Received 21 August 2022
Revised 8 November 2022
Accepted 11 December 2022

Rhetorical structure parallels research topic in LIS articles: a temporal bibliometrics examination

Wen Lou
Department of Information Management, Faculty of Economics and Management, East China Normal University Minhang Campus, Shanghai, China and Key Laboratory of Advanced Theory and Application in Statistics and Data Science (East China Normal University), Ministry of Education of China, Shanghai, China
Jiangen He
The University of Tennessee Knoxville, Knoxville, Tennessee, USA
Qianqian Xu, Zhijie Zhu and Qiwen Lu
East China Normal University, Shanghai, China, and
Yongjun Zhu
Yonsei University, Seoul, Republic of South Korea

Abstract

Purpose – The effectiveness of rhetorical structure is essential to communicate key messages in research articles (RAs). The interdisciplinary nature of library and information science (LIS) has led to unclear patterns and practice of using rhetorical structures. Understanding how RAs are constructed in LIS to facilitate effective scholarly communication is important. Numerous studies investigated the rhetorical structure of RAs in a range of disciplines, but LIS articles have not been well studied.

Design/methodology/approach – In this study, the authors encoded rhetorical structures to 2,216 articles in the *Journal of the Association for Information Science and Technology* covering a period from 2001 to 2018 with the approaches of co-word analysis and visualization. The results show that the predominant rhetorical structures used by LIS researchers follow the sequence of Introduction-Literature Review-Methodology-Result-Discussion-Conclusion (ILMRDC).

Findings – The authors' temporal examination reveals the shifts of evolutionary pattern of rhetorical structure in 2008 and 2014. More importantly, the authors' study demonstrates that rhetorical structures have varied greatly across research areas in LIS community. For example, scholarly communication and scientometrics studies tend to exclude literature review in articles.

Originality/value – The present paper offers a first systematic examination of how rhetorical structures are used in a representative sample of a LIS journal, especially from a temporal perspective.

Keywords Literature review, Scholarly communication, JASIST, Research articles, Rhetorical structure, Writing pattern

Paper type Research paper

Aslib Journal of Information Management
Vol. 76 No. 3, 2024
pp. 416-434

2050-3806
DOI 10.1108/AJIM-08-2022-0370

This research, an extended paper of ISSI 2021 research-in-progress paper entitled "The Effects of Rhetorical Structure on Citing Behavior in Research Articles," is funded by Shanghai Planning Office of Philosophy and Social Science Youth Project (No: 2021ETQ002), National Social Science Foundation (No: 21FTQB009), China Scholarship Council (No: 202106145010) and Open Research Fund Program of KMKS, National Press and Publication Administration (No: 2021KMKS01)], and Jiangen He wishes to thank the support of the National Science Foundation (No: 2122691).

Introduction

To effectively communicate scientific knowledge and meet the expectation of audiences, researchers typically follow conventional structures in the high-level organization of a research article (RA). To facilitate scholarly communication in the context of globalization and international collaboration in science, conventional rhetorical structures are used by scientists in the field of their scientific interest (Kanoksilapatham, 2005). The most well-known structure is the IMRaD format, in which RAs are typically divided into sections covering the Introduction, Methods, Results and Discussion (IMRD) (Swales, 1990). Authors can construct the sections themselves according to what content ought to be included and what inessential details may be omitted. Essentially, building a rhetorical structure in academic writing is a linguistic task of constructing a series of "moves" to persuade readers of the key messages. Article structure is not uniform because the key messages that authors need to deliver for effective scholarly communication may vary across disciplines and fields. Journals and publishers also influence article structure by mandating or guiding structured abstracts and article sections.

As a highly interdisciplinary field, library and information science (LIS) may exhibit unusually diverse rhetorical structures in RAs. Understanding the conventional rhetorical structures in LIS is necessary to inform authors and reviewers about the communicative effectiveness of structures in the community. Although the rhetorical structures of abstracts and full-text articles have been examined in many different fields, such as agriculture (Milagros Del Saz Rubio, 2011), biochemistry (Kanoksilapatham, 2005), computer science (Soler-Monreal *et al.*, 2011), applied linguistics (Yang and Allison 2004) and cross-disciplinary study (Lin and Evans, 2012), the rhetorical structures of LIS articles have not been fully examined. A few studies have conducted rhetorical move analysis of abstracts in LIS areas (Rashidi and Meihami, 2018), but very few have analyzed full-text of RAs. Given this, it is worthwhile to examine the rhetorical structures of LIS articles to facilitate more effective scholarly communication. Moreover, the research landscape of LIS has changed rapidly in recent decades. In response to these changes, authors in LIS may adjust the way they organize their articles at a high level or adopt rhetorical structures that are conventional in other fields. Thus, investigating the evolution of rhetorical structures in LIS may improve our understanding of the current rhetorical structures commonly used by authors in this evolving field.

Although the main purpose of rhetorical structure analysis conducted by linguists is to educate non-native speakers in academic writing (Rashidi and Meihami, 2018), rhetorical structure analysis can lead to a better understanding of the process of scholarly communication, which may contribute to some central problems in scientometrics studies. Scientometrics addresses the quantitative aspects of science study (Mingers and Leydesdorff, 2015). A thorough analysis of how scientific knowledge is communicated from the perspective of rhetorical structures may benefit further quantitative studies of science and scientific research. With the availability of open scholarly datasets and computational approaches for analyzing big scholarly datasets (Jamali and Nabavi, 2015; Saier and Färber, 2020), full-text analysis of scientific publications became an emerging area in the community of scientometrics. Unlike most of the existing studies, we analyzed the rhetorical structures of full text instead of abstracts of scientific publications. A better understanding of how text is organized and structured in scientific publications derived from our rhetorical structure analysis can provide a framework for full-text and citation context analysis (Bertin *et al.*, 2016; Zhang *et al.*, 2013).

An effective structure of an article is essential in delivering its core message. Since the choice of an appropriate rhetorical structure contributes to the effective delivery of research content, authors' selection of rhetorical structure may be partly determined by the research content and the expectation of the audience. The association between linguistic patterns and research content has been examined by linguistics, as has the cross-discipline difference in rhetorical structures, but, to the best of our knowledge, there is no study providing examination and evidence concerning the relationship between the research content and rhetorical structure of RAs. There are some norms

of rhetorical structures for scientific writing in some of scientific domains, including implicit norms that are widely accepted by the community or explicit norms suggested by publication venues. However, the norms for interdisciplinary domains can be confusing to the authors, reviewers and audiences of scientific publications. The interdisciplinary nature of LIS leads to diverse research topics and methods. Among the most prominent research topics are information retrieval, library services, scientific communication and information seeking (Milojević *et al.*, 2011). Due to the lack of studies, the norms in LIS are not clearly known and we don't know if the norms differ across various research topics either. We aim to use a highly interdisciplinary domain, LIS, as a case study to shed light on how the norms form over time and how they are associated with research content. Understanding how interdisciplinarity shapes the rhetorical structure in scientific writing is critical for scientometrics researchers to build an accurate mapping of scholarly communication in the current increasingly interdisciplinary era.

Given these research gaps, this paper addresses the following research questions to address our contributions, including characterizing the norms of using rhetorical structures in LIS and the formation of the norms and understanding the relationship between research content and rhetorical structures:

RQ1. What are the main rhetorical structures used to organize RAs in LIS?

RQ2. How did the distribution of commonly used rhetorical structures in LIS evolve over time?

RQ3. Do the rhetorical structures vary across different research topics in LIS?

Related work

There has been remarkable scholarly interest in the structural features of RAs in recent decades. SWales developed a groundbreaking framework for analyzing RAs, the "Create-a-Research-Space" (CARS) model (Swales, 1990). SWales' move-based approach to genre analysis has enhanced our understanding of rhetorical structure of RA that follow the conventional IMRaD (Introduction, Methods, Results and Discussion) structure (hereafter as IMRD). Many studies stimulated by SWales' approach have analyzed the rhetorical structure of individual sections of RAs in various disciplines. Tessuto (2015) found how structure reflects legal RAs by using the prototypical IMRD model to identify the discourse structure in tests across a representative sample of genres. Particularly, the abstract and introductory sections have been focuses of attention.

Some common moves/structures of abstracts have been identified. Loré (2004) found two major types of rhetorical structures, the IMRD type and the CARS type by an analysis of RA abstracts from linguistics journals. Pho's (2008) study of abstracts in applied linguistics and educational technology has shown there are three obligatory moves in these two fields – presenting the research, describing the methodology and summarizing the results. Some study dealt with interpersonality in RA abstracts by analyzing interactional metadiscourse, the degree of interpersonality realize by hedges, boosters and attitude markers diminishes over time (Gillaerts and Van de Velde, 2010). The introductory sections have been examined extensively by genre-based analysis to understand rhetorical structures and the relationship between function and structure in RAs. The CARS model may not adequately account for the structure of all RA introductions. Anthony's (1999) study of introductions in the field of software engineering found that the CARS model doesn't characterize some essential features of introductions, such as research evaluation. To address the limitation of the CARS models, Samraj (2002) proposed a revised CARS model with a greater degree of flexibility in embedding. New approaches have been applied with CARS model for rhetorical analysis. Milagros Del Saz Rubio (2011), for instance, used CARS model and Hyland's and Tse (2004) model of metadiscourse to conduct a pragmatic two-level rhetorical analysis of moves and steps of introductions. Besides the extensive studies of abstracts and introductions, various other parts of RAs have been investigated about

their overall organization. Brett (1994) analyzed the communicative moves in the results section of RAs in sociology. Basturkmen (2012) examined steps in 'commenting on results' moves in dentistry RAs to understand how arguments about the meaning and significance of results in the discussions were constructed. It is also important to understand the distribution of commonly used rhetorical structures over time. Many studies only addressed the change in the impact of papers over time. Lu's study (2017) investigates how one highly cited essay, such as Hirsch's "h-index" article (H-article) published in 2005, has been cited by other articles. Content-based citation analysis is applied to trace the dynamics of the article's impact changes from 2006 to 2014.

Beyond the rhetorical analysis of various parts of RAs, Swales' analytical framework and other researchers' using Swales' approach have popularized the importance of understanding how RAs are constructed (Kanoksilapatham, 2005) and applied it to various fields. Ding (2007) conducted a multi-level discourse analysis on a corpus of 30 medical/dental school application letters, five recurrent moves were identified. The studies that analyzed rhetorical structure found the limitation of Swales' framework and proposed more flexible, micro-level, or domain-specific frameworks. The genre analysis by Nwogu (1997) identified 11 moves in medical RAs. Nine were found to be normally required and two optional. Each move contains several constituent elements or sub-moves. Posteguillo's (Posteguillo, 1999) study on computer science indicated that IMRD pattern cannot be applied to the computer science RAs. Yang and Allison (2003) proposed a macro-structure framework for secondary studies in applied linguistics apart from the conventional IMRD framework. Kanoksilapatham (2005) provided a two-level rhetorical structure (moves and steps) based on the findings of analyzing 60 biochemistry articles. Their framework includes 15 distinct moves: three moves for the Introduction section, four for the Methods section, four for the Results section and four for the Discussion section. Lin and Evans (2012) analyzed a larger set of RAs that included 433 empirical articles from 39 disciplines, in which they found the most frequently used structural pattern is Introduction-Literature Review-Method-Results and Discussion-Conclusion (ILM[RD]C) and identified other prominent patterns. In a study on Chemistry RAs by Stoller and Robinson (2013), the pattern of Abstract, Introduction, Methods, Results, Discussion and Conclusion (A-IMRDC) sections was found to be the most frequently used one. More recently, Ye (2019) studied energy engineering RAs written by Chinese expert writers and found 86% of the articles used the IM[RD]C macro-structure.

Variations across disciplines and languages have been addressed by many studies. Samraj's study (2002) revealed disciplinary variation in the structure of introductions between Wildlife Behavior and Conservation Biology. Loi (2010) found that the rhetorical moves and steps were employed in fewer Chinese introductions compared to the English. The variability of rhetorical structure has also been investigated where a general trend of rhetorical structure evolution could be periodical. Xie (2020), by adopting multidimensional analysis, investigated the diachronic variation of styles in the Master thesis by Chinese students across 10 years. The results exhibited that a general trend is fluctuated in a 3-year-cycle pattern. Ozturk (2007) investigated the differences between two subdisciplines of applied linguistics and found the two subdisciplines seemed to employ different and almost unrelated move structures. Some studies have shown that rhetorical structures change with the development of disciplines. Nathan *et al.* (2002) examined whether textbooks exhibited a symbol precedence view of mathematical development and found textbooks published after 1900 contained far fewer symbol-only sections. As a highly interdisciplinary area, LIS's variability of rhetorical structure may be expected but have not been fully investigated. Very few studies analyzed the rhetorical structure of RAs in LIS (Rashidi and Meihami, 2018), and to our best knowledge, no study addressed the relationship between rhetorical structure and research content.

Research design

Data collection and sample selection

We selected the *Journal of the Association for Information Science and Technology* (*JASIST*), a well-known peer-reviewed journal in the LIS field, as the data source. Full text and metadata of 2,916 *JASIST* articles from 2001 to 2018 were downloaded in June 2019. We excluded several types of articles including brief communication, opinion paper, call for papers and biography, leaving 2,610 articles in the preliminary set. However, not all articles were indexed with Author Keywords or Keywords Plus, a complementary keyword index generated by Web of Science. We merged Keywords Plus to Author Keywords, hereafter simply "keywords," when the latter was missing. 2,216 articles in total, each with at least one keyword, were finally considered in our study.

Rhetorical structure identification

To consistently understand the meanings of components in terms of rhetorical structure among collaborators, we randomly chose 81 articles (4% of articles each year by a random number generator) as coding samples. Two collaborators manually read through all the articles and summarized the similarities in reference distribution throughout an article based on the literal names of section headings. The resemblances eventually formed into six components of rhetorical structure, i.e. Introduction, Literature Review, Methodology, Results, Discussion and Conclusion. The sequence of components was recorded for each article (e.g. "IMRD" or "ILMRD") to compare the coders' results (see Table 1). The resulting inter-coder reliability was 0.67, a score deemed "substantial" by the standard of Landis and Koch (1977). To train the single coder who would eventually code the rest of the articles, another 50 random articles were assigned to the same two coders as before. The inter-rater reliability (0.877) was "almost perfect" this time. The rest of the *JASIST* RAs were then coded. All the observed combinations of rhetorical structures (37 in total) are shown in Table 3.

Keyword analysis

Keywords were first standardized via Apache NLP 1.5.3, a vocabulary tool that uses natural language processing (Ding *et al.*, 2001; Waltman *et al.*, 2010) to facilitate such tasks as tokenization, sentence segmentation, chunking, parsing and co-reference resolution. We proceeded to use the toolkit to singularize plural forms of nouns and to identify prefix and suffix forms of nouns. Eventually, 2,371 distinct keywords were obtained. We applied VOSViewer (Van Eck and Waltman, 2010) to group into 57 clusters before two LIS domain experts select representative keywords in each cluster, which returned a total of 1,294 keywords with 3,998 occurrence frequency (41.31% of total keywords frequency) and sharing 1,743 (73.51%) of total papers. Among 1,294 keywords, we selected top five keywords in each cluster if the frequency is higher than five times (54 keywords in total) to conduct co-occurrence analysis in the analysis section.

Table 1.
The six components of rhetorical structure

Component	Abbr	Examples of other expressions in headings
Introduction	I	Background
Literature review	L	Related work, Related research, Previous research, Review of related literature, Related studies and Previous studies
Methodology	M	Approach and Research method
Results	R	Findings
Discussion	D	Discussion, Discussion and conclusion
Conclusion	C	Conclusion, Conclusion and discussion, Summary, Concluding remarks and Future work

Findings

Overview

Figure 1 displays the distribution of sampled articles from 2001 to 2018. The line above indicates the filtered proportion of original *JASIST* papers, presenting a direct impression. As for the numerical distribution, it is obvious that there is a stable increase from 2004 to 2016, despite some fluctuations; the annual number of publications grew by around 230% during this period. The trend has reversed with a sharp drop after 2016, shifting down to only 92 papers in 2018, indicating a stricter selection process. Meanwhile, the ratio of total papers retained in our sample ranges from 65% to 99%. The ups and downs of the proportion are in line with the main tendency of the numerical distribution. By reviewing the composition of *JASIST* volumes before and after 2016, we found that the decrease in selected samples was caused by a cut-off number of total publications in *JASIST*, whose annual article count halved in the years after 2016. The proportion of sampled articles remains relatively high (around 80%) in general and our sample can be reasonably taken to represent rhetorical structure trends in the journal.

Rhetorical structure evolution

Overview

Table 2 summarizes the statistics of major rhetorical structures in our sample. The most commonly used rhetorical structure is Introduction-Literature Review-Methodology-Result-Discussion-Conclusion (ILMRDC) (42.82%), with an average of 52.72 papers per year, indicating that scholars in the field of information science use this pattern most frequently. In fact, ILMRDC is more likely to be employed by other academics, such as those in applied linguistics, theoretical linguistics and management and marketing (Lin and Evans, 2012). Moreover, the top five rhetorical structures – ILMRDC, ILMRC, IMRDC, ILMRD and IMRC – happen to be used every year. Together they account for 87.99% of total usage, indicating that they are the mainstream rhetorical structures in information science. What interests us is that the conventional combination of IMRD turns out to account for only 3.43% of total usage. This finding partly accord with those of Posteguillo (1999) who noted that IMRD cannot be applied to RAs in computer science systematically. All major rhetorical structures also include an introduction, indicating that this section plays an essential role in papers and except for ILC (1.26%), major rhetorical structures also contain methodology and results

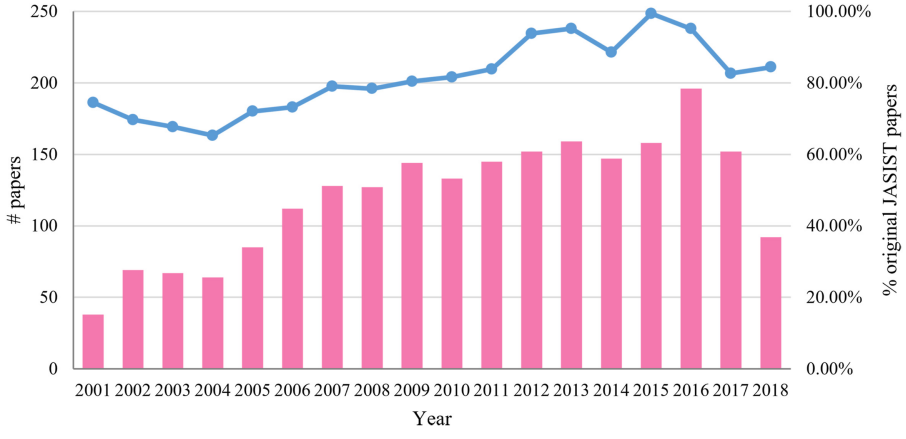

Figure 1.
Sampled articles and proportion of total *JASIST* papers included in sample

sections. Similarly, most of the major rhetorical structures contain conclusion and/or discussion sections. Other parts (3.61%) of the sample include a multiplicity of less-conventional patterns such as IL and IM.

Table 3 briefly describes the statistics of rhetorical structures as grouped by number of components. Analyzing usage patterns based on the components will help us understand the importance and feasibility of each component in the articles over all rhetorical structures. The largest variety of combinations are those consisting of four components, but these 13 combinations are not as widely used overall as those combinations with five and six components. The six types of rhetorical structures in the six-component category accounted for the largest proportion (44.90%), among which ILMRDC accounted for 95.38% of occurrences. This means ILMRDC is the dominant pattern and the other five rhetorical structures play a supporting role for it.

Considering the components individually, nearly all articles (99.82%) contain an introduction section, with an average of 122.89 articles per year (compare to the overall yearly average of 123.11) (Shown in Table 4). Only four articles lack a clearly indicated "Introduction" or similarly worded section heading; their rhetorical structures all begin with an L to compensate the lack of introduction, indicating that the introduction section plays an important role in establishing the background and research significance of articles. Other sections, including methodology, results and conclusion/discussion, appear in more than 97% of the articles, which means they are indispensable for RAs. Comparatively speaking, literature review (81.99%) is less used, perhaps because its content may instead form a part of introduction or discussion.

Table 2.
Statistics of major rhetorical structures by overall frequency

Rhetorical structure	Total #	% Total	Years attested	# per year
ILMRDC	949	42.82%	18	52.72
ILMRC	571	25.77%	18	31.72
IMRDC	178	8.03%	18	9.89
ILMRD	143	6.45%	18	7.94
IMRC	109	4.92%	18	6.06
IMRD	76	3.43%	17	4.47
ILMRCD	37	1.67%	13	2.85
ILMR	31	1.40%	14	2.21
ILC	28	1.26%	12	2.33
IMRCD	14	0.63%	9	1.56
Others	80	3.61%	17	4.71
Total	2,216	100.00%	18	123.11

Table 3.
Statistics of rhetorical structures by component count

# Components	# RS	Rhetorical structure	Total #	% Total	Years attested	# per year
2	2	IL, IM	4	0.18	4	1.00
3	7	IDC, ILC, ILD, IMD, IMR, IRC and IRD	46	2.08	17	2.71
4	13	ILDC, ILMC, ILMD, ILMR, ILRC, ILRD, IMDC, IMRC, IMRD, IMRL, IRDC, LMRC and LMRD	248	11.19	18	13.78
5	9	ILMDC, ILMRC, ILMRD, ILRDC, IMLRC, IMRCD, IMRDC, IMRLC and LMRDC	923	41.65	18	51.28
6	6	ILDMRC, ILMRCD, ILMRDC, IMLRDC, IMRDLC and IMRLDC	995	44.90	18	55.28

Temporal analysis of rhetorical structure usage

Figures 2–5 provide evolutionary views of rhetorical structure usage over the past 2 decades. The views of individual rhetorical structures are first presented in Figures 2 and 3; Figures 4 and 5 then give the results for integrated types of rhetorical structures.

Figure 2 is in line with Table 2 in illustrating annual changes of the major rhetorical structures. We make three main observations. First, the position of each rhetorical structure shows the highest proportion it attained within the total. It is obvious the trend changed in 2008: prior to this year, an average of 13 rhetorical structures were used, compared with 11 rhetorical structures after 2008. The minority structures ("Others" in the graph) deceased intensely after 2008. Similarly, ILMRC saw the highest use in 2003 and 2004, afterward facing a steady decrease. This finding leads to the second one, i.e. after 2008 the most prominent rhetorical structures account for an ever-greater proportion while the minority structures decrease over the years. A third finding is that rhetorical structures without D tend to peak before 2008; those with D, however, all peak after 2008, with the exception of ILMRD in 2007. This indicates that adding a discussion section has become a trend in *JASIST* RAs since 2008.

Rhetorical structure		Total #	% Total	# per year
I	With	2,212	99.82%	122.89
	Without	4	0.18%	1.33
L	With	1817	81.99%	100.94
	Without	399	18.01%	22.17
M	With	2,156	97.29%	119.78
	Without	60	2.71%	3.33
R	With	2,159	97.43%	119.94
	Without	57	2.57%	3.17
DC	With	2,171	97.97%	120.61
	Without	45	2.03%	2.50

Table 4.
Inclusion and omission of specific rhetorical structures

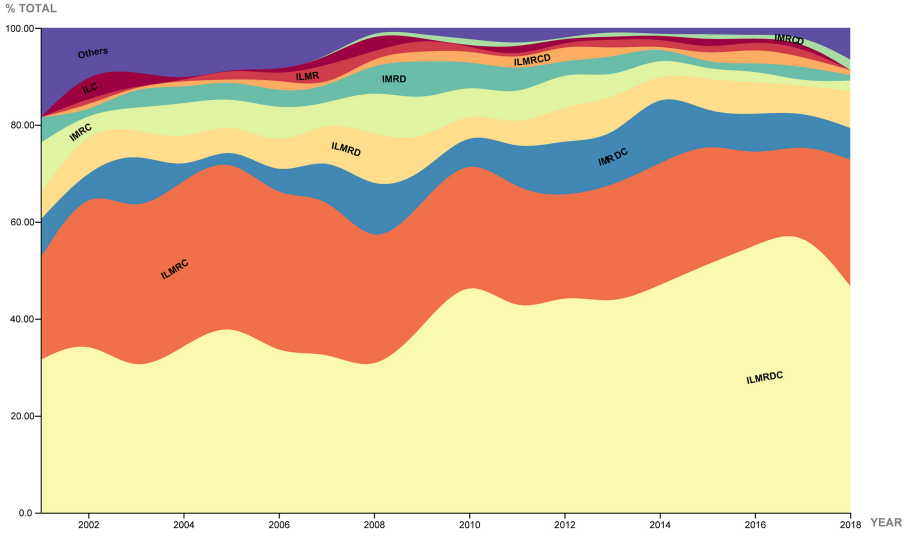

Figure 2.
Temporal timeline for major rhetorical structures

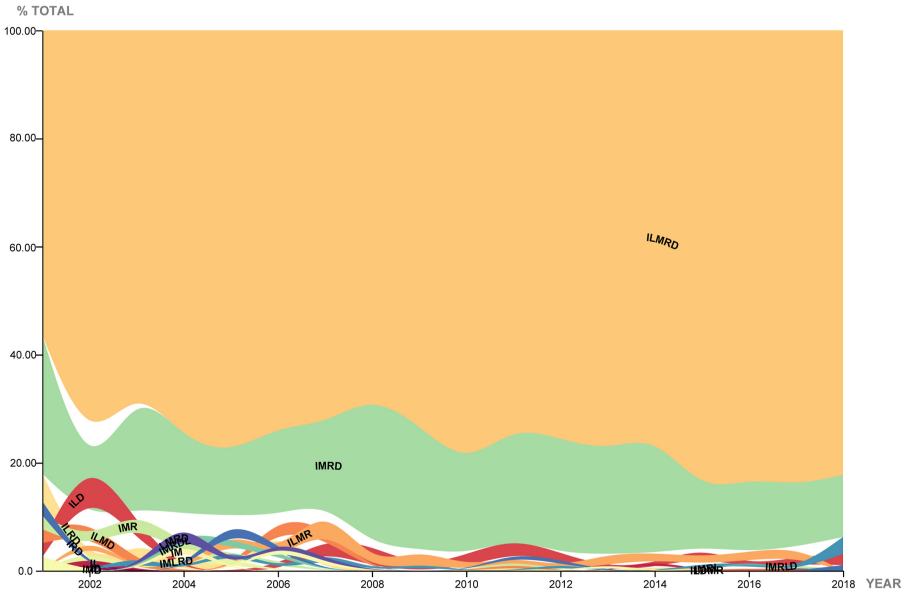

Figure 3.
Percent timeline for rhetorical structures with integrated DC

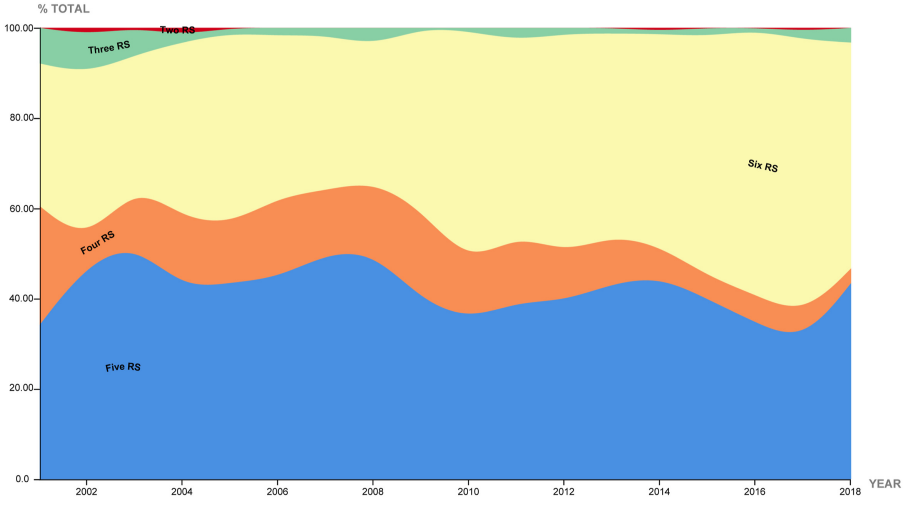

Figure 4.
Percent timeline for rhetorical structures by component count

Following the analytic logic of the discussion section, we put forward a hypothesis that discussion has played an important role in the temporal change. We then integrate conclusion with discussion as D, reducing the 37 observed types of rhetorical structures into 18 types as shown in Figure 3. This gives a markedly different distribution from Figure 2—one highly skewed toward ILMRD and IMRD, which represent eight of the original rhetorical structures. This supports our hypothesis regarding the importance of the discussion section. Viewed together, Figures 2 and 3 show that the majority structures have been used more and more over the period of interest.

Figure 5.
Percent timeline for rhetorical structures with and without specific sections

We next examine trends in the component count of the structures. In line with Table 3, Figure 4 shows that the most commonly used rhetorical structures fall into the groups with five components (Five RS) and six components (Six RS). The usage of Six RS increased over the years while other structures with fewer components saw less and less use. Five RS shows stable usage with a slight tapering-off after 2008. Moreover, apart from Six RS, peaks for each group appeared before 2004. The rhetorical structures of RAs have thus tended to become more complicated over the past 2 decades. A Pearson test shows a slight correlation ($r = 0.1813$, $p < 0.01$) between time and component count. This also gives evidence that rhetorical structures are more complex now than decades ago and indicates that the pattern of RA writing has become more and more detailed.

Figure 5 gives insight into the different roles played by each component in the rhetorical structures, which will help us to see the dynamic changes that the components have undergone every year. To demonstrate the differences, we introduced two indicators, "With" and "Without," for each component. "With" means that the combination includes a given component and "Without" means the combination omits that component. The percentage shown on the Y axis represents the proportion one group of rhetorical structures occupies in the total annual usage of rhetorical structures for each of the 18 years. Thus, the annual distribution shows the proportional changes across years. Since only four papers in our sample omitted an introduction, we exclude "With I" and "Without I" groups from the analysis in this figure.

The proportions for groups in Figure 5 range from 0% (1.43% if excluding the two 0s, Without DC in 2018 and Without R in 2005) to 15.79%. In most years, the proportions concentrated between 2 and 10%. The median ratios for all groups fall into a range from 4.44% to 6.19% and the average ratio of each year is 5.56%, showing a moderately concentrated distribution. The groups Without M and Without R were found to be normal

distributions ($p < 0.1$). All With groups are highly correlated to each other (average $r = 0.9897, p < 0.01$). Together, these phenomena show that all components have been playing similarly critical roles in the composition of RAs.

Comparing With and Without groups, we noticed that With groups have been more stably increasing than Without groups from the beginning of the study period. The groups Without M, R and DC have encountered ups and downs. Coincidentally, the three of them each decreased in prominence to some degree. It is worth testing whether the increase in the With groups is related to the decrease in the Without groups, but the results do not support this hypothesis. There are no significant overall relationships between With and Without groups. There is, however, a moderate positive relationship between With L and Without L groups ($r = 0.5382, p < 0.05$).

Moreover, alongside the other shifts observed in 2008, a major turning point occurred among these With and Without groups. Prior to 2008, Without groups in M, R and DC were larger than their respective With groups, indicating that researchers have paid more attention to the inclusion of detailed section titles since 2008. Unlike the other three groups, however, the literature review section exhibits two distinct shifts in 2008 and 2014. Before 2008, articles were more or less evenly divided in terms of literature review inclusion. During the next six years, articles without literature review became the mainstream. The mainstream then reversed and literature review has become an indispensable part of the scholarly communication standard, especially since 2014. These phenomena imply that methodology, result and discussion and conclusion sections can be applied more flexibly, while trends in the literature review section more closely correspond with the passage of time.

The results here interest us. In our previous study (Lou *et al.*, 2021), we found that literature review has increasingly been given a dedicated section in papers. The fact that *JASIST* started to limit word count in the early 2010s and recently limited articles to 7,000 words led us to presume the new policy would have affected the inclusion of a separate literature review, which is normally considered to be a word-count-intensive section. The results falsified our assumption, though the reason can also be explained by the results of this study: word count limitations do not restrict the content of RAs as long as these are well organized via detailed rhetorical structures. The quality of articles in many other highly prestigious journals is not affected even though they too are requested to limit their word count (PNAS, 2021).

Relationship between rhetorical structure and research content

The following sections will explore rhetorical structure usage preferences in terms of research areas in *JASIST* articles. For this purpose, Table 5 shows the top 10 clusters from VOSViewer clustering results, accounting for 68.62% of all keywords and occurring in 90.57% of total sampled articles. Topics 1 to 10, sequentially coordinating with Table 5, address Scholarly Communication and Scientometrics, Information Retrieval and Behavior, Information System, Knowledge Organization and Management, Information Economics, Human Computer Interaction, Social Network Analysis, Digital Libraries, Biomedical Informatics and Information Visualization. Representative keywords were manually examined by two LIS domain experts with the help of Zhu *et al.* (2016).

The probabilities in Table 5 give the likelihood of using each combination of rhetorical structures in an article on a certain research topic. They are calculated as follows. Each article can be categorized into several topics based on which topics its keywords belong to. To this end, this article's rhetorical structure can be assigned across several topics. For instance, article A has keywords a and b, clustered into topics 1 and 2, respectively. It will thus be given a 50% probability in topics 1 and 2. If article A applied the IMRD structure, then IMRD is deemed to have a 50% chance to be used in topic 1 and 2. Therefore, the total probability of one rhetorical structure being used in one topic is:

Topics	# Key words	Representative keywords/Usage probability (%)									
		ILC	ILMR	ILMRC	ILMRCD	ILMRD	ILMRDC	IMRC	IMRCD	IMRD	IMRDC
1	463	citation analysis, impact factor, scholarly communication, scientometrics, scientific collaboration									
		53.61	53.05	59.11	64.00	40.64	51.71	70.08	75.30	66.74	69.49
2	418	retrieval, information seeking, relevance, search process, cognitive model									
		75.69	56.74	56.69	51.73	48.65	55.95	42.96	12.50	44.13	45.79
3	222	decision making, self-efficacy, technology acceptance model, information systems, open-source software									
		18.13	38.97	46.31	55.67	35.35	37.09	46.47	60.00	42.78	37.91
4	144	text mining, automatic classification, knowledge representation, knowledge modeling, semantic web									
		100.00	47.28	67.74	70.67	67.20	64.56	74.58	30.00	77.69	68.04
5	87	performance, job satisfaction, social influences, turnover, interorganizational collaboration									
		17.36	10.00	32.71	42.14	27.04	26.26	18.75	26.67	18.75	49.00
6	65	human computer interaction, virtual communities, computer simulation, informal communication, assisted searching									
		50.00	33.33	48.87	12.50	75.66	52.94	68.89	33.33	75.00	63.33
7	65	social network, disclosure, myspace, Facebook use, information sharing									
		0.00	37.14	46.41	0.00	37.70	36.30	22.92	0.00	29.52	46.67
8	58	scholarly publishing, open access publication, publication lags, commercial publishers, digital rights management									
		0.00	100.00	54.93	60.00	88.89	53.58	55.56	0.00	55.00	68.06
9	56	breast cancer, prostate cancer, Alzheimer's disease, medical literature, literature-based discovery									
		50.00	0.00	52.01	0.00	48.52	26.94	51.11	0.00	0.00	45.67
10	49	color, display, motif, graphical information presentation, visual search									
		0.00	0.00	47.88	0.00	33.33	37.68	46.67	0.00	61.11	0.00

Table 5.
Keyword clusters and usage probability of top 10 rhetorical structures

$$P\left(rs_{a_i} \in cl_j \middle| a_i\right) = \frac{C\left(kw_{a_i} \in cl_j\right)}{C\left(kw_{a_i}\right)}$$

Where a_i is *ith* RA in the collection, rs_{a_i} is the rhetorical structure used by a_i, cl_j is the *jth* cluster, kw_{a_i} is the keyword of a_i, $C(kw_{a_i})$ is the keyword count of a_i and $C(kw_{a_i} \in cl_j)$ is the count of keywords of a_i that belongs to cl_j.

$$P(rs_k \in cl_j) = \frac{\sum_{i=1}^{n} P\left(rs_{a_i} \in cl_j, rs_{a_i} = rs_k \middle| a_i\right)}{\sum_{i=1}^{n} C\left(rs_{a_i} = rs_k, cl_j \in cl_{a_i} \middle| a_i\right)}$$

Where rs_k is the *kth* rhetorical structure, cl_{a_i} is the clusters of a_i, n is the number of RAs in the collection, $\sum_{i=1}^{n} P(rs_{a_i} \in cl_j, rs_{a_i} = rs_k | a_i)$ is the probability that the rhetorical structure used by a_i is rs_k and belongs to cl_j and $\sum_{i=1}^{n} C(rs_{a_i} = rs_k, cl_j \in cl_{a_i} | a_i)$ is the count of a_i that uses rs_k and whose clusters includes cl_j.

Inspired by the gender detection method based on full names (Santamaría and Mihaljević, 2018), which does provide the probability of gender instead of affirmation of gender, we intend to demonstrate that the usage patterns of rhetorical structure in different research topics are not in a fixed relationship but a selective one. In this way, we can analyze the correlation relationship between research topics and rhetorical structures. As they are both categorical data, we convert them into a contingency table. Pearson's chi-squared test results show that they are highly related ($\chi^2 = 2{,}432.2$, $p = 5.452\text{e-}08$) with a Cramér's V of 0.602 (Ferguson, 1981; Smith and Albaum, 2005). This indicates that researchers in different research areas prefer to apply various rhetorical structures. As shown in Table 5, the two most commonly used structures, ILMRDC and ILMRC, are comparably equally distributed in every topic. They are most likely to be used by researchers in text analysis areas. Articles about Knowledge Organization and Management would also have a high likelihood of applying ILC and those in Digital Libraries of applying ILMR, but not vice versa. Articles in Scholarly Communication and Scientometrics prefer IMRC, IMRCD, IMRD and IMRDC as rhetorical structures, all of which are missing a distinct literature review section. Conversely, articles in Information Retrieval and Behavior prefer ILC, ILMR, ILMRC and ILMRDC, all of which include a literature review section.

The coupling cluster view in Figure 6 clearly shows how rhetorical structures are positioned with respect to *JASIST* research areas. 10 major rhetorical structures and 54 representative keywords were gathered into nine clusters based on the co-occurrence between structures and keywords. In the figure, a keyword can co-occur with multiple rhetorical structures. The three biggest rhetorical structures are ILMRDC, ILMRC and ILMRD showing their major statuses in the LIS domain. They were commonly used in research areas such as digital libraries (keywords: digital libraries, open access publication, scholarly publishing, etc.), information retrieval and behavior (keywords: information seeking, information access, user behavior, etc.) and human computer interaction (keywords: human computer interaction, computer-mediated communication, virtual communities, etc.) IMRDC, IMRCD, IMRD and IMRC, which has the common characteristic of not having the literature review section, were mainly used in scholarly communication and scientometrics (keywords: citation, h-index, bibliometrics, etc.). Compared to others, the area does not show much interest in including literature review in RAs; this reflects bibliometrics' standing as a tool to review literature itself (Rousseau, 2021).

Figure 7 shows the cluster map of relationships between 18 types of rhetorical structures and high-frequency keywords. After integrating conclusion with discussion and compared to Figure 6, the picture has transformed from a tricentric to a unicentric map. ILMRD is the most used rhetorical structure here, interacting with all high-frequency keywords at an average of 54 co-occurrences. Basically, ILMRD becomes the only dominant rhetorical structure, eclipsing even the classic combination, IMRD. However, IMRD was also applied in research areas focusing on all high-frequency keywords, especially in scientometrics research. Between them, these two rhetorical structures occupy 94.29% of co-occurrence with high-frequency keywords, making them the mainstream rhetorical structures in *JASIST* articles of the past 2 decades.

Clustering by With and Without groups allows us to further depict the importance of each component of rhetorical structure in different research areas (See Figure 8). Eight such groups are combined into six clusters; Without M and Without R were clustered into the same group. This result is predictable since the results section of a RA often reports the results of applying the proposed methods. They share keywords such as *library*, *performance* and *information* need. The keywords are usually used in non-technical research areas where the proposal of new research methods is usually not necessary. Without L and Without DC were clustered into the same group with common keywords such as *bibliometrics*, *citation*, *h-index*, etc. However, these keywords were also connected with the group of With L and With DC, showing that the wring styles of researchers in the area of scholarly communication and scientometrics are selective.

Pearson correlation coefficients show that most of the negative correlations are not significant, except for the battle between methodology and results. The relationship between

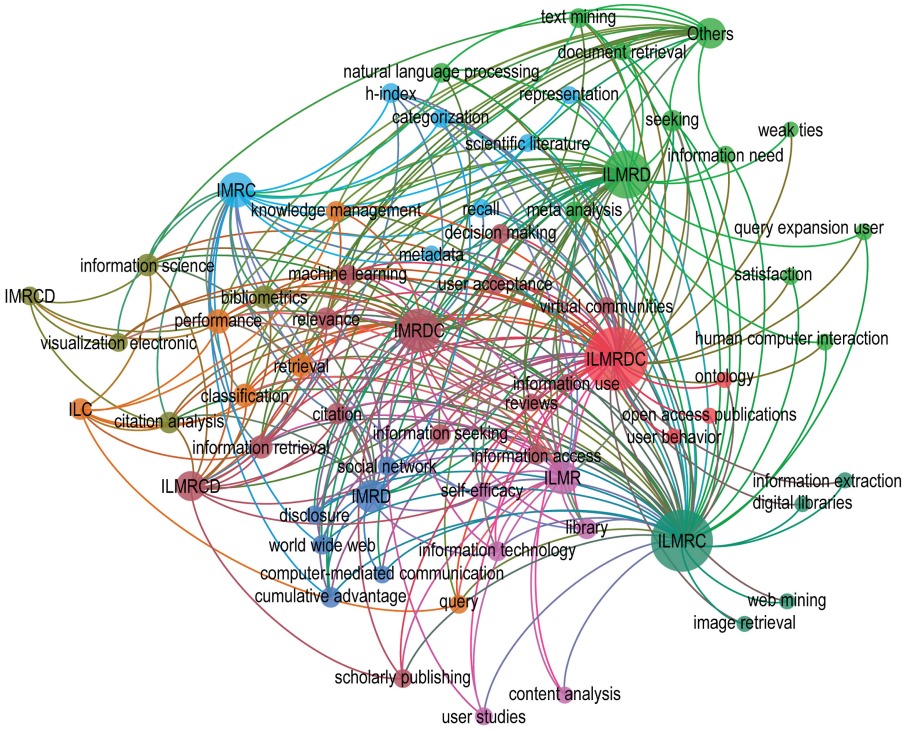

Figure 6.
Clusters of major rhetorical structures and high-frequency keywords

With M and Without R and the relationship between Without M and With R are both strong negatively related ($r = -0.6876, p < 0.001$). This leads to the great distance between these two groups on the map. The fact that With R is posited in the center reflects the importance of results presentation in RAs. Moreover, articles with methodology and discussion and conclusion sections are situated together, while With L and Without L are clustered into different categories from the others, underscoring the special function of literature review.

The same map shows that researchers select the components of rhetorical structure differently in different research areas. The lines between nodes in Figure 8 represent the ratio of usage of one rhetorical structure type by a certain keyword to the total occurrence of the keyword. All keywords applied methodology, research, discussion and conclusion in at least 86% of their articles. Moreover, these three groups are highly related, therefore, they are not indicative of a difference among research areas. On the other hand, the With and Without L groups can tell more. Research areas such as knowledge organization, digital libraries and user study are particularly likely to include literature review in their articles; scientometrics studies focusing on bibliometrics and citation analysis use literature review sections comparatively little.

Discussion and conclusion

Rhetorical structures play key roles in every research area and every journal. Previous studies have explored the usage patterns of rhetorical structures in many venues, yet we have very limited knowledge of whether and how rhetorical structures are connected to different research areas. To address this gap, we manually coded the rhetorical structures used in more than 2,000 *JASIST* articles to understand how these structures are used in a prestigious LIS journal and to identify usage trends across time and LIS sub-topics.

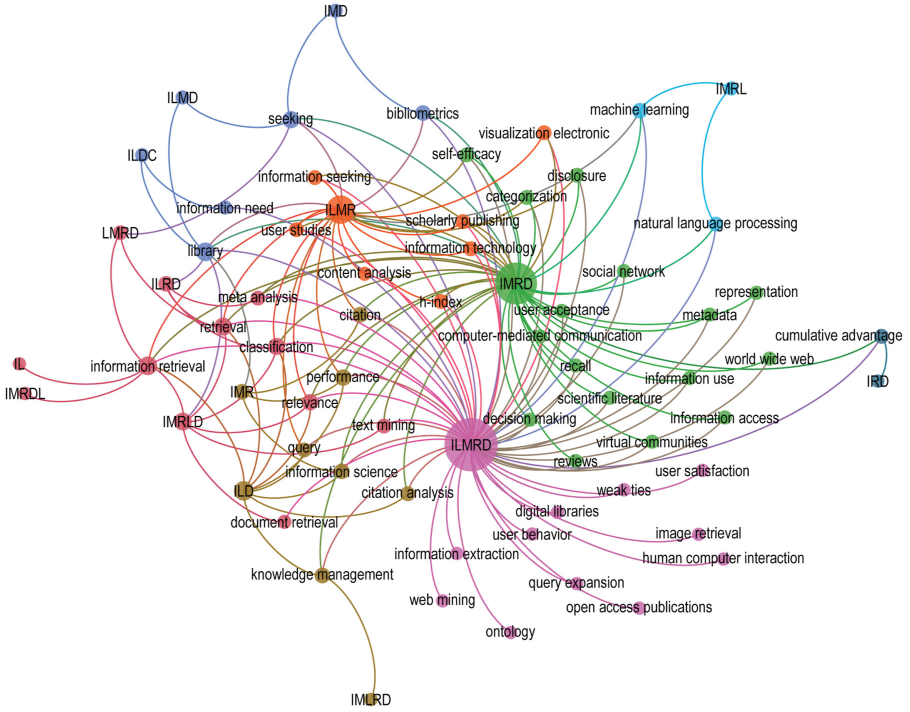

Figure 7.
Clusters of rhetorical structures (integrated DC) and representative keywords

We found that ILMRDC has contributed a large proportion of *JASIST* publications in this field, in contrast to other empirical results on rhetorical structures (Kovačević *et al.*, 2012). Moreover, the components in this rhetorical structure have been flexibly formed into five other types that have gained momentum as well.

Another important finding is that the use of all the rhetorical structures undergoes significant changes over time. First of all, rhetorical structures have been less and less diverse but more and more detailed, with evidence of fewer and fewer combinations, a predominance of five- and six-component structures and the emergence of ILMRDC groups. Secondly, the year 2008 is a milestone in rhetorical structure usage patterns: the prominent ILMRDC structure has definitively sprung up since 2008 along with its sibling, while ILMRC started to show decline. This evidence prompts a necessary investigation of the history of *JASIST* and even LIS community. For instance, *JASIST* changed its name in 2014, and from our evidence, there is a clear shift in literature review section usage since that time.

One possible explanation for the temporal changes in rhetorical structure use is that writing patterns within the scientific community have themselves changed. We concluded in our literature review section that even though IMRD is the conventional rhetorical structure, there are significant differences across research areas. These we attempt to explain from the perspective of changes in the editorial team. In the past 2 decades, *JASIST* has assigned different groups of editorial teams, and it is evident that journal editorial boards have to some extent set the pace of research topics (Kim, 2014; Walters, 2016), which have been proved in this study to be highly related to the choice of rhetorical structure. We do not speculate further on this matter here, but consider it as a limitation of the present paper and a target for further studies.

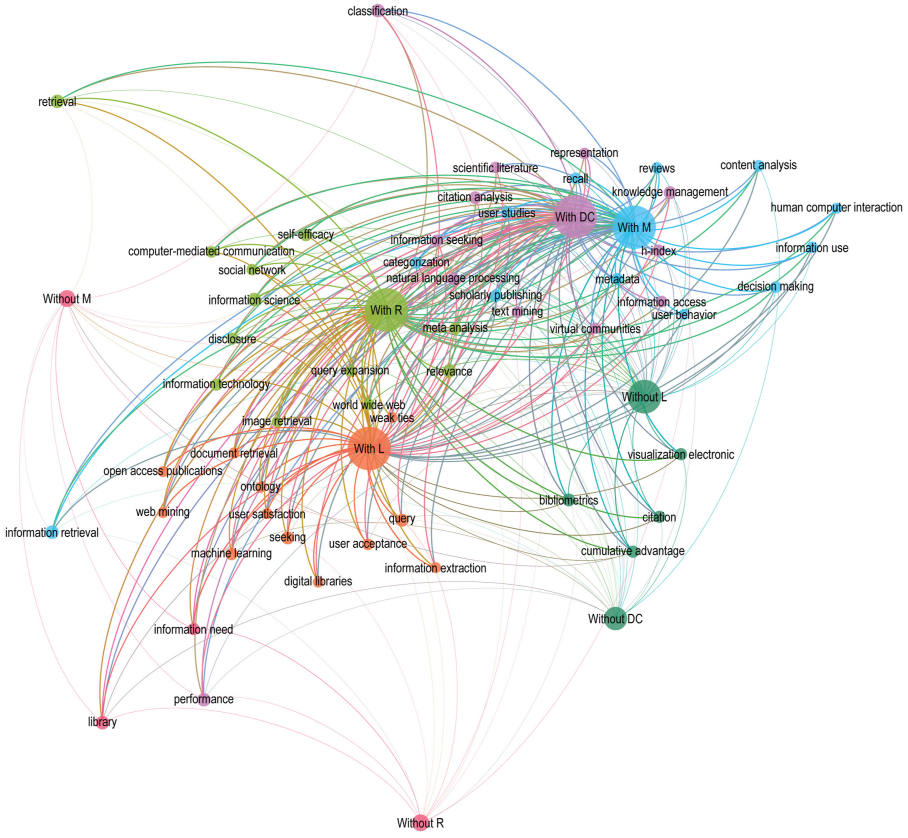

Figure 8.
Cluster of eight groups of rhetorical structures and representative keywords

The other possibility, which we try to emphasize in this article, is that the patterns reflect the differing focuses of research topics in *JASIST*. Based on our results, a reasonable conclusion is that researchers have selected rhetorical structures based on their research specialties. Information retrieval, knowledge organization and scientometrics exhibit significant differences in rhetorical structure usage patterns; respectively, they prefer to apply ILMRDC, ILMRC and IMRD to structure articles. Moreover, literature review and discussion sections tend to be on the controversial center: our results demonstrate a strong preference against including literature review in scientometrics research but a preference for it in the fields of information retrieval and user studies. Interestingly, researchers in the field of technological topics have avoided the discussion section. Together, these observations indicate that researchers have strived to adjust to certain norms in the sub-field of the LIS community by reflecting it in their choice of rhetorical structures. In our 18-year sample, *JASIST* has published research ranging widely from information retrieval, web (Internet) analysis and social (media/networks), to scholarly publishing and citation analysis (Agarwal and Islam, 2020). The temporal dissemination of topics coincides with our results on the usage differences in rhetorical structures over time.

The factors on choosing rhetorical structure can be complicated more than research topic, which we attempt to demonstrate in this article. Some investigations suggest a commonly occurrence during reviewing process is that reviewers tend to strongly suggest the authors

revise papers as their expectations of how the paper is structured (Kumar *et al.*, 2011). With the help of natural language process, this type of negotiations on articles' structure can be examined quantitatively in a larger sample. This further study is not in our aim but even though the final formats of publications have been reformed by reviewers out of authors' original choices, it reflects the pattern in a certain time.

To sum up, the present paper offers a first systematic examination of how rhetorical structures are used in a representative sample of a LIS journal, especially from a temporal perspective. We find evidence consistent with prior literature and our knowledge of the progress of the LIS research community. The contribution of this research is not only in expanding the sample size but also in bridging the gap between rhetorical structure and specific research areas. Our present approach is an intermediate step towards solving a significant follow-up question for further study, i.e. illuminating the relationship between writing patterns and individual knowledge structures and the internal usage pattern in individual rhetorical moves.

References

Agarwal, N.K. and Islam, M.A. (2020), "Journal of the association for information science and technology: analysis of two decades of published research", *Proceedings of the Association for Information Science and Technology*, Vol. 57 No. 1, pp. 1-18, doi: 10.1002/pra2.220.

Anthony, L. (1999), "Writing research article introductions in software engineering: how accurate is a standard model?", *IEEE Transactions on Professional Communication*, Vol. 42 No. 1, pp. 38-46, doi: 10.1109/47.749366.

Basturkmen, H. (2012), "A genre-based investigation of discussion sections of research articles in Dentistry and disciplinary variation", *Journal of English for Academic Purposes*, Vol. 11 No. 2, pp. 134-144, doi: 10.1016/j.jeap.2011.10.004.

Bertin, M., Atanassova, I., Sugimoto, C.R. and Lariviere, V. (2016), "The linguistic patterns and rhetorical structure of citation context: an approach using n-grams", *Scientometrics*, Vol. 109 No. 3, pp. 1417-1434, doi: 10.1007/s11192-016-2134-8.

Brett, P. (1994), "A genre analysis of the results section of sociology articles", *English for Specific Purposes*, Vol. 13 No. 1, pp. 47-59.

Ding, H. (2007), "Genre analysis of personal statements: analysis of moves in application essays to medical and dental schools", *English for Specific Purposes*, Vol. 26 No. 3, pp. 368-392, doi: 10.1016/j.esp.2006.09.004.

Ding, Y., Chowdhury, G.G. and Foo, S. (2001), "Bibliometric cartography of information retrieval research by using co-word analysis", *Information Processing and Management*, Vol. 37 No. 6, pp. 817-842.

Ferguson, G.A. (1981), *Statistical Analysis in Psychology and Education*, McGraw-Hill, New York, NY.

Gillaerts, P. and Van de Velde, F. (2010), "Interactional metadiscourse in research article abstracts", *Journal of English for Academic Purposes*, Vol. 9 No. 2, pp. 128-139, doi: 10.1016/j.jeap.2010.02.004.

Hyland, K.E.N. and Tse, P. (2004), "Metadiscourse in academic writing: a reappraisal", *Applied Linguistics*, Vol. 25 No. 2, pp. 156-177, doi: 10.1093/applin/25.2.156.

Jamali, H.R. and Nabavi, M. (2015), "Open access and sources of full-text articles in Google Scholar in different subject fields", *Scientometrics*, Vol. 105 No. 3, pp. 1635-1651, doi: 10.1007/s11192-015-1642-2.

Kanoksilapatham, B. (2005), "Rhetorical structure of biochemistry research articles", *English for Specific Purposes*, Vol. 24 No. 3, pp. 269-292, doi: 10.1016/j.esp.2004.08.003.

Kim, H. (2014), "The JASIST editorial board members' research areas and keywords of JASIST research articles", *Journal of the Korean Society for Information Management*, Vol. 31 No. 3, pp. 227-247.

Kovačević, A., Konjović, Z., Milosavljević, B. and Nenadic, G. (2012), "Mining methodologies from NLP publications: a case study in automatic terminology recognition", *Computer Speech and Language*, Vol. 26 No. 2, pp. 105-126.

Kumar, P., Rafiq, I. and Imam, B. (2011), "Negotiation on the assessment of research articles with academic reviewers: application of peer-review approach of teaching", *Higher Education*, Vol. 62 No. 3, pp. 315-332.

Landis, J.R. and Koch, G.G. (1977), "The measurement of observer agreement for categorical data", *Biometrics*, Vol. 33 No. 1, pp. 159-174.

Lin, L. and Evans, S. (2012), "Structural patterns in empirical research articles: a cross-disciplinary study", *English for Specific Purposes*, Vol. 31 No. 3, pp. 150-160, doi: 10.1016/j.esp.2011.10.002.

Loi, C.K. (2010), "Research article introductions in Chinese and English: a comparative genre-based study", *Journal of English for Academic Purposes*, Vol. 9 No. 4, pp. 267-279, doi: 10.1016/j.jeap.2010.09.004.

Loré, R. (2004), "On RA abstracts: from rhetorical structure to thematic organisation", *English for Specific Purposes*, Vol. 23 No. 3, pp. 280-302, doi: 10.1016/j.esp.2003.06.001.

Lou, W., Su, Z., Zheng, S. and He, J. (2021), "The Effects of rhetorical structure on citing behavior in research articles", *18th International Conference on Scientometrics and Informetrics*, pp. 705-710.

Milagros Del Saz Rubio, M. (2011), "A pragmatic approach to the macro-structure and metadiscoursal features of research article introductions in the field of Agricultural Sciences", *English for Specific Purposes*, Vol. 30 No. 4, pp. 258-271.

Milojević, S., Sugimoto, C.R., Yan, E. and Ding, Y. (2011), "The cognitive structure of library and information science: analysis of article title words", *Journal of the American Society for Information Science and Technology*, Vol. 62 No. 10, pp. 1933-1953.

Mingers, J. and Leydesdorff, L. (2015), "A review of theory and practice in scientometrics", *European Journal of Operational Research*, Vol. 246 No. 1, pp. 1-19, doi: 10.1016/j.ejor.2015.04.002.

Nathan, M.J., Long, S.D. and Alibali, M.W. (2002), "The symbol precedence view of mathematical development: a corpus analysis of the rhetorical structure of textbooks", *Discourse Processes*, Vol. 33 No. 1, pp. 1-21, doi: 10.1207/S15326950DP3301_01.

Nwogu, K.N. (1997), "The medical research paper: structure and functions", *English for Specific Purposes*, Vol. 16 No. 2 SPEC. ISS., pp. 119-138, doi: 10.1016/s0889-4906(97)85388-4.

Ozturk, I. (2007), "The textual organisation of research article introductions in applied linguistics: variability within a single discipline", *English for Specific Purposes*, Vol. 26 No. 1, pp. 25-38, doi: 10.1016/j.esp.2005.12.003.

Pho, P.D. (2008), "Research article abstracts in applied linguistics and educational technology: a study of linguistic realizations of rhetorical structure and authorial stance", *Discourse Studies*, Vol. 10 No. 2, pp. 231-250.

PNAS (2021), "Submitting your manuscript | PNAS", *Submitting Your Manuscript*, available at: https://www.pnas.org/authors/submitting-your-manuscript (accessed 29 October 2021).

Posteguillo, S. (1999), "The schematic structure of computer science research articles", *English for Specific Purposes*, Vol. 18 No. 2, pp. 139-160, doi: 10.1016/s0889-4906(98)00001-5.

Rashidi, N. and Meihami, H. (2018), "Informetrics of Scientometrics abstracts: a rhetorical move analysis of the research abstracts published in Scientometrics journal", *Scientometrics*, Vol. 116 No. 3, pp. 1975-1994.

Rousseau, R. (2021), "Why I do not like literature review sections", *ISSI Newsletter*, Vol. 17 No. 1, pp. 13-14.

Saier, T. and Färber, M. (2020), "unarXive: a large scholarly data set with publications' full-text, annotated in-text citations, and links to metadata", *Scientometrics*, Vol. 125 No. 3, pp. 3085-3108, doi: 10.1007/s11192-020-03382-z.

Samraj, B. (2002), "Introductions in research articles: variations across disciplines", *English for Specific Purposes*, Vol. 21 No. 1, pp. 1-17, doi: 10.1016/S0889-4906(00)00023-5.

Santamaría, L. and Mihaljević, H. (2018), "Comparison and benchmark of name-to- gender inference services", *PeerJ Computer Science*, Vol. 2018 No. 7, pp. 1-29, doi: 10.7717/peerj-cs.156.

Smith, S.M. and Albaum, G.S. (2005), *Fundamentals of Marketing Research*, Sage.

Soler-Monreal, C., Carbonell-Olivares, M. and Gil-Salom, L. (2011), "A contrastive study of the rhetorical organisation of English and Spanish PhD thesis introductions", *English for Specific Purposes*, Vol. 30 No. 1, pp. 4-17.

Stoller, F.L. and Robinson, M.S. (2013), "Chemistry journal articles: an interdisciplinary approach to move analysis with pedagogical aims", *English for Specific Purposes*, Vol. 32 No. 1, pp. 45-57, doi: 10.1016/j.esp.2012.09.001.

Swales, J. (1990), *Genre Analysis: English in Academic and Research Settings*, 1st ed., Cambridge University Press.

Tessuto, G. (2015), "Generic structure and rhetorical moves in English-language empirical law research articles: sites of interdisciplinary and interdiscursive cross-over", *English for Specific Purposes*, Vol. 37, pp. 13-26, doi: 10.1016/j.esp.2014.06.002.

Van Eck, N.J. and Waltman, L. (2010), "Software survey: VOSviewer, a computer program for bibliometric mapping", *Scientometrics*, Vol. 84 No. 2, pp. 523-538.

Walters, W.H. (2016), "The research contributions of editorial board members in library and information science", *Journal of Scholarly Publishing*, Vol. 47 No. 2, pp. 121-146, doi: 10.3138/jsp.47.2.121.

Waltman, L., Van Eck, N.J. and Noyons, E.C.M. (2010), "A unified approach to mapping and clustering of bibliometric networks", *Journal of Informetrics*, Vol. 4 No. 4, pp. 629-635, doi: 10.1016/j.joi.2010.07.002.

Xie, S. (2020), "Multidimensional analysis of Master thesis abstracts: a diachronic perspective", *Scientometrics*, Vol. 123 No. 2, pp. 861-881, doi: 10.1007/s11192-020-03408-6.

Yang, R. and Allison, D. (2003), "Research articles in applied linguistics: moving from results to conclusions", *English for Specific Purposes*, Vol. 22 No. 4, pp. 365-385, doi: 10.1016/S0889-4906(02)00026-1.

Yang, R. and Allison, D. (2004), "Research articles in applied linguistics: structures from a functional perspective", *English for Specific Purposes*, Vol. 23 No. 3, pp. 264-279, doi: 10.1016/S0889-4906(03)00005-X.

Ye, Y. (2019), "Macrostructures and rhetorical moves in energy engineering research articles written by Chinese expert writers", *Journal of English for Academic Purposes*, Vol. 38, pp. 48-61, doi: 10.1016/j.jeap.2019.01.007.

Zhang, G., Ding, Y. and Milojević, S. (2013), "Citation content analysis (CCA): a framework for syntactic and semantic analysis of citation content", *Journal of the American Society for Information Science and Technology*, Vol. 64 No. 7, pp. 1490-1503, doi: 10.1002/asi.22850.

Zhu, Y., Yan, E. and Song, M. (2016), "Understanding the evolving academic landscape of library and information science through faculty hiring data", *Scientometrics*, Vol. 108 No. 3, pp. 1461-1478.

Corresponding author
Yongjun Zhu can be contacted at: zhu@yonsei.ac.kr