Annotation of Rhetorical Moves in Biochemistry Articles

Mohammed Alliheedi
Al Baha University, KSA
University of Waterloo
Waterloo, Ontario, Canada
`mallihee@uwaterloo.ca`

Robert E. Mercer
Dept. of Computer Science
University of Western Ontario
London, Ontario, Canada
`mercer@csd.uwo.ca`

Robin Cohen
Dept. of Computer Science
University of Waterloo
Waterloo, Ontario, Canada
`rcohen@uwaterloo.ca`

Abstract

This paper focuses on the real world application of scientific writing and on determining rhetorical moves, an important step in establishing the argument structure of biomedical articles. Using the observation that the structure of scholarly writing in laboratory-based experimental sciences closely follows laboratory procedures, we examine most closely the Methods section of the texts and adopt an approach of identifying rhetorical moves that are procedure-oriented. We also propose a verb-centric frame semantics with an effective set of semantic roles in order to support the analysis. These components are designed to support a computational model that extends a promising proposal of appropriate rhetorical moves for this domain, but one which is merely descriptive. Our work also contributes to the understanding of argument-related annotation schemes. In particular, we conduct a detailed study with human annotators to confirm that our selection of semantic roles is effective in determining the underlying rhetorical structure of existing biomedical articles in an extensive dataset. The annotated dataset that we produce provides the important knowledge needed for our ultimate goal of analyzing biochemistry articles.

1 Introduction

Scientists must routinely review the scholarly literature in their fields to keep abreast of current advances and to retrieve information relevant to their research. However, the volume of online scientific literature is immense, and rapidly increasing. In the biomedical field, the National Center for Biotechnology Information (NCBI) developed a literature search engine, PubMed[1], to access various databases such as MEDLINE (journal citations and abstracts for biomedical literature), full-text life science e-journals, and online

books. Between 2010 and 2018 PubMed repositories increased from more than 20 million citations for biomedical literature (Lu, 2011) to more than 28 million[2]. As a consequence, it has become extremely challenging for biomedical scientists to keep current with information in their fields. This challenge has attracted Natural Language Processing researchers to develop resources and automated tools for performing various tasks in Information Extraction and Text Mining using online corpora of biomedical articles, and thus enable biomedical researchers to better manage and exploit this volume of data (Hunter and Cohen, 2006).

The types of tasks currently handled by Biomedical Natural Language Processing (BioNLP) systems have generally been aimed at extracting very specific and limited information, for example, protein and gene names and relations (Cohen and Demner-Fushman, 2014), and so have been able to rely on relatively simple forms of information extraction. Although these approaches fulfil some information needs, more in-depth and comprehensive information contained in biomedical texts would be highly valuable to scientists. This type of information can enable validating scientific claims, tracing current research directions, reproducing scientific procedures, and so forth. Recently, a new and more challenging information extraction task has been introduced as a means of obtaining this type of information: identifying the argumentation structure in biomedical articles (e.g., (Green, 2014, 2015)).

The essence of argumentation can be considered as influencing others to gain their adherence to a particular idea (Perelman and Olbrechts-Tyteca, 1973). Arguments have an explicit logical structure, for example, claims that are backed with reasons, which in turn are supported by evidence,

---

[1]http://www.ncbi.nlm.nih.gov/pubmed

[2]http://www.ncbi.nlm.nih.gov/books/NBK3827/

leading to conclusions (Toulmin, 2003). Argumentation analysis is the recognition and identification of the different forms of argumentative structures in texts. Various studies have used recurrent patterns of text organization called *rhetorical moves* (i.e., text segments that are rhetorical and perform specific communicative goals) to analyze argumentative organization of texts manually (Swales, 1990) or automatically (Teufel and Moens, 2002). Swales' CARS model targets the Introduction section[3] of scientific articles. Teufel's interests are concentrated on rhetorical moves associated with defining the research space and suggesting the knowledge claims for computational linguistics and chemistry articles (Teufel, 2010). Kanoksilapatham (2003) adds to these works by providing the first comprehensive set of rhetorical moves for complete biochemistry articles.

With our long-term goal being analyzing argumentation in biochemistry articles, our mid-term research goal is to provide a computational model for Kanoksilapatham's descriptive rhetorical move taxonomy. Our research agenda is to design algorithms which would produce a representation of rhetorical moves in a biochemistry article and in this paper we outline the proposed semantic categories to be used, and discuss how we were able to guide human annotators to provide their interpretations of the analysis (to later be used as a gold standard in order to test our solutions).

Initially, our focus is on the Methods section of the taxonomy since this provides a description of the procedures followed in the experiment and the analysis of the results of the experiment thereby giving a framework for analyzing the moves in the remainder of the article. Because the experimental process is procedural, the moves tend to follow the verbs describing the steps in the experimental process. In other words, argumentation structure and scientific method both consist of rhetorical moves and experimental process, respectively. When a scientist describes her/his method in the writen article, it contains a list of experimental steps which are described by verbs (actions). These verbs evoke (initiate) the rhetorical moves in the writing. To understand the moves, we need information about the semantic roles associated with these procedural verbs. Two well known databases contain-

---

[3]Experimental articles in the biomedical sciences are normally organized in the IMRaD style: Introduction, Methods, Results, and Discussion.

ing semantic role information, Framenet (Baker et al., 1998) and Verbnet (Schuler, 2005), do not provide the information appropriate for the verbs found in this scientific domain. Our goal is to provide FrameNet and VerbNet-like information for the specialized domain of biochemistry.

So, the focus of this paper is to introduce the semantic roles that we are proposing for this domain, some of which are the same as those normally found and some which are new and we suggest are required for this domain. With these semantic roles and the Methods section rhetorical moves, we have begun annotating a corpus of the Methods sections from biochemistry articles. The annotation consists of the semantic roles and the rhetorical moves associated with each verb.

The paper is structured as follows: First, an overview of some theoretical and computational approaches to argumentation are presented in Section 2. Then, our proposed approach to argumentation analysis is described in Section 3. Next, a description of our annotation scheme is given in Section 4. A description of an annotation study conducted along with the creation of a dataset is given in Section 5. Finally, the future work and a conclusion of this paper is given in Section 6.

2 Related Work

2.1 Theoretical Approaches to Rhetorical Moves and Argumentation

Swales (1990) proposed the Create-A-Research-Space (CARS) model that uses intuition about the argumentative structure of scientific research articles. Swales defined rhetorical moves as text segments that convey communicative goals. However, despite the widespread influence of the CARS model, some researchers observed two problems: (i) the inconsistent assignment of rhetorical moves to text segments because the identification of the rhetorical moves relies on overall text comprehension, and (ii) a lack of empirical validation of moves in linguistic terms (Kanoksilapatham, 2003).

To overcome these problems, Kanoksilapatham (2003) advanced Swales' approach to move analysis by developing a framework that combines his original CARS model with the use of Biber's (1991) multidimensional analysis to enrich the model with additional information about linguistic characteristics. Although Kanoksilapatham provides an extension to the Swales move analysis

study and attempted a validation of these moves in biochemistry articles, she only provides a descriptive analysis about rhetorical moves without defining an explicit method for analyzing and recognizing these moves in texts.

2.2 Annotating Rhetorical Moves and Argumentation Schemes

Argumentative Zoning (AZ) was developed by Teufel and Moens (1999) to categorize sentences based on their contextual information (e.g., determining authorship of knowledge claims). The AZ scheme classifies sentences into seven categories including the ones from the CARS model (Swales, 1990). The data set consisted of 48 computational linguistic papers. Three annotators were involved in the study to extract sentences that fell into these seven categories. The results showed kappa scores of 83% and 82% between the annotators in the first and second schemes, respectively. The AZ scheme was later modified to suit the characteristics of biology articles (Mizuta et al., 2006). Furthermore, Teufel et al. (2009) and Teufel (2010) proposed a revised version of AZ to include more categories for annotating scientific articles such as chemistry. This revised version was planned to model all experimental sciences, which is challenging, since the style of scientific writing varies across disciplines. Most recently, Teufel (2015) has proposed a modified version of AZ to recognize rhetorical moves in scientific articles.

Liakata et al. (2012) developed an annotation scheme called Core Scientific Concepts (CoreSC) to classify sentences into scientific categories (e.g., "related to author's other work"). The authors use Machine Learning classifiers (i.e., Conditional Random Fields and Support Vector Machines) to automatically classify sentences into the CoreSC categories. The data set consisted of 265 biochemistry and chemistry articles. The authors were only able to achieve an accuracy around 50% in categorizing sentences in the appropriate CoreSC scientific categories indicating that this is a very difficult task.

Overall, these different approaches based on argumentation theories for analyzing and recognizing argumentative elements, including move analysis ((Kanoksilapatham, 2003) (Swales, 1990)), argumentative zoning (Teufel et al., 1999), and epistemic topoi (Gladkova, 2011), lacked a formal knowledge representation which could be used

computationally for in-depth argumentation analysis and mining.

Another problem in identifying argumentative elements is that relatively few biomedical related corpora annotated with argumentation structures currently exist for use in training or evaluating Machine Learning classifiers.[4] This has encouraged researchers to begin developing annotated corpora for use by the Computational Argumentation community ((Green, 2014, 2015), in particular).

Green (2014) proposed a plan for creating an annotated corpus of biomedical genetics research articles. Importantly, in justifying the need for such a corpus, Green strongly argued for domain knowledge as a requisite of argumentation recognition in the experimental sciences. Green (2015) specified a set of argumentation schemes for scientific claims in genetics research articles. The author used a corpus of unannotated genetics research articles, and identified the components (e.g., premises, conclusions) of an argument as well as its type of scheme. Overall, the author's ultimate goal for this initial study was to develop annotation guidelines for creating corpora for argumentation mining research.

None of these previous approaches to automated argumentation analysis and mining provided a formal knowledge representation that could be used in detecting and recognizing argumentative elements. We believe that developing a formal representational framework based on verb semantics in procedural scientific discourse will enable a more in-depth analysis of argumentative elements in a computationally feasible manner. We intend to provide such knowledge for the biochemistry domain to achieve this goal. This paper discusses the annotation of a corpus of biochemistry text, the first step in this longer term enterprise.

3 Procedurally Rhetorical Verb-centric Frame Semantics

In this research we will work on the biochemistry domain to develop a formal knowledge representation, *procedurally rhetorical verb-centric frame semantics*, that can be used for in-depth argumen-

---

[4]We note, however, increasing attention to this concern, with the design of such corpora as The Internet Argument Corpus (IAC) for research in political debates on internet forums (Walker et al., 2012) and the Dr. Inventor Multi-Layer Scientific Corpus (DRI) for computer graphics articles (Lauscher et al., 2018).

tation analysis, is computationally feasible to implement, and will enable argumentation mining of more-detailed scientific knowledge than is currently available. This will be an important step towards providing researchers in Computational Argumentation working in domains with similar discourse structure with a means of using and evaluating the metrics we will develop. To the best of our knowledge, no research has proposed or incorporated the idea of a semantic frame based on verb analysis to assist in the analysis of argumentation in biochemistry articles.

We have introduced various methods for detecting rhetorical moves in Section 2. We hypothesize that recognizing and detecting rhetorical moves would provide additional information to our framework of argumentation analysis. We also hypothesize that the Methods sections in biochemistry articles contain rhetorical moves which can be correlated with the author's experimental procedures. These moves can be used to determine salient information about the elements of the article's argumentative structure (e.g., premises) and can contribute to the overall understanding of the author's scientific claims. A key aspect of our hypothesis is that development of a frame-based knowledge representation can be based on the semantics of the verbs associated with these procedures. This representation can provide detailed knowledge for understanding these rhetorical moves, which will in turn facilitate analysis of argumentation structure. In other words, we propose that a *procedurally rhetorical verb-centric frame semantics* can be used to obtain a deeper analysis of sentence meaning than is currently the case with simple methods of Information Extraction (e.g., shallow syntactic pattern) and in a computationally feasible manner. Hence our focus on this critical section as a starting point for confirming the value of our chosen model for rhetorical moves and semantic roles.

Scientific argument[5] is defined as a process that scientists follow by using certain procedures to obtain empirical data which will either support or defeat their claims, hence leading to the intended conclusion. The strength of a scientific argument depends on its reproducibility and consistency. For a scientific argument to be strong, a scientist should identify and explain all the proce-

dures in their experiment, i.e., reproducibility, so that another researcher who follows the same procedures will reach the same conclusion, i.e., consistency. Thus, for a well-constructed scientific article, a scientist should expect the same conclusion if she follows the same procedures in the same sequence as described in the Methods section.

Scientific writing in the biochemistry domain has certain characteristics that made it ideal for our purposes. In this domain, experimental procedures describe the sequence of actions the biochemist performs to carry out an experiment to derive scientific conclusions, to demonstrate science experiments as can be seen in the experimental manuals (e.g., (Boyer, 2012; Sambrook and Russell, 2001)). Verbs play an essential role as indicators of these experimental procedures. These procedures can be viewed as corresponding to the elements of the scientific argumentation structure. For example, when examining a biological substance (e.g., a certain type of bacteria) in order to prove a hypothesis (e.g., this bacteria is correlated with a certain disease) the biochemist would perform a sequence of certain procedures to arrive at a conclusion. Essentially, biochemists create an argumentation framework through the scientific methodology they follow—how they perform their experiments is how they argue. We can observe that this genre— biochemistry articles—is procedure-oriented since the scientific procedures that are described are parallel to the scientific argumentation in the text. For example:

Example 1 *"Beads with bound proteins were washed six times (for 10 min under rotation at 4 C) with pulldown buffer and proteins harvested in SDS-sample buffer, separated by SDS-PAGE, and analyzed by autoradiography." (Ester and Uetz, 2008).*

In this example, the verbs "washed", "harvested", "separated", and "analyzed" are used to illustrate the procedure steps in sequential order. Such an experiment can be reproduced if one follows these steps.

Fillmore (1976) introduced the notion of frame semantics as a theory of meaning. A *semantic frame* is defined as 'any coherent individuatable perception, memory, experience, action or object' by Fillmore (1977), in other words, coherently structured concepts that are related to each other to represent a complete knowledge of world events or experiences. For example, to un-

derstand the word "buy", one would access the knowledge contained in the commercial transaction frame which includes words such as the person who buys the goods (buyer), the goods that are being sold (goods), the person who sells the goods (seller), and the currency that the buyer and seller agree on (money).

Following Fillmore's theory of frame semantics, FrameNet (Baker et al., 1998) was developed to create an online lexical resource for English. This framework includes more than 170,000 manually annotated sentences and 10,000 words. The computational linguistics community has been attracted to the concept of frame semantics and has developed computational resources using this concept, such as VerbNet (Schuler, 2005), an on-line verb lexicon for English and PropBank (Palmer et al., 2005), an annotated corpus with basic semantic propositions.

Following the notion of frame semantics, we propose to build a knowledge representation framework to analyze verbs in a procedure-oriented genre. Our concept of procedurally rhetorical verb-centric frame semantics is intended to address this lack of a formal framework by developing a computationally feasible knowledge representation that will enable argumentation analysis. The knowledge contained in the frame semantics will facilitate the extraction of elements of arguments, i.e., argumentation mining. To reiterate, our hypothesis is that procedurally rhetorical verb-centric frame semantics can provide a knowledge representation framework for analyzing and representing the meanings of the verbs used in biochemistry articles. In turn, these frames will facilitate the identification of argumentation structure in the discourse describing experimental procedures by highlighting the important steps in the experiment which are used to argue for the author's claims.

4 Annotation Scheme for Experimental Events

We have developed a new annotation scheme for identifying the structured representation of knowledge in a set of sentences describing the experimental procedures in the Method sections of biochemical articles. Several researchers have developed other forms of schemes (e.g., "bio-events" (Thompson et al., 2008)) to extract biological information (e.g., gene regulation). However, a bio-event is different from our definition of an experimental event. On the one hand, a bio-event is concerned with detection of bio-molecular events within the biomedical literature, such as the identification of events that are related to given proteins (Thompson et al., 2008). In our case, an experimental event is concerned with processes and procedures that are used to investigate biological events. The experimental event is also concerned with the recognition of the biochemist's reasoning of standard biochemical procedures such as using certain instruments or specific biological materials. Our annotation scheme consists of two tiers of information. A *rhetorical move* is on the sentence or clause level while *semantic role* is on the word or phrase level. The following subsections describe these two tiers of information.

Annotators are allowed to select the text span for labeling units (e.g., rhetorical moves and semantic roles) with some constraints as follows:

1. For a sentence or clause to be qualified as a rhetorical move, it must include a main verb and stand on its own. For example:

   Example 2 *"Beads with bound proteins were **washed** six times (for 10 min under rotation at $4°C$) with pulldown buffer ..." (Ester and Uetz, 2008).*

2. A sentence or clause that is qualified as a rhetorical move, it should have at least one or more semantic roles. Given the previous example, one could label the sentence as follows: - "Beads with bound proteins" as a *theme* - "were washed" as a *predicate*, - "six times", "for 10 min", "under rotation", and "at $4°C$" as protocol-details (repetition, time, condition, and temperature respectively).

4.1 Annotation for Rhetorical Moves

We have developed a set of rhetorical moves following Kanoksilapatham's (2003; 2005) work. That is, we have adapted and modified some of Kanoksilapatham's moves, as well as adding new more fine-grained moves to our annotation scheme. In combination, there are four major rhetorical moves concerned with the Methods section in biochemistry articles as can be seen in Table 1.

The clause given in Example 2, which is part of a complete sentence that contains several verbs, should be labeled as "Description-of-method".

Move type	Definition
Description-of-method	Concerned with sentences that describe experimental events.
Appeal-to-authority	Concerned with sentences that discuss the use of well-established methods.
Background information	Concerned with all background information for the experimental events such as "method justification, comment, or observation, exclusion of data, approval of use of human tissue" as defined by Kanoksilapatham (2003).
Source-of-materials	Concerned with the use of certain biological materials in the experimental events.

Table 1: Rhetorical Moves in the Method Sections of Biochemistry Articles

Semantic role	Definition
Agent	Generally a human or an animate subject.
Patient	Participants that have undergone a process.
Predicate	A word that initiates the frame.
Theme	Participants in a location or undergoing a change of location.
Goal	Identifies a thing toward which an action is directed or a place to which something moves.
Factitive	A referent that results from the action or state identified by a verb.
Location	The physical place where the experiments took place.
Protocol-Detail:	
Time	Identifies the time or a duration of an experimental process.
Temperature	Identifies the temperature of an experimental process.
Condition	Identifies the condition of how an experimental process is performed.
Repetition	Identifies the number of times an experimental process is repeated.
Buffer	Identifies the buffer that was used in an experimental process.
Cofactor	Identifies the cofactor that was used in an experimental process.
Instrument:	
Change	Describes objects (or forces) that come in contact with an object and cause some change.
Measure	Describes an object or protocol that can measure another object(s).
Observe	Describes an object which can be used to observe another object(s).
Maintain	Describes an object or protocol which can be used to maintain the state of object(s).
Catalyst	Describes an object that can be used as a catalytic "facilitator" for an experimental event to occur.
Reference	Refers to a method or protocol that is being used.
Mathematical	Describes a mathematical or computational instrument

Table 2: Semantic Roles in the Annotation Scheme of our Experimental Event

4.2 Annotation for Semantic Roles

As described earlier, our experimental event scheme was inspired by the annotation scheme for bio-events (Thompson et al., 2011). We based our experimental event scheme for verb arguments on the inventory of semantic roles in VerbNet (Schuler, 2005) and modified and added new semantic roles to define our scheme. Our experimental event scheme includes: *Theme*, *Patient*, *Predicate*, *Agent*, *Location*, and *Goal*. The complete set of semantic roles and their definitions in our experimental event scheme is presented in Table 2.

Working with a biochemist, we have extended the VerbNet definition of the semantic role *Instrument* from simply "an object or force that comes in contact with an object and causes some change in them" (Schuler, 2005) to include a variety of sub-categories corresponding to various types of biological and man-made instruments used in a biochemistry laboratory. We have also added *Protocol detail* as a set of semantic roles that identify certain types of information about experimental processes such as time and temperature.

5 Annotation

5.1 Data Set

We have created a data set consisting of 105 text files. These files include only the Methods sections from biochemistry journal articles which were randomly selected from PubMed Central. To prepare the data set for our task, all files were converted to plain text files that included one sentence per line and all figures and tables were omitted. We have used this data set for our initial text analysis that we described in Section 3. We also extended our data set to include 3499 articles between the years 2013 to 2015 from the top nine journals in biochemistry (Cell, Genome Research, Molecular Cell, Molecular Biology and Evolution, Molecular Aspects of Medicine, Nature Medicine, Nature Methods, Nature Structural & Molecular Biology, and Nature Chemical Biology).

5.2 Annotation Guidelines

We have created guidelines for annotating the Methods section in biochemistry articles. The guidelines include a description and the necessary background information of the task. The guidelines also include examples for each type of semantic role and their occurrence in the text. A list of questions supplements the guidelines to help annotators classify each sentence into its proper category. This task is done for semantic role labeling at the word level and rhetorical move labeling at the sentence level. We further supplemented the guidelines with a list of common co-factors and buffers that are normally used in the experimental procedures. Essentially, each annotator is asked to read the guidelines and if at any point she/he has a question or needs clarification, we can illustrate by providing more examples. We set up a meeting with the annotators either by Skype or in person to answer their questions. In fact, the guidelines have been revised and updated several times to reflect the annotators' feedback.

Our plan is to hire experts in the biomedical domain to label the Methods section in all of the articles in our dataset using our annotation scheme. Due to resource limitations, only 5% of the total number of articles have been annotated by two annotators, to date. We have hired ten annotators with a variety of backgrounds (Biochemistry, Bioinformatics, Biology) and different academic levels ranging from Bachelor to PhD degree. The annotators have engaged in various training sessions that were led by the authors. We have provided different resources that can help and support the annotators in this project. These resources include frequent meetings, the annotation guidelines, a list of questions and answers about the annotation, our biochemistry expert (a PhD student working with us), and the use of web-based software called Slack[6] which allows annotators to post questions, comments, or illustrate an example from the data set. We have also created a demo video[7] that shows annotators step by step how to use the GATE tool[8] and how to use the schema to label texts. Annotators are asked to use the GATE tool as an interface which gives them access to our developed schema for the semantic roles.

Each article is labeled by two annotators. The labeling is done on a verb basis rather than a full-sentence basis. In other words, each sentence with more than one verb is divided into smaller text spans (Annotation Units (AUs)), which are composed of a verb and the text containing its semantic roles. The annotators identity the verb in that AU and label all associated semantic roles for that verb

---

[6]https://slack.com/

[7]The demo video and guidelines are available at https://uwaterloo.ca/scholar/mallihee/links/gate-annotation-demo-and-annotation-guidelines

[8]https://gate.ac.uk/

Configuration	Kappa score
Original annotation	61.3%
Theme combined with patient and all instrument roles combined	68.9%
Protocol details combined	71.6%
Adjudicated	93.6%

Table 3: Inter annotator agreement κ-score for semantic role labeling

Configuration	Kappa score
Original	42.0%
Adjudicated	98.2%

Table 4: Inter annotator agreement κ-score for rhetorical move identification

within that AU. The annotators decide which constituent is a semantic role. Then, annotators label the entire AU with appropriate rhetorical moves. Each annotation is stored in an XML file. Figure 1 shows an example of some sentences annotated for both rhetorical moves and semantic roles.

5.3 Inter-annotator Agreement

Identification of semantic roles: We measured the inter-annotator agreement for semantic role labeling between the two annotations of the same article using the κ-score (Cohen, 1960). To have a matching label, both the semantic role category and the text span must be the same. Then, we measured the κ-score after the adjudication step which was done by one of the authors. The adjudication step's main goal is to resolve any disagreement in annotations (Palmer et al., 2005). We have also measured the kappa score for different configurations of the data set as shown in Table 3. "Original annotation" is the annotation that was provided by the annotators. "Theme combined with patient and all instrument roles combined" indicates theme and patient were combined as one role and all instrument subcategories were considered as one. "Protocol detail combined" indicates that in addition to the previous merging of semantic roles, all protocol detail subcategories were combined as one role. "Adjudicated" means that the disagreements in the original annotations were resolved and any missing semantic roles were added. All of the κ-scores in Table 3 are rated substantial (Landis and Koch, 1977; McHugh, 2012). The results are very promising.

Identification of rhetorical moves: We also measured the inter-annotator agreement for rhetorical move identification between the two annotations of each article using the κ-score. Here again, the rhetorical move and text span must be the same to be considered a match. As seen in Table 4, we

have measured the kappa-score for two configurations. "Original" is the annotation provided by the annotators, while "Adjudicated" means that the disagreements in the original annotations were resolved. The Adjudicated step was done by one of the authors. The result, shown in Table 4, shows a moderate to almost perfect agreement (Landis and Koch, 1977; McHugh, 2012). We have calculated the confusion matrix for the original annotation of rhetorical moves. During our adjudication step, we noticed some commonly mislabeled instances by some annotators. For example:

Example 3 *"The hierarchical cluster analyses were performed in MATLAB (Release 2012a), and the bar graphs were produced in Microsoft Excel 2010." (Davies et al., 2015).*

This sentence should be labeled "Description-of-method" since it clearly describes steps of the authors' method, i.e., using tools to perform analyses and produce graphs. However, one annotator mislabeled it as "Appeal-to-authority".

Example 4 *"Constructs comprising new opsin sequences cloned in pMT4 were transiently transfected into Neuro-2a cells with GeneJuice reagent (Novagen), according to the manufacturer's instructions (for further information, see Supplemental Material)." (Davies et al., 2015).*

This sentence was labeled incorrectly as "Description-of-method" whereas it should be labeled as "Appeal-to-authority" since it refers to an "established" method. We have concluded that our annotation guidelines need to be updated to better aid our annotators to properly select the right rhetorical move for each candidate AU.

6 Conclusion and Future Work

In this paper, we have presented the semantic roles that we have suggested to be necessary for this scientific domain and which will be used in our annotation scheme. This Experimental Event Scheme, which is based on the proposed semantic roles, is the first step towards developing an automated rhetorical move analysis. We have also presented
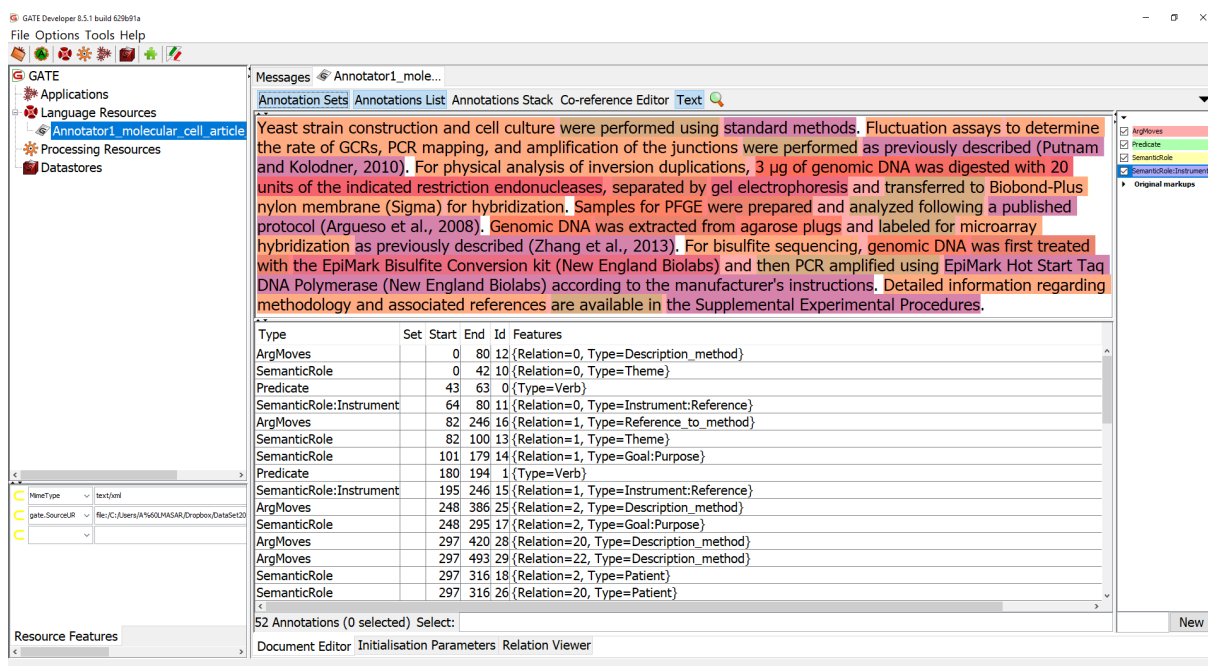
Figure 1: Snippet extracted from one article (Deng et al., 2015) of our annotated dataset showing the labelling of the rhetorical moves and semantic roles using the GATE tool.
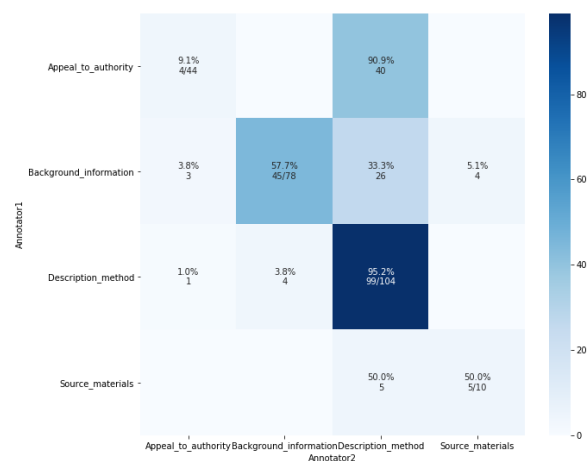

Figure 2: Confusion matrix for rhetorical move identification.

the most common rhetorical moves based on our observations of biochemistry procedures. We also have described our annotation study along with the dataset used. Ultimately, we aim to develop a framework to analyze argumentation structure in biochemistry procedures using the rhetorical moves.

We note that while there is substantial agreement among annotators in our results with respect to semantic roles, the agreement regarding rhetorical moves is more modest. One reason why this might be the case is the fact that the anno-

tated dataset to date is relatively small and annotators might actually have more inherent insight into recognizing the differences between rhetorical moves. Since these moves have spans which range from clauses to full sentences, whereas semantic roles are confined to at most a few words, the guidelines for annotation that were developed focused more on this simpler case. We anticipate expanding these guidelines in order to improve inter-annotator agreement regarding rhetorical moves in the future.

As future work, in parallel with annotating the complete data set, we will develop a computational model to label the rhetorical moves for this domain. As well, from our experience with annotating the biochemistry articles with our experts, we recognized that not all of the information needed to interpret the move structure is available in the text. What is needed is an ontology that captures the knowledge that a working biochemist would have regarding biochemistry experimental procedures, especially the sequence of events that are normally undertaken in these procedures. We have begun building such an ontology and future development will involve some automation.

References

Collin F. Baker, Charles J. Fillmore, and John B. Lowe. 1998. The Berkeley FramNnet project. In *Proceedings of the 17th international conference on Computational linguistics-Volume 1*, pages 86–90. Association for Computational Linguistics.

Douglas Biber. 1991. *Variation across speech and writing*. Cambridge University Press.

Rodney F. Boyer. 2012. *Biochemistry Laboratory: Modern Theory and Techniques*. Prentice Hall.

Jacob Cohen. 1960. A coefficient of agreement for nominal scales. *Educational and psychological measurement*, 20(1):37–46.

Kevin Bretonnel Cohen and Dina Demner-Fushman. 2014. *Biomedical natural language processing*, volume 11. John Benjamins Publishing Company.

Wayne IL Davies, T Katherine Tamai, Lei Zheng, Josephine K Fu, Jason Rihel, Russell G Foster, David Whitmore, and Mark W Hankins. 2015. An extended family of novel vertebrate photopigments is widely expressed and displays a diversity of function. *Genome research*, 25(11):1666–1679.

Sarah K Deng, Yi Yin, Thomas D Petes, and Lorraine S Symington. 2015. Mre11-sae2 and rpa collaborate to prevent palindromic gene amplification. *Molecular cell*, 60(3):500–508.

Claudia Ester and Peter Uetz. 2008. The ff domains of yeast u1 snrnp protein prp40 mediate interactions with luc7 and snu71. *BMC biochemistry*, 9(1):29.

Charles J Fillmore. 1976. Frame semantics and the nature of language. *Annals of the New York Academy of Sciences*, 280(1):20–32.

Charles J Fillmore. 1977. Topics in lexical semantics. *Current issues in linguistic theory*, 76:138.

Olga Gladkova. 2011. *Identification of epistemic topoi in a corpus of biomedical research articles*. Ph.D. thesis, University of Waterloo.

Nancy Green. 2014. Towards creation of a corpus for argumentation mining the biomedical genetics research literature. In *Proceedings of the first workshop on argumentation mining*, pages 11–18.

Nancy Green. 2015. Identifying argumentation schemes in genetics research articles. In *Proceedings of the 2nd Workshop on Argumentation Mining*, pages 12–21.

Lawrence Hunter and K. Bretonnel Cohen. 2006. Biomedical language processing: What's beyond pubmed? *Molecular Cell*, 21(5):589–594.

Budsaba Kanoksilapatham. 2003. *A corpus-based investigation of scientific research articles: Linking move analysis with multidimensional analysis*. Ph.D. thesis, Georgetown University.

Budsaba Kanoksilapatham. 2005. Rhetorical structure of biochemistry research articles. *English for Specific Purposes*, 24(3):269–292.

J Richard Landis and Gary G Koch. 1977. The measurement of observer agreement for categorical data. *biometrics*, pages 159–174.

Anne Lauscher, Goran Glavaš, and Simone Paolo Ponzetto. 2018. An argument-annotated corpus of scientific publications. In *Proceedings of the 5th Workshop on Argument Mining*, pages 40–46, Brussels, Belgium. Association for Computational Linguistics.

Maria Liakata, Shyamasree Saha, Simon Dobnik, Colin Batchelor, and Dietrich Rebholz-Schuhmann. 2012. Automatic recognition of conceptualization zones in scientific articles and two life science applications. *Bioinformatics*, 28(7):991–1000.

Zhiyong Lu. 2011. Pubmed and beyond: a survey of web tools for searching biomedical literature. *Database*, 2011.

Mary L McHugh. 2012. Interrater reliability: the kappa statistic. *Biochemia medica: Biochemia medica*, 22(3):276–282.

Yoko Mizuta, Anna Korhonen, Tony Mullen, and Nigel Collier. 2006. Zone analysis in biology articles as a basis for information extraction. *International journal of medical informatics*, 75(6):468–487.

Martha Palmer, Daniel Gildea, and Paul Kingsbury. 2005. The proposition bank: An annotated corpus of semantic roles. *Computational linguistics*, 31(1):71–106.

C. Perelman and L. Olbrechts-Tyteca. 1973. *The New Rhetoric: A Treatise on Argumentation*. University of Notre Dame Press.

Joseph Sambrook and David W. Russell. 2001. *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press.

Karin Kipper Schuler. 2005. *Verbnet: a broad-coverage, comprehensive verb lexicon*. Ph.D. thesis, University of Pennsylvania.

John Swales. 1990. *Genre analysis: English in academic and research settings*. Cambridge University Press.

S. Teufel. 2010. *The Structure of Scientific Articles: Applications to Citation Indexing and Summarization*. CSLI Studies in Computational. Center for the Study of Language and Information.

Simone Teufel. 2015. Scientific argumentation detection as limited-domain intention recognition. In *Proceedings of the Workshop on Frontiers and Connections between Argumentation Theory and Natural Language Processing*, volume 1341 of *CEUR Workshop Proceedings*, page 9pp.

Simone Teufel, Jean Carletta, and Marc Moens. 1999. An annotation scheme for discourse-level argumentation in research articles. In *Proceedings of the ninth conference on European chapter of the Association for Computational Linguistics*, pages 110–117. Association for Computational Linguistics.

Simone Teufel and Marc Moens. 2002. Summarizing scientific articles: experiments with relevance and rhetorical status. *Computational linguistics*, 28(4):409–445.

Simone Teufel, Advaith Siddharthan, and Colin Batchelor. 2009. Towards discipline-independent argumentative zoning: Evidence from chemistry and computational linguistics. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing*, pages 1493–1502.

Paul Thompson, Philip Cotter, John McNaught, Sophia Ananiadou, Montemagni, Andrea Simonetta, Trabucco, and Giulia Venturi. 2008. Building a bio-event annotated corpus for the acquisition of semantic frames from biomedical corpora. In *Proceedings of the Sixth International Language Resources and Evaluation*. European Language Resources Association.

Paul Thompson, Raheel Nawaz, John McNaught, and Sophia Ananiadou. 2011. Enriching a biomedical event corpus with meta-knowledge annotation. *BMC Bioinformatics*, 12(1):393.

Stephen E Toulmin. 2003. *The uses of argument*. Cambridge university press.

Marilyn Walker, Jean Fox Tree, Pranav Anand, Rob Abbott, and Joseph King. 2012. A corpus for research on deliberation and debate. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC-2012)*, pages 812–817, Istanbul, Turkey. European Language Resources Association (ELRA).