

Analysis of the Macro-Level Discourse Structure of Literature Reviews

Christopher S.G. Khoo, Jin-Cheon Na and Kokil Jaidka

Wee Kim Wee School of Communication and Information, Nanyang
Technological University, Singapore

Abstract

Purpose – The purpose of this study is to analyze the macro-level discourse structure of literature reviews found in information science journal papers, and to identify different styles of literature review writing. Though there have been several studies of human abstracting, there are hardly any studies of how authors construct literature reviews. This study is carried out in the context of a project to develop a summarization system to generate literature reviews automatically.

Design/methodology/approach – A coding scheme was developed to annotate the high-level organization of literature reviews, focusing on the types of information. Two sets of annotations were used to check inter-coder reliability.

Findings – It was found that literature reviews are written in two distinctive styles, with different discourse structures. *Descriptive literature reviews* summarize individual papers/studies and provide more information on each study, such as research methods, results and interpretation. *Integrative literature reviews* provide fewer details of individual papers/studies, but focus on ideas and results extracted from these papers. They provide critical summaries of topics, and have a more complex structure of topics and sub-topics. The reviewer's voice is also more dominant.

Originality/value – The coding scheme is useful for annotating the macro-level discourse structure of literature reviews, and can be used for studying literature reviews in other fields. The basic characteristics of two styles of literature review writing are identified. The results have provided a foundation for further studies of literature reviews—to identify discourse relations and rhetorical functions employed in literature reviews, and their linguistic expressions.

Keywords – Information science, literature reviews, discourse structure

Paper type – Research paper

Introduction

Every research paper is expected to have some kind of literature review. Papers in good research journals are expected to have a fairly substantive literature review section. A literature review is not just a summary or overview of previous literature on the subject but serves particular objectives including situating the reported study in a specific research tradition or community, providing a rationale for carrying out the study, delineating gaps in the literature the study is expected to fill, and providing support for various aspects of the study such as methodology and factors investigated.

A literature review therefore serves important functions in a research report. Though many books on research methods and writing give prescriptive advice on how to write good literature reviews, there are few empirical studies of how authors actually construct a literature review—select information to include in the literature review, organize it, shape the arguments to support the research study, and select the linguistic expressions and rhetorical devices to present it. There are few formal analyses of published literature reviews.

This study is carried out in the context of a project to develop a summarization system to generate literature reviews automatically. A literature review can be considered a multi-document summary of previous research. Previous multi-document summarization research has focused on summarizing news articles (McKeown et al., 2002; Barzilay and McKeown, 2005). Literature review generation is radically different because it identifies relevant information in different source documents that are relevant to the current study and use them to justify various aspects of the study.

However, to mimic human literature review generation, it is necessary to understand how humans construct literature reviews. The literature review writing task is typically performed by researchers who are trained to construct literature reviews in their field. They employ certain strategies and techniques. A study of the structure and characteristics of real literature reviews would be of potential use in creating a literature review rubric.

Analyzing human-generated literature reviews is an obvious place to start as there is an abundance of examples in research journals. A study of literature review discourse would involve understanding the components that make up the text, such as textual units and discourse relations. Textual units make up the content of the literature review. Discourse relations hold between textual units and are used to shape the author's logical and rhetorical argument and define the text structure. Discourse analysis has been defined as "a cross-disciplinary method of inquiry which studies the structure of texts and considers both their linguistic and socio-cultural dimensions in order to determine how meaning is constructed" (Barsky, 2001 pp. 1). Our focus is on the types of information found in literature reviews, and how they are used in the author's logical and rhetorical arguments. These logical and rhetorical arguments are implemented in the discourse relations that may involve cohesive and rhetorical relations (Marcu, 2000), cue phrases (McKeown et al., 2002), cognitive resources drawn upon by writers (Endres-Niggemeyer et al., 1995), and so on.

We are carrying out in-depth content analyses and linguistic analyses of literature review sections in information science journals. The project is divided into three parts:

1. Analysis of how information is selected from the source papers and organized in the literature review
2. Analysis of the macro-level discourse structure of literature reviews at the sentence level
3. Analysis of the micro-level structure within sentences and clauses.

This paper reports the results of an initial study of the macro-level discourse structure of sample literature reviews found in the *Journal of the American Society for Information Science*

& *Technology* (JASIST). JASIST is one of the leading journals in the field and carries high-quality research articles with substantive literature review sections. The study will later be extended to a few other information science journals. A coding scheme was developed to annotate the high-level organization of the literature review content focusing on the types of information. Two sets of annotations were used to check inter-coder reliability.

In analyzing the sample literature reviews, two distinct styles of literature review was found:

1. *descriptive literature reviews* that focus on describing individual studies/papers
2. *integrative literature reviews* that focus on the main ideas and results from the cited papers—extracting results and ideas from the papers and synthesizing a summary of the research trends and milestones.

The macro-level discourse structure of descriptive and integrative literature reviews were compared to find out differences in types of information found and how the literature reviews were structured.

Related Work

In the words of (Cooper, 1988, p. 107), “A literature review seeks to describe, summarise, evaluate, clarify and/or integrate the content of primary reports”. The purpose it serves is to allow a comparison of the opinions of other researchers, compare and contrast their contributions in a balanced, structured way (Bourner, 1996). In the context of thesis writing, literature reviews have been portrayed as serving similar or overlapping functional purposes as introduction sections (Kwan, 2006); thesis writing manuals often discuss the literature review and introduction section together (e.g. Rudestam and Newton, 2001). Indeed introduction sections sometimes contain mini-literature reviews.

However, a literature review can serve many purposes. The following is a list culled from the literature:

- *Source identification*. Identifying and pinpointing documents of relevance (Rowley and Slack, 2004).
- *Comparing and contrasting previous research*. Identifying other people working in the same field (Bourner, 1996), comparing their contributions, identifying the relationships amongst studies and contrasting opposing views (Hart, 1998).
- *Identifying gaps in the literature*. Distinguishing what has already been achieved from what still needs to be understood and accomplished, thus identifying existing gaps in literature and raising questions that hold potential for exploration (Boote and Beile, 2005; Bourner, 1996).
- *Identifying issues*. Providing “a clearly organized, well argued statement of what the literature does and does not say about some key issue or question” (Elmore, 2006 pp. 3).
- *Defining the proposed research contributions*. Highlighting the intended contributions and justifying the research (Jonsson, 2006; Massey, 1996; Bruce, 1994)

- *Building the foundation*. Providing “the foundation stone on which one’s own work is built” (Massey, 1996), as well as the background to the research (Bruce, 1994).
- *Situating the work in the research literature*. Helping “the reader understand your thinking and what has impacted it, [adding] credibility to your expertise as an author, and [situating] your writing in the larger scholarly dialogue” (Hinchliffe, 2003 pp. 163).
- *Reinterpretation of results*. Creating “new understandings of the topic” and building a premise about what the existing evidence suggests for future inquiry (Boote and Beile, 2005; Torraco, 2005; Hart, 1998).

Thus it is clear that, besides providing an overview of a topic, a literature review serves many purposes and can support a research study in many ways. It is anticipated that the functional purposes of a literature review will be manifested in its rhetorical organization and linguistic expression.

In broad terms, a literature review can be considered as having the following constituent sections (Cooper, 1988):

- *Introduction*, including the research questions that the writer wishes to answer
- *Body*, which can be organized in the following ways:
 - Chronologically—to highlight trends
 - Thematically—to identify the scope and themes of groups of studies
 - Methodologically—to identify the types of research methods used
- *Conclusion*, including a critique stating flaws and gaps in the literature to justify the author’s own work (Molina, 1995).

Swales (1981; 1990) offered one approach to studying the discourse structure of the introduction section of research articles—the Create a Research Space (CARS) model describes the schematic structure of introduction sections as the application of three “moves”. Moves are rhetorical movements in text that define its schematic structure and create a logical progression. Each move may employ several “steps” to achieve its purpose.

Swales’ model has been used by other researchers for discourse analysis of academic text. Most recently, Kwan (2006) applied an extended CARS model developed by Bunton (2002) to propose a move structure for the “thematic units” or topics in the literature review chapters of theses (see Table 1). She identified the “strategies” in text which were the ways employed to fulfill each rhetorical move. Kwan reported that the schema of a literature review chapter usually follows an Introduction-Body-Conclusion structure. Move 1 (*Establishing one part of the territory of one’s own research*) and Move 2 (*Creating a research niche*) are usually recursively applied. Move 3 (*Occupying the research niche*) was rarely found to occur in the set of literature reviews considered by Kwan. A new type of move encountered by Kwan was the writer’s affirmative stance towards the surveyed research. Also, none of the move elements followed a specific progression. Through her work, Kwan demonstrated that literature review chapters are distinct from introduction chapters in structure, purpose and rhetorical organization.

Table 1. The modified CARS model for literature review chapters (Kwan, 2006)

Move 1	Establishing one part of the territory of one's own research by
Strategy A	surveying the non-research-related phenomena or knowledge claims
Strategy B	claiming centrality
Strategy C	surveying the research-related phenomena
Move 2	Creating a research niche (in response to Move 1) by
Strategy A	counter-claiming
Strategy B	gap-indicating
Strategy C	asserting confirmative claims about knowledge or research practices surveyed
Strategy D	asserting the relevancy of the surveyed claims to one's own research
Strategy E	abstracting or synthesizing knowledge claims to establish a theoretical position or a theoretical framework
Move3 (optional)	Occupying the research niche by announcing:
Strategy A	research aims, focus, research questions or hypotheses
Strategy B	theoretical positions/theoretical frameworks
Strategy C	research design/processes
Strategy D	interpretations of terminology used in the thesis

This study adopts a different approach to Kwan. The macro-level discourse analysis focuses on the types of information/content and the high-level organization of the content. The analysis of rhetorical functions that is on-going does not adopt an *á priori* model such as CARS, but a bottom-up approach of analyzing the text and linguistic expressions, and inferring the intended rhetorical purpose of the expressions found. The rhetorical functions inferred are thus more low-level and closer to the meaning of the linguistic expressions than CARS *moves*. They are at about the same level as CARS strategies.

Research method

Sampling

Twenty research articles were sampled from 8 volumes of *The Journal of the American Society for Information Science and Technology* (JASIST) (2001-2008), as it is a leading journal in the field and carries high-quality research articles with substantive literature reviews. Two or three articles were haphazardly taken from each year (volume) of the electronic version of JASIST. Their literature review sections were extracted for analysis. Ten literature reviews were analyzed to identify the major elements of a literature review and to develop a coding scheme to annotate literature reviews. The second set of ten literature reviews were used to assess the completeness of the coding scheme and to identify further improvements.

Subsequently, a test sample of 30 literature reviews from JASIST, the *Journal of Documentation* and *Online Information Review* (10 from each source) were used to test the inter-coder

reliability of the coding scheme. A random sample of 10 literature reviews was taken from the 2000-2008 issues of each journal, using a random number generator.

Coding Scheme

An XML coding scheme was developed to annotate the structure of literature reviews at the sentence level. Given that a literature review contains summaries of previous studies, it was assumed that a literature review must, by definition, have the following basic structure:

- There is a series of *topic* sections, and subtopics arranged hierarchically
- A *topic* section contains a statement indicating *what* the topic is, and one or more cited *studies*
- A *study* section contains a statement of *what* the study is about, and additional *description* of the cited research, e.g. research *method* and research *result*
- The reviewer's comments, interpretation or critique of the cited studies are often present. The reviewer's comments can serve various purposes: to provide an overview summary of the research in the field, to critique or re-interpret the results of a cited study, to comment on the significance of a study, or to make a comparison of two or more studies.

Ten literature reviews, referred to as the training set, were analyzed by the three authors to apply the basic framework outlined above, and to identify additional structural elements in the document. The following additional elements were found:

- Some sentences mention a few topics/studies within the same sentence. These were referred to as *brief-topics*.
- Some reviews describe the results of a study and go on to provide the cited author's *interpretation* of the results. In the coding scheme, these were distinguished as research *results* and *interpretation*.
- The reviewer may refer to and make comparison with the reviewer's current study. This is referred to as *current-study*.

Based on these findings, an XML schema was constructed to represent the expected elements and their hierarchical relationships of a literature review. Table 2 lists the structural elements of the literature review. A typical structure of a literature review is illustrated in Fig. 1.

Broadly, there are two types of elements. The elements beginning with *meta* (e.g. *meta-critique* and *meta-summary*) indicate the *citing* author's (i.e. reviewer's) comments on the cited studies. In contrast, the *result* and *interpretation* elements refer to the *cited* author's results and interpretation.

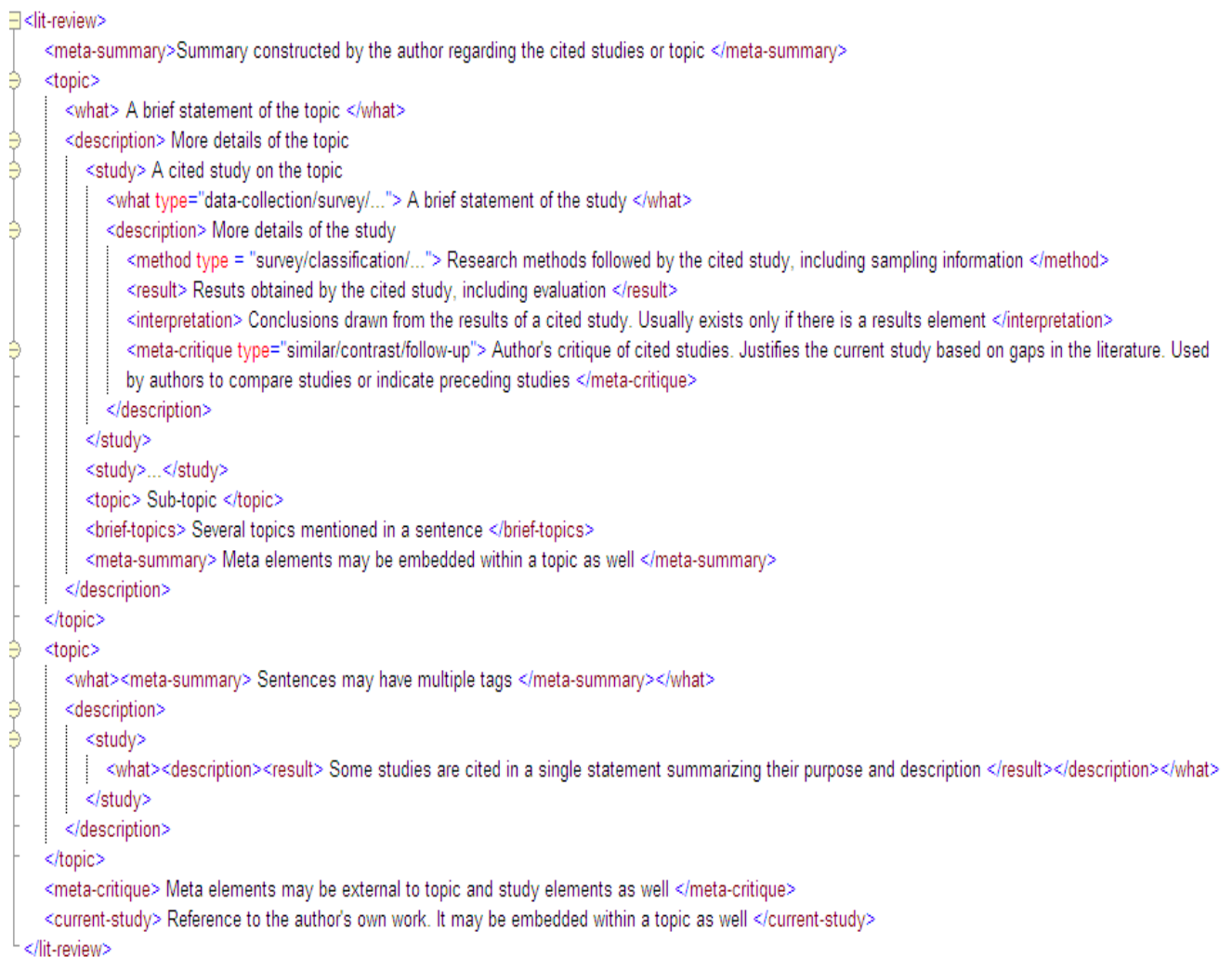


Figure 1. Typical structure of a literature review

As seen in Fig. I, a *topic* element includes a *what* element indicating what the topic is about, followed by a *description* element which provides more details or one or more studies. A *study* element contains a *what* element followed by a *description* element which provides more details such as *method* and *result*. The *topic*, *meta-summary*, *meta-critique*, *brief-topics* and *current-study* elements can occur anywhere in the document structure. The *study* element is allowed to occur only within *topic*. The *what* and *description* elements can occur only within *topic* or *study*. The *method*, *result* and *interpretation* elements are child elements of *study/description*. An example annotated literature review is given in Fig. II.

Table 2. Structural elements of a literature review

Name	Purpose
lit-review	Acts as the overall wrapper for the literature review text. This is a general element, compulsorily used to enclose the entire source text as one element.
topic	Encloses the literature reviewed for a particular topic. It is a mandatory element and may contain embedded topic elements indicating sub-topics. It has an <i>id</i> attribute allocated serially for reference purposes.
study	Encloses the text describing a cited study. It also has an <i>id</i> attribute, used principally for reference purposes in the meta-critique elements.
what	Encloses a short statement of the topic or study. It is a mandatory element embedded within a <i>topic</i> or a <i>study</i> element. It has a <i>type</i> attribute used to specify the higher level genre of the study, for example, algorithm, model, survey etc.
description	Encloses more details of the <i>topic</i> or <i>study</i> element within which it is embedded. It must be preceded by a <i>what</i> element. It also contains other embedded elements which belong within a <i>topic</i> or a <i>study</i> element.
meta-critique	Encloses the author's critique of cited studies. It may contain justification for the current study, critical comparison or an indication that the current study is a follow-up of a previous study and may be embedded within the <i>description</i> element of a <i>topic</i> .
meta-summary	Encloses an objective, noncritical summary of cited studies, constructed by the author and may be embedded within a <i>description</i> element of a <i>topic</i> .
brief-topics	Encloses text which has several studies cited within the span of a single sentence.
current-study	Encloses references to the author's current work (being reported in the paper).
method	Encloses information about the research method or procedure used in a cited study, including sampling information. It is embedded within the <i>description</i> element of a <i>study</i> .
result	Encloses the results of the cited study, including evaluation results if any. It is embedded within the <i>description</i> element of a study and is usually preceded by a <i>method</i> element.
interpretation	Encloses the high-level interpretation, generalizations or conclusions drawn in the cited study. It is embedded within a <i>description</i> element of a <i>study</i> and it usually follows a result element.

«topic»

«what» The existing studies have focused more often on evaluating mental models against a predefined framework - presumably a conceptual model of the system under study. »what»

«description» «meta-critique» Although great efforts have been made to elicit mental models using interviews, few studies have illustrated mental models explicitly in terms of composite elements and structures grounded in the transcript. »meta-critique»

«study» «what» For example, Dimitroff ([1992]) constructed an eight-concept scale based on system documentation and her own experience. »what»

«description» The eight concepts covered the most basic features of an OPAC system, such as database contents, interactive nature of the system, multiple files, and use of controlled vocabulary.

«method» The completeness of mental models was measured by comparing participants' interview transcripts with the scale. In the exploration of the effects of users' characteristics on their mental models, Zhang ([1998]) arrived at nine essential concepts and attributes of IR systems by consulting a group of experts using the repertory grid technique. The nine concepts identified important components of an IR system, such as browsing, classification, data structure, and interface. The three attributes - format/process, targeted/untargeted, and specific to IRs/applicable to all information systems - were the dimensions on which concepts were judged. »method»

«result» Subjects' mental models were measured by their ratings on the concepts and attributes. »result»

»description»

»study»

«study»

«what» Taking a data-grounded approach, Borgman ([1986]) developed a holistic index to evaluate users' mental models based on the interview transcripts. »what»

«description» The index consists of three measurements: completeness of the model, accuracy of the model, and level of abstraction.

«method» Completeness of the model was based on the number of features mentioned. Accuracy of the model measured internal consistency of the models. Level of abstraction was a ranking based on whether the participant articulated an independently invented metaphor. »method»

«meta-critique» In the study, she did not evaluate participants' mental models against a predefined conceptual framework as the two studies reviewed earlier; however, she also was not intent to construct participants' mental models from the interview transcripts. »meta-critique» »description» »study»

»description» »topic»

Figure 2. Sample coded literature review

Reliability Testing of the Coding Scheme

The coding scheme was applied to a new set of 10 literature reviews from JASIST to assess its completeness and reliability. One of the authors coded 10 literature reviews to provide one set of coding. The other two authors collaborated to provide the second coding set. Two undergraduate students were also employed to provide two sets of coding, after being trained on the training set. It was, however, found that they had difficulty identifying some elements because of lack of familiarity with literature reviews. Since only researchers are likely to be familiar with the different aspects and purposes of a literature review, high-quality coding can be provided only by researchers.

Cohen's *Kappa* (Cohen, 1960) inter-coder reliability measure were calculated for each literature review and averaged across the ten reviews. In the calculation, the following procedure was followed:

- For every sentence, the XML element used to tag the sentence was compared for the two sets of coding.

- Sentences with multiple tags were double-counted, i.e., once each as the member of each element. This made the reliability score more sensitive to element hierarchy and presence of embedded elements in the coding.
- In the case of the “container” element *topic*, only the first sentence in each enclosed text was considered as bearing the *topic* tag. Also, the sub-topics embedded within a *topic* were considered as tagged by a single *topic* tag. The reason is that the *topic* tag is all-encompassing; considering all embedded text as an annotation match would inappropriately raise the reliability score.

Cohen’s *Kappa* represents inter-coder agreement after chance agreement is removed from consideration (Cohen, 1960). The equation for κ is (Cohen, 1960):

$$\kappa = \frac{f_o - f_e}{N - f_e}$$

where, f_o is the number of matched sentences from among the two sets of coding and N is the total number of sentences. f_e is the frequency of chance agreement calculated as the joint probability of the two coders randomly making the same decision. The κ value ranges from 0 (only chance agreement) to 1 (complete agreement). The average Kappa value obtained for the 10 literature reviews was 0.836.

To assess the inter-coder reliability for each element in the coding scheme, the *Jaccard Similarity Coefficient* was also calculated for each element. This measures the similarity or overlap between the two sets of sentences coded as that element. The Jaccard Coefficient is calculated as:

$$J = \frac{A \cap B}{A \cup B}$$

This is the size of the intersection between set A and set B (i.e. number of common sentences in the two sets, tagged with the same element), divided by the size of the union of the two sets (the total number of sentences tagged with the element in *either* set A or B). The results are provided in Table 3.

From Table 4, it can be seen that the inter-coder agreement is high for all the structural elements except *brief-topics*. The high agreement is prominent in *current-study* (100%) and *meta-critique* (81%) and *meta-summary* (76%). While the coding sets mostly agreed with each other on defining the high-level elements like *topic* (86%), *study* (89%), *topic/description* (83%) and *study/description* (85%) there was some disagreement on labeling the embedded elements within a *study*, such as *interpretation* (67%). The inter-coder agreement for the *brief-topics* tag is a low 40%.

An example of correct and incorrect tagging for *brief-topics* is as follows:

- **Incorrect:** <*brief-topics*> *Experimental studies have repeatedly shown that when people are asked to think of a term to describe a particular concept, they respond with a variety of terms (Bates, [1998]; Furnas, Landauer, Gomez, & Dumais, [1987]; Gomez, Lochbaum, & Landauer, [1988]; Saracevic, Kantor, Chamis, & Trivison, [1988]).*</*brief-topics*>

- **Correct:** <brief-topics> Researchers were also able to obtain interesting findings on the information needs and search behavior of users, such as the trends in **Web searching** (Spink et al., [2002]), **sexual information searching** on the Web (Spink, Ozmutlu, & Lorence, [2004]), and **question format Web queries** (Spink & Ozmutlu, [2002]).</brief-topics>

The *brief-topic* elements were sometimes confused with the *meta-summary* element where studies were briefly mentioned as examples in topical discussions. There was also some ambiguity about whether a *brief-topics* element was constituted by multiple topics mentioned in a single sentence or multiple studies cited in one sentence. It was decided to combine *brief-topics* with *meta-summary* in the coding scheme.

The coders also faced some difficulty distinguishing interpretations by the cited author (*interpretation* tag) from the critical discussion by the reviewer (*meta-critique* tag), as illustrated in the following examples:

- **Incorrect:** <interpretation> As it is known, these three collections have been extensively used in the past, and consist in a limited number of documents, which are abstracts of scientific papers in medicine, computer science, and aerodynamics, respectively. However the limited effectiveness of stemming was not the only result of Harman's and it would appear discouraging if it were taken without further analysis.</interpretation>
- **Correct:** <interpretation> She pointed out that none of the tested stemmers significantly improves the overall performance of a system which retrieves documents from three test collections: Medlars, CACM, and Cranfield.</interpretation>

It was also sometimes difficult to distinguish *interpretation* from *result*.

The final coding scheme was applied to the test set of 30 literature reviews randomly sampled from JASIST, Journal of Documentation and Online Information Review. Two sets of coding were provided by the authors, in the same way as before. The average inter-coder reliability score (Kappa) obtained was 0.848, with a 95% confidence interval of ± 0.032 . This is considered high agreement (Wood, 2007; Landis & Koch, 1977).

The detailed Jaccard Coefficient values for each element are given in Table 4. There continues to be some confusion between *interpretation*, *meta-critique* and *meta-summary*.

During the coding exercise, the researchers observed that literature reviews were of two distinct styles. Some literature reviews were descriptive and summarized individual papers/studies. Other literature reviews were integrative in style in that the authors extracted the content and results of the cited papers and synthesized a summary of trends and milestones in the research area. Integrative literature reviews contain fewer quotes and expressions taken directly from the cited papers, and is written in the individual style of the author.

Table 3. Inter-coder agreement for 10 literature reviews from JASIST

Element	Jaccard Coefficient
topic	0.83
topic/what	0.83
topic/description	0.72
meta-critique	0.70
meta-summary	0.64
brief-topics	0.40
current-study	1.00
study	0.83
study/what	0.83
study/description	0.78
study/description/method	0.82
study/description/result	0.62
study/description/interpretation	0.61

Table 4. Inter-coder agreement for the test set of 30 literature reviews

Element	Jaccard Coefficient
topic	0.87
topic/what	0.87
topic/description	0.83
meta-critique	0.62
meta-summary	0.58
current-study	0.87
study	0.90
study/what	0.90
study/description	0.84
study/description/method	0.86
study/description/result	0.80
study/description/interpretation	0.48

Torraco (2005, pp. 356) characterized integrative literature review articles as “a form of research that reviews, critiques and synthesizes representative literature on a topic in an integrated way such that new frameworks and perspectives on the topic are generated.” On the other hand, Knott defines descriptive literature reviews articles as similar to annotated bibliographies which briefly summarize the research questions of a research study, its major methods of investigation and its main conclusions (Knott, 1999).

Results: Comparison of Descriptive and Integrative Literature Reviews

The 20 literature reviews from JASIST were individually categorized by the three authors as 'integrative' or 'descriptive' or 'mixed'. The three coders agreed on the categorizations of 17 of the 20 literature reviews. For 2 literature reviews, two authors agreed on a category and the category was assigned by majority agreement. There was complete disagreement for 1 literature review, which was then replaced by a new literature review.

The literature reviews were analyzed to compare the frequencies of various discourse elements in the descriptive versus integrative literature reviews. The *topic* and *study* blocks were counted once per occurrence since they are the subsuming parent elements, while the other elements were counted as the total number of sentences coded as belonging to that particular element in the text. Table 5 provides the frequency of occurrence of the various discourse elements. Table 5 does not take into account the varying lengths of literature reviews, which ranged from a word count of 700-2200 in the set considered. Therefore, Table 6 considers the content of discourse elements as word count normalized by the word count of the entire text being considered. T-Tests were conducted to identify significant differences.

It is seen from Table 6 that integrative literature reviews have significantly more of the following elements:

- *topic/what*
- *topic/description*
- *meta-summary*
- *brief-topics*.

Their normalized average frequencies are more than twice the frequencies for descriptive literature reviews. In contrast, descriptive literature reviews have significantly more of the following elements:

- *study/what*
- *study/description*
- *study/description/method*
- *study/description/result*.

It should be noted that the low average number of *study* elements in integrative literature reviews does not imply a lower number of references. Rather, it reflects a difference in referencing styles. In integrative literature reviews, papers are often cited briefly as examples of typical results found in the literature.

Thus it is clear that the major difference between the two categories is the way in which previous research is cited. While integrative literature reviews provide high-level summaries of topics and cite studies as examples illustrating the author's argument, descriptive literature reviews describe individual studies and provide more information on each study—methods, results and interpretation. Integrative reviews contain more meta-summary and meta-critique elements, indicating that the reviewer's voice is more dominant (rather than using quotes and expressions from the cited papers).

Table 5. Frequency of occurrence of various discourse elements

in integrative and descriptive literature reviews						
Elements	Min Occurrence		Max Occurrence		Avg Frequency	
	*Int	Desc	Int	Desc	Int	Desc
topic	3	1	10	6	6	3
study	3	5	9	13	5	8
topic/what	3	1	10	6	5	3
topic/description	11	0	47	26	25	9
meta-critique	0	0	15	9	7	3
meta-summary	2	0	28	10	9	2
brief-topics	0	0	11	2	3	0
current-study	0	0	7	3	2	1
study/what	3	5	9	13	6	8
study/description	0	8	15	38	7	23
study/description/method	0	0	4	18	1	8
study/description/result	0	1	4	14	1	6
study/description/interpretation	0	0	2	9	0	2

* Note: Int: Integrative literature reviews
 Desc: Descriptive literature reviews

Table 6. Normalized proportion of literature reviews tagged with various discourse elements in the integrative and descriptive literature reviews

Elements	Normalized Avg Frequency (%)		Std Deviation	
	I	D	I	D
topic/what	5.79**	2.47	3.1	1.14
topic/description	39.7**	12.7	10.5	12.7
meta-critique	8.57	4.05	6.08	4.83
meta-summary	11.7**	1.17	4.74	1.93
brief-topics	4.85*	0.7	4.61	1.36
current-study	2.41	1.95	3.24	2.75
study/what	9.05	15.8**	4.52	4.88
study/description	11.9	32.8**	6.69	7.87
study/description/method	2.18	13.3**	2.9	7.71
study/description/result	2.57	11.4**	2.3	6.87
study/description/interpretation	1.3	3.73	2.42	5.21

* significant at the 0.05 level

** significant at the 0.01 level

Table 7. Discourse structure of four sample literature reviews

Descriptive Literature Reviews Simple structure		Integrative Literature Reviews Complex Structure	
D1	D2	I1	I2
lit-review	lit-review	lit-review	lit-review
topic	topic	meta-summary	topic
study	study	topic	meta-critique
study	study	meta-summary	study
study	study	topic	study
study	study	study	topic
study	meta-critique	meta-critique	topic
topic	study	topic	brief-topics
study	...	study	study
study	study	meta-critique	topic
study	study	topic	meta-summary
	study	...	topic
	study	study	
	meta-critique	meta-critique	
	current-study	topic	
		meta-critique	
		study	
		...	
		meta-critique	
		topic	
		...	
		meta-critique	
		current-study	

Table 7 illustrates the typical structure of descriptive and integrative literature reviews with examples from the collection. Table 7 corroborates the findings in Tables 5 and 6. Documents D1 and D2 have numerous *study* elements, and I1 and I2 have numerous *meta-critique* and *meta-summary* elements. It is also seen that D1 and D2 have fewer *topic* elements and are simpler in structure. There is only one level of topic elements, i.e. no sub-topics (no embedded *topic* elements). On the other hand, I1 and I2 have numerous *topic* elements and are relatively complex in structure. There are a few levels of embedded *topic* elements. In I1, the topic elements are embedded to the depth of four (i.e. four levels of topics and sub-topics). I2 has topic elements embedded to the depth of three.

We analyzed the maximum depth of the topic elements in the 20 literature reviews, and found that integrative literature reviews have an average maximum depth of 2.5, compared to 1.4 for descriptive literature reviews. The difference is significant at the 0.01 level.

Conclusion

We have developed a coding scheme for analyzing the macro-level discourse structure of literature reviews, focusing on types of content. The main elements of the coding scheme are *topic* and *study*. The scheme distinguishes between the cited author's interpretations and the citing author's (reviewer's) comments, represented by *meta-summary* and *meta-critique* elements. An inter-coder reliability test based on a random sample of 30 literature reviews obtained a Kappa value of 0.848, which is considered high agreement.

The study found that literature reviews are written in two distinctive styles, with different discourse structures. *Descriptive literature reviews* summarized individual papers/studies and provide more information on each study—methods, results and interpretation. *Integrative literature reviews* have few study elements and instead provide high-level critical summaries of topics. They have more of *meta-summary* and *meta-critique* elements, indicating that the reviewer's voice is dominant. They have more levels of topics and sub-topics, with a more complex structure. They tend to cite studies as examples illustrating the author's argument.

It should be pointed out, however, that most literature reviews have elements of both descriptive and integrative elements. The two styles represent two poles of a continuum, and particular literature reviews can probably be placed on this continuum to represent the relative proportion of descriptive and integrative elements. Since literature reviews are traditionally expected to discuss previous work in the field and frame an argument through the meta elements, it is expected that a good literature review will have balance of meta elements and description elements.

As future work, this study will be extended to other information science journals so that a comparison can be made between the literature reviews written for different journals. Work is ongoing to analyze the discourse relations between text units (mainly clauses), their rhetorical functions and the linguistic expressions used to realize them. Content analysis is also being carried out to find out how information is selected from the cited papers (source) to include in the literature review, its location in the source papers and the extent the information is paraphrased in the literature review. It is hoped that the results of these studies will shed light on how automatic methods can be developed to generate multi-document summaries of research literature, with the characteristics of literature reviews.

References

- Barsky, R. (2001), "Discourse Analysis Theory", *McGill University*, available at: http://74.125.155.132/scholar?q=cache:BVJ2ODu6KT0J:scholar.google.com/&hl=en&as_sdt=2000 (accessed May 13, 2010).
- Barzilay, R., and McKeown, K. R. (2005), "Sentence Fusion for Multidocument News Summarization", *Computational Linguistics*, Vol. 31 No. 3, pp. 297-328.
- Boote, D. N., and Beile, P. (2005), "Scholars before researchers: On the centrality of the dissertation literature review in research preparation", *Educational Researcher*, Vol. 34 No. 6, pp. 3-15.
- Bourner, T. (1996), "The research process: four steps to success", In Greenfield, T. (Ed), *Research Methods: Guidance for Postgraduates*, Arnold, London, pp. 7-11.

- Bruce, C. (1994), "Research student's early experiences of the dissertation literature review", *Studies in Higher Education*, Vol. 19 No. 2, pp. 217-229.
- Bunton, D. (2002), "Generic moves in Ph.D Introduction chapters", In Flowerdew, J. (Ed), *Academic Discourse*, Longman, London, pp. 57-75.
- Cohen, J. (1960), "A coefficient of agreement for nominal scales", *Educational and Psychological Measurement*, Vol. 1 No.1, pp. 37-46.
- Cooper, H. (1988), "The structure of knowledge synthesis", *Knowledge in Society*, Vol. 1, pp. 104-126.
- Elmore, R. (2006). "Some guidance on doing a literature review", *Harvard University*, available at: http://isites.harvard.edu/fs/docs/icb.topic178762.files/Literature_Reviews/Some_Guidance_on_Doing_a_Literature_Review.doc (accessed May 13, 2010).
- Endres-Niggemeyer, B., Maier, E., and Sigel, A. (1995), "How to implement a naturalistic model of abstracting: Four core working steps of an expert abstractor", *Information Processing & Management*, Vol. 31 No. 5, pp. 631-674.
- Hart, C. (1998), *Doing a Literature Review*, Sage, London.
- Hinchliffe, L. (2003), "Having your say in a scholarly way", *Research Strategies*, Vol. 19 No. 3-4, pp. 163-164.
- Jonsson, S. (2006), "On academic writing", *European Business Review*, Vol. 18 No. 6, pp. 479-490.
- Knott, D. (1999), "Writing an annotated bibliography", *University of Toronto*, available at: <http://www.writing.utoronto.ca/advice/specific-types-of-writing/annotated-bibliography> (accessed May 13, 2010).
- Kwan, B. S. C. (2006), "The schematic structure of literature reviews in doctoral theses of applied linguistics", *English for Specific Purposes*, Vol. 25 No. 1, pp. 30-55.
- Landis, J., & Koch, G. (1977), "The measurement of observer agreement for categorical data", *Biometrics*, Vol. 33 No. 1, pp. 159-174.
- Marcu, D. (2000), *The Theory and Practice of Discourse Parsing and Summarization*, MIT Press, Cambridge, MA.
- Massey, A. (1996), "Using the Literature: 3 x 4 Analogies", *The Qualitative Report*, Vol. 2 No. 4, available at: <http://www.nova.edu/ssss/QR/QR2-4/massey.html> (accessed August 28, 2009).
- McKeown, K. R., Barzilay, R., Evans, D., Hatzivassiloglou, V., Klavans, J. L., Nenkova, A., Sable, C., Schiffman, B., Sigelman, S. (2002), "Tracking and summarizing news on a daily basis with Columbia's Newsblaster", In *Proceedings of the second international conference on Human Language Technology Research*, Morgan Kaufmann Publishers Inc., San Diego, California, pp. 280-285.
- Molina, M. P. (1995), "Documentary abstracting: towards a methodological model", *Journal of the American Society for Information Science*, Vol. 46 No. 3, pp. 225-234.
- Rowley, D. J., and Slack, D. F. (2004), "Conducting a Literature Review", *Management Research News*, Vol. 27 No. 6, pp. 31.
- Rudestam, K. E. and Newton, R. (2001), *Surviving Your Dissertation*, Sage Publications Inc, Thousand Oaks, CA.
- Swales, J. (1981), "Aspects of articles introductions", *Aston ESP Reports No. 1*, The University of Aston, Birmingham.
- Swales, J. (1990), *Genre Analysis: English in academic and research settings*, Cambridge University Press, New York.
- Torraco, R. J. (2005) "Writing Integrative Literature Reviews: Guidelines and Examples", *Human Resource Development Review*, Vol. 4 No. 4, pp. 356-367.
- Wood, J.M. (2007), "Understanding and computing Cohen's Kappa: A Tutorial", *WebPsychEmpiricist*, available at <http://www.wpe.info/vault/wood07/Wood07.pdf> (accessed May 13, 2010).