



저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

문학석사학위논문

An Automatic Analysis of
Argumentation Schemes of Korean Texts

한국어 텍스트 논증 구조의 자동 분석 연구

2016 년 2 월

서울대학교 대학원

언어학과 언어학전공

이 상 아

Abstract

An Automatic Analysis of Argumentation Schemes of Korean Texts

Lee, Sangah

Department of Linguistics

The Graduate School

Seoul National University

These days, there is an increased need to analyze mass opinions using on-line text data. These tasks need to recognize the argumentation schemes and main contents of subjective, argumentative writing, and the automatization of the required procedures is becoming indispensable.

This thesis constructed the text data using Korean debates on certain political issues, and defined the types of discourse relations between basic units of text segments. The discourse relations are classified into two levels and four subclasses, according to the standards which determine whether the two segments are related to each other in a context, whether the relation is coordinating or subordinating, and which of the two units in a pair is supported by the other as a more important part.

The relations between basic text units are predicted based on machine learning and rule-based methods. The features for the prediction of discourse

relations include what the author of a text wants to claim and argumentative strategies comprising grounds for the author's claim, using linguistic properties shown in texts. The strategies for argument are observed and subcategorized into Providing Examples, Cause-and-Effects, Explanations in Detail, Restatements, Contrasts, Background Knowledge, and more. These subclasses compose a broader class of discourse relations and became the basis for features used during the classification of the relations.

Some linguistic features refer to those of previous studies, they are reconstituted in a revised form which is more appropriate for Korean data. Thus, this study constructed a Korean debate corpus and a list of connectives specialized to deal with Korean texts to include in the experiment features. The automated prediction of discourse relations based on those features is suggested in this study as a unique model of argument mining.

According to the results of experiments predicting discourse relations, the features defined and used in this study are observed to improve the performance of prediction tasks through positive interactions with each other. In particular, some explicit connectives, dependent sentence structures based on lack of certain components, and whether the same meanings are restated clearly contributed to the classification tasks.

The discourse relations between basic text units are related and combined with each other to comprise a tree-form argumentation structure for the overall document. Regarding the argumentation structure, the topic sentence of the document is located at the root node in the tree, and it is assumed that the nodes of sentences or clauses right below the root node contain the most important contents as grounds for the topic unit. Therefore, extraction of the

text segments directly supporting the topic sentence may help in obtaining the important contents in each document. This can be one of the useful methods in text summarization. Additionally, applications to various fields may also be possible, including stance classification of debate texts, extraction of grounds for certain topics, and so on.

Key words: Argument Mining, Argumentation Scheme, Argumentation Structure, Discourse Relation, Reasoning Strategy, Korean Text Processing

Student Number: 2013-20032

Contents

1	Introduction	1
1.1	Purposes	1
1.1.1	A Study of Korean Texts with Linguistic Cues	1
1.1.2	Detection of Argumentation Schemes in Debate Texts	2
1.1.3	Extraction of Important Content in Argumentation Schemes of Texts	2
1.2	Structure	3
2	Previous Work	5
2.1	Argumentation Mining Tasks	7
2.1.1	Argument Elements	7
2.1.2	Argumentation Schemes	9
2.2	Argumentation Schemes in Various Texts	14
2.2.1	Dialogic vs. Monologic Texts	14
2.2.2	Debate Texts vs. Other Texts	15
2.2.3	Studies in Other Languages	17
2.3	Theoretical Basis	18
2.3.1	Argumentation Theory	18
2.3.2	Discourse Theory	21
3	Identifying Argumentation Schemes in Debate Texts	25
3.1	Data Description	25
3.2	Basic Units	27

3.3	Discourse Relations	29
3.3.1	Strategies for Proving a Claim	29
3.3.2	Definition	35
4	Automatic Identification of Argumentation Schemes	41
4.1	Annotation	41
4.2	Baseline	46
4.3	Proposed Model	50
4.3.1	O vs. X Classification	51
4.3.2	Convergent Relation Rule	61
4.3.3	NN vs. NS vs. SN Classification	65
4.4	Evaluation	67
4.4.1	Measures	67
4.4.2	Results	68
4.5	Discussion	74
4.6	A Pilot Study on English Texts	81
5	Detecting Important Units	87
6	Conclusion	99

List of Figures

2.1	A Diagram of Toulmin's Argumentation Scheme	19
2.2	An Example of the RST Diagram: Evidence	22
3.1	An Example of the Strategies for Argument	34
3.2	A Two-Level Classification of Discourse Relations	36
3.3	Diagrams of Discourse Relations	40
4.1	An Example of Data Distribution and Hyperplanes	49
4.2	The Overall Flow of the Proposed Model	50
4.3	Two Unrelated Basic Units in the Argumentation Structure . .	63
5.1	An Example of the Tree-Structured Argumentation Scheme . .	88

List of Tables

2.1	A type of Walton’s Argumentation Scheme: Appeal to Expert Opinion ¹	20
2.2	A type of Mann and Thompson’s Rhetorical Structure: Evidence ²	22
3.1	Post Counts for each Topic	26
3.2	Unit Counts for each Post	29
4.1	Distribution of Discourse Relations O and X	44
4.2	Distribution of Discourse Relations NN, NS and SN	44
4.3	The Overall Performances of O vs. X Classification	69
4.4	Individual Feature Contributions of O vs. X Classification . . .	71
4.5	The Overall Performances of NN vs. NS vs. SN Classification .	72
4.6	Individual Feature Contributions of NN vs. NS vs. SN Classification	74
4.7	The Overall Performances of O vs. X Classification in the Pilot Experiment	83
4.8	The Overall Performances of NN vs. NS vs. SN Classification in the Pilot Experiment	84
5.1	The Depth 1 Units Detection Metrics	90

1 Introduction

The development of the Internet made the expression of personal opinions free. People write documents on-line to express their ideas to many, sometimes unspecified readers, and persuade them to accept the ideas. Accordingly, mass opinions are increasingly determined by analyzing on-line texts. For example, companies care about consumer responses to products, and politicians want to know how many voters support a certain candidate.

However, documents written on-line are so large in quantity, and varied in their topics and forms. So it is difficult to collect and understand the opinions of many people. With the amount of data increasing, automation is indispensable. That is, a procedure automatically extracting and collecting important information from texts is necessary. This study mainly aims to identify the structure of a subjective, persuasive text with an automated system.

1.1 Purposes

The goals of this study are as follows.

1.1.1 A Study of Korean Texts with Linguistic Cues

Studies on Korean texts, being in their infancy, are not that large in quantity. Korean text data exist sporadically, and just a few of them are systematically constructed and archived. This study collected debate posts from Korean on-line forums, which are appropriate to study and experiment on in the field of argument mining.

In addition, reinforcing a list of Korean shifters provided by Jang and Shin (2010), this study constructed a richer list of Korean explicit connectives indicating rhetorical relations between text segments.

1.1.2 Detection of Argumentation Schemes in Debate Texts

The frameworks to identify and constitute an argument scheme proposed by previous studies mostly do not make use of distinct properties of persuasive, subjective texts. Even if they consider the characteristic properties of debate texts, most of them are somewhat extralinguistic: information of the forum users, connections of rebuttal posts, and so on.

This study extracts and uses the linguistic cues which are characteristically used in texts for expression of personal opinions. Some typical cues like the length of sentences or the absolute position of a certain sentence in the document are excluded. Instead, this study attempted to construct a feature set reflecting strategies for persuasion which are usually used in argumentative writing, and to identify the rhetorical relations of discourse between text segments. These relations form an argument, first in a small scale, including conclusions, premises, or both. These smaller arguments are collected and combined to obtain the overall structure of the opinion suggestion corresponding to a text.

1.1.3 Extraction of Important Content in Argumentation

Schemes of Texts

The whole internal structure of documents is analyzed after the important parts in discourse relations are already identified through the first stage

of classification. The argumentation structure of texts which is in a reversed tree form is assumed to have relatively important nodes of text segments on the upper side of the structure.

This study tries to determine the sentences or clauses which directly support the topic sentence, as they are located directly below the topic sentence node in a tree structure. This is supposed to be helpful in catching some structured reasons which the author wants to suggest for a topic and his/her stance toward it. The information extracted here would be available for applications including stance classification, reason detection, text summarization, and so on.

1.2 Structure

This thesis is organized as follows: first, Chapter 2 introduces several previous studies which tried to identify argumentation structures of texts automatically. The studies are classified by several criteria: the type of tasks the studies pursue, the genre of the texts, and the language used in the texts. And this chapter also shows the theoretical bases of structured argument models and the concept of discourse relations which comprise them. Chapter 3 explains the basic unit of text segment assumed in this study and discourse relations between them to detect the argumentative structure of texts. Chapter 4 starts the actual procedure of extracting argumentation schemes from documents with an automated module. The three levels of experiment where discourse relations between basic text segments are automatically classified and

their results are stated here. Then a pilot study on English debate texts will be introduced additionally. In Chapter 5, the proposed model in this study constructs an overall argumentation structure of a text by combining the classified discourse relations, and extracts some important content from the structure. Finally, Chapter 6 concludes the thesis by summarizing the idea and experiments of this study, mentioning some issues to be considered in the future.

2 Previous Work

Argumentation Mining includes tasks about detecting the logical structures of texts, where authors propose a claim and provide additional statements supporting, supplementing and correcting the claim. Studies about argumentation deal with various subtasks: identifying argumentative units of sentences or clauses, relationships between the argumentative units, definite types of the relationships, etc. These works ultimately aim to find ways of comprising an argumentation structure corresponding to a document with the atomic relations between basic units of the text.

In the case of expressing a subjective claim to a given topic, the author would arrange proper grounds to support the claim at the appropriate position according to some logical, argumentative structure. As an example, there is a short debate text used in this study. It is a post answering the topic question, “Should a trashcan be installed on the street?” It insists that “A trashcan should be installed.”

- (1) 길거리에 휴지통이 있어야 합니다. 옛날과 달라서 길거리를 걸으면서 음료 등을 먹는 사람도 많고 24시 마트도 많아서 쓰레기가 많이 발생합니다. 휴지통이 없다 보면 아무데나 버리면 그만큼 길거리가 더러워지고 그것을 치우려면 청소하시는 분들이 힘들고 세금 낭비가 있습니다. 적극적으로 찬성합니다.

There should be a trashcan on the street. Different from older days, many people drink beverages while walking streets and there are lots of 24-hour markets, making tons of garbages. Having no trashcan, if people throw their garbages anywhere, the streets become dirty, the refuse

collectors have a hard time cleaning them, and a waste of tax money occurs. I actively agree.

The author proposes several reasons to make the claim more persuasive, saying (1) people drinking on the streets and 24-hour markets make garbage, (2) the street becomes dirty, (3) the refuse collectors have a hard time cleaning garbage, and (4) a waste of tax money occurs.

Many previous studies defined and tried to classify the topic sentence of these kind of texts as a *conclusion* or *claim*, and the reasons given above as a *premise* (Ghosh et al., 2014; Govier, 2009; Llewellyn, 2012; Schneider and Antipolis, 2014). Given the grounds supporting the claim as its sub-statements, more and more studies were done to find the supporting or attacking relationships between statements in texts, and to identify the argumentation structure of the texts where one or more premises provide support for a conclusion (Mukherjee and Bhattacharyya, 2012; Palau and Moens, 2009; Polanyi and Berg, 2003; Potter, 2008; Potter, 2007; Rosé et al., 2008; Schneider and Antipolis, 2014; Walker, Vazirova, and Sanford, 2014).

Argumentation mining tasks have been used in various types of applications in Natural Language Processing. First, some studies tried to summarize and extract the main, important contents of a document by analyzing an argumentation scheme used in the text (Kikuchi et al., 2014; Misra et al., 2015). Other studies detected the orientation of the polarity a text has about a given topic, based on the components and their combinations in the document (Faulkner, 2014; Ranade, Sangal, and Mamidi, 2013; Yessenalina, Yue, and Cardie, 2010). Some approaches visualized the argumentation structure of texts. In addition, the types of applications also differ according to the genre

of texts. In the case of student essays, an evaluating task where an essay is scored by various criteria including its structure, is dealt with importantly (Ong, Litman, and Brusilovsky, 2014; Song et al., 2014; Stab and Gurevych, 2014). With scientific or legal texts, the argumentation structure can be helpful to catch specific contents in the texts, since these types of texts are relatively standardized compared to texts in other genres (Graves et al., 2014; Green, 2014; Lawrence et al., 2014; Palau and Moens, 2009; Walker, Vazirova, and Sanford, 2014). Product reviews, microblogs, tweets, Wikipedia articles, etc. can also be treated as the text data for argumentation mining tasks (Alistair and Diana, 2005; Chenlo, Hogenboom, and Losada, 2013; Zhang et al., 2013).

2.1 Argumentation Mining Tasks

Traditionally, arguments consist of conclusions and their premises. That is, an argument is basically assumed to be a relation where one unit of description supports another. To study arguments, it is important to know which of the sentences or clauses is a conclusion or a premise in the relation, and how they are connected to be a complete structure of a document.

2.1.1 Argument Elements

Some studies identify the sentences or clauses interacting with each other in a text and classify them according to their functions. This is the stage where the separate text segments are identified by their roles in the whole

text expressing a subjective idea. The roles of the individual statements are defined through some criteria which are various, but similar to each other. Basically, the *conclusion* or *claim* indicates the main content an author wants to say in a document, while the reasons or grounds supporting the claim are called the *premise*. Most studies are based on these concepts, sometimes with slightly different terminology.

Govier (2009) described the concept of arguments and their practical examples and extracted conclusions and premises from texts. Observing that the extracted elements construct a structure in certain patterns, i.e. a deductive argument or an inductive argument, the text was summarized as a general model form. Schneider and Antipolis (2014) also identified claims and premises in online debate texts, and provided types of relations between the claims and premises: one supporting or attacking another. Somewhat differently, Stab and Gurevych (2014) proposed three types of arguments for each sentence or clause in a text, i.e. major claim which is the main topic sentence of the text, claim, and premise, then checked if one of them supports another in a relation. Additionally, Llewellyn (2012) automatically identified the structure of arguments, defining two levels of argumentation. A micro-level argumentation classified claims into subclasses by whether the claim has grounds or not. Whereas, a macro-level argumentation provided four basic types of argumentation components, i.e. argument or proposition, counter argument, integration or reply, and non-argumentative moves which include questions or coordinating moves, etc.

2.1.2 Argumentation Schemes

After defining and identifying the basic text segments in argumentation, many approaches assumed that the relationships between the text segments constitute certain structures and tried to analyze the structure of texts, i.e. the *argumentation scheme*, automatically. Those relations between the components in a document are called *discourse relations*. Discourse relations are sometimes expressed by text coherence or context, since certain components in texts show close relationships in their meanings.

The types of possible relations differ somewhat by study, even though, in many cases, they are similar to those of other studies. In particular, the criteria of grouping and classifying the relation types can vary by study. Marcu, Echihab, and Rey (2001) defined and recognized four types of discourse relations between arbitrary spans of text: contrast, explanation-evidence, condition, and elaboration. Wolf and Gibson (2005) provided coherence relations between discourse segments, including cause-effect, condition, elaboration, example, generalization, similarity, contrast, violated expectation, attribution, and same. Similarly, Saito, Yamamoto, and Sekine (2006) detected discourse relations between two successive sentences in Japanese text. The discourse relations have seven types of subclasses, including elaboration, contrast, cause-effect, equivalence, change-topic, example, and other. The study tried to identify these relations with some phrasal patterns. On the other hand, Asher, Benamara, and Mathieu (2008) defined detailed categories of opinions and rhetorical relations. The rhetorical relations appear when the opinion of a document is expressed, and are divided into five subclasses: contrast, correction, support, result, and continuation. Potter (2008) proposed some natu-

ral patterns of discourse which are the structural basis of explanatory, argumentative reasoning. Three categories of rhetorical relations, i.e. inferential, synthetic, and multinuclear, are introduced here, and the combinations of the tendencies of those relations are used to know whether two or more premises supporting a conclusion are linked to or convergent from each other. Further, Ghosh et al. (2014) identified the agreeing or disagreeing relations between pre-defined basic segments, i.e. an Argumentative Discourse Unit.

While some previous works only detected the atomic discourse relations between text segments, other studies implemented the next stage, too. After the stage of identifying discourse relations, the relations in the debate posts are collected together and finally constitute the post's overall structure, which applies argumentation schemes. Polanyi and Berg (2003) defined the main discourse rule of coordination and subordination, and mentioned additional types of question-answer pairs, greetings, and some preposed context specifiers. Palau and Moens (2009) set the elementary units of argumentation and classified the relations between the units in two stages. First, the relations were divided into those that are argumentative relations and those which are not argumentative, and then the argumentative relations were further divided into coordination, subordination, and multiple argumentation relations. Biran and Rambow (2011) identified the strategies which authors of written dialogs use to justify their claims. Using some indicators from Rhetorical Structure Theory (Mann and Thompson, 1988), when the first text segment in a segment pair is a claim, whether the second one is a justification of the first one or not is automatically detected. In addition, Peldszus and Stede (2013) provided various forms of argument structure which are obtained by synthesizing

premises and conclusions. The types of argument structure are first classified into support, attack, and counter-attack. Those structures are then divided again into subclasses based on varied strategies, including linked, serial, or separated relations between several supports, and attacks by rebuttal, undercutting, and supporting a rebutter. In the works above, the classified relations or structures are collected to construct a tree graph of a document’s argumentation scheme.

Finally, some works implemented automatic discourse parsers which perform the analysis of the argumentation schemes above. Feng and Hirst (2014) built an automatic parser whose discourse relations are based on Rhetorical Structure Theory. Lin, Ng, and Kan (2014) also implemented an automatic discourse parser, using the discourse information from Penn Discourse Tree-Bank (Prasad et al., 2007).

In summary, many previous studies build individual frames to identify basic discourse segments in texts and to structurize the argumentation schemes of the texts.

On the other hand, conclusions (claims) and premises are supposed to be somewhat relative concepts according to the levels of relations between them. For example, a sentence which was a conclusion of another sentence may become a premise of another conclusion sentence at a higher level. Being a conclusion or a premise corresponds to the relative concept in the discourse relation indicating which of the two text segments in the relation is a nucleus or a satellite in its meaning. Therefore, it is not possible to decide the role of a text segment only by determining whether the segment is argumentative or not. That’s because a claim does not always become the supported conclusion

of another text segment, even if it expresses a strong opinion about a certain topic. Also, even though a text segment is extremely descriptive, it may show an argumentative nuance in its meaning. It is still difficult to deal with these cases automatically using the binary criteria of argumentativity.

Instead, it is possible to judge whether a text segment supports or attacks another, and if so, which one is the conclusion in the relation. Thus, this study took the concept of conclusions and premises, but did not perform the process of detecting the span of the text segments themselves. The ultimate goal of this study is to obtain the overall argumentation structure corresponding to a document, so it is important to know if the text segments are related to each other, and if so, to determine which form of relation the related segments have.

This study does not follow the two-stage approach, where claims and premises are detected, then discourse relations between them are identified, along with the argumentation scheme of the text. Therefore, the pre-marked segment types, i.e. claims and premises, cannot be used as training features in the classification of discourse relations. Instead, the distinction between inter-related text segment pairs from those not related to each other is introduced. This procedure checks whether the two segments in the relation are relevant to each other in their meanings or not, that is, whether the segments are in a coherent relation or not. Then the specific types of discourse relations in coherence are defined.

The idea of the 2-level classification of discourse relations in this study is different from the standards used in other state-of-the-art studies. Previous works include the cases where a distinction between relations based on

coherence is not needed since only successive or adjacent text segments are considered. Also, some studies focus on the argumentativity of text segments before the full-out classification of discourse relations.

Given text segments related in the same context, this study defined subclasses of discourse relations according to whether the text segments are in a multinuclear (coordinating) or subordinating relation, and according to the location of the nucleus in the two elements in the relation, i.e. the latter segment supporting the former one, or vice versa. As Williams and Power (2008) proposed, rhetorical relations are assumed to appear different based on their position in the rhetorical structure tree, that is, a higher level or a lower level. These categories correspond to some groups which cluster detailed rhetorical relations or argumentation strategies. The labels for classes seem to be similar to previous works, but linguistic cues which differ according to the location of nuclei and the type of rhetorical relations are sensitively considered. Some studies including Somasundaran (2010) and Ranade, Sangal, and Mamidi (2013) use information about an author's intention in each sentence or clause. Authors state some positive or negative evaluations about certain stances on a topic, which finally becomes parts of the strategies for reasoning. For instance, a positive opinion about a stance and a negative opinion about the stance of the opposing side are grouped together into 'a positive evaluation about the current idea.'

Based on this idea, all pairs of text segments are supposed to have one of the combinations of 'conclusion-premise', 'premise-conclusion', 'conclusion-conclusion', and 'premise-premise.' Connecting all those assigned relations, a bottom-up method of structuration is used to constitute a document.

2.2 Argumentation Schemes in Various Texts

There are various types of texts to be studied in terms of argumentation mining. First, texts are divided into dialogic texts and monologic texts according to the number of authors involved in the text document. And by genre, texts are classified into debate texts, student essays, reviews, scientific texts, and legal texts, etc. Finally, even though most studies deal with English texts, many researchers are studying data written in other languages.

2.2.1 Dialogic vs. Monologic Texts

Debate texts show one or more people's opinions about certain topics. Most of these texts are comprised of online debating between internet users. People express their ideas and read others' opinions, sometimes agreeing with them and sometimes arguing against them. Many websites provide online forums, where their users freely comment and support opinions about various topics.

Many of these debates consist of opinion posts and continuous replies or comments on them. Two or more participants give feedback on each other's ideas, in the form of a continuous chain of comments (Boltuzic and Snajder, 2014; Ghosh et al., 2014; Misra et al., 2015). As these texts are written and collected on-line, it is much less burdensome for people to participate in the debates compared to writing a formal persuasive essay or an editorial. Posts tend to be free-style and include many colloquial expressions. This kind of data, however, includes posts which are too short to look into their internal structure. Continuous comments to each other are just like dialogues through

those texts, so a post in a long chain might be too dependent on its context. The fact it is not an independent post is not helpful in identification of the post's argumentation scheme.

Instead, a monologic form of debate is used in this study. Some debate texts are mainly written in separate posts, with few replies and comments. People individually write a new, separate post about a given topic. In this case, each post becomes a persuasive essay, without conversations between users, which are direct feedback for previous posts. This kind of debate post tend to be longer and be more organized (Peldszus, 2014; Stab and Gurevych, 2014; Williams and Power, 2008).

2.2.2 Debate Texts vs. Other Texts

Debate texts have much more varied targets of description and opinion than texts from other genres or previous works. Many studies are done on product or movie reviews (Alistair and Diana, 2005; Zhang et al., 2013). In review texts, the authors use a relatively limited set of grounds for their conclusions. For example, if an author wants to claim that a specific model of camera is good, he or she would have to talk about the price, the image quality, the lens, etc. of the camera. Other reviewers of the same product would also think about similar aspects.

Many studies have been done on student essays. Essays are texts which present a subjective idea about a certain topic in a relatively standardized form. Essays usually consist of several paragraphs including an introduction, body, and conclusion. Additionally, essays commonly observe a relatively limited form of structure constructed by the certain types of paragraphs. There-

fore, it is easier to catch a positional feature for a text segment of claim and premise, or for discourse relations. Using ideas from studies of argumentation schemes, some studies identify the structure of argumentation in essays written by students, and sometimes they contain a procedure for evaluating them (Faulkner, 2014; Ong, Litman, and Brusilovsky, 2014; Song et al., 2014; Stab and Gurevych, 2014).

Texts from legal or scientific fields are in a more standardized structure. Subsections in the texts are decided in advance, and their forms are mostly limited. For example, scientific writing would require dispensable components in its logical structure, e.g. hypotheses and evidence. Therefore, to some degree, those texts can be analyzed using simple phrasal patterns. Lawrence et al. (2014) dealt with the structure of philosophical texts from the 19th century, while Palau and Moens (2009) consider legal texts. Walker, Vazirova, and Sanford (2014) defined four types of logical connectives used in evidential writing, i.e. medical and legal texts, and provided argument types using these connectives. Studies detecting argumentation schemes to express a subjective idea effectively in biomedical texts have been done continuously (Faiz and Mercer, 2014; Graves et al., 2014; Green, 2014, 2015; Houngbo and Mercer, 2014). Additionally, Several studies dealing with news articles also consider the characteristic structures of texts (Kiesel et al., 2015; Sardianos et al., 2015).

Debate texts have distinctive properties that are different from those texts. First, debate texts are written in relatively free forms compared to essays, or legal or scientific texts. Since most debate texts studied in terms of argumentation mining are user-generated free writing in online debates, they

are not often written in several tightly connected paragraphs, instead showing only one paragraph for the whole text. Also, although many subjective texts tend to have their topic sentences located in the beginning or the end of the document showing inductive or deductive argumentation, it is not always the case. Moreover, the topic sentences do not explicitly appear in many cases, making the positional features for argumentation schemes less effective.

Second, debate texts have an infinite range of attributes about the topic. What would one say if he or she is asked whether God exists? The attributes used as reasons are not easily summarized as simple words and classified to separate fixed groups, compared to those of product reviews. So a simple matching of words or grammatical patterns would not be that helpful in detecting the attributes for debate topics. Thus, organizing patterns of argumentation in debate texts are expected to be more influential.

Many previous works in argumentation mining have studied debate texts, including Boltuzic and Snajder (2014); Ghosh et al. (2014); Lin and Hauptmann (2006); Yessenalina, Yue, and Cardie (2010), etc.

2.2.3 Studies in Other Languages

Studies on argumentation mining are also done about many languages other than English. Argumentation units, discourse relations, argumentation structures, etc. have been identified and predicted by many studies, some of those considering the language-specific discourse markers or connectives. Those previous works treated German texts (Adam and Dalmas, 2012; Peldszus, 2014; Trevisan et al., 2014), Spanish texts (Taboada and Angeles, 2012), French texts (Adam and Dalmas, 2012; Braud and Denis, 2014), Japanese

texts (Saito, Yamamoto, and Sekine, 2006), and Chinese texts (Huang et al., 2014), etc.

However, the argumentation mining studies on languages other than English are not that common at the present time. In particular, studies on Korean texts are nearly in their infancy. There are some previous works which deal with Korean texts including Jang and Shin (2010) where Korean discourse markers are arranged and categorized for sentiment analysis in Korean. But studies about the argumentation schemes of debate texts in Korean have almost never been done yet. Therefore, this study collected and reconstituted the list of discourse markers in Korean by referring to several separate sources. Additionally, numerous properties observed in Korean documents and sentences are reflected in this study of argumentative structure, which finally comprises a model for the overall scheme.

2.3 Theoretical Basis

This study proposes an argumentation scheme and a module to automatically detect it from texts. The procedures refer to some theoretical bases regarding an argument and its components.

2.3.1 Argumentation Theory

Earlier studies observed the nature of arguments, defining the types of elements which comprise the structure of argument. Toulmin (1958) defined six types of components that are involved in the process of making a claim.

According to the study, an argumentation scheme includes the *claim*, which corresponds to the topic sentence of a text, the *data*, which is the basic grounds of the claim, and *backing*, *warrant*, *qualifier* and *rebuttal* to connect the data to the claim as the supporting components.

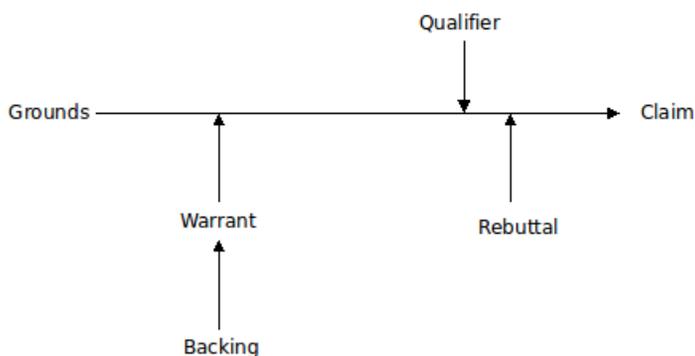


Figure 2.1: A Diagram of Toulmin's Argumentation Scheme

The warrant is logic that relates the claim and the data, while the backing is an additional explanation supporting the warrant. The qualifier limits the degree of force the claimed argument shows by setting conditions on it. Finally, the rebuttal indicates a counter-statement against the claim. These components are connected and combined to construct the logical structure of argumentation by supporting the proposed idea and controlling its strength.

Douglas, Christopher, and Fabrizio (2008) provided argumentation schemes which differ by the author's intention or purpose expressed in texts. Various forms of argumentation schemes are made up of a conclusion and several sub-categorized premises combined in diverse ways. In some cases, an author proposes a conclusion and justifies it by citing an expert opinion about the topic as grounds.

Appeal to Expert Opinion (Version II)

Major Premise	Source E is an expert in subject domain S containing proposition A.
Minor Premise	Asserts that proposition A (in domain S) is true (false).
Conditional Premise	If source E is an expert in a subject domain S containing proposition A, and E asserts that proposition A is true (false), then A may plausibly be taken to be true (false).
Conclusion	A may plausibly be taken to be true (false).

Table 2.1: A type of Walton’s Argumentation Scheme: Appeal to Expert Opinion¹

In the example frame (Table 2.1), the minor premise functions as direct grounds for the conclusion, while two other types of premises act as the logical basis for the proposition of the minor premise to support its conclusion.

Additionally, many logical strategies for arguments are defined and patternized in Toulmin’s work. The argumentative strategies include witness testimony, popular opinion, popular practice, example, analogy, alternatives, cause to effect, evidence to a hypothesis, consequences, fear appeal, need for help, etc.

It is not easy to automatically implement those argumentative structures with high accuracy. Dealing only with explicitly expressed texts without any markers of the authors’ internal ideas, the practical approaches mostly find only two of the argumentation components described above. The conclusion and its grounds are considered in priority because almost always, only these

¹Walton, D. and C. Reed (2002). “Argumentation schemes and defeasible inferences”. In: Working Notes of the ECAI’2002 Workshop on Computational Models of Natural Argument, pp. 45–55.

two components are apparent in texts. Sometimes the links between the conclusion and the grounds, i.e. the warrant or backing, are written explicitly, sustaining the argument.

This illustrates some bases to justify the strategies for proving one's idea, which the present study provides. As Douglas, Christopher, and Fabrizio (2008) described in detail, there are various ways of connecting and relating grounds to a claim. Even if those methods are not explicitly shown in texts, they are supposed to be reasonable cues for identifying and analyzing the relationships between text segments for argumentation. The relationships contributing to arguments are described in refined frameworks by discourse theories in Section 2.3.2.

2.3.2 Discourse Theory

Mann and Thompson (1988) proposed subdivided classes organizing an argumentation scheme. Several classes of rhetorical relations between two adjacent text segments are defined and displayed based on functional and semantic judgements: Circumstance, Solutionhood, Elaboration, Background, Enablement, Motivation, Evidence, Justify, Volitional Cause, Non-Volitional Cause, Volitional Result, Non-Volitional Result, Purpose, Antithesis, Concession, Condition, Otherwise, Interpretation, Evaluation, Restatement, Summary, Sequence, and Contrast. The study states the constraints on each text segment in the relation, the effect of the relation expressed, the locus of the effect, and a few more details for each rhetorical relation. For example, the statement about a rhetorical relation *EVIDENCE* reads as below, accompanied with a visualized diagram:

Evidence

relation name	EVIDENCE
constraints on N	R might not believe N to a degree to W
constraints on S	The reader believes S or will find it credible
constraints on N + S combination	R's comprehending S increases R's belief of N
the effect	R's belief of N is increased
locus of the effect	N

Table 2.2: A type of Mann and Thompson's Rhetorical Structure: Evidence²

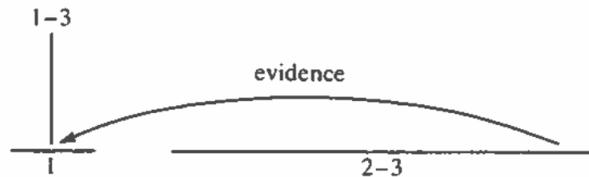


Figure 2.2: An Example of the RST Diagram: Evidence

The study also shows an example combination of text segments in the relationship of evidence.

(2) An Example of a Rhetorical Relation: Evidence

- (a) The program as published for calendar year 1980 really works.

²Mann, 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.

- (b) In only a few minutes, I entered all the figures from my 1980 tax return and got a result which agreed with my hand calculations to the penny.

The concept of a nucleus and a satellite is introduced and used here. Given a text segment pair like (2), the distinction between the nucleus and the satellite is obvious. Since the supportive evidence (2b) assists (2a) which is a major statement in the discourse, the two segments are mapped to the satellite and the nucleus, respectively.

According to the study, certain rhetorical relations exist between text components, reflecting the distinction between supporting and attacking relations, and between main and auxiliary contents in texts. These relations in discourse are composed to finally represent a tree-like structure corresponding to a text document.

A large number of the studies which followed referred to the Rhetorical Structure Theory working on discourses or argumentation schemes. They regarded the rhetorical relations as fundamental concepts and utilized them. The present study also accepts the relations and introduces argument strategies based on them. The detailed rhetorical relations described in previous work are used as cues for individual strategies and features of the proposed scheme. Though, in this study, they are clustered into somewhat rougher classes. The classes significantly reflect the concepts of nucleus and satellite illustrated in the Rhetorical Structure Theory. In this study, the framework is helpful to identify and finally extract text segments which are important in terms of argument, applying the rhetorical relations and nuclearity in the meaning of the text.

3 Identifying Argumentation Schemes in Debate Texts

3.1 Data Description

439 monologic posts were crawled from an online forum.¹ On this website, people individually express their personal ideas about a given topic. There are hundreds of issues to be discussed, and two of them are selected for this study. Both topics are about new policies in Seoul.

1. 길거리에 쓰레기통이 있어야 할까, 없어야 할까? ²

Should trashcans be placed on streets?

2. 한강둔치서도 ‘삼겹살 파티’ 벌일 수 있다? ³

Will barbecue be allowed at the riverside of the Han River?

Not all the posts under the selected topics are included in the data. Posts which are not appropriate to use are ruled out: too long, too short, not related to its given topic, etc.

Users of this website must write a title for their posts, and then the contents of post itself. Most of the titles are used as the topic sentence of their idea. In this case, the text starts with a topic sentence or a conclusion followed

¹<http://bbs1.agora.media.daum.net/gaia/do/agora/issue/list?bbsId=I001>

²<http://bbs1.agora.media.daum.net/gaia/do/agora/participant/list?bbsId=C001&issueArticleId=344&issueBbsId=I001>

³<http://bbs1.agora.media.daum.net/gaia/do/agora/participant/list?bbsId=C001&issueArticleId=343&issueBbsId=I001>

Topic	Pros	Cons	Total
a trashcan on streets	190	70	260
barbecue allowed at riverside	42	137	179

Table 3.1: Post Counts for each Topic

by its premises or grounds, giving the posts a deductive structure. 289 posts in the data are written with the title or the first sentence as the topic sentence, which comprise 65.8% of all data.

(3) A Debating Post Example

Title 쓰레기통이 있어야 합니다.

Body

길 가다가 편의점이나 슈퍼 같은 데서 아이스크림이나 음료 과자 같은 거 사 먹고 나면 쓰레기가 발생하는데 쓰레기통이 없으니 많이 불편하죠.

예전에 그 많던 쓰레기통 그 쓰레기통 주변이 더럽다고 치워 버렸는데 쓰레기통을 즉각 즉각 비우면 더럽혀질 일이 별로 없을 겁니다.

도심의 길거리 쓰레기통 설치하고 관리하는 미화원 배치하면 일자리도 생기고 거리도 깨끗해지고 좋을 것 같네요.

Title There should be trashcans.

Body

Trash is created when people have ice cream, beverages, or snacks

at convenience stores or supermarkets on streets, and it is inconvenient if there's no trashcans.

While the trashcans which formerly were large in number were removed because surrounding area was dirty, this would not be the case if trashcans are emptied often.

If sanitation workers are allocated to install trashcans on downtown streets and care for them, it would be good with more jobs and cleaner streets.

Some posts of the data have other users' comments or new posts replying themselves. In this study, the comments are excluded because they tend to be too short. Additionally, comments usually are simple reactions about some given texts. On the other hand, the replying posts are regarded as separate debate posts.

After the crawling procedure, the posts were edited a little to minimize errors in their form. Typos, word spacing errors, and emoticons were corrected or removed to make the automatic procedures, which the data will undergo, easier.

3.2 Basic Units

This study defines a *basic unit* in argumentation structure as the minimum unit which expresses one idea consistently. A basic unit is basically a sentence, but sometimes the sentence is split into several clauses to become

separate basic units. That is, a basic unit may contain a full sentence, but in some cases it is one of the clauses in a sentence. An example list of defined basic units are shown in (4).

(4) A List of Basic Units in an Example Post

- 0 쓰레기통이 있어야 합니다.
 - 1 길 가다가 편의점이나 슈퍼 같은 데서 아이스크림이나 음료 과자 같은 거 사 먹고 나면 쓰레기가 발생하는데
 - 2 쓰레기통이 없으니 많이 불편하죠.
 - 3 예전에 그 많던 쓰레기통 그 쓰레기통 주변이 더럽다고 치워 버렸는데
 - 4 쓰레기통을 즉각 즉각 비우면 더럽혀질 일이 별로 없을 겁니다.
 - 5 도심의 길거리 쓰레기통 설치하고 관리하는 미화원 배치하면 일자리도 생기고 거리도 깨끗해지고 좋을 것 같네요.
-
- 0 There should be trashcans.
 - 1 Trash is created when people have ice cream, beverages, or snacks at convenience stores or supermarkets on streets,
 - 2 and it is inconvenient if there's no trashcans.
 - 3 While the trashcans which formerly were large in number were removed because surrounding area was dirty,
 - 4 this would not be the case if trashcans are emptied often.

- 5 If sanitation workers are allocated to install trashcans on downtown streets and care for them, it would be good with more jobs and cleaner streets.

Basic units in posts are split and assigned manually, since an automatic identification of those spans is a difficult, sizable task in itself.

Topic	Pros	Cons
a trashcan on streets	7.8684	7.9714
barbecue allowed at riverside	8.0238	7.8394

Table 3.2: Unit Counts for each Post

As Table 3.2 shows, the length of a post does not seem to be distinctive in different topics or stances. This means that the length of posts would not function as a variable in identifying posts' argumentative structure.

3.3 Discourse Relations

3.3.1 Strategies for Proving a Claim

This study defines *discourse relations* which combine together and organize an argumentation scheme of a document, based on several methods of proving a claim in debate texts. In terms of organization, debate texts are expected to have some specific types of strategy to prove themselves, since they are written to convince readers. The strategies are realized as relations between two or more basic units. For instance, if a basic unit gives some de-

tails about the idea of another one, that relationship is named as one type of strategy, e.g. ‘Giving Details.’ Such strategies are observed in various debate texts, and this study arranged and named them as the specific types according to their functions: Giving Reasons, Background Knowledge, Providing Examples, Cause and Effect, Giving Details, Giving Opinions or Reactions, Paraphrasing with the Same Meaning, Enumeration, Simple Serialization, Concession, Contradiction, Switching, and more.

(5) Giving Reasons

- 5 분리수거를 철저히 하면 일단 수거비용은 뺐을 수 있다고 생각합니다.
- 6 재활용이 가능하기 때문예요.
- 5 It seems to be able to get the pick-up cost for now if garbage collections are thoroughly separated.
- 6 It's because recycling is available.

(6) Background Knowledge

- 1 요즘 취업이 안 돼서 난리인대요.
- 2 쓰레기통 관리인 지자체 별로 적당선 고용해서 운영하면 실업을 해소도 되고 시민들은 편해서 좋고 일거양득이네요.
- 1 It's hard to get a job these days.
- 2 Hiring a trashcan keeper appropriately at each local government would lower the unemployment rate and it'll be convenient for citizens.

(7) Providing Examples

- 1 쓰레기는 쓰레기가 생긴 곳에서 처리하면 될 것입니다.
- 2 예를 들어 슈퍼 등에서 물건을 산 경우라면 그 곳에 버리면 될 것이라 생각합니다.
- 1 Trash is to be cleaned up at the place where it was created.
- 2 For example, if you bought something at a supermarket, you may throw it away there.

(8) Cause and Effect

- 2 잔디에 떨어진 기름.
- 3 그럼 잔디는 다 죽겠죠.
- 2 Oil dropped on the grass.
- 3 Then the grass would die.

(9) Giving Details

- 6 쓰레기들 천지입니다.
- 7 먹다 만 음료수, 캔맥주, 치킨 쪼가리, 사발면, 기타 등등.
- 6 It's fraught with trash.
- 7 Half-consumed beverages and canned beer, pieces of chicken, instant cup ramen, etc.

(10) Giving Opinions or Reactions

- 2 우리나라에서도 쓰레기통 폭탄 테러를 방지하기 위해 길거리에서 쓰레기통을 죄다 없앴거든?
- 3 근데 내가 보기에 이걸 좀 오바였지 싶어.
- 2 This country also removed the trashcans on the street to prevent bomb terrorism in trashcans.
- 3 But I think it was a little excessive.

(11) Paraphrasing with the Same Meaning

- 0 쓰레기통 있어야 합니다.
- 1 쓰레기통 당연히 있어야지요.
- 0 A trashcan should be located.
- 1 Of course a trashcan should be located.

(12) Enumeration

- 7 둘째, 공공기관이나 각 기업체는 화장실에 비데를 설치하여 휴지 사용을 억제하여야 합니다.
- 8 셋째, 학교에 50% 이상 남자 선생님이 있어야 합니다.
- 7 Second, public institutions or business entities should limit use of wastepaper by installing bidets in the restrooms
- 8 Third, male teachers should comprise more than 50% of school faculties.

(13) Simple Serialization

4 갈 때마다 보는 것은 한강 둔치에 버려져 있는 캔맥주 깡통과 피쳐 병들.

5 그리고 담배 꽁초.

6 배달 시켜 먹고 버려진 닭뼈.

4 Every time I go there, I see cans and bottles of beer which are thrown away at the riverside of the Han River.

5 And a cigarette butt.

6 Abandoned chicken bones from delivery.

(14) Concession

2 길거리뿐만 아니라 공원 주차장 등 곳곳에 마련되어야 합니다.

3 물론 그로 인한 비용도 만만치 않겠지만요.

2 It should be created everywhere, including the parking lots at parks not only on the streets.

3 Of course, it would require heavy charges.

(15) Contradiction

1 예전엔 길거리에 휴지통은 당연히 있어야 한다고 생각했을 겁니다.

2 하지만 요즘은 그 생각에 의문이 많이 듭니다.

1 Formerly, I would have thought that there should be trashcans on streets.

2 However, I often doubt the idea these days.

(16) Switching

- 4 쓰레기통 길가에 반드시 설치해야 합니다.
- 5 근데 저는 서울시에서 쓰레기통 디자인을 잘못 했다고 생각합니다.
- 4 Trashcans should be installed on streets.
- 5 By the way, I think that Seoul made a mistake designing the trashcans.

These strategies combine together to constitute argumentation schemes used in a debate post. In other words, a topic unit has some reasons, examples, countermeasures, or something else related to it to convince its readers to believe and accept its idea. The provided grounds may have additional units to support itself, which are also in the form of the strategies above. Figure 3.1 illustrates an example of the hierarchical structure of supporting and being supported as argumentation schemes.

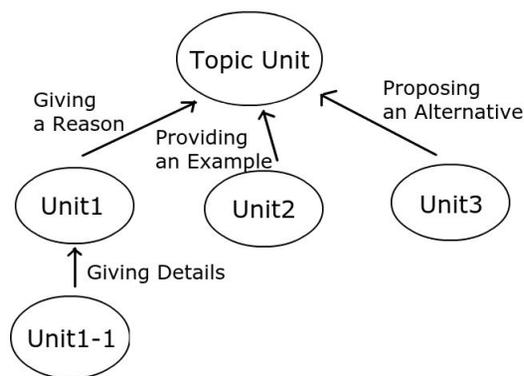


Figure 3.1: An Example of the Strategies for Argument

The strategies to prove one's idea are classified by the form of the relationships between the two basic units: parallel, vertical, or others.

Vertical Relation One of the two basic units taking part in the relation is more important than another.

Cases: Giving Reasons, Examples, Cause and Effect, Alternative and Countermeasure, Details, Opinions, Reactions, Background, etc.

Parallel Relation Both of the two basic units in the relation are equal in their importance.

Cases: Repeating the same phrases, Paraphrasing, Enumeration, Simple Serialization, etc.

Otherwise No relation is defined between two basic units.

Cases: Irrelevant, Unfocused statements, Switching topics, etc.

The groups classified here become the categories of the actual discourse relations defined in this study. A series of procedures for identifying and organizing the strategies of making a claim would be helpful to understand the argumentation schemes of given texts, especially with approaches which are specialized to deal with arguments.

3.3.2 Definition

All possible combinations of basic units constitute a list of pairs in a post. This study defines five types of *discourse relations* between the two basic units in the basic unit pairs. The most basic text segments assemble to

gether and are related to each other by several patterns, to comprise argumentation schemes provided in this study.

Discourse relations are classified through two separate levels. First, as depicted in Figure 3.2, discourse relations are classified as O and X, which means that the basic units are relevant or irrelevant to each other, respectively. Then only the discourse relations marked as O are classified again into NN, NS, and SN. NN(Nucleus-Nucleus) is defined as an equivalent relation where both basic units are Nuclues, NS(Nucleus-Satellite) indicates a relation where the second unit in the pair supports the first one, and SN(Satellite-Nucleus) means a reversed relation where the first unit in the pair is subordinate the second one.

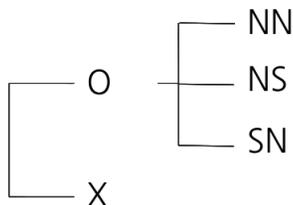


Figure 3.2: A Two-Level Classification of Discourse Relations

This two-level partition is needed because the debate texts are crawled from on-line sources and are relatively informal. The free-style writing is less likely to have concrete structure. That is, not all the basic unit pairs in the text show specific discourse relations. For example, statements which are irrelevant to the topic should have no relation with other relevant units. Moreover, if the two basic units are included in different contexts respectively, any contextual

relation would not be assigned between them.

Therefore, before detecting specific discourse relations between the two basic units in the pair, it is important to make sure that they are directly related in terms of the context. If not, applying classification standards to unrelated basic units would cause lots of errors in the classification procedure.

O The two basic units are contextually related.

In this case, the two units in the basic unit pair should be related in terms of the content that they claim. This class includes both vertical and parallel strategy relations described in Section 3.3.1. The two basic units in the unit pair may represent the same form, be paraphrased, and be in a vertical relation where one basic unit gives additional statements for another.

Relations in this class are further divided into three subclasses: NN, NS, SN. These classes are defined according to their detailed properties, including whether a basic unit is a Nucleus or a Satellite.

NN The two basic units are equivalent.

In this discourse relation, both of the two basic units are nuclei. NN is an acronym for ‘Nucleus-Nucleus Relation.’ The basic units express just the same meaning, and are on an equal level of relation. No one basic unit dominates the other one. This category includes strategies of restatement, paraphrasing, and simple serialization of statements, etc.

(17) Given a basic unit pair (UNIT1, UNIT2),

UNIT1 길거리 휴지통 있어야 합니다.

UNIT2 길거리에 휴지통이 있어야 합니다.

UNIT1 There should be trashcans on the streets.

UNIT2 There should be trashcans on the streets.

NS The former basic unit is supported by the latter one.

The relation marked as NS, short for ‘Nucleus-Satellite Relation’, means that the former element of the basic unit pair is a nucleus, and the latter element is a satellite. As a vertical relationship, the former basic unit dominates the latter one. The satellite basic unit provides grounds for the idea of the former nucleus basic unit, which includes explanation of details, reasons, examples, etc.

(18) Given a basic unit pair (UNIT1, UNIT2),

UNIT1 한류 열풍 어찌고 하는 게 낫 뜨거울 정도로 명동이나
종로 거리는 온갖 쓰레기의 전시장입니다.

UNIT2 이런저런 이유를 대 가면서 미화요원들 잘라낸 때문이
지요.

UNIT1 Myeong-dong or Jongno streets are full of trash, making
me ashamed about the Korean Wave.

UNIT2 That’s because they fired the street cleaners for one rea-
son or another.

SN The latter basic unit is supported by the former one.

As an acronym of ‘Satellite-Nucleus Relation,’ the former element of the basic unit pair is a satellite, and the latter one is a nucleus. That is, SN is a vertical relation just like NS, with the former satellite basic unit

being subordinate to the latter nucleus one. The satellite gives reasons, examples, background knowledge, etc. to support what the nucleus basic unit says.

(19) Given a basic unit pair (UNIT1, UNIT2),

UNIT1 길 가다가 편의점이나 슈퍼 같은 데서 아이스크림이나
음료 과자 같은 거 사 먹고 나면 쓰레기가 발생하는데

UNIT2 쓰레기통이 없으니 많이 불편하죠.

UNIT1 Trash is created when people have ice cream, beverages,
or snacks at convenience stores or supermarkets on streets,

UNIT2 and it is inconvenient if there's no trashcans.

X The two basic units are contextually unrelated.

In this case, the two basic units in the pair have little relation in terms of the context. Even though both of them appear in the same post and can be regarded as indirectly related to each other, they may not show any direct and explicit relation in terms of the context they provide. Thus, the basic unit pair is classified as X.

Also, switching of contexts is regarded as having no relationship, where a basic unit provides a different subtopic from the previous ones.

(20) Given a basic unit pair (UNIT1, UNIT2),

UNIT1 설치하는 게 좋다고 생각함.

UNIT2 60대 중반입니다.

UNIT1 I think it's good to install it.

UNIT2 I'm in my mid-60s.

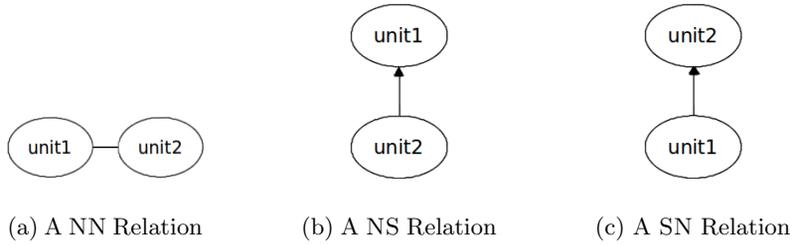


Figure 3.3: Diagrams of Discourse Relations

Figure 3.3 shows a graphic representation of the discourse relations defined above. Here, each oval indicates a basic unit. Arrows describe a vertical relationship between two basic unit, where the arrowhead goes from the subordinate satellite basic unit to the nucleus basic unit (NS, SN, illustrated in Figures 3.3b and 3.3c, respectively). The direction of the arrow shows the direction of support. And the straight line without an arrowhead indicates a parallel, equivalent connection between two basic units (NN, as Figure 3.3a shows). On the other hand, if there is nothing marked between two ovals in the tree structure of a document, it means that the two basic units are not directly related to each other.

After all the pairs of basic units are assigned their discourse relations above, an argument structure corresponding to a debate post is constructed. Through a simple algorithm, the discourse relations between the basic units gather together and synchronize themselves, and finally form the overall structure of a post. This procedure will be reported in detail later in Chapter 5.

4 Automatic Identification of Argumentation

Schemes

4.1 Annotation

All posts in the data are manually annotated. Given the posts whose basic units are already split and defined, annotators marked information about each post as below:

- **Topic Unit**

First, an annotator is required to read the post and pick its *topic unit*. The topic unit is a unit which directly expresses agreement or disagreement with a given issue, or which represents the most important claim in the post. A post can have one or more topic units, or no explicit topic unit.

Only one topic unit in a post Since many posts have the topic unit as its title, they are regarded as having a deductive structure. This case also includes posts with different structures, i.e. an inductive structure.

Two or more topic units in a post A post may have topic units both at the beginning and the end. And another may mention its issue repeatedly anywhere in the post. In those cases, all the topic units indicate the same meaning, so there would be no problem regarding the topic units as the same units. The topic units would

be connected with the discourse relation NN defined in Section 3.3.2.

No topic unit in a post This is the case where a post obviously is supporting or attacking a stance, but does not have an explicit topic unit form. Without any direct statements about the issue, reasons or grounds for that issue are provided in this type of post. In this case, an alternative marker is used in order to form a temporary topic unit. Instead of not having any topic units, the post is regarded as having a zero-form topic unit. Then the basic units with main grounds for the topic of the post will be attached right below the temporary topic unit, in case of a tree-form structure.

- **Discourse Relations**

After marking the topic units of the posts, discourse relations for pairs of the basic units are annotated. All the possible pairs of the basic units are marked with one of the four possible options of discourse relations: X, NN, NS, and SN. Annotators are required to judge what kind of relation the two basic units have, based on the content and meaning of the basic units.

The discourse relation O, which means that the two basic units in an unit pair are contextually related, is not used in the annotation procedure. Annotators just use the subclasses NN, NS, and SN. These subclasses will be grouped into O afterward, per the O vs. X classification procedure in Section 4.3.1.

- **A “C” mark**

A “C” mark represents the initial consonant of concession and contrast. When a basic unit pair has a vertical discourse relation, i.e. NS and SN, the two basic units may have different stances regarding a given issue, in relation to each other. Posts only in this case are marked with “C”, following the annotation of discourse relations.

This case includes a concession, where a basic unit accepts the idea of another one, while still supporting its own stance. A contrast also is relevant to this case, where a basic unit denies the stance of another one by giving contrary opinions.

(21) Concession

- 12 물론 기초질서의식이 아예 없는 경우도 많지만
- 13 하여튼 쓰레기통은 사람들이 필요한 곳에 적절히 두고
- 12 Of course, even if some people lack a basic sense of order,
- 13 trashcans ought to be located at places needed anyway.

(22) Contrast

- 10 시민이 덜 불편하다.
- 11 그러나 쓰레기를 길거리 휴지통에 버릴 수 있으니 길거리 휴지통에 잡다한 쓰레기가 넘칠 것이다.
- 10 Citizens would feel less inconvenience.
- 11 But they may throw garbage in the street trashcans, thus making street trashcans full of random garbage.

Especially in case of annotating discourse relations between basic units, it was recommended to the annotators to draw the argumentation scheme of the document as a tree structure. Scanning the overall structure in a visualized tree graph is helpful to identify the type of relations between the nodes in the tree.

Finally, 21,083 basic unit pairs are annotated and assigned to one of the four pre-defined types of discourse relations, which were formed from 3,468 basic units in 439 documents.

The classes of annotated discourse relations show the distribution, like tables 4.1 and 4.2 illustrate.

	O	X
Number of Relations	3651	17432

Table 4.1: Distribution of Discourse Relations O and X

	NN	NS	SN
Number of Relations	842	2047	762

Table 4.2: Distribution of Discourse Relations NN, NS and SN

However, it can be dangerous to construct a theory or a scheme based on just one set of annotations, since the data used in the analysis may show a random distribution. Thus, a procedure devised to compare data labels from two or more annotators to confirm whether the annotation is reliable or not. If several annotators selected the same tag for the same data segment, the

annotation is assumed to be reliable, and used to develop a theory based on it.

In this study, two annotators, including the author of this work, performed the annotation task, both being experts in linguistics. The experiments which are described below took the author's annotations as the gold standard. Another annotation was used to calculate the degree of agreement. Cohen's Kappa (Cohen, 1960) is used as a method of measurement in this case.

Cohen's Kappa is a measure which compares the labels assigned by two annotators, which takes into consideration the probability of a coincidental agreement if the two labels match.

$$\hat{\kappa} = \frac{p_o - p_e}{1 - p_e} \quad (4.1)$$

The equation (4.1) includes p_o which indicates an observed proportional agreement, and p_e , i.e. an expected agreement by chance. p_e shows the probability where the two annotators are independent, which infers the coincidental agreement of two annotators.

In the formula of Kappa, if the κ value grows close to 1, the two annotators are expected to show a very good agreement, while a lower κ value close to 0 means a poor agreement between the two annotators.

In this study, the agreement of the two annotators shows the κ value of 0.428, which means an agreement in the level of *fair to good*. This level of agreement is not that high, since the annotators tended to select different topic units from each other in the text. Thus all the discourse relations

attached to the topic units show disagreement between the two annotators. The κ value indicating the agreement of topic unit assignments between the annotators shows a value of 0.285, which is somewhat poor. This is because the number of basic units assigned to be a topic unit is not limited. It can be more than two, not just one, allowing the range of the available basic units for topic units to vary.

Therefore, the κ value was calculated again for the discourse relations assigned only to the basic unit pairs except the pairs where the annotated topic units are involved. It showed 0.478, a fair increase.

4.2 Baseline

The detailed tasks of this study are a little different from those of previous works. Also, this study proposes a new type of argumenation structure dealing with monologic debate texts and, thus, a new annotation scheme. So it is impossible to directly compare the performance of the proposed model to that of any existing models. Additionally, the debate texts used in this study are written in Korean, while studies dealing with Korean texts are nearly unprecedented.

More indirect and fundamental standards of comparison are needed. Therefore, a baseline study is provided in this study, which is the standard for comparison to know how much the performance of the task improved due to the proposed model.

Using the simplest feature separately can be a baseline method. For the

only feature for a baseline, this study uses unigrams referring to many previous works. This feature marks whether a specific word appears in the text or not, and if so, the frequency of the word is also included. After converting each basic unit into a vector of content words, a united list of content words of the two basic units in the basic unit pair is constructed. That is, a union of the two vectors is formed here. All the words in this list are checked for their existences and frequencies. Three kinds of counts are marked for this feature: (1) How many times does the word appear in the first basic unit? (2) How many times does the word appear in the second basic unit? (3) How many times does the word appear in both basic units?

(23) Unit Pairs

- 7 차라리 쓰레기통을 만들고 쓰레기통에 안 버리고 주위에 쓰레기 무단 투기하는 사람들을 신고하는 파파라치 제도를 만드세요.
- 8 먹고 살아야 하기 때문에 파파라치가 단속반들보다 더 죽어라고 무단 투기자들을 잡아낼 겁니다.
- 7 Rather, install trashcans and make a paparazzi system where people who don't use the trashcan and illegally throw garbages away can be reported.
- 8 People will discover the illegal trash more thoroughly than squads because of economic incentives.

(24) After Converting into Content Word Vectors

- 7 [차라리, 쓰레기통, 만들, 쓰레기통, 안, 버리, 주위, 쓰레기, 무단, 투기, 사람, 신고, 파파라치, 제도, 만들]

8 [떡, 살, 하, 파파라치, 단속, 반, 더, 죽, 무단, 투기자, 잡아내]

(25) A Union of Content Word Vectors

[차라리, 쓰레기통, 만들, 안, 버리, 주위, 쓰레기, 무단, 투기, 사람, 신고, 파파라치, 제도, 떡, 살, 하, 단속, 반, 더, 죽, 투기자, 잡아내]

As shown in Examples (23), (24) and (25), each word in the united content word vector is counted and used as the unigram feature. These features just use the text itself, checking the existence and counting the frequency of specific words. That is, no additional motivation or assumption is required to deal with the data, except the target text itself. Therefore, bag-of-words, unigrams, or word pair features have been regarded and used as the most fundamental feature in many previous studies (Lawrence et al., 2014; Lin, Ng, and Kan, 2014; Palau and Moens, 2009; Peldszus, 2014; Saito, Yamamoto, and Sekine, 2006; Song et al., 2014; Stab and Gurevych, 2014).

After constructing a feature dictionary like above, a machine learning algorithm is used to train the annotated data and predict the answer of the newly provided test data. This study uses Support Vector Machine to train and predict discourse relations between basic units.

Support Vector Machine (SVM) is one of the learning machines which basically deal with 2-class classification problems. The SVM classifies the input data of given tasks, based on the feature space information where each piece of data is mapped. As shown below in Figure 4.1, the machine sets a linear surface between the data points located on a high-dimensional feature space, classifying the data points into one of the two divided parts. The linear surface, being a decision boundary of the classification procedure, is also

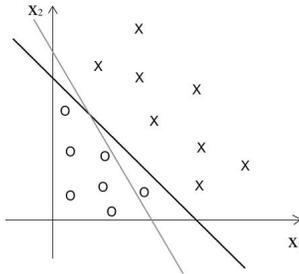


Figure 4.1: An Example of Data Distribution and Hyperplanes

called a separating hyperplane. The way of defining and using this hyperplane depends on the configured mathematical formulas.

$$y = g(w^T x + b) \tag{4.2}$$

The class which a data point would be assigned is decided by the value of $g(w^T x + b)$ in equation (4.2). The values of y in equation (4.2) show the type of classes to be decided, as equation (4.3) illustrates.

$$g(z) = \begin{cases} 1 & \text{if } z \geq 0 \\ -1 & \text{otherwise} \end{cases} \tag{4.3}$$

The ability of SVM to automatically learn given features of data and predict some new data is implemented as a scikit-learn package. Scikit-learn

is a Python tool for Machine Learning, including various implementations of Support Vector Classification procedures. Among those, this study selected LinearSVC, since it is possible to deal with a multi-class classification task. As the first stage of classification has only two options, O or X, it is not meaningful to use a multi-class classification method, just functioning the same as the binary classification method. The second stage, however, has three classes to choose from, which are NN, NS, and SN. In this case, the multi-class classification method is indispensable. The LinearSVC is based on the “one-vs.-the-rest” classification strategy, training three models: NN vs. the rest, NS vs. the rest, and SN vs. the rest.

4.3 Proposed Model

A series of automatic procedures predicting discourse relations between basic units and detecting argumentation schemes of debate texts based on the predicted relations are designed as in Figure 4.2.

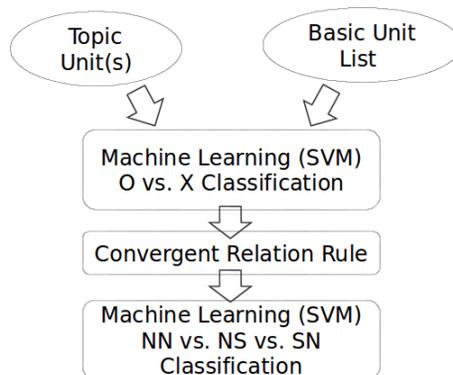


Figure 4.2: The Overall Flow of the Proposed Model

First, with the basic units defined from post texts and their topic unit(s), an SVM-based classifier decides which of the discourse relation, O or X (as defined in 3.3.2) should be assigned to each basic unit pair. Whether the two basic units are related in the same context or not is identified in this stage.

Now given an O or X label for each basic unit pair, a rule-based post-processing method is applied to improve the performance of the O vs. X classification above. A rule can help correcting some cases which were misclassified by the automatic classifier.

After that, the second stage of classification is performed on only the basic unit pairs which were classified as O in the first stage. If there is some relation between the two basic units, this classifier identifies the type of relation among the discourse relations NN, NS, and SN (as described in 3.3.2).

If all the basic unit pairs in a post are given one of the defined discourse relations, they are summed up and constitute the overall structure of argumentation which indicates the post. This procedure will be explained in detail in Chapter 5.

4.3.1 O vs. X Classification

Both of the two classification stages designed in this study undergo the training and predicting procedures using Support Vector Machine, as the baseline did in Section 4.2.

In this section, a classification model is proposed to identify the discourse relation between basic units, annotated as O or X, and to predict the discourse relations to be assigned to the newly given basic unit pairs for testing. Features used for this procedure are learned and generalized based on studies of

debate texts and argumentation schemes. Several features refer to the concept of text cohesion from Halliday and Hasan (1976), which defined four cohesion types including reference, substitution, ellipsis, and conjunction.

The features for this stage of classification are defined below. Most of the features except EXPCON, DEPENDRULES and UNIGRAMS are newly defined and introduced in this thesis, observing and patternizing characteristic properties of debate texts and sometimes of Korean sentences.

ExpCon The proposed model in this study defines a list of specific explicit connectives and mark them if those explicit connectives appear at some pre-defined position between the two basic units. If so, the proposed model counts how many times the explicit connective appears in the text, although the frequency is usually one.

Explicit Connectives are elements markedly used between sentences or clauses, showing their connections and relations. As an example, if a sentence follows another, and begins with ‘For instance’, it is easy to expect that this latter sentence is likely to support the former one by providing an example about it. Since prediction tends to be more accurate, explicit connectives were often used in various previous works of Natural Language Processing (Alistair and Diana, 2005; Fraser, 1999; Huang et al., 2014; Jang and Shin, 2010; Morsy and Rafea, 2012; Peldszus, 2014; Prasad et al., 2007; Stab and Gurevych, 2014; Taboada and Angeles, 2012; Wellner and Pustejovsky, 2007).

In the case of Korean, those explicit connectives mostly appear as connectives at the beginning of sentences or at the end of clauses. Thus, Korean explicit connectives are collected based on the properties of po-

sition and usage. This study referred to a collected list of shifters of Jang and Shin (2010) and connectives used in Korean news articles in Jang, Kim, and Shin (2013).

This feature consists of lexical cues according to the types of relations between two basic units. As described in Section 3.3.1, the detailed types of relations include reasons, background knowledge, examples, cause-and-effect, details, and so on. Those strategies corresponding to discourse relations between two basic units influence which explicit connective is shown in the basic units. This clustering of explicit connectives are assumed to be helpful to the classification task. Some of them are proposed below.

Enumeration/Serialization 또한, 더불어, 게다가, 더욱이, 더구나, 더군다나, 더더군다나

Reason 왜냐하면, 가령, 이를테면, 때문, 덕분에, 까닭, ...

Example 예컨대, 특히, 예-, 예를 들-, ...

Cause-and-Effect 1 따라서, 그러니까, 그러므로, 그리하여

Cause-and-Effect 2 아서, 어서, 니, 므로, 으므로, 으니, 라서, 으니까, 니까, ...

Concession/Contrast 지만, 는데, 아도, 라도, ㄴ데, 더라도, 그러나, 하지만, 그런데, 그렇지만, 허나, 그래도, 그런데도, ...

Switching 한편, 혹은, 아무튼, 어쨌든, 하여튼, ...

Each discourse relation has a separate list of explicit connectives according to the clustering procedure above. In this section, the lists of explicit

connectives are assigned to the discourse relations O and X. The same lists mapped to the subclasses NN, NS, and SN, are also defined. Since they are subclasses of the relation O, all of them are used to classify the basic unit pairs as having properties of discourse relation O at this stage. The connective lists of those subclasses will be described in detail in the second classification stage in Section 4.3.3.

Unit2woNV This is a boolean feature which marks whether the second basic unit in a basic unit pair has marked subjects, objects and predicates. If a basic unit does not have any proper subjects, objects or predicates, the basic unit would have limited information. In this case, the omitted components of the basic unit would refer to the information in former units. Thus, it is likely that the basic unit is dependent on another basic unit, usually to the former one. This feature is defined only between basic units which are adjacent to each other.

(26) An Example of UNIT2WOINV

- 0 쓰레기통 설치하고 돈 받아라.
- 1 10원이라도.
- 0 Install trashcans and charge for it.
- 1 At least 10 won.

Based on the idea above, if a basic unit depends on another, the discourse relation between two basic units is expected to be classified as O.

The value of this feature is judged based on the parsed results of a Korean Dependency Parser (Lee, Kim, and Kim, 2014). The existence

of a specific sentence component is shown through the tagged part-of-speech of each word.

Unit2Anaphor This feature has a boolean value which indicates whether the latter unit in the basic unit pair includes pronouns or not. A pronoun should have its reference somewhere in the previous context, depending on the contents of the previous basic units. The coreference resolution task, which identifies the reference entity of the pronoun, was not implemented in this study. However, a basic unit that includes pronouns would never appear without any relation to previous statements. Thus, that basic unit with pronouns has some contextual relation with its former one. Thus, it is expected to be in a discourse relation O.

(27) An Example of UNIT2ANAPHOR

7 겨울철 눈이 와도 자기집 대문 밖 제설작업을 하는 사람은 손에 꼽을 정도이다.

8 이런 국민 의식하에서 길거리에 휴지통을 만든다면 어떻게 될까?

7 Few people clear snow from the roads near their house even if it snows.

8 Under such national consciousness, what will happen with trashcans on the street?

AreSameSent This feature is defined only between adjacent basic units. A boolean value checks whether both of the two basic units in a basic unit pair are included in the same sentence or not. If both basic units

are components of the same sentence, it is more likely that the two basic units are related in the same context. If the value shows TRUE, the discourse relation between the two basic units is expected to be O.

(28) An Example of ARESAMESENT

- 1 삼겹살 파티만 하는 것이 아니라
- 2 술이 따라 올 것입니다.
- 1 Not only a barbecue party,
- 2 there must be some drinks too.

BothAreRoot This feature indicates whether the two basic units are both topic units of the post or not, as a boolean value. If both of the two basic units are topic units, these two units are equivalent each other and state the same meaning. Thus, the discourse relation between the basic units tends to be classified as O at this stage of O vs. X classification.

(29) An Example of BOTHAREROOT

- 0 쓰레기통이 적당히 있어야 한다고 생각합니다.
- 7 쓰레기통은 있는 게 맞다고 생각합니다.
- 0 I think that trashcans should be located appropriately.
- 7 I think it's reasonable for trashcans to be located.

ParenOpen A boolean feature checks if the second unit of the basic unit pair starts with an opening parenthesis. Statements in the parenthesis usually give additional or detailed explanations for the previous statements. Therefore, if this boolean feature has a TRUE value, the proposed

model is likely to predict the discourse relation between the two basic units to be O, with the latter basic unit explaining more about the former one.

(30) An Example of PARENOPEN

11 그러나 쓰레기를 길거리 휴지통에 버릴 수 있으니 길거리 휴지통에 잡다한 쓰레기가 넘칠 것이다.

12 (음식물, 플라스틱 커피통과 커피, 담배꽂초, 예측할 수 없는 쓰레기 등등)

11 But they may throw (their personal) garbage in the street trashcan, thus making the street trashcans full of random garbage.

12 (food waste, a plastic can with coffee in it, cigarette butts, other kinds of unpredictable garbage, etc.)

ReAction This boolean feature is defined to indicate if the second basic unit in the unit pair includes any reactions or subjective opinions about the first one. This feature uses another dimension of lexical information related to the argumentation strategies. Sometimes a statement is followed by some evaluations based on personal, subjective opinions. This case also includes the basic units in the position of the second unit, whose form explicitly seems not to be especially dependent on previous basic units. However, those second basic units showing subjective evaluations or opinions towards the first one in the pair are expected to use adjectives with subjectivity. The list of adjectives showing subjective polarities refers to the subjectivity lexicon of Jang and Shin (2010). If the latter basic unit in the pair contains at least one of the subjective

adjectives or adjectival nouns, the proposed model is expected to classify the basic unit pair as having the discourse relation O.

(31) An Example of REACTION

3 그 결과 곳곳에 쓰레기가 보이고 더럽고 취해 자는 사람들도 있습니다.

4 꼴불견이죠.

3 As a result, trash is here and there, it is dirty, and some people lays (on streets) drunken.

4 Ugly.

NeedsReason This feature has a boolean value indicating whether the first basic unit in the basic unit pair requires additional explanation or not. This is designed for the case where the first basic unit in the pair is a very subjective statement, expecting the second one to provide some grounds or elaboration for the former basic unit. If the first basic unit includes one or more subjective verbs, adjective verbs, verbal nouns, or adjectival nouns (Jang and Shin, 2010), the feature is assigned a TRUE value, thus the discourse relation between the two basic units is likely to be classified as O.

(32) An Example of NEEDSREASON

0 더위 먹은 미친 행정.

1 담배꽁초, 음료 마신 빈 컵 하나 휴지통에 제대로 버릴 줄 모르는 사람들이. 삼겹살 파티!

0 A crazy, insane administration.

- 1 For the people who don't know how to throw the cigarettes and empty beverage cups into trashcans, a barbecue party!

AreBothNR This boolean feature checks if the both of the units in the basic unit pair start with numbers. This study assumes that debating texts tend to systematically provide one or more grounds to support their conclusions. As a systematical way of reasoning, many written texts number each basic unit of grounds.

(33) An Example of AREBOTHNR

- 7 단점 1. 통 주변이 너무 지저분합니다.
- 8 2. 악취가 1년 내내 보행자들을 괴롭힙니다.
- 7 Cons 1. It's too dirty around the trashcan.
- 8 2. A bad smell annoys pedestrians all the year over.

Therefore, if both basic units are numbered statements, they are expected to be separate grounds which support the same conclusion. In such cases, the proposed classifier would decide the discourse relation between the two basic units to be X, since those units are not likely to influence each other directly.

ParentToNR In a basic unit pair whose basic units are adjacent to each other, this feature checks if the second basic unit starts with a number as explained above, especially as the first element of a list of the numbered units. At the same time, the first basic unit of the pair is assumed not to begin with a number. In this case, the former basic unit is expected to function as a nucleus unit being supported by the latter, which becomes

a satellite of the former. In other words, if a basic unit begins with a number, i.e. ‘1’ or ‘first’, the immediately preceding basic unit is likely to be the common conclusion of the numbered units which would be listed later in the text.

(34) An Example of PARENTTONR

- 1 문제점만 적어 보죠.
- 2 1. 잔디에 떨어진 기름.
- 1 I’ll provide the problems only.
- 2 1. Oil dropped on the grass.

Assume two basic units x , y form a basic unit pair $U(x, y)$. If the latter basic unit y in the pair U begins with a number other than the number ‘1’ or ‘first’, y mostly states the second, third, or later elements of the numbered basic units. In such cases, assuming the two basic units are not adjacent, if the former basic unit x in that pair is not numbered and a new unit z which directly follows x begins with the number ‘1’ or ‘first’, then x would be the common conclusion of z and y .

(35) Another Example of PARENTTONR; discourse relation between

Units ① and ⑭

- 1 문제점만 적어 보죠.
- 2 1. 잔디에 떨어진 기름.
- 14 5. 주차난.
- 1 I’ll provide the problems only.
- 2 1. Oil dropped on the grass.

If the feature confirms that the latter basic unit in the pair is a subordinate, numbered unit providing grounds for the former one, it would help the proposed model predict the discourse relation between the basic units to be O.

DependRules This feature refers to Lin, Ng, and Kan (2014) marking the existence of specific dependency relations between components of each basic unit in a basic unit pair. Given the parsed results of dependency in each basic unit of the unit pair, the list of the production rules in the internal structure of units are constructed respectively. Then the two lists are united to form a combined list of production rules of dependency. For each rule in the list, three boolean values are assigned: (1) if the production rule appears at the first basic unit, (2) if the production rule appears at the second basic unit, and (3) if the production rule appears at both of the basic units in the basic unit pair.

Unigrams This is the same basic feature which was used in Section 4.2 when the baseline experiment was performed.

4.3.2 Convergent Relation Rule

A rule is defined to raise the accuracy of the O vs. X classification described in Section 4.3.1. Named a CONVERGENT RELATION RULE, the rule defined here detects the basic units which should have a convergent relation, both ultimately supporting the same topic but not having any direct contextual relation with each other.

The rule first predicts the discourse relation O or X between the basic units which are close to the beginning of the post, and takes advantage of this prediction to identify the discourse relations between other basic units. This procedure is applied to every basic unit pair, increasing the indices of basic units from 0 to the maximum.

The cases where the two basic units in the pair should have the discourse relation X are defined and picked out by the convergent relation rule. If a basic unit pair is to be assigned a discourse relation, the relations which the former unit of the pair are involved in are considered as the condition of decision. In case the former basic unit in the pair already began to constitute a new subtree of context, the latter one in the pair is not included in that subtree too. Thus, the latter basic unit is not related to the basic units in the previous contextual subtree, as they all have the discourse relation X. If the basic unit pair is assigned the relation O in the automatic classification stage, this rule forces the pair to have the relation X.

This procedure is represented by a simple algorithm, written in pseudo-code, as shown below in Algorithm 1.

Algorithm 1 Convergent Relation Rule

```

for  $j = 1$  to  $\max(\text{unitIndex}) + 1$  do
  for  $i = 0$  to  $j$  do
    for  $x = i$  to  $0$  do
      if  $\text{relation}(x, i) == "x"$  then
         $\text{relation}(x, j) \leftarrow "x"$ 
      end if
    end for
  end for
end for

```

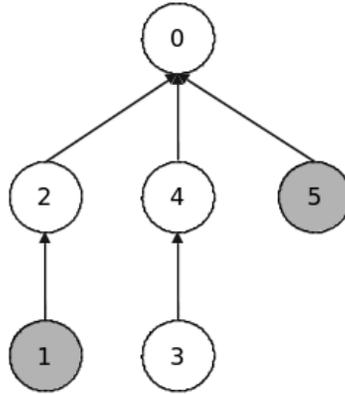


Figure 4.3: Two Unrelated Basic Units in the Argumentation Structure

For example, if the rule wants to check the predicted discourse relation between the basic units ① and ⑤ in the example tree structure (Figure 4.3), it would be classified as X, indicating that the two basic units are not related to each other in a given context. The basic units ① and ⑤ are included in the separate subtrees, as they are involved in different discourses from each other. Since even the basic units ③ and ④ already form a new, distinct context subtree, the unit ⑤ of the other subtree cannot be related to the former subtree of the units ① and ②. That is, the contents of a certain discourse are already completed before discovery of ⑤, so it does not have a chance to be connected with them.

Figure 4.3 illustrates the structure of a document split and numbered at (4) (restated in (36)). The basic unit ① describes the possible source of garbage, while the basic unit ⑤ proposes a solution regarding the given topic. Thus, the two basic units are dealing with different subcontents from each other, even if both of them are describing the same topic. Therefore, the

basic units ① and ⑤ are not necessarily related to each other, and, thus, are not connected in the tree structure of the argumentation scheme in the post.

(36) A List of Basic Units in an Example Post

- 0 쓰레기통이 있어야 합니다.
 - 1 길 가다가 편의점이나 슈퍼 같은 데서 아이스크림이나 음료 과자 같은 거 사 먹고 나면 쓰레기가 발생하는데
 - 2 쓰레기통이 없으니 많이 불편하죠.
 - 3 예전에 그 많던 쓰레기통 그 쓰레기통 주변이 더럽다고 치워 버렸는데
 - 4 쓰레기통을 즉각 즉각 비우면 더럽혀질 일이 별로 없을 겁니다.
 - 5 도심의 길거리 쓰레기통 설치하고 관리하는 미화원 배치하면 일자리도 생기고 거리도 깨끗해지고 좋을 것 같네요.
-
- 0 There should be trashcans.
 - 1 Trash is created when people have ice cream, beverages, or snacks at convenience stores or supermarkets on streets,
 - 2 and it is inconvenient if there's no trashcans.
 - 3 While the trashcans which formerly were large in number were removed because surrounding area was dirty,
 - 4 this would not be the case if trashcans are emptied often.
 - 5 If sanitation workers are allocated to install trashcans on downtown streets and care for them, it would be good with more jobs and cleaner streets.

This rule is supposed to help the proposed model check whether the discourse relation between units ① and ⑤ above are correctly predicted, and improve the performance of classification by revision. Even if the prediction procedure mistakenly decided the two basic units have an O relation, the CONVERGENT RELATION RULE can correct it to be X.

4.3.3 NN vs. NS vs. SN Classification

After a rule-based procedure improves the prediction of the former O vs. X classification, another procedure to classify the relation as NN, NS or SN is performed, also using the Support Vector Machine. This stage of classification deals with only the basic unit pairs which were predicted as having the discourse relation O in the former stage.

This stage of classification constructed and used the features of the discourse relations NN, NS and SN, respectively. Similar to the features in Section 4.3.1, the features used in this stage are also based on the strategies of argumentation. However, since the separation between O and X is no longer needed, the features which simply check contextual relatedness between two basic units are not used here. Thus, features designed to distinguish between O and X, e.g. the identification of coherence, are excluded.

ExpCon This features marks if certain explicit connectives appear at some pre-defined position between the two basic units. If so, the proposed model counts how many times the explicit connective appears in the text, although the frequency is usually one.

BothAreRoot This feature shows whether the two basic units are both the topic units of the post or not, as a boolean value. If this feature is

assigned a TRUE value, the discourse relation between the basic units tends to be classified as NN, because the two basic units indicate the same topic content.

ParenOpen This feature checks whether the second unit of the basic unit pair starts with an opening parenthesis or not. Usually statements in the parenthesis show additional or detailed explanations of the previous content. Therefore, if this boolean feature has a TRUE value, the discourse relation between the two basic units is classified as NS.

ReAction This feature is defined to mark if the second basic unit in a basic unit pair contains any reactions or subjective opinions about the first unit. In this case, the two basic units in the pair can be grouped into a segment inferring the same content. Thus, this case would help the proposed model to classify the discourse relation into NN.

NeedsReason This binary feature indicates whether the first basic unit in the basic unit pair requires additional explanation or not. If the first basic unit in the pair is a very subjective statement which expects the second one to provide some grounds or elaboration, then the discourse relation in the pair is likely to be judged as NS.

ParentToNR This feature checks whether the second basic unit in a unit pair is dependent on the first unit. In the case of enumerated elements with numbers in front of each segment, the marked numbers may be a clue for connecting them to each other. This feature relates a numbered basic unit to its contextual headline, letting the discourse relation between them be classified as NS.

DependRules This feature uses the list of production rules in the structure, as implemented in Section 4.3.1.

Unigrams This is the same basic feature which was used in Sections 4.2 and 4.3.1, when the baseline experiment was performed.

In case of the same features from Section 4.3.1, only the types of candidate discourse relations are different from those of the previous descriptions.

4.4 Evaluation

4.4.1 Measures

All the experiments designed above are done using the Support Vector Machine. The posts in the collected data are divided randomly into ten pieces, each containing the same number of posts, to form a 10-fold cross validation method. 90% of all the documents are used as training data, while the remaining 10% constitute the test data whose labels are to be predicted. The experiment is repeated 10 times for all 10 pieces of the divided posts, finally obtaining the average performance.

Performances of the proposed model are usually measured by a *F1-measure*. The F1-measure includes precision, recall, and F1-score to express, and expresses how well a proposed model predict class labels through classification procedures.

- *Precision* shows how many predictions done by the model are correct.

For each class label, given the number of predictions classified as the

class, precision indicates how many of the predictions are actually mapped to the class.

$$precision = \frac{TruePositive}{TruePositive + FalsePositive} \quad (4.4)$$

- *Recall* indicates how many predictions correctly find the cases which have a certain class label in the gold standard. Thus, for each class, recall shows how many of the annotated elements are detected by the model.

$$recall = \frac{TruePositive}{TruePositive + FalseNegative} \quad (4.5)$$

- *F1-score* is calculated using precision and recall.

$$F1 = \frac{2 \times precision \times recall}{precision + recall} \quad (4.6)$$

4.4.2 Results

This section illustrates the results of experiments designed in Section 4.3. First, tables 4.3 and 4.4 show the performance of the tasks designed to distinguish between the discourse relations O and X. The baseline, the case calculated with all the defined features, and the case revised by the CONVERGENT RELATION RULE are included in the table. The experiment was performed for 20,810 basic unit pairs after removing some exception cases.

		precision	recall	f1-score	support
BASELINE	O	0.24	0.39	0.29	3603
	X	0.84	0.71	0.77	12707
	avg/total	0.73	0.66	0.69	20810
FEATURES	O	0.24	0.43	0.31	3603
	X	0.84	0.70	0.76	17207
	avg/total	0.74	0.65	0.68	20810
FEATURES-LESS	O	0.26	0.41	0.31	3603
	X	0.85	0.73	0.78	17207
	avg/total	0.74	0.68	0.70	20810
RULE	O	0.41	0.28	0.33	3603
	X	0.85	0.90	0.87	12707
	avg/total	0.77	0.80	0.78	20810

Table 4.3: The Overall Performances of O vs. X Classification

In Table 4.3, BASELINE shows the performance which was trained and tested by the unigram features of Section 4.2. FEATURES illustrates the experimental results with all the features described in Section 4.3.1, while FEATURES-LESS means the result trained with the features like FEATURES, but without the feature DEPENDRULES which is a list of production rules extracted from the dependency structures of texts. Finally, RULE means the performance calculated after revising the predicted result of FEATURES with the CONVERGENT RELATION RULE.

Table 4.4, below, is the list of performances which show the contribution

of individual features used in FEATURES in Table 4.3. The classification of discourse relations O and X are performed with each feature accompanied by the unigram features only.

		precision	recall	f1-score	support
BASELINE	O	0.24	0.39	0.29	3603
	X	0.84	0.71	0.77	12707
	avg/total	0.73	0.66	0.69	20810
EXPCON	O	0.24	0.39	0.29	3603
	X	0.84	0.71	0.77	17207
	avg/total	0.73	0.66	0.69	20810
UNIT2WOINV	O	0.39	0.48	0.42	3603
	X	0.88	0.83	0.85	17207
	avg/total	0.79	0.83	0.85	20810
UNIT2ANAPHOR	O	0.24	0.39	0.29	3603
	X	0.84	0.71	0.77	17207
	avg/total	0.73	0.66	0.69	20810
ARESAMESENT	O	0.24	0.39	0.29	3603
	X	0.84	0.71	0.77	17207
	avg/total	0.73	0.66	0.69	20810
BOTHAREROOT	O	0.24	0.40	0.29	3603
	X	0.84	0.71	0.77	17207
	avg/total	0.74	0.66	0.69	20810
	O	0.24	0.39	0.29	3603

PARENOPEN

	X	0.84	0.71	0.77	17207
	avg/total	0.73	0.66	0.69	20810
AREBOTHNR	O	0.24	0.39	0.29	3603
	X	0.84	0.71	0.77	17207
	avg/total	0.73	0.66	0.69	20810
PARENTTONR	O	0.24	0.39	0.29	3603
	X	0.84	0.71	0.77	17207
	avg/total	0.73	0.66	0.69	20810
REACTION	O	0.24	0.39	0.29	3603
	X	0.84	0.71	0.77	17207
	avg/total	0.73	0.66	0.69	20810
NEEDSREASON	O	0.24	0.39	0.29	3603
	X	0.84	0.71	0.77	17207
	avg/total	0.73	0.66	0.69	20810
DEPENDPARSED	O	0.22	0.42	0.29	3603
	X	0.84	0.66	0.74	17207
	avg/total	0.73	0.62	0.66	20810

Table 4.4: Individual Feature Contributions of O vs. X Classification

On the other hand, Table 4.5 describes the performance of the classification of discourse relations NN, NS and SN. The classification of NN, NS and SN targets the discourse relations which were classified as O in the first stage of classification. However, since it is impossible to get the pure performance of this stage if the class O is assigned by prediction, the second stage of clas-

sification is performed with the 3,603 basic unit pairs which were manually annotated as O.

		precision	recall	f1-score	support
BASELINE	NN	0.34	0.37	0.35	828
	NS	0.70	0.70	0.70	2027
	SN	0.43	0.40	0.41	748
	avg/total	0.56	0.56	0.56	3603
FEATURES	NN	0.40	0.44	0.42	828
	NS	0.71	0.74	0.73	2027
	SN	0.51	0.43	0.46	748
	avg/total	0.60	0.61	0.60	3603
FEATURES-LESS	NN	0.41	0.45	0.43	828
	NS	0.71	0.72	0.72	2027
	SN	0.48	0.43	0.45	748
	avg/total	0.60	0.60	0.59	3603

Table 4.5: The Overall Performances of NN vs. NS vs. SN Classification

Like the cases above, BASELINE means the experiment using only the unigram features. FEATURES and FEATURES-LESS illustrates the performances of the experiment with all the features and the features without the dependency production rules, respectively.

In addition, Table 4.6 shows the individual contributions of the features defined in Section 4.3.3 and used in FEATURES in Table 4.5. Each experiment

was performed to classify the discourse relation of a basic unit pair into NN, NS and SN, adding each feature to the unigram features.

		precision	recall	f1-score	support
BASELINE	NN	0.34	0.37	0.35	828
	NS	0.70	0.70	0.70	2027
	SN	0.43	0.40	0.41	748
	avg/total	0.56	0.56	0.56	3603
EXPCON	NN	0.34	0.37	0.35	828
	NS	0.70	0.71	0.70	2027
	SN	0.45	0.41	0.42	748
	avg/total	0.57	0.57	0.56	3603
BOTHAREROOT	NN	0.41	0.43	0.42	828
	NS	0.71	0.72	0.72	2027
	SN	0.46	0.42	0.43	748
	avg/total	0.59	0.59	0.59	3603
PARENOPEN	NN	0.34	0.37	0.35	828
	NS	0.70	0.70	0.70	2027
	SN	0.43	0.41	0.41	748
	avg/total	0.56	0.56	0.56	3603
PARENTONR	NN	0.34	0.37	0.35	828
	NS	0.70	0.70	0.70	2027
	SN	0.43	0.40	0.41	748
	avg/total	0.56	0.56	0.56	3603

REACTION	NN	0.34	0.37	0.35	828
	NS	0.70	0.70	0.70	2027
	SN	0.43	0.40	0.41	748
	avg/total	0.56	0.56	0.56	3603
NEEDSREASON	NN	0.34	0.37	0.35	828
	NS	0.70	0.70	0.70	2027
	SN	0.43	0.40	0.41	748
	avg/total	0.56	0.56	0.56	3603
DEPENDRULES	NN	0.32	0.36	0.34	828
	NS	0.70	0.71	0.70	2027
	SN	0.45	0.38	0.41	748
	avg/total	0.56	0.56	0.56	3603

Table 4.6: Individual Feature Contributions of NN vs. NS vs. SN Classification

4.5 Discussion

First, the distinction between the discourse relations O and X shows some rise in performance from the baseline to the results performed by features and the CONVERGENT RELATION RULE, with the highest average values of precision, recall, and F1-score being 0.77, 0.80, and 0.78, respectively. As the average precision and recall exceed 0.70, this stage of classification identifying whether two text segments are related or not seems to be reasonable. All of the precisions, recalls, and f1-scores in the overall performances in Table 4.3

shows the significant differences ($p < 0.05$) between the baseline and the case with features, and between the baseline and the case revised by the CONVERGENT RELATION RULE.

The performance of discourse relation O is relatively low while that of X is very high. This is because the frequency of X is high in the training data (82.7%, or 17207 of 20810 basic unit pairs). It is a natural distribution because if a new context subtree is constructed, the discourse relations are usually not constructed across the subtrees.

The overall performance is a little higher when the CONVERGENT RELATION RULE is applied to the experiment, compared to the case when only the features are used. The average precision rises from 0.74 to 0.77 and the average recall surges from 0.65 to 0.80 by using the revision rule. The difference seems not that significant, because during the application of the rule to the predicted results, some of the properly-predicted O are sacrificed by changing the O's which are thought to be mispredicted to X. If the rule is applied based on the predicted labels, not on the gold standard, the incorrect prediction may negatively influence the performance. However, removing the mispredicted O's raised the precision of O sharply, from 0.24 to 0.41. The wrong O which was originally to be assigned as X, was changed to X, raising the recall of X from 0.70 to 0.90. The precision of X also grew a little. But O's recall dropped from 0.43 to 0.28. The proposed model in this study performs the second stage of classification with the predicted result of O and X's distinction, so the precision of the former task seems to be more important than recall. This becomes a reasonable trade-off.

In addition, as the features combine to raise the performance, it is likely

the features combine and interact positively with each other. Some of the features largely contributed to increase the performance.

ExpCon Some categories of connectives contributed to raise the recall of O. The subclasses of effective connectives explicitly illustrate the strategies which provide the cause-and-effect, reasons, and contrasts in the procedure of arguments. Those strategies help to find that two text segments are related in a discourse. Thus, the existence and subcategories of argument strategies which are proposed in this study are proven to be appropriate.

However, when all ten subclasses of explicit connectives are included in the feature set, the rise of the performance is not significant. This is because the explicit usage of connectives is unexpectedly little, even if authors want to connect and relate text segments to each other.

Unit2woNV The precision on average grows from 0.73 to 0.79, and the average recall from 0.66 to 0.83 by using this feature. The feature alone improves the performance rapidly, and it is assumed to have the largest contribution to the model among all the features used in the O vs. X classification. This confirms the hypothesis that a text segment without the important components of sentences would depend on the former segment in terms of meaning.

BothAreRoot This feature develops the recall of O and the precision on average. The topic units of texts are defined first to be in the relation of paraphrasing or just the same form, indicating the same meaning. The improvement is not that significant in terms of numerical values, but

this is because the absolute number of discourse relations between the topic units is so few.

DependRules This raised the recall of O from 0.39 to 0.42, but lowered O’s precision and X’s recall. The range of the decrease is bigger than the increase, so it seems not to be a reasonable trade-off.

The experiment with all the features except the dependency production rules performs like in Table 4.3, raising the average recall from 0.65 to 0.68. The precision of O and X also improved. Thus, the type and existence of dependency structures in sentences do not have any significant influence on the decision of relatedness in a discourse.

Additionally, the NN vs. NS vs. SN classification is performed for the 3,603 basic unit pairs annotated as O from the 20,810 total pairs. It is only 17.3% total pairs, and the number of relations annotated as NS is less, being at 2,027 basic unit pairs. Thus, the training data is somewhat biased.

For all three subclasses NN, NS, and SN, both the precision and recall improved by using the features in the experiment. On average, precision grew from 0.56 to 0.60, while recall increased from 0.56 to 0.61. Especially in case of NN and SN, the performance improvement is drastic (NN’s precision grew from 0.34 to 0.40, recall from 0.37 to 0.44, while SN’s precision from 0.43 to 0.51, recall from 0.40 to 0.43). This difference of performance effected by the features is significant ($p < 0.05$), in terms of all the precision, recall, and f1-score values.

At this stage, the features also interact positively with each other, as the performance is the highest when many features are collected altogether. And

some of the features greatly effect the model to improve the performance.

ExpCon When this feature is used, the precision and recall on average grow a little (both 0.56 to 0.57). The recall of NS rises in a small range, while SN improves both precision and recall. In the classification of NN, NS, and SN, the subclasses of explicit connectives are supposed to function more sensitively than the binary classification of relatedness. Especially, the subclasses of connectives which indicate Results, Cause-and-Effect, and Contrast evidently affected the overall contribution of explicit connectives. In this stage of classification also, the elaborate categories of debate strategies are proven to be helpful to the identification of discourse relations in argumentation schemes.

BothAreRoot This feature contributes the most to the improvement of the performance in this stage of classification, raising both the average precision and recall from 0.56 to 0.59. The identification and precision of all three subclasses improved, with NN growing the most (precision from 0.34 to 0.41, recall from 0.37 to 0.43), since this is the feature defining the relation between topic units.

DependRules When the dependency production rules are included in the feature set, SN's precision and NS's recall rise a little, while the performance of other parts including NN drops. The overall performance on average does not show significant difference. However, when the proposed model performed the same classification without the dependency rules in the feature set, the performance of NS and SN worsens a little even though that of NN increases a little.

Therefore, the production rules form dependency structures do not seem to have a large effect on the classification of NN, NS, and SN.

Regarding the overall performances, it is observed that the cues illustrating the features above are not that explicit, unexpectedly. First, the list of explicit connectives infrequently matched to the real text data. Authors mostly express discourse relations by enumerating text segments without any explicit discourse markers, instead just relying on the meaning and content.

The subcategories of argumentation strategy defined in this study at least reflect the specific properties according to the genres and internal content of texts. Many previous works tend to depend on the distribution of raw text as continuous strings or characters. There were so many attempts to determine the implicit relations in discourse through various approaches, but their performances are not that reasonable yet. Lin, Ng, and Kan (2014) also tried to identify non-explicit relations with information about connectives, obtaining about 39.6% precision in maximum.

Also, it is obvious that a certain explicit connective contributes to the improvement of precision in the classification procedure, if the connective is confirmed to be used in a specific context in texts. However, it is difficult to extract the connectives used in the specific context exactly. For example, in the case of the connective *and*, it usually means the addition of content in a series of meaning. But it can be used in somewhat different contexts, e.g. beginning other subtrees with slightly different content, or adding explanations in detail, even with a low probability. That a connective can be used for various meanings and purposes is a well-established fact. That is, using connectives as a feature has the danger of having a low probability of being

matched as the string itself, and of not being captured with its correct meaning.

Second, the case of numbered elements is similar to that of explicit connectives. Numbering text segments like *first, second, ...* or *1., 2., ...*, etc. is a characteristic of persuasive writing, and the discourse relations between text segments with the numbered forms tend to be fixed. However, the actual data did not show many of them. This seems to be because the documents in this study are not essays or editorials whose forms are restrictive, but instead, they written in some free-style forms. Of course, the numbering of elements is not necessary for the enumeration of text segments, so the features regarding the numbered texts may not work well in those cases.

On the other hand, some features functioned not that effectively because of parametric properties of languages and existence of data sources in specific languages. The features REACTION and NEEDSREASON defined in Section 4.3.1 showed positive influences in a pilot experiment of English data, which will be described in Section 4.6. The two features are constructed based on the Subjectivity Lexicon (Wilson, Wiebe, and Hoffmann, 2005). In the case of Korean, a counterpart to this kind of lexicon does not exist. Instead, Jang and Shin (2010) constructed a dictionary for sentiment analysis, so this study took the list of subjective words in the dictionary for those features. The list of subjective words do not include the distinction of intensity of subjectivity, i.e. strong or weak, and the usage of subjective words would show different aspects in Korean and English. Therefore, the two features based on the subjectivity of words would have functioned less in Korean texts than in English texts.

Additionally, since the parsing of the syntactic structure of Korean sentences are not yet available, some features used for classification in previous studies of English texts were not able to be applied in this study. One of those is production rules in syntactically parsed results. Referring to the features Lin, Ng, and Kan (2014) defined, this feature shows the existence of the syntactic production rules observed in the syntactic structure of two basic units in a basic unit pair, contributing to the classification of discourse relations between text segments.

4.6 A Pilot Study on English Texts

A pilot experiment on English debate texts was additionally performed to justify the framework of argumentation schemes provided in this study. It is supposed that the ways of writing subjective texts and developing thoughts are unchangeable, regardless of the language used. All authors would persuade readers of their claims by suggesting appropriate grounds. Therefore, the pilot experiment used the same concepts of argumentation schemes and methods as in the main experiment of this study, to observe if the current study still shows reasonable results with texts written in English.

The experiment used the Political Debate Corpus of MPQA (Somasundaran and Wiebe, 2010) as the debate texts to be analyzed by the automatic module of this thesis. While the corpus includes subjective writings on several topics, the pilot experiment here used 287 posts dealing with only one topic,

that is, the reelection of President Obama.

This corpus also required the procedure of manually splitting raw texts into basic units. Then an annotator who is an English native speaker constructed the gold standard for this experiment by annotating the topic sentence and the discourse relations between basic units for each post.

The pilot experiment is designed as similar to the main experiment of this thesis with Korean debate texts as possible. There are also two stages of classification tasks, O vs. X, and NN vs. NS vs. SN, the former prediction revised by the CONVERGENT RELATION RULE to improve the performance of the proposed model. Both stages of classification are based on a 10-fold machine learning algorithm using Support Vector Machine. The baseline experiments are done with only the unigram features, and additional features from the main study in Chapter 4 are applied and sometimes revised for use with English texts: EXPCON, UNIT2WONV, UNIT2ANAPHOR, ARESAMESENT, BOTHAREROOT, PARENOPEN, REACTION, NEEDSREASON, AREBOTHNR, PARENTONR, DEPENDRULES. In particular, the feature EXPCON, i.e. explicit connectives, refers to the list of categorized explicit connectives of the Penn Discourse TreeBank (Prasad et al., 2007) and was classified again according to position and function of the connectives. UNIT2WONV and UNIT2ANAPHOR used the English part-of-speech counterparts which function the same as the Korean ones introduced in this thesis. In addition, the list of subjective words used in the features ReAction and NeedsReason refers to the Subjectivity Lexicon (Wilson, Wiebe, and Hoffmann, 2005). Only the words marked with strong subjectivity were chosen as subjective words to use.

Additionally, a feature which was not available for Korean texts is used

here for English texts. As the feature `DEPENDRULES` did, referring to the features Lin, Ng, and Kan (2014) defined, this new feature, `SYNTAXRULES`, shows the existence of the syntactic production rules observed in the syntactic structure of two basic units in a basic unit pair. In this study, the production rules in the syntactic structure of each basic unit are collected relatively, using the Stanford Parser (Klein and Manning, 2003). Then the union of two lists of production rules is used as a feature bag. For each of the production rule in the union, the proposed model defines three boolean values: (1) if the production rule appears at the first basic unit, (2) if the production rule appears at the second basic unit, and (3) if the production rule appears at both of the basic units in the basic unit pair.

The results of the pilot experiments described above follow.

		precision	recall	f1-score	support
BASELINE	O	0.30	0.45	0.36	1075
	X	0.72	0.58	0.64	2708
	avg/total	0.60	0.54	0.56	3783
FEATURES	O	0.34	0.50	0.41	1075
	X	0.75	0.61	0.67	2708
	avg/total	0.64	0.58	0.60	3783
RULE	O	0.45	0.44	0.45	1075
	X	0.78	0.78	0.78	2708
	avg/total	0.68	0.68	0.68	3783

Table 4.7: The Overall Performances of O vs. X Classification in the Pilot Experiment

Table 4.7 shows the results of the O vs. X classification for the discourse relation of 3,783 pairs of basic units. FEATURES, which is the result of experiment with all the defined features, can be compared to BASELINE, the most basic experiment using only unigram features. It can be observed that the overall performance has improved with the addition of features. In addition, RULE, which the result with the CONVERGENT RELATION RULE to the predicted class labels of FEATURES, shows a slight increase in performance by greatly improving precision of the class O. This pattern is similar to that of performance of experiments on Korean texts of this thesis described in Section 4.4.

Then, the results of the NN vs. NS vs. SN classification follow.

		precision	recall	f1-score	support
BASELINE	NN	0.03	0.02	0.02	56
	NS	0.83	0.93	0.88	882
	SN	0.20	0.10	0.13	137
	avg/total	0.71	0.78	0.74	1075
FEATURES	NN	0.12	0.08	0.08	56
	NS	0.84	0.95	0.89	882
	SN	0.20	0.07	0.09	137
	avg/total	0.72	0.79	0.75	1075

Table 4.8: The Overall Performances of NN vs. NS vs. SN Classification in the Pilot Experiment

Table 4.8 shows the results of the NN vs. NS vs. SN classification for 1,075 basic unit pairs whose discourse relations were predicted as O in the previous stage of classification. The results indicate that this stage has a problem. The FEATURES, which is the result with all the features defined for this stage, shows a slightly improved overall performance compared to the result of BASELINE. However, the absolute numerical value for performance is very bad. In both cases, the performance of the NS class is overwhelmingly good, while the model hardly functions for the classes NN and SN. This seems to be affected by the biasedness of data, where most of the discourse relations annotated as O at the stage of O vs. X classification were predicted to have the class NS. That is, 882 of 1,075 (or about 82.04%) O relations were classified as NS. The biased data can prevent the proposed automatic model from training itself properly.

Furthermore, this pilot experiment showed some more problems like the following. First, the collected Korean texts are in the form of a post with a title. The title is regarded as the first sentence of the post and tends to include the condensed, important contents of the whole post. Thus, many posts are written in a deductive structure, and each post can be thought as a complete, individual document. However, the pilot study on English texts includes posts replying to other authors, or rebuttals. Even though each post can be regarded as a separate document, several posts have an author's personal opinion or reactions to others' posts. These are sort of noises in the data used in this study, and they might have negatively affected the result of this pilot experiment.

Secondly, unexpectedly few explicit features were used in this corpus of

English debate texts. This is much less than that of Korean texts in this thesis. This seems to be affected by the differences in language and cultural properties. In Korean, suffixes or connectives are used in most cases when connecting two sentences or clauses, regardless of whether it is written or spoken language. It is thought that continuing sentences without any connectives is unnatural. In contrast, connectives are less used in English texts, especially in informal texts. More implicit cues in developing logic in texts should be reflected in the model.

These problems including experimental methods and differences in language and culture will have to be resolved in a follow-up study in the future.

5 Detecting Important Units

When all the basic unit pairs are assigned their discourse relations through the model proposed above, they are collected and summed up to constitute a structure formed like a tree. This procedure was implemented by a relatively simple algorithm, which is provided in Algorithm 2.

Algorithm 2 Summing Up the Discourse Relations

depth : the vertical distance from the root node
parent: the list of the basic units to which the current unit is subordinated
children: the list of the basic units which are subordinated to the current unit
friend: the list of the basic units which are equivalent with the current unit

```
for all basic unit in the post do  
  if unit == root then  
    depth  $\leftarrow$  0  
  else if unit is more than two then  
    update friend (each other, all root units)  
  end if  
end for  
for all unit pair in the post do  
  if relation == NS then  
    depthunit2  $\leftarrow$  depthunit2 + 1  
    update parentunit2  $\leftarrow$  unit1  
    update childrenunit1  $\leftarrow$  unit2  
  else if relation == SN then  
    swap depthunit1 and depthunit2  
    update parentunit2  
    update parentunit1  
    update childrenunit2  
  else if relation == NN then  
    depthunit2  $\leftarrow$  n  
    update friendunit1 and friendunit2  
    update parentunit2  
  end if  
end for
```

Now all the basic unit nodes have an appropriate depth, lists of parent nodes, children units, and friend units used to construct a tree structure illustrating the argumentation scheme of the document. Using these properties, the tree structure indicating a post’s argumentation scheme can be visualized, as shown in Figure 5.1.

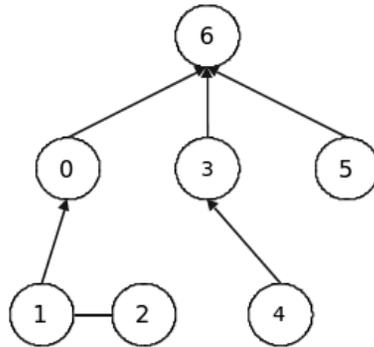


Figure 5.1: An Example of the Tree-Structured Argumentation Scheme

In the structure, with each node illustrating a basic unit, the root node indicates the topic unit of the post. The direct children nodes of the topic unit support it, and the children nodes of those children nodes again support their parent nodes. This relation continues to the final leaf node of the tree structure. For example, in the case of the tree structure shown in Figure 5.1, the root node is the sixth text segment, (6), of the post, with three subordinating nodes of text segment (0), (3) and (5). These three nodes directly support the topic sentence (6) of the document. Additionally, (0) is again supported by (1) and (2), while (3) directly supported by (4).

In particular, (1) supports (0) directly, while (2) supports (0) *indirectly*. The relation between (0) and (2) was expected to be X, indicating no relation-

ship in terms of content, in the procedure of automatic prediction. However, given enough contextual information from the surrounding text segments, it can be thought of slightly differently. ① supports ①, and ② is assumed to have the same meaning as ①, enabling nodes ① and ② to be thought as a merged node. Thus, the meaning indicated by ② is also supposed to back up ①. This study defines this case as a relation of indirect support, considered in the overall structure of a document.

The closer a basic unit is to the root topic unit in the structure, the more important the content of that basic unit is assumed to be. Therefore, this structure shows, at a glance, which part of the document is more important, and which basic units are the grounds for the important idea of the post. Since this study analyzes debate texts, it is expected for the argumentation schemes to show the topic unit and its main grounds in the post. This idea is also proposed by Marcu (1999), where a detected discourse structure contributes to score basic units in texts based on their depths in the tree-form structure. The scores are used to pick important contents from the text and finally summarize it. Thus, the procedure in this section also tried to summarize documents, especially selecting only the basic units whose depth in the overall tree structure is 1 (given the depth of the root node is 0).

In this case, the basic units with depth 1 may be directly connected to the topic unit or not. Some depth 1 units support their parent units indirectly. Also, the topic unit itself would not explicitly appear in some posts. All those cases are considered and are dealt with to get the basic units whose depths are 1 equally, as important grounds of the document.

The basic units which have the depth of 1 in a tree structure of an argu-

mentation scheme are collected. Then the depth 1 basic units are evaluated, since they are extracted from the structure which is automatically predicted by the proposed model in this study. The counterparts in the gold standard’s structure are used for comparison with the predicted ones, through the F1-measure, as in the previous evaluation procedure in Section 4.4.

	precision	recall	f1-score
performance	0.868	0.799	0.832

Table 5.1: The Depth 1 Units Detection Metrics

As the result of the experiment shown in Table 5.1, the precision came out to be 0.868 with the recall 0.799, yielding an f1-score of 0.832. That is, the proposed model detected about 79.9% of the actual basic units with depth 1, with 86.8% of them being correct predictions. This result lays in a quite reasonable range.

In fact, the important thing in the tree structure of argumentation schemes is the author’s intention and the topic of the document. The reasons why the author claims that idea are also included in the range of important contents of texts. In the frame of this study, where the concepts of nucleus and satellite in the discourse relation are considered, the basic units located in a deep position in the tree structure are supposed to be less important. This is because the units hung close to the topic unit, which is in the root position of the tree structure, indicate some weighty contents that relate to the topic unit more directly. Thus, even if the discourse relations of subordinate basic units disagree with those of the gold standard influencing the overall performance neg-

atively, some of them may be trivial errors in terms of argumentation structures. Therefore, this section shows the result of an experiment which checks the performance of predictions regarding only the basic units of depth 1.

Some examples of extracting important, depth 1 units from a full text follow.

(37) An Example of Post

- 0 쓰레기통 있어야 합니다.
- 1 쓰레기통 있을 때는 길 가다가 쓰레기(담배꽂초, 아이스크림 곁종이 등)가 발생하면 그래도 손에 쥐고 있다가 쓰레기통이 보이면 버렸는데
- 2 거리에 쓰레기통이 없으니 거리도 쓰레기로 지저분하고 처치도 곤란하고 그냥 버립니다.
- 3 죄의식 없이.
- 4 쓰레기통 있다고 집안 쓰레기 들고 나와 버려 봐야 얼마나 버리겠습니까?
- 5 그런 사람들이 쓰레기통 없다고 길에다 집안 쓰레기 안 버릴까요?
- 0 There should be trashcans.
- 1 When there were trashcans, we hold trash (e.g. cigarette butts, ice cream packages, etc.) on streets and threw them in trashcans,
- 2 but without trashcans on streets, the streets are dirty with trash, and we just throw them because it's hard to deal with them.
- 3 Without feeling guilty.

- 4 How much domestic trash would be thrown into the street trashcans?
- 5 Would those people not throw domestic trash on streets if there are no trashcans?

(38) Extracted Basic Units from (37)

- 2 거리에 쓰레기통이 없으니 거리도 쓰레기로 지저분하고 처치도 곤란하고 그냥 버립니다.
- 4 쓰레기통 있다고 집안 쓰레기 들고 나와 버려 봐야 얼마나 버리겠습니까?
- 5 그런 사람들이 쓰레기통 없다고 길에다 집안 쓰레기 안 버릴까요?
- 2 but without trashcans on streets, the streets are dirty with trash, and we just throw them because it's hard to deal with them.
- 4 How much domestic trash would be thrown into the street trashcans?
- 5 Would those people not throw domestic trash on streets if there are no trashcans?

Example (37) is a post which supports the installation of street trashcans, while (38) is the extracted depth 1 basic units through the proposed model in this study. That is, the basic unit ① in the example above is the topic sentence, being the root node in the constructed tree structure of the overall post. And the basic units ②, ④ and ⑤ are the depth 1 units which support the root unit 0. Unit ① is a fundamental statement to support what ② means, while ③ provides an additional explanation of ②. Therefore, it is

possible to understand which grounds the author proposes for the post saying ‘There should be trashcans on streets.’ just by reading the extracted basic units ②, ④, and ⑤.

(39) Another Example of Post

- 0 한강에는 지금도 술 마시는 사람들 많습니다.
- 1 지금도 한강에 가면 술도 팔고 각종 음식들 팔고 있습니다.
- 2 또한 식구들이나 친구들과 함께 그늘막 치고 함께 즐기고 합니다.
- 3 그 결과 곳곳에 쓰레기가 보이고 더럽고 취해 자는 사람들도 있습니다.
- 4 꿀볼건이죠.
- 5 그렇지만 이들 몇몇 때문에 요즘 같이 더운 날씨에 한강만큼 좋은 곳이 없는데 한국 사람들이 그렇게 좋아하는 삼겹살에 소주 좀 마신다고 문제될 것 있나요?
- 6 차라리 시민들이 즐기게 하고 약간의 비용을 부담시켜 주변을 깨끗이 청소하게 하는 것이 낫다고 봅니다.
- 7 오히려 못 하게 하면 할수록 숨어서 하게 되고 다른 산이나 강을 더럽히게 됩니다.
- 8 그렇다면 생각을 달리해서 이들을 취사장으로 유인해서 삼겹살도 구워먹게 하고 술도 마시게 하되 입장료를 받고 경비와 경찰의 감시 하에 정해진 장소와 시간에만 허용하게 하는 것이 오히려 좋을 수도 있습니다.
- 9 예약제도 가능합니다.

- 10 뭐 한 가지가 맘에 안든다고 국민들의 욕구를 무조건 못 하게 해서
서는 아니됩니다.
- 11 이것이야말로 행정편의적 발상이고 독재적 발상입니다.
- 0 There are many people now drinking at the Han River.
- 1 Even now, liquids and all sorts of foods are sold at the Han River.
- 2 Also, people draw canopies and enjoy it with their family or friends.
- 3 As a result, trash appears here and there, it is dirty, and there are
some drunk people sleeping.
- 4 Ugly.
- 5 However, as Han River is the best place in the hot weather of these
days, will there be problems about drinking a little Soju with Pork
Belly which Korean people love, because of those certain people?
- 6 Rather, it would be better to allow citizens to enjoy and impose on
them small fee to clean the surroundings.
- 7 If prohibited, people will have a barbecue secretly, making other
mountains or rivers dirty.
- 8 If so, as a different idea, it would be better to attract them to
kitchens to have Pork Belly and drinks, but charge an admission fee
and allow barbecue at fixed places and times, under the supervision
of guards and police.
- 9 A subscription system is also available.
- 10 You must not restrict the desires of the public because of only one
thing you don't like.

11 This is just the idea of administrative opportunism and autocracy.

(40) Extracted Basic Units from (39)

- 0 한강에는 지금도 술 마시는 사람들 많습니다.
- 1 지금도 한강에 가면 술도 팔고 각종 음식들 팔고 있습니다.
- 2 또한 식구들이나 친구들과 함께 그늘막 치고 함께 즐기고 합니다.
- 3 그 결과 곳곳에 쓰레기가 보이고 더럽고 취해 자는 사람들도 있습니다.
- 6 차라리 시민들이 즐기게 하고 약간의 비용을 부담시켜 주변을 깨끗이 청소하게 하는 것이 낫다고 봅니다.
- 7 오히려 못 하게 하면 할수록 숨어서 하게 되고 다른 산이나 강을 더럽히게 됩니다.
- 8 그렇다면 생각을 달리해서 이들을 취사장으로 유인해서 삼겹살도 구워먹게 하고 술도 마시게 하되 입장료를 받고 경비와 경찰의 감시 하에 정해진 장소와 시간에만 허용하게 하는 것이 오히려 좋을 수도 있습니다.
- 9 예약제도 가능합니다.
- 10 뭐 한 가지가 맘에 안든다고 국민들의 욕구를 무조건 못 하게 해서 는 아니됩니다.
- 0 There are many people now drinking at the Han River.
- 1 Even now, liquids and all sorts of foods are sold at the Han River.
- 2 Also, people draw canopies and enjoy it with their family or friends.
- 3 As a result, trash appears here and there, it is dirty, and there are some drunk people sleeping.

- 6 Rather, it would be better to allow citizens to enjoy and impose on them a small fee to clean the surroundings.
- 7 If prohibited, people will have a barbecue secretly, making other mountains or rivers dirty.
- 8 If so, as a different idea, it would be better to attract them to kitchens to have Pork Belly and drinks, but charge an admission fee and allow barbecue at fixed places and times, under the supervision of guards and police.
- 9 A subscription system is also available.
- 10 You must not restrict the desires of the public because of only one thing you don't like.

Example (39) is a post supporting barbecue at the riverside of the Han River, and (40) shows the important units in the post extracted by the proposed model in this study.

This example does not have an explicit topic sentence. However, it is also possible to get the important grounds for the major claim of this post by extracting the basic units whose depth in the tree structure is 1, just the same as Examples (37) and (38) did. Basic unit (4) is the subjective opinion of the author about the situation given in (3), while (5) is the background of the idea of unit (6). In addition, (11) expresses the subjective opinion about the statement of (10). Thus, the basic units (4), (5) and (11) are not included in the summarized text (40). It is possible to grasp the major grounds provided for barbecue at the riverside of the Han River, which the author claims implicitly, with just the extracted text segments.

This experiment verified that the proposed model can extract important content from texts well. This may be a methodology of text summarization. Actually, one method of the text summarization task takes the extraction strategy, automatically picking up some important, weighted sentences from among all the sentences comprising a whole text.

6 Conclusion

This thesis suggested an automatic frame to analyze argumentation schemes of texts with subjective stances. This study proposed a new frame which identifies relationships between text segments of sentences or clauses in a document, and constructs argumentation schemes of the document based on a machine learning algorithm and a manually constructed rule.

The proposed module defined four types of discourse relations between text segments according to text cohesion and weight of importance, and automatically organized them into a systematic structure. The model first identified the relatedness between text segments by the machine learning method. The predicted results were revised by a rule-based method, followed by the subclassification procedure of discourse relations into detailed subclasses.

This study constructed the data using Korean texts which contain subjective opinions towards certain political topics. Several linguistic properties including lexical cues function as features in the provided module. They were designed to consider the author's intentions, not only the simple characteristics of a sequence of strings. The usefulness of those features and the rule was confirmed by the experiments in the study. Additionally, this study constructed a list of explicit connectives appropriate for Korean and included it in the feature set.

The argumentation structure of texts helped to extract the important reasons directly backing up the topic sentence(s). Those main grounds may be applied to various tasks in Natural Language Processing. For example, they can be one method for automatic text summarization, or an automated classification task of stances toward a certain given topic expressed in texts.

To summarize, collecting important contents and recognizing mass ideas, i.e. the authors of debating posts, are the common and ultimate goal of those studies, regardless of the detailed subtasks.

For future research, some core keywords are to be extracted from the main, important text units in the argumentation scheme, since the units are still in the form of raw text. And a proper evaluation is required to justify that the extracted basic units indicate important content in the original post well. This is a fundamental step to apply the model extracting basic units to the task of text summarization, which requires a gold standard that is manually constructed.

In addition, this study analyzed only the texts containing personal opinions about certain political issues, in a wide range of debate texts or argumentative writing. In the future, the study will gradually expand the object of analysis to various genres of texts including newspaper editorials and persuasive essays. These types of texts are also supposed to be based on expressing subjective ideas, in longer, more structured form.

Finally, the suggested model in this study might be applied to texts written in languages other than Korean, since the frame of argumentation is supposed to consist of the universal properties of languages and ways of reasoning. In the future, the study will be expanded to properly treat English texts, composing a language-independent module. Actually, Section 4 introduced a pilot study on English data, which indicates that there are still many problems to overcome. The model to deal with the internal structures of subjective documents needs to reflect some unchangeable properties of debate texts, not affected by the parametric differences between languages. Also, more implicit

cues will have to be considered and applied to improve the performance when dealing with debate texts.

Bibliography

- Adam, S. and M. Dalmas (2012). “Discourse markers in French and German: Reasons for an asymmetry”. In: *Linguistics and the Human Sciences* 6.1-3, pp. 77–98.
- Aharoni, E. et al. (2014). “A Benchmark Dataset for Automatic Detection of Claims and Evidence in the Context of Controversial Topics”. In: *Proceedings of the First Workshop on Argumentation Mining*, pp. 64–68.
- Alistair, K. and I. Diana (2005). “Sentiment Classification of Movie and Product Reviews Using Contextual Valence Shifters”. In: *Proceedings of FINEXIN*.
- Anand, P. et al. (2011). “Cats rule and dogs drool!: Classifying stance in online debate”. In: *Proceedings of the 2nd Workshop on Computational Approaches to Subjectivity and Sentiment Analysis WASSA 2011*, pp. 1–9.
- Asher, N., F. Benamara, and Y. Mathieu (2008). “Distilling opinion in discourse: A preliminary study”. In: *Coling 2008: Companion volume: Posters and Demonstrations* August, pp. 5–8.
- Biran, O. and O. Rambow (2011). “Identifying justifications in written dialogs by classifying text as argumentative”. In: *International Journal of Semantic Computing* 5.04, pp. 363–381.
- Boltuzic, F. and J. Snajder (2014). “Back up your Stance : Recognizing Arguments in Online Discussions”. In: *Proceedings of the First Workshop on Argumentation Mining*, pp. 49–58.
- Braud, C. and P. Denis (2014). “Combining Natural and Artificial Examples to Improve Implicit Discourse Relation Identification”. In: *Proceedings*

- of *COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*, pp. 1694–1705.
- Chardon, B. et al. (2013). “Measuring the Effect of Discourse Structure on Sentiment Analysis”. In: *Computational Linguistics and Intelligent Text Processing 7817*, pp. 25–37.
- Chen, D. and C. Manning (2014). “A fast and accurate dependency parser using neural networks”. In: *Proceedings of EMNLP*.
- Chenlo, J. M., A. Hogenboom, and D. E. Losada (2013). “Sentiment-Based Ranking of Blog Posts Using Rhetorical Structure Theory”. In: *Proceedings of the 18th international conference on applications of Natural Language to Information Systems (NLDB’13)*, pp. 13–24.
- Cohan, A. (2015). “Scientific Article Summarization Using Citation-Context and Article’s Discourse Structure”. In: *Emnlp2015* September, pp. 390–400.
- Cohen, J. (1960). “A coefficient of agreement for nominal scales.” In: *Educational and Psychological Measurement*.
- De Marneffe, M.-C., B. MacCartney, C. D. Manning, et al. (2006). “Generating typed dependency parses from phrase structure parses”. In: *Proceedings of LREC*. Vol. 6. 2006, pp. 449–454.
- Douglas, W., R. Christopher, and M. Fabrizio (2008). *Argumentation schemes*. Springer.
- Emeren, F. H. van and R. Grootendorst (2004). *A Systematic Theory of Argumentation*, p. 216.

- Faiz, S. I. and R. E. Mercer (2014). “Extracting Higher Order Relations From Biomedical Text”. In: *Proceedings of the First Workshop on Argumentation Mining*, pp. 100–101.
- Faulkner, A. R. (2014). “Automated Classification of Argument Stance in Student Essays: A Linguistically Motivated Approach with an Application for Supporting Argument Summarization”. PhD thesis, p. 232.
- Feng, V. W. and G. Hirst (2012). “Text-level Discourse Parsing with Rich Linguistic Features”. In: *Acl July*, pp. 60–68.
- (2014). “A Linear-Time Bottom-Up Discourse Parser with Constraints and Post-Editing”. In: *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*, pp. 511–521.
- Fraser, B. (1999). “What are discourse markers ?” In: *Journal of Pragmatics* 31.November 1996, pp. 931–952.
- (2006). “Towards a theory of discourse markers”. In: *Approaches to discourse particles* 1, pp. 189–205.
- Ghosh, D. et al. (2014). “Analyzing Argumentative Discourse Units in Online Interactions”. In: *Proceedings of the First Workshop on Argumentation Mining*, pp. 39–48.
- Govier, T. (2009). *A practical study of argument*, p. 417.
- Graves, H. et al. (2014). “Titles That Announce Argumentative Claims in Biomedical Research Articles”. In: *Proceedings of the First Workshop on Argumentation Mining*, pp. 98–99.

- Green, N. L. (2014). “Towards Creation of a Corpus for Argumentation Mining the Biomedical Genetics Research Literature”. In: *Proceedings of the First Workshop on Argumentation Mining* 1998.
- (2015). *Identifying Argumentation Schemes in Genetics Research Articles*, pp. 12–21.
- Halliday, M. A. K. and R. Hasan (1976). “Cohesion in English”. In: *RELIC Journal* 7.2, pp. 65–66.
- Hasan, K. S. and V. Ng (2012). “Predicting Stance in Ideological Debate with Rich Linguistic Knowledge”. In: *Coling2012* December 2012, pp. 451–460.
- Hatzivassiloglou, V. et al. (1997). “Predicting the semantic orientation of adjectives”. In: *Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics and Eighth Conference of the European Chapter of the Association for Computational Linguistics* pages, p. 181.
- Hernault, H., H. Prendinger, and Others (2010). “HILDA: a discourse parser using support vector machine classification”. In: *Dialogue and Discourse* 1.3, pp. 1–33.
- Houngbo, H. and R. E. Mercer (2014). “An automated method to build a corpus of rhetorically-classified sentences in biomedical texts”. In: *Proceedings of the First Workshop on Argumentation Mining*, pp. 19–23.
- Huang, H.-H. et al. (2014). “Interpretation of Chinese Discourse Connectives for Explicit Discourse Relation Recognition”. In: *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*, pp. 632–643.

- Hyland, K. (2010). “Metadiscourse: Mapping Interactions in Academic Writing”. In: *Nordic Journal of English Studies* 9.2, pp. 125–143.
- Jain, T. I. and D. Nemade (2010). “Recognizing Contextual Polarity in Phrase-Level Sentiment Analysis”. In: *International Journal of Computer Applications* 7.5, pp. 12–21.
- Jang, H., M. Kim, and H. Shin (2013). “KOSAC: A Full-fledged Korean Sentiment Analysis Corpus”. In: *Proceedings of the 27th Pacific Asia Conference on Language, Information and Computation*.
- Jang, H. and H. Shin (2010). “Effective Use of Linguistic Features for Sentiment Analysis of Korean”. In: *Pacific Asia Conference on Language, Information and Computation*, pp. 173–182.
- Kiesel, J. et al. (2015). “A Shared Task on Argumentation Mining in Newspaper Editorials”. In: *Proceedings of the 2nd Workshop on Argumentation Mining*. Association for Computational Linguistics, pp. 35–38.
- Kikuchi, Y. et al. (2014). “Single Document Summarization based on Nested Tree Structure EDU sentence subtree sentence selection selection”. In: *Acl* 2, pp. 315–320.
- Kirschner, C., J. Ecker-Köhler, and I. Gurevych (2015). “Linking the Thoughts : Analysis of Argumentation Structures in Scientific Publications”. In: *Proceedings of the 2nd Workshop on Argumentation Mining*, pp. 1–11.
- Klein, D. and C. D. Manning (2003). “Accurate unlexicalized parsing”. In: *Proceedings of the 41st Annual Meeting on Association for Computational Linguistics-Volume 1*. Association for Computational Linguistics, pp. 423–430.

- L., E. P. and S. Cho (2014). “KoNLPy : Korean natural language processing in Python”. In: *Proceedings of the 26th Annual Conference on Human and Cognitive Language Technology*, pp. 4–7.
- Larson, M. (2012). “Automatic Summarization”. In: *Foundations and Trends® in Information Retrieval* 5.3, pp. 235–422.
- Lawrence, J. and C. Reed (2015). “Combining Argument Mining Techniques”. In: *Proceedings of the Second Workshop on Argumentation Mining*, pp. 127–136.
- Lawrence, J. et al. (2014). “Mining arguments from 19th century philosophical texts using topic based modelling”. In: *Acl 2014*, p. 79.
- Lee, C., J. Kim, and J. Kim (2014). “Korean Dependency Parsing using Deep Learning”. In: *Proceedings of the 26th Annual Conference on Human and Cognitive Language Technology*.
- Lin, W.-H. and A. Hauptmann (2006). “Are these documents written from different perspectives?” In: *Proceedings of the 21st International Conference on Computational Linguistics and the 44th annual meeting of the ACL - ACL '06 2*, pp. 1057–1064.
- Lin, Z., M.-y. Kan, and H. T. Ng (2009). “Recognizing Implicit Discourse Relations in the Penn Discourse Treebank”. In: *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing*, pp. 343–351.
- Lin, Z., H. T. Ng, and M.-Y. Kan (2014). “A PDTB-styled end-to-end discourse parser”. In: *Natural Language Engineering* 20.02, pp. 151–184.

- Litvak, M. et al. (2015). “Krimping Texts for Better Summarization”. In: *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing* September, pp. 1931–1935.
- Liu, P., S. Joty, and H. Meng (2014). “Fine-grained Opinion Mining with Recurrent Neural Networks and Word Embeddings”. In: *Emnlp*.
- Llewellyn, C. (2012). “Using Argument Analysis to Define a Structure for User Generated Content”. In: *Joint Conference on Digital Libraries 2012*.
- Mann, W. C. and S. A. Thompson (1988). “Rhetorical structure theory: Toward a functional theory of text organization”. In: *Text-Interdisciplinary Journal for the Study of Discourse* 8.3, pp. 243–281.
- Marcu, D. (1999). “Discourse trees are good indicators of importance in text”. In: *Advances in Automatic Text Summarization*, pp. 123–136.
- (2000). *The Theory and Practice of Discourse Parsing and Summarization*. MIT Press, pp. 81–83.
- Marcu, D., A. Echihabi, and M. Rey (2001). “An unsupervised approach to recognizing discourse relations”. In: *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics - ACL '02* July, p. 368.
- Misra, A. et al. (2015). “Using Summarization to Discover Argument Facets in Online Ideological Dialog”. In: *Naacl2015*, pp. 430–440.
- Morsy, S. and A. Rafea (2012). “Improving document-level sentiment classification using contextual valence shifters”. In: *Natural Language Processing and Information . . .* Pp. 253–258.

- Mukherjee, S. and P. Bhattacharyya (2012). “Sentiment Analysis in Twitter with Lightweight Discourse Analysis.” In: *Coling* December 2012, pp. 1847–1864.
- Murakami, A. and R. Raymond (2010). “Support or Oppose ? Classifying Positions in Online Debates from Reply Activities and Opinion Expressions”. In: *Coling* August, pp. 869–875.
- Nguyen, H. V. and D. J. Litman. “Extracting argument and domain words for identifying argument components in texts”. In: *Proceedings of the 2nd Workshop on Argumentation Mining* June, pp. 22–28.
- Ong, N., D. Litman, and A. Brusilovsky (2014). “Ontology-Based Argument Mining and Automatic Essay Scoring”. In: *Proceedings of the First Workshop on Argumentation Mining*, pp. 24–28.
- Palau, R. M. and M. F. Moens (2009). “Argumentation mining: The Detection, Classification and Structuring of Arguments in Text”. In: *Belgian/Netherlands Artificial Intelligence Conference*, pp. 351–352.
- Park, J. and C. Cardie (2012). “Improving Implicit Discourse Relation Recognition Through Feature Set Optimization”. In: *Proceedings of the 13th Annual Meeting of the Special Interest Group on Discourse and Dialogue* July, pp. 108–112.
- (2014). “Identifying Appropriate Support for Propositions in Online User Comments”. In: *Proceedings of the First Workshop on Argumentation Mining*, pp. 29–38.

- Peldszus, A. (2014). “Towards segment-based recognition of argumentation structure in short texts”. In: *Proceedings of the First Workshop on Argumentation Mining*, pp. 88–97.
- Peldszus, A. and M. Stede (2013). “From Argument Diagrams to Argumentation Mining in Texts:” in: *International Journal of Cognitive Informatics and Natural Intelligence* 7.1, pp. 1–31.
- (2015). “Joint prediction in MST-style discourse parsing for argumentation mining”. In: *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pp. 938–948.
- Polanyi, L. and M. V. D. Berg (2003). “Discourse Structure and Sentential Information Structure”. In: *Journal of Logic, Language and Information* 12, pp. 337–350.
- Potter, A (2008). “Linked and Convergent Structures in Discourse-Based Reasoning”. In: *Proceedings of the 3rd International Explanation Aware Computing Workshop (ExaCt 2008), Patras, Greece*, pp. 72–83.
- Potter, A. (2007). “A discourse approach to explanation aware knowledge representation”. In: *Explanationaware computing Papers from the 2007* 1958, pp. 56–63.
- Prakken, H. (2010). “An abstract framework for argumentation with structured arguments”. In: *Argument & Computation* 1.2, pp. 93–124.
- Prasad, R. et al. (2007). *The Penn Discourse Treebank 2.0 Annotation Manual*.
- Ranade, S., R. Sangal, and R. Mamidi (2013). “Stance Classification in Online Debates by Recognizing Users ’ Intentions”. In: *Proceedings of the SIG-DIAL 2013 Conference August*, pp. 61–69.

- Reisert, P. et al. (2015). “A Computational Approach for Generating Toulmin Model Argumentation”. In: *Proceedings of the 2nd Workshop on Argumentation Mining*, pp. 45–55.
- Rinott, R. et al. (2015). “Show Me Your Evidence – an Automatic Method for Context Dependent Evidence Detection”. In: *Proceedings of the 2015 Conference on Empirical Methods in NLP (EMNLP)* September, pp. 440–450.
- Rosé, C. et al. (2008). “Analyzing collaborative learning processes automatically: Exploiting the advances of computational linguistics in computer-supported collaborative learning”. In: *An Official Publication of the International Society of the Learning Sciences* 3.3, pp. 237–271.
- Saito, M., K. Yamamoto, and S. Sekine (2006). “Using phrasal patterns to identify discourse relations”. In: *Proceedings of the Human Language Technology Conference of the NAACL, Companion Volume: Short Papers*. Association for Computational Linguistics, pp. 133–136.
- Sardianos, C. et al. (2015). “Argument Extraction from News”. In: *Proceedings of the 2nd Workshop on Argumentation Mining* June.
- Schiffrin (2001). *Discourse Markers: Language, Meaning, and Context*. Vol. 80. 4.
- Schneider, J. and I. S. Antipolis (2014). “Automated argumentation mining to the rescue? Envisioning argumentation and decision-making support for debates in open online collaboration communities”. In: *Proceedings of the First Workshop on Argumentation Mining* c, pp. 59–63.

- Socher, R. et al. (2013). “Parsing with compositional vector grammars”. In: *ACL 2013 - 51st Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference 1*, pp. 455–465.
- Somasundaran, S. (2010). “Discourse-Level Relations for Opinion Analysis”. PhD thesis, p. 237.
- Somasundaran, S., J. Ruppenhofer, and J. Wiebe (2007). “Detecting Arguing and Sentiment in Meetings”. In: *Proceedings of the 8th SIGdial Workshop on Discourse and Dialogue*, pp. 26–34.
- Somasundaran, S. and J. Wiebe (2010). “Recognizing stances in ideological on-line debates”. In: *Proceedings of the NAACL HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text*. Association for Computational Linguistics, pp. 116–124.
- Somasundaran, S. et al. (2009). “Supervised and Unsupervised Methods in Employing Discourse Relations for Improving Opinion Polarity Classification”. In: *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing* August, pp. 170–179.
- Song, Y. et al. (2014). “Applying Argumentation Schemes for Essay Scoring”. In: *Proceedings of the First Workshop on Argumentation Mining 2011*, pp. 69–78.
- Sporleder, C. and A. Lascarides (2008). “Using automatically labelled examples to classify rhetorical relations: an assessment”. In: *Natural Language Engineering* 14.03.
- Stab, C. and I. Gurevych (2014). “Identifying Argumentative Discourse Structures in Persuasive Essays”. In: *Proceedings of the 2014 Conference on*

- Empirical Methods in Natural Language Processing (EMNLP)*, pp. 46–56.
- Taboada, M. and G.-G. M. Angeles (2012). “Discourse markers and coherence relations : Comparison across markers , languages and modalities”. In: *Linguistics and the Human Sciences* 6, pp. 17–41.
- Taboada, M., K. Voll, and J. Brooke (2008). “Extracting sentiment as a function of discourse structure and topicality”. In: *Technical Report* 20, pp. 1–22.
- Toulmin, S. (1958). *The uses of argument*. Cambridge University Press.
- Trevisan, B. et al. (2014). “Indicators of Argument-conclusion Relationships. An Approach for Argumentation Mining in German Discourses”. In: *Proceedings of the First Workshop on Argumentation Mining*, pp. 104–105.
- Turney, P. D. (2002). “Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews”. In: *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL)* July, pp. 417–424.
- Vanzo, A., D. Croce, and R. Basili (2014). “A context-based model for Sentiment Analysis in Twitter”. In: *Conference on Computational Linguistics (COLING)*, pp. 2345–2354.
- Walker, M. A. et al. (2012). “That is your evidence?: Classifying stance in online political debate”. In: *Decision Support Systems* 53.4, pp. 719–729.
- Walker, V. R., K. Vazirova, and C. Sanford (2014). “Annotating Patterns of Reasoning about Medical Theories of Causation in Vaccine Cases: Toward

- a Type System for Arguments”. In: *Proceedings of the First Workshop on Argumentation Mining 2005*, pp. 1–10.
- Walton, D. and C. Reed (2002). “Argumentation schemes and defeasible inferences”. In: *Working Notes of the ECAI’2002 Workshop on Computational Models of Natural Argument*, pp. 45–55.
- Wellner, B. and J. Pustejovsky (2007). “Automatically Identifying the Arguments of Discourse Connectives”. In: *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL)*, pp. 92–101.
- Williams, S. and R. Power (2008). “Deriving Rhetorical Complexity Data from the RST-DT Corpus”. In: *Proceedings of the Sixth International Language Resources and Evaluation (LREC’08)*.
- Wilson, T., J. Wiebe, and P. Hoffmann (2005). “Recognizing contextual polarity in phrase-level sentiment analysis”. In: *Proceedings of the conference on human language technology and empirical methods in natural language processing*. Association for Computational Linguistics, pp. 347–354.
- Wolf, F. and E. Gibson (2005). “Representing Discourse Coherence: A Corpus-Based Study”. In: *Computational Linguistics* 31.2, pp. 249–287.
- Wyner, A. and T. V. Engers (2010). “Web-based Mass Argumentation in Natural Language”. In: *International Conference on Knowledge Engineering and Knowledge Management 2010*, pp. 3–4.
- Yessenalina, A., Y. Yue, and C. Cardie (2010). “Multi-level structured models for document-level sentiment classification”. In: *Proceedings of the*

2010 Conference on Empirical Methods in Natural Language Processing, pp. 1046–1056.

Zhang, Q. et al. (2013). “Discourse Level Explanatory Relation Extraction from Product Reviews Using First-Order Logic”. In: *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, pp. 946–957.

Zhou, Z. M. et al. (2010). “Predicting discourse connectives for implicit discourse relation recognition”. In: *Proceedings of the 23rd International Conference on Computational Linguistics: Posters August*, pp. 1507–1514.

초록

한국어 텍스트 논증 구조의 자동 분석 연구

최근 온라인 텍스트 자료를 이용하여 대중의 의견을 분석하는 작업이 활발히 이루어지고 있다. 이러한 작업에는 주관적 방향성을 갖는 텍스트의 논증 구조와 중요 내용을 파악하는 과정이 필요하며, 자료의 양과 다양성이 급격히 증가하면서 그 과정의 자동화가 불가피해지고 있다.

본 연구에서는 정책에 대한 찬반 의견으로 구성된 한국어 텍스트 자료를 직접 구축하고, 글을 구성하는 기본 단위들 사이의 담화 관계의 유형을 정의하였다. 하나의 맥락 안에서 두 개의 문장 혹은 절이 서로 관계를 갖는지, 관계를 갖는다면 서로 동등한 관계인지, 그렇지 않은 경우 어느 문장(절)이 더 중요한 부분으로서 다른 하나의 지지를 받는지의 기준에 따라 담화 관계를 두 개의 층위로 나누어 이용하였다.

이러한 기본 단위들 사이의 관계는 기계 학습과 규칙 기반 방식을 이용하여 예측된다. 이 때 각 글의 저자가 표현하고자 하는 의도, 자신의 주장을 뒷받침하기 위해 제시하는 근거의 종류, 그리고 그 근거를 이루는 논증 전략 등이 텍스트의 언어적 특징과 함께 중요한 자질로 작용된다. 논증의 전략으로는 예시, 인과, 세부 사항에 대한 설명, 반복 서술, 정정, 배경 지식 제공 등이 관찰되었다. 이들 세부 분류는 담화 관계의 대분류를 구성하고, 그 담화 관계를 예측하는 데 쓰이는 자질의 기반이 되었다.

또한 일부 언어적 자질들은 기존 연구를 참고하여 한국어 자료에 적용할 수 있는 형태로 재구성하였다. 이를 이용하여 한국어 코퍼스를 구축하고 한국어 연구에 특화된 접속사 및 연결어의 목록을 구성하여 자질 목록에 포함시

켰다. 이러한 자질들에 기반해서 담화 관계를 예측하는 과정을 이 연구에서 독자적인 모델로서 자동화하여 제안하였다.

예측 실험의 결과를 보면 본 연구에서 정의하여 이용한 자질들은 긍정적인 상호 작용을 통해 담화 관계 예측의 성능을 향상시킨다는 것을 알 수 있었다. 그 중에서도 일부 접속사 및 연결어, 문장 성분의 유무에 따른 의존적인 문장 구조, 그리고 같은 내용을 반복 서술하는지의 여부 등이 특히 예측에 기여하였다.

텍스트를 이루는 기본 단위들 사이에 존재하는 담화 관계들은 서로 연결, 합성되어 텍스트 전체에 대응되는 트리 형태의 논증 구조를 이룬다. 이렇게 얻은 논증 구조에 대해서는, 트리의 가장 위쪽인 루트 노드에 글의 주제문이 위치하고, 그 바로 아래 층위에 해당하는 문장(절)들이 근거로서 가장 중요한 내용을 담고 있다고 가정할 수 있다. 따라서 주제문을 직접적으로 뒷받침하는 문장(절)을 추출하면 글의 중요 내용을 얻게 된다. 이는 곧 텍스트 요약 작업에서 유용하게 쓰이는 방식이 될 수 있다. 또한 주제에 따른 입장 분류나 근거 수집 등 다양한 분야에서도 응용이 가능할 것이다.

주요어: 논증 마이닝, 논증 구조, 담화 관계, 논증 전략, 한국어 텍스트 분석

학번: 2013-20032