A Decision Tree for Annotating Argumentation Scheme Corpora

JOHN LAWRENCE
Centre for Argument Technology, University of Dundee, UK
j.lawrence@dundee.ac.uk

JACKY VISSER
Centre for Argument Technology, University of Dundee, UK
j.visser@dundee.ac.uk

DOUGLAS WALTON
Centre for Research in Reasoning, Argumentation and Rhetoric, the University of Windsor, Canada
dwalton@uwindsor.ca

CHRIS REED
Centre for Argument Technology, University of Dundee, UK
c.a.reed@dundee.ac.uk

Quantitative approaches, necessary for e.g. computational-linguistic methods such as argument mining, require large annotated corpora of argumentative discourse. Publicly available corpora of argumentation schemes often only cover a small selection of example schemes and suffer from low inter-annotator agreement. To address this, we present a heuristic decision tree for the classification of Walton's top-level taxonomy of 60 schemes. An annotation study on 505 arguments resulted in a 97% classification covering 38 schemes (Cohen's κ 0.723).

KEYWORDS: annotation, argument analysis, argumentation schemes, corpora, decision tree, software

1. INTRODUCTION

Data about argumentative practice both informs descriptive approaches to argument and provides a testing ground for normative models. This data can come from the qualitative appraisal of selected examples, but

quantitative approaches, while labour intensive, are gaining traction, motivated by the rise of computational-linguistic methods such as argument mining. Such quantitative approaches require large corpora of argumentative discourse annotated using theories of argumentation.

Argumentation schemes capture the passage of (typically presumptive) inference from a set of premises to a conclusion representing stereotypical patterns of human reasoning. As such, they form a historical descendant of the topics of Aristotle (Aristotle, 1958) and, much like Aristotle's topics, play a valuable role in both the construction and evaluation of arguments. Various attempts have been made to identify and classify schemes and though these sets of schemes overlap, both their granularity and comprehensiveness vary greatly. As a result, annotated corpora of argumentation schemes tend to contain a selection of examples from only one scheme set, with those based on Douglas Walton's typology (Walton, 1996) being the most common.

Despite several proposals to systematise the Walton scheme set by imposing some ordering principle on the typology, to our knowledge, no exhaustive and systematic account currently exists. This absence is reflected in the publicly available argumentation scheme corpora, all of which suffer from a combination of either low inter-annotator agreement, or lack of exhaustive coverage, with, in many cases, only those examples that clearly fit a particular pattern annotated.

In the current paper, we present a decision tree for the classification of Walton's scheme set. Whilst intended primarily as a guide for annotators, the decision tree captures a detailed systematisation of the scheme set, with each of the top-level branches representing divisions into general categories (for example, arguments based on character, or on opinion), before breaking these down further by following a path of simple questions until a definitive scheme classification is reached. To ensure a comprehensive coverage, the decision tree is based on Walton, Reed and Macagno's 2008 book *Argumentation Schemes*, which describes over 60 schemes.

In order to test the applicability of the decision tree to challenging, real-world data, an annotation study was carried out to classify all occurrences of inference relations in an existing argumentative analysis of the first US presidential election debate between Hillary Clinton and Donald Trump. The annotation resulted in substantial inter-annotator agreement. These results suggest that the application of the decision tree to argumentation scheme annotation constitutes a significant improvement to both reliability and breadth of coverage when compared to previous scheme annotation work.

The rest of the paper is structured as follows. In Section 2, we discuss relevant existing annotations of argumentation schemes. In Section 3, we introduce the decision tree heuristic for annotating

argumentation schemes. in Section 4, we describe and evaluate the annotation study. In Sections 5, we discuss ways of further improving the annotation of argumentation schemes, by considering scheme clusters and a systematisation of the Walton scheme set. In Section 6, we conclude the paper.

2. ANNOTATING ARGUMENTATION SCHEMES

The annotation of argumentation schemes comprises the classification of the inferential relations between premises and conclusions of arguments in accordance with a particular typology. While we start from Walton's typology, alternative approaches are also employed for scheme identification: (Green, 2015) presents ten custom argumentation schemes for genetics research articles, (Musi, Ghosh, & Muresan, 2016) explore annotation guidelines on the basis of the Argumentum Model of Topics (Rigotti & Greco, 2019), and (Visser, Lawrence, Wagemans, & Reed, 2019) annotate argumentation schemes on the basis of the Periodic Table of Arguments (Wagemans, 2016).

Existing annotations on the basis of Walton's typology tend to use a restricted set of scheme types, and struggle to obtain replicable results. For example, (Duschl, 2007) initially adopts a selection of nine argumentation schemes described by (Walton, 1996), for his annotation of transcribed middle-school student interviews about science fair projects. Later, however, he collapses several schemes into four more general classes no longer directly related to particular scheme types. This deviation from Walton's typology appears to be motivated by the need to improve annotation agreement. The validation of the annotation method does not account for chance agreement, by only providing percentage-agreement scores (instead of resorting to, e.g., a κ or α metric). Out of a total of 17 texts, the inter-annotator agreement is reported on two as 90% and 84%, without any further detailing of the sampling method.

Similarly, (Song, Heilman, Beigman Klebanov, & Deane, 2014) base their annotation on a modification of Walton's typology, settling on a restricted set of three more general schemes: policy, causal, and sample – resulting in Cohen's κ scores for inter-annotator agreement ranging from 0.364 to 0.848. (Anthony & Kim, 2015) employ a bespoke set of nine coding labels modified from the categories used by (Duschl, 2007) and nine schemes described in a textbook by (Walton, 2006). They do not measure any inter-annotator agreement, opting for a fully open collaborative annotation without any testing of the reliability of the methods.

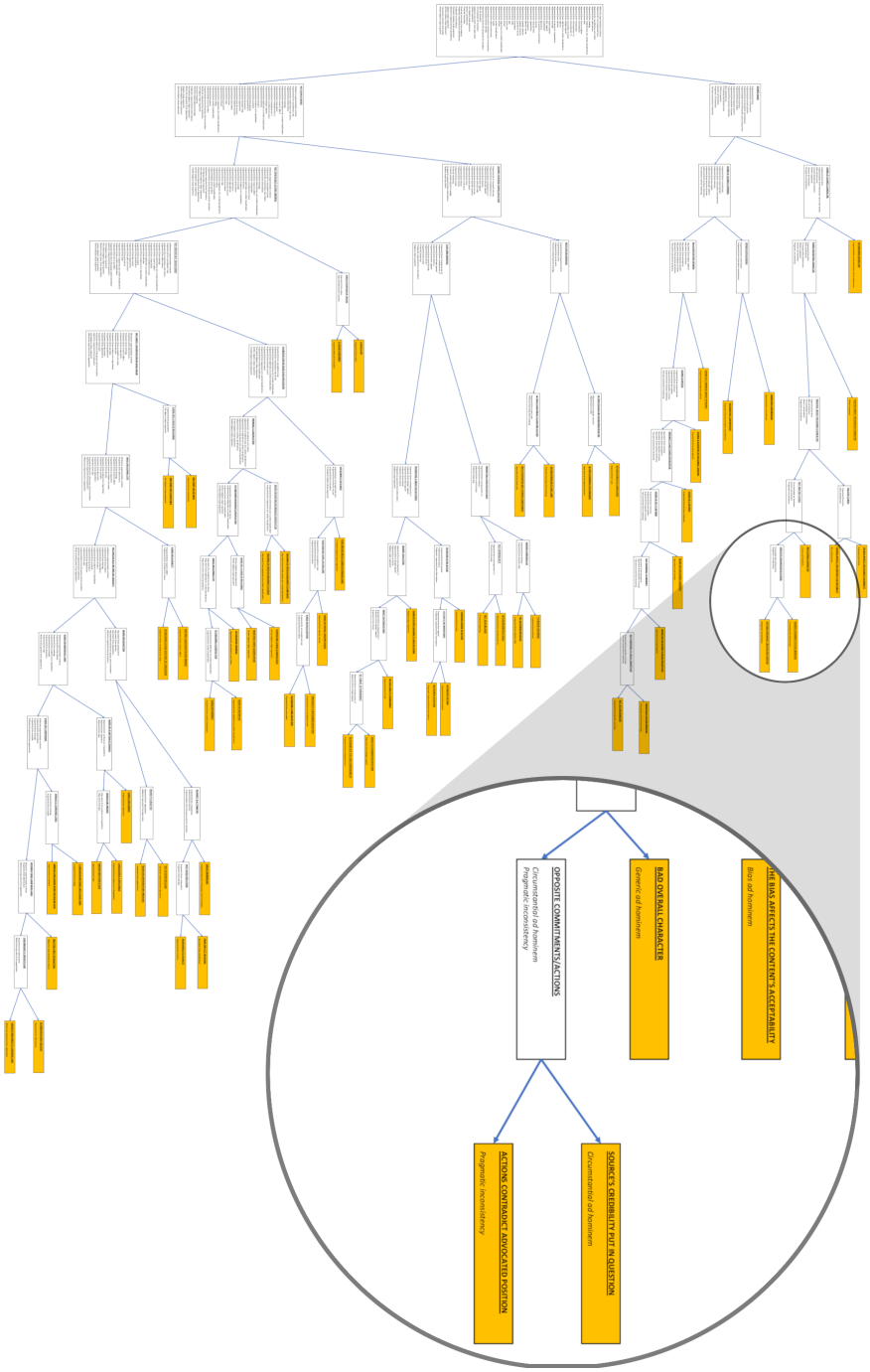

Figure 1 – Decision tree for argumentation scheme annotation

3. ARGUMENTATION SCHEME DECISION TREE

To facilitate the process of annotating the main 60 argumentation schemes described by Walton, Reed and Macagno in their 2008 book *Argumentation Schemes*, we developed a classification decision tree: an indicative heuristic for the annotators, to intuitively support their analytical task. The decision tree interprets the book's Chapter 9 (Walton et al., 2008, pp. 308–346) *A User's Compendium of Schemes* as the starting point of annotation guidelines. The main principle guiding the annotation is the clustering of argumentation schemes on the basis of intuitively clear features recognisable for annotators, resulting in the decision tree of Figure 1. The decision tree constitutes a dichotomous identification tree that leads the analyst through a series of disjunctive choices based on the distinctive features of a 'species' of argumentation scheme to the particular type. Starting from the distinction between source-based and other arguments, each further choice in the tree leads to either a particular argumentation scheme or to a further distinction.

In annotating Example (1), an analyst using the tree follows a sequence of numbered characteristics to identify the argument as an instance of *practical reasoning from analogy*: Argument does not depend on a source's opinion or character; Conclusion is about a course of action; Argument hinges on another motivation for the action [other than its outcome]; Course of action is compared to a similar or alternative action; Action is directly compared to another.

> (1) Hillary 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.

Figure 1 (available online at: http://arg.tech/~john/scheme-tree.png) visualises the decision procedure, with each leaf representing an argumentation scheme label, and all internal nodes showing clusters of schemes that share particular characteristic properties. For each binary decision point, the tree branches into two, thus leading the annotator from the full set of schemes, through their binary choices, to one (and only one) leaf – i.e. an argumentation scheme classification.

4. ANNOTATION STUDY

In order to test the applicability of the decision tree to challenging, real-world data, an annotation study was carried out to classify all occurrences of inference relations in US2016G1tv (Visser et al., 2019), an existing argumentative analysis of the first election debate (26

September 2016, Hempstead, NY) between Hillary Clinton and Donald Trump. The US2016G1tv corpus (stored in AIFdb (Lawrence et al., 2012), and available online at corpora.aifdb.org/US2016G1tv) is annotated on the basis of Inference Anchoring Theory (IAT) (Reed & Budzynska, 2011), resulting in an Argument Interchange Format (AIF) (Chesñevar et al., 2006) compliant corpus.

Two annotators used the argumentation scheme decision tree to classify 55% of the 505 inferential relations within the corpus; for example classifying Example (1) as an instance of the argumentation scheme *practical reasoning from analogy*. The two annotations resulted in an overlapping sample of 10.2% of the corpus annotated by both annotators. For these annotations a Cohen's κ (Cohen, 1960) of 0.723 was achieved; well within the category of 'substantial agreement' (Landis & Koch, 1977). Some classes of argumentation scheme turned out to be particularly difficult to distinguish; e.g., Example (2) was classified by one annotator as *practical reasoning*, related to promoting goals, and by the other as *argument from values*, related to promoting values.

> (2) 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 argumentation schemes are collected in the US2016G1tvWALTON corpus – available online at corpora.aifdb.org/US2016G1tvWALTON. Of the 505 inferences in the original US2016G1tv corpus, a total of 491 are annotated with one of the 60 argumentation scheme types in Walton's classification, leaving only 14 as unclassified *default inference*. The most common scheme, by some margin, is *argument from example*. The *argument from expert opinion* scheme, an often used example, is remarkably rare with only three occurrences. Full results of the annotation and the corpus are discussed by Visser et al. (2018).

5. PRINCIPLES OF ARGUMENTATION SCHEME CLASSIFICATION

The annotation study on the US2016G1tv corpus using the decision tree resulted in substantial inter-annotator agreement. The argumentation schemes in the decision tree are organised according to their distinctive features allowing annotators to distinguish between them. To further improve the decision tree, we aim to take into account the guiding principles underlying the Waltonian taxonomy of argumentation schemes, and the possible clustering of schemes on that basis.

The classification of argumentation schemes should not be regarded as a completed structure, but as a work in progress that is continually being subject to readjustment and refinement as the concepts defining the schemes are formulated in a more precise way and applied to new examples. We will explain the research procedure of improving a classification system of schemes as a process of continuing adjustment between collecting data, sharpening criteria that enable coders to identify a scheme, and used to refine the typology to assist the continuing collection of data.

The 2008 classification system (Walton et al., 2008, pp. 349-350) divided schemes into three general categories, reasoning, source-based arguments and applying rules to cases. Under reasoning five subcategories were distinguished: deductive reasoning, inductive reasoning, practical reasoning, and abductive reasoning. Under the general heading of source-based arguments, four schemes were listed: arguments from position to know, arguments from commitment, arguments attacking personal credibility, and arguments from popular acceptance. The third general category was called applying rules to cases. It had four subcategories: arguments based on cases, defeasible rule-based arguments, verbal classification arguments and chained arguments connecting rules in cases. Each of these second-level types of schemes contained categories at a finer level of granularity. These third level schemes include many of the schemes that are so highly familiar to researchers on argumentation. For example, the third category under source-based arguments contains the following three schemes: argument from allegation of bias, poisoning the well by alleging group bias and *ad hominem* arguments.

The annotators in our annotation study made use of chapter 9, the *user's compendium of schemes* (Walton et al., 2008, pp. 308-346), as guidelines to build a classification decision tree intended to be used as an annotation heuristic. In this heuristic, the top level branches represent divisions and the general categories, while the lower branches break these categories down further by following a path of binary questions. As each question is answered the user is directed down the tree until a definitive scheme classification is arrived at. However, in chapter 10 there was given a proposed classification system for argumentation schemes (Walton et al., 2008, pp. 349-350). One might wonder what the relationship is between this early classification system and the classification decision tree presently offered as an annotation heuristic. One might also wonder whether the 2008 classification system has changed over the ten year interval in the continuing research on schemes classification systems. Finally one might wonder about the current state of this research. This section is designed to answer those questions.

5.1 Clusters of argumentation schemes

It is important to be aware, as stated (Walton et al., 2008, p. 348) that because of the difficulty of defining the concepts that any classification system of schemes has to be based on (such concepts as knowledge, causation, threat, and so forth), any attempt to classify schemes faces conceptual difficulties in adequately defining the contested concepts used at the top levels of the tree structure. For this reason readers were warned that the 2008 system of classifying schemes was to be regarded as a provisional hypothesis that should be subject to improvement as further empirical and analytical work on schemes classification continues. This warning is especially important now, because in the ten year interval the explosion of research on argument mining has raised many fine-grained questions about how particular groups of schemes should be fitted together into the larger picture of any general classification system.

Some subsequent work (Walton & Macagno, 2016) presented a survey of the literature on scheme classification, as well as outlining how the 2008 system needs to be modified in order to accommodate current research in artificial intelligence and computational linguistics on argument mining. In the 2016 paper, it was shown how the procedure of developing and using classification systems can only move forward by combining two approaches. One of these is a top-down approach that begins with concepts formulated at a high level of abstraction, then moves to particular types of schemes that fit under these general categories, and then finally moves to schemes representing the types of arguments we are already so widely familiar with. But at the same time, as research on argument mining continues, it is also necessary to have a bottom-up approach (Walton, 2012) that begins with real examples of arguments at the ground level of cases that distinguish in a very particular way between subtypes of a given scheme. What happens at this bottom-up level is that so-called clusters of schemes are fitted together into larger groups, and then these groups have to be fitted into more general classifications of schemes.

To get a general idea of how clusters of schemes fit into an encompassing schemes classification system, it is useful to examine a graph structure representing a typical example showing how a scheme classification system can be represented visually as a graph. Such a graph is shown in Figure 2. The familiar types of arguments associated with schemes, such as “argument from expert opinion”, are shown in the rectangles with rounded corners. Other categories useful for classifying schemes, such as “source-dependent arguments” are shown in the rectangles with sharp corners.

Two examples of clusters of schemes are indicated by the darkened borders of the rectangles and the arrows in the components of the cluster. The cluster displayed on the right depicts the various kinds of arguments that come under the general category of practical reasoning. Practical reasoning is a distinctive type of argument in its own right, and has its own scheme, but it also as subspecies such as instrumental practical reasoning and value-based practical reasoning. The cluster of schemes under the general heading of *ad hominem* arguments is displayed at the left of Figure 2. This cluster is shown as incomplete. Under the general heading of *ad hominem* arguments some schemes are shown such as the direct *ad hominem* argument type, sometimes called the abusive *ad hominem* in the logic textbooks, and the circumstantial *ad hominem* type. However, as is well known in the literature, there are many other types of *ad hominem* arguments that are not shown here (Walton et al., 2008, p. 352). All that is shown is an elliptical node at the bottom left indicating that there are other *ad hominem* arguments that need to be classified within this cluster. This particular graph is not meant to represent Walton's classification system in a finished or comprehensive form. It is merely an example meant to show what clusters of schemes look like and how the clusters can fit into a more comprehensive classification system.

Note that the graph in Figure 2 is meant to be only a fragment of a larger graph which could include other categories of kinds of arguments that are not defeasible, such as deductive *modus ponens*. Further note that the partial classification system is also incomplete at the bottom level. For example, some types of *ad hominem* arguments are classified at the left of the graph, but the elliptical node, other *ad hominem* arguments, indicates that further sub classifications are possible. For example, in (Walton et al., 2008, p. 352) a graph structure is visually presented that displays seven particular types of *ad hominem* arguments, including the poisoning the well type, the guilt by association type, the *tu quoque* type and the circumstantial *ad hominem* type, subsumed under the more general categories, such as argument from inconsistent commitment and the ethotic or personal type of *ad hominem* argument which is a direct attack on the arguer's character in order to discredit his or her argument.

As shown in Figure 2, argument from precedent combines with the basic slippery slope type of argument to produce a species of slippery slope argument called the precedent type of slippery slope argument. To explain how this works, let us look at the scheme for the basic slippery slope argument.

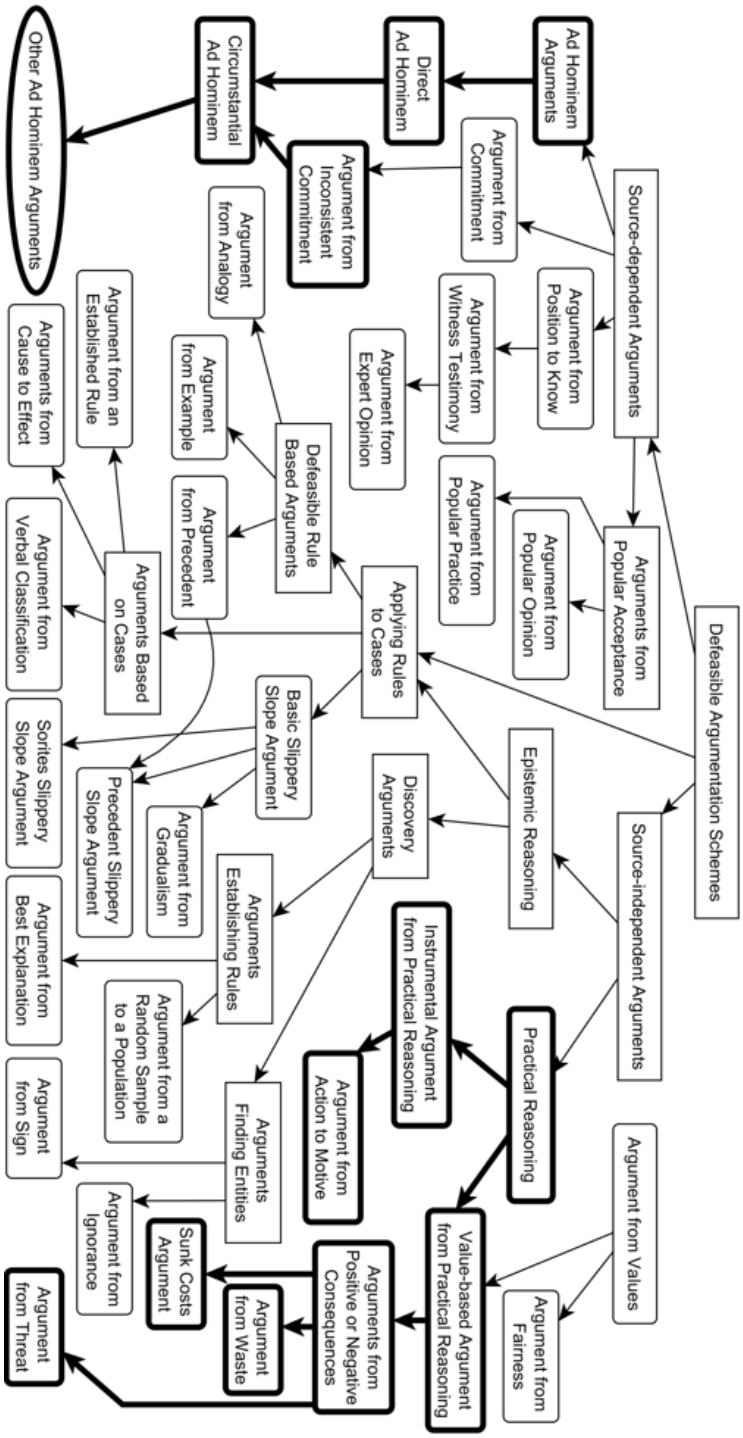

Figure 2 – Graph representing a fragment of a classification system

Initial Premise: An agent α is considering carrying out an action A0.
Sequential Premise: Carrying out A0 would lead to A1, which would in turn lead to carrying out A2, and so forth, through a sequence A2, ..., Ax, ..., Ay, ..., An.
Indeterminacy Premise: There is a sequence A0, A1, A2, ..., Ax, ..., Ay, ..., An that contains a subsequence Ax, ..., Ay called the gray zone where x and y are indeterminate points.
Control Premise: α has control over whether to stop carrying out the actions in the sequence until α reaches some indeterminate point in the gray zone Ax, ..., Ay.
Loss of Control Premise: Once α reaches the indeterminate point in the gray zone Ax, . . . , Ay, α will lose control and will be compelled to keep carrying out actions until she reaches An.
Catastrophic Outcome Premise: An is a catastrophic outcome that should be avoided if possible.
Conclusion: A0 should not be brought about (Walton, 2015, p. 288).

There are various types of slippery slope argument that can be built by extending the basic scheme, and one of these, the precedent type of slippery slope argument, which generates a sequence whereby one case is a precedent for a second one, and the second one is a precedent for a third one, and so forth. As shown in Figure 2, the precedent slippery slope argument combines argument from precedent with the basic slippery slope type of argument. This means that the precedent slippery slope argument is formed as a cluster from other types of arguments, as shown in Figure 3.

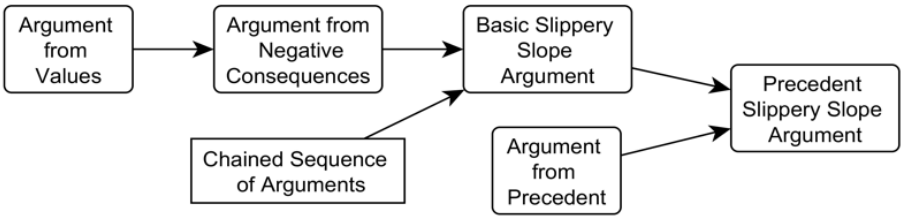

Figure 3 – A slippery slope cluster

Being aware of how this cluster of arguments is formed is helpful for enabling annotators to distinguish between a precedent slippery slope argument and a run-of-the-mill argument from precedent that should not be classified as a slippery slope argument.

5.2 Identification conditions of argumentation schemes

A central practical problem inherent in existing corpus-linguistic work on argumentation schemes is that the annotators lack enough specific guidance on how to decide whether an argument found in a real natural language text can properly be said to fit a particular scheme or not. An early study which used schemes to classify kinds of arguments put forward by candidates in a provincial election in Canada (Hansen & Walton, 2013) classified 256 arguments using 14 schemes and a category called "none of the above". A group of six annotators, two of them experts in argumentation theory, collected arguments found in newspaper articles reporting arguments commenting on issues being debated in the campaign. The difficulty they encountered was that the four non-expert annotators, as well as the two experts in some instances, found it difficult to classify arguments in some instances because of the open texture of key terms used in the schemes. For example, annotators found it difficult to differentiate between a circumstantial *ad hominem* argument and an argument from inconsistency, a species of argument from commitment. Both kinds of arguments allege that an opposed arguer has put forward an argument, or part of an argument, that conflicts with a prior commitment of the arguer. But only the circumstantial *ad hominem* type of argument uses the commitment to derive a secondary inference to the conclusion that the arguer has exhibited some ethical defect of character, indicated by the use of a keyword, such as 'hypocrite' or 'liar'.

The solution to this problem recommended in (Walton, 2012) was to devise a set of so-called identification conditions that can be used to offer annotators additional guidance on whether a particular scheme fits a particular case or not. There were 24 of these identification conditions formulated by Walton (2012, pp. 49-56). A current project is to refine these conditions to make them more precise and easier to use. To give the reader an idea of what these kinds of conditions are like, here are six of the reformulated ones.

> (IC1) Argument from Inconsistent Commitments: (1) There has to be evidence from the way a has put A forward as a claim (assertion) in a dialogue exchange to indicate that a is committed to A, and (2) there has to be evidence from the way a has put A forward as a claim (assertion) in a dialogue exchange or the circumstances of a case to indicate that a is committed to not-A. (4) The conclusion is drawn on the basis of (1) and (2) that a is committed to (A and not-A).

(IC2) Direct *Ad Hominem* Argument: there has to be (1) not only an attack on the arguer's ethical character (ethos), but (2) this attack has to be used to discredit the arguer's credibility (personal trustworthiness as a source), (3) in order to try to defeat his argument.

(IC3) Circumstantial *Ad Hominem* Argument: there has to be (1) an attack on the arguer's ethical character, but (2) this attack has to be based on an alleged inconsistency among the arguer's commitments (3) which has to be used to discredit the arguer's credibility (personal trustworthiness as a source), and (4) the premises (1), (2) and (3) have to be put forward to try to defeat his argument.

(IC4) Argument from Values: (1) The audience to whom the argument is addressed is thought by the arguer to hold a positive (or negative) value with respect to a proposition, and (2) appeal to this value is used by the arguer as a means of supporting (or attacking) the commitment of the audience to some goal or policy he advocates.

(IC5) Argument from Positive Consequences: (1) A proposal in favour of carrying out action A is put forward, (2) pro A and con A arguments are being considered, (3) the claim is made that A, if carried out, with have positive consequences, (4) use of the term 'positive' means that the action is claimed to have positive value for the audience the argument is directed to, and (5) on this basis it is claimed that the audience should support the proposal to carry out A.

(IC6) Argument from Negative Consequences: (1) A proposal against carrying out action A is put forward, (2) pro A and con A arguments are being considered, (3) the claim is made that A, if carried out, with have negative consequences, (4) use of the term 'negative' means that the action is claimed to have negative value for the audience the argument is directed to, and (5) on this basis it is claimed that the audience should decline to support the proposal to carry out A.

The other identification conditions have the same general format, except that some of them are more complex forms of argument that contain simpler forms of argument, such as the seven types represented above. One scheme can be shown to contain another scheme, using the identification conditions, and the structure of this relationship can be visually displayed as a graph. By this means, for example, a complex scheme, such as the slippery slope argument, can be shown to contain another simpler form of argument, such as argument from negative consequences, embedded within its structure.

5.3 Identification conditions applied to clusters

This vagueness and ambiguity about how to more precisely define these three types of arguments pervaded Hansen and Walton's (2013) election project because we found numerous examples of argument from inconsistent commitments, some of them arguably being *ad hominem* arguments, some arguably not. A first step toward resolving the problem is to treat the circumstantial *ad hominem* argument (CA) as being a subspecies of the wider category of argument from inconsistent commitments (IC), which is in turn a subspecies of argument from commitment (AC) and the direct *ad hominem* argument (DA) – see Figure 4.

This way it doesn't matter so much if there was a difficulty in trying to make a decision based on the text of discourse of the example on which category the argument should be placed into. If there is doubt whether the argument is really meant to be a personal attack, then we can classify in into the more general category of argument from inconsistent commitment. If there is no doubt, we can classify it into the circumstantial *ad hominem* category.

To cite another example, elements (5) and (6) of the scheme for argument from values (AV) are carried over and incorporated into the structure for argument from value-based practical reasoning (VBPR) – see Figure 5. Similarly, on the left, elements of the scheme for instrumental practical reasoning (IPR) are incorporated into the scheme for value-based practical reasoning (VBPR).

Using this approach, the structure that holds a cluster together is derived from the identification conditions for a particular scheme, showing how that scheme is related to other neighbouring schemes to form a cluster.

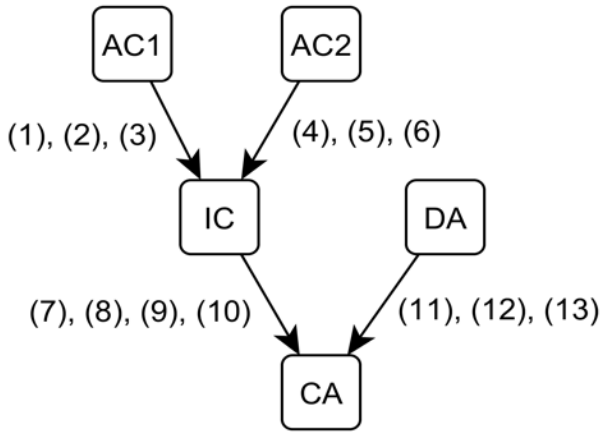

Figure 4 – Part of the *ad hominem* cluster

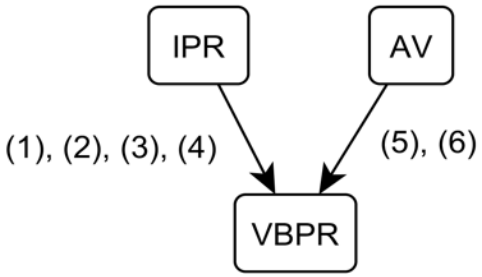

Figure 5 – The value-based practical reasoning cluster

5.4 Leveraging clusters in argumentation scheme classification

Once this method is used to form several clusters of schemes, the clusters can be all put together into a larger classification graph of the kind illustrated in Figure 2. But it is not certain what will happen when this is done on a very large scale. It may be that there are connections between two different clusters in a classification graph, or even connections between several different clusters. At the state this kind of research has reached now, none of this has yet been explored. As more and more examples of arguments fitting a given structure are collected and classified, the clusters can be expected to grow in complexity. For instance, as more and more examples of *ad hominem* arguments are collected and analysed, new types of *ad hominem* arguments are likely to be discovered. Part of this line of discovery will be the formulation of identification conditions for each of these new schemes.

This procedure is circular in nature, but in this instance the circularity is not evidence that a fallacy has been committed. It shows that the activity of collecting data from natural language corpora, and using that to refine the classification system, is a defeasible but scientific way of collecting evidence for or against a hypothesis and improving it by feedback. The recursive application of the procedure improves the accuracy of the formulation of the schemes.

It is shown how the production of an evolving taxonomy that takes increasing sophistication of sub-schemes into account through the use of identification conditions. Essentially the procedure consists in the refinement and evolution of the taxonomy as it is tested against the data by being continuously applied to real examples of naturally occurring arguments.

At the higher levels, the identification conditions can be used to sharpen the general concepts, providing precise and definitions of these terms. By this means, both tools, the identification conditions and the clusters they generate, can be applied to real examples in order to

improve a given classification system for schemes, making it both more precise and more applicable to identifying types of arguments found in discourse.

6. CONCLUSION

Theory-driven applications of computational models of argument, and empirically oriented work alike, rely on data about the actual use of argumentation in practice. The availability of large, reliable, and representative datasets of argumentation scheme usage is essential both to the empirical study of such schemes, and to the development of automated classifiers and argument mining techniques (Budzynska & Villata, 2017). In this paper, we present a decision tree heuristic for annotating argumentation schemes. The decision tree supports annotation which is both comprehensive in the range of schemes it covers, and reliable in the results obtained. Finally, we have considered future directions for improving the decision tree, taking into account the guiding principles underlying the Waltonian taxonomy of argumentation schemes, and the possible clustering of schemes on that basis.

ACKNOWLEDGEMENTS: This research was supported in part by the Engineering and Physical Sciences Research Council (EPSRC) in the United Kingdom under grant EP/N014871/1, and in part by Social Sciences and Humanities Research Council of Canada Insight Grant 435-2012-0104.

REFERENCES

Anthony, R., & Kim, M. (2015). Challenges and remedies for identifying and classifying argumentation schemes. *Argumentation*, *29*(1), 81–113.

Aristotle. (1958). Topics (W. D. Ross, Trans.). Oxford University Press.

Budzynska, K., & Villata, S. (2017). Processing argumentation in natural language texts. In P. Baroni, D. Gabbay, M. Giacomin, & L. van der Torre (Eds.), *Handbook of formal argumentation*. College Publications.

Chesñevar, C., Modgil, S., Rahwan, I., Reed, C., Simari, G., South, M., ... others (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.

Duschl, R. A. (2007). Quality argumentation and epistemic criteria. In *Argumentation in science education* (pp. 159–175). Springer.

Gordon, T. F., Prakken, H., & Walton, D. (2007). The carneades model of argument and burden of proof. *Artificial Intelligence, 171*(10), 875–896.

Green, N. (2015, June). Identifying argumentation schemes in genetics research articles. In *Proceedings of the 2nd workshop on argumentation mining* (pp. 12–21). Denver, CO: ACL.

Hansen, H. V., & Walton, D. (2013). Argument kinds and argument roles in the Ontario provincial election. *Journal of Argumentation in Context, 2*(2), 226–258.

Janier, M., Lawrence, J., & Reed, C. (2014). OVA+: An argument analysis interface. In S. Parsons, N. Oren, C. Reed, & F. Cerutti (Eds.), *Proceedings of the fifth international conference on computational models of argument (comma 2014)* (pp. 463–464). Pitlochry: IOS Press.

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

Lawrence, J., Bex, F., Reed, C., & Snaith, M. (2012). AIFdb: Infrastructure for the argument web. In *Proceedings of the fourth international conference on computational models of argument (comma 2012)* (pp. 515–516). Vienna: IOS Press.

Musi, E., Ghosh, D., & Muresan, S. (2016, August). Towards feasible guidelines for the annotation of argument schemes. In *Proceedings of the 3rd workshop on argumentation mining*. Berlin: ACL.

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)* (pp. 1633–1645). SicSat.

Reed, C., & Rowe, G. (2004). Araucaria: Software for argument analysis, diagramming and representation*. International Journal on Artificial Intelligence Tools, 13*(4), 961–980.

Rigotti, E., & Greco, S. (2019). *Inference in argumentation: A topics-based approach to argument schemes*. Springer.

Song, Y., Heilman, M., Beigman Klebanov, B., & Deane, P. (2014). Applying argumentation schemes for essay scoring. In *Proceedings of the first workshop on argumentation mining* (pp. 69–78). Association for Computational Linguistics. doi: 10.3115/v1/W14-2110

van Gelder, T. (2007). The rationale for rationale. *Law, probability and risk, 6*(1-4), 23–42.

Visser, J., Konat, B., Duthie, R., Koszowy, M., Budzynska, K., & Reed, C. (2019). Argumentation in the 2016 US presidential elections: annotated corpora of television debates and social media reaction. *Language Resources and Evaluation*. doi: 10.1007/s10579-019-09446-8

Visser, J., Lawrence, J., Wagemans, J., & Reed, C. (2018). 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. doi: https://doi.org/10.3233/978-1-61499-906-5-313

Visser, J., Lawrence, J., Wagemans, J., & Reed, C. (2019). An annotated corpus of argument schemes in us election debates. In B. Garssen, D. Godden, G.

Mitchell, & J. Wagemans (Eds.), *Proceedings of the 9th conference of the international society for the study of argumentation (issa)* (pp. 1101–1111). Amsterdam, The Netherlands: Sic Sat.

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)* (pp. 1–12). OSSA.

Walton, D. (1996). *Argumentation schemes for presumptive reasoning*. Mahwah, New Jersey: Lawrence Erlbaum Associates.

Walton, D. (2006). *Fundamentals of critical argumentation*. Cambridge University Press.

Walton, D. (2012). Using argumentation schemes for argument extraction: A bottom-up method. *International Journal of Cognitive Informatics and Natural Intelligence, 6*(3), 33–61.

Walton, D. (2015). The basic slippery slope argument. *Informal Logic, 35*(3), 273–311.

Walton, D., & Macagno, F. (2016). A classification system for argumentation schemes. *Argument and Computation, 6*(3), 219–245.

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