# An annotated corpus of argument schemes in US election debates

Visser, Jacky; Lawrence, John; Wagemans, Jean; Reed, Chris

*Published in:*
Proceedings of the 9th Conference of the International Society for the Study of Argumentation (ISSA), 3-6 July 2018

*Publication date:*
2019

*Document Version*
Publisher's PDF, also known as Version of record

# An annotated corpus of argument schemes in US election debates

JACKY VISSER[1], JOHN LAWRENCE[1], JEAN H. M. WAGEMANS[2], CHRIS REED[1]

[1] *Centre for Argument Technology*
*University of Dundee*
*Dundee DD1 4HN*
*United Kingdom*
*{jacky/john/chris}@arg.tech*

[2] *Speech Communication, Argumentation Theory, and Rhetoric*
*University of Amsterdam*
*1012 VB Amsterdam*
*The Netherlands*
*j.h.m.wagemans@uva.nl*

ABSTRACT: We present a corpus comprising the first general election debate between Clinton and Trump (17,190 words) annotated with types of argument on the basis of the Periodic Table of Arguments. This extends the annotation of an existing corpus (97,999 words) of transcripts of television debates and associated reactions on the Reddit social media platform, annotated on the basis of Inference Anchoring Theory with relations of inference, conflict and rephrase, and their illocutionary discourse anchoring.

KEYWORDS: annotation, argument schemes, corpus, Inference Anchoring Theory, Periodic Table of Arguments, political debate, television debate, US presidential elections

## 1. INTRODUCTION

In this paper we present the US2016G1tvWAGEMANS corpus, a freely available resource of empirical data on argument schemes. The corpus consists of transcripts of television debates leading up to the 2016 presidential elections in the United States, combined with reactions to these debates on the Reddit social media platform. The annotation of the corpus consists of two layers. The whole corpus (97,999 words) was annotated (Visser et al. 2018a) on the basis of Inference Anchoring Theory (Reed & Budzynska, 2011), with argumentative relations of inference, conflict and rephrase, and their dialogical anchoring by means of illocutionary connections (see Section 3). The annotation of a sub-corpus containing the first general election television debate (17,190 words) is extended by classifying the argument schemes on the basis of a factorial approach to argument classification called the *Periodic Table of Arguments* (Wagemans, 2016) (see Section 2).

The Periodic Table of Arguments classification is based on three discriminating properties: first-/second-order arguments; predicate/subject arguments; propositions of fact/value/policy. The individual propositions and inference relations in all arguments are annotated with the three properties (see Section 4). The results on the three distinctions are then combined for each argument and mapped to the Periodic Table of Arguments' technical classifications (such as '1 pre FF', a first-order predicate argument relating two factual propositions) (see Section 5).

The US2016G1tvWAGEMANS corpus serves as an open resource of empirical data on argument schemes (see Section 6). It can inform the identification of isotopes of the 36 systematic characterisations of arguments in the Periodic Table, and be used to review the appropriateness of the three discriminating properties. The corpus also serves as a resource for argument mining (the automated reconstruction of argumentative discourse). The development of the machine learning techniques that are popular for argument mining is dependent on the availability of large quantities of uniformly annotated data, which US2016G1tvWAGEMANS provides.

## 2. THE PERIODIC TABLE OF ARGUMENTS

For identifying the types of arguments in the corpus, we made use of a factorial approach to argument classification called the *Periodic Table of Arguments* (Wagemans, 2016). Within this approach, an argument type is conceived as a characterization of an inference relation, i.e., the specific way in which a premise supports a conclusion. The theoretical framework of the table consists of the following three independent, partial characteristics of arguments.

### *2.1 First-order arguments and second-order arguments*

The approach assumes that premises and conclusions of arguments are expressed by categorical propositions consisting of a subject term (S) and a predicate term (P), giving an argument the general form "$S_C$ is $P_C$, because $S_P$ is $P_P$". The distinction between first-order and second-order arguments hinges on the possibility of breaking down the subject term of the proposition expressed in the premise of the argument ($S_P$). If this element cannot be broken down any further, the argument is characterised as a first-order argument ("1"). An example is "The suspect was driving fast, because he left a long trace of rubber on the road", which has "he" as the subject of the premise. If this element can be broken down, for instance because it consists of the categorical proposition expressed in the conclusion ($S_C$ is $P_C$), the argument is characterized as a second-order argument ("2"). An example is "We only use 10% of our brain, because Einstein said so", which has the conclusion functioning as the subject of the premise and "is said by Einstein" as the predicate of the premise. In this case, the general form is instantiated as "$S_C$ is $P_C$, because ($S_C$ is $P_C$) is $P_P$".

### *2.2 Predicate arguments and subject arguments*

If the subject of the proposition expressed in the premise is identical to that in the conclusion, the underlying mechanism of the argument is based on a relation between the (different) predicates. Such an argument is characterized as a predicate argument ("pre") and has the general form "*a* is *X*, because *a* is *Y*". Both examples mentioned above can be viewed as predicate arguments ("being true" functioning as the unexpressed predicate of the conclusion in the case of the second-order argument). If the predicate of the proposition expressed in the premise is identical to that in the conclusion, the underlying mechanism of the argument is based on a relation between the (different) subjects. In this case, the argument is characterised as a subject argument ("sub") and has as its general

form "*a* is *X*, because *b* is *X*". An example is "Biking on the lawn is forbidden, because walking on the lawn is forbidden".

### *2.3 Types of propositions*

Finally, arguments are characterized on the basis of the specific combination of types of propositions they instantiate. For this purpose, the approach distinguishes between propositions of fact (F) such as "Investing in solar energy will diminish $CO_2$-emission", propositions of value (V) such as "Investing in solar energy is a good idea", and propositions of policy (P) such as "The UK should invest in solar energy". After determining the type of proposition expressed in the conclusion and the premise, the argument is characterised as a PP, PV, PF, VP, VV, VF, FP, FV, or FF argument.

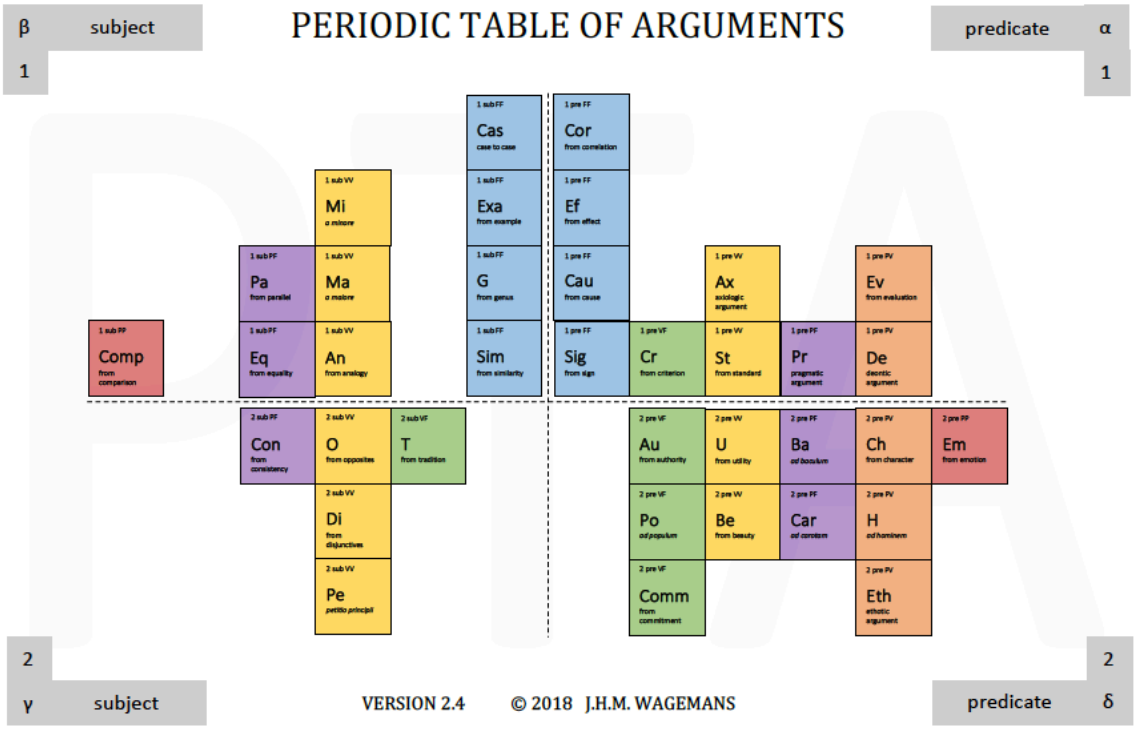


Figure 1. Visualisation of the four quadrants of the Periodic Table of Arguments.

### *2.4 Factorial full characterisation*

After having performed these three analyses, they are combined into a full characterisation of the argument type. The resulting possible characterisations can be visualised in a Periodic Table based on the distinctive partial argument characteristics. Figure 1 shows a subset of argument types classified into four quadrants: the α quadrant groups the first-order predicate argument types ("1 pre"), the β quadrant contains the first-order subject arguments ("1 sub"), the γ quadrant covers the second-order subject arguments ("2 sub"), and the δ quadrant the second-order predicate arguments ("2 pre"). The argument types in the four quadrants are then further subdivided based on the combinations of different types of propositions employed as conclusion or premise. For

example, “The suspect was driving fast, because he left a long trace of rubber on the road”, would be characterised as a “1 pre FF” argument, i.e. a first-order predicate argument (α quadrant) combining a proposition of fact with another proposition of fact. This approach to argument classification is factorial in that the theoretical framework of the *Periodic Table of Arguments* leads to 2 x 2 x 9 = 36 full characterisations of arguments.[1]

## 3. A CORPUS OF TELEVISED ELECTION DEBATES

### *3.1 The 2016 US presidential election debates*

The US2016 corpus comprises transcripts of televised debates for the 2016 presidential elections in the United States of America. These debates constitute a type of communicative activity within the political domain. The context in which communication takes place influences the argumentative activity, as it determines, e.g., the outcomes aimed for, the roles of the participants involved, and the rules or conventions with respect to the argumentative means available to them (van Eemeren, 2010). The interests and values of the individual participants further shape the practice (Fairclough, 2006): the context of televised election debates is heavily influenced by the candidates' objective to persuade the electorate to vote for them, and the broadcasting networks' aim of providing a fair and well-viewed platform for doing so.

Ever since the first televised election debate between the then US presidential candidates John F. Kennedy and Richard Nixon in 1960, the debates have played an important role in the democratic process in many countries (Kraus, 2013). The general election and the associated television debates between Hillary Clinton and Donald Trump as the candidates for the two dominant political parties in the US (respectively the Democratic Party and the Republican Party) took place in the Autumn of 2016.[2] Prior to the general elections, both main parties held primary elections and caucuses to elect their party’s candidate for the presidency. These primaries were also preceded by television debates between the leading prospective candidates in 2015 and 2016.

While the format of each of the debates is slightly different, there are some recurring characteristics. Being television debates, the discourse is spoken, with transcripts available retrospectively through a variety of sources, and video recordings broadcast live and available afterwards. The participants are expected to use language that is appropriate for the occasion. A selection of a limited number of candidates is invited to these events, moderated by anchors and journalists from the television networks that air them (among others, CBS, CNN, Fox News, and NBC).

The television networks' moderators pose questions to the invited candidates, and guide the debate (for example by keeping time and order), while the candidates make opening statements, answer the moderators' (and occasionally the public's) questions, defend their views and challenge those of their political opponents, in an attempt to garner more support among the electorate. For the general elections, three television

[1] More information about the Periodic Table of Arguments is also available online at www.periodic-table-of-arguments.org.

[2] In the current paper, we focus exclusively on the debates between the (prospective) candidates of the two dominant parties in US politics.

debates were organised between Democratic candidate Clinton and Republican candidate Trump, and one debate between their candidate vice-presidents. For the primaries, the Republican party held 12 debates for the front-runners and seven so-called `undercard' debates between the next tier of candidates. The Democratic party held 10 primary debates. As time went on and more of the candidates withdrew their candidacy, the number of participants declined over the course of these series of debates.

The argumentation encountered in the debates is not always nicely signalled linguistically, or even and intuitively clear. The television debates are a spoken genre of discourse, such that the history of the dialogue is not entirely available to the participants – depending on their memory, of course – which may lead to repetitions and contradictions of what was said earlier. Furthermore, candidates cannot always rely on their prepared and practised lines and topics, but have to respond to unexpected turns and twists, and to interaction with the other candidates and moderators. Because responding well to such dynamic situations is expected to instil the voters' confidence in the candidate, candidates receive support to varying degrees from communication professionals in their preparation and training, and rely on their experience in political debating.

The context of televised election debates fosters a mixture of well-structured and well-presented argumentation that appears to have been prepared in advance, and impromptu argumentation originating from the need to cope with the interactional dynamics. The level of noise in the data – in terms of e.g. crosstalk, unconventional use of discourse markers, and low discourse cohesion – poses a challenge in the analysis of the argumentation. Consider Example (1), advanced by then prospective candidate (now President) Trump.[3] Trump anticipates his claim about the topic of immigration to not be accepted outright. He therefore supports it with multiple statements, but does so in a non-straightforward fashion. Upon closer inspection, Trump's support relies mostly on the rhetorical device of repetition, with several of his assertions constituting a relation of rephrase rather than inference. By relying on varying ways of presenting the same content within a superficially inferential reasoning structure, Trump introduces an element of circularity.

(1) Donald Trump: *So, if it weren't for me, you wouldn't even be talking about illegal immigration, Chris. You wouldn't even be talking about it. This was not a subject that was on anybody's mind until I brought it up at my announcement. And I said, Mexico is sending. Except the reporters, because they're a very dishonest lot, generally speaking, in the world of politics, they didn't cover my statement the way I said it.*

In addition to the annotated transcripts of the television debates, the US2016 corpus contains annotated social media posts extracted from Reddit. The communicative context of social media posts leads to different conventions and communicative characteristics. Since, in the current paper, we only focus on a television debate sub-corpus of US2016, we will not go into detail on the Reddit sub-corpora – these are reported on elsewhere (Visser et al., 2018a).

[3] The annotation of example (1) – taken from our corpus of the first Republican primaries television debate on 6 August 2015 in Cleveland, Ohio – is available online at aifdb.org/argview/10829.

*3.2 Annotation with Inference Anchoring Theory*

Four annotators, trained in the use of Inference Anchoring Theory (IAT) (Reed & Budzynska, 2011), annotated the US2016 corpus that we take as a case in point in the current paper. Building on insights from discourse analysis and argumentation studies, IAT explains argumentative conduct in terms of the anchoring of argumentative reasoning in communicative interaction. Drawing on Speech Act Theory (Austin, 1962; Searle, 1969), the anchoring is theoretically conceptualised as the `illocutionary connection' between locutions in dialogue and their propositional content. IAT then allows the analytical concepts and annotations to be represented in terms of the Argument Interchange Format ontology (Chesñevar et al., 2006), resulting in a graph-based representation that facilitates the computational processing of data and procedures.

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.

- *Segments* divide the (transcribed) text into locutions, consisting of a speaker designation and an `argumentative discourse unit' (a text span with discrete argumentative function) (Peldszus & 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*.

The annotation has been validated by means of calculating the inter-annotator agreement on a 11.3% sample, resulting in a Cohen's (1960) κ of 0.610, and a CASS (Duthie et al., 2016) κ of 0.752 – both indicating substantial agreement according to Landis and Koch's 1977) standard interpretation of the kappa metric. The resulting annotated US2016 corpus is freely available online at corpora.aifdb.org/US2016. We compiled some of the quantitative characteristics of the US2016 corpus in Table 1. Additionally, the table contains the properties of the US2016G1tv sub-corpus, the extended annotation of which we discuss in the current paper. Aside from a basic word count, Table 1 comprises counts of, e.g., locutions (text segments), some illocutions, arguments ('inference'), and counterarguments ('conflict').

| Corpus | Word tokens | Locutions | Illocutions | Propositions | Inference | Conflict | Rephrase | Arguing | Disagreeing | Restating |
|---|---|---|---|---|---|---|---|---|---|---|
| US2016G1tv | 17190 | 1584 | 2285 | 1473 | 505 | 79 | 140 | 507 | 62 | 121 |
| US2016 | 97999 | 8937 | 13331 | 8099 | 2830 | 942 | 764 | 2788 | 907 | 576 |

Table 1. Quantitative annotation properties of the US2016 and US2016G1tv corpora.

## 4. ANNOTATION WITH THE PERIODIC TABLE OF ARGUMENTS

The annotation of argument schemes on the basis of the Periodic Table of Arguments is treated as an extension of the existing IAT-annotated argument structure of US2016G1tv. Because the typology of the Periodic Table of Arguments is based on the interplay between three distinguishing characteristics of the arguments, the annotation task has been deconstructed into three partial classification sub-tasks. Two annotators trained in annotation with the Periodic Table of Arguments, each carried out the three classification sub-tasks on 55% of the inferential relations and the related propositions of the US2016G1tv corpus. Based on those partial results an aggregated final classification of the argumentative inferences is produced with one of the 36 possible main types of the Periodic Table of Arguments (e.g. *1 pre FF*). If any of the inference relations or propositions involved in an argument cannot be classified, this leads to a classification as *Default Inference* in the final aggregation step. Similarly, any inference relation involving several premises without a dominant proposition type is labelled *Default Inference*.

### *4.1 Annotation guidelines*

- *First-order and second-order arguments:* An inference relation is classified as *first-order* if it connects two propositions each containing a subject-predicate pair. An inference relation is classified as *second-order* if its premise is a locution (often the result of reported speech), or if the premise is otherwise applying a predicate to the full proposition in the conclusion.

- *Predicate and subject arguments:* An inference relation is classified as a *predicate argument* if the propositions involved share the same subject term to which different predicates are applied, and as a *subject argument* if vice versa. This classification is made more complicated by the fact that natural language generally does not neatly follow the subject-predicate structure of categorical propositions, while the IAT analysis does not mandate such reconstruction of propositions either. This means that the annotator has to make a reconstructive interpretation of the proposition as if it were a categorical proposition, to then categorise it – in order to respect the starting point of not changing the original annotation aside from classifying the types of argumentative inferences, i.e. argument schemes.

- *Propositions of fact, value and policy:* A proposition is classified as a *proposition of fact* if its veracity can be verified through empirical observation, as a *proposition of value* if it contains some evaluation (whether ethical, aesthetical, legal, or logical), and as a *proposition of policy* if it expresses an act or policy to be carried out.

### *4.2 Validation*

The annotation guidelines are validated by calculating the inter-annotator agreement for the three partial classifications, as well as for the final aggregated schemes. For the classification of *first-order* and *second-order* arguments, a random sample of 10.0% was annotated by both annotators, resulting in a Cohen's κ of 0.658. While generally not considered a low κ – still amounting to substantial agreement (Landis & Koch, 1977) – this is the lowest inter-annotator agreement of all three sub-tasks. The lower score is a result of the set of arguments being unbalanced with a preponderance of *first-order* arguments: 481 *first-order* to only 11 *second-order* arguments). This imbalance throws off the calculation of the Cohen's κ metric, as becomes clear when calculating the corresponding percentage agreement of 98.0% between the two annotators.

Also on a 10.0% sample, the classification of *predicate/subject* arguments results in a Cohen's κ of 0.851. The classification of propositions as *fact/value/policy* yields a Cohen's κ of 0.778 on a 13.4% sample. The inter-annotator agreement for the aggregated argument scheme classification is based on a 10.4% sample, resulting in a Cohen's κ of 0.689. This means that the partial and final annotations all fall within the range of substantial to almost perfect agreement.

## 5. THE US2016G1tvWAGEMANS CORPUS

The annotation of argument schemes based on the Periodic Table of Arguments is compiled in the US2016G1tvWAGEMANS corpus (available online at corpora.aifdb.org/US2016G1tvWAGEMANS). Each of the previously annotated argumentative inference relations in the US2016G1tv corpus has been labelled on the basis of the three distinguishing characteristics defined by the Periodic Table. For example, Clinton's

argument in (2) is classed as a *1 pre PV* argument.[4] Clinton defends a policy proposal (*P*) (people on the terrorist watch list should be restricted from buying a gun), by drawing an analogy to a value statement (*V*) (people on the terrorist watch list are too dangerous to fly). She does so by means of two first-order propositions (*1*), that share a common subject term (*pre*) (people on the terrorist watch list).

(2) CLINTON: *And we finally need to pass a prohibition on anyone who's on the terrorist watch list from being able to buy a gun in our country. If you're too dangerous to fly, you are too dangerous to buy a gun.*

| | **Total number** | **First-order argument** | **Second-order argument** | **Subject argument** | **Predicate argument** | **Proposition of value** | **Proposition of policy** | **Proposition of fact** |
|---|---|---|---|---|---|---|---|---|
| Inferences | 505 | 481 | 11 | 124 | 308 | | | |
| Propositions | 1473 | | | | | 383 | 110 | 289 |

Table 2. Results for annotation sub-tasks of distinguishing *first-*/*second-order* and *subject*/*predicate* arguments, and propositions of *value*/*policy*/*fact*

Quantitative metrics of the annotated corpus are compiled in Tables 2 and 3. Table 2 shows the counts for the three sub-annotations, while Table 3 contains the aggregated results. Notably low is the proportion of *second-order arguments*: accounting for only 11 out of a total of 505 inference relations. On the other end of the scale, the number of *default inference* classifications is uncomfortably high: 85 out of 505 inference relations have remained unclassified, amounting to 17% of the corpus. The main reason for the high number of *default inference*s is that a failure to classify a proposition or relation in any of the three annotation sub-tasks will cause the combination of the three sub-tasks to default into an unlabelled classification. In other words, if a proposition cannot be classified in terms of *policy*/*value*/*fact*, for example because it is too vague, or if the relation is not clearly *first-* or *second-order*, or if the propositions are incomplete to the extent that it's not clear whether the *subject* or *predicate* is responsible for the transferring of justificatory force, then the aggregated final classification of the argument as a whole fails and defaults.

[4] The annotation of example (2) – taken from our corpus of the first General Election television debate on 26 September 2016 in Hempstead, New York – is available online at aifdb.org/argview/10850.

| Argument scheme | Count |
|---|---|
| Default inference | 85 |
| 1 pre VV | 78 |
| 1 pre VF | 61 |
| 1 sub VV | 50 |
| 1 pre FF | 47 |
| 1 pre FV | 27 |
| 1 pre PP | 27 |
| 1 pre PV | 25 |

| Argument scheme | Count |
|---|---|
| 1 sub VF | 23 |
| 1 sub FV | 17 |
| 1 pre PF | 15 |
| 1 sub FF | 10 |
| 1 pre VP | 8 |
| 1 sub PF | 7 |
| 1 pre FP | 5 |
| 1 sub PP | 5 |

| Argument scheme | Count |
|---|---|
| 1 sub VP | 4 |
| 1 sub PV | 3 |
| 2 pre FV | 3 |
| 2 pre VF | 2 |
| 2 pre VV | 2 |
| 2 pre FF | 1 |

Table 3. Aggregated results for the annotation of US2016G1tv with the Periodic Table of Arguments

## 6. IMPLICATIONS AND CONCLUSION

The US2016G1tvWAGEMANS is the first corpus of argumentative discourse annotated on the basis of the Periodic Table of Arguments, and it is one of the largest and most reliably annotated corpora of argument schemes publicly available. We intend the corpus to provide a resource for the quantitative study of argument schemes, and for computational approaches to argumentation in particular. Elsewhere (Visser et al., 2018b), we have reported on the annotation of the same source material (i.e. the US2016G1tv corpus) on the basis of Walton's (Walton, Reed & Macagno, 2008) typology of argument schemes. The dual annotation of the same source material with two distinct typologies of argument schemes makes it possible to do comparative studies. For example, the dual annotation can be used to map the technical names of Wagemans' Periodic Table of Arguments (e.g., *1 pre FF*) to the colloquial names of argument schemes familiar from Walton's typology (e.g., *Argument from sign*). This will expand the range of 'isotopes' in the Periodic Table: identifying the various sub-types of the larger classes delineated by the technical types (thereby creating the individual boxes in the four quadrants of Figure 1). A co-occurrence matrix, such as Table 4, can be used to look for regularities in the annotations based on the two typologies: it shows the number of arguments classified for the more common combination of the two typologies.

Corpus-based studies can provide new insights into the dialogical nature of argument schemes, and how they are employed in different communicative contexts. The frequency of particular schemes can further characterise the argumentative preconditions of the activity type, and feed into the study of which prototypical argumentative patterns are actually stereotypical (van Eemeren, 2017). A closer look at the linguistic surface structure associated with a particular type of argument scheme can lead to a greater insight into the use of discourse markers indicative of argumentation (van Eemeren, Houtlosser & Snoeck Henkemans, 2007); which in turn would be highly valuable for argument mining – the automated reconstruction of argumentative content from a natural language text (Lawrence & Reed, 2015). Also for approaches to argument mining that do not rely on discourse markers, the availability of a robust annotated dataset is of great value for machine learning purposes (Peldszus & Stede, 2013).

| | Argument from alternatives | Argument from analogy | Argument from bias | Argument from cause to effect | Argument from composition | Argument from consequences | Argument from danger appeal | Argument from example | Argument from fear appeal | Argument from popular opinion | Argument from popular practice | Argument from position to know | Argument from sign | Argument from values | Argument from verbal classification | Circumstantial ad hominem | Ethotic argument | Generic ad hominem | Practical reasoning | Practical reasoning from analogy | Pragmatic argument from alternatives | Default Inference |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **1 pre FF** | 1 | 2 | 1 | 6 | 1 | 5 | · | 13 | · | · | 1 | 4 | 3 | · | 2 | · | 1 | 1 | 2 | · | 1 | 2 |
| **1 pre FV** | · | · | · | 2 | · | 4 | · | 4 | · | · | 1 | 1 | · | · | 3 | 2 | 1 | 4 | 1 | · | · | · |
| **1 pre FP** | · | · | · | · | · | · | · | 3 | · | · | · | · | · | · | · | · | · | 1 | · | · | · | · |
| **1 pre VF** | · | · | 1 | 11 | 1 | 6 | · | 12 | · | · | · | 2 | 13 | · | 3 | 4 | · | 4 | 1 | · | 1 | · |
| **1 pre VV** | · | 1 | 2 | 9 | 1 | 2 | 1 | 13 | · | · | · | 1 | 7 | 3 | 11 | 4 | · | 6 | 9 | · | 3 | 2 |
| **1 pre VP** | · | · | · | · | · | · | · | 3 | · | · | · | · | · | 2 | 1 | · | · | · | 2 | · | · | · |
| **1 pre PF** | · | · | · | · | · | 2 | · | 3 | · | · | · | · | · | · | · | · | · | · | 2 | 2 | 3 | · |
| **1 pre PV** | 3 | · | 1 | · | · | 3 | · | · | · | 2 | · | 1 | · | 5 | · | · | · | · | 5 | 1 | 3 | · |
| **1 pre PP** | · | · | · | · | 2 | 2 | 1 | 3 | · | · | · | · | · | · | 1 | · | · | · | 14 | · | 4 | · |
| **1 sub FF** | · | 2 | 1 | · | · | 1 | · | 4 | · | 1 | · | 1 | · | · | · | · | · | · | · | · | · | · |
| **1 sub FV** | 2 | 1 | · | · | · | 3 | · | · | 1 | · | 2 | · | · | · | 1 | 2 | 1 | 2 | 1 | · | 1 | · |
| **1 sub VF** | · | 2 | · | 4 | · | 1 | · | 4 | 1 | 1 | · | · | 2 | · | 2 | 3 | · | · | · | · | 1 | · |
| **1 sub VV** | · | · | 2 | 6 | 1 | 2 | 1 | 7 | · | 1 | · | · | 4 | 3 | 4 | 8 | · | 6 | 1 | · | · | 1 |
| **1 sub VP** | · | · | · | · | · | · | · | 2 | · | · | · | · | · | 1 | · | · | · | · | 1 | · | · | · |
| **1 sub PF** | · | · | · | · | · | 1 | · | 2 | 1 | · | · | · | 1 | · | · | · | · | · | · | · | 1 | · |
| **1 sub PP** | · | · | · | · | · | 2 | · | 1 | 2 | · | · | · | · | · | · | · | · | · | · | · | · | · |
| **Default Inference** | 3 | · | · | 10 | · | 6 | 4 | 7 | 6 | 2 | 2 | 2 | 7 | 1 | 2 | 1 | 2 | 3 | 6 | 1 | 4 | 12 |

Table 4. Co-occurrence matrix of the most common argument scheme annotations in US2016G1tvWALTON (columns) and US2016G1tvWAGEMANS (rows)

An unavoidable difficulty in annotating argument schemes is posed by the fact that natural-language argumentation is not expressed in terms of categorical propositions or other abstractions. A reconstructive step is required to get from naturally expressed argument to the schematic abstractions that the typologies of argument schemes are based on. While technically possible, IAT annotation (see Section 3.2) ordinarily does not go into that reconstructive depth. The different approaches available to address this issue will be explored in future work.

ACKNOWLEDGEMENTS: This research was supported in part by the Engineering and Physical Sciences Research Council (EPSRC) in the UK under grant EP/N014871/1.

REFERENCES

Austin, J. L. (1962). *How to do things with words*. Clarendon Press.

Chesñevar, C., McGinnis, J., Modgil, S., Rahwan, I., Reed, C., Simari, G., South, M., Vreeswijk, G., & Willmott, S. (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.

Duthie, R., Lawrence, J., Budzynska, K., & Reed, C. (2016). The CASS technique for evaluating the performance of argument mining. In, *Proceedings of the 3rd workshop on argument mining* (pp. 40-49). Association for Computational Linguistics.

Fairclough, N. (2006). Genres in political discourse. In K. Brown (Ed.), *Encyclopedia of language & linguistics (second edition)* (pp. 32-38). Oxford: Elsevier.

Kraus, S. (2013). *Televised presidential debates and public policy*. Taylor & Francis.

Landis, J., & Koch, G. (1977). The measurement of observer agreement for categorical data. *Biometrics 3*, 159-174.

Lawrence, J., & Reed, C. (2015). Combining argument mining techniques. In, *Proceedings of the 2nd workshop on argumentation mining* (pp. 127-136). ACL.

Peldszus, A., & Stede, M. (2013). From argument diagrams to argumentation mining in texts: A survey. *International journal of cognitive informatics and natural intelligence*, *7*(1), 1-31.

Reed, C., & Budzynska, K. (2011). How dialogues create arguments. In F. H. van Eemeren, B. Garssen, D. Godden, & G. Mitchell (Eds.), *Proceedings of the 7th conference of the international society for the study of argumentation (ISSA)*. Amsterdam: SicSat.

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

van Eemeren, F. H. (2010). *Strategic maneuvering in argumentative discourse: Extending the pragma-dialectical theory of argumentation*. Amsterdam: John Benjamins.

van Eemeren, F. H. (2017). *Prototypical argumentative patterns: Exploring the relationship between argumentative discourse and institutional context*. Amsterdam: John Benjamins.

van Eemeren, F. H., Houtlosser, P., & Snoeck Henkemans, A. F. (2007). *Argumentative indicators in discourse: A pragma-dialectical study*. Dordrecht: Springer.

Visser, J., Duthie, R., Lawrence, J., & Reed, C. (2018a). Intertextual correspondence for integrating corpora. In N. Calzolari et al. (Eds.), *Proceedings of the eleventh international conference on language resources and evaluation (LREC 2018), Miyazaki, Japan* (pp. 3511-3517). European Language Resources Association (ELRA).

Visser, J., Lawrence, J., Wagemans, J., & Reed, C. (2018b). Revisiting computational models of argument schemes: Classification, annotation, comparison. In S. Modgil, K. Budzynska & J. Lawrence (Eds.), *Proceedings of the 7th international conference on computational models of argument (COMMA 2018), Warsaw, Poland*. IOS Press.

Wagemans, J. H. M. (2016). Constructing a periodic table of arguments. In P. Bondy & L. Benacquista (Eds.), *Argumentation, Objectivity, and Bias: Proceedings of the 11th international conference of the Ontario society for the study of argumentation (OSSA), 18-21 May 2016* (pp. 1-12). Windsor, ON: OSSA.

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