Information and argument patterns in the Introduction sections of sociology research papers

Wei-Ning Cheng & Christopher S. G. Khoo
Shanghai University (China) & Nanyang Technological University (Singapore)
wcheng009@e.ntu.edu.sg & chriskhoo@pmail.ntu.edu.sg

Abstract

This study analysed the information-argument structure of the Introduction sections of sociology research papers, to identify differences across three types of sociology research: *Investigative research*, *Development and Evaluation research*, and *Descriptive research*. The information-argument analysis framework bears some resemblance to rhetorical structure frameworks following Swales' CARS model, but focuses on the argumentative aspect of text and how information is used to support argument claims. The coding scheme specifies information types, subdivided into those that imply an argument claim and those that play the role of argument support. Seventy papers were sampled from ten sociology journals for analysis. Sequential association rule mining was used to identify sequential information-argument patterns. The study identified significant differences in information-argument profile across the three types of research papers, as well as differences in sequential patterns. Methodology contributions of the paper include the coding scheme for information-argument types in research papers, and the method of analysing sequential patterns.

Keywords: academic writing, argument structure, argument structure analysis, argumentation, academic argument.

Resumen

Patrones de información y argumentos en las secciones de Introducción de los trabajos de investigación en sociología

El presente estudio analiza la estructura de la información y argumentos en la introducción de los trabajos de investigación en sociología, con el objetivo de

identificar diferencias entre los siguientes tres tipos de investigación: la *investigación "investigadora", la investigación de desarrollo y evaluación,* y *la investigación descriptiva*. El marco para el análisis de la información y los argumentos guarda cierta semejanza con los marcos de estructura retórica que siguen el modelo de Swales, CARS, pero, a diferencia de ellos, se centra en la argumentación del texto y las maneras en que respaldar las afirmaciones. El esquema de codificación implementado diferencia entre dos tipos de información: los que incluyen una afirmación de los argumentos y los que contienen un apoyo. Como muestra de análisis se escogieron setenta artículos de diez revistas de sociología. Se han utilizado diferentes reglas de asociación secuencial para poder identificar posibles patrones secuenciales formadas por una estructura de información y argumentos. El estudio ha identificado diferencias significativas en el perfil de la estructura información-argumento entre los tres tipos de trabajos de investigación y entre los patrones secuenciales. Las contribuciones metodológicas del artículo incluyen un esquema de codificación para los tipos de argumentos de información en artículos de investigación y un método de análisis para patrones secuenciales.

Palabras clave: escritura académica, estructura del argumento, análisis de estructura argumental, argumentación, argumento académico.

1. Introduction

This is part of a broader study of the information and argument structure of research papers, focusing initially on sociology research papers. This paper reports our analysis of the Introduction section of sociology research papers, using our information-argument analysis framework and coding scheme. The framework is derived partly from rhetorical structure analysis frameworks following Swales's (1990) Creating a Research Space (CARS) model. Rhetorical structure analysis is a kind of discourse analysis focusing on the author's persuasive strategy and tactics (i.e., rhetorical steps or rhetorical functions). However, the rhetorical purpose must be to persuade or convince the reader about something, namely information and arguments. Our perspective is that discourse analysis of academic writing focusing on its argumentative aspect and how information is used in argument claims will yield a deeper understanding of the nature of a research study, that is, how it constructs new knowledge using a network of argument claims, supports and information.

Our coding scheme for information-argument types overlaps with the list of rhetorical steps in Swales's (1990) CARS model (listed in Table 1), and

subsequent elaborations by researchers in genre studies. Many of the steps in the CARS framework lack an obvious persuasive function. For example, *research objective* is a type of information and statement of research intent, rather than a persuasive strategy. Of the eleven rhetorical steps in the CARS framework, five clearly indicate argument claims: *Claiming centrality, Counter-claiming, Indicating a gap*, *Making topic generalisations* and *Announcing principal findings*. The other six rhetorical steps appear to refer to types of information, but may be considered to contain embedded arguments. For example, *Reviewing items of previous research* may include conclusions synthesised or generalised from previous research papers. *Continuing a tradition* may include justifications of why that is useful. Thus, the rhetorical steps mix persuasive functions with information types and argument claim types.

Our coding scheme (listed in Table 2) starts with a list of information types, but groups them into those that carry or imply an argument claim (Group 1), those that typically take the role of argument support (Group 2), those that are not argumentative but indicate types of descriptive information (Group 3), and those that indicate types of context information (Group 4). Focusing on the argumentative aspect of a text allows us to analyse the text in terms of `support => claim` links, bearing in mind that a claim can function as a support to another claim (i.e., `claim => claim`), and furthermore a support statement can occur in the text after the claim statement (`claim <= support`). Our analysis identified frequently occurring sequences of argument claims in the text (referred to as *sequential argument patterns*). We also analysed logical links between argument claims and argument supports that are independent of the sequence in which they appear in the text, thus conflating `support => claim` and `claim <= support` sequences We refer to these as *logical argument patterns*.[1]

We attempt to show in this paper that by foregrounding the argumentative aspect of discourse and its relation to information types, we gain - by linking information-argument elements into a network structure of discourse. In an earlier study (Cheng, 2020), we identified five types of sociology research represented in our corpus. As three of the research types were found in only a few papers and most of these studies employed qualitative methods, we combined them into one *Descriptive research* category. Thus, we carried out information-argument structure analysis for three types of sociology research papers, to identify discourse differences between them: Investigative research, Development and Evaluation research, and Descriptive research (defined later in Table 4).

The objective of this study was therefore to analyse the information-argument structure of the Introduction sections of sociology research papers, to identify differences across three types of sociology research. More specifically, we sought to find out:

1. How their information-argument profiles differ—in terms of the relative proportions of the different argument claim, support and information types.
2. How the *sequential* and *logical* argument patterns found in the papers are different.

One innovation in our analyses was the use of sequential pattern mining (a data mining technique) to identify frequent sequential patterns of argument claim, support and information types.

Moves	Steps
Move 1: Establishing a territory	Step 1: Claiming centrality
	Step 2: Making topic generalisations
	Step 3: Reviewing items of previous research
Move 2: Establishing a niche	Step 1A: Counter-claiming
	Step 1B: Indicating a gap
	Step 1C: Question-raising
	Step 1D: Continuing a tradition
Move 3: Occupying the niche	Step 1A: Outlining purposes
	Step 1B: Announcing present research
	Step 2: Announcing principal findings
	Step 3: Indicating the Structure of the Research Article

Table 1. Swales' cars model for Introduction sections of research papers.

Coding scheme	
Code	Group 1. Information types that carry/imply an argument claim
ArgC01	Research issue
ArgC02	Research gap
ArgC03	Research question
ArgC04	Research objective
ArgC05	Research method
ArgC06	Research hypothesis
ArgC07	Research result
ArgC08	General result
ArgC09	Research contribution/recommendation
ArgC10	Topic centrality
ArgC11	General statement
ArgC12	Research idea/approach
ArgC13	Address research limitation

ArgC14	Develop something new
ArgC15	Concept/theory/model–apply
ArgC16	Concept/theory/model–compare/relate
ArgC17	Concept/theory/model–derive
ArgC18	Concept/theory/model–define
ArgC19	Concept/theory/model–evaluate/critique
ArgC20	Literature–summarise/generalise
ArgC21	Literature–derive/infer
ArgC22	Literature–highlight claim
ArgC23	Literature–concession
ArgC24	Literature–counter argument
ArgC25	Literature–evaluate/critique
ArgC26	Literature–difference to current research
ArgC27	Literature–compare/relate
ArgC28	Literature–highlight scholar/concept
ArgC29	Concession/counter argument
ArgC30	Novelty
Code	Group 2. Information types playing an argument support role
ArgS01	Research method
ArgS02	Practical problem
ArgS03	Data analysis
ArgS04	Example
ArgS05	Established knowledge
ArgS06	Inference chain
ArgS07	Important implications
ArgS08	Theoretical mechanism
ArgS09	Explanation/Elaboration
ArgS10	Cited author's claim/opinion
ArgS11	Research motivation/justification
ArgS12	Theoretical framework
ArgS13	Extension of previous research
Code	Group 3. Descriptive information types
InfD01	Research scope
InfD02	Research area
InfD03	Outline the structure
InfD04	Data context
InfD05	Single study description
InfD06	Own previous work description
InfD07	Literature signpost
InfD08	Concept/theory/model-details
Code	Group 4. Context information types
InfC01	Practical background
InfC02	Historical background
InfC03	Personal background
InfC04	Theoretical background
InfC05	Topic classification structure

Table 2. Coding scheme for information-argument analysis framework.

2. Analysis framework: Information-argument analysis framework in relation to rhetorical structure frameworks

Our coding scheme for information-argument types was developed bottom-up from an analysis of 20 sociology research papers, and confirmed and refined using an additional 50 papers. The focus of the coding was on identifying types of information and types of argument claim and support that appear to be important to social science research, especially in arguing the validity of the research results and that a significant research contribution was made. Thus, *Research result* and *Research contribution* are two important information types, which carry the implicit claim that they are valid. "Claim" suggests that the propositions in the *Research result* and *Research contribution* statements are not self-evidently true; the author has to adduce support for the claims to show that the claims are valid or at least reasonable. Thus, other information types are used as argument supports for the claims.

We consider the main information-argument types in research papers to be *Research issue*, *Research gap*, *Research objective*, *Research result*, and *Research contribution*. These coding categories indicate information or content types, but they also embody certain implicit claims. The research objective statement implies the following claims:

1. it is well-formed: the statement is a clear research objective, and is researchable
2. it is well-founded: founded on theory, the literature or common knowledge (i.e., prior knowledge)
3. it is well-worth researching: addresses a research gap, research issue or practical problem.

These embedded claims in a research objective statement are, we assume, common knowledge of experienced researchers, often explained in research method textbooks (e.g., Munyua, 2021), and taught in PhD programs. That the research objective is "well-formed" (i.e., clear, unambiguous and researchable) is supported by offering definitions, explanations, and elaborations. Thus, the following sequence of information-argument elements indicate support that the research objective is well-formed: `Explanation/Elaboration => Research objective`.

The following sequences provide support for the research objective's well-foundedness (i.e., founded on theory, the literature or common knowledge):

```
Theoretical framework => Research objective
Literature-summarise/generalise => Research objective
Established knowledge => Research objective.
```

The following sequences argue that the research objective is well-worth researching:

```
Topic centrality => Research objective
Research gap => Research objective.
```

Examples of these `support => claim` sequences from our corpus are available from DR-NTU (Data)—the data repository of the Nanyang Technological University.[2]

It is clear from the above `support => claim` sequences and the examples that the argument support statement clarifies or makes explicit which claim is supported, for example, whether the research objective is well-formed, well-founded or well-worth researching. In other words, the argument support statement does not merely support the claim, but also highlights which claim it supports, if there are multiple implied claims. In fact, for some researchers in the field of argumentation, the term *argument* refers to the argument support (van Eemeren et al., 2014; Tindale, 1999).

The research result statement carries the implied claim that it is valid and, for quantitative research studies, replicable. These claims are often supported by research method and data analysis statements: `Research method & Data analysis => Research result`.

In the coding scheme, *Research method* is the only category that appears in both Group 1 (a claim) and Group 2 (a support). This is because research method can be used to support the research result, but may itself need to be defended as appropriate to address the research objective. Most of the other Group 1 information-argument types carry the implication that the claim is valid or reasonable.

Although our coding scheme was developed bottom-up based on analysing sample texts, our prior experience in rhetorical structure coding has

obviously influenced us to create several categories that are similar to rhetorical steps, for example *Topic centrality* (sometimes called *Centrality/Importance of the topic*) and *General statement* (or *Topic generalisation*). Indeed, several authors in the field of argumentation have pointed out the interrelatedness of argumentation and rhetoric. Hinton (2019) noted that the "distinction [between persuasion and argument] is difficult to maintain (…) and it is clear that any investigation into how language is used to put across arguments cannot remain aloof from considerations of rhetorical impact" (p. 96).

A comparison of our information-argument coding scheme with extant rhetorical structure coding schemes (i.e., Kanoksilapatham, 2005; Kathaplia & Khoo, 2020; Swales,1990; Zhang et al., 2011) can usefully highlight the characteristics of our framework.[3] The extra categories in our coding scheme include: *Research issue, Research idea, Develop something new*, and *Novelty*. These types of claims are clearly important to researchers. Our coding scheme has more refined categories for theory: *Concept/theory/model–apply, Concept/theory/model–compare/relate*, and *Concept/theory/model–derive*. The coding scheme includes the following categories as argument support: *Theoretical mechanism*, and *Theoretical framework*. It has an extensive list of categories related to literature review, including: *Literature–derive/infer, Literature–highlight claim, Literature–counter argument*, and *Literature–difference to current research*.

The detailed coding scheme for information-argument types, and resources to support xml tagging and display of coded text in a Web browser, have been deposited in dr-ntu (Data).[4]

3. Argument structure

Information-argument structure analysis starts with tagging text units (usually sentences) with categories in the coding scheme. A text unit that is tagged with an argument claim or argument support type (i.e., Group 1 and Group 2 categories in Table 2) is referred to as an *argument discourse unit*, following Peldszus and Stede (2013).

We then extract sequences of argument claims and supports, such as `claim => claim => claim` and `support => claim => claim`. We refer to these as *argument chains*. The shortest argument chain is the two-step pattern `claim => claim` or `support => claim` called an *argument step*.

A frequently occurring sequential argument chain is referred to as a *sequential argument pattern*. We also identified sequential argument patterns where some claims in the argument chain are skipped over (explained later).

As mentioned earlier, the support for a claim may appear after the claim statement in a text (i.e., `claim <= support`). We attempted to analyse support-claim relations independent of text presentation order, thus conflating `support => claim` and `claim <= support`. Furthermore, the support for a claim may not appear immediately before or after the claim statement, but may occur some distance away. We refer to a `support => claim` relation independent of presentation order as a *logical argument step*. This kind of relation is difficult to code as it has to be inferred based on the coder's understanding. Nevertheless, it is important to try as this can uncover a deeper argument structure that may be obscured by sequential presentation order.

Argument structure can refer to the internal logical structure of propositions (called *microstructure*). Our concern is with what Freeman (2011) termed *macrostructure,* which is concerned with:

> how its component statements (together perhaps with other elements) fit together as wholes to allegedly lend support to some claim or claims. Which statements are put forward to support which other statements in the course of an argument and how, if at all, are those claims of support qualified?. (p. 1)

We use *argument structure* to refer to a set of argument elements (i.e., claims and supports) in a section of a document, linked together (by logical argument steps) into a network structure. *Argumentation* refers to the way of developing an argument structure in a research paper. Table 3 lists our definitions of these terms.

Term	Definition
Argument claim	The proposition that the author seeks to convince the reader is true, valid or reasonable.
Argument support	A proposition that supports the argument claim. It may also clarify or make explicit the claim that is supported (out of a set of possible implied claims).
Sequential argument step	A `claim=>claim` or `support=>claim` sequence in presentation order in the text.
Sequential argument chain	A list of claims and supports in order in which they are presented in the text, not necessarily spanning the whole section of a research paper.
Sequential argument pattern	A frequently occurring argument step, or sequential argument chain, possibly allowing some claims in the chain to be skipped.
Logical argument step	A `support=>claim` or `claim=>claim` relation independent of presentation order, but deemed conceptually linked.
Logical argument pattern	A frequently occurring logical argument step.
Argument structure	The set of all the argument elements (i.e., claims and supports) found in a section of a research paper, linked into a network structure by logical argument steps. The argument structure supports the coherence of the section of the paper.
Argumentation	A way or method to develop the argument structure in a research paper.

Table 3. Definitions of argument-related terms.

4. Argument structure in linguistics literature

There is substantial interest in argument structure analysis in genre studies: a literature search in the Linguistics and Language Behavior Abstracts carried out in July 2022 found 555 peer-reviewed publications published from 2010 to 2022 with the query *ab (argument* PRE/0 (structure? OR pattern? OR scheme?))*. Surprisingly, we have not identified any paper focusing on argument structure of research papers. Limiting the search to research papers (by adding the query phrase *ab((academic or research or journal) PRE/0 (paper? or article? or report?))*) retrieved only one paper that focused on hedging strategies.

Linguistics scholars have proposed different approaches for analysing argument structure. Folli and Harley (2013) proposed an analysis method from a syntactic perspective, using a set of grammatical rules to identify relations between arguments; Moretta, Feltracco, Jezek, and Magnini (2018) proposed a method from a semantic perspective, using different signifiers such as words or phrases to indicate semantic roles. Some researchers (e.g., Kohen et al., 2011; Sheinfux, 2017) investigated the structure between the syntactic and semantic level. Some studies focused on the argument flow (i.e., sequence of arguments) in student essays without categorising types of arguments (i.e., just identifying the argument claims) (Jalilifar et al., 2017; Lee, 2013; Rusfandi, 2015; Liu & Furneaux, 2015).

These approaches are unsatisfactory as they are not general methods that can be applied to any academic paper. They are akin to case studies of individual papers to understand the author's arguments, but do not try to identify patterns that can be generalised to other papers.

Erduran, Simon and Osborne (2004) and Simon (2008) used Toulmin's framework to develop indicators of the quantity and quality of argumentation in science classroom discussion, as well as to measure improvement in argumentation after changes in instructional support. Several authors (e.g., Wingate, 2012) have noted that the framework can model only micro-level argument instances, and not macro-level argument structures.

Most of the studies of academic arguments are for designing scientific research, and hardly any for scientific research papers and none that we know of for social science research. Specifically, we lack studies investigating types of arguments relating to research objectives, results and contributions.

A commonly used argument framework in the field of academic writing and genre studies is Toulmin's (2003) model of argument, which indicates that the argument claim may have qualifiers and potential rebuttals, and the argument supports include the data or evidence (forming the basis for the claim), the warrant (that authorises the step from the data to making the claim), and backing (that provides support for the warrant). However, in a research paper, some of these elements are not explicitly stated or are implied (van Eemeren et al., 2014). For example, the warrant is often implied prior knowledge in the research community, and is provided only when the research method is new or when the warrant is controversial. We include data, warrant, and backing under *support*.

5. Types of sociology research

Social science research can be divided into different types, the simplest division being between quantitative and qualitative studies. Our pilot study (Cheng, 2020) identified five types of sociology research: *Investigative research*, *Development & Evaluation research*, *Descriptive research*, *Historical analysis*, and *Identification research*. They are defined in Table 4. These types of research have quite different end goals and we expected the authors to use quite different information and argument structures. Because of the small number of papers belonging to the last three research types (Descriptive research, Historical analysis, and Identification research), these were combined into

one Descriptive research category for the analysis, as most of them used qualitative research methods.

The different research types reflect different epistemic paradigms, seeking different kinds of knowledge and presenting different kinds of arguments. Investigative research, which investigates a research relation (usually a cause-effect relation) between two concepts or entities, typically employ quantitative research methods. They can be characterised as adopting a positivist paradigm of research, rather than a postmodernist, constructivist, interpretivist or critical theory paradigm (Lincoln et al., 2011; Schwandt, 1998), which is more likely to characterise Descriptive research.

The research type can be identified solely from the research objective and research method statements, which contain different indicative keywords. For example, as Investigative research seeks to investigate a causal or association relation between two concepts, its indicative keywords include *affect*, *influence*, and *improve*. Using these definitions and indicative keywords, even undergraduate coders were able to identify the types of research quite accurately. Intercoder reliability results are provided in the next section.

Research Type	Definition
Investigative research	The research seeks to investigate a causal or associative relation between concepts or entities, often by using quantitative research methods (e.g., questionnaire survey).
Development & Evaluation research	The research seeks to develop a complex concept (i.e., theory, method or system), or to evaluate such a concept, usually using a quantitative research method.
Descriptive research	The research seeks to explicate a phenomenon or an event, often using qualitative research methods (e.g., ethnography).
Historical analysis	The research seeks to explicate a change in a particular social, economic, cultural and institutional phenomenon (including the development of an entity) over a period of time, often based on the analysis of historical data.
Identification research	The research seeks to identify a subtype or instance of a concept (e.g., pattern), usually using a qualitative research method.

Table 4. Definition of each type of sociology research.

6. Method

6.1. Scope of the study

Seventy papers were annotated with the coding scheme: 30 reporting Investigative research, and 20 each for Development and Evaluation research and Descriptive research. The information-argument profiles for the three types of sociology research were identified, characterised by frequency distributions and percentages for the different types of claims, supports, and information. Significant differences across the three types of sociology

research were identified. Sequential association rule mining (a data mining method) was used to identify frequently occurring patterns of claims and supports, considering the presentation order in the text (referred to as *sequential argument patterns*). Frequently occurring `support => claim` patterns occurring in adjacent sentences, but irrespective of presentation order, were also identified (referred to as *logical argument patterns*).

6.2. Corpus

The corpus for this study comprises research papers from ten journals with the highest impact factor in InCites Journal Citation Reports for sociology, listed in Table 5. The articles were published in the late 2015 or early 2016 volumes of the journals. Only articles reporting research that involved data analysis were included. Journal articles that report literature surveys or philosophical/theoretical discussions were excluded.

Code	Journal title	Sample
S01	American Journal of Sociology	6
S02	Annals of Tourism Research	9
S03	Cornell Hospitality Quarterly	7
S04	European Sociological Review	6
S05	Gender Society	6
S06	Information Communication Society	8
S07	Journal of Marriage and Family	7
S08	Social Networks	10
S09	Qualitative research	5
S10	American sociological review	6
Total		**70**

Note: The different sample sizes is because Development and Evaluation research and Descriptive research papers are less common, and additional papers in these two categories were sampled from journals that published more of such research papers.

Table 5. Number of research papers selected from sociology journals.

6.3. Coding of information-argument types

We carried out manual annotation of argument elements in the text using oXygen XML Editor version 18.1 (a software for annotating and editing XML files). The coding was usually at the sentence level. Only one code was assigned to each text unit. If multiple argument types were possible (depending on the reader's interpretation), the coder had to select the best code. Figure 1 gives an example of annotated text. Argument claims are

shown with red background colour; argument supports with yellow background.

6.4. Sequential association rule mining

To analyse sequential argument patterns, we carried out sequential association rule mining using IBM/SPSS Modeler version 18.1 to identify common sequences of argument elements that occurred in at least 20% of the papers. We analysed sequences of two to five adjacent elements, as well as sequences of up to five elements that are not necessarily adjacent (i.e., with any number of intervening elements), thereby allowing some claims in an argument chain to be skipped.

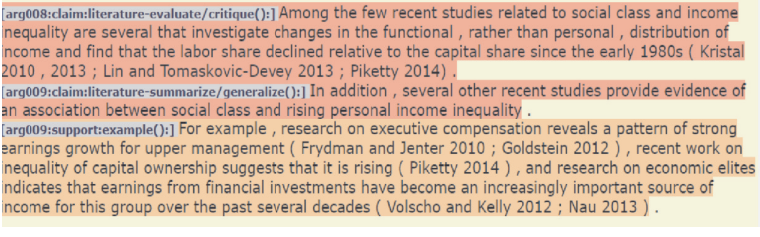
[arg008:claim:literature-evaluate/critique():] Among the few recent studies related to social class and income inequality are several that investigate changes in the functional , rather than personal , distribution of income and find that the labor share declined relative to the capital share since the early 1980s (Kristal 2010 , 2013 ; Lin and Tomaskovic-Devey 2013 ; Piketty 2014) .
[arg009:claim:literature-summarize/generalize():] In addition , several other recent studies provide evidence of an association between social class and rising personal income inequality .
[arg009:support:example():] For example , research on executive compensation reveals a pattern of strong earnings growth for upper management (Frydman and Jenter 2010 ; Goldstein 2012) , recent work on inequality of capital ownership suggests that it is rising (Piketty 2014) , and research on economic elites indicates that earnings from financial investments have become an increasingly important source of income for this group over the past several decades (Volscho and Kelly 2012 ; Nau 2013) .

Figure 1. Example argument element coding.

In addition to mining sequential argument patterns, we also analysed logical argument patterns. It is difficult to analyse logical argument patterns because the argument supports may be presented a few sentences before or after the claim. Furthermore, the argument support may not be explicitly expressed, but implied by adding citations in parenthesis, or by tacit appeal to general knowledge. We focused our analysis on explicit `support => claim` and `claim <= support` sentence pairs that have a clear relationship. We view this explicit `support => claim` relation as a kind of coherence relation, where the reader readily makes out a logical or conceptual relation between the sentences.

6.5. Inter-coder reliability

In this study, argument element coding was carried out by the two authors. The first author annotated the text first, which was reviewed and revised by the second author (deemed the more experienced researcher). This two-phase coding compensates for individual oversights and biases, as the second coder would be well aware of the first coder's perspective and will take that into

consideration before changing a code. We found that coding of argument claims and supports has to be done by experienced researchers: it is difficult to train undergraduate and graduate students to do this analysis (Cheng, 2020).

In total, 21% of the codes were changed by the second coder. Intercoder analysis of the changes made by the second author reveals differences of interpretation, but also highlights the assumptions underlying the final coding. The error analysis suggests that the first author was looking at more superficial features (word markers that suggest the code), whereas the second author was looking at deeper intentions (that may need to be inferred). Some of the coding conflicts reflect difficulties with the following types of argument claims and supports:

1. *Concept/theory/model-related* claims and *Literature review-related* claims. This reflects a common confusion of whether a cited paper is related to a particular concept/theory/model;
2. *Research result*, *General result* and *Research contribution/recommendation* claims. This reflects the common difficulty of understanding general (broader) and detailed (narrower) research findings.

A small intercoder reliability test was carried out on this coding. A sample of 10 articles from the corpus was coded by both authors. Cohen's kappa coefficient of 0.782 was obtained by analysing 261 argument and non-argument elements. Cohen's kappa measures the proportion of agreement after removing chance agreement between two coders, and our value of 0.782 can be considered substantial agreement. Some authors considered kappa values of 0.21 to 0.40 to represent fair agreement, 0.41 to 0.60 moderate agreement, and 0.61 to 0.80 substantial agreement (Cohen, 1960; Landis & Koch, 1977; Viera & Garrett, 2005).

An analysis of the conflicts found that the following pairs of argument types tended to be confused:

1. *Topic centrality* and *General statement* claims. Both are broad overview claims, and whether there is an intention to highlight the importance of the research topic is a judgement call.
2. *Literature-evaluate/critique*, *Concept/theory/model/-evaluate/critique*, and *Research gap* claims. These three claims may overlap: a research gap may be highlighted by critiquing previous papers and theories.

3. *Literature review-related* claims versus *Literature review-related* supports. For example:
    - *Literature-highlight claim* is considered a claim, whereas *Cited authors claim/opinion* can be coded as supporting a claim, such as a *General statement*.
    - *Literature-summarise/generalise* is considered a claim, but it may be coded as *Explanation/elaboration* and *Established knowledge* to support a claim. This highlights the issue of whether to code a cited paper as a support or a claim.

As for identifying the five types of research, six undergraduate students were recruited to do the analysis. The average inter-coder reliability scores (percentage agreement between coders, and Cohen's Kappa) were substantial at 0.72 and 0.60, showing that it not difficult to distinguish between the research types.

7. Findings

7.1. Information-argument profile across the three research types

One-way Analyses of Variance (ANOVA) were performed to identify significant differences in the information-argument type frequencies across the three types of research. Table 6 lists the information-argument types with significant differences (α=0.05).

The information-argument profile for the three research types can be characterised as follows:

1. For Investigative research: *Research gap*, *Research hypothesis, Research method*, *Research purpose* (especially *Research objective*), and *Research finding* (including research result and research contribution) elements had the highest frequencies.
2. For Development and Evaluation research: *Concept/theory/model-related* elements (especially *Concept/theory/model-derive* and *Concept/theory/model-evaluate/critique*) had the highest frequencies. The non-argumentative *Descriptive information* element occurred on average about six times, especially *Concept/theory/model-details* and *Outline the structure*.

3. For Descriptive research: the *Research result* element had a slightly higher frequency. *Descriptive information* elements occurred about five times on average.

This reflects different information and argument strategies needed for different research types:

1. Investigative research papers usually argue for a research gap to justify the research objective. Research finding (particularly research contribution) is also claimed in the Introduction.
2. Development and Evaluation research papers have to argue for the novelty (idea) for the theory or method proposed. As most of these studies in sociology sought to develop a new theory or improve an old theory, it is not surprising that there were more theory-related elements. An outline of the paper structure was often found at the end of the Introduction section, suggesting that there is no standard structure for the whole paper.
3. Descriptive research papers do not have a distinctive argument strategy. However, they have more descriptive elements and contextual information. A summary of the research results is often provided. An outline of the paper structure is often provided at the end of the Introduction section.

Information-argument type		Type of research		
		Investigative (N=30)	Development and Evaluation (N=20)	Descriptive (N=20)
Main research claims	-	-	-	-
Research gap	Mean	**2.03**	1.00	0.80
	SD	1.65	1.97	1.11
Research purpose claims	Mean	**4.57**	2.05	2.45
	SD	1.91	1.70	2.21
Research objective	Mean	**2.77**	1.45	1.80
	SD	1.63	1.19	1.36
Research hypothesis	Mean	**0.90**	0.16	0.00
	SD	1.09	0.50	0.00
Research method	Mean	**1.87**	0.55	0.80
	SD	2.16	0.89	1.06
Research finding claims	Mean	**3.00**	2.05	1.80
	SD	3.41	3.52	2.07
Research result	Mean	1.07	1.05	**1.30**
	SD	2.08	1.67	1.95
Literature review-related claims	-	-	-	-
Literature-summarise/generalise	Mean	**2.53**	1.50	0.80
	SD	2.11	1.99	0.83
Concept/theory/model-related claims	Mean	3.23	**7.90**	3.20
	SD	3.48	7.76	3.76
Concept/theory/model-derive	Mean	0.07	**1.00**	0.10
	SD	0.25	1.34	0.31
Concept/theory/model-evaluate/critique	Mean	0.17	**1.45**	0.05
	SD	0.59	2.42	0.22
Descriptive information	Mean	0.97	**5.40**	**4.75**
	SD	1.61	5.13	5.16
Concept/theory/model-details	Mean	0.10	**1.45**	0.15
	SD	0.55	3.19	0.67
Outline the structure	Mean	0.63	**2.30**	1.50
	SD	1.40	2.64	1.91

Note: Significant differences across the research types (α=0.05).

Table 6. Common information-argument types in the introduction sections (frequency distribution).

7.2. Opening move, middle game and concluding move

We analysed the information-argument patterns in terms of which information-argument types tended to appear at the beginning, middle and end of the Introduction section - in other words, the opening move, middle game and concluding move (see Table 7).

For all types of research, the Introduction section often opened with *Topic centrality* (20% to 40% of the sample) or *General statement* (20% to 30%). Descriptive research sometimes opened with non-argumentative Descriptive or Context information (20% to 25%).

At the end of the Introduction section, all research types had *Outline the structure* as the final element: more often for Development and Evaluation research and Descriptive research (40% to 45% of sample) than for *Investigative research* (23%). Investigative research and Descriptive research concluded with *Research result* and *Research contribution/recommendation* 20% to 37% of the time. In contrast, Development and Evaluation research made *Concept/theory/model-related* claims about 20% of the time.

In the middle of the Introduction sections, Investigative research and Descriptive research typically made use of this sequence of elements: *Literature-summarise/generalise* … *Research gap* … *Research objective*. Descriptive research sometimes added a *Concept/theory/model-apply* claim. However, *Development and Evaluation research* can have any kind of elements (i.e., there is no dominant pattern).

Type of research	Position	Common argument elements
Investigative research	Begin	Topic centrality or General statement
	Middle	Literature-summarise/generalise …
		Research issue or Research gap or Research question … Research objective …
	End	Research contribution/recommendation
	Inserted element	- *Research method*:
		anywhere in the Introduction section
		- *Concept/theory/model-define*:
		anywhere in the Introduction section
Development and Evaluation research	Begin	Topic centrality or General statement or
		(Topic centrality -> Concept/theory/model-define)
	Middle	(any argument element)
	End	Outline the structure
	Inserted element	- Any argument element: in the middle of the pattern
Descriptive research	Begin	Topic centrality or Descriptive information or Context information
	Middle	Literature-summarise/generalise … Research gap (Concept/theory/model-apply)? … Research objective …(Concept/theory/model-apply)? …
	End	Outline the structure
	Inserted element	- *Research method:* before or after the Research objective
		- *Concept/theory/model-define*: anywhere before the Research objective

Table 7. Common argument elements at the beginning, middle and end of Introduction sections, across the research types.

7.3. Sequential argument patterns

We analysed the sequences of information-argument types starting with 2-step and 3-step adjacent patterns. We then analysed sequences of information-argument types with any number of intervening elements (i.e., allowing some claims/supports in an argument chain to be skipped). Sequential argument patterns in the Introduction sections are too numerous to list here. They are listed in the first author's PhD thesis (Cheng, 2020).

We attempted to derive one basic or overall sequential argument pattern for the Introduction section, by linking and combining the frequent sequences of adjacent and non-adjacent elements. The assumption is that researchers doing a particular type of research should be familiar with these basic argument patterns. In practice, researchers are expected to modify or extend the basic argument pattern to present a strong case for their research. The basic argument patterns that we derived for each research type can be considered a theory underlying the argument structure of each research type. We summarise the basic sequential argument patterns separately for each research type below. Examples of the argument patterns from our corpus are available from dr-ntu (Data).[5]

7.3.1. Investigative research

The sequential pattern mining did not detect any pattern with a high frequency (over 50%). The most common pattern was `Research objective -> Research method -> Research result` (40%). Moreover, *General statement* was the first argument claim in about 30% of the Introduction sections; and *Topic centrality* in about 20%. The results indicate the following basic argument pattern in the Introduction sections:

```
(General statement or Topic centrality) …
Literature-summarise/generalise …
(Research issue or Research gap or Research
question) … Research objective … Research
contribution/recommendation.
```

In addition, we found four other more specialised patterns:

1. The basic argument pattern, with additional literature review to support *Research gap*. A *Literature-summarise/generalise* claim may be

inserted before or after a *Research gap* claim, or between two *Research gap* claims.

2. Multiple *Research gap* claims, supported by literature review.
3. Extended literature review, leading to the *Research objective*.
4. Additional *Research method* element, with the *Research method* appearing after the *Research objective* or between two *Research objective* claims.

7.3.2. Development and Evaluation research

The sequential argument patterns were more elaborate than those for Investigative research, as Development and Evaluation research papers had more *Concept/theory/model-related* claims (20%). A *Concept/theory/model-evaluate/critique* claim (20%) was often the final element. Thus, the following basic sequential argument pattern was derived:

(Research objective or **Concept/theory/model-define/derive/apply)**…
Research idea/approach … Research objective … (Research result or **Concept/theory/model-evaluate/critique** or Research contribution/recommendation).

In the above pattern, *Research idea/approach* replaces the *Research gap* in the argument pattern derived for Investigative research, because Development and Evaluation research generally proposes a new idea or approach, rather than address a *Research gap* in the literature.

7.3.3. Descriptive research

Non-argumentative elements (i.e., *Descriptive information* and *Context information*) played an important role in Descriptive research. The following elements each comprise 10% to 15% of the Introduction sections:

1. *Descriptive information* (especially *Data context*, and *Outline the structure*);
2. *Context information* (especially *Practical background*, and *Historical background*)

3. *Literature review-related claims*;
4. *Concept/theory/model-related claims.*

This indicates that descriptive research papers have more descriptions of qualitative data, which is based on a theory/model/framework and previous research.

The analysis of sequential argument patterns did not find any pattern with a high frequency (over 50%). The results indicated the following overall/basic argument pattern in the Introduction sections:

(Topic centrality or **Descriptive information** or **Context information**) … Literature-summarise/generalise … Research gap … (**Concept/theory/model-apply**)? … Research objective … (**Concept/theory/model-apply**)? … Outline the structure.

Concept/theory/model-define claim can be inserted anywhere before the *Research objective* claim; the *Research method* claim may be inserted before or after the *Research objective* claim.

7.4. Logical argument steps in the Introduction sections

We also analysed logical argument steps irrespective of sequential order in the text. However, this analysis was limited to argument elements occurring in adjacent sentences, so that the relation between argument claim and argument support is direct and explicit. Thus, only a relatively small number of explicit argument steps were found in each Introduction section.

We found that authors exerted considerable effort to support the *Research objective* with *Research gap* (18 of 70 papers) or *Research motivation/justification* (13) or *Topic centrality* (7). Investigative research and Descriptive research papers tended to support the *Research objective* with a *Research gap* claim. Investigative research also used *Topic centrality*, whereas Descriptive research used *Practical problem* to support a *Topic centrality*.

Two common logical argument steps were found in all research types:

1. Literature-summarise/generalise => Research gap (11 of 70 papers)

2. `Explanation/elaboration => Topic centrality` (10).

This indicates that authors typically support a research gap by summarising previous studies, and support the importance (centrality) of the research topic with additional explanations. The common logical argument steps across the research types are shown in Figure 2.

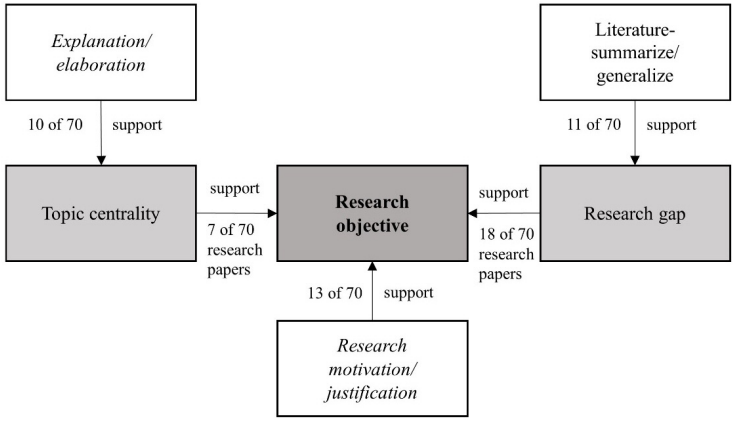

Figure 2. Common logical argument steps across the types of research.

8. Conclusion

We have analysed the types of information and argument claims and supports used in the Introduction sections of sociology research papers, as well as common patterns of claims and supports. We identified three types of sociology research: Investigative research, Development and Evaluation research, and Descriptive research. A coding scheme was developed for information-argument types, divided into those that carry or imply an argument claim (Group 1), those that play the role of argument support (Group 2), descriptive information types (Group 3) and context information types (Group 4).

We identified significant differences in information-argument profile (i.e., distribution of information-argument types) among the three types of research papers, as well as differences in sequential argument patterns and logical argument patterns.

The Introduction sections reflect different information and argument

strategies across the three research types: Investigative research papers usually argue for a research gap to support the research objective. Research contribution is also often claimed in the Introduction. Development and Evaluation research papers typically argue for the novelty of the proposed theory or method. There are also more theory-related elements to support the new theory or method, and more descriptive elements to give details of the theory. Descriptive research papers do not have a distinctive argument strategy, but have more descriptive elements and context information, especially at the beginning of the Introduction. A summary of the research results is often provided. Both Development and Evaluation research and Descriptive research papers have *Outline the structure* at the end of the Introduction nearly half the time, suggesting that there is no standard paper structure, and a paper outline is helpful to the reader.

The results of sequential pattern mining of Introduction sections indicate the following basic argument pattern in the Introduction sections: `(General statement or Topic centrality) … Literature-summarise/generalise … (Research issue or Research gap or Research question) … Research objective … Research contribution/recommendation`. Development and Evaluation `research` and Descriptive research papers may modify the basic argument pattern with *Concept/theory/model-related* claims. Development and Evaluation research papers often replace *Research gap* with *Research idea/approach*, whereas Descriptive research papers open with *Descriptive information* (especially *Data context*) or *Context information* (especially *Practical background*, and *Historical background*) about a quarter of the time.

From an analysis of the logical argument steps irrespective of sequential order, but presented in adjacent sentences, we derived a typical argument structure for the Introduction section, represented in Figure 2.

Our method of argument element coding using XML tags has allowed us to carry out quantitative analysis and sequential pattern mining to gain insights into the information-argument structure of academic texts. We expect that applying this analysis method to other sections of the research paper (i.e., literature review, results, discussion/conclusion) will yield more interesting results. We believe our results for Introduction sections would generally hold for other social science fields, though this is to be confirmed in subsequent studies. Scientific and engineering disciplines may have other research types that have evolved other kinds of information-argument structure.

Our information-argument structure analyses were carried out mainly at the sentence level to find patterns in what Freeman (2011) termed *argument macrostructure*. In complementary work, we are analysing the information microstructure of argument claims and supports at the level of concepts, entities and semantic roles, to identify information links between claims and their supports. An initial study of how *comparison* information structures and *cause-effect* information structures are linked together to support the validity of the cause-effect claim (in Research result) has been reported in Cheng and Khoo (2021).

Acknowledgement

This study was funded partly by the Singapore Ministry of Education research grant MOE2015-1-TR05.

Article history:
Received 10 February 2022
Received in revised form 30 August 2022
Accepted 11 September 2022

References

Cheng, W. -N. (2020). *Argument and information structures in sociology research papers: Analysis of the abstract and introduction sections*. Doctoral thesis, Nanyang Technological University, Singapore.

Cheng, W. -N., & Khoo, C. S. G. (2021). Information structures in sociology research papers: Modeling cause-effect and comparison relations in research objective and result statements. *Journal of the American Society for Information Science and Technology*, *72*(11), 1367-1385. <https://doi.org/10.1002/asi.24486>

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, *20*(1), 37-46.

Folli, R., & Harley, H. (2013). The syntax of argument structure: Evidence from Italian complex predicates. *Journal of Linguistics*, *49*(1), 93-125. <https://doi.org/10.1017/S0022226712000072>

Freeman, J. B. (2011). *Argument structure: Representation and theory*. Springer.

Erduran, S., Simon, S., & Osborne, J. (2004). TAPing into argumentation: Developments in the use of Toulmin's argument pattern in studying science discourse. *Science Education*, *88*(6), 915-933. <https://doi.org/10.1002/sce.20012>

Hinton, M. (2019). Language and argument: A review of the field. *Research in Language*, *17*(1), 93-103. <https://doi.org/10.2478/rela-2019-0007>

Jalilifar, A., Keyvan, M., & Don, A. (2017). The effect of writing prompts on developing argumentative essays of Iranian university students of English. *Journal of Modern Languages*, *27*, 48-76. <https://ejournal.um.edu.my/index.php/JML/article/view/6847>

Kanoksilapatham, B. (2005). Rhetorical structure of biochemistry research articles. *English for Specific Purposes, 24*(3), 269-292. <https://doi.org/10.1016/j.esp.2004.08.003>

Kathpalia, S. S., & Khoo, C. S. G. (2020). Generic structure and citation functions in introductions of biological science articles in English-medium international journals. *Ibérica, Journal of European Association of Languages for Specific Purposes*, *40*, 123-148. <https://revistaiberica.org/index.php/iberica/article/view/61>

Kohen, F., Milsark, G., & Martin, N. (2011). Effects of syntactic and semantic argument structure on sentence repetition in agrammatism: Things we can learn from particles and prepositions. *Aphasiology*, *25*(6-7), 736-747. <https://doi.org/10.1080/02687038.2010.537348>

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*(1), 159-174.

Lee, S. H. (2013). Argument structure as an interactive resource by undergraduate students. *Linguistics and the Human Sciences*, *9*(3), 277-306. <https://doi.org/10.1558/lhs.v9i3.277>

Lincoln, Y. S., Lynham, S. A., & Guba, E. G. (2011). Paradigmatic controversies, contradictions, and emerging confluences. In N. K. Denzin & Y. S. Lincoln (Eds.), *The Sage handbook of qualitative research* (pp. 97-128). Sage.

Liu, X., & Furneaux, C. (2015). Argument structures in Chinese university students' argumentative writing: A contrastive study. *English Text Construction*, *8*(1), 65-87. <https://doi.org/10.1075/etc.8.1.03liu>

Munyua, J. M. (2021). *Basics of research ang academic writing: A text book for research students and lectures.* Exceller Books.

Moretta, F. D., Feltracco, A., Jezek, E., & Magnini, B. (2018). Designing a methodology for semantic type tagging of argument positions. *IJCoL: Italian Journal of Computational Linguistics*, *4*(2), 57-72. <https://doi.org/10.4000/ijcol.508>

Peldszus, A., & Stede, M. (2013). From argument diagrams to argumentation mining in texts: A survey. *International Journal of Cognitive Informatics and Natural Intelligence*, *7*(1), 1-31. <https://doi.org/10.4018/jcini.2013010101>

Rusfandi. (2015). Argument-counterargument structure in Indonesian EFL learners' English argumentative essays: A dialogic concept of writing. Regional Language Centre. *RELC Journal*, *46*(2), 181-197. <https://doi.org/10.1177/0033688215587607>

Schwandt, T. A. (1998). Constructivist, interpretivist approaches to human inquiry. In N. K. Denzin & Y. S. Lincoln (Eds.), *The Landscape of Qualitative Research* (pp. 221-259). Sage.

Sheinfux, L. H., Melnik, N., & Wintner, S. (2017). Representing argument structure. *Journal of Linguistics*, *53*(4), 701-750.

Simon, S. (2008). Using Toulmin's argument pattern in the evaluation of argumentation in school science. *International Journal of Research & Method in Education*, *31*(3), 277-289. <https://doi.org/10.1080/17437270802417176>

Swales, J. (1990). *Genre analysis: English in academic and research settings*. Cambridge University Press.

Tindale, C. W. (1999). *Acts of arguing: A rhetorical model of argument*. SUNY.

Toulmin, S. (2003). *The uses of argument* (2nd ed.). Cambridge University Press.

van Eemeren, F. H., Garssen, B., Krabbe, E.C.W., Snoeck Henkemans, A.F., Verheij, B., & Wagemans, J.H.M. (2014). *Handbook of argumentation theory*. Springer.

Viera, A. J., & Garrett, J. M. (2005). Understanding interobserver agreement: The kappa statistic. *Family Medicine*, *37*(5), 360-363.

Wingate, U. (2012). 'Argument' helping students understand what essay writing is about. *Journal of English for Academic Purposes, 11*(2), 145-154. <https://doi.org/10.1016/j.jeap.2011.11.001>

Zhang, L., Kopak, R., Freund, L., & Rasmussen, E. (2011). Making functional units functional: The role of rhetorical structure in use of scholarly journal articles. *International Journal of Information Management, 31*(1), 21-39. <https://doi.org/10.1002/meet.14504701215>

Dr Wei-Ning Cheng received her PhD in Information Studies from the Wee Kim Wee School of Communication and Information, Nanyang Technological University, Singapore. Her research interests are in argument and information structures analysis, knowledge organisation and graph visualisation.

Dr Christopher Khoo is an Associate Professor in the Wee Kim Wee School of Communication and Information, Nanyang Technological

University, Singapore. He teaches courses in knowledge organisation, and data analytics. His research interests are in knowledge graph, text mining, and multidisciplinary study of academic writing & thinking.

NOTES

[1] By "logical" we do not mean logical entailment as used in formal logic (propositional logic or first order logic), but that the `support => claim` links are independent of the surface (sequential) presentation in the text. The support and claim need not even appear close together in the text. Thus, a *logical argument structure* is a representation of the conceptual structure that is hypothesised to be constructed in the reader's mind after reading the research paper.

[2] https://researchdata.ntu.edu.sg/file.xhtml?fileId=96819&version=2.0

[3] https://researchdata.ntu.edu.sg/file.xhtml?fileId=96821&version=2.0

[4] https://doi.org/10.21979/N9/LD3EBQ

[5] https://researchdata.ntu.edu.sg/file.xhtml?fileId=96819&version=2.0