TEI compliant	generated form participants forms
Signals	sound recordings	mono, 96000Hz sample rate 24-bit sample format	1 channel per speaker
	wav files	mono, 16-bit sample format	cut per speaker/per turn
	Kinect tracking	xml, 30 frames per second avi videos	tracked per speaker recorded per speaker
Automatic Speech Recognition Transcriptions	turn (id, start, end, string) turn (id, start, end, string) utterance (id, start, end, string) functional segments (id, start, end, pointers)	plain text plain text xml, TEI compliant xml, TEI compliant	automatic manual automatic automatic
DA annotations	dialogue act (sender, dimension, communicative function, qualifier functionalDependenceRelation feedbackDependenceRelation) rhetoricalLinks	Anvil and DiAML	manual

Table 3: Metalogue Debate Trainee Corpus overview.

(Petukhova et al., 2016) where negotiation semantics is defied[21], debate domain-specific semantics can be plugged into DiAML. For example:

```
<dialogueAct xml:id="da1" sender="#p1"
  addressee="#p2" dimension="task"
  communicativeFunction="inform"
  target="#fs38"
  qualifier="certain">
    <DebateSemantics>
     <Argument type="for"/>
        <Topic>tax\_increase</Topic>
    </DebateSemantics>
</dialogueAct>
```

6. Corpus Overview

The Metalogue DTC corpus comprises signals, tracking data, transcriptions, meta-data, semantic and pragmatic annotations in standard xml-format. Table 3 provides corpora overview specifying type of data planned for release.

Six types of semantic annotations were performed by two trained and one expert annotators. In total, the Metalogue DTC corpus contains about 10.000 annotated entities. The Metalogue corpus will be published in the ELRA catalogue[22] and be available to the community for research purposes in 2018.

7. Acknowledgements

8. Bibliographical References

Ashley, K., Pinkwart, N., Lynch, C., and Aleven, V. (2007). Learning by diagramming supreme court oral arguments. In *Proceedings of the 11th International Conference on Artificial Intelligence and Law*, ICAIL '07, pages 271–275, Stanford, California. ACM.

Bai, X. (2011). Predicting consumer sentiments from online text. *Decision Support Systems*, 50(4):732–742.

Brüninghaus, S. and Ashley, K. D. (2005). Generating legal arguments and predictions from case texts. In *Proceedings of the 10th International Conference on Artificial Intelligence and Law*, ICAIL '05, pages 65–74. ACM.

Bunt, H., Alexandersson, J., Choe, J.-W., Fang, A., Hasida, K., Petukhova, V., Popescu-Belis, A., and Traum, D. (2012). Iso 24617-2: A semantically-based standard for dialogue annotation. In *Proceedings 8th International Conference on Language Resources and Evaluation (LREC 2012)*, pages 430–437, Istanbul, Turkey. ELRA, Paris.

Crismore, A., Markkanen, R., and Steffensen, M. (1993). Metadiscourse in persuasive writing: A study of texts written by american and finnish university students. *Written communication*, 10(1):39–71.

Haider, F., Luz, S., and Campbell, N. (2017). Data collection and synchronisation: Towards a multiperspective multimodal dialogue system with metacognitive abilities. In *Dialogues with Social Robots*, pages 245–256. Springer.

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

ISO. (2006). *TEI-ISO 24610-1:2006 Language resource management: Feature structures, Part 1: Feature structure representation*. ISO, Geneve.

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

Lawrence, J. and Reed, C. (2014). AIFdb Corpora. In *COMMA*, pages 465–466.

Murakami, A. and Raymond, R. (2010). Support or oppose?: classifying positions in online debates from reply activities and opinion expressions. In *Proceedings of the 23rd International Conference on Computational Linguistics: Posters*, pages 869–875. Association for Computational Linguistics.

Napoles, C., Tetreault, J., Pappu, A., Rosato, E., and Provenzale, B. (2017). Finding good conversations online: The yahoo news annotated comments corpus. In

---

[21] See also (Petukhova et al., 2017a)
[22] http://catalog.elra.info/

Proceedings of the 11th Linguistic Annotation Workshop, pages 13–23.

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

Petukhova et al. (2016). Modelling multi-issue bargaining dialogues:data collection, annotation design and corpus. In *Proceedings 9th International Conference on Language Resources and Evaluation (LREC 2016)*, pages 3133–3140. ELRA, Paris.

Petukhova, V., Malchanau, A., and Bunt, H. (2016). Modelling argumentative behaviour in parliamentary debates: data collection, analysis and test case. In M. Baldoni, et al., editors, *Principles and Practice of Multi-Agent Systems. Lecture Notes in Artificial Intelligence*, pages 26–46. Springer, Berlin.

Petukhova, V., Bunt, H., and Malchanau, A. (2017a). Computing negotiation update semantics in multi-issue bargaining dialogues. In *Proceedings of the 21st Workshop on the Semantics and Pragmatics of Dialogue (SemDial 2017 - SaarDial)*, pages 114– 124, Saarbrücken, Germany.

Petukhova, V., Mayer, T., Malchanau, A., and Bunt, H. (2017b). Virtual Debate Coach Design: Assessing multimodal argumentation performance. In *Proceedings of the 2017 ACM on International Conference on Multimodal Interaction*, Glasgow, UK. ACM.

Petukhova, V., Raju, M., and Bunt, H. (2017c). Multimodal markers of persuasive speech : designing a Virtual Debate Coach. In *Proceedings of the 18th Annual Conference of the International Speech Communication Association (INTERSPEECH)*, pages 142–146, Stockholm, Sweden. International Speech Communication Association (ISCA), Baixas, France.

Povey, D. (2011). The Kaldi speech recognition toolkit. In *Proceedings of the 2011 IEEE Workshop on Automatic Speech Recognition and Understanding*, Big Island, HI, US. IEEE Signal Processing Society.

Prasad, R., Dinesh, N., Lee, A., Miltsakaki, E., Robaldo, L., Joshi, A., and Webber, B. (2008). The Penn Discourse Treebank 2.0. In *Proceedings of the 6th International Conference on Language Resources and Evaluation (LREC 2008)*, Marrakech, Maroc.

Purpura, S., Cardie, C., and Simons, J. (2008). Active learning for e-rulemaking: Public comment categorization. In *Proceedings of the 2008 International Conference on Digital Government Research*, pages 234–243. Digital Government Society of North America.

Reed, C. (2006). Preliminary results from an argument corpus. *Linguistics in the twenty-first century*, pages 185–196.

Rinott, R., Dankin, L., Perez, C. A., Khapra, M., Aharoni, E., and Slonim, N. (2015). Show me your evidence - an automatic method for context dependent evidence detection. In *EMNLP*, pages 440–450, Lisbon, Portugal. The Association for Computational Linguistics.

Rouvier, M., Dupuy, G., Gay, P., Khoury, E., Merlin, T., and Meignier, S. (2013). An open-source state-of-the-art toolbox for broadcast news diarization. In *INTERSPEECH-2013*, pages 1477–1481.

Schön, D. A. (1983). The reflective practitioner: How professionals think in action. In T. Smith, editor, *Basic Books*. Temple Smith, London.

Somasundaran, S. and Wiebe, J. (2010). Recognizing stances in ideological on-line debates. In *Proceedings of the NAACL HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text*, pages 116–124. Association for Computational Linguistics.

Strangert, E. and Deschamps, T. (2006). The prosody of public speech - a description of a project. *Lund Unversity Working Papers*, 52:121–124.

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

Walker, M. A., Anand, P., Abbott, R., Tree, J. E. F., Martell, C., and King, J. (2012a). That is your evidence?: Classifying stance in online political debate. *Decision Support Systems*, 53(4):719–729.

Walker, M. A., Tree, J. E. F., Anand, P., Abbott, R., and King, J. (2012b). A corpus for research on deliberation and debate. In *Proceedings 8th International Conference on Language Resources and Evaluation (LREC 2012)*, pages 812–817.

Wiley, J. and Voss, J. F. (1999). Constructing arguments from multiple sources: Tasks that promote understanding and not just memory for text. *Journal of Educational Psychology*, 91:301–311.

Zohar, A. and Nemet, F. (2002). Fostering students' knowledge and argumentation skills through dilemmas in human genetics. *Journal of Research in Science Teaching*, 39(1):35–62.