Analysis of Argument and Information Structures in Research Papers

Wei-Ning Cheng
Wee Kim Wee School of Communication and Information,
Nanyang Technological University, Singapore
Email: *wcheng009@e.ntu.edu.sg*

Abstract

The overall objective of this multi-disciplinary study is to investigate how information is structured in different types of arguments in research papers. Specifically, the author analyzes the relation and difference among argument types, rhetorical functions and information types across five types of research (i.e. *Investigative research*, *Development/evaluation research*, *Descriptive research*, *Historical analysis*, and *Identification research*) that have been identified in research articles published in high-impact sociology journals.

The approach is based on applied linguistics and information science: rhetorical analysis in the field of genre analysis, argument analysis based on the developed argument scheme in this study which is derived from Toulmin's Model of Argumentation (2003), and information analysis based on Frame Semantics Theory (Fillmore, 1976) from the field of knowledge representation and natural language processing. The scope of the study is limited to the Abstract and Introduction sections of research papers in three different domains: sociology, biological science, and mechanical engineering. The corpus consists of 150 research papers published in 30 journals—10 top journals from each of three domains.

By developing argument and information annotation schemes, and identifying argument and information structures across domains, this study will contribute to a deeper understanding of academic argumentation as reflected in research papers, with practical implications for the teaching of academic writing as well as for information literacy instruction. The developed *Research Information Model* (i.e. a set of research semantic frames) can theoretically be applied in (1) automatic text summarization in digital libraries and (2) deriving information for various argumentative purposes such as argument mining.

INTRODUCTION

Academic writing involves an intrinsically complicated mental process integrating thinking and writing: developing arguments by linking pieces of information, and expressing the arguments in sequential form in text to convince and persuade the reader. This is challenging not only for teachers to teach, but for students and scholars to learn, especially for new scholars who are non-native English speakers.

Many studies in the area of genre studies (a specialty within applied linguistics) have investigated the rhetorical structure of various text genres, and the rhetorical or persuasive function of each part (clause, sentence or paragraph) of the text. It is important to note that arguments and information have to be presented in text not only *sequentially* following rhetorical patterns expected of the genre, but the information has to be organized in way to help the reader to grasp the logic of the arguments. A vital demand to postgraduate students – especially international students – is to cultivate the ability of developing research ideas (Leong , 2015) and synthesizing information (Phakiti & Li, 2011) while writing research papers.

It is no surprise that some scholars have emphasized the importance of argument while writing a research paper (e.g. Hillocks, 2010). A well-known argument model applied in academic writing and genre studies is the British philosopher Stephen Toulmin's model of

argumentation. Toulmin (2003) indicated that the argument claim may be qualified by qualifiers and rebuttals as well as supported by data, warrant (the inference connection between the claim and the data), and backing (the support for the warrant).

However, although Toulmin's (2003) model has been used to analyze and evaluate the argument structure in various contexts (e.g., analyzing research papers and teaching academic writing), it can only model micro-level argument instances instead of macro-level argument structures. Toulmin, Rieke and Janik (1984, p. 14) referred to this macro-level argument as "the sequence of interlinked claims and reasons that … establish content and force of the position for which a particular speaker is arguing." Thus, this study analyzes two types of argument structures: *textual* (i.e. a list of argument claims in the sequential order) and *logical argument structure* (i.e. a complete set of arguments found in a research paper, linked by directed edges from argument supports to argument claims). This study mainly investigates the textual sequence of two or more argument elements (referred to as *textual argument chain*), and the relational structure of argument claims (referred to as *logical argument structure*) at the macro level. More specially, this study analyzes how the information structure changes over a logical sequence of two argument claims (referred to as *argument step*). This micro-level analysis is to understand how authors build arguments by selecting and organizing information and research thoughts logically to convince the readers.

In research papers, an *argument* refers to an assertion that the author seeks to convince the readers of with support such as evidence or logical reasoning. The Common Core State Standards (CCSS) Initiative defines an argument as "A reasoned, logical way of demonstrating that the writer's position belief, or conclusion is valid." (NGA, 2010, p. 23). In this study, an argument is divided into two types: *argument claim* (e.g., that a research contribution is significant) and *supporting argument* (e.g., a research gap).

An argument in a research paper has semantic or information content). To understand an argument (i.e. argument claim and supporting argument), we have to comprehend the argument elements: the concepts and the relations linking the concepts. I refer to a set of concepts linked together by conceptual relations as an *information structure*. Some scholars have highlighted the importance of understanding the information structure while writing a research paper. For example, Mallia (2017) pointed out that an author needs to think about how to organize information in the text structure as the "before writing tasks".

In my earlier study (Cheng, Khoo, & Kathpalia, 2017), I have identified that sociology research can be divided into five types: *Investigative research, Development/evaluation research*, *Descriptive research*, *Historical analysis*, and *Identification research*. I also found that the different types of research have different argument structures using various types of information, as well as involving other non-argument elements (i.e. *descriptive information* and *context information*). Thus, the research questions are developed as follows:

- RQ1: What are the types of *argument claims, supporting arguments*, *descriptive information* and *context information* are found in the five types of research – across two sections of research papers (i.e. Abstract and Introduction), and in three disciplines (i.e. sociology, biological science and mechanical engineering)?
- RQ2: What are the common textual sequences of argument claims (i.e. *textual argument chains*) in each section across five types of research and three disciplines?
- RQ3: What are the common logical structure of argument claims (i.e. *logical argument structures*) in each section across five types of research and three disciplines?

- RQ4: What are the common *information structures* found in research papers – associated with the five types of research, across the two sections of research papers, and in three disciplines?
- RQ5: What types of information structure (i.e. semantic frames) and what types of information (within the semantic frames) are associated with the argument claim and the supporting argument - for the five types of research? What are the differences across the five types of research, two sections of research papers, and the three disciplines?

METHODOLOGY

Study approach

This multidisciplinary study conducts different approaches based on applied linguistics and information science. It incorporates: rhetorical analysis in the field of genre analysis, argument analysis based on a scheme derived from Toulmin's model of argumentation (2003) in the field of philosophy, and information analysis based on Frame Semantics theory (Fillmore, 1976) from the field of knowledge representation and natural language processing.

In Toulmin's model of argumentation (2003), the *claim* may be qualified by *qualifiers* and *rebuttals* and supported by *data*, *warrant* (the inference connection between the claim and the data) and *backing* (the support for the warrant). However, although the model has been widely used to analyze and evaluate argument structures in the applied linguistics field, I found that most arguments in research papers do not structure as Toulmin's prediction, which is consistent with other scholars' criticism of Toulmin's model (e.g., Lunsford & Ruszkiewicz, 2010). For example, most research papers only have the claim supported by the data without an explicitly specified warrant. Thus, I developed my own argument scheme along with Toulmin's idea (2003): an *argument* in this study can be divided into two parts:

1. an *argument claim* refers to a proposition that the author believes to be true and is seeking to convince the readers (e.g., that a research objective is well-founded and worth investigating).
2. a *supporting argument* refers to a set of propositions leading up to the claim and provides evidence or logical reasoning (e.g., a research gap).

Frame Semantics theory (Fillmore, 1976) is applied in this study to model the information structures found in the text of research papers. The basic assumption is that: the meaning of a word should be understood based on the essential knowledge relevant to the word. All relevant information about the background knowledge of the word are embedded in the syntactic contexts, being carried by a semantic frame, which is "schematic representations of the conceptual structures and patterns of beliefs, practices, institutions, images, etc. that provide a foundation for meaningful interaction in a given speech community" (Fillmore, Johnson, & Petruck, 2003, p. 235). Based on the Frame Semantic theory, a set of research semantic frames has been developed in this study, which I term the *Research Information model*. In general, the model is used to analyze the information content in the Abstract and Introduction sections of research papers.

Corpus

The corpus for this study consists of 150 research papers published in 30 journals in three fields: biological science, sociology, and mechanical engineering. The fields are chosen due to the

corpus of a previous research project; however, these three fields can represent the diversity of important disciplines covering science, social science, and applied science.

The analysis is conducted in two sections (i.e. Abstract and Introduction) of research papers from the three fields. To ensure standardization in each discipline, 50 research articles are taken from 10 top journals listed in InCites Journal Citation Reports in each field. The articles were published in late 2015 or early 2016 volumes of the journals. Only articles reporting research that involves data analysis are included. Hence, journals and journal articles that report literature surveys or philosophical/theoretical discussions are not included in the study.

Frameworks

An argument scheme is developed and applied in this study. Twenty-nine types of argument claims and eleven types of supporting argument were identified (see Table 1 and 2). Due to the limited space, this paper only lists common types of argument claim and supporting argument.

Moreover, seven types of *descriptive information* (i.e. *Research scope, Research area, Data collection method, Theoretical framework, Extension of previous research, Outline the structure*, and *Data content*) and five types of *context information* (*Practical background, Historical background, Theoretical background, General research trend*, and *Topic classification structure)* were also found.

Table 1. Types of argument claims found in sociology research papers

Code	Type of argument claim	Definition
ArgC01	Research issue	Indicates a research issue or problem in the field. It is usually broader or more abstract than a research gap.
ArgC02	Research gap	Indicates a research gap indicating that a specific research issue has not been adequately investigated and deserves further study.
ArgC03	Research question	Indicates a research question of the current study, implying that is worthwhile investigating.
ArgC04	Research objective	Indicates a research objective of the current study, implying that it is well-founded.
ArgC05	Research method	Indicates a research method, implying that it is appropriate for addressing a research objective.
ArgC06	Research hypothesis	Indicates a specific hypothesis of the current study, either claiming that it is reasonable (based on theory or the author's belief) or that is a well-defined option related to the research objective.
ArgC07	Research result	Indicates a research result of the current study.
ArgC08	General result	Indicates a more general or abstract research result.
ArgC09	Research contribution/recommendation	Indicates a research contribution of the current study, or recommendation arising from the research results.
ArgC10	Topic centrality	Indicates the importance of the research topic or an aspect.
ArgC11	General statement	Indicates a broad or sweeping claim, i.e. more abstract or covers a wide area. It is sometimes a generalization previous research results.

Table 2. Types of supporting argument found in sociology research papers

Code	Types of supporting argument	Definition
ArgS01	Research method	A research method/approach to address a research objective, especially relating to the type of data and data collection.
ArgS02	Research gap/problem	The research gap/problem that the current study will address, which is a reason to conduct the current study. The research gap is given as an aside to support the research objective/contribution, but stated prominently as a claim.
ArgS03	Data analysis	Indicates the characteristics of the dataset and how it is analyzed. [Prefer this to research method]
ArgS04	Example	A specific example to illustrate or exemplify a claim.
ArgS05	Established knowledge (from previous studies)	Accepted knowledge or common understanding in the field (usually with multiple citations) to support a claim.

The *Research Information model* (see Figure 1) is developed according to the feature of research papers. Basically, different research relations (i.e. cause-effect, association, comparison, development, evaluation, and description) and information content (i.e. theory, model, framework and measurement) are established and covered in research papers across different types of research. A semantic frame assigns a role and meaning to each piece of information, in relation to the research relation. Thus, the model consists of three main research semantic frames representing different types of information associated with each type of research:

- *Research-relation frame* is associated with Investigative research.
- *Development/evaluation frame* is associated with Development/evaluation research.
- *Descriptive frame* is associated with Descriptive research, Historical analysis, and Identification research.

Another three semantic frames represent common information across the five types of research:

- *Comparison frame* to represent the comparative relation among concepts and the result.
- *Theory/model/framework frame* to represent the adopted theory, model, framework and concepts.
- *Measurement frame* to represent the way to measure the main entity of research.

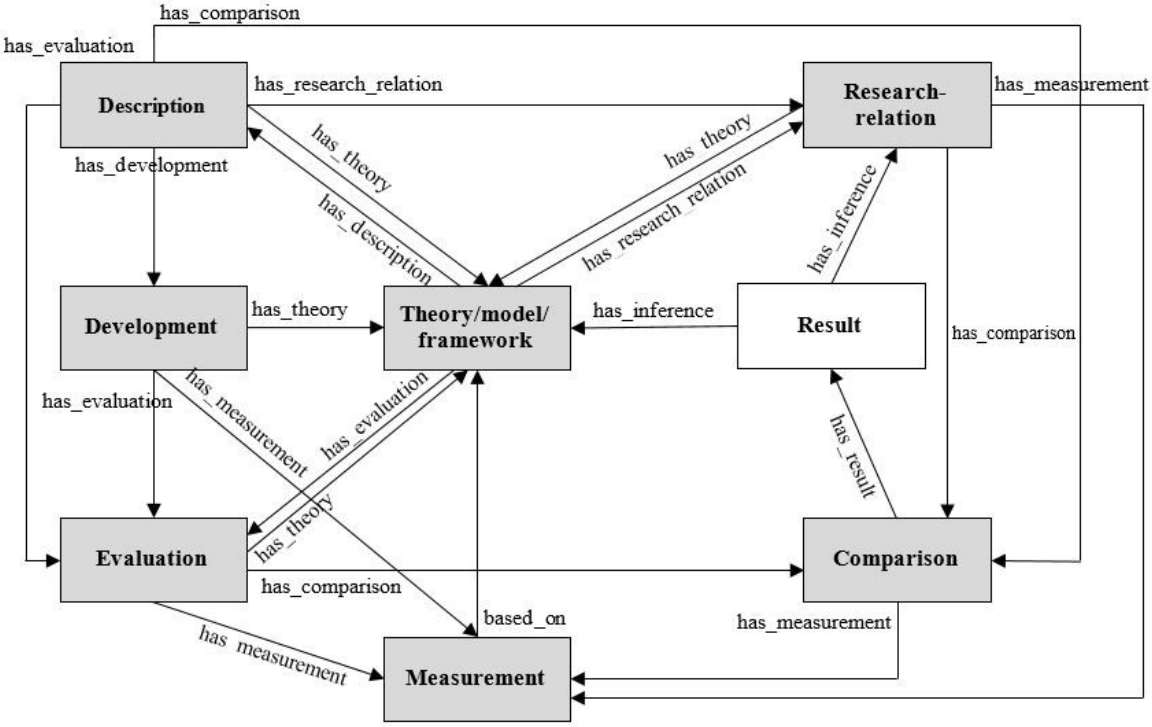

Figure 1. The *Research Information Model*

CONTRIBUTIONS

Significant contributions will be made in this multi-disciplinary study as follows:

1. The study will contribute to argumentation research by identifying the common types of argument claims and supporting arguments , which can be a coding scheme for further research in argument analysis. Moreover, the argument chains and argument structures across different domains will also be identified, which can be applied in other relevant studies or extended to other sections (i.e. Method and Conclusion) of research papers.
2. The research results will have practical implications for the teaching of academic writing and for information literacy instruction. On the one hand, teachers can teach students how to develop a well-structured research paper by presenting arguments with different types of information coherently in different sections, based on the argument chains, argument structures and information structures. On the other hand, librarians can teach students how to use different words or phrases to present research information. Moreover, how and when cited information should be used in arguments as support can also be taught.
3. The developed *Research Information Model* (i.e. a set of research semantic frames) can theoretically be applied in:
   - Automatic text summarization in digital libraries. The model can be used to shorten the length of research papers that 15 cover important arguments and information, creating summaries for database users.
   - Deriving information for various argumentative purposes such as argument mining. In the argument mining field, the model can be applied to offer structured data by extracting arguments automatically from unstructured documents.

REFERENCES

Cheng, W. N, Khoo, C. S. G., & Kathpalia, S. S. (2017). Argument structure of sociology research abstracts: An exploratory study. In Proceedings of the 8th Asia-Pacific Conference on Library & Information Education & Practice 2017 (A-LIEP 2017). Bangkok, Thailand: Faculty of Arts, Chulalongkorn University.

Fillmore, C. J. (1976). Frame semantics and the nature of language. *Annals of the New York Academy of Sciences: Conference on the Origin and Development of Language and Speech*, *280*, 20-32.

Fillmore, C.J., Johnson, C. R., & Petruck, M. R. L. (2003). Background to FrameNet. *International Journal of Lexicography*, *16*, 235-250. doi: 10.1093/ijl/16.3.235

Leong, A. P. (2015). Topical themes and thematic progression: the "picture" of research articles, *Text & Talk, 35*, 289-315. doi: 10.1515/text-2015-0001

Lunsford, A. A. & Ruszkiewicz, J. J. (2010). Everything is argument (5th ed.). Boston, MA :Bedford/St. Martin's.

Mallia, J. (2017). Strategies for developing English academic writing skills. Arab World English Journal, 8, 3-15. doi: 10.24093/awej/vol8no2.1

National Governors Association Center for Best Practices & Council of Chief State School Officers. (2010). Appendix A: Research supporting key elements of the standards. In *Common Core State Standards for English language arts and literacy in history/social studies, science, and technical subjects*. Washington, DC: National Governors Association Center for Best Practice, Council of Chief State School Officers.

Phakiti, A., & Li, L. (2011). General academic difficulties and reading and writing difficulties among Asian ESL postgraduate students in TESOL at an Australian university. *RELC Journal, 42*, 227-264. doi: 0.1177/0033688211421417

Tindale, C. W. (1999). *Acts of Arguing: A rhetorical model of argument*. New York: SUNY Press.

Toulmin, S. (2003). *The uses of argument*. Cambridge: Cambridge University Press.

Toulmin, S., Reike, R., & Janik, A. (1984). *An introduction to reasoning* (2nd ed.). New York, NY: Macmillan.