



University of Dundee

Intertextual correspondence for integrating corpora

Visser, Jacky; Duthie, Rory; Lawrence, John; Reed, Chris

Published in:

LREC 2018, Eleventh International Conference on Language Resources and Evaluation

Publication date:

2018

Document Version

Publisher's PDF, also known as Version of record

[Link to publication in Discovery Research Portal](#)

Citation for published version (APA):

Visser, J., Duthie, R., Lawrence, J., & Reed, C. (2018). Intertextual correspondence for integrating corpora. In H. Isahara, B. Maegaard, S. Piperidis, C. Cieri, T. Declerck, K. Hasida, H. Mazo, K. Choukri, S. Goggi, J. Mariani, A. Moreno, N. Calzolari, J. Odijk, ... T. Tokunaga (Eds.), LREC 2018, Eleventh International Conference on Language Resources and Evaluation (pp. 3511-3517)

General rights

Copyright and moral rights for the publications made accessible in Discovery Research Portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from Discovery Research Portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain.
- You may freely distribute the URL identifying the publication in the public portal.

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Intertextual Correspondence for Integrating Corpora

Jacky Visser, Rory Duthie, John Lawrence, Chris Reed

Centre for Argument Technology
University of Dundee, Dundee DD1 4HN, United Kingdom
{j.visser, r.w.duthie, j.lawrence, c.a.reed}@dundee.ac.uk

Abstract

We present intertextual correspondence (ITC) as an integrative technique for combining annotated text corpora. The topical correspondence between different texts can be exploited to establish new annotation connections between existing corpora. Although the general idea should not be restricted to one particular theoretical framework, we explain how the annotation of intertextual correspondence works for two corpora annotated with argumentative notions on the basis of Inference Anchoring Theory. The annotated corpora we take as examples are topically and temporally related: the first corpus comprises television debates leading up to the 2016 presidential elections in the United States, the second corpus consists of commentary on and discussion of those debates on the social media platform Reddit. The integrative combination enriches the existing corpora in terms of the argumentative density, conceived of as the number of inference, conflict and rephrase relations relative to the word count of the (sub-)corpus. ITC also affects the global properties of the corpus, such as the most divisive issue. Moreover, the ability to extend existing corpora whilst maintaining the level of internal cohesion is beneficial to the use of the integrated corpus as resource for text and argument mining based on machine learning.

Keywords: argument, corpus, debate, dialogue, intertextuality, Reddit, US presidential elections

1. Introduction

In many regards, a larger corpus is a better corpus (all other things being equal). Additionally, for some purposes, a diverse corpus covering various communicative genres or data from different modalities can be of great value. In a study of argumentative conduct in television debates leading up to the 2016 presidential elections in the United States, we are looking for both. We need a large corpus for the purpose of training machine learning algorithms for argument mining (Palau and Moens, 2009; Feng and Hirst, 2011; Stab and Gurevych, 2014) – with a size of 97,999 words (tokens), our US2016 corpus is the largest corpus of dialogical argumentation annotated for both inferential and discursive structure. We also want the corpus to be amenable to automated comparison between the characteristics of argumentative discussions in different communicative genres, as a means of quantifying the uptake certain topics from the television debates find in online social media reactions.

To construct a large corpus, it might be sufficient to collate various (preferably consistently annotated) corpora into one super-corpus. However, if any of the studies for which the text corpus is used relies upon the relationships between different parts of the corpus, then mere collation will not suffice. To properly integrate corpora, we propose that intertextual correspondence can be used as a method to preserve dialogical cohesion by functionally connecting content from two or more corpora on the basis of their topical correspondence.

In the present paper, we introduce the notion of ‘intertextual correspondence’ (ITC) in relation to existing work (Section 2.), we outline the general idea behind ITC and explain how we employ ITC in the creation of the US2016 corpus of election debate material (3.). The integration of both transcripts of television debates between the presidential candidates, and social media reactions to the debates, allows us to exploit the large-scale properties of the corpus

to create Argument Analytics that provide new insight into the dynamics of argumentative reality (4.).

2. Related Work

The notion of ‘intertextuality’, introduced by Kristeva (1977), is one of the cornerstones of the postmodern tradition particularly in literary studies and religion studies (Alfaro, 1996). Intertextuality is used to explain that any text can only be properly understood in its relation to the larger body of other texts. To put it differently: context – broadly conceived – should be taken into account when interpreting (the meaning of) a text.

In our repurposing of the term, we divorce it from its associated philosophical connotations, to refer to the topical correspondence that exists between the contents of independent text corpora. While our focus is exclusively on corpora of argumentative discourse and on the intertextual inference dependencies that exist between them, we do not intend to constrain ITC to such corpora. The corpus which we consider in Section 4. comprises argumentation in the political domain. In her Critical Discourse Analysis studies of political communication, Wodak (2009) makes a distinction between ‘intertextuality’ and ‘interdiscursivity’. Because of the shared focus on topical interconnection, the latter notion is closer to what we mean here with ‘intertextual correspondence’.

Connecting various topically related parts of analysed argumentative discourse is directly related to the (World Wide) Argument Web: a vision of “a large-scale Web of interconnected arguments posted by individuals to express their opinions in a structured manner” (Rahwan et al., 2007). The foundation underpinning the Argument Web is the Argument Interchange Format (AIF) (Chesñevar et al., 2006). The AIF is an ontology designed to capture the variety of theory-dependent notions, properties and constructs used in the study of argumentation. As such, the AIF is intended to serve as an interlingua in which the different theoretical

conceptions of argumentatively relevant terms can be translated. Reed et al. (2010) extended the AIF to account for the dialogical dimension of argumentation in the AIF ontology.

3. Intertextual Correspondence

3.1. Integrating Corpora

We propose a method for integrating annotated text corpora on the basis of the correspondences that exist between the topics at issue in each corpus. Of course, this is not to imply that any two corpora can always be integratively merged. The merging is not the issue, two corpora can just be collocated into one. Integrating them, however is where the challenge lies. This requires identification of correspondences between elements of the two corpora. Such correspondences can be weaker, e.g., when the same topic is being discussed without further direct communicative relation, or stronger, e.g., when a direct reference is made by means of reported speech to some statement made elsewhere.

The underlying principles of ITC can be used to integratively combine corpora annotated with a variety of theoretical approaches. A large, but partially disjoint corpus such as Abbott et al. (2016)'s Internet Argument Corpus (version 2) which contains content from different online sources could be amenable to the techniques we discuss to strengthen the internal cohesion. In the remainder of this paper, however, we build on the theoretical foundation upon which our corpus annotation is based, with application to the US2016 corpus.

3.2. Corpus Annotation Guidelines

Four annotators were extensively trained in the use of Inference Anchoring Theory (IAT) (Budzynska and Reed, 2011) to analyse the television debates and Reddit discussions constituting the US2016 corpus that we take as a case in point in the current paper. Building on insights from discourse analysis and argumentation studies (van Eemeren et al., 2014), IAT explains argumentative conduct in terms of the anchoring of argumentative reasoning in dialogical interaction. Drawing on Speech Act Theory (Austin, 1962; Searle, 1969), the anchoring is theoretically conceptualised by means of the 'illocutionary connection' between locutions in dialogue and their propositional content. Importantly, IAT complies with the standards of the AIF ontology discussed in Section 2.).

The annotation guidelines, summarised below, are based on IAT. The full version of the guidelines (available online at arg.tech/US2016-guidelines) deals with, among others: anaphoric references, epistemic modalities, repetitions, punctuation, discourse indicators, interposed text, reported speech, and how to deal with context-specific peculiarities.

Segmentation divides the (transcribed) text into locutions. A locution consists of a speaker designation and an 'argumentative discourse unit' (a text span with discrete argumentative function) (Peldszus and Stede, 2013).

Transitions capture the functional relationships between locutions, reflecting the dialogue protocol – a high level specification of the set of transition types that are available in a particular communicative activity.

Illocutionary connections embody the intended communicative functions of locutions or transitions, such as: *Agreeing*, *Arguing*, *Asserting*, (three sub-types of) *Challenging*, *Disagreeing*, (three sub-types of) *Questioning*, *Restating*, and *Default Illocuting* (when none of the other types suffice). Some types of illocutionary connection lead to the reconstruction of a propositional content.

Inferences are directed relations between propositions, reflecting that a proposition is meant to supply a reason for accepting another proposition. A specific argument scheme (e.g., *Argument from Example* or *Argument from Expert Opinion*) can be specified; failing that, it is labelled as *Default Inference*.

Conflicts are directed relations between propositions, reflecting that a proposition is meant to be incompatible with another proposition or relation. Such incompatibility may depend on, e.g., logical contradiction or pragmatic contrariness, or the annotated relation may default to *Default Conflict*.

Rephrases are directed relations between propositions, reflecting that a proposition is meant to be a reformulation of another proposition. Such reformulation may involve, e.g., *Specialisation*, *Generalisation* or *Instantiation*, or the relation defaults to *Default Rephrase*.

3.3. Intertextual Correspondence Guidelines

The ITC between the comments on Reddit and the television debates to which they react were annotated by two annotators trained in the application of the instruction described in this section. Because the original US2016 corpus is annotated on the basis of IAT, the annotation guidelines for ITC are generally the same as those described in Section 3.2.. However, the contextual characteristics of the two genres of television debates and social media discussion lead to five annotation patterns that can be expected to typically occur in ITC in particular (although variations are possible). In the explanation of these patterns, we will use the suffixes '-tv', '-reddit', and '-itc' when we refer to the elements of the annotations that are part of, respectively, the television debate sub-corpus (US2016tv), the Reddit discussion sub-corpus (US2016reddit), and the intertextual correspondence sub-corpus that serves to bridge between them (US2016itc).

As part of the annotation of ITC, no new propositional content is annotated. Rather, new relations are introduced that connect content (both propositions and locutions) in the US2016reddit corpus to their counterparts in the television corpus. Due to the temporal ordering of the Reddit commentary following the proceedings in the television debate, by necessity, the transitions and most of the other relations will be directed from US2016reddit to US2016tv (one notable exception will be introduced below).

The first common pattern reflects rephrases of what is said in the television debates. Although propositions can be quoted literally, in most cases there is some degree of reformulation. In these cases, an ITC rephrase relation is introduced. An example of such a rephrase relation is visualised in Figure 1. In this example, Reddit user Mr_Jensen reformulates Bernie Sanders' claim about college affordability in order to explain what the then candidate for the

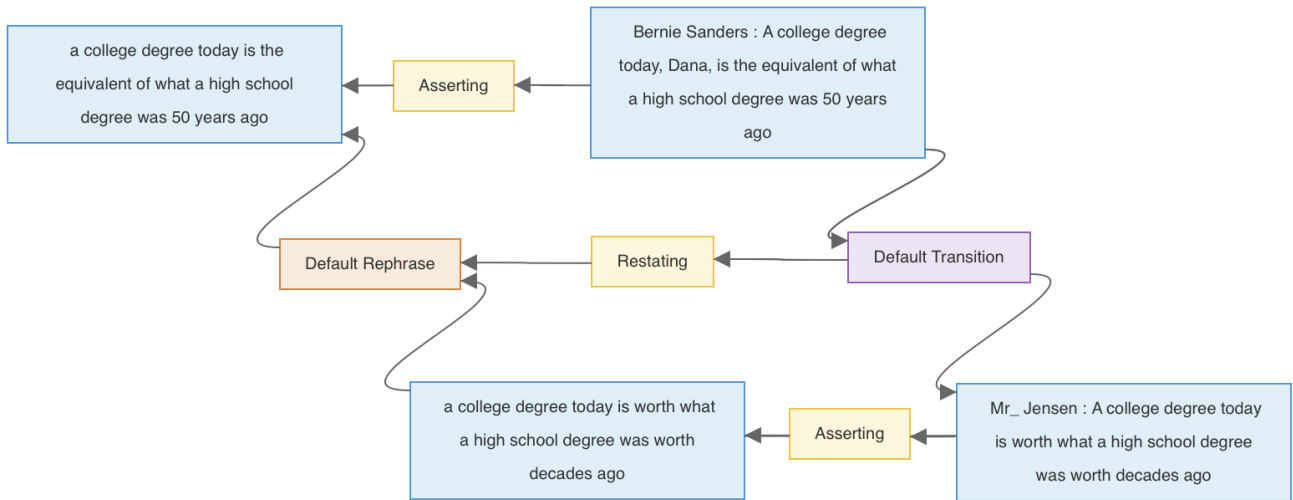


Figure 1: Diagrammatic visualisation of common ITC rephrase pattern.

Democrat’s nomination meant. The top row of Figure 1 shows the connections between proposition-tv, illocution-tv and locution-tv; the bottom row shows proposition-reddit, illocution-reddit and locution-reddit; the middle row contains the connecting relations introduced as part of the annotation of ITC. The rephrase-itc relation from proposition-reddit to proposition-tv reflects the intertextual rephrase. The illocution-itc of *Restating* anchors the *Default Rephrase* relation between the two propositions in the *Default Transition* between the two locutions.

The next three common patterns are structurally equivalent to the first, although they introduce different relations between the proposition-reddit and the proposition-tv, and hence employ different illocutionary connections. A Reddit user can engage in an argumentative interaction with the content of the television debate by either supplying (additional) reasons for a claim made during the television debate, or by drawing a conclusion on the basis of what was said on television. In both cases the transition-itc connects locution-reddit to locution-tv, and in both cases the illocution-itc of *Arguing* anchors the inference-itc. In the first case (providing additional reasons), the inference-itc goes from proposition-reddit to proposition-tv. In the second case (drawing a conclusion), this direction is reversed. Similarly to arguing, a Reddit user can express disagreement with what is said on television. In this case, the illocution-itc of *Disagreeing* anchors the conflict-itc in the transition-itc.

Instead of disagreeing, a Reddit user can agree with what is said during the television debate. The resulting structure of the ITC is different from that in Figure 1, because the associated illocution-itc of *Agreeing* does not anchor any propositional relation (of rephrase, inference or conflict). Rather, the annotated ITC only introduces two relations: the by now expected transition-itc from locution-reddit to locution-tv, and an illocution-itc from locution-reddit to proposition-tv.

3.4. Validation

To validate the ITC annotation guidelines, the two annotators both annotated an 11% subset of the corpus. This subset was sampled by taking every second US2016reddit excerpt until the number of propositions was greater than 10% of the overall corpus. The same process was used for the validation of the US2016tv and US2016reddit corpora yielding a Cohen’s κ (Cohen, 1960) of 0.73 for the IAT annotation. In the case of ITC annotation, validating using percentage agreement becomes difficult. The high proportion of agreed negatives, due to a low amount of overall connections made, and a high number of possible connections of $178,940$ (overall locutions and propositions in $US2016tv \times$ possible illocutionary relations \times possible propositional relations) make validation with this metric impractical as a result of the weight of negative examples (see Table 1).

Annotation	% Agreement	Cohen’s κ
Relations	0.99	0.50
Excerpts	0.99	0.83
Combined	0.98	0.62

Table 1: Percentage agreement and Cohen’s κ for annotation of ITC-relations, ITC-excerpts and both combined.

Cohen’s κ was used for the inter-annotator agreement calculation to take into account chance agreement and lower the weight of the agreed negatives. Although this can effectively give an agreement score for relations, it undermines the complexity of the ITC task due to the need for exact matches between both annotators. In ITC this is particularly difficult due to the high number of possible relations that can be made by an annotator. A more effective measure is employed which first considers that two excerpts, one from television and one from Reddit, are agreed to have an ITC relation connecting them by both annotators (see excerpts in Table 1). By using κ we account for chance agreement between the annotators and therefore the evalu-

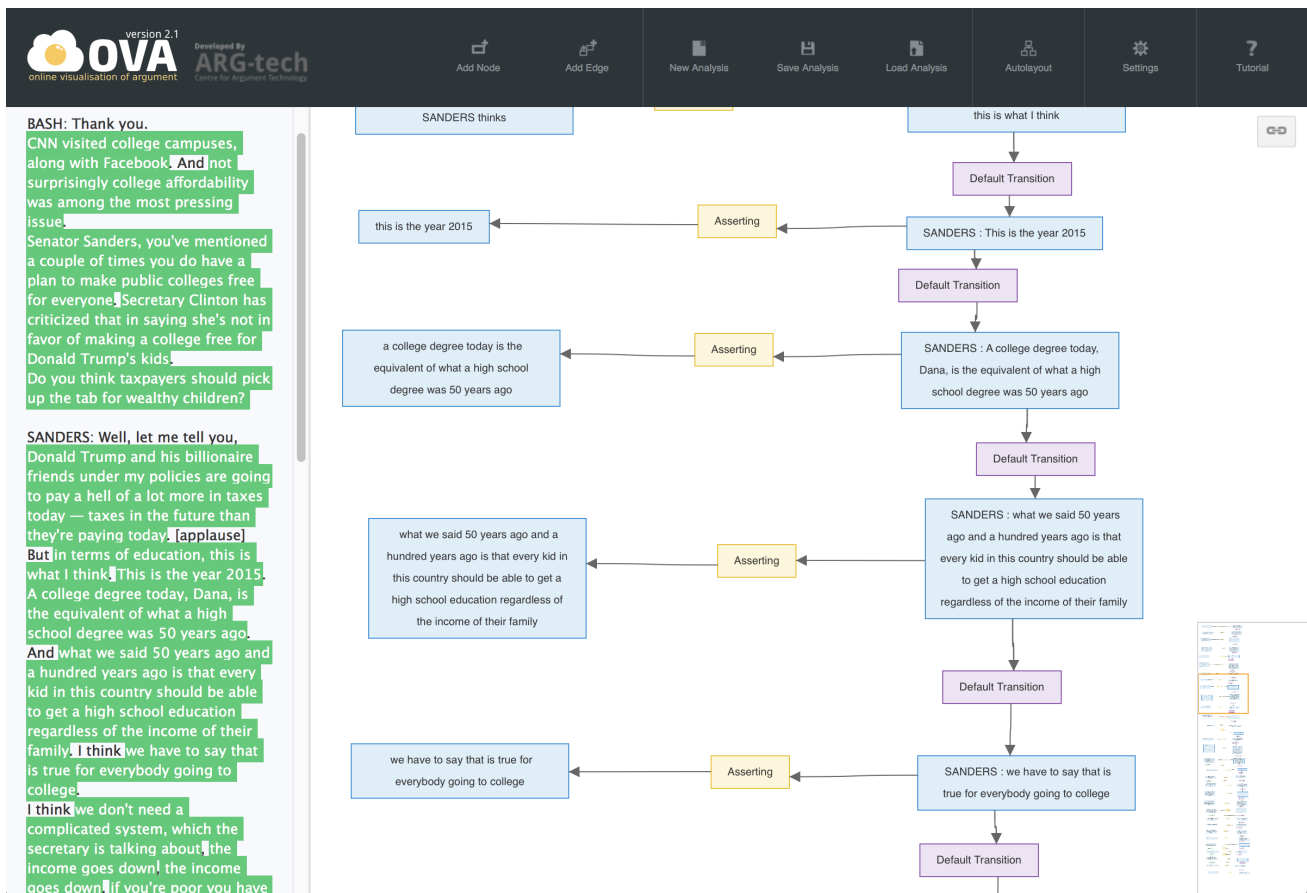


Figure 2: IAT analysis in OVA of an excerpt of US2016tv.

ation is not skewed by false negatives like in the percentage agreement calculation. Following this, κ is used to evaluate the relations made between two excerpts with a more harsh agreement policy where all locutions, propositions, and ITC relations must be identical (see relations in Table 1). Finally, to give an overall κ value we combine the results from the excerpt linking task and the ITC labelling task using the harmonic mean which gives a κ score of 0.62 ((see combined in Table 1)) seen as “good” agreement by (Landis and Koch, 1977).

3.5. Software tools

In the annotation of the US2016 corpus, annotators use an infrastructure of software tools for the manipulation and archiving of argumentation. For the IAT annotation of the television debates and the Reddit discussions, annotators use the Online Visualisation of Argument (OVA) tool (see Figure 2 and available online at ova.arg.tech) (Janier et al., 2014). OVA allows raw natural language text to be highlighted (left of Figure 2) to create locutions, propositions and illocutionary connections in an IAT diagram when the canvas on the right is clicked. Relations between locutions or propositions can be created by toggling the add edge button on the OVA menu and can be edited by double clicking. The full IAT diagram is displayed in the canvas and can be navigated through a small overview box (see bottom right of Figure 2).

The annotation of ITC is facilitated by the ITClinker tool

available at arg.tech/ITClinker. Figure 3 shows a screenshot of the online tool, which is still in the academic prototype stage. Two IAT diagrams are loaded side by side, and locutions and propositions can be selected to create ITC relations. After selection, the user is prompted to decide on the relation type which is displayed below the two IAT diagrams. There is also an option to delete ITC relations. The finished annotations are stored in AIFdb (Lawrence et al., 2012), an online repository of analysed arguments (available at aifdb.org). AIFdb Corpora (Lawrence and Reed, 2014) is used to manage the annotated corpora, and now includes more than 130 corpora of varying size and purpose (e.g. to support education, for particular academic studies, or for our collaboration with the BBC) (corpora.aifdb.org). The global argumentative and discursive properties of a corpus can be automatically extracted by using Argument Analytics (analytics.arg.tech) (Lawrence et al., 2016).

4. Intertextual Correspondence between 2016 US Election TV Debates and Reactions on Reddit

4.1. The US2016 Election Debate Corpus

The US2016 corpus comprises annotated texts from two genres: transcripts of television debates leading up to the 2016 US presidential elections, and online reactions to those debates on the Reddit social media plat-

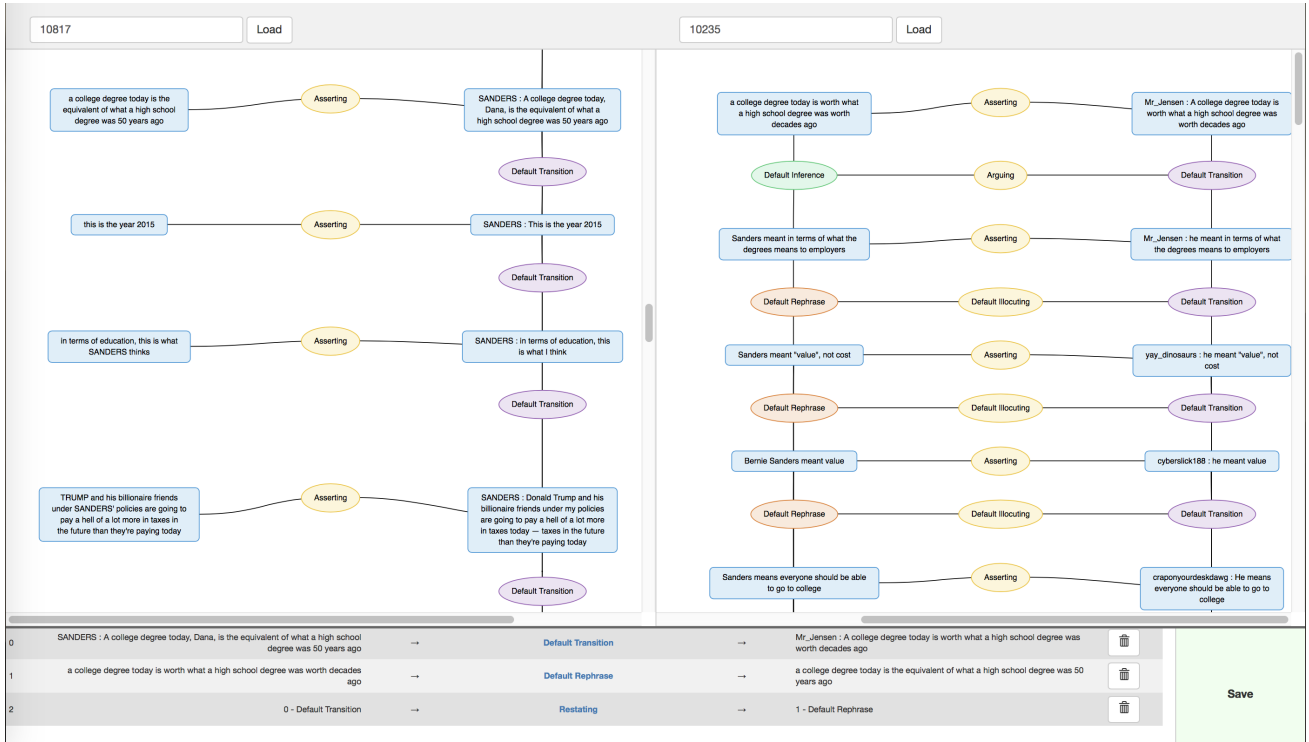


Figure 3: ITC annotation with ITClanker between graph-based IAT annotations from US2016tv and US2016reddit, showing the ITC relations connecting them underneath.

form (reddit.com). The full US2016 corpus is freely available online at corpora.aifdb.org/US2016, and contains a television debate sub-corpus (US2016tv), and a Reddit sub-corpus (US2016reddit) – both independently available online at respectively corpora.aifdb.org/US2016tv and corpora.aifdb.org/US2016reddit.

The US2016tv sub-corpus comprises annotated transcripts of three television debates leading up to the 2016 US presidential elections: the first Republican primaries debate (Peters and Woolley, 2015b), the first Democrat primaries debate (Peters and Woolley, 2015a), and the first general election debate between Hilary Clinton and Donald Trump (Peters and Woolley, 2016). The US2016reddit sub-corpus comprises manually retrieved excerpts from the Reddit mega-threads on the topic of the respective television debates (a new mega-thread being opened during the debates every 30 minutes), selecting sub-threads corresponding to time windows of increased dialogical interaction in the television debate while excluding topically irrelevant sub-threads (e.g. those about practical issues of a technical nature).

The IAT annotation of the US2016tv and sub-corpora (see Section 3.2.) means that the corpus is annotated with dialogical structures consisting of locutions and illocutionary connections, as well as with argumentative relations between the propositional contents of the locutions. Table 2 shows the total counts of annotated illocutionary connections and propositional relations (of inference, conflict and rephrase) in the various (sub-)corpora.

Corpus	Inference	Conflict	Rephrase	Illocution
US2016tv	1551	194	333	6617
US2016reddit	1203	629	287	6348
US2016itc	76	119	144	366
US2016	2830	942	764	13331

Table 2: Propositional relation and illocutionary connection counts for the US2016 corpus.

Corpus	Most divisive proposition
US2016D1tv	Bernie Sanders is tough enough on guns.
US2016D1reddit	Bernie Sanders is really doing that “bad” as the comments suggest.
US2016D1	Bernie Sanders wants Colleges to get their costs down.

Table 3: Most divisive issues for US2016D1tv, US2016D1reddit, and US2016D1 including ITC.

4.2. Intertextual Correspondence in US2016

The election debates and the online reactions are in many ways entirely different argumentative activities: one is broadcast on live television, the other is a social media platform with user-generated content, there is no overlap in participants, and both have their own genre conventions. The two constitutive parts of our corpus are however re-

lated both topically and temporally: the *raison d'être* of the Reddit threads is to discuss what happens in the television debate, and this online commentary is happening while the election debate is being broadcast (in a typical second-screen fashion). This makes the US2016 corpus very suitable for integrative annotation with ITC.

Before ITC annotation, the US2016 corpus contains a total of 4,197 argumentative relations between propositions (2,754 inferences, 823 conflicts, 620 rephrases) and 12,965 illocutionary connections. Table 2 shows the increase of interconnectivity between the television and Reddit sub-corpora of US2016.

As well as increasing the overall counts of propositional relations within the US2016 corpus, ITC also provides a more complete overview of a debate and unifies the issues discussed. This effect on the corpus' global properties becomes evident when calculating, for example, the 'divisiveness' measure. This graph-based Argument Analytic allows the automatic calculation of the most divisive issues within a corpus (Konat et al., 2016). In Table 3, the most divisive issue is shown for the first Democratic primaries debate (US2016D1tv), for the corresponding reactions on Reddit (US2016D1reddit), and for the compilation of all material related to the first Democratic primaries debate (US2016D1) – television debate, Reddit discussion and ITC. While in all three cases the most divisive issue is related to Bernie Sanders, it is clear that the addition of cross-corpus relations of inference, conflict and rephrase by means of ITC results in a shift; in this case moving from opinions about Bernie Sanders himself to a more general divisive issue.

5. Conclusion

Intertextual correspondence (ITC) provides a powerful method of increasing the value of existing annotated corpora. We have demonstrated how ITC is used in our US2016 corpus of US presidential election debates and associated reactions on Reddit. By integrating previously independent corpora, ITC enables the connecting of constellations of (argumentative) content in different corpora to get a better perspective of the interplay between reasons for and against claims in general, regardless of the genre, context or time period in which the claim was made. ITC can be used to connect discourses, establishing detailed, specific and computational meaning to the notions of intertextuality and context. Our initial work has shown how three televised debates and the live reaction to them on Reddit can be connected together not only to build a much larger corpus of dialogue on the topic, but more importantly, to allow investigation of phenomena such as popularity and divisiveness that hinge upon this connection being well established. In the longer term, ITC will allow connections between more distant corpora that are thematically or contextually related, enabling more thorough and wide-ranging computational modelling of complex debates in which discussion takes place across a diverse set of fora: climate change discussions, for example, that take place partly in the scientific literature, partly in popular science media, partly in the political sphere, partly on social media, and so on. ITC offers a key step towards being able to model these debates in their

entirety allowing audiences and contributors (whether, in this case, scientists, politicians or the public) access to the entire debate in a structured and navigable format. In this way, ITC forms a critical part in the current effort to realise the vision of the Argument Web (Rahwan et al., 2007).

6. Acknowledgements

This research was supported by the Engineering and Physical Sciences Research Council in the UK under grants EP/M506497/1 and EP/N014871/1.

7. Bibliographical References

- Abbott, R., Ecker, B., Anand, P., and Walker, M. (2016). Internet Argument Corpus 2.0: An SQL schema for dialogic social media and the corpora to go with it. In Nicoletta Calzolari, et al., editors, *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*, pages 4445–4452, Paris, France. European Language Resources Association (ELRA).
- Alfaro, M. J. M. (1996). Intertextuality: Origins and development of the concept. *Atlantis*, 18(1/2):268–285.
- Austin, J. L. (1962). *How to Do Things with Words*. Clarendon Press.
- Budzynska, K. and Reed, C. (2011). Whence inference. Technical report, University of Dundee.
- Chesñevar, C., Modgil, S., Rahwan, I., Reed, C., Simari, G., South, M., Vreeswijk, G., Willmott, S., et al. (2006). Towards an argument interchange format. *The Knowledge Engineering Review*, 21(04):293–316.
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20(1):37–46.
- Feng, V. W. and Hirst, G. (2011). Classifying arguments by scheme. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1*, pages 987–996. Association for Computational Linguistics.
- Janier, M., Lawrence, J., and Reed, C. (2014). OVA+: An argument analysis interface. In S. Parsons, et al., editors, *Proceedings of the Fifth International Conference on Computational Models of Argument (COMMA 2014)*, pages 463–464, Pitlochry. IOS Press.
- Konat, B., Lawrence, J., Park, J., Budzynska, K., and Reed, C. (2016). A corpus of argument networks: Using graph properties to analyse divisive issues. In *LREC 2016*, pages 3899–3906. European Language Resources Association.
- Kristeva, J. (1977). Word, dialogue and novel. In Leon S. Roudiez, editor, *Desire in language: A semiotic Approach to Literature and Art*, pages 64–91. Columbia University Press.
- Landis, J. R. and Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1):159–174.
- Lawrence, J. and Reed, C. (2014). AIFdb corpora. In S. Parsons, et al., editors, *Proceedings of the Fifth International Conference on Computational Models of Argument (COMMA 2014)*, pages 465–466, Pitlochry. IOS Press.

- Lawrence, J., Bex, F., Reed, C., and Snaith, M. (2012). AIFdb: Infrastructure for the Argument Web. In *Proceedings of the Fourth International Conference on Computational Models of Argument (COMMA 2012)*, pages 515–516.
- Lawrence, J., Duthie, R., Budzynska, K., and Reed, C. (2016). Argument Analytics. In *the Sixth International Conference on Computational Models of Argument (COMMA 2016)*, pages 371–378.
- Palau, R. M. and Moens, M.-F. (2009). Argumentation mining: the detection, classification and structure of arguments in text. In *Proceedings of the 12th international conference on artificial intelligence and law*, pages 98–107. ACM.
- 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.
- Peters, G. and Woolley, J. T. (2015a). Democratic candidates debate in Las Vegas, Nevada, October 13, 2015. Accessed 11 Aug. 2017.
- Peters, G. and Woolley, J. T. (2015b). Republican candidates debate in Cleveland, Ohio, August 6, 2015. Accessed 11 Aug. 2017.
- Peters, G. and Woolley, J. T. (2016). Presidential debate at Hofstra University in Hempstead, New York, September 26, 2016. Accessed 11 Aug. 2017.
- Rahwan, I., Zablith, F., and Reed, C. (2007). Laying the foundations for a world wide argument web. *Artificial Intelligence*, 171:897–921.
- Reed, C., Wells, S., Budzynska, K., and Devereux, J. (2010). Building arguments with argumentation: the role of illocutionary force in computational models of argument. In P. Baroni, et al., editors, *Proceedings of the 3rd International Conference on Computational Models of Argument (COMMA 2010)*, pages 415–426. IOS Press.
- Searle, J. R. (1969). *Speech acts: An essay in the philosophy of language*. Cambridge University Press.
- Stab, C. and Gurevych, I. (2014). Identifying argumentative discourse structures in persuasive essays. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 46–56, Doha, Qatar. Association for Computational Linguistics.
- van Eemeren, F. H., Garssen, B., Krabbe, E. C. W., Snoeck Henkemans, A. F., Verheij, B., and Wagemans, J. H. M. (2014). *Handbook of argumentation theory*. Springer.
- Wodak, R. (2009). *The discourse of politics in action. Politics as usual*. Palgrave Macmillan.