



## **Comparing Rhetorical Structure of Research Article Abstracts Written by Native English Writers and Chinese Writers**

**Beibei Xu**

*Korea University, Korea*

### **Introduction**

Writing abstracts in English is a requisite for the publication of research articles in both international and domestic academic journals. Research article abstracts, as a faithful and factual description of the whole research, draw people's attention because they will decide whether the accompanying article deserves further reading (Bhatia, 2014). There are numerous studies concerning research article abstracts due to its importance, ranging from the microlevel of linguistic features to the macrolevel of textual organization. On the microlevel, many researchers are concerned with the linguistic features of abstracts, such as hedges, boosters, evaluative words, modality, tense and so on (e.g., Hu & Cao, 2011; Hyland & Tse, 2005; Pho, 2008). On the macrolevel, most studies are about the generic structure of research article abstracts following Swales' (1990) Introduction-Method-Results-Discussion (IMRD) model, where the focus falls on the different moves of abstracts in different disciplines (e.g., Lorés, 2004; Martin, 2003; Salager-Meyer, 1992; Samraj, 2005).

Rhetorical Structure Theory (RST) was conceived by Mann and Thompson (1988) to "characterize text and textual relations" (Taboada & Mann, 2006b, p. 567). Although RST has been widely used as a linguistic approach to describe textual organization (Taboada & Mann, 2006a), few studies have been done on the general rhetorical structure of research article abstracts. This work aims to compare the coherence relations of research article abstracts written by native English writers and Chinese writers under the framework of RST. With the comparison of rhetorical structures, it is hoped to find out the similarities and differences in coherence relations and thus help Chinese researchers to polish their writing so that their research can be better accepted in the international academic circle. By analysing twenty English abstracts written by native speakers and non-native speakers, the researcher presents the macrostructure and microstructure of research article abstracts. Based on RST, two research questions have been formulated:

- 1) What are the global rhetorical structures of research article abstracts written by native English writers and Chinese writers?
- 2) What are their similarities and differences?

### Theoretical Background

The basic notions in RST are *rhetorical relations* and *spans*. The structure and coherence of a natural text can be “realized by various rhetorical relations that hold between different spans” (Mann & Thompson, 1988, p. 243). The text units which are in the minimal form of clauses or sentences are termed as *spans*. Spans can also be composed of other spans. They are related recursively by rhetorical relations (e.g., Elaboration, Background, Antithesis, Evidence, Purpose, Contrast, etc.).

The relative status of *spans* in RST is termed as *nuclei* and *satellites* respectively. In a rhetorical relation, if a span is considered more essential, it is a nucleus; if a span is on a secondary position and provides extra information about the nucleus, it is a satellite. The distinction is made between nuclear relations and multinuclear relations. The nuclear relations denote a hypotactic relationship between a nucleus and one or more satellites, e.g., circumstance, elaboration, purpose, etc. Paratactic relations which are composed of two or more nuclei are referred to as multinuclear relations, for example, List, Contrast, Joint, and so on. There are two kinds of nuclear relations: subject-matter relations and presentational relations. The classification of rhetorical relations is presented in Table 1.

TABLE 1  
Classification of RST Relations

Nuclear relations		Multinuclear relations
Subject-Matter relations	Presentational relations	
Circumstance, Condition, Elaboration, Evaluation, Interpretation, Means, Non-volitional cause, Non-volitional result, Otherwise, Purpose, Solutionhood, Unconditional, Unless, Volitional cause, Volitional result	Antithesis, Background, Concession, Enablement, Evidence, Justify, Motivation, Preparation, Restatement, Summary	Conjunction, Contrast, Disjunction, Joint, List, Sequence

The definition of an RST relation is determined by four factors: 1) constraints on the nucleus, 2) constraints on the satellite, 3) constraints on the combination of the nucleus and satellite, and 4) the effect achieved on the reader (Mann & Thompson, 1988, p. 245). Besides Relations, RST also defines Schemas, which specify how spans of text can co-occur to construct RST structures. Five kinds of schemas are recognized in RST (see Figure 1).

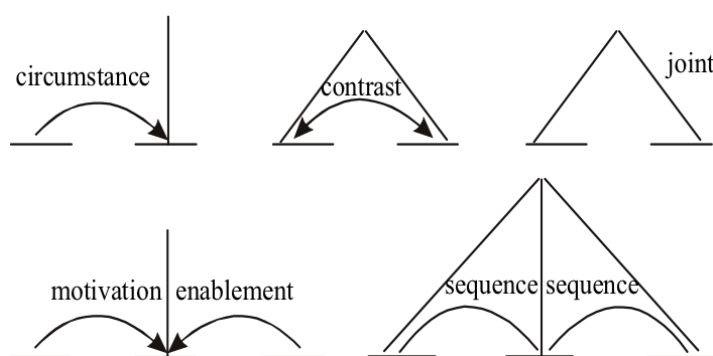


Figure 1. Examples of five schema types. (Mann & Thompson, 1988, p. 247)

Mann and Thompson (1988) point out that the relation judgement cannot avoid subjectivity because the judgment, which is based on the researchers’ understandings of context and intentions of the writer, is about plausibility but not certainty. An RST analysis is a set of schema applications and some constraints

are also prescribed, such as completeness, connectedness, uniqueness and adjacency (Mann & Thompson, 1988, pp. 248-249).

RST has been widely applied in various fields, including computational linguistics, discourse analysis and evaluation of second language writing (Taboada & Mann, 2006b). The well-defined RST relations and the clear analysing procedures ensure RST as a useful analytical tool to interpret different types of text, be it argumentative, editorial, news report and so on (e.g., Azar, 1999; Mu, 2017). It has also been tested as an effective tool in contrastive rhetoric (Taboada & Mann, 2006b). For instance, Sarjala (1994) compared English and Finnish psychology article in terms of reason and cause relations. Abelen, Redeker and Thompson (1993) compared the rhetorical relations of fundraising letters between English and Dutch. While there have been numerous analyses of various texts by applying RST, there are hardly any systemic studies revealing the textual rhetorical structure of research article abstracts. There is only one study done by Rimrott (2007), who tentatively examined ten research article abstracts and summarized two universal structures, i.e., the elaboration structure and the background-other structure. The comparison study of research article abstracts under the framework of RST, however, has not been found in the literature. The lack of empirical findings in this area motivates the current study.

## Method

### Data Collection

This is a preliminary study, so the corpus size is very small. Our corpus contains twenty English abstracts which were selected from the empirical research articles in the field of second language acquisition. Among the twenty, ten English abstracts written by native speakers (NEA) are taken from *Second Language Research* and *Applied Linguistics*, which are internationally influential SSCI journals. Some articles are written by more than one writer. The researcher googled their background information to ensure that at least one of them was a native speaker. Ten abstracts written by Chinese speakers (CEA) were taken from *Foreign Language Teaching and Research* and *Modern Foreign Languages*, which are two of the leading journals in China. A brief description of the two corpora is found in Table 2.

TABLE 2  
*Description of the Corpus*

	NEA corpus	CEA corpus
words	1537	1348
sentences	57	53
words per sentence	26.96	25.53

The abstracts in the NEA corpus ranged in length from 87 to 190 words with the average being 153.7 words per abstracts; the abstracts in CEA corpus ranged from 89 to 168 with the average being 134.8 words per abstracts. Their mean sentence length is also quite similar, 26.96 words per sentence for NEA and 25.53 words per sentence for CEA. Appendix 1 provides detailed information about the abstracts.

### Rhetorical Annotation

The analysis process included the following steps: the text segmentation which involves the definition of elementary discourse unit (EDU), rhetorical annotation which involves the identification of relations and spans, and generation of tree-like diagrams via RSTTool 3.0 developed by Michael O'Donnell. During this process, the researcher herself was responsible for all judgments made, so subjectivity seems unavoidable. With the help of annotation guidelines for rhetorical structure (Stede & Taboada, 2017), the

researcher tried to minimize it by analysing the data repeatedly (three times) based on the clear understanding of the mononuclear and multinuclear relations as defined by Mann and Thompson (1988).

## **Text Segmentation**

The first step in generating an RST diagram was to segment the text into EDUs. The units can be as short as clauses or as long as paragraphs based on the researchers' needs. Although unit size is arbitrary, the division of the text into units was based on theory-neutral classification (Mann & Thompson, 1988, p. 248). In this work, units are essentially clauses. Independent, subordinate, non-restrictive clauses are considered distinct units whereas clausal subjects and complements and restrictive relative clauses are considered parts of their host clause units. Concerning units appearing in parentheses, the researcher segmented them when they contained a finite verb and clearly showed some rhetorical relations with other units. The following is an example of the segmentation of units.

### **Sample Text of Segmentation:**

[In order to identify the causes of inflectional variability in adult second language (L2) acquisition,] [this study investigates lexical and syntactic aspects of gender processing in realtime L2 production and comprehension.] [Twenty advanced to near-native adult first language (L1) English speakers of L2 German and 20 native controls were tested in a study comprising two experiments.] [In elicited production, we probe accuracy in lexical gender assignment.] [In a visual-world eye tracking task, we test the predictive processing of syntactic gender agreement between determiners and nouns.] [The findings show clear contingencies (1) between overall accuracy in lexical gender assignment in production and target predictive processing of syntactic gender agreement in comprehension] [and (2) between the speed of lexical access and predictive syntactic gender agreement.] [These findings support lexical and computational accounts of L2 inflectional variability] [and argue against models positing representational deficits in morphosyntax in late L2 acquisition and processing.]

## **Annotation Tool**

The annotation of the texts and the generation of tree-like diagrams were done with the help of RST Tool version 3.0. Different sets of rhetorical relations are embedded in this software: 1) the classic one containing 24 relations as defined by Mann and Thompson (1988), 2) the extended one of 30 relations (Mann & Taboada, 2010), and 3) Marcu's classification of 136 relations (Carlson, Marcu, & Okurowski, 2002). For the convenience of the analysis, the extended classification was chosen and the researcher used result relation to replace non-volitional result since all the result relations in abstracts are non-volitional. After segmenting the abstracts into EDUs by using RSTTool 3.0, the researcher did the annotation work.

## **Results and Discussion**

This section compares the distribution of rhetorical relations used in NEA and CEA respectively. The similarities and differences between the two corpora are compared from quantitative and qualitative perspectives. The global rhetorical structure of English abstracts has been generated.

## **Comparison of Rhetorical Relations**

Table 3 and Table 4 show the distribution of different RST relations in the two corpora. With respect to the types and numbers of rhetorical relations, the NEA corpus contains 81 RST relations distributed in 16

types, while the CEA corpus contains only 65 relations distributed in 14 types. In NEA, the number of relations used ranges from 6 to 12 with the average 8.1 while in CEA the number ranges from 3 to 13 with an average of 6.5. The result indicates that the rhetorical structures of English abstracts written by native speakers are much more complex than those written by Chinese speakers, taking into consideration the similar mean sentence length and average sentences per abstract.

TABLE 3  
*Distribution and Frequency of RST Relations in EEA*

RST relation	01	02	03	04	05	06	07	08	09	10	Total	Percent
Elaboration	2	1	2	3	1	-	3	-	3	1	16	19.8%
Result	1	1	1	1	1	1	1	1	1	1	10	12.3%
Interpretation	-	1	2	1	1	1	1	1	1	1	10	12.3%
Means	1	1	-	1	1	-	1	1	1	1	8	9.9%
Contrast	-	-	1	3	-	1	-	-	2	1	8	9.9%
List	-	2	1	1	1	1	1	-	-	-	7	8.6%
Background	1	-	1	1	2	-	-	1	-	-	6	7.4%
Concession	-	-	-	-	-	1	1	1	-	1	4	4.9%
Sequence	1	-	-	-	-	-	-	1	-	1	3	3.7%
Purpose	-	1	-	-	-	-	-	1	-	-	2	2.5%
Joint	-	-	2	-	-	-	-	-	-	-	2	2.5%
Antithesis	-	1	-	-	-	-	-	-	-	-	1	1.2%
Circumstance	-	-	-	1	-	-	-	-	-	-	1	1.2%
Cause	-	-	-	-	-	1	-	-	-	-	1	1.2%
Condition	-	-	-	-	-	-	-	-	1	-	1	1.2%
Motivation	-	-	-	-	-	-	-	-	-	1	1	1.2%
Total	6	8	10	12	7	6	8	7	9	8	81	100%

TABLE 4  
*Distribution and Frequency of RST Relations in CEA*

RST relation	01	02	03	04	05	06	07	08	09	10	Total	Percent
List	1	1	3	1	1	1	1	1	1	-	11	16.9%
Result	1	1	1	1	1	1	1	1	1	1	10	15.4%
Elaboration	1	1	2	2	1	1	-	-	1	-	9	13.8%
Circumstance	-	1	1	2	-	1	-	-	1	2	8	12.3%
Interpretation	-	1	1	1	1	1	-	-	-	1	6	9.2%
Joint	-	1	-	-	-	-	1	1	1	1	5	7.7%
Means	-	-	1	-	-	1	1	-	-	1	4	6.2%
Contrast	-	-	2	-	-	-	-	-	-	1	3	4.6%
Background	-	-	1	-	-	-	1	-	-	-	2	3.1%
Concession	1	-	1	-	-	-	-	-	-	-	2	3.1%
Sequence	-	-	-	-	2	-	-	-	-	-	2	3.1%
Motivation	-	-	-	-	1	-	-	-	-	-	1	1.5%
Evidence	1	-	-	-	-	-	-	-	-	-	1	1.5%
Preparation	1	-	-	-	-	-	-	-	-	-	1	1.5%
Total	6	6	13	7	7	6	5	3	5	7	65	100%

Concerning the frequency of rhetorical relations, native speakers and Chinese L2 speakers of English also show different preferences. Table 5 shows the top 5 most frequent relations in NEA and CEA. The most frequent 5 relations used by both groups occupy more than half of all the relations. However, it should be noted that Chinese speakers rely more heavily on the 5 most frequent relations, which suggests Chinese speakers' lack rhetorical means to convey their intentions. It also proves the simplicity of English abstracts written by Chinese writers in terms of rhetorical structures.

TABLE 5  
*Most Frequent RST Relations in EEA & CEA*

Rhetorical relations	EEA	Rhetorical relations	CEA
Elaboration	19.8%	List	16.9%
Result	12.3%	Result	15.4%
Interpretation	12.3%	Elaboration	13.8%
Means	9.9%	Circumstance	12.3%
Contrast	9.9%	Interpretation	9.2%
Total	51.9%	Total	67.6%

One interesting finding is that both native speakers and L2 speakers prefer to use Elaboration, Result and Interpretation most frequently. This result is not surprising given the function of abstracts. An abstract should state what the research is about and include the major sections of an academic paper. In addition, the importance and value of the research should also be stated in an abstract. The high frequency of these relations is related to the generic nature of abstracts. This generic nature can be better illustrated by the most essential rhetorical relations appearing in NEA and CEA. The essentiality is measured by the number of abstracts containing a certain relation, as shown in Table 6.

TABLE 6  
*Most Essential RST Relations in EEA & CEA*

Rhetorical relations	No. of abstracts (EEA)	Rhetorical relations	No. of abstracts (CEA)
Result	10	Result	10
Interpretation	9	List	9
Means	8	Elaboration	7
Elaboration	8	Interpretation	6
List	6	Circumstance	6

Some scholars (Lorés, 2004; Salager-Meyer, 1992; Samraj, 2005) point out that the generic structure of an abstract should mirror the IMRD structure of a research article. That is to say, the IMRD structure should also be used in writing abstracts, i.e., the Introduction, Method, Results and Discussion parts are usually included in an abstract, particularly for an empirical research study. It is clear that Result is the most essential rhetorical step both in NEA and CEA with full appearance in all of the 20 abstracts. This is because the most basic part of an empirical research is its results. It highly matches the generic nature of an abstract. In NEA, Interpretation and Means are also very necessary and they represent the Method and Discussion moves in writing an abstract. However, means relation is not regarded as an essential relation by Chinese writers. The most essential RST relations in CEA coincide with their frequency although the order may vary. Those relations such as List, Elaboration and Circumstance are quite multi-functional and cannot reflect the basic moves in abstracts. It seems the rhetorical relations employed by Chinese speakers are not very consistent with the generic structure of abstracts.

## Global Rhetorical Structure

Since the abstracts in the NEA corpus were selected from the internationally leading and influential SSCI journals, their rhetorical patterns are considered as a kind of standard. The above discussions demonstrate that the abstracts in the NEA are more closely related to the basic moves needed in writing an abstract. So, when the global rhetorical structures are concerned, the discussion is mainly based on the analysis of NEA. The following two figures demonstrate two alternative global rhetorical structures of English abstracts for empirical studies.

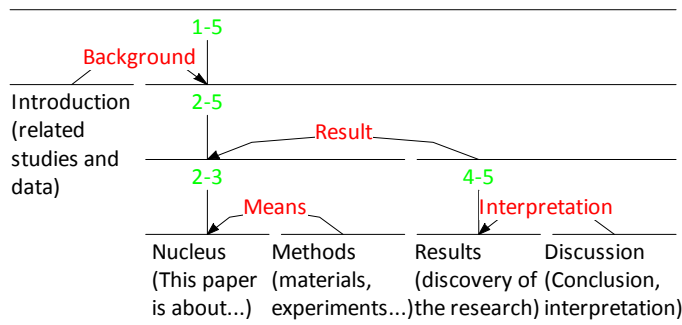


Figure 2. Type 1 global rhetorical structure of English abstracts.

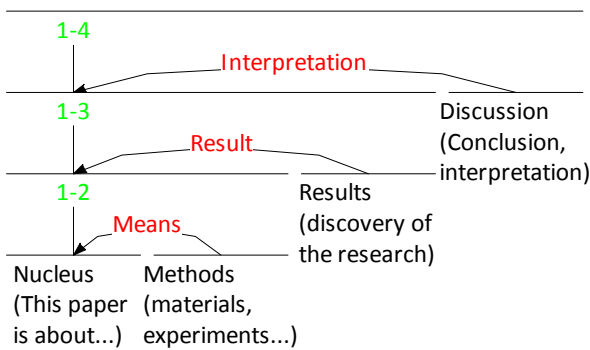


Figure 3. Type 2 global rhetorical structure of English abstracts.

In the Type 1 structure, the satellite which indicates Background RST relations occurs in half of our data although it is not counted as one of the essential relations. The satellite indicating Interpretation relation is connected to the results part. The two combine together as a larger span with the results part as the nucleus to be the satellite for the Nucleus of the whole abstract, which is the main content of an abstract; whereas in the Type 2 structure, the discussion part showing Interpretation relation is directly related to the Nucleus of the text. Type 1 structure coincides with the IMRD macrostructure of abstracts in genre studies.

Compared with NEA, the global structure of CEA can be illustrated by the following figure.

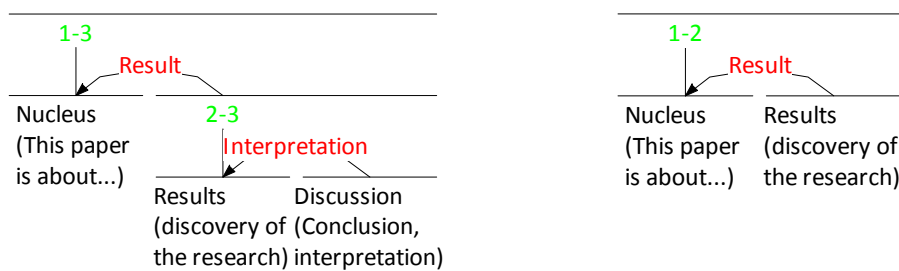


Figure 4. Global rhetorical structure of CEA.

It is clear via comparison of these figures that some information has been missing in CEA, which causes difficulties in understanding. Some English abstracts written by Chinese writers only contain the result relation, which suggests that Chinese writers just report the findings of their studies without claiming the importance or value of their studies.

## Problems with CEA

The research has more difficulties in identifying relations for CEA. In some cases, for better understanding, the researcher must refer to the Chinese version of those abstracts for more precise judgment. However, when referring to the Chinese versions, there are more problems arising. Some Result relations in Chinese are turned into Interpretation relations in English. This phenomenon is partly due to the syntactic and semantic mismatch in the English abstracts written by Chinese writers. The comparison between the Chinese version and English version of CEA 02 can be an example to illustrate the problem.

Chinese version as a multinuclear relation of List:

... 在习得照应语的初始阶段, 学习者不仅依靠正面证据和母语迁移, 而且更多地依赖语篇层面的信息。

English version as a nuclear-satellite relation of Interpretation:

... It is suggested that at the initial stage of the acquisition of the English reflexives, not only do pre-intermediate learners rely on positive evidence and first-language transfer, but they also rely strongly on discourse-level information.

In the English version, by adding 'It is suggested...', the writer changes the results in the Chinese version into interpretations of the result. Actually, in Chinese, there are no such signalling words as 'it is suggested...' to show the RST relation. The sentence is just listed as part of the findings. It is unclear whether the English versions are translated from the Chinese versions or written directly in English by the writers. If they are translated versions, there must be translation strategies involved to explain the differences. In Chinese, coherence is mainly achieved through semantic meanings, that is to say, cohesive devices are seldom used between sentences since Chinese is a language paying less attention to form. While English is a language which pays more attention to form, Chinese writers try to add some cohesive devices in order to make the English version more coherent. However, this kind of addition leads to the mismatch between semantic meaning and syntactic representation. In this sense, it is implied that there are some problems in Chinese writers' use of cohesive devices to construct a coherent text.

## Conclusion

This study compares RST relations between English abstracts written by native speakers and L2 speakers whose L1 is Chinese. Through the comparative analysis, it is found that the two groups show different global rhetorical structure of English abstracts. Native speakers' use of rhetorical relations is more complicated and colourful than their Chinese counterparts. The study also shows that the global rhetorical structures of NEA mirror the generic IMRD structure of abstracts while the rhetorical structures of CEA show the incomplete moves of the same genre. There also exist some problems in analysing RST relations of the CEA.

The study suggests that the Chinese writers need to be more concerned with the rhetorical structure of their English abstracts in order to improve the coherence and accessibility of their English abstracts. For this purpose, the association between RST relations and the content they express should be made clear.

This work is just a very preliminary pilot study for comparing different coherent relations in the framework of Rhetorical Structure Theory. The corpus contains only 20 abstracts and for more objective findings, it is necessary to expand the corpus. The coherence relations are reflected not only by the RST relations but also via some other factors, such as connectives and other cohesive devices. As for future work, it is encouraging to compare the connectives used by native speakers and Chinese speakers to further investigate the differences between NEA and CEA. It is hoped that such studies can help Chinese



researchers to better organize their English abstracts and then improve their influence in the academic circle.

### Acknowledgements

The research was supported by the China Scholarship Council (No. 201508260003) and Qingdao Agricultural University (Grant No. 1114Y50).

### The Author

Beibei Xu is a PhD student at Korea University. Her research interests cover language assessment, second language teaching and learning, and discourse analysis.

Department of English Education  
Korea University  
Seoul, 02841, Korea  
Tel: +82 232902350  
Mobile: + 82 1051380126  
Email: hsubeibei@gmail.com

### References

- Abelen, E., Redeker, G., & Thompson, S. A. (1993). The rhetorical structure of US-American and Dutch fund-raising letters. *Text: Interdisciplinary Journal for the Study of Discourse*, 13(3), 323-350. doi:10.1515/text.1.1993.13.3.323
- Azar, M. (1999). Argumentative text as rhetorical structure: An application of rhetorical structure theory. *Argumentation*, 13(1), 97-114. doi:10.1023/A:1007794409860
- Bhatia, V. K. (2014). *Analysing genre: Language use in professional settings*. New York: Routledge.
- Carlson, L., Marcu, D., & Okurowski, M. E. (2002). *RST discourse treebank*. Philadelphia: Linguistic Data Consortium.
- Hu, G., & Cao, F. (2011). Hedging and boosting in abstracts of applied linguistics articles: A comparative study of English-and Chinese-medium journals. *Journal of Pragmatics*, 43(11), 2795-2809. doi:10.1016/j.pragma.2011.04.007
- Hyland, K., & Tse, P. (2005). Hooking the reader: A corpus study of evaluative *that* in abstracts. *English for Specific Purposes*, 24(2), 123-139. doi:10.1016/j.esp.2004.02.002
- Lorés, R. (2004). On RA abstracts: From rhetorical structure to thematic organisation. *English for Specific Purposes*, 23(3), 280-302. doi:10.1016/j.esp.2003.06.001
- Mann, W. C., & Taboada, M. (2010). *RST web site*. Available at: <http://www.sfu.ca/rst>.
- Mann, W. C., & Thompson, S. A. (1988). Rhetorical structure theory: Toward a functional theory of text organization. *Text: Interdisciplinary Journal for the Study of Discourse*, 8(3), 243-281. doi:10.1515/text.1.1988.8.3.243
- Martín, P. M. (2003). A genre analysis of English and Spanish research paper abstracts in experimental social sciences. *English for Specific Purposes*, 22(1), 25-43. doi:10.1016/S0889-4906(01)00033-3
- Mu, Teng (2017). *A contrastive study of English and Chinese disaster news discourses in the framework of Rhetorical Structure Theory* (Unpublished master's thesis). Huazhong Normal University, Wuhan.

- 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*, 10(2), 231-250. doi:10.1177/1461445607087010
- Rimrott, A. (2007). The discourse structure of research articles abstracts: A rhetorical structure theory (RST) analysis. In *Proceedings of the 22nd Northwest Linguistics Conference (NWLC) at Simon Fraser University*. Canada: Linguistics Graduate Student Association (pp. 207-220).
- Salager-Meyer, F. (1992). A text-type and move analysis study of verb tense and modality distribution in medical English abstracts. *English for Specific Purposes*, 11(2), 93-113. doi:10.1016/S0889-4906(05)80002-X
- Samraj, B. (2005). An exploration of a genre set: Research article abstracts and introductions in two disciplines. *English for Specific Purposes*, 24(2), 141-156. doi:10.1016/j.esp.2002.10.001
- Sarjala, M. (1994). Signalling of reason and cause relations in academic discourse. *Anglicana Turkuensia*, 13, 89-98.
- Stede, M., Taboada, M., & Das, D. (2017). *Annotation guidelines for rhetorical structure*. Retrieved from [http://www.sfu.ca/~mtaboada/docs/research/RST\\_Annotation\\_Guidelines.pdf](http://www.sfu.ca/~mtaboada/docs/research/RST_Annotation_Guidelines.pdf)
- Swales, J. (1990). *Genre analysis: English in academic and research settings*. Cambridge: Cambridge University Press.
- Taboada, M., & Mann, W. C. (2006a). Rhetorical structure theory: Looking back and moving ahead. *Discourse Studies*, 8(3), 423-459. doi.org/10.1177/1461445606061881
- Taboada, M., & Mann, W. C. (2006b). Applications of rhetorical structure theory. *Discourse Studies*, 8(4), 567-588. doi:10.1177/1461445606064836