

Towards Creation of a Corpus for Argumentation Mining the Biomedical Genetics Research Literature

Nancy L. Green

Dept. of Computer Science
U. of N. Carolina Greensboro
Greensboro, NC 27402, USA
nlgreen@uncg.edu

Abstract

Argumentation mining involves automatically identifying the premises, conclusion, and type of each argument as well as relationships between pairs of arguments in a document. We describe our plan to create a corpus from the biomedical genetics research literature, annotated to support argumentation mining research. We discuss the argumentation elements to be annotated, theoretical challenges, and practical issues in creating such a corpus.

1 Introduction

Argumentation mining is a relatively new challenge in corpus-based discourse analysis that involves automatically identifying argumentation within a document, i.e., the premises, conclusion, and type of each argument, as well as relationships between pairs of arguments in the document. To date, researchers have investigated methods for argumentation mining of non-scientific text and dialogue. However, the lack of appropriately annotated corpora has hindered research on argumentation mining of scientific research articles. Using the term ‘argument’ in a related but different sense than here, researchers have investigated annotation of scientific abstracts and full-text articles (e.g. Teufel, 2002; Mizuta et al., 2005; Liakata et al., 2012). However, the annotated corpora they have created are not designed for argumentation mining in the above sense.

Our goal is to create a freely available corpus of open-access, full-text scientific articles from the biomedical genetics research literature, anno-

tated to support argumentation mining research. The corpus also would provide a rich new resource for researchers in related areas including information retrieval, information extraction, summarization, and question-answering. There is a critical need for automated analysis of the rapidly growing genetics research literature. Availability of the corpus should promote the development of computational tools for use by biomedical and genetics researchers. In the future, e.g., a tool enabled by argumentation mining could be used to automatically summarize arguments in the research literature that a certain genetic mutation is a cause of breast cancer. Methods developed from experimentation with this corpus should be adaptable to other scientific domains as well.

Section 2 of this paper discusses some terms from argumentation theory that are relevant to our goals and surveys related work. Section 3 discusses examples of argumentation in the target literature. The next three sections discuss challenges, practical issues, and future plans for creating the corpus.

2 Background

2.1 Argumentation Theory

Traditionally, an argument is said to consist of a set of *premises* and a *conclusion*, and a formal model such as deductive logic is used to determine whether the argument is valid. An argument can be attacked by refuting a premise or by presenting an argument for a conclusion in contradiction to the original conclusion. However Toulmin (1998), who was concerned with modeling arguments in fields such as law and science, argued that logical validity is too restrictive a criterion for determining argument acceptability. Toulmin distinguished two types of premis-

es: *data*, i.e., observations or conclusions of other arguments, and *warrant*, i.e., a field-dependent accepted principle (such as a legal rule or a “law” of science).

Argumentation schemes are abstract descriptions of forms of argument that are used to construct acceptable arguments in everyday conversation, law, and science (Walton et al., 2008). Argumentation schemes may describe non-deductively valid arguments, and their conclusions may be retracted when more information is obtained. For example, an abductive argumentation scheme, often used in genetic counseling (Green et al., 2011), is reasoning from observations to a hypothesized cause. *Critical questions* associated with argumentation schemes play an important role in evaluating argument acceptability (Walton et al., 2008). For example, one of the critical questions of the abductive argumentation scheme is whether there is an alternative, more plausible explanation for the observation used as a premise. An *enthymeme* is an argument with implicit premises or conclusion. Argumentation schemes are sometimes useful in reconstruction of missing components of enthymemes.

2.2 Argumentation Corpora

A corpus of genetic counseling patient letters was analyzed in several ways to design a computational model for generation of arguments from healthcare experts to patients (Green et al., 2011). An annotation scheme was developed to describe the conceptual model of genetic disease and inheritance communicated to patients (Green, 2005a). Formal argumentation schemes describing arguments found in the corpus were defined (Green et al., 2011). Analyses of pragmatic features included rhetorical relations (Green, 2010a), ordering constraints and discourse markers (Green et al., 2011), point of view (Green 2005b), and use of probability expressions (Green 2010b). However, it was not a goal of that project to provide a publicly available corpus.

The Araucaria argumentation diagramming tool was developed to aid human analysts and students to visualize and annotate naturally occurring arguments (Reed and Rowe, 2004). Diagrams can be stored as text files with stand-off annotation of premises and conclusions, argumentation schemes, and relationships between arguments. The Araucaria project has created a publicly available corpus of annotated argumentation from newspaper articles, parliamentary records, magazines, and on-line discussion

boards (Reed et al., 2010). The corpus has been used in some argumentation mining research (Mochales and Moens, 2011; Feng and Hirst, 2011; Cabrio and Villata, 2012).

2.3 Argumentation Mining

To date, researchers have investigated methods for argumentation mining of non-science content: legal documents (Mochales and Moens, 2011; Bach et al., 2013; Ashley and Walker, 2013; Wyner et al., 2010), on-line debates (Cabrio and Villata, 2012), product reviews (Villalba and Saint-Dizier, 2012; Wyner et al., 2012), and newspaper articles and court-cases (Feng and Hirst, 2011). Here we summarize the work that is most relevant to our project.

Mochales and Moens (2011) experimented with the Araucaria corpus and a legal corpus. They developed a multi-stage approach to argumentation mining. The first stage, *argumentative information detection*, addresses the problem of classifying a sentence (or sentential subunit) as being part of an argument or not. Next, *argument boundary detection*, or segmentation, is the problem of determining the boundaries of each argument. Third, *argumentative proposition classification* labels the sentences in an argument according to their role as a premise or the conclusion. Lastly, *argumentation structure detection* is the problem of detecting the relationships between arguments, i.e., whether two atomic arguments are “chained” (the conclusion of one is a premise of another), whether multiple arguments are provided in support of the same conclusion, and whether one argument attacks another argument in some way. Statistical techniques were used for the first three stages, while manually constructed context-free grammar rules were used for *argumentation structure detection*.

Cabrio and Villata (2012) used an approach to *argumentation structure detection* based on calculating textual entailment (Dagan 2006) to detect support and attack relations between arguments in a corpus of on-line dialogues stating user opinions.

Feng and Hirst (2011) focused on the problem of *argumentation scheme recognition* in the Araucaria corpus. Assuming that the conclusion and premises of an argument have been identified already, classification techniques achieved high accuracy for two argumentation schemes described in (Walton et al., 2008), *argument from example* and *practical reasoning*. Those schemes are less likely to be useful in analysis of scientific texts however.

In fact, since scientific research articles substantially differ from the genres that have been explored for argumentation mining so far, it is an open question what techniques will be successful in the scientific literature.

2.4 Argumentative Zoning and Related Annotation Schemes

Some NLP researchers have studied ways to automatically identify discourse structure in scientific text. The motivation is to provide contextual information that will improve automatic information access without the need to represent or reason about domain knowledge (Teufel, 2010). These researchers have developed several annotation schemes.

The argumentative zoning (AZ) annotation scheme was developed for automatically classifying the sentences of a scientific article in terms of their contribution of new knowledge to a field (Teufel and Moens, 2002; Teufel, 2010). Applied to articles in computational linguistics, AZ labels “zones” or variable-length sequences of sentences with one of seven categories: AIM (the research goal of the article), BASIS (the contribution of existing knowledge to a knowledge claim of the article), CONTRAST (criticizing or negatively contrasting competitors’ knowledge claims to a knowledge claim of the article), TEXTUAL (indicating the structure of the article), BACKGROUND (generally accepted background knowledge), OTHER (existing knowledge claims), and OWN (describing any aspect of a new knowledge claim made by the authors).

An extension of AZ (AZ-II) developed for application to chemistry articles, refined AZ’s distinctions into fifteen categories (Teufel, 2010). In another extension of AZ developed for genetics articles (Mizuta et al., 2005), the AZ OWN category was replaced by categories distinguishing descriptions of methodology (MTH), experimental results (RSL), insights from experimental results or previous work (INS), and implications (such as conjectures and applications) of experimental results or previous work (IMP).

The CoreSC (Core Scientific Concepts) annotation scheme was developed for automatic classification of sentences in terms of the components of a scientific investigation: Hypothesis, Motivation, Goal, Object, Background, Method, Experiment, Model, Observation, Result and Conclusion (Liakata et al., 2012a). An automatic classifier for CoreSC was developed and evalu-

ated on a corpus of 265 full-text articles in biochemistry and chemistry. A comparison study (Liakata et al., 2012b) in which articles were annotated with both AZ-II and CoreSC “found that CoreSC provides finer granularity ... while the strength of AZ-II lies in detecting the attribution of knowledge claims and identifying the different functions of background information” (Liakata et al. 2012b, p. 45). Liakata et al. (2012b) compared CoreSC to two other scientific discourse annotation schemes (Thompson et al., 2011; De Waard and Pander Maat, 2009). The three schemes were found to be complementary, operating at different levels of granularity.

However, none of the above annotation schemes address argumentation as described in section 2.3. They are not designed to identify the premises and conclusion of each argument (including missing components of enthymemes) and the argumentation scheme, nor relationships between pairs of arguments. Nevertheless, we plan to coordinate our efforts with that research community to benefit from their expertise and to ensure that our corpus will ultimately provide a valuable resource for their research.

3 Examples

In this section we discuss examples of some of the arguments in an article (Schrauwen et al., 2012) that is representative of the articles to be included in the corpus. The main claim of this article is that a c.637+1G>T mutation of the *CABP2* gene in the region 11q12.3-11q13.3 (DFNB93) is a cause of autosomal recessive non-syndromic hearing loss (arNSHL) in humans. The article’s body is divided into four sections: Introduction, Material and Methods, Results, and Discussion. The following examples in Table 1 are from the first subsection of the Results section (under the subheading “Next-Generation Sequencing of the DFNB93 Region Identifies a Splice-Site Mutation in *CABP2*”). The excerpt has been manually segmented into regions of text conveying arguments. Adjacent segments not conveying arguments have been omitted to save space; the approximate number of omitted lines is given in square brackets. Also, for readability, alternative identifiers of genetic variants have been replaced by ellipses.

1	¶The DFNB93 region contains more than 300 annotated and hypothetical genes, and several genes are expressed in the mouse and human inner ear. Because there are many strong candidate genes in the region, we sequenced all genes and noncoding genes in this region by using a custom DNA capture array to identify the disease-causing mutation in one affected individual from the family. [skip next 5 lines]
2	¶After the identified homozygous variants were filtered through the 1000 Genomes Project November 2010 release and dbSNP131, 47 previously unreported variants remained and included two exonic mutations, one splicing mutation, six nontranslated mutations, 16 intergenic (downstream or upstream) mutations, and 22 intronic mutations.
3	The two exonic variants included one nonsynonymous variant, c.1379A>G ... in PPFIA1 ... and synonymous variant c.174G>A ... in GAL3ST3 ... The splice-site variant, c.637+1G>T ... was located at the 5' donor site of intron 6 of CABP2 (Figure 1 and Figure S1, available online). ¶The variants in PPFIA1 and CABP2 were subsequently validated by Sanger DNA sequencing, which only confirmed the splicing variant in CABP2. [skip next 4 lines]
4	Next, we checked the inheritance of the CABP2 variant in the entire Sh10 family (Figure 1) and screened an additional 100 random Iranian controls to ensure that the variant is not a frequent polymorphism. The mutation was not detected in any of the controls, and inheritance was consistent with hearing loss in the family.

Table 1. Excerpt from (Schrauwen et al., 2012)

In an annotation scheme such as AZ, the first sentence of segment 1 might be classified as BKG (background) and the second as MTH (methodology). In CoreSC, the second sentence might be classified as Hypothesis and Method. However, the following argument is also communicated in (1) to the intended audience of scientists. (A genetics researcher has confirmed our interpretation of the arguments in this paper.) Note that in the following analyses in our paper,

square brackets indicate implicit information derivable from the discourse context or domain knowledge. In the following argument, two of the premises are implicit, i.e., this is an example of an enthymeme. Also, premises are distinguished as Data or Warrant, where the former type of premise corresponds to old or new evidence or a conclusion of another argument in the article, and the latter to generally accepted principles or assumptions in genetics. It is understood by the intended audience that warrants may have exceptions and that the conclusions of the following arguments are tentative.

Note that the conclusion of Argument 1 has been recovered from the phrase *there are many strong candidate genes in the region*. The argument can be analyzed in terms of a type of abductive argumentation scheme, i.e., reasoning from effect (arNSHL) to plausible cause (a mutation in the DFNB932 region). For a specification of the argumentation schemes identified in the genetics paper, see (Green and Schug, in preparation).

Argument 1:

Data: Several genes in the DFNB93 region are expressed in the human inner ear.

Data: [arNSHL involves the inner ear]

Warrant: [If a gene is expressed in a tissue related to a genetic condition then a mutation of that gene may be a cause of that condition]

Warrant: [Autosomal recessive genetic conditions are caused by homozygous mutations.]

Conclusion: A [homozygous] mutation of a gene in the DFNB93 region may be a cause of arNSHL in humans.

In an annotation scheme such as AZ, the subordinate clause at the beginning of segment 2 might be classified as MTH, and the main clause as RSL (results). However it has been analyzed in Argument 2 as an instance of an argumentation scheme involving the elimination of candidates. Note that the identity of the arNSHL-affected individual whose DNA was tested (V:14) and the family to which she belonged (Sh 10) was not specified in this section, but was given in the Material and Methods section. Also note that the first premise in Argument 2 is the conclusion of the preceding Argument 1. In our paper, this is indicated by providing the previous argument's identifier in parentheses.

Argument 2:

Data: (Argument 1) [A homozygous mutation of a gene in the DFNB93 region may be a cause of arNSHL in humans]

Data: [In a DNA sample from one arNSHL-affected individual, identified as V:14 of family Sh10] 47 previously unreported [i.e. not frequent polymorphisms] homozygous variant alleles in the DFNB93 region were identified.

Warrant: [If a variant is a frequent polymorphism then it is not a cause of a genetic condition]

Conclusion: [One of the 47 variants may be the cause of arNSHL in individual V:14]

Various clauses in segment 3 might be classified as MTH or RSL in a scheme such as AZ. In an argumentation analysis, however, it conveys an argument that the *CABP2* mutation may be the cause of arNSHL in one individual (:V14), after the elimination of the other candidates.

Argument 3

Data: (Argument 2) [One of the 47 variants may be the cause of arNSHL in individual V:14]

Data: Only splice-site variant c.637+1G>T of *CABP2* was confirmed.

Warrant: [Only confirmed exonic or splice-site variants may be the cause of arNSHL.]

Conclusion: [The c.637+1G>T variant of *CABP2* may be the cause of arNSHL in individual V:14]

Segment 4 uses two different sets of data to argue that the c.637+1G>T variant of *CABP2* may be the cause of arNSHL in the family of V:14, Sh10. In a scheme such as AZ, the first sentence would probably be described as MTH and the second as RSL. However, an argumentation analysis provides two arguments, 4a and 4b. They each support the same conclusion, which is not explicitly stated in the text.

Argument 4a

Data: (Argument 3) [The c.637+1G>T variant of *CABP2* may be the cause of arNSHL in individual V:14]

Data: Inheritance of the variant segregates with arNSHL in family Sh10.

Warrant: [A mutation that is present in one affected family member may be the cause of an autosomal recessive genetic condition in the rest of the family if the mutation segregates with the genetic condition in the family (i.e., the mutation is present in all and only the family members who have the genetic condition, and the oc-

currence of the condition is consistent with autosomal recessive inheritance)]

Conclusion: [The c.637+1G>T variant of *CABP2* may be the cause of arNSHL in family Sh10]

Argument 4b

Data: Inheritance of the variant c.637+1G>T of *CABP2* segregates with arNSHL in family Sh10.

Data: The variant c.637+1G>T of *CABP2* is not found in the DNA of a control group of 100 individuals [who are not in family Sh10 and who are not affected with arNSHL]

Warrant: [If a variant segregates with an autosomal recessive condition in a family but is not found in the DNA of a control group of individuals who are not affected with the condition, then it may be the cause of the condition in that family]

Conclusion: [The c.637+1G>T variant of *CABP2* may be the cause of arNSHL in family Sh10]

In addition to identifying individual arguments, argumentation mining addresses relationships between pairs of arguments. Arguments 1-4a illustrate a chain of arguments, i.e., where the conclusion of Argument *i* is a premise of Argument *i+1*. Also, arguments 4a and 4b illustrate two arguments in support of the same conclusion. Note that, individually, Arguments 1-3 are relatively weak. However, Argument 1 might be useful in answer to a query such as *What regions may carry a mutation leading to arNSHL?* Arguments 2-3 might be useful in answer to a query such as *Have any individual cases of arNSHL been attributed to a mutation of CABP2?* Arguments 1-4a and Argument 4b could be given as the answer to the query *What mutation may be the cause of arNSHL in an affected family?* (Note that in an interactive query facility, instead of presenting the user with a chain of arguments, the system could leave it up to the user to “drill down” to see the subarguments in a chain.)

The above arguments are provided here for purposes of illustration. In the remainder of the genetics article the main claim (that the *CABP2* mutation is a cause of arNSHL in humans) is supported by arguments that the mutation is the cause of arNSHL in two other families. Also, using a different type of argumentation, it provides a biochemical explanation for how the mutation may cause an abnormality in the inner ear that could cause hearing loss. In addition to the main claim, the article contains several other supported claims, e.g., that the c.637+1G>T variant of *CABP2* may be a founder mutation.

4 Challenges

Argumentation mining of this type of discourse will be challenging. A challenge that is shared with BioNLP text mining in general is dealing with the extensive use of biological, chemical, and clinical terminology in the BioNLP domain. A number of challenges specific to argumentation mining are discussed next.

To specify an argument it is necessary to identify the premises (or data and warrant), conclusion, and argumentation scheme. However, as illustrated in the previous examples, arguments with implicit components (enthymemes) are common, e.g., where a conclusion is implicit or used as an implicit premise of another argument. A related challenge is to supply domain knowledge for reconstructing implicit warrants in this genre. Another related challenge is the need to make use of discourse context to supply missing information, e.g., where context is required to supply the identity of individual V:14 in Argument 2. Note that in that case, it was necessary to read the previous Materials and Methods section to supply that information.

Another problem illustrated in the example is that argument boundaries do not coincide with sentential subunit boundaries. For example, segment 4 contains parts of Argument 4a and 4b in the first sentence and parts of those two arguments in the second sentence. Furthermore, identification of argument components does not appear to be systematically associated with discourse markers such as ‘therefore’. However, the arguments contain lexical items relating to scientific discovery (e.g., ‘confirmed’, ‘detected’, ‘consistent with’, ‘indicate’, ‘is likely that’, ‘expected to’, ‘showed’, ‘suggests’) that may aid in automatic identification of the components.

Our analysis of argumentation in genetic counseling (Green et al., 2011) and in the genetics research literature (Green and Schug, in preparation) has identified other (and more specific) argumentation schemes and critical questions than those listed in (Walton et al., 2008). Since some of the argumentation schemes we have identified are causal, lexical patterns of causality may be useful features for use in argumentation mining.

5 Practical Considerations for Creating the Corpus

In order to ensure that the future corpus can be freely disseminated, we will select articles from journals that are fully open-access, i.e., that are published under the Creative Commons attribu-

tion license “which allows articles to be re-used and re-distributed without restriction, as long as the original work is correctly cited” (<http://www.biomedcentral.com/about>). To date, we have identified the following fully open-access journals that contain biomedical genetics research articles:

- BMC <http://www.biomedcentral.com> journals: BMC Genetics, BMC Genomics, BMC Medical Genetics, BMC Medical Genomics and BMC Molecular Biology,
- PLoS <http://www.plos.org/> journals: Genetics, Biology, Medicine

A number of other journals (e.g. *American Journal of Human Genetics*), indexed by PubMed (<http://www.pubmedcentral.nih.gov>), make a subset of their articles available as open-access.

After selecting articles for the corpus, we will define and evaluate the intercoder reliability (Arstein and Poesio, 2008) of the following types of annotations:

- Data, warrant, and conclusion and argumentation scheme of each argument,
- Multiple arguments for the same conclusion, and
- Chained relationships between arguments, i.e., where the conclusion of an argument is the premise of a subsequent argument.

Note that we plan to employ graduate students with a background in genetics and biochemistry as coders.

Identifying *implicit* components of arguments will be challenging for coders. However, there are a number of constraints that will be given in the instructions to help the coders. First, they will be given a list of commonly accepted principles of genetics as possible warrants, such as Mendel’s laws, the concept of segregation in a pedigree, etc. Second, coders will be instructed to look for chained arguments, i.e., where the premises/conclusions of chained arguments can be reconstructed from the relationship between two arguments. Third, coders will be given a description of argumentation schemes, which also constrain the interpretation of argument components.

A pilot annotated corpus and associated documentation of the argumentation coding scheme will be made available to other researchers on a temporary basis for the purpose of publicizing the planned corpus and getting feedback from potential stakeholders.

An important consideration is the selection of corpus annotation tools to facilitate argumentation mining research. On the one hand, the text

mining community uses linguistic annotation tools such as GATE (<http://gate.ac.uk/>), UIMA (<http://www.ibm.com/research/uima>), and OpenNLP tools (<http://opennlp.sourceforge.net>). It would be advisable to use tools that would allow that community to benefit from the argumentation corpus, as well as to allow argumentation mining researchers to use previously developed tools. For example, argumentation mining researchers may find it useful to automatically preprocess the corpus with linguistic annotations as well as the annotation schemes described in section 2.4. BioNLP researchers may find it useful to consider argumentation annotations as well. Just as modality and negation currently are used for BioNLP tasks, a text segment's participation in argumentation as outlined in this paper may provide useful context at a deeper level of analysis.

On the other hand, the argumentation and educational community uses tools for diagramming argumentation, e.g.

Araucaria (<http://arg.computing.dundee.ac.uk>) and LASAD (<http://cscwlab.in.tu-clausthal.de/lasad>).

It is important to maintain compatibility between argumentation mining corpora developed with linguistic annotation tools and corpora developed with diagramming tools.

6 Conclusion

This paper described our plan to create a freely available corpus of open-access, full-text scientific articles from the biomedical genetics research literature, annotated to support argumentation mining research. It discussed the argumentation elements to be annotated, theoretical challenges, and practical issues in creating such a corpus. We hope this workshop will provide an opportunity for us to get feedback from potential users (or contributors) to this effort, and possibly even identify synergistic research opportunities.

Acknowledgments

We thank Dr. Malcolm Schug of the Biology Department of the University of North Carolina Greensboro for verifying our interpretation of the arguments in the genetics article.

References

Artstein, R. and Poesio, M. 2008. Inter-Coder Agreement for Computational Linguistics. *Computational Linguistics* 34(4): 555-596.

Ashley, K.D. and Walker, V.R. 2013. Towards Constructing Evidenced-Based Legal Arguments Using Legal Decision Documents and Machine Learning. In *Proc. ICAIL 2013*, June 10-14, Rome.

Bach, N.X., Minh, N.L., Oanh, T.T., and Shimazu, A. 2013. A Two-Phase Framework for Learning Logical Structures of Paragraphs in Legal Articles. *ACM Trans. Asian Lang. Inform. Process.* 12, 1, Article 3 (March 2013).

Cabrio, E. and Villata, S. 2012. Generating Abstract Arguments: A Natural Language Approach. In Verheij, B., Szeider, S., and Woltran, S. (eds.) *Computational Models of Argument: Proceedings of COMMA 2012*. Amsterdam, IOS Press, 454-461.

Dagan, I., Dolan, B., Magnini, B., and Roth, D. 2009. Recognizing textual entailment: Rationale, evaluation, and approaches. *Natural Language Engineering* 15(4): i-xvii.

De Waard, A. and Pander Maat, H. 2012. Knowledge Attribution in Scientific Discourse: A Taxonomy of Types and Overview of Features. In *Proc. of the ACL 2012 Workshop on Detecting Structure in Scientific Discourse*.

Feng, V.W. and Hirst, G. 2011. Classifying Arguments by Scheme. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics*, Portland, OR, 987-996.

Green, N. 2005a. A Bayesian Network Coding Scheme for Annotating Biomedical Information Presented to Genetic Counseling Clients. *Journal of Biomedical Informatics* 38: 130-144.

Green, N. 2005b. Analysis of Linguistic Features Associated with Point of View for Generating Stylistically Appropriate Text. In J. G. Shanahan, James G., Qu, Y., and Wiebe, J. (Eds.) *Computing Attitude and Affect in Text: Theory and Applications*, 33-40. Secaucus, NJ: Springer-Verlag.

Green, N. 2010a. Representation of Argumentation in Text with Rhetorical Structure Theory. *Argumentation* 24(2): 181-196.

Green, N. 2010b. Analysis of communication of uncertainty in genetic counseling patient letters for design of a natural language generation system. *Social Semiotics*. 20(1):77-86.

Green, N., Dwight, R., Navoraphan, K., and Stadler, B. 2011. Natural Language Generation of Transparent Arguments for Lay Audiences. *Argument and Computation* 2(1): 23-50.

- Green, N. and Schug, M. In preparation. Modeling Argumentation in Scientific Discourse.
- Liakata, M., et al. 2012a. Automatic recognition of conceptualization zones in scientific articles and two life science applications. *Bioinformatics* 28(7).
- Liakata, M., et al. 2012b. A Three-Way Perspective on Scientific Discourse Annotation for Knowledge Extraction. In *Proc. of the ACL 2012 Workshop on Detecting Structure in Scientific Discourse*, 37-46.
- Mizuta, Y., Korhonen, A., Mullen, T. and Collier, N. 2005. Zone Analysis in Biology Articles as a Basis for Information Extraction. *International Journal of Medical Informatics* 75(6): 468-487.
- Mochales, R. and Moens, M. 2011. Argumentation mining. *Artificial Intelligence and Law* 19, 1-22.
- Monteserin, A. and Amandi, A. 2010. Building user argumentative models. *Applied Intelligence* 32, 131-145.
- Reed, C. and Rowe, G. 2004. Araucaria: Software for argument analysis, diagramming and representation. *International Journal of Artificial Intelligence Tools* 14, 961-980.
- Reed, C., Mochales-Palau, R., Moens, M., and Milward, D. 2010. Language resources for studying argument. In *Proceedings of the 6th Conference on Language Resources and Evaluation, LREC2008, ELRA*, 91-100.
- Schrauwen et al. 2012. A Mutation in CABP2, Expressed in Cochlear Hair Cells, Causes Autosomal-Recessive Hearing Impairment. *The American Journal of Human Genetics* 91, 636-645, October 5, 2012.
- Teufel, S. and Moens, M. 2002. Summarizing Scientific Articles: Experiments with Relevance and Rhetorical Status. *Computational Linguistics* 28(4), 409-445.
- Teufel, S. 2010. *The Structure of Scientific Articles: Applications to Citation Indexing and Summarization*. Stanford, CA, CSLI Publications.
- Thompson, P., Nawaz, R., McNaught, J. and Ananiadou, S. 2011. Enriching a biomedical event corpus with meta-knowledge annotation. *BMC Bioinformatics*, 12: 393.
- Toulmin, S. E. 1998. *The Uses of Argument*, Cambridge, UK: Cambridge University Press.
- Villalba, M.P.G. and Saint-Dizier, P. 2012. Some Facets of Argument Mining for Opinion Analysis. In *Proc. COMMA 2012*, 23-34.
- Walton, D., Reed, C., and Macagno, F. 2008. *Argumentation Schemes*. Cambridge University Press.
- Wyner, A., Mochales-Palau, R., Moens, M-F, and Milward, D. 2010. Approaches to Text Mining Arguments from Legal Cases. In *Semantic Processing of Legal Texts*, 60-79.
- Wyner, A., Schneider, J., Atkinson, K., and Bench-Capon, T. 2012. Semi-Automated Argumentative Analysis of Online Product Reviews. In *Proc. COMMA 2012*, 43-50.