

Analysis of Thematic Structure in L2 Writing: A Systemic Functional Perspective

Kwanghyun Park

Myongji University

Daehyeon Nam*

Ulsan National Institute of Science and Technology

Abstract

In academic writing, it is essential that the text is organized to convey information through the logical development of ideas. The key element of the logical structure is *Theme*, a textual unit that condenses known information and establishes grounds for further development. Although theme has long been researched in text and register analysis, it is an emerging area of research in L2 writing—mainly due to the lack of availability of large L2 corpora and a computational means. Using an automatic software program for analyzing thematic structure, this study investigates themes in essays written by Korean students in contrast to themes in data taken from the Michigan Corpus of Upper-level Student Papers (MICUSP), a 1.6 million-word corpus comprised of advanced academic papers. The themes in 138 essays written by Korean students were analyzed focusing on two learner variables: the academic discipline (humanities vs. science), and L2 proficiency. The results show a moderate difference in analysis of academic discipline, but highly significant differences in analysis of proficiency. The follow-up qualitative analysis found the limited repertoire of thematic choices in the essays written by Korean students. Based on this finding, the possible use of the theme in L2 writing research is discussed.

Keywords: L2 writing, Corpus, Theme, thematic structure, method of development

* Corresponding author (dnam@unist.ac.kr)

I. Introduction

In academic writing at a tertiary level, the logical development of text is paramount because it serves as a vehicle to present information in a coherent manner. Common academic writing tasks can come in the form of various types of essays including process, narrative, descriptive, and argumentative. Regardless of the genre, students must organize their writing so that information is condensed into a logical and coherent progression. However, despite the emphasis on the logical progression, it remains a daunting challenge for L2 students to accomplish a high degree of logical development in the text. Typically, an advanced writer organizes the text so it starts from a less complex idea that is shared with the reader, then develops it to an increasingly complex one that has the potential to shape into a new idea. The effective structure for delivery of information relates to how incremental progression is achieved by what is presented in the 'given-new' information structure (Halliday & Hasan, 1976; Martin, 1992; Schleppegrell, 2004; Ventola & Mauranen, 1996).

In L2 writing research, promoting logical development has long been an area of interest (Johns, 1986; Lovejoy, 1991; Schleppegrell, 2001; Silva, 1993). Genre-based approaches to writing have been particularly influential because they tend to linguistically define the elements of the development (Halliday, 1994; Hyland, 2003; Johns, 2002; Martin, 1992; Swales, 1990). Although genre-based approaches differ in terms of their theoretical orientation, they share the key characteristics that emphasize a language-focused perspective, and form-function dynamics that links the author's communicative purpose and textual convention. Pedagogical approaches to L2 writing that are genre-based and language-focused are powerful strategies because they can show clearly the linguistic resources (e.g. vocabulary, grammar, and rhetorical style) that contribute to the logical development of the text. This allows the writer to deliver information accurately and efficiently.

Using Systemic Functional Linguistics (SFL) as a theoretical framework, this paper investigates how L2 writers present information in their writing. SFL is a genre-based perspective that catalogues

information structure into two parts: *Theme* and *Rheme*. The theme is the initial position in the sentence with a new idea or requisite information. The rheme is the other half of the sentence that contains new idea or piece of information. Because it draws upon background or established knowledge that is either known to the reader or previously mentioned in the passage, the theme is critical to the logical progression of the textual structure.

Given the central role in creating logical connections in the text, theme has strong educational implications for teaching writing. Korean L2 learners, for instance, can be taught explicitly how to construct an effective progression of information between sentences. However, relatively little research has been done on the theme in the writing of Korean students leaving the thematic progression a largely uncharted area. This paper reports an exploratory analysis of thematic structure in their English essays as groundwork for understanding their use of theme. Specifically, this paper explores thematic structure in relation to L2 proficiency, and by comparing the writing of students majoring in two academic disciplines: humanities and science. Accordingly, two research questions have been formulated to guide this study:

- (1) Is there a significant difference in how students majoring in the humanities use theme compared to those majoring in the sciences?
- (2) Does L2 proficiency significantly affect how students use theme?

In the following sections, this paper reports the thematic variation in relation with these two variables in student essays collected from six Korean universities and a corpus of essays collected from a large American university.

II. Literature review

A. Definition of Theme in SFL

SFL is a theory of language as a system of meaning that focuses

on the function of the language: how it is used to create and convey meaning with linguistic resources (for a complete review, see Halliday, 1994). In SFL, theme is a key resource for textual cohesion and coherence that underlies the logical development of ideas and information. Structurally, the theme is simply the beginning part of a sentence. Semantically, the theme represents a conceptual background or requisite information, the knowledge that is shared by both the writer and reader (Halliday, 1994; Halliday & Matthiessen, 2004; Thompson, 2004). Using the theme, the writer prepares to introduce new information in the other half of the clause or sentence known as the rheme. The essential role that the theme plays in creating a logical and coherent progression is reflected in what Halliday and Matthiessen (2004) define as “the point of departure” (p. 64), a cognitive origin of textual progression.

The thematic structure, sometimes referred as the theme-rheme structure, contributes to the logical progression of information significantly. To illustrate, the following examples are taken from Halliday and Matthiessen (2004, p. 66):

	Theme	Rheme
(1)	<i>The duke</i>	has given my aunt that teapot.
(2)	<i>My aunt</i>	has been given that teapot by the duke.
(3)	<i>That teapot</i>	the duke has given to my aunt.

In these examples *The duke*, *My aunt*, and *That teapot* are pieces of information understood to be old information because they are mutually known between the writer and the reader. The rheme part of the sentence is the new idea that the reader does not yet know: hence new information. As such, the theme plays an important role in creating a logical transition from one sentence to the next. Working within the same sentence structure, the following examples have been adapted from Halliday and Matthiessen (2004):

- (4) I have an honor to serve the duke of Wales. He gave my aunt that teapot.
- (5) Yesterday, at the market, the duke met my aunt. She, then, was given that teapot by him.

- (6) Do you see the teapot on the cupboard? That teapot, the duke gave to my aunt.

In the examples above, the rheme of the first sentence connects with the theme of the following sentence. In connecting the sentences, the initial position of the theme orients the reader's cognitive reception of information in the surrounding rhemes. In SFL, this capacity to connect sentences into a coherent whole is a key element of development, and is referred to as the *textual function* (Halliday, 1994).

B. Theme in writing research

Theme has been a focal area of research on textual cohesion/coherence and information structure (Halliday, 1994; Halliday & Hasan, 1976; Halliday & Matthiessen, 2004). A number of studies examined the theme in an L1-L2 comparison (Hirose, 2003; Kubota, 1998; Matsuda, 1997; McCabe, 1999; Péry-Woodley, 1991), in feature-based analysis of L2 texts (Chiang, 1999; Granger & Tyson, 1996; Ventola, 1992, 1994), and more recently by using computational techniques (Crossley & McNamara, 2011, 2012; Graesser, McNamara, & Kulikowich, 2011). Though a small number of studies have examined the writing of L2 learners to identify the characteristics of the thematic structure (Belmonte & McCabe, 1998; Hannay, 2007; Herriman, 2011; Mauraanen, 1996), few studies have investigated thematic structure in L2 writing both as a descriptive feature and an index of development.

Among the few studies on the L1-L2 comparison of the themes, Herriman (2011) compared a collection of essays written by highly proficient Swedish students to essays written by British students. Her findings reported the themes used by Swedish students had more characteristics of conversational interaction: e.g. first and second person pronouns and emotional vocabulary. As such, their use of theme was rendered as a less expository text. She also found little thematic variation in their writing compared to the use of themes by the British students. Though her study raises an interesting question on the

thematic variation, her study was limited to only one type of theme, the topical theme. As a result, the other two themes—the interpersonal theme and the textual theme—went unreported. As Halliday (1994) demonstrated in his seminal work, all three themes are essential in SFL. In a Korean context, Lee (2008) reported a similar finding. By comparing 20 essays written by Korean students to 15 essays written by Australian students, the study found the themes in Korean students' essays tended to be more personal and interactional.

In investigating the role of proficiency on the thematic structure, Belmonte and McCabe (1998) compared the English essays written by 25 Spanish speaking students to similar kinds of text written by professional writers. They reported a number of problems in analysis of thematic structure in the essays written by the students: (1) rhetorical/coherent issues in the selection of the inappropriate theme and an over-use of a particular item to start a sentence (e.g. 'there'); (2) use of too many themes without presenting ideas from preceding themes; (3) lexico-grammatical issues involving sentences that had too many themes, or sentences that did not have any rheme. All of these observations confuse the reader with either grammatically incorrect sentences or incoherent use of theme, or sometimes both. These findings, however, contrast with Witte and Faigley (1981). Their study compared low-rated to high-rated essays written by American students and found the difference pertained mainly to issues in coherence at the rhetorical level, rather than at lexico-grammatical level.

Another pertinent variable in the analysis of thematic structure is a student's academic discipline. This approach was adopted by North's (2005) study of thematic structure comparing the writing from British students majoring in the sciences to those majoring in the humanities. She found a significant difference in the thematic structure between the two groups reporting that the humanities students used more orienting themes, a type of theme that incorporates two or more themes, than science students. She argued that the difference can be attributed to the tendency of students in the humanities to present information as constructed, rather than as factual as is preferred by students majoring in the sciences.

Although these studies made a significant contribution to our

understanding of the thematic structure, they are not without limitations, which is not surprising in the sense that, as this thematic structure in the analysis of L2 writers is just an emerging avenue of research. Aside from Herriman (2011), findings from these studies are limited by the rather small amount of data with the population size between only two and twenty-five students and none of these studies have examined all three functions of the theme: ideational, interpersonal, and textual. Furthermore, these studies do not adequately detail the methods employed on how the actual analysis of the theme was conducted, making it difficult, if not impossible, to conduct a replicate study in a Korean context. In response to these gaps in the literature, this paper begins by exploring the theme use of Korean learners based on two collections of essays: (1) essays written by Korean students totaling approximately 400,000 words, and (2) a 1.6 million-word corpus of advanced students' papers collected in an American university.

III. Method

A. The data

Three kinds of data were used in the present study: (1) Korean students essays (science majors), (2) Korean students essays (humanities majors), and (3) student papers that received an 'A' collected from an American university. Data set 1 was a collection of 138 essays (33,947 words) written in English by Korean students majoring in science and technology (KR-science). These essays were collected over one semester in an academic writing course offered at a large national university. Data set 2 consisted of 358 essays written by humanities students (200,469 words) were used as (KR-humanities), which were collected from six universities in Korea. For the actual theme analysis, subsets of 2,000 sentences were randomly extracted from both data sets totaling of 32,751 words (KR-science) and 31,809 words (KR-humanities). Table 1 summarizes the Korean students' data.

Table 1. The Korean Students' Essays

	Essays by science majors	Essays by humanities majors
Data set handle	KR-science	KR-humanities
Total number of words	33,947	200,469
Total number of essays	138	358
Sentences analyzed	2,000	2,000
Words analyzed	32,751	31,809
Discipline	Science and technology (e.g. natural sciences, nuclear science, life science, and mechanical engineering,)	English language and literature, and English related majors (e.g. English education, English translation)

Data set 3 was taken from the Michigan Corpus of Upper-level Student Papers (MICUSP), a collection of 1.6 million-word advanced academic papers written by students attending University of Michigan (Römer & Wulff, 2010). The MICUSP was the high-proficiency English corpus to be compared with the L2 English essays by Korean students.

Table 2. Essays Collected in Korean Universities and an American University

	Essays from Korean universities	Essays from an American university
Data set handle	KR-essays	US-essays
Total words	64,560	563,692 (argumentative papers only)
Total number of essays	496	187
Sentence analyzed	4,000	2,000
Words analyzed	64,560	35,008
Data Source	Argumentative essays written by Korean university students majoring in both science and humanities	A-grade argumentative papers written by highly proficient students at an American University

Out of 1.6 million words, only argumentative papers were extracted totaling 563,692 words. For the actual theme analysis, a subset of 2,000 sentences was randomly extracted from both data sets totaling of 64,560 words (KR-essays) and 35,008 words (US-essays). Table 2 presents a comparison of the essays written by Korean students to the MICUSP.

B. Analytical procedure

In the first step of analysis, four key features of the theme were identified based on the key research literature in SFL (Eggins, 2004; Halliday & Matthiessen, 2004; Thompson, 2004). Then, these features were formally defined for data coding. Table 3 lists the key features and provides descriptions and examples.

Table 3. Four Key Features of Theme

Feature name	Category	Definition and example
Theme length	A number	The number of words in the theme
Theme function	Textual	The theme links to the previous sentence (e.g. <i>and, but, however</i>)
	Interpersonal	The theme expresses the speaker's attitude (e.g. <i>honestly, surprisingly, naturally</i>)
	Topical (ideational)	The theme expresses the contents (e.g. <i>The time</i> [flies], <i>We</i> [laughed], <i>Rain</i> [stopped])
Theme plurality	Single	A single theme has only one function, that is, topical (ideational)
	Multiple	A multiple theme has multiple functions: textual + interpersonal functions (e.g. <i>But, honestly</i>)
Theme role	Subject	<i>We</i> saw that one in the market.
	Object	<i>That one</i> , we saw in the market.
	Adjunct	<i>In the market</i> , we saw that one.

The theme is *italicized*.

In the next step of analysis, two Korean data sets (KR-science set

and the KR-humanities set) were examined for the four thematic features. Then, in the third step, Korean essays were compared to essays written by students attending an American university—these students chosen are highly proficient; their papers have received high marks (grade A) or recommendation from their instructors.

C. Software for automatic theme analysis

In this study, the theme features were analyzed in 800 sentences totaling approximately 130,000 words. Without a computerized tool, an analysis of this size would be extremely time consuming and difficult. The software program used in this study is the theme analyzer first introduced in Park and Lu (2015). It is a program written in the Python programming language and implemented for command line on Unix-like environment including Linux and OSX. There are two steps to obtain theme information from a sentence using this program. Step 1 involves part-of-speech tagging and syntactic parsing, the procedure to break down the sentence into grammatical constituent. Figure 1 shows an example of a parsed sentence using a parser that is available for free at <http://stanfordnlp.github.io/CoreNLP>.

```

Sea turtles are not so many to not care of them.

Tagging

Sea/NNP turtles/NNS are/VBP not/RB so/RB
many/JJ to/TO not/RB care/NN of/IN them/PRP
./

Parse

(ROOT
 (S
  (NP (NNP Sea) (NNS turtles))
  (VP (VBP are) (RB not)
    (ADJP (RB so) (JJ many)
      (PP (TO to)
        (NP
          (NP (RB not) (NN care))
          (PP (IN of)
            (NP (PRP them)))))))
  (. .)))

```

Figure 1. A parsed sentence

In Step 2, the parsed sentence is fed into the theme analyzer. This returns the results of analysis as shown in Figure 2.

```
Sent 1: Sea turtles are not so many to not care of them.
THEME: Sea turtles
length: 2 (words)
theme_func: topical
theme_role: Subject
theme_plurality: single
```

Figure 2. Output of the automatic theme analyzer

When processed, the output of the analyzer can be saved and used for further statistical processing. In this study, SPSS was the statistical package used to interpret the data.

IV. Results and Discussion

A. Theme length

A one-way ANOVA test was conducted to examine the theme lengths in the data. The results in Table 4 show a statistically significant difference among the three groups ($F=4,015.706$ (2, 5,327), $p=.000$).

Table 4. ANOVA Test of Theme Length: KR-humanities x KR-science x US-essays

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	511,923.7	2	255,961.858	4,015.706	.000
Within Groups	339,544.0	5,327	63.740		
Total	851,467.7	5,329			

* $p < .05$

To locate which groups contribute to the main effects of the significant difference, comparisons among means were tested using Tuckey HST. The results in Table 5 show that there are statistically significant differences in length of theme between essays written by both groups of Korean students (KR-science and KR-humanities) to the

data taken from the MICUSP (US-essays). No statistically significant difference, however, was found due to discipline in the comparison between KR-science and KR-humanities.

Table 5. Post Hoc Tests of Theme Length

Group (I)	Group (J)	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
KR-humanities	KR-science	.296	.252	.471	-.3	.89
	US-essays	-22.479*	.282	.000	-23.16	-21.84
KR-science	KR-humanities	-.296	.252	.471	-.89	.30
	US-essays	-.296	.282	.000	-23.46	-22.16
US-essays	KR-humanities	22.479*	.282	.000	21.84	23.16
	KR-science	22.793*	.282	.000	22.13	23.46

Note. Tuckey HSD. Dependent variable: length. * $p < .05$

The results suggest that Korean students, regardless of their majors, use relatively shorter themes in their writing, compared to the students at the American university. This finding may come as a surprise to some writing teachers who believe that humanities students write longer, and arguably better as observed in North (2005). The analysis of the theme length offers a counter-argument against such belief. Table 6 shows the examples of long and short themes in Korean essays compared to American essays.

Table 6. Long and Short Themes in Korean and American Essays

Long Themes	Theme	Rheme
Korean (Humanities)	In an effort to lessen major crimes such as rape, robbery and homicide etc,	the government has purposed many policies to prevent us from being exposed in the offenders.
Korean (Science)	But the reason I do not agree with the statement completely	is that the kind of tasks is varying with time.
American	The ability of bio-invaders to hybridize with native species to produce a spectrum of pseudo-exotics	exacerbates difficulties faced.
Short Themes		
Korean (Humanities)	The answer	will be various but most university students want to be hired in big companies.
Korean (Science)	Sea turtles	are not so many to not care of them.
American	We	launch invasive removal projects to “save the natives”.
	This	is most common among plants, which can also undergo extremely rapid evolution.

Although these examples by no means represent the entire data, they do raise a hypothetical point to consider. While both short and long themes appear in Korean and American essays, the impression is that Korean students tend to use the long theme as an adverbial (e.g. ‘In an effort to . . .’) and clausal (e.g. ‘But the reason . . .’), while the longer themes in the American data appear to be nominalized and phrasal, which are typical features of information-intense texts (Schleppegrell, 2001). It is not clear at this point, however, if this impression can be supported by statistical evidence, and thus the hypothesis remains to be tested.

B. Theme function and plurality

The analysis of the theme function found no significant difference in comparing the Korean students majoring in the humanities to those majoring in the sciences. However, there are significant differences between Korean and American data. Table 7 presents the results.

Table 7. Theme Function: KR-humanities x KR-science x US-essays

		Interpersonal	Textual	Topical	N/A
KR-humanities	Count	33 (32.97)	470 (467.17)	1442 (1439.05)	16 (21.81)
	Std. Res.	0.01	0.13	0.08	-1.24
KR-science	Count	32 (32.39)	451 (453.83)	1395 (1397.95)	27 (21.19)
	Std. Res.	-0.01	-0.13	-0.08	1.26
KR-humanities	Count	33 (35.75)	470 (345.01)	1442 (1570.71)	16 (9.53)
	Std. Res.	-0.46	6.73	-3.25	2.09
US-essays	Count	27 (24.25)	109 (233.99)	1194 (1065.29)	0 (6.47)
	Std. Res.	0.56	-8.17	3.94	-2.54
KR-science	Count	32 (34.74)	451 (329.77)	1395 (1524.59)	27 (11.10)
	Std. Res.	-0.47	6.68	-3.32	2.78
US-essays	Count	27 (24.26)	109 (230.23)	1194 (1064.59)	0 (11.10)
	Std. Res.	0.56	-7.99	3.97	-3.33

Note. N/A = Incomplete clauses for theme analysis. Numbers in the parenthesis denote expected counts. Humanities Set x KR-science: $\chi^2=3.189$; $df=3$; $p=.363$. Humanities Set x US-essays: $\chi^2=149.524$; $df=3$; $p=.000$. KR-science x US-essays: $\chi^2=154.574$; $df=3$; $p=.000$.

Regarding the major contribution of the differences, the standardized residuals among the groups are related with the textual function and the topical functions. The textual theme accounts for 6.73 and -8.17 between KR-humanities and US-essays, and 6.68 and -7.99 between KR-science and US-essays. The topical theme accounts for -

3.25 and 3.94 between KR-humanities and US-essays, and -3.32 and 3.97 between KR-science and US-essays.

Similar to the results of the length and function, there is no significant difference in the theme plurality due to discipline: KR-humanities vs. KR-science majors. However, the chi-square test confirms a significant difference between KR-essays and US-essays. Table 8 presents the results.

Table 8. Theme Plurality: KR-humanities x KR-science x US-essays

		Multiple	Single	N/A
KR-humanities	Count	503 (500.14)	1442 (1439.05)	16 (21.81)
	Std. Res.	0.13	-0.08	-1.24
KR-science	Count	483 (485.86)	1395 (1397.95)	27 (21.19)
	Std. Res.	-0.13	-0.08	1.26
KR-humanities	Count	503 (380.76)	1442 (1570.71)	16 (9.53)
	Std. Res.	6.26	-3.25	2.09
US-essays	Count	136 (258.24)	1194 (1065.29)	0 (6.47)
	Std. Res.	-7.61	3.94	-2.54
KR-science	Count	483 (364.51)	1395 (1524.59)	27 (15.90)
	Std. Res.	6.21	-3.32	2.78
US-essays	Count	136 (254.49)	1194 (1064.41)	0 (11.10)
	Std. Res.	-7.43	3.97	-3.33

Note. N/A = Incomplete clauses for theme analysis. Numbers in the parenthesis denote expected counts. KR-humanities x KR-science: $\chi^2=3.188$; $df=2$; $p=.203$. KR-humanities x US-essays: $\chi^2=134.057$; $df=2$; $p=.000$. KR-science x US-essays: $\chi^2=139.326$; $df=2$; $p=.000$.

According to the standardized residuals of the analyses, both the multiple and single themes are the major contributors of the differences: 6.26 and -7.61 of multiple for KR-humanities and US-essays, -3.25 and 3.94 of single for KR-science and US-essays, 6.21 and -7.43 of multiple for KR-science and US-essays; -3.32 and 3.97 of single for KR-science and US-essays.

Results show that there is no significant difference between KR-humanities and KR-science. Therefore, at least for Korean students,

discipline is not an influential factor for theme functions and plurality. However, proficiency seems to play a role in creating the thematic variation. When compared with US-essays, Korean students used significantly more multiple themes than students of the American university. Table 9 shows the examples.

Table 9. Examples of theme function: Korean and American Students' Writing

	Textual Theme	Interpersonal Theme	Topical Theme
Korean (Humanities)	Moreover,		some critics
		Obviously,	the power to change
Korean (Science)	But		these helps
		In my opinion,	I
American		Fortunately for Partulids in French Polynesia,	at least one extant member of each endangered claude
		However, aside from the anoxia in the water,	all of these

This finding is interesting, as it is not line with the finding in North (2005) who found that humanities students use more multiple themes than science students. On a quick inspection of the multiple themes in the Korean essays, however, the textual themes by Korean students seem to be repetitive and formulaic. This is another hypothesis to test out in a larger-scale analysis considering not only the quantity of multiple themes but also the diversity.

C. Theme role

Consistent with the previous findings, the analysis of the theme role shows no significant difference due to discipline: KR-humanities vs. KR-science and significant differences between Korean and

American data. Table 10 presents the comparison of theme role in the essays written by the students.

Table 10. Theme Role: KR-humanities x KR-science x US-essays

		Adjunct	Subject	Others	N/A
KR-humanities	Count	1085 (1086.10)	771 (764.68)	48 (45.63)	23 (30.59)
	Std. Res.	-0.03	0.23	0.35	-1.37
KR-science	Count	1081 (1079.90)	754 (760.32)	43 (45.37)	38 (30.41)
	Std. Res.	0.03	-0.23	-0.35	1.38
KR-humanities	Count	1085 (940.13)	771 (917.06)	48 (50.88)	23 (18.93)
	Std. Res.	4.72	-4.82	-0.40	0.93
US-essays	Count	504 (648.87)	779 (632.94)	38 (35.12)	9 (13.07)
	Std. Res.	-5.69	5.81	0.49	-1.13
KR-science	Count	1081 (935.57)	754 (904.88)	43 (47.81)	38 (27.74)
	Std. Res.	4.75	-5.02	-0.70	1.95
US-essays	Count	504 (649.43)	779 (628.12)	38 (33.19)	9 (19.26)
	Std. Res.	-5.71	6.02	0.84	-2.34

Note. N/A = Incomplete clauses for theme analysis. Numbers in the parenthesis denote expected counts. KR-humanities x KR-science: $\chi^2=4.129$; $df=3$; $p=0.248$. KR-humanities x US-essays: $\chi^2=114.173$; $df=3$; $p=.000$. KR-science x US-essays: $\chi^2=127.009$; $df=3$; $p=.000$.

Based on the standardized residuals, Adjunct and Subject are identified as major sources of the difference: 4.72 and -5.69 of Adjunct for KR-humanities and US-essays; 4.75 and -5.71 of Adjunct for KR-humanities and US-essays; -4.82 and 5.81 of Subject for KR-science and US-essays; -5.02 and 6.02 of Subject for KR-science and US-essays. Other categories including complements, finite operator, predicator, or WH-elements are contributors of the differences, as well.

It is notable that the theme role is the only feature that differs between KR-humanities and KR-science: the difference, however,

resides in 'Others' category, which consists of Object, Predicator, and non-analyzable sentences or learner errors (marked as N/A). KR-humanities has more items from the others category than KR-science ($p < .05$). It is difficult to tell, however, whether the difference is due to an extensive use of the theme as Object and Predicator, or simply due to the grammar errors either by the students. A grammar error will confuse the parser and an erroneous output from the parser will, in turn, affect the performance of the theme analyzer.

When the essays written by Korean students are compared with US-essays, there is, again, very clear difference. In Korean data, there are significantly more themes used as Adjunct and fewer themes used as Subject. In contrast, there are more Subject-type themes in US-essays with fewer Adjuncts. Regarding the observation, an emerging hypothesis is that the subject-type themes are not random, but rather reflect the use of long themes in US-essays, which were nominal and phrasal. The long subject themes in US-essays then can be related with their effective use of the theme as a carrier of dense information setting the conceptual background.

In sum, no significant difference was found due to academic discipline. However, there was a very significant difference due to proficiency. This suggests, unlike North's (2005) findings, that proficiency is the major factor for the theme variation in Korean students' writing. When compared with the themes in essays from the American university, the themes identified in Korean students' writing are significantly shorter (length) with more textual elements and interpersonal elements (function and plurality) and the preference for adverbial (role). In general, the differences reflect the monotonous repetition of the same theme and an ineffective use of the themes. This may be due to the limited knowledge of diverse thematic patterns.

Although the findings constitute only two pieces of the puzzle in analysis of writing from Korean students, they can still serve as useful questions to ask at a later stage of L2 thematic structure research. We know now that discipline is not likely an important factor, at least for students with lower levels of proficiency. The next logical question is whether discipline will significantly affect thematic variation in the essays of L2 writers who have higher levels of proficiency. In other

words, would highly proficient students majoring in science use the theme differently highly proficient students majoring in humanities? What level of proficiency is the threshold for thematic variation between disciplines?

As for proficiency, a strong factor with definitive influence, this paper proposes a tentative conclusion that the thematic structure may work as a gate-keeper that L2 students must pass in order to obtain a higher level of textual coherence. In other words, there is a good possibility that the effective thematic structure is a hallmark of an advanced L2 writer, and thus needs to be taught to intermediate-level learners. This finding, thus, can provide an insight into the timing of the explicit instruction. As a feature of advanced writing, the theme should be taught when students approach upper-intermediate or advanced levels of proficiency. In a standard 16-week semester, for example, the theme should be taught toward the later stage of the course, perhaps following instruction on connectives, cohesive devices, and clause combining strategies.

V. Conclusion

This paper has explored the thematic structure in L2 writing in relation with two learner variables: academic discipline and proficiency. Theme is the chosen unit of analysis, as it is the key element of textual cohesion and coherence. Analysis shows that discipline is a marginal factor only affecting the semantic role of the theme. In a sharp contrast, proficiency has been found to be a very influential factor that significantly affects all four key areas of the theme. Findings of this study, when considered from a pedagogical perspective, imply that teachers should offer an explicit instruction of the thematic structure at an appropriate timing in later units of curriculum. Through the explicit instruction of the theme, students should be able to understand that some sentences flow logically while some do not; and the ‘flow’ can be controlled by the effective use of the theme. This way, instead of telling—and wishing—they to write logically, teachers can show straightforwardly *how* they can produce a logical text with concrete

examples of the theme.

In future research, a few limitations of the current paper need to be addressed. First, the size of the data set used in this study was still only around 20,000-30,000 words, which is relatively small to draw a definite conclusion. Second, the discipline of students who contributed to American data set has not been considered in this paper. Therefore, there could be an interaction of the discipline with their theme use. Third, there were some unanalyzable sentences in the data due to grammar errors. It is hoped that this limitation will be addressed with the next version of grammatical parser and the theme analyzer. The next study on the theme needs to include an extended consideration of learner variables and an extensive qualitative analysis to reveal the relationship between overall quality of the writing and use of thematic structure. The richness of an extensive theme analysis will add to our knowledge and provide teachers a clearer picture of the challenges that Korean students must confront and serve as a concrete linguistic tool to enhance their textual coherence.

References

- Belmonte, I. A., & McCabe, A. (1998). Theme-Rheme patterns in L2 writing. *Didactica*, 10, 13–31.
- Chiang, S. Y. (1999). Assessing grammatical and textual features in L2 writing samples: The case of French as a foreign language. *Modern Language Journal*, 83(2), 219–232.
- Crossley, S. A., & McNamara, D. S. (2011). Understanding expert ratings of essay quality: Coh-Metrix analyses of first and second language writing. *International Journal of Continuing Engineering Education and Life Long Learning*, 21(2), 170–191.
- Crossley, S. A., & McNamara, D. S. (2012). Predicting second language writing proficiency: The roles of cohesion and linguistic sophistication. *Journal of Research in Reading*, 35(2), 115–135.
- Eggs, S. (2004). *An introduction to systemic functional linguistics* (2nd ed.). New York: Continuum.
- Graesser, A. C., McNamara, D. S., & Kulikowich, J. M. (2011). Coh-Metrix: Providing multilevel analyses of text characteristics. *Educational Researcher*, 40(5), 223–234.
- Granger, S., & Tyson, S. (1996). Connector usage in the English essay writing of native and non-native EFL speakers of English. *World Englishes*, 15(1), 17–27.
- Halliday, M. A. K. (1994). *An introduction to functional grammar* (2nd ed.). London: Edward Arnold.
- Halliday, M. A. K., & Hasan, R. (1976). *Cohesion in English*. New York: Longman.
- Halliday, M. A. K., & Matthiessen, C. (2004). *An introduction to functional grammar* (3rd ed.). London: Edward Arnold.
- Hannay, M. (2007). Patterns of multiple theme and their role in developing English writing skills. In C. S. Butler, R. H. Downing, & J. Lavid (Eds.), *Functional perspectives on grammar and discourse: In honour of Angela Downing* (pp. 257–278). Amsterdam/Philadelphia: John Benjamins.
- Herriman, J. (2011). Themes and theme progression in Swedish

- advanced learners' writing in English. *Nordic Journal of English Studies*, 10(1), 1–28.
- Hirose, K. (2003). Comparing L1 and L2 organizational patterns in the argumentative writing of Japanese EFL students. *Journal of Second Language Writing*, 12(2), 181–209.
- Hyland, K. (2003). Genre-based pedagogies: A social response to process. *Journal of Second Language Writing*, 12(1), 17–29.
- Johns, A. M. (1986). Coherence and academic writing: Some definitions and suggestions for teaching. *TESOL Quarterly*, 20(2), 247–265.
- Johns, A. M. (2002). *Genre in the classroom: Multiple perspectives*. Mahwah, NJ: Lawrence Erlbaum.
- Kubota, R. (1998). An investigation of L1–L2 transfer in writing among Japanese university students: Implications for contrastive rhetoric. *Journal of Second Language Writing*, 7(1), 69–100.
- Lee, S.-W. (2008). Investigating ideational thematic content in Korean university student essay writing. *English Teaching*, 63(2), 71–94.
- Lovejoy, K. B. (1991). Cohesion and information strategies in academic writing: Analysis of passages in three disciplines. *Linguistics and Education*, 3(4), 315–343.
- Martin, J. R. (1992). *English text: System and structure*. Amsterdam/Philadelphia: John Benjamins.
- Matsuda, P. K. (1997). Contrastive rhetoric in context: A dynamic model of L2 writing. *Journal of Second Language Writing*, 6(1), 45–60.
- Mauranen, A. (1996). Discourse competence-evidence from thematic development in native and non-native texts. In E. Ventola & A. Mauranen (Eds.), *Academic writing: Intercultural and textual issues* (pp. 195–230). Amsterdam/Philadelphia: John Benjamins.
- McCabe, A. M. (1999). *Theme and thematic patterns in Spanish and English history texts* (Unpublished doctoral dissertation). Aston University, Birmingham.
- North, S. (2005). Disciplinary variation in the use of theme in

- undergraduate essays. *Applied Linguistics*, 26(3), 431–452.
- Park, K., & Lu, X. (2015). Automatic analysis of thematic structure in written English. *International Journal of Corpus Linguistics*, 20(1), 82–102.
- Péry-Woodley, M.-P. (1991). Writing in L1 and L2: Analysing and evaluating learners' texts. *Language Teaching*, 24(2), 69–83.
- Römer, U., & Wulff, S. (2010). Applying corpus methods to written academic texts: Explorations of MICUSP. *Journal of Writing Research*, 2(2), 99–127.
- Schleppegrell, M. (2001). Linguistic features of the language of schooling. *Linguistics and Education*, 12(4), 431–459.
- Schleppegrell, M. (2004). *The language of schooling: A functional linguistics perspective*. Mahwah, NJ: Lawrence Erlbaum.
- Silva, T. (1993). Toward an understanding of the distinct nature of L2 writing: The ESL research and its implications. *TESOL Quarterly*, 27(4), 657–677.
- Swales, J. M. (1990). *Genre analysis: English in academic and research settings*. Cambridge: Cambridge University Press.
- Thompson, G. (2004). *Introducing functional grammar* (2nd ed.). London: Edward Arnold.
- Ventola, E. (1992). Writing scientific English: Overcoming intercultural problems. *International Journal of Applied Linguistics*, 2(2), 191–220.
- Ventola, E. (1994). Finnish writers' academic English: Problems with reference and theme. *Functions of Language*, 1(2), 261–293.
- Ventola, E., & Mauranen, A. (1996). *Academic writing: Intercultural and textual issues*. Amsterdam/Philadelphia: John Benjamins.
- Witte, S. P., & Faigley, L. (1981). Coherence, cohesion, and writing quality. *College Composition and Communication*, (32), 189–204.

Authors**Park, Kwanghyun**

Myongji University, 1st author
kpark@mju.ac.kr

Nam, Daehyeon

Ulsan National Institute of Science and Technology, corresponding author
dnam@unist.ac.kr

Received: 14th November, 2015

Reviewed: 21st December, 2015

Revised version accepted: 28th December, 2015