

ARGUMENT RESEARCH CORPUS

Joel Katzav, Chris Reed and Glenn Rowe
Department of Applied Computing
University of Dundee DD1 4HN

1. Introduction

Argumentation is a relatively recent field of research. It aims to analyse, describe, and evaluate real-world, natural language arguments. It also features as a topic in many undergraduate syllabi, where it aims to teach students both to think critically about the arguments of others, and to create better, more measured arguments of their own. Both research into argumentation and teaching critical thinking could benefit from resources that are, at present, unavailable. In particular, there are no text corpora that are dedicated to capturing real world arguments. In what follows, we discuss the need for such a corpus and our ongoing work towards creating one. Our aim is to construct an argumentation text corpus that will serve both as a resource for the community of researchers and teachers working on argumentation, and as a test case for the development of similar corpora in the future.

2. Argumentation Theory

Informal logicians and argumentation theorists tend to view themselves as reacting against post-Fregean formal logic. Taking their cue from the work of Toulmin (1958) in philosophy, and of Perelman and Olbrechts-Tyteca (1969) in rhetoric, they argue that formal logic is a poor choice for representing and characterizing natural – i.e. real-world - language and argument. As a result, they have attempted to develop an empirically driven understanding of everyday argumentation. This attempt has proceeded along three main paths. To begin with, having concluded that it is implausible to identify everyday arguments with deductive proofs, argumentation theorists have developed new theories about what everyday arguments are. For example, one recent influential view is the pragmatic conception of argument. On this view, an argument is a speech act intended to convince a given audience (see (Walton, 1990)).

Parallel to the development of new conceptions of arguments, argumentation theorists have noted that discourse often proceeds by using certain stereotypical forms or types of argument. Thus, argumentation theorists have attempted to develop theories of argument form or type. Here writers have returned to Aristotle's *Topics* in an attempt to outline and classify the alternative forms of argument that everyday argument takes.

Finally, with the move away from a purely deductive conception of arguments, and thus with the development of the view that the formal criterion of soundness is inadequate when it comes to evaluating everyday argumentation, there has been an attempt to develop a new theory of good arguments and of good forms or types of argument. Some have also proposed a theory of fallacious arguments, that is to say a theory of forms or types of faulty argumentation.

It should be emphasised that the recent interest in everyday argumentation has been driven largely by teaching needs. Teaching critical thinking in a way that is relevant to everyday problems and argumentative practices has been one of the primary driving forces behind the attempt to develop an adequate theory of argumentation, argument types and argument evaluation for everyday arguments. Thus, alongside the development of argumentation theory, there has also been considerable interest in developing techniques for teaching argumentation skills. Among these, for example, there is the technique of diagramming arguments as a

series of nodes linked by relations of support, rather than formalizing arguments using the tools of propositional or predicate calculus.

3. Uses of Argument Corpora

The renewed interest in argumentation suggests the possibility of using specially designed corpora of arguments both for research and for teaching. In particular, online corpora could be useful tools for research into a wide variety of aspects of argumentation. Here are some examples of research-oriented uses of argument corpora:

- (1) Comparative research into argument usage in a discourse field over time. For example, comparison of newspaper argumentative practices at different times. Thus, one could investigate whether a given newspaper has used the same types of argument at different times.
- (2) Comparative research into argument use and evaluation across discourse fields. For example, comparison of argumentative practices in newspapers and legislative bodies, or of argumentative practices in everyday discourse and science. Indeed, it would be of particular interest to use argument corpora in order to determine whether there are types of argument that are used across all fields of discourse.
- (3) Comparative research into different strategies for interpreting arguments. For example, comparison of different classification schemes of arguments or of different ways of determining which implicit premises arguments use. Thus, one might compare Rhetorical classifications with semantic classifications to see which, if any, semantic properties correspond to a given rhetorical type of argument.
- (4) Developing descriptive argument typologies or classification schemes for existing discourse fields. An argument corpus could, for example, be used to evaluate the exhaustiveness of a classification scheme.
- (5) Critical application of normative models of argumentation to extant discourse fields. For example, one could apply a normative model of argumentation in order to evaluate the rationality of argumentative practices in legislative bodies.

Argument corpora could also facilitate decision-making and teaching. Here are some of the primary ways in which this might be done:

- (1) Facilitation of decision-making within communities through the use of 'good practice' argument corpora. By 'good practice' argument corpora we mean corpora that contain paradigmatic cases of good, realistic arguments. The idea, then, would be to assist the decision making processes of communities by familiarising them with corpora that contain paradigmatic instances of 'good' reasoning about the kinds of problem they face.
- (2) Facilitating the learning of critical thinking by students through the use of 'good practice' argument corpora.
- (3) Facilitating the learning of critical thinking by students through the use of corpora that contain typical modes of fallacious reasoning.
- (4) Facilitating the learning of critical thinking by students through the use of corpora as sources of data upon which they can practice their analytical skills.
- (5) Automating the process of marking students' work in critical thinking. For example, it would be easy to develop software that evaluates students' performance on multiple answer questions about how arguments within an online corpus should be analysed.

Unfortunately, as things stand, the use of text corpora in argumentation is limited. For the most part, and despite the empirical orientation of recent argumentation research, our views of everyday argumentation rely largely on anecdotal evidence. Indeed, to our knowledge, there is only one corpus, namely the Free Britain Corpus developed at the University of Durham, that has been created, at least in part, in order to study arguments. This

corpus, however, focuses only on discourse about UK integration into the EU, and does not include the analysis or reconstruction of arguments.¹

One could, of course, use text corpora that have not been designed specifically for research into argumentation for such research. However, as we will see, doing so leaves much to be desired.

4. Argument Corpus Description, Advantages and Disadvantages

What is needed, then, is an argumentation corpus specifically designed for a wide variety of uses in argumentation research and teaching. We are in the process of creating such a corpus. The corpus will capture arguments that are accessible to the general public, as opposed to arguments from specialist fields of discourse such as physics or biology. Initially, it will be comprised of 20 sets of 30 analysed sections of text. Each analysed section of text will contain a chain of one or more arguments. Each set of analysed texts will be drawn from a different online source. The sources are from a number of countries and include online new services, parliamentary debates, court judgements, discussion boards and so on.

Each chain of arguments will be presented in its 'raw' form along with information about the source from which it was taken, and each argument will also appear in one or more reconstructed and marked-up versions. Specifically, its implicit premises will be made explicit and it will be classified as to what type of argument it is. So too, an argument's position within a chain of arguments will be made explicit.

There are a number of advantages to a relatively small argument corpus such as the one we propose. Let us spell out three of these. To begin with, such a corpus offers analysed arguments alongside raw text. By contrast, using an existing corpus that is not dedicated to argumentation merely supplies raw text.

In addition, if generic corpora are used for research into argumentation, the examination of a substantial quantity of text is required in order to select material. Given the limited resources available to researchers in argumentation, this would force the use of automated text scanning methods. However, such methods are, at least at the moment, unreliable at detecting the presence of arguments and types of argument within a text as they use a very narrow range of syntactical cues in order to do so. By creating a small argumentation corpus, we are able to use trained researchers to develop a ready resource of appropriately selected arguments and types of argument, and so enable other argumentation theorists to avoid the unreliability of scanning.

A final advantage of a small corpus dedicated to argumentation consists in the appropriateness of the sources used in its creation. In contrast to these sources, the texts of large, general-purpose corpora are not selected in order to give a fair representation of everyday arguments used in English.

Some worries do arise about relying on a small manually created corpus. Specifically, there are worries about bias in manual argument selection and reconstruction, and about the adequacy of the sample of arguments used. In order to deal with these worries, we are using sources that can be accessed, and so reassessed, by other researchers at a later date, thus allowing a critical evaluation of our selection and analyses.

Similarly, we leave open the possibility that the corpus will, in the future, be expanded so as to include additional material, including material from other researchers and additional marked-up versions of the original argument set, that is to say additional marked-up versions that represent alternative views about how arguments should be classified and reconstructed.

Finally, we have endeavoured to reduce individual bias in the selection and analysis of arguments by setting out, in advance, guidelines for the extraction and reconstruction of arguments. Partly, these guidelines have been determined by a theory of the nature of arguments that we have developed, and by a corresponding classification system for arguments. Our theory of arguments assists us in being consistent in our determinations of which arguments a text contains, and our classification system for arguments assists us in being consistent in analysing or reconstructing arguments by guiding the making of implicit premises explicit.

We are, however, aware that since the creation of corpora that are dedicated to argumentation is as yet untried, we will not be able to deal adequately with all the worries that arise in doing so. Thus, we have as a secondary aim of creating our corpus the learning of lessons in how to construct such corpora. We propose to use our corpus in order to carry out research into argumentation, and in doing so to evaluate the corpus. Once we

¹ The Free Britain Corpus can be found at <http://www.dur.ac.uk/SMEL/depts/german/euro-arc.htm>. For some of the research that relates to this corpus see (Musolff, Good, Wittlinger and Points, 2001).

have finished creating our corpus, we plan to use it to evaluate the success of the classification scheme used to mark-up arguments as to their type. Leaving aside what we will thus learn about arguments and their classification, we hope to be able to learn something about the limitations of our corpus as a tool for research into argumentation and about how we might go about producing better corpora in the future.

5. Project Outline

We now turn to discuss the stages of the creation of our argument research corpus and the extent to which these stages have been completed. The initial stage of this project has been completed and proceeded along two tracks. On the one hand, it involved developing a theory of arguments and, in light of this theory, a classification system or scheme for arguments that will be used to mark-up the raw arguments in our corpus. This theory of arguments and the corresponding classification system is discussed elsewhere (Katzav and Reed, 2003).

The initial stage of the project also involved selecting appropriate online sources of argument. As already stated, we are using 20 sources of arguments, each of which is in English, accessible to other researchers at later dates and understandable without specialist education.

In addition, the sources we are using have been selected in light of the aim that they comprise a heterogeneous set, that is to say they have been selected so that they allow us to collect arguments that are representative both of the wide variety of argument types that are used in general argumentation in English and of the relative frequencies in which the argument types in question are used. We aimed to ensure a heterogeneous set partly by initially examining a broad set of forty-five sources of argument from which the twenty final sources that are being used were selected.²

Having selected the appropriate sources, we are proceeding with the second stage of our project. This stage consists in the selection and mark-up of arguments from each source. We plan to collect arguments over a period of approximately one year. In order to minimise bias, we require that the first argument encountered in examining a source on a single occasion be selected. And, for the same reason, we try to ensure that each time we access a source, a different section of it is examined.

The reconstruction and mark-up of arguments is proceeding alongside the collection of arguments. When an argument is selected, we determine its type, make its implicit premises explicit in light of this determination and mark it up accordingly. Information about the source from which an argument is taken is stored together with it, and comments about the argument reconstruction are entered alongside this information. So too, information about the adequacy of the system of classification used in the mark-up is collected. In particular, we evaluate the completeness of this classification system and expand it where necessary.

The software used to capture arguments both in their raw and in their analysed form is called *Araucaria*, and has been designed by us specifically in order to allow the user to mark-up arguments as he or she pleases. It supports the use of different conceptions of arguments and classification systems, and uses the standard method of representing arguments as a series of nodes connected by relations of support. It also allows each analysed argument to be stored in a database using our Argumentation Mark-up Language. This language is implemented using XML, thus facilitating searching through the corpus according to argument type, argument content and argument structure.³

With the completion of the corpus, we will arrive at the final stage of our project. In this stage, recall, we hope to learn both lessons for future argument corpus building and lessons about what can be learned from argument corpora. We will also develop a WWW interface to the corpus, thus facilitating analysis and expansion by other researchers.⁴ As this stage of the project has not yet begun, we cannot offer systematic or final conclusions. Nevertheless, we will outline a number of tentative, anecdotal results of our work.

To begin with, it appears that a corpus such as ours will allow us to learn something substantial about how, and to what purpose, arguments are used in non-specialist contexts. Thus, for example, our work so far suggests that there is almost no appeal either to statistical data or to any kind of experimental evidence in the arguments

² See Appendix I for a list of the sources from which the final selection was made.

³ Araucaria, along with supporting documentation, can be downloaded at <http://www.computing.dundee.ac.uk/staff/creed/araucaria/index.html>.

⁴ The evolving corpus is, it should be noted, already accessible online via Araucaria.

that are found in newspapers, parliamentary debates and court judgements. To offer a second example, the types of argument that are found, and the relative frequency with which these types are found, in newspapers, parliamentary debates and court judgements seem to be similar. Finally, let us offer an example of what we have learned about the role of everyday argumentation. One of the most common types of argument we have found proceeds from a description of what is the case to a conclusion about what ought to be the case. Thus, it would seem that justifying normative stances is one of the primary functions of everyday argumentation.

Beyond assisting us in learning about the use and purpose of argumentation, our corpus has allowed us to improve our classification scheme by alerting us to a diversity of types of argument that we were not aware of at the outset of our project. So too, we have learned that it is very difficult, even for trained researchers, to be consistent in their strategies for reconstructing missing premises. Thus, there is a need to develop more stringent rules governing such reconstruction.

6. Conclusion

Research and teaching in argumentation could benefit significantly from empirical resources such as argument research corpora. To this effect, among others, we have started to develop a general-purpose argument research corpus. Our corpus is available online as a resource for argumentation theorists and others. Our hope is that it will thus not only assist us in our work in argumentation, but will also further the work of the wider community of argumentation researchers and serve as a step towards creating additional resources both for the teaching of critical thinking and for research into argumentation.

APPENDIX I**ONLINE ARGUMENT SOURCES****NEWSPAPERS:****New York Times:**

<http://www.nytimes.com/>

Washington Post:

<http://www.washingtonpost.com/>

The Independent (UK):

<http://www.independent.co.uk/>

The Times (UK):

<http://www.timesonline.co.uk/?999>

The Mirror (UK):

<http://www.mirror.co.uk/>

The Telegraph (UK):

<http://www.telegraph.co.uk>

Financial Times (UK):

<http://news.ft.com/home/uk/>

The Times of India:

<http://timesofindia.indiatimes.com/>

The Hindustan Times:

<http://www.hindustantimes.com/>

The Indian Express:

<http://www.indian-express.com/>

The Age (Australia):

<http://www.theage.com.au/#>

The Australian:

<http://www.theaustralian.news.com.au/>

The International Herald Tribune:

<http://www.ihf.com/frontpage.html>

Mail and Guardian Online (South Africa):

<http://www.mg.co.za/11.jsp>

The Japan Times:

<http://www.japantimes.co.jp/start.htm>

PARLIAMENTARY RECORDS:**UK House of Parliament debates:**

<http://www.parliament.uk/hansard/hansard.cfm>

UK House of Lord debates:

http://www.parliament.uk/about_lords/about_lords.cfm

US Congress Congressional Record:

<http://thomas.loc.gov/home/r107query.html>

Indian Parliament Debates:

<http://alfa.nic.in/>

OTHER GOVERNMENTAL ORGS:**World Trade Organisation:**

<http://www.wto.org/>

LEGAL:**High Court of England and Wales (Judgments):**

<http://www.courtservice.gov.uk/>

House of Lords Judgments:

<http://www.parliament.the-stationery-office.co.uk/pa/ld/ldjudinf.htm>

United States Supreme Court:

<http://www.supremecourtus.gov/>

United States Circuit Courts of Appeals:

<http://www.findlaw.com/casecode/courts/1st.html>

Judgments of the Canadian Supreme Court:

<http://www.lexum.umontreal.ca/csc-scc/en/index.html>

WEEKLY MAGAZINES:**Time Magazine:**

<http://www.time.com/time>

The Economist:

<http://www.economist.co.uk/>

Outlook India:

<http://www.outlookindia.com/>

ONLINE MAGAZINES:**Grist Magazine:**

<http://www.gristmagazine.com/>

POPULAR SCIENCE:**Scientific American:**

<http://www.sciam.com/>

Science:

<http://www.sciencemag.org/>

TRANSLATED SOURCES:**Frankfurter Allgemeine Zeitung:**

<http://www.faz.com/IN/INtemplates/eFAZ/default.asp>

People's Daily (China):

<http://english.peopledaily.com.cn/>

Pravda (Russia):

<http://english.pravda.ru/>

Ha'aretz (Israel):

<http://www.haaretzdaily.com/>

DISCUSSION FORA:**BBC (Talking Point):**

<http://news.bbc.co.uk/>

CramSession Discussion Boards:

<http://boards.cramsession.com/boards/default.asp>

**Christian Apologetics & Research Ministry
Discussion Boards:**

<http://www.carm.org/boards.htm>

MSNBC discussion forum:

<http://weblab.msnbc.com/>

NPR discussion boards:

<http://www.npr.org/yourturn/>

UK Ebay discussion boards:

<http://pages.ebay.co.uk/community/index.html>

The Call to Islam:

<http://groups.yahoo.com/group/thecalltoislam/>

The New Muslimah:

<http://groups.yahoo.com/group/thenewmuslimah/>

Global Greens:

<http://www.globalgreens.org/index.php>

CAUSE INFORMATION SOURCES:**Human rights watch:**

<http://www.hrw.org/>

The True Religion:

<http://thetruereligion.org/>

**Global warming information page (cooler heads
coalition):**

<http://www.globalwarming.org/index.htm>

Independent Media Centre:

<http://uk.indymedia.org/>

REFERENCES

- Aristotle (1928). *Topics*. W. D. Ross (ed.). Oxford: Clarendon Press.
- Eemeren, F.H. van, Grootendorst, R. and F. Snoeck-Henkemans (1996). *Fundamentals of Argumentation Theory*. Mahwah: Lawrence Erlbaum Associates.
- Katzav, J. and C. Reed (2003). "On Argumentation Schemes and the Natural Classification of Arguments". *Forthcoming*.
- Musolff, A., Good, C., Wittlinger, R. and P. Points (2001). *Attitudes towards Europe. Language in the Unification Process*. Aldershot: Ashgate.
- Perelman, Ch. and L. Olbrechts-Tyteca (1969). *The New Rhetoric*. Notre Dame: University of Notre Dame Press.
- Toulmin, S.E. (1958). *The Uses of Argument*. Cambridge: Cambridge University Press.
- Walton, D. (1990). "What is Reasoning? What is Argument?" *Journal of Philosophy* 87: 399-419.