Revisiting Computational Models of Argument Schemes: Classification, Annotation, Comparison

Jacky VISSER. [a,1], John LAWRENCE [a], Jean WAGEMANS [b] and Chris REED [a]

[a] *Centre for Argument Technology, University of Dundee, United Kingdom*
[b] *Argumentation and Rhetoric, University of Amsterdam, The Netherlands*

Abstract. In this paper, we present an in-depth comparative analysis of two classi-fications of argument schemes: Walton's typology and Wagemans' Periodic Table of Arguments. We describe annotation guidelines for each classification and apply these to a corpus of arguments from the 2016 US presidential debates. In so doing, we achieve substantial inter-annotator agreement, and produce what, to the best of our knowledge, are the two largest and most reliably annotated corpora of argument schemes in dialogical argumentation publicly available. In describing the creation and comparison of these corpora, we discuss the strengths of each, with an eye towards both computational modelling and argument mining.

Keywords. annotation, argument mining, argument schemes, classification, corpus

1. Introduction

The notion of 'argument scheme', referring to the conventionalised grounds for an argumentative inference, is essential to a proper interpretation and evaluation of argumentation.[2] Although the concept was developed for different purposes, it has found uptake with the computational community: the importance of schemes for the computational modelling of argumentation is reflected in the inclusion of chapters or sections devoted to computational models of argument schemes in Rahwan and Simari's overview volume *Argumentation in Artificial Intelligence* [26], the *Handbook of Argumentation Theory* [31], and the *Handbook of Formal Argumentation* [2].

Theory-driven applications of computational models of argument and empirically-oriented work alike, rely on data about the actual use of argumentation in practice. This data can come from the qualitative appraisal of selected examples, but quantitative approaches, while labour intensive, are gaining traction, especially motivated by the rise of argument mining [5]. Quantitative approaches require (preferably large) corpora of actual argumentative discourse annotated with the necessary theoretical concepts. In the

---

[1]Corresponding Author: School of Science and Engineering, University of Dundee, Nethergate, Dundee, DD1 4HN, United Kingdom; E-mail: j.visser@dundee.ac.uk

[2]In the literature, various authors use different terms to signify the same general idea: argument scheme, argumentation scheme, argumentative scheme. In the present paper, we favour the term 'argument scheme'.

current paper, we adopt such a corpus-linguistics perspective to revisit the computational models and classifications of argument schemes (§2). To this avail, we present two annotated corpora of argument schemes, each based on a different schemeset: annotating once with Douglas Walton's typology of argument schemes (§4), and once with Jean Wagemans' Periodic Table of Arguments (§5). We extend the annotation of the existing US2016 corpus (§3), to construct what, to the best of our knowledge, are the two largest and most reliably annotated corpora of argument schemes in dialogical argumentation publicly available. Moreover, because both parallel annotations are done on the same original dataset, we open up the possibility to carry out comparative experiments on the different ways of classifying argument schemes – whether with a focus on computational scheme modelling or on automated classification and mining (§6).

2. Automatically Classifying Argument Schemes

2.1. Classifications of Argument Schemes

Explicating the inferential principles underpinning argumentation has been a scholarly objective since Antiquity [29]. As one such explanation, the notion of 'argument scheme' was introduced during the second half of the 20th century [10]. Although Perelman and Olbrechts-Tyteca introduced the similar notion of 'argumentative scheme' in their New Rhetoric [24], the current interpretation of argument scheme goes back to Hastings' PhD dissertation [12] and the independent conceptualisation in van Eemeren et al.'s first handbook of argumentation theory [33]. Argument schemes capture the conventionally acceptable patterns of reasoning that are applied in persuasive communication, substantiating the inferential connection between premises and conclusion. The defeasibility of the schemes sets them apart from the strict reasoning patterns of classical formal logic (e.g., Modus Ponens), and the dialogical nature of the schemes is evident in their association with 'critical questions' used to evaluate the acceptability of an applied argument scheme.[3]

Since their introduction, argument schemes have become a central issue in modern argumentation studies, leading to a variety of classifications, e.g., by van Eemeren and Grootendorst [32], Kienpointner [14], and Schellens [30]). In the computational community, it has particularly been Walton's [36,39] take on argument schemes that found uptake. Walton's classification comprises a great variety of schemes, described in some detail, but with the flexibility to allow adjustments in order to fit a scheme to a desired domain-specific application (see, e.g., the revisions and extensions of the *practical reasoning* scheme by Atkinson and Bench-Capon [1], and Kokciyan et al. [15].)

2.2. Mining Argument Schemes

The automatic identification of argument scheme occurrences remains a major challenge. As previously discussed, a large number of scheme classifications exist, with additional domain specific schemes utilised in specialised areas: Green [11] lists ten custom argument schemes targeted at genetics research articles; Musi et al. [21] present a set of

---

[3]Not all current accounts of argument schemes emphasise this communicative angle, favouring a reinterpretation of the interactional critical questions as static exceptions or defeaters.

guidelines for the annotation of argument schemes based on the Argumentum Model of Topics [28]; Wyner et al. [40] describe a consumer argument scheme, with the structure of this scheme used to guide an argument identification process.

Walton [37] addresses these challenges, presenting a systematic approach to identifying arguments and their schemes by first identifying the arguments occurring in a piece of text, followed by identification of specific known argument schemes. Walton, however, points out that beyond this initial identification there are likely to be issues differentiating between similar schemes and he suggests the development of a corpus of borderline cases to address the issue.

The work of Feng and Hirst [9] follows a similar path, aiming to reconstruct enthymemes by first classifying to an argument scheme, then fitting the propositions to an associated template, and finally inferring the enthymemes. For the first step of fitting one of the top five most commonly occurring Waltonian argument schemes to a pre-determined argument structure, accuracies of 0.63–0.91 are achieved in one-against-others classification and 0.80–0.94 in pairwise classification.

Another approach to identifying the occurrence of schemes is given by Lawrence et al. [20], where, rather than considering features of the schemes as a whole, the individual scheme components are identified and then grouped together into a scheme instance. In this case, only two schemes (*expert opinion* and *positive consequences*) are considered, and classifiers are trained to identify their individual component premises and conclusion. By considering the features of the individual types of these components, F-scores between 0.75 and 0.93 are recorded for identifying at least one component part of a scheme.

With the data currently available, the ontologically rich information provided by argument schemes has been demonstrated to be a powerful component of a robust approach to argument mining. Collaboration amongst analysts as well as the further development of tools supporting argument schemes is essential to growing the datasets required to improve on these techniques. Clear annotation guidelines and the development of custom argument schemes for specific domains will hopefully result in a rapid growth in the material available and further increase the effectiveness of schematic classification.

3. Source Data

3.1. The 2016 US presidential election debates

The argument scheme annotations are extensions of the existing US2016 corpus (available at `corpora.aifdb.org/US2016`). This corpus contains argumentative debate and discussion centred around the 2016 presidential elections in the United States of America. It comprises annotations of transcripts of televised election debates and online reaction to those debates on the Reddit social media platform, and the intertextual correspondence between those two genres [34]. The transcripts of the television debates cover full debates from the primaries of the two political parties dominating US politics, and from the general election — all collected from The American Presidency Project [25]. The social media commentary is manually retrieved from Reddit mega-threads dedicated to the discussion of the ongoing election debates. To the best of our knowledge, the combined US2016 corpus is the larges of its kind: combining detailed argumenta-

tive and discursive annotation on dialogical text genres. In the argument scheme annotation, we focus our attention on the first general election debate between Hilary Clinton (Democrat) and Donald Trump (Republican). This US2016G1tv sub-corpus is available at `corpora.aifdb.org/US2016G1tv`.

3.2. *Annotation with Inference Anchoring Theory*

The US2016 corpus has been annotated on the basis of Inference Anchoring Theory (IAT) [4]. IAT builds on insights from discourse/conversation analysis, speech act theory, and argumentation studies, as a way of explaining how the propositional reasoning that is appealed to in argumentation is anchored in discourse (whether written or spoken). IAT annotation results in an Argument Interchange Format (AIF) [6] compliant graph representation of both the reconstructed argumentation structure and its discursive anchoring in the analysed text segments. The resulting AIF graph is a constellation of information nodes (I-nodes) and scheme nodes (S-nodes) connected with unlabelled directed edges, and is stored online in the AIFdb argument repository at `aifdb.org` [17].

IAT underpins the annotation guidelines used by the four expert annotators involved in the annotation of the US2016 corpus. Based on a 11.3% sample, the agreement between the annotators was substantial (according to the Landis and Koch [16] interpretation), with a Cohen's κ [7] of 0.610. Duthie et al. [8] have, however, argued that Cohen's κ misrepresents the interdependency between some of the sub-tasks involved in the annotation process. To do justice to such interdependency, Duthie et al. propose to calculate a Combined Argument Similarity Score (or CASS-κ) by combining independent agreement scores for the sub-tasks of text segmentation, discourse annotation, and propositional annotation. When taking into account the interplay between these constitutive tasks, the average inter-annotator agreement in terms of CASS-κ is 0.752.

The four annotators made use of the OVA analysis software [13] (freely available at `ova.arg.tech`). While the full annotation guidelines available at `arg.tech/US2016-guidelines` deal with complex issues such as anaphoric references, epistemic modalities, repetition, punctuation, discourse indicators, interposed text, and reported speech, we summarise below those aspects of the annotation that are essential for a proper understanding of the corpus study.

Locutions: The original text is segmented into locutions. A locution consists of a speaker designation and an 'argumentative discourse unit' (ADU) [23], a text span with discrete argumentative function (often directly resulting in the introduction of an inference, conflict or rephrase in the argumentation structure – see below). In accordance with the AIF ontology [27], locutions are modelled as L-nodes, a sub-type of I-node.

Transitions: Functional discourse relationships are represented as transitions connecting the segmented locutions. The transitions reflect the dialogue protocol underpinning the discourse. Transitions, or TA-nodes, are a type of S-node that connects L-nodes.

Illocutionary connections: The communicative intention encapsulated in a locution is annotated by means of illocutionary connections that relate the locutionary to the propositional dimension of the analysis. In AIF terms, illocutionary connections are YA-nodes, a sub-type of S-node.

Propositions: Most illocutionary connections lead to the reconstruction of the propositional content of the associated locution. Propositions are modelled as I-nodes.

Inference, conflict and rephrase: Generally connecting one proposition to another, the argumentative relations of inference, conflict and rephrase respectively indicate jus-

tificatory defence, refutatory incompatibility, and revisionary reformulation. The propositional relations are modelled as sub-types of S-nodes: as RA-, CA-, and MA-nodes.

3.3. The US2016G1tv corpus

In Table 1, we have collected the most relevant properties of the 17,190-word (tokens) US2016G1tv corpus. For reference, we also include those of the full US2016 corpus comprising almost 100,000 words. The properties are retrieved automatically using the Argument Analytics module [18] of the Argument Web [3] at `analytics.arg.tech`. Both corpora are freely available online through AIFdb Corpora [19] (at `corpora.aifdb.org`). In Table 1, we include counts of *Arguing*, *Disagreeing* and *Restating* as the illocutionary connections most commonly used to anchor argumentatively relevant relations between propositions.

Table 1. Properties of the US2016 and US2016G1tv corpora.

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

4. Annotation with Walton's Classification of Argument Schemes

4.1. Walton's Classification of Argument Schemes

Walton's longstanding scholarly engagement with the topic of argument schemes, within various domains and from various angles, has resulted in an eclectic collection of schemes conventionally occurring in argumentative practices, ranging from colloquial discussion to argumentation in the legal domain (see, e.g., [36,39]). Some of Walton's schemes are commonly distinguished in dialectical or informal-logical approaches to argumentation (e.g. *argument from sign* or *argument from cause to effect*). Others, however, are more exotic or highly specialised (e.g. *argument from arbitrariness of a verbal classification* or *argument from plea for excuse*), are closer to modes of persuasion in a rhetorical perspective on argumentation (e.g. *ethotic argument*), or would by some be readily relegated to the realm of fallacies (e.g. *hasty generalisation*). The list also includes composite schemes that combine aspects from various schemes into one (e.g. *practical reasoning from analogy* combining *practical reasoning* and *argument from analogy*).

Despite several proposals to systematise Walton's schemeset by imposing some ordering principle on the resulting typology (see, e.g., the distinction between the classes of 'reasoning', 'source-based arguments' and 'applying rules to cases' in [39, pp.347–363], and the subsequent [38]), to the best of our knowledge, no exhaustive and systematic account exists to date. As the starting point for our annotation of argument schemes based on Walton's typology, we therefore resort to the collection in the 2008 book "Argumentation Schemes" by Walton, Reed and Macagno [39]. Depending on what is counted as

a type of argument scheme (i.e. whether sub-types are counted or not), the book contains upwards of 60 schemes. The schemes are presented with their distinctive pattern of premises and conclusion, and with an associated list of critical questions, mostly drawn from Walton's previous work.

4.2. Annotation Guidelines for Walton's Argument Schemes

Two expert annotators trained in argumentation analysis and with prior knowledge of Walton's typology of argument schemes each classified 55% of the RA-nodes in the US2016G1tv corpus in accordance with Walton's typology. Only the main schemes from the 2008 *Argumentation Schemes* book [39] are considered, which still results in a choice from 60 possible labels to be applied to each of the more than 500 previously analysed inference relations in the corpus (see §3.3).

To facilitate the process, the annotators made use of a classification decision tree: an indicative heuristic for the annotators, to intuitively support their coding task. The fragment of the heuristic in Figure 1 shows the indication of the grounds for making a decision between various action-oriented argument schemes. The decision tree ties into the actual guidelines consisting of Chapter 9 of [39, pp. 308–346]: *A User's Compendium of Schemes*. Since the annotation relies on the existing annotated argumentation structure, in some cases, the schemes are applied in a simplified, condensed or partial manner, to fit the original annotation. In addition, one auxiliary class is introduced for arguments not fitting any of the 60 main schemes: *default inference*.
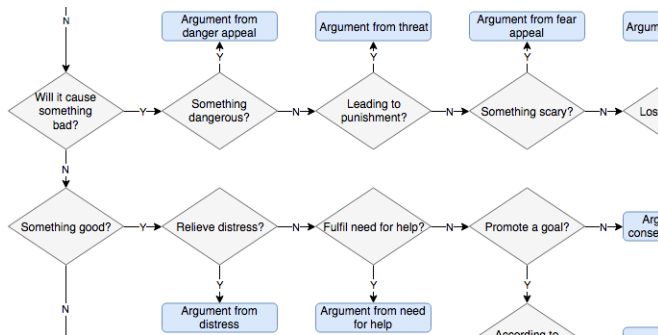

Figure 1. Distinguishing between action-oriented argument schemes with the decision tree heuristic.

4.3. Results of the Annotation with Walton's Argument Schemes

A sample of 10.2% of the corpus was annotated by both annotators, resulting in a Cohen's κ [7] of 0.723; well within substantial agreement [16]. Some classes of argument scheme turned out to be particularly difficult to distinguish: e.g., Example (1) was classified by one annotator as *practical reasoning*, related to promoting goals, and by the other as *argument from values*, related to promoting values.

(1)   Hilary Clinton: *What I have proposed would be paid for by raising taxes on the wealthy [...] I think it's time that the wealthy and corporations paid their fair share to support this country.*

The results of the annotation in accordance with Walton's classification of argument schemes are collected in the US2016G1tvWALTON corpus (available online at `corpora.aifdb.org/US2016G1tvWALTON`). Figure 2 shows an example of the *practical reasoning from analogy* scheme mentioned in §4.1 as applied in the corpus. Of the 505 RA-nodes in the original US2016G1tv corpus, a total of 491 are annotated with one of the 60 argument scheme types in Walton's classification, leaving only 14 as *default inference*. The most common scheme, by some margin, is *argument from example*. The *argument from expert opinion* scheme, a scholarly favourite, is remarkably rare with only three occurrences.
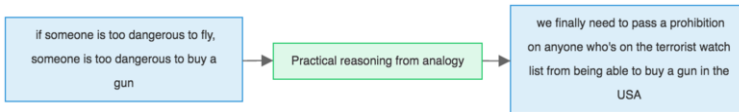

Figure 2. OVA visualisation of *practical reasoning from analogy* in US2016G1tvWALTON.

Table 2. Counts of argument schemes in the US2016G1tvWALTON corpus.

Argument scheme	Count	Argument scheme	Count
Argument from example	81	Ethotic argument	5
Argument from cause to effect	48	Practical reasoning from analogy	4
Practical reasoning	45	Argument from commitment	3
Argument from consequences	40	Argument from expert opinion	3
Argument from sign	38	Argument from waste	3
Argument from verbal classification	32	Argument from gradualism	2
Generic ad hominem	28	Argument from need for help	2
Circumstantial ad hominem	24	Argument from oppositions	2
Pragmatic argument from alternatives	23	Argument from perception	2
Argument from values	15	Argument from correlation to cause	1
Default inference	14	Argument from definition to verbal classification	1
Argument from position to know	13	Argument from division	1
Argument from fear appeal	11	Argument from ignorance	1
Argument from alternatives	9	Argument from rules	1
Argument from bias	9	Argument from vagueness of verbal classification	1
Argument from analogy	8	Argument from witness testimony	1
Argument from popular opinion	8	Argumentation from interaction of act and person	1
Argument from danger appeal	7	Pragmatic inconsistency	1
Argument from popular practice	7	Two-person practical reasoning	1
Argument from composition	6		

5. Annotation with Wagemans' Periodic Table of Arguments

5.1. Wagemans' Typology of Argument Schemes

The Periodic Table of Arguments is a classification of argument proposed by Wagemans [35] as a theoretically sound and practically useful alternative for the traditional multitude of incomplete, informal and sometimes even inconsistent descriptions of types of argument. The framework of the Table consists of three distinguishing characteristics of arguments, the superposition of which yields a factorial typology of argument that can be used for the purpose of analysing, evaluating, and generating arguments in natural language.

The first distinction is between first-order arguments and second-order arguments. The approach assumes that premises and conclusions of arguments can be reconstructed in terms of categorical propositions consisting of a subject term (*a*, *b*, etc.) and a predicate term (*X*, *Y*, etc.). If the subject term of the proposition expressed in the premise of an argument cannot be broken down any further, the argument is characterised as a first-order argument with the general form "*a* is *X*, because *b* is *Y*". If the subject term in the premise can be broken down since it consists of the categorical proposition expressed in the conclusion, then the argument is characterised as second-order, having the general form "*a* is *X*, because (*a* is *X*) is *Y*".

The second distinction is that between 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 characterised as a predicate argument and has the general form "*a* is *X*, because *a* is *Y*". 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 and has as its general form "*a* is *X*, because *b* is *X*".

Finally, arguments are characterised on the basis of the specific combination of proposition types they instantiate. For this purpose, the approach distinguishes between propositions of fact such as "investing in solar energy will diminish CO_2-emission", propositions of value such as "investing in solar energy is a good idea", and propositions of policy such as "the UK should invest in solar energy".

These three classifications are combined into a full characterisation of the argument type. The prefixes *1* and *2* indicate first-order and second-order arguments. The infixes *pre* and *sub* indicate predicate arguments and subject arguments. Finally, combinations of *P*, *V* and *F* as suffix distinguish the various combinations of propositions of policy, value and fact, respectively. For example, "unauthorized downloading is not theft, because it doesnt deprive the original owner of use" would be characterised as a *1 pre VF* argument, i.e. a first-order predicate argument combining an evaluative conclusion with a factual premise.

Combining the three distinguishing characteristics of arguments, the Periodic Table of Arguments contains 36 (= 2*2*(3*3)) main types of arguments. The 'technical' names of the 36 types can subsequently be related to corresponding 'trivial' names known from the literature on argument schemes and related typologies.

5.2. *Annotation Guidelines for Wagemans' Periodic Table of Arguments*

The procedure followed is similar to that for annotation with Walton's typology (see §4.2). Again, the annotation of argument schemes is treated as an extension of the existing annotated argument structure of US2016G1tv. However, 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 expert annotators trained in annotation with the Periodic Table of Arguments, each carried out the three classification sub-tasks on 55% of the RA-nodes and the related I-nodes of the US2016G1tv corpus. Based on those partial results an aggregated final classification of the RA-nodes 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 I-nodes or RA-node involved in an argument cannot be classified, this leads to a classification of the RA-node as *default inference* in the final aggregation step. Similarly, any RA-node involving several premises without a dominant proposition type is labelled *default inference*.

First-order and second-order arguments: An RA-node is classified as *first-order* if it connects two I-nodes each containing a subject-predicate pair. An RA-node is classified as *second-order* if its premise is an L-node (i.e. a locution, often resulting from reported speech), or if the premise is otherwise applying a predicate to the full proposition in the conclusion.

Predicate and subject arguments: An RA-node is classified as a *predicate argument* if the I-nodes 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, and neither does the IAT analysis mandate such reconstruction of I-nodes. This means that the annotator makes a reconstructive interpretation of the I-node 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 RA-nodes.

Propositions of fact, value and policy: An I-node is classified as a *proposition of fact* if it 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.

5.3. Results of the Annotation with Wagemans' Typology

The annotation guidelines are validated by means of the calculation of 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 [7] of 0.658. Also on a 10.0% sample, the classification of *predicate/subject* arguments results in a Cohen's κ of 0.851. The classification of I-nodes 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 fall within the range of substantial to almost perfect agreement [16]. The lower score for *first-/second-order* arguments is due to an unbalanced set with a predominance of *first-order* arguments, signalled by the corresponding percentage agreement of 98.0%.

We compile the counts of the aggregated argument scheme classification of the US2016G1tvWAGEMANS corpus (available online at `corpora.aifdb.org/US2016G1tvWAGEMANS`) in Table 3. Notably low is the proportion of *second-order arguments*: accounting for only 8 out of a total of 505 inference relations. Conversely, there is a high number of *default inference* classifications, especially when compared to the corresponding count in Table 2, which is why in §6 we will discuss this relative variation further.

Table 3. Counts of argument schemes in the US2016G1tvWAGEMANS corpus.

Argument scheme	Count	Argument scheme	Count	Argument scheme	Count
Default inference	85	1 sub VF	23	1 sub VP	4
1 pre VV	78	1 sub FV	17	1 sub PV	3
1 pre VF	61	1 pre PF	15	2 pre FV	3
1 sub VV	50	1 sub FF	10	2 pre VF	2
1 pre FF	47	1 pre VP	8	2 pre VV	2
1 pre FV	27	1 sub PF	7	2 pre FF	1
1 pre PP	27	1 pre FP	5		
1 pre PV	25	1 sub PP	5		

6. Comparative Discussion of Annotation Results

Table 4 shows the co-occurrence of classifications according to the two typologies of §4 and §5. Only scheme classifications occurring more than thrice are considered. The co-occurrence demonstrates how the two typologies by Walton and Wagemans relate to each other. Notable is that most of the *default inferences* in US2016G1tvWALTON are also classified as such in US2016G1tvWAGEMANS, but not vice versa. An explanation for the preponderance of *default inferences* when annotating with Wagemans' typology is that the Periodic Table of Arguments focuses on atomic arguments consisting of one premise and one conclusion, whereas the structural IAT annotation of US2016 allows multiple premises per argument. Furthermore, the aggregation process falls back on the *default inference* class if one of the three constitutive sub-classifications does not yield a positive result (i.e. a *null* label).

Table 4. Co-occurrence matrix of argument schemes in US2016G1tvWALTON and US2016G1tvWAGEMANS

	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

An advantage of Wagemans' approach is that it provides the additional value of the partial annotations. The classification of proposition types, for example, has clear

intrinsic value (see, e.g., [22]). Of the 798 propositions in the corpus, the majority of 376 is classified as *value*, followed by 298 propositions of *fact*, and 108 classifications as *policy* (with a Cohen's κ [7] of 0.778 on a 13.4% sample).

7. Conclusion

Any computational modelling of argument schemes relies upon the theoretically motivated classification or typology that the modelling starts from. We have considered two classifications of argument schemes: the popular classification of Walton [39], and the newly developed Periodic Table of Arguments by Wagemans [35]. On the basis of the two approaches, we extended an existing corpus with argument scheme annotation, resulting in two large reliably annotated parallel corpora of argumentation schemes in a dialogical discourse genre. The inter-annotator agreement in both annotations is comparable and substantial, respectively resulting in a Cohen's κ of 0.723 and 0.689.

The dialogical nature of the corpora opens up a promising future line of research in exploring the discursive aspects of argument schemes and critical questions in corpus-based studies. The corpora also provide invaluable training and test datasets for argument mining techniques. In particular, the US2016G1tvWAGEMANS corpus opens up new avenues in automatic scheme identification by providing the means to break down the objective into simpler classification tasks.

References

[1] K. Atkinson and T. Bench-Capon. Taking account of the actions of others in value-based reasoning. *Artificial Intelligence*, 254:1 – 20, 2018.

[2] P. Baroni, D. Gabbay, M. Giacomin, and L. Van der Torre. *Handbook of formal argumentation, Vol. 1*. College Publications, 2018.

[3] F. Bex, J. Lawrence, M. Snaith, and C. Reed. Implementing the argument web. *Communications of the ACM*, 56(10):66–73, 2013.

[4] K. Budzynska and C. Reed. Whence inference. Technical report, University of Dundee, 2011.

[5] K. Budzynska and S. Villata. Processing argumentation in natural language texts. In P. Baroni, D. Gabbay, M. Giacomin, and L. van der Torre, editors, *Handbook of Formal Argumentation*. 2017.

[6] C. Chesñevar, S. Modgil, I. Rahwan, C. Reed, G. Simari, M. South, G. Vreeswijk, S. Willmott, et al. Towards an argument interchange format. *The Knowledge Engineering Review*, 21(04):293–316, 2006.

[7] J. Cohen. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20(1):37–46, 1960.

[8] R. Duthie, J. Lawrence, K. Budzynska, and C. Reed. The CASS Technique for Evaluating the Performance of Argument Mining. *Proceedings of the 3rd Workshop on Argument Mining, Association for Computational Linguistics.*, pages 40–49, 2016.

[9] V. W. Feng and G. Hirst. Classifying arguments by scheme. In *Proceedings of the 49th Annual Meeting of the ACL: Human Language Technologies-Volume 1*, pages 987–996. ACL, 2011.

[10] B. J. Garssen. Argument schemes. In F. H. van Eemeren, editor, *Crucial concepts in argumentation theory*, pages 81–99. Amsterdam University Press, 2001.

[11] N. Green. Identifying argumentation schemes in genetics research articles. In *Proceedings of the 2nd Workshop on Argumentation Mining*, pages 12–21, Denver, CO, June 2015. ACL.

[12] A. C. Hastings. *A Reformulation of the Modes of Reasoning in Argumentation*. PhD thesis, Northwestern University, 1963.

[13] M. Janier, J. Lawrence, and C. Reed. OVA+: An argument analysis interface. In S. Parsons, N. Oren, C. Reed, and F. Cerutti, editors, *Proceedings of the Fifth International Conference on Computational Models of Argument (COMMA 2014)*, pages 463–464, Pitlochry, 2014. IOS Press.

[14] M. Kienpointner. *Alltagslogik. Struktur and Funktion von Argumentationsmustern [Everyday logic. Structure and functions of specimens of argumentation].* Fromman-Holzboog, 1992.

[15] N. Kokciyan, I. Sassoon, A. Young, M. Chapman, T. Porat, M. Ashworth, V. Curcin, S. Modgil, S. Parsons, and E. Sklar. Towards an argumentation system for supporting patients in self-managing their chronic conditions. In *AAAI Joint Workshop on Health Intelligence (W3PHIAI 2018)*, 2018.

[16] J. Landis and G. Koch. The measurement of observer agreement for categorical data. *Biometrics*, 3:159–174, 1977.

[17] J. Lawrence, F. Bex, C. Reed, and M. Snaith. AIFdb: Infrastructure for the argument web. In *Proceedings of the Fourth COMMA*, pages 515–516, 2012.

[18] J. Lawrence, R. Duthie, K. Budzynska, and C. Reed. Argument Analytics. In *the Sixth International Conference on Computational Models of Argument (COMMA 2016)*, pages 371–378, 2016.

[19] J. Lawrence and C. Reed. AIFdb Corpora. In S. Parsons, N. Oren, C. Reed, and F. Cerutti, editors, *Computational Models of Argument*, pages 465–466. IOS Press, 2014.

[20] J. Lawrence and C. Reed. Combining argument mining techniques. In *Proceedings of the 2nd Workshop on Argumentation Mining*, pages 127–136, Denver, CO, June 2015. ACL.

[21] E. Musi, D. Ghosh, and S. Muresan. Towards feasible guidelines for the annotation of argument schemes. In *Proceedings of the 3rd Workshop on Argumentation Mining*, Berlin, August 2016. ACL.

[22] J. Park and C. Cardie. Identifying appropriate support for propositions in online user comments. In *Proceedings of the First Workshop on Argumentation Mining*, pages 29–38, Baltimore, MD, 2014. ACL.

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

[24] C. Perelman and L. Olbrechts-Tyteca. *The New Rhetoric: A Treatise on Argumentation*. University of Notre Dame Press, 1969.

[25] G. Peters and J. T. Woolley. The American Presidency Project, 1999. Accessed 11 Aug. 2017.

[26] I. Rahwan and G. R. Simari. *Argumentation in artificial intelligence.* Springer, 2009.

[27] C. Reed, S. Wells, G. Rowe, and J. Devereux. AIF+: Dialogue in the argument interchange format. In P. Besnard, S. Doutre, and A. Hunter, editors, *Proceedings of the 2nd International Conference on Computational Models of Argument (COMMA 2008)*, pages 311–323. IOS Press, 2008.

[28] E. Rigotti and S. G. Morasso. Comparing the argumentum model of topics to other contemporary approaches to argument schemes: the procedural and material components. *Argumentation*, 24(4):489–512, 2010.

[29] S. Rubinelli. *Ars Topica: the Classical Technique of Constructing Arguments from Aristotle to Cicero.* Springer, 2009.

[30] P. J. Schellens. *Redelijke argumenten. Een onderzoek naar normen voor kritische lezers [Reasonable arguments. A study of norms for critical readers].* Foris, 1985.

[31] F. H. van Eemeren, B. Garssen, E. C. W. Krabbe, A. F. Snoeck Henkemans, B. Verheij, and J. H. M. Wagemans. *Handbook of argumentation theory.* Springer, 2014.

[32] F. H. van Eemeren and R. Grootendorst. *Argumentation, communication, and fallacies: A pragma-dialectical perspective.* Lawrence Erlbaum Associates, 1992.

[33] F. H. van Eemeren, R. Grootendorst, and T. Kruiger. *Argumentatietheorie [Argumentation theory].* Het Spectrum, 1978.

[34] J. Visser, R. Duthie, J. Lawrence, and C. Reed. Intertextual correspondence for integrating corpora. In *11th edition of Language Resources and Evaluation Conference (LREC)*, pages 1–7, 2018.

[35] J. Wagemans. Constructing a periodic table of arguments. In P. Bondy and L. Benacquista, editors, *Argumentation, Objectivity, and Bias: Proceedings OSSA 11*, pages 1–12. OSSA, 2016.

[36] D. Walton. *Argumentation Schemes for Presumptive Reasoning*. Erlbaum, 1996.

[37] D. Walton. Argument mining by applying argumentation schemes. *Studies in Logic*, 4(1):38–64, 2011.

[38] D. Walton and F. Macagno. A classification system for argumentation schemes. *Argument and Computation*, 6(3):219–245, 2015.

[39] D. Walton, C. Reed, and F. Macagno. *Argumentation Schemes*. Cambridge University Press, 2008.

[40] A. Wyner, J. Schneider, K. Atkinson, and T. Bench-Capon. Semi-automated argumentative analysis of online product reviews. In *Proceedings of COMMA 2012*, pages 43–50, 2012.