

# Remnants of Culture in Journal Article Titles: A Comparison between the United States and Korea in the Field of Social Sciences

논문 제목상의 문화적 흔적: 한국과 미국의 사회과학분야 비교

Eungi Kim\*

## <Contents>

I. Introduction	V. Types of Titles
II. Literature Review	VI. N-gram Comparison
III. Methodology	VII. Conclusion
IV. Title Length	

## ABSTRACT

Most academic journals in the world today typically require submission of journal article titles in English. However, most authors and reviewers are insensitive to the fact that cultural differences at a national level exist in writing titles. In this paper, journal article titles that have been published in the United States and Korea were compared in order to find cross-national cultural characteristics in these titles. To conduct this study, sample titles in the field of social sciences were obtained from two bibliographic databases — *Scopus* and *RIS*. A frequency count on number of variables was used: length of title, types of titles and n-gram phrases. In addition, a variety of similarities and differences found from this study including the type of words and phrases that Korean authors tend to favor in journal articles. The results showed that there is a considerable amount of cultural related variability in the construction of journal article titles. This study suggests that cross national characteristics of journal article titles should be emphasized in the future.

Keywords: Journal articles, Title, Sub-title, Culture, Research paper, Social sciences

## 초 록

국내 대다수의 학술지는 논문 제출시 한글로 쓴 논문 제목과 함께 영어로 된 논문 제목을 제출할 것을 요구한다. 그러나 일반적으로 논문작성자가 논문 심사자들조차도 영문 논문 제목에 큰 비중을 두지 않는다. 본 논문은 미국과 한국에서 출판된 학술지의 논문 제목에서 국가간 문화적 특성과 차이를 알아보기 위한 연구이다. Scopus와 RIS 데이터베이스를 이용하여 사회과학분야와 관련된 논문제목들을 발췌하여 이 연구를 진행하였다. 연구 대상이 된 논문의 제목은 제목의 길이, 종류 및 n-gram 문구 등의 여러 유형을 사용하여 분석 및 비교하였고, 샘플 논문 제목을 분석한 결과, 한글로 작성된 학술지 논문의 제목에서는 특별히 선호되는 단어와 구문의 유형을 발견하였다. 이러한 유형들의 관습적 사용들으로써 논문의 제목에서 국가간의 문화적 차이가 나타난다는 결과를 도출하였다. 이 연구는 영문 제목을 붙일 때에 국내 논문에서 보여주는 전형적인 양상을 지적하고 국가간의 문화적 차이에 대해 좀더 인식하고 염두 해야 한다는 점을 주장한다.

키워드: 문화, 연구, 논문 제목, 사회과학

\* 계명대학교 문헌정보과 조교수 (egkim@gw.kmu.ac.kr)

· 논문접수: 2015년 1월 22일 · 최초심사: 2015년 3월 6일 · 게재확정: 2015년 3월 18일  
· 한국도서관정보학회지 46(1), 345-372, 2015. [http://dx.doi.org/10.16981/kliss.46.201503.345]

## I . Introduction

Every academic journal articles published in the world requires a title. Regardless of the language used to write the journal article, most journals around the world normally require authors to submit journal article titles in English. In Korea, an English title must also be submitted along with a Korean title. Such a requirement appears to be partly motivated by the fact that titles written in English helps to meet international academic standards. In addition, the titles of journal articles in English can increase marketability and visibility to readers outside of Korea. Thus having a journal article title in English, in general, satisfies various requirements known to the academic communities in Korea.

Despite the promotion of the use of English in the academic community, writing a journal article title in English for novice Korean researchers can be a difficult task. Such a case is especially true for authors who prefer to submit their manuscripts to internationally recognized journals, since international journals prefer every aspect of manuscript to be well written. Considering the fact that title is usually the first element that readers usually read before reviewing an article, it is not surprising to know that the title of the journal article must be carefully constructed to attract readers.

An effective title needs to represent the content of the article in a most concise and proper manner. A journal article title should be written using terms and phrases that are as clear and acceptable as possible within the specific disciplinary community. For this reason, it requires some level of mastery in the language that the title is written in. The vocabularies and linguistic expressions should be appropriately and carefully selected for the title. Furthermore, the title has to attract members of a particular academic community and concisely deliver the articles' thesis. Most successful authors would agree with these suggestions on writing journal article titles.

Despite this, most authors and reviewers are insensitive to the fact that cultural differences at a national level exist in writing titles. Moreover, a vast majority of academic journal in Korea are written in the Korean language rather than English. Due to difficulties in writing academic titles in English, Korean authors might write the title in Korean first. Then, after

this initial process, authors in Korea are likely to translate the title of the paper along with the other required metadata just before the final journal article submission.

In general, titles written in English for Korean journals can be considered as a guidepost in order to discover the conventional practices of Korean authors. Sewell (2005) explained that culture can be viewed as a system and practice. Looking from this point of view of culture, it is this author's contention that the discourse of the academic community tends to differ from one nation to another. As authors tend to work and publish journal articles in their country, the practice of using a particular type of linguistic expressions are disseminated and shared in common within the particular academic community that he or she typically resides. The resulting linguistic form ends up representing a certain sense of normality by becoming molded into the particular academic standards. This paper argues that the language used in the academic community can be examined based on the countries of the published journals in addition to the disciplines.

Meanwhile, from an information science perspective, titles are closely associated with the notion of information. Buckland (1991) argued that the notion of information can be best viewed as a process, knowledge, or thing. Theoretically, creating a metadata is often motivated by the fact that metadata can provide essential information about the object (National Information Standards Organization, 2004). Among the essential metadata, a journal title can be considered as a signpost for an article. However, not only titles often display the essential information about the articles, there could also be other elements of information that are often hidden; these information are ignored by ordinary academics. For this reason, researchers should investigate various types of information that can be obtained by examining titles. Most practitioners and academics appear to be unaware of the fact that the titles can show certain elements of information other than the description of the content.

As an interesting element of titles that can be discovered, this paper intends to show the remnants of culture in the journal article titles. Apart from this, the intention of this paper is to promote understanding of the cultural aspects of the titles. Discovering how researchers in a country in a particular discipline write titles has practical implications since we live in a globalized academic community. As will be discussed in the subsequent section, previous research has shown that titles can potentially provide insights into various aspects of a particular discipline. For the purpose of the this paper, the study has focused on comparing

U.S. and South Korea in an attempt to reveal characteristics of journal article titles in the context of culture.

In order to find the cross-national characteristics between Korea and the U.S., different elements of the titles were quantified and examined. More specifically, the journal paper titles published in Korean journals and the U.S. journals are compared by quantitatively analyzing the title types as well as the elements of titles such as common terms and phrases. As for the data, two bibliographic databases — *Scopus* (<http://www.scopus.com>) and Resource Information Sharing Service, known as *RISS* (<http://www.riss.kr>) — were used in this study. This research is based on 11,639 titles drawn from each database.

## II. Literature Review

Much of previous research focused on disciplinary characteristics of titles suggested that the titles of academic journal articles differ across disciplines. For instance, Haggan (2004) examined the difference in titles among journal articles published in the field of literature, linguistics, and science. The author found some noteworthy differences in the frequency counts of variables (e.g., number of words, full sentence, noun phrase, etc.) among the different disciplines. In comparing the field of literature to science, Haggan (2004, 313) pointed out that literature titles often attempted to attract the reader through certain verbal flirtation, enticing the reader with suggestive and tantalizingly enigmatic hints of the delight that would follow. In comparison, scientific journal article titles tend to be up-front and straight-forward in presentation of information. A possible explanation for this observed finding is that the field of science values facts and data whereas the field of literature favors flowery language and other literary elements that are typical of the discipline.

Nevertheless, the most common quantified element, as shown in Table 1, in the previous research was the length of title. The earlier research suggests that the average length of title in "hard" science might be longer than most other fields. In spite of previous works in this area, an average length of titles was not reported for the field of social sciences.

&lt;Tab. 1&gt; Average Length of Titles Reported Previous Studies

Study	Field	Sample Size	Average Length
Anthony (2001)	Computer Science	600	8.9
Haggan (2004)	Literature	237	9.4
Haggan (2004)	Linguistics	207	8.8
Gesuato (2009)	Linguistics	250	10.8
Jalilifar (2010)	Applied Linguistics	997	10.60
Wang and Bai (2007)	Medical	417	10.9
Mendez et al. (2014)	Astro-physics	300	13.49
Haggan (2004)	Sciences	307	13.8
Cianflone (2013)	Food Science	126	15.3

Another title element that earlier research investigated was the uses of punctuation marks. In particular, the previously conducted research examined the use and implications of using colons. For instance, Hartley (2007) have shown that the use of colon can vary by discipline and argued that some readers even prefer to read titles with colons. There have been studies that examined whether the colon has an effect on citation rate as discussed in Jamali and Nikzad (2011). A general recommendation for increasing citation based on colon usage cannot be easily determined as Subotic and Mukherjee (2014) argued recently. Despite all of the previous works on punctuation marks, punctuation marks in titles have not been investigated in the context of discovering national characteristics.

Other than the length and punctuation marks, prior research studies in this area were carried out for the purpose of examining different forms of titles (e.g., question title). However, most of these studies were conducted from a linguistic point of view with emphasis on discovering common discourse pattern in a particular discipline. The sample data drawn for these studies were generally small due to human decision required in identifying and sorting each of the linguistic elements being studied. For instance, Wang and Bai (2007) examined the type of nominal groups in a title. Incidentally, nominal title is a title that is based on a single phrase and does not contain a conjugated verb. In any case, a human identification of the nominal title was employed. Identifying nominal titles computationally is more difficult, and algorithmic solutions generally produce imprecise results.

For this study, a more relevant issue is the type of title and the sort of title elements that readily exhibit cultural characteristics. Arguably, finding title elements that show cultural

features could be linked to common linguistic expressions that authors in a given country tend to favor. While Dahl (2004) discussed cultural aspects in academic writing and pointed out the fact that academic writers leave traces of national and disciplinary characteristics, focusing on cultural differences, Rath (2010) investigated the cultural aspects of writing journal article titles and showed differences that existed between native Chinese writers and native English writers. Rath (2010) suggested that journal articles written in English by native Chinese writers seemed to reflect both Chinese research writing form and Anglo-American form.

In the case of Korea, for instance, the word "study" in Korean is prevalent in journal article titles. Such choice of words in the Korean academic community may seem like a trivial issue to some academics. Yet, it is an indication of a unique cultural form reflected in a linguistic form that is contrastive to the Western culture. Because of this, traces of culture could be found, especially if this type of claim can be supported with an ample amount of empirical evidences. Unfortunately, as Soler (2007) pointed out, the quantitative evidence-based formulation of acceptable titles for journal articles in the context of culture is not readily available. In fact, this research extends Soler's (2007) point one step further. The author of present study argues that titles drawn from a particular academic community can reveal remnants of the national culture.

In a somewhat related study, Dahl (2004) argued that academic authors are influenced by their native languages' culture as well as by the disciplinary culture. The remnants regarding linguistic selections that are used in titles can be traced back to Korean Confucianism and its long history. In the case of Korea, contemporary style of conducting research was adopted from the West and Japan, but China has an enormous influence on Korean writing system. Since Korean is typically translated into English, such evidence of the Korean culture, which have been developed over time may surface in the title.

The discourse of the academic community supposedly transcends national boundaries. For the most part, in academia, national cultural characteristics should not be a major factor in assessing the quality of research. Academic communities, in general, do not want to emphasize the distinction between native English forms versus a non-native of English forms as suggested in (Romer and Arbor 2009). Yet, more than likely, reviewers of academic journal articles would favor articles written in a form closer to native English than non-native English. The distinction between native English author and non-native English authors can be made based

on the type of syntactic errors that one makes in the use of the language (Wong et al. 2009). In addition to being in a native-like form, the expressions in the title need to be more native to a particular discourse academic community.

In sum, the majority of previously conducted research focused on discovering disciplinary differences from a linguistic standpoint. For this reason, most sample sizes that the authors used were rather small. Although there has been a number of previous research studies on titles, previous works on examining nationality-based cultural aspects of journal article titles have been conducted sufficiently. Rather, this study will be emphasizing the cultural similarities as well as the differences between two countries — South Korea and the U.S. — with respect to journal article titles in the field of social sciences.

### III. Methodology

The main approach that was employed in this study is to compare the characteristics of journal article titles by comparing the various title elements between the journals published in Korea and the United States. To accomplish this goal, two bibliographic databases — *Scopus* (<http://www.scopus.com>) and Resource Information Sharing Service, known as *RISS* (<http://www.riss.kr>) — were used for this study to create two equal quantities of datasets. According to *RISS* website, *RISS* contains approximately 3.5 million journal articles. For simplicity in analyzing the data, an equal amount of approximately 12,000 journal article titles were drawn from each of the database. Based on these drawn samples, the final dataset was created with each dataset containing 11,639 titles. We refer the final datasets as *RISS* and *Scopus* datasets in the foregoing. The detailed process of selecting data for this research is as follows.

Both *RISS* and *Scopus* databases require initial keywords for searching. *RISS* required an initial keyword or keywords for searching the database. We selected two keywords: “this” and “that” and matched against the abstract field. These keywords were merely selected to reduce the size of the sample pool. The retrieved results seemed to have little correlation to the particular type of title as they were matched against the abstract rather than the title. To further reduce the retrieved results, the following are some of the additional conditions that were applied: a) disciplines were selected as the social science, b) only *KCI* indexed journals were

selected, the article language text selected was Korean, and c) the date selected was year 2013. The preprocessing step was applied to remove duplicate titles and titles that were erroneous. For the approximately 12,000 retrieved results using the above conditions, the first 11,639 journal article titles were selected from the RISS website. These selected datasets were from 620 journals, which were in the field of the social sciences.

Similarly, approximately 12,000 titles from journal articles were randomly drawn from the *Scopus* website initially. To obtain quantitatively equal amount of sample pool to the *RISS* sample, which was 11,639, the following procedure was applied. First, in creating the *Scopus* dataset, the same two keywords that were selected used to against abstract field in *RISS* were used. These are “this” and “that”. Additionally, the following conditions were imposed: a) the discipline selected was the social sciences, b) the “United States” was selected for the country, c) “Article” was selected as article type, “journals” was selected for source type, d) “English” was selected for the language, and e) “2013” was selected for the year of publication. After this preprocessing step, only first 11,639 journal article titles were finally selected based on the retrieved results. For the *Scopus* database, the selected dataset comprised of 273 journals.

For the purpose of uncovering cultural similarities and differences, the created datasets were analyzed based on the following criteria:

1. Length of title
2. Title types: colonic, question, and assertive title
3. Frequently used words and phrases based on n-gram

The primary tool to process the downloaded raw data was a UNIX based shell scripting language such as *sed* and *awk* (Dougherty and Robbins 1997). Furthermore, Microsoft Excel was used to obtain the descriptive statistics from the collected frequency counts.

Moreover, the limitations of this study need to be stated. First, this study relied upon *RISS* and *Scopus* to categorize journal articles. There could be interdisciplinary journals that could be more difficult to categorize according to categories listed in *RISS* and *Scopus*. A categorization of the journal may be subjective and may have influenced the results. Second, in spite of many selected journals in the field of social sciences, each created dataset for this study may not be representative of entire field of social sciences. Third, social science journals in Korea and U.S. may cover different disciplines. Library and Information Science papers, for instance, are published mainly in Social Science journals in Korea, but are likely to be



published in multiple domains (e.g., computer science, informatics) in the U.S. This may perhaps be relevant for the title formation since it could influence the result of the study depending on the disciplines of the journals. Lastly, the n-gram approach, which described later in this paper, appears to require more data in some instances to generalize the findings.

#### IV. Title Length

The purpose of examining the length of the title was to notice the disciplinary characteristics of the social sciences as well as the cultural characteristics between the two countries - South Korean and the U.S. The length of the title was performed by counting the total number of words within the title. For the purpose of this study, a word is defined as a series of character strings that has is bound by a blank space.

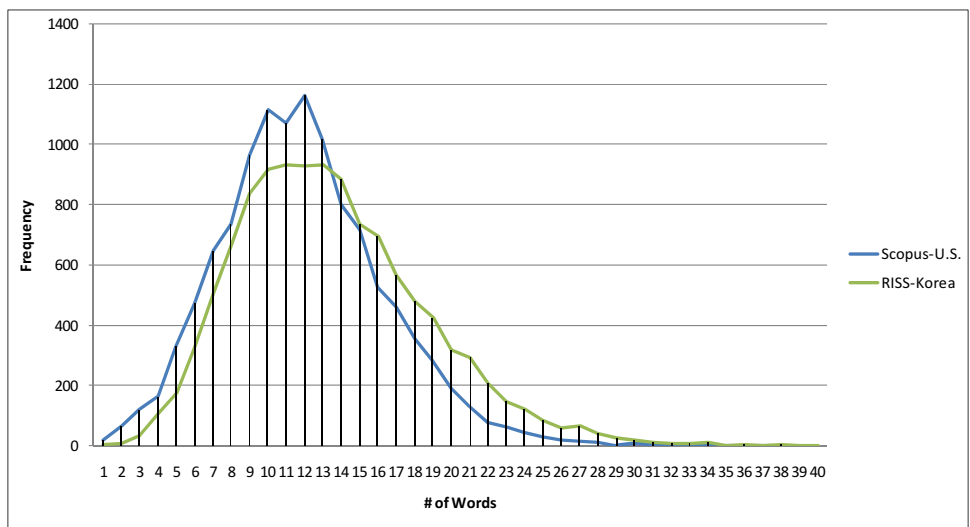
The average length of the title was the following:

- 14.5 words for the *RISS* database, which contains journal articles published in South Korea, and
- 11.9 words for *Scopus* database, which contains journal articles published in the U.S.

As shown in Figure 1, the length of the title appears to be 30 words or less for both datasets. Since most journal article titles contain less than 30 words, titles having longer than this length might be viewed as exceptionally lengthy. Reviewers might not view such articles having this long title favorably. In contrast, in this same figure, journal articles bearing exceptionally short titles are also revealed. The shortest title in *RISS* was two, while the shortest title in *Scopus* dataset was one. For instance, the title “Divination” has only one word. Despite its short length, this one-word word was present in the Scopus dataset. Yet, as shown in Figure 1, a length of title, which is less than 2 or longer than 30, was rare.

The result shows that the overall length of journal article title in Korea is longer than in the U.S. The average title length between the two countries is 13.1. Overall Korean authors tend to write longer titles in comparison to authors who published in the U.S. based journals. Based on Table 1, it is reasonable to assume that the average length of title in the field of social sciences lies somewhere between computer science and food science. Yet, a point that is more important should be made regarding the length of the title in social sciences. That is,

it is difficult to determine the overall length of titles in social sciences because it may depend on each type of discipline within the social science (e.g., psychology versus sociology) and the datasets that are used to find the average length of title. To this end, generalizing based on the *Scopus* database alone for finding the average length of titles are not sufficient, as it may not represent an entire field of the social sciences.



<Fig. 1> Ordered Distribution of Title Length

In addition, as shown in Table 2, a t-test was performed to determine if there was a difference between two means of title lengths. The  $p$ -value ( $< 0.05$ ) indicates that the average length difference between the *RISS* and *Scopus* dataset is statistically significant. *RISS* dataset showed a higher variance than *Scopus* database, suggesting relatively greater variability in the length of titles.

Further observation can be made by considering the probability distribution of the datasets. Table 3 shows the kurtosis and the skewness measurement of two datasets — *RISS* and *Scopus*. The skewness measurement of both datasets indicates that there were positive values for skewness, since the tail of the data distribution graph is larger to the right. Both databases show a moderate level of skewness since skewness is not found to be substantial as the skewness is  $< 1.0$ . The skewness measurement of *RISS* dataset is slightly larger than that of the *Scopus* dataset. Furthermore, the kurtosis measure of the *Scopus* dataset is larger than the *RISS*, indicating that *Scopus* has a higher peak than *RISS*. As shown in Figure 1, the *Scopus*

dataset has the highest peak when the number of words reaches 12. In other words, the common length of titles is around 12 for both datasets since the mode, which is shown as highest peak, is approximately 12.

<Tab. 2> t-Test: Paired Two Sample for Means

	RISS	Scopus
Mean	14.51	11.89
Variance	26.77	19.89
Observations	11639.00	11639.00
Hypothesized Mean Difference	0.00	
df	11638.00	
t Stat	41.23	
P(T<=t) one-tail	0.00	
t Critical one-tail	1.64	
P(T<=t) two-tail	0.00	
t Critical two-tail	1.96	

A possible explanation for the lengthier titles in *RISS* dataset should be mentioned. Perhaps English being the second language, a Korean author is likely to write a title of journal article in Korean prior to translating it into English. More than likely, a translated title from Korean to English tends to be longer than the original Korean version of title. Frankenberg-Garcia (2009) pointed out that translated text tends to be longer than the original language due to the process known as explicitation. Explicitation is the process of explicitly describing implicitly stated information in the source text to the target text (Vinay and Darbelnet 1958).

<Tab. 3> Measurement of Kurtosis and Skewness

	RISS - Korea	Scopus - U.S.
Kurtosis	0.983211	1.278893
Skewness	0.765722	0.59063

## V. Types of Titles

### 1. Colonic Titles

The utilization of colon in titles can be simply defined as a title having a colon “:” to separate words or phrases based expressions in a title. Hartley (2007) used the term “colonic titles” in order to refer to titles having such colons in a title. There were substantial differences between the numbers of colonic titles between the two databases — *Scopus* than in *RISS*. As shown in Table 4, the result shows that 50.3% of journal article titles in *Scopus* dataset contained colons. From here and on forward, the percentage indicated in parenthesis is obtained by dividing the total number of titles, which is 11,639, in a dataset. This result is not surprising considering that the commonality of the use of the colon in titles was reported in Lewison and Hartley (2005). They reported that colons could range from 7% to 50% depending on the discipline. For the field of social sciences, we found out that 20.1% journal article titles in *RISS* dataset contained colons whereas 50.3% of titles in *Scopus* database contained colons. Even though the same field was chosen in this study, an exceedingly contrasting result was obtained regarding the use of colons.

The average of title length is also shown in Table 4. In both *RISS* and *Scopus* dataset, the average length of colonic titles is longer than the non-colonic titles. Because the colon implies the use of a sentence or phrase before and after it, perhaps it is the reason why the average length of colonic titles is longer than the non-colonic titles. In addition, there seems to be positive correlation between title types and the length of the title. However, the length of *RISS* was consistently longer than that of the *Scopus* in terms of both colonic and non-colonic titles.

The result shows that the length of colonic titles is longer than the non-colonic titles. Also, in this table, the average number of words in the colonic titles pertaining to each dataset is shown. For both colonic and non-colonic titles, the length of journal article titles in *RISS* database was much longer than *Scopus* database.

<Tab. 4> Colonic Titles

Title Types	Example Title	RISS - Korea		Scopus - U.S.	
		Instances	# of words	Instances	# of words
Colonic Titles	Choosing between lotteries: Remarkable coordination without communication	2329 (20.1%)	17.4	5850 (50.3%)	13.3
Non-Colonic Titles (No Presence of Colon)	Family structure and child anemia in Mexico	9310 (79.9%)	13.77	5789 (49.7)	10.43

## 2. Question Titles

Titles in the form of questions or “question title” allows writers to pose a question and to offer one answer to a question and can raise readers’ curiosity (Soler 2007). As shown in Table 5, all three types of question titles appeared in both databases. In contrast to question titles in *Scopus* database, in the *RISS* dataset, the total instances of question titles appear to be low; only 162 instances (1.42%) were found.

The number of these question titles should be categorized along with the previous works on title length. Ball (2005) has shown that question titles are on the rise in the field of life sciences, medicine, and physics. Whether the increase is applicable to field of social sciences is not yet clear. A study conducted by Ball (2005) has reported that the field of medicine had 6.7% of question titles, which was the highest percentage of question titles among the three fields that the author examined. However, to our knowledge, Ball (2005) published his work in 2005, and, since then, the field of social sciences had not been examined in terms of question titles.

In this study, 7.9% of question titles were found in the *Scopus* dataset. More importantly, the result shows that there seem to be a clear difference between Korea and the U.S. The low instances of question titles in the *RISS* dataset suggests that Korean authors are still likely to take more conservative approaches when selecting a title type. The result indicates that Korean authors do not use the question titles as often as the authors who published in U.S. journals.

<Tab. 5> Question Titles

Question Title Types	Example Title	RISS - Korea		Scopus - U.S.	
		Instances	# of words	Instances	# of words
Single Question with Ending “?”	How many queries will resolve common randomness?	52 (0.45%)	11.3	208 (1.79%)	9.9
Compound Question with Left Side “?”	What about language while equitably assessing science?: Case studies of preservice teachers' evolving expertise	90 (0.77%)	17.1	473 (4.06%)	14.7
Compound Question with Right Side “?”	3D printing for children: What to build next?	20 (0.17%)	14.3	239 (2.05%)	12.5
Total Instances		162 (1.4%)		920 (7.9%)	
Average # of Words			14.9		13.1

### 3. Assertive Titles

In addition to colons and question marks, this study examined the use of exclamation marks in the titles. For the purpose of this study, assertive titles are defined as titles that contain an exclamation mark !. As in question titles, assertive titles can easily draw attention of readers. The appropriateness of assertive title in journal articles is open to debate. Peat (2002) pointed out an interesting reason why exclamation should not be used in academic journal articles. The author pointed out that assertive title should be avoided since proving beyond reasonable doubt is not realistic and bold conclusion often tends to misinform readers. As Peat (2002) suggested, the results should not be stated imperatively as there are often risks associated stating the findings bluntly.

In any case, whether assertive title is appropriate or not appropriate in journal titles is beyond discussion of this paper. In this study, assertive titles were examined in the context of finding the differences between the two countries and to recognize the pattern in the field of social science. The frequency count of assertive titles is shown Table 6. Assertive titles appeared more often in the *Scopus* database than *RISS* database. However, due to a small number of total instances, we could assume that assertive titles are rare in the titles in the field of social sciences.

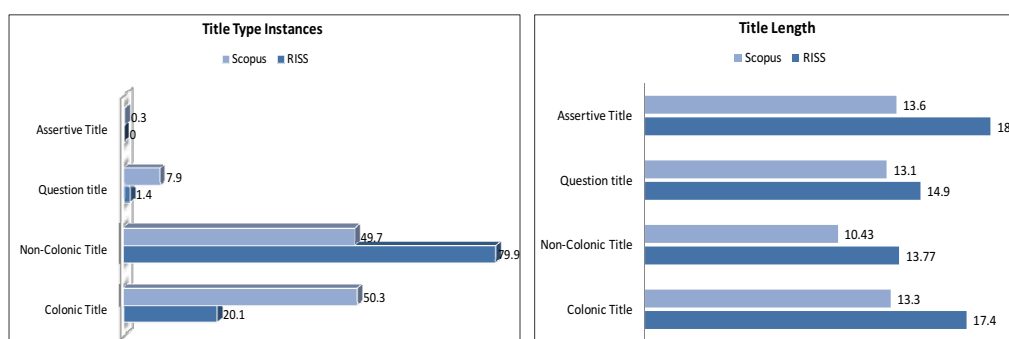
<Tab. 6> Assertive Titles

Assertive Title Types	Example Title	RISS - Korea		Scopus - U.S.	
		Instances	# of words	Instances	# of words
Single or Colon Compound Assertive Title (!)	WE HAVEN'T TALKED IN 30 YEARS!: Relationship reconnection and internet use at midlife	3 (0.0%)	18.7	31 (0.3%)	13.6

### 4. Comparison of Title Types

In a nutshell, Figure 2 illustrates the length of titles based on the title types and title type instances found in *RISS* and *Scopus* datasets. This figure has been generated from the data shown in Table 4, 5, and 6. The differences of the average length of titles can be noticed among the title types and between the *RISS* and *Scopus* datasets. As shown in this figure, the

length of titles in RISS dataset is consistently longer than *Scopus* dataset for each type of titles. Among the different title types that were examined, non-colonic titles were the most common title type in RISS. The majority of titles in *RISS* are in non-colonic form and a non-colonic form. The instances of title types in *Scopus* dataset is more evenly distributed throughout the title types. Also, the assertive title instances are negligible in both datasets. The result also implies that Korean authors use the punctuation mark in all of mentioned titles far less frequently than authors who submitted to U.S. journals in the social sciences.



<Fig. 2> Comparison of Title Types

## VI. N-gram Comparison

N-gram can be often used to discover useful information in data collection set (Sureka and Goyal, 2010). In this study, n-gram can reveal various uses of linguistic expressions. Phrases are made of different number of words (e.g., 1, 2, 3, etc.) and n-gram refers to the number of words appearing as a phrase (unigram, bigram, trigram, etc). The number of words in n-gram can increase from 1 and onward, exhibiting different types of characteristics. A frequency of unigram (1 word) is the simplest to generate, but the context of words can become lost during the process. For this reason, primarily bigram (2 words) and trigram (3 words) were employed to discover some of characteristics between Korean journal articles and the U.S. journal articles. Notably, the frequency counts of bigram and trigrams were calculated and compared to each datasets. As a result, topical differences and linguistic characteristics between the two datasets were identified. In calculating the n-gram, punctuation marks were

removed and all upper cases were normalized to lower cases letters.

A top 50 frequently used bigram appearing in *RISS* database and *Scopus* database is shown in Table 7. Bigram phrases that do not appear in the opposing column are highlighted in the table. Both the bigram phrases that are unique in the column but do not appear in the opposite column are highlighted. Thus, bigrams that are not highlighted represent the phrases that are presented in both top 50 list. Before continuing the discussion on the bigram approach in order to discover the uses of unique phrases, limitations of this study should also be emphasized. To a certain extent, this study has its limitations in that it will be difficult to generalize the result in other cases. The result may depend on the accuracy of random samples, and if a sample is drawn, it will likely to produce a slightly different result. However, the procedure described in this section is to demonstrate the overall procedure, which is a relatively reasonable approach used to discover differences between the two datasets. In this sense, the exact number of matching and un-matched bigram phrases is not particularly important for this study. Nevertheless, out of 50 bigram phrases, 24 bigram phrases appear in both datasets. In other words, approximately 1/2 of the listed bigram phrases are in both datasets. The bigram phrases may still appear in the opposing dataset since only the top 50 bigrams were considered.

<Tab. 7> Top 50 Bigram Phrase List

Rank	RISS-KOREA	FREQ	%	SCOPUS-U.S.	FREQ	%
1	on the	3351	28.79%	in the	924	7.94%
2	study on	2693	23.14%	and the	774	6.65%
3	of the	2016	17.32%	of the	755	6.49%
4	in the	843	7.24%	of a	241	2.07%
5	effects of	775	6.66%	in a	227	1.95%
6	effect of	746	6.41%	on the	213	1.83%
7	and the	580	4.98%	role of	212	1.82%
8	analysis of	567	4.87%	analysis of	198	1.70%
9	study of	503	4.32%	study of	191	1.64%
10	focusing on	499	4.29%	effects of	188	1.62%
11	focused on	482	4.14%	from the	166	1.43%
12	for the	387	3.33%	evidence from	153	1.31%
13	based on	323	2.78%	impact of	153	1.31%
14	relationship between	235	2.02%	to the	145	1.25%
15	analysis on	228	1.96%	the united	132	1.13%
16	and its	227	1.95%	for the	131	1.13%
17	to the	221	1.90%	as a	130	1.12%
18	case of	203	1.74%	the role	126	1.08%
19	of a	203	1.74%	case of	101	0.87%
20	of korean	202	1.74%	the case	96	0.82%
21	focus on	183	1.57%	of social	92	0.79%
22	the korean	182	1.56%	a case	91	0.78%
23	of social	172	1.48%	implications for	91	0.78%



24	case study	171	1.47%	perceptions of	91	0.78%
25	impact of	169	1.45%	from a	83	0.71%
26	influence of	167	1.43%	and social	80	0.69%
27	a study	160	1.37%	at the	79	0.68%
28	in korea	160	1.37%	in an	78	0.67%
29	the relationship	160	1.37%	case study	77	0.66%
30	role of	159	1.37%	the impact	77	0.66%
31	satisfaction and	158	1.36%	use of	75	0.64%
32	development of	157	1.35%	influence of	73	0.63%
33	characteristics of	156	1.34%	of an	72	0.62%
34	the effect	149	1.28%	effect of	71	0.61%
35	the case	144	1.24%	and its	70	0.60%
36	empirical study	135	1.16%	approach to	70	0.60%
37	according to	126	1.08%	united states	67	0.58%
38	students with	123	1.06%	differences in	66	0.57%
39	the effects	123	1.06%	african american	65	0.56%
40	comparative study	116	1.00%	the us	65	0.56%
41	from the	109	0.94%	a new	63	0.54%
42	as a	108	0.93%	relationship between	63	0.54%
43	early childhood	104	0.89%	development of	60	0.52%
44	job satisfaction	104	0.89%	comparison of	59	0.51%
45	improvement of	103	0.88%	change in	57	0.49%
46	the moderating	102	0.88%	children with	56	0.48%
47	the mediating	100	0.86%	examination of	56	0.48%
48	in korean	99	0.85%	model of	56	0.48%
49	elementary school	98	0.84%	politics of	53	0.46%
50	application of	97	0.83%	the effects	52	0.45%

Note: Phrases that appear in both *RISS* and *Scopus* are highlighted in gray. The percentage was calculated based on 11,639, which is the total instance of each dataset.

As shown, the most frequently appearing bigram in the *RISS* dataset is “on the” while the most frequently appearing bigram phrases in *Scopus* dataset is “in the”. What is more interesting is the use of bigram phrase such as “study on”. The bigram “study on” was the second most commonly appearing bigram in *RISS* (23.14%) while this bigram did not appear in the *Scopus* database. This implies that most Korean authors prefer using the phrase “study on” in the title, while the U.S. authors prefer not to use these phrases frequently in the journal article titles. Other notable bigram phrases in *RISS* include the word "focus". The result shows that 10% title has the variation of the word focus: “focus on”, “focusing on”, and “focused on”. Further probe into these phrases are desirable.

In Table 7, a variation of bigram phrases that contains the particular word “Korea” (e.g., “of korean”, “the korean”, “in korea”, “in korean”) are shown in the *RISS* dataset list. One note that is worth mentioning is that the capitalized letters are normalized to lower case letters in order to reduce variation associated with letters. Nonetheless, phrases that has the word

“korea” is an indication of a topical matter that is concerned with Korea. Similarly, we can see that the bigram phrases, namely “the United” and “United States”, are merely decomposed phrases of “the United States”. These phrases reflect the topical matters most likely relevant to the United State. Thus, the highlighted bigram phrases, frequently appearing one database as opposed to another, may indicate topical matters unique to a particular dataset.

Table 8 shows a top 50 frequently used trigram phrases in the *RISS* and *Scopus* datasets. Unlike bigram phrases in Table 7, this table shows that a fewer trigram phrases are matched with each other. There were 11 matched trigram phrases that mutually exist in two of the top 50 lists. As the list of words in n-gram become longer, the matched n-gram list should become shorter. Thus, 11 matched trigram phrases are far less prevalent than 24 matched bigram phrases. The higher number of words based n-gram will lead to fewer matching phrases because of the probability associated with the number of words in the n-gram (Thanopoulos et. al. 2002). As the number of the n-gram increases, the probability of having particular collocation of words decreases.

<Tab. 8> Top 50 Trigram Phrase List

Rank	RISS-Korea	FREQ	%	Scopus-U.S.	FREQ	%
1	study on the	1664	14.30%	the role of	123	1.06%
2	focusing on the	198	1.70%	in the united	105	0.90%
3	focused on the	178	1.53%	the case of	85	0.73%
4	study of the	160	1.37%	the impact of	73	0.63%
5	analysis of the	159	1.37%	a case study	62	0.53%
6	the effect of	139	1.19%	the united states	53	0.46%
7	a study on	134	1.15%	the effects of	52	0.45%
8	the relationship between	133	1.14%	case study of	46	0.40%
9	analysis on the	124	1.07%	analysis of the	40	0.34%
10	the case of	124	1.07%	the relationship between	35	0.30%
11	empirical study on	117	1.01%	the influence of	34	0.29%
12	the effects of	117	1.01%	a study of	32	0.27%
13	based on the	104	0.89%	a comparison of	31	0.27%
14	on the relationship	103	0.88%	evidence from the	31	0.27%
15	on the effect	86	0.74%	the effect of	28	0.24%
16	case study on	82	0.70%	the politics of	28	0.24%
17	on the effects	80	0.69%	in the us	26	0.22%
18	moderating effect of	77	0.66%	evidence from a	25	0.21%
19	mediating effect of	74	0.64%	the development of	25	0.21%
20	moderating effects of	74	0.64%	the context of	23	0.20%
21	case study of	72	0.62%	an analysis of	22	0.19%
22	korean study on	71	0.61%	an examination of	22	0.19%

23	comparative study on	70	0.60%	study of the	22	0.19%
24	effect of the	65	0.56%	united states and	21	0.18%
25	mediating effects of	64	0.55%	case of the	20	0.17%
26	focus on the	59	0.51%	in the context	20	0.17%
27	in korean study	59	0.51%	the study of	19	0.16%
28	a focus on	58	0.50%	the use of	19	0.16%
29	of the korean	57	0.49%	examination of the	17	0.15%
30	the development of	57	0.49%	the emergence of	17	0.15%
31	effects of the	54	0.46%	with and without	17	0.15%
32	according to the	53	0.46%	and use of	16	0.14%
33	the impact of	52	0.45%	in new york	16	0.14%
34	the influence of	51	0.44%	the importance of	16	0.14%
35	the role of	49	0.42%	to the special	16	0.14%
36	exploratory study on	48	0.41%	differences in the	15	0.13%
37	for students with	48	0.41%	the rise of	15	0.13%
38	the moderating effect	45	0.39%	the roles of	15	0.13%
39	a case study	44	0.38%	a response to	14	0.12%
40	the improvement of	44	0.38%	and the politics	14	0.12%
41	on the improvement	43	0.37%	as predictors of	14	0.12%
42	job satisfaction and	42	0.36%	have sex with	14	0.12%
43	the perspective of	42	0.36%	the effectiveness of	14	0.12%
44	on the mediating	41	0.35%	a reply to	13	0.11%
45	in the korean	39	0.34%	and the role	13	0.11%
46	on the factors	39	0.34%	in the new	13	0.11%
47	the moderating effects	38	0.33%	introduction to the	13	0.11%
48	the mediating effect	37	0.32%	new york city	13	0.11%
49	research on the	36	0.31%	role of social	13	0.11%
50	of elementary school	36	0.31%	who have sex	13	0.11%

Note: Phrases that appear in both *RIS*S and *Scopus* are highlighted in gray. The percentage was calculated based on 11,639, which is the total instance of each dataset.

In Table 8, the matched trigram phrase that has a highest rank in the *RIS*S dataset was “study of the”, whereas the matched trigram phrase that has a highest rank in the *Scopus* dataset was “the role of”. As in bigram phrases, the trigram phrases that have the word “study” have a higher frequency count in the *RIS*S trigram phrase list than in the *Scopus* dataset. Since the frequency of “study of the” is 1.37%, no matched trigrams were more than 1.37%. Similar to bigram phrases, the unmatched trigram phrases seem to exhibit the frequently discussed topics as well as expressions that are more unique to a particular nation. Trigrams such as “in new york” or “in the us” indicate the traces of topics that the authors in the U.S. might be interested in. The same point was made earlier with bigram phrases. Moreover, unmatched

trigram phrases such as “analysis on the” suggests that Korean author's being less native-like use of English language. A mistaken use of collocations, any common pair of words, would consequently be a characteristic feature of non-nativeness (Ballesteros 2005).

For further illustration, the frequency counts of the two selected words are shown as in Table 9. In particular, two words — “study” and “focus” — should be examined with further analysis. The reason for their selection is that “study” and “focus” are among the most frequently appearing words in the *RISS* dataset. These words can be referred to as cue words. A wild card operator “\*” indicates that any pattern after the word string was matching. As shown in Table 9, 31% of titles in *RISS* dataset contain the word “study”, while 3.87% of Scopus dataset contain the word “study”. Similarly, the variation of word “focus” appeared in *RISS* dataset; the frequency was considerably higher than in the *Scopus* dataset.

In general, the unmatched bigram phrases are of following characteristics:

- the nature of topic tends to be in the best interests of a nation-based group of individuals
- common collocation of words tends to be used more frequently in one country than another

Bigram or trigram phrases can be categorized based on these criteria. In the later section, there are further illustrations of these points. In essence, bigrams or trigrams can be categorized based on these criteria.

<Tab. 9> Frequently Used Cue Words

Index	RISS-KOREA	FREQ	%	SCOPUS-U.S.	FREQ	%
1	study*	3618	31.09%	study*	451	3.87%
2	focus*	1232	10.59%	focus*	39	0.34%

Table 10 is a top 20 list of bigram that contains the word “study”. As shown, the overall frequency count of *RISS* was much higher than that of the *Scopus* dataset. In regards to the *RISS* dataset, the overall distribution pattern of top 20 bigram containing the word “study” appears to be similar to bigram graph that was shown in Figure 3, if graphed. In fact, the graph of the bigram that contains the word “study” would be lower than the graph of the entire bigram. The reason is that the frequency counts in the bigram containing the word “study” is apparently less than the frequency counts of the overall bigram phrase. The word “study” is accompanied by a collocated word. The parts of speech (POS) of the collocated word could

be categorized as adjectives or other than adjectives (e.g., preposition, conjunction, articles, etc.). The other than adjective POS of collocated word can be shown in phrases such as “study on”, “study of”, and “the study”.

<Tab. 10> Top 20 List of Bigram Containing the Word “Study”

Rank	RISS – Korea				Scopus - U.S.			
	‘Study’ with Other Than Adjective	‘Study’ Followed by an Adjective Collocated Word	FREQ	%	‘Study’ with Other Than Adjective	‘Study’ Followed by an Adjective Collocated Word	FREQ	%
1	study on		2693	23.14%	study of		191	1.64%
2	study of		503	4.32%		case study	77	0.66%
3		case study	171	1.47%	the study		20	0.17%
4		empirical study	135	1.16%	study in		16	0.14%
5		comparative study	116	1.00%	study from		15	0.13%
6		exploratory study	64	0.55%		exploratory study	12	0.10%
7	study for		63	0.54%		longitudinal study	12	0.10%
8		qualitative study	25	0.21%		qualitative study	11	0.09%
9	study about		22	0.19%		comparative study	10	0.09%
10		legal study	17	0.15%	study abroad		8	0.07%
11	the study		17	0.15%	study on		8	0.07%
12		phenomenological study	15	0.13%	study using		6	0.05%
13		critical study	14	0.12%		pilot study	6	0.05%
14	study in		12	0.10%		experimental study	5	0.04%
15		longitudinal study	10	0.09%	to study		4	0.03%
16		preliminary study	10	0.09%		crosscultural study	4	0.03%
17	study to		10	0.09%		empirical study	3	0.03%
18	study and		9	0.08%	for studying		3	0.03%
19	study between		9	0.08%	of study		3	0.03%
20		comparison study	8	0.07%	shortterm study		3	0.03%

Note: If the *RISS* bigram phrase is matched with *Scopus* bigram phrase, it is highlighted in gray.

The % was calculated based 11,639, which is the total instance of each dataset.

By comparing various syntactic elements in titles between the two countries, features related to native usage and non-native usage could be discovered. For example, the phrases “study

for” and “study about” were present 63(0.54%) and 22(0.19%) times respectively in the *RISS* dataset. In light of small percentage, it is worth noting that the number of occurrences in the title, the bigram phrases did occur more than 20 times in the *RISS* dataset. In Table 10, no instances are found in the *Scopus* sample. As indicated in the right side of list, which is *Scopus*, the most appropriate expression is more than likely “study of”. This particularly phrase appears 191 times (1.64%) in the *Scopus* dataset. Thus, expressions such as “study for” and “study about” in the titles are due to the non-native usage of the English language. In turn, although the expressions might be non-native, the cultural characteristics are evident through frequently used words and phrases.

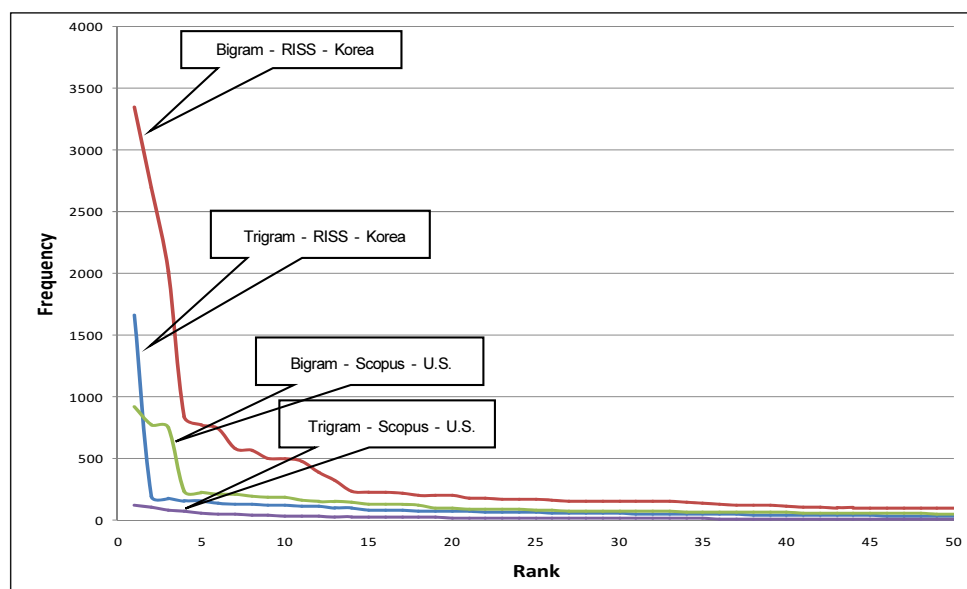
Meanwhile, other bigram phrases that contain the word “study” are used in conjunction with an *adjective* collocated word and could be used to explore and identify the common type of studies that are appropriate for the titles. However, this approach should be adopted with caution since the occurrence of the type of studies in the U.S. is fewer compared to Korea. Based on the frequency count of bigram that contains the word “study”, the sense of native usage to the English language or the norm of discourse community tends to be less popular among authors in the U.S. compared with Korea. An exception to this generality is the use of “case study” as the phrase appeared 77(0.66%) in the *Scopus* dataset. The bigram phrases that are highlighted may possibly indicate the cultural academic phenomenon that is unique to the country since the highlighted bigram phrases are not present in the opposing dataset. Again, a highlighted bigram phrase could still present in the comparing dataset; the highlighted bigram phrase is not present *only* in the top 20 of opposing list of bigram phrases. To increase the number of matched bigram phrases, we can simply increase the number of rank in the opposing list. One can experiment with the list and identify phrases that are more common and rare, etc. For instance, since the bigram phrase “pilot study” is highlighted, it means that the phrase is present only in the *Scopus* dataset and not in the *RISS* dataset. Since “pilot study” could be still present in *RISS* dataset (e.g., if the rank is 21), we could increase the matching threshold value (e.g., rank 50) in matching the bigram phrases to each other.

The drawback of this approach is that while each phrase is checked for its uniqueness against increasing amount of matching pool of list, it is possible to miss some of useful bigram phrases that may be unique to a certain degree. The reason is that the process of gaining more matched phrases actually reduces the bigram phrases that appear to be unique in one dataset.

The frequency count must be considered in the process of exploring the bigram phrase as the rank alone does not indicate the amount of frequently appearing phrases in the dataset.

In sum, the unmatched phrases with high frequency seem to often suggest various areas and topics in the field of social sciences that authors' interested in each country. The earlier examples demonstrate the application of using n-grams in the context of discovering elements of the culture that are present in the form of linguistic expression. When properly analyzed, a collection of titles can show a vast amount of real world expressions that could be used in constructing titles. The approach that was used in this research can also be suitable for suggesting lexical expressions that can be used in a title. This type of strategy and suggestion, particularly developed as a tool, can be useful in particular for non-native English authors, who may lack abilities to select the appropriate vocabularies and phrases for his or her research topic. However, a larger collection of data is required to develop and validate the n-gram based approach described in this study.

Finally, it is useful to compare the distribution of bigram phrases to the distribution of trigram phrases (Finally, comparing the distribution of bigram phrases with the distribution of trigram phrases are useful for this study). Figure 3 depicts the distribution of top 50 frequently used bigram and trigram phrases found in *RISS* and *Scopus* datasets. The y-axis represents the



<Fig. 3> Top 50 Bigram and Trigram Distribution

frequency count of n-grams, while x-axis represents rank pertaining to each n-gram type. As shown, the gap among bigram and trigram frequency counts become narrower as the rank becomes lower. In fact, it has shown that the distribution of n-gram follows Zif's law (Ha, et al. 2002). In the context of this research, differences between these n-gram lines are notable visually. In this figure, the rank of bigrams and trigrams becomes higher in the left side of the figure, the n-gram counts of *RISS* become contrastively higher than of *Scopus*. The fact that frequency of the bigram and the trigram phrases in the *RISS* dataset are substantially higher than *Scopus* datasets suggests that Korean authors tend to use highly common phrases in the construction of journal articles.

## VII. Conclusion

So far, the journal paper titles published in Korean journals and the U.S. journals have been compared by quantitatively analyzing a number of variables: length, title types, and frequently appearing n-gram matchability. This research has shown that various aspects of cultural phenomenon are reflected in the form of the title. By comparing the *RISS* and *Scopus* as two datasets, we showed that various characteristics of title (e.g., the average length) might be dependent on the country in which the article is published.

As a whole, the frequent use of words and phrases in the *RISS* seem to suggest that Koreans tend to be more conservative in choosing words for article title. In contrast, the result seems to suggest that the authors who published in the U.S. journals seem to use a wider range of vocabularies and phrases. Another intriguing question is whether Korean authors are too restrictive when it comes to choosing words and phrases for titles. Seemingly over-used phrases in the Korean academic journals raises the level of proficiency in selecting terms and phrases to articulate and to find a focus for their journal article. Moreover, this raises a question of lexical strengths of titles portrayed in English by the Korean academics. Lexical strengths can be measured in terms of flexibility and intuition (Nakamaru 2010). Lexical flexibility is an ability to have access alternative words, whereas ability to judge the correctness based on listening is referred to as lexical intuition.

Even so, how these ideas are linked to the quality of the journal articles is not clear, as



it is difficult to find answers to this type of question. Still, the frequent use of words such as “study” in the Korean based journal articles suggests that introspection to writing titles in English seems to be necessary as academic community becomes more globalized. To a certain extent, it appears that Korean authors' description of title is quite limited. Also, we have to recognize the fact that there could be a relationship between lexical strengths and articulating in a title. This possible explanation should be further questioned by academic researchers in this area.

Furthermore, of the interest to Korean academics, quantitative information title such as bigrams and trigrams, if extended, can be useful for those who intend to publish in another country. General guidelines and articles related to writing a research paper and a research article title are available. Yet, quantified information is more useful and has a didactic value for those who want to publish in both countries. While some of differences could be viewed as minor and hard to detect, in the peer review process stylistic titles that still accurately portray the content of the article could be helpful.

For instance, recognizing the tendency of author's formulation of research paper titles in the U.S may help authors who want to avoid *cliché* in writing. Thus, the process of discovering alternative expressions may encourage Korean authors to broaden the lexical terms and expressions in constructing the titles. In other cases, by understanding the tendencies of journal articles published in the U.S., Korean authors may want to broaden the styles of the academic discourse. After all, academic communities may not be culturally neutral, showing bias toward another country.

Thus, much information can be useful in assisting authors in the construction of titles, particularly for non-native authors who are less familiar with intricacies of English language.

From the standpoint of improving the quality of academic journals, reviewers of international journals should become aware of cultural elements that tend to exhibit in the form of titles. Korean authors as non-native English authors should make efforts to broaden the appropriate expressions in writing the journal article titles, although the ultimate goal is to use language that is most clear and acceptable in a specific disciplinary community.

Considering all of the above-mentioned points, a number of future areas for research can be recommended. First, further research is needed to confirm the finding in this study since the sample size of the dataset still appears to be small in applying n-gram based approach.

Second, additional countries and languages can be examined further in order to validate some of findings in this research. Investigating characteristics of titles in other countries can aid in finding commonly used linguistic expressions as whole. Finally, a study to examine the equivalency of titles in Korean to titles written in English can be carried out as the degree to which the Korean influences authors in writing titles in English is still not certain.

In a nutshell, this paper has shown that there is a considerable amount of culturally related variability in constructing the title of journal articles. Future studies related to titles should be more sensitive to cultural elements. Generalizing the title related research in the context of nationality should be more appropriately mentioned in order to avoid making over-generalized statements in regards to journal article titles.

## References

- Anthony, Laurence. 2001. "Characteristic Features of Research Article Titles in Computer Science." *IEEE Transactions, on Professional Communication*, 44(3): 187-194.
- Ball, Rafael. 2009. "Scholarly Communication in Transition: The Use of Question Marks in the Titles of Scientific Articles in Medicine, Life Sciences and Physics 1966-2005." *Scientometrics*, 79(3): 667-679.
- Ballesteros, Francisco, Rica, Pedro, Jeff and Joanne Neff. 2005. "The ICLE Error Tagging Project: Analysis of Spanish EFL Writers." *Studies in Contrastive Linguistics*. Proceedings of the 4th International Contrastive Linguistics Conference, Santiago de Compostela. 89-97.
- Buckland, Michael K. 1991. "Information as Thing." *JASIS*, 42(5): 351-360.
- Cianflone, Eugenio. 2013. "Framing Research in Food Science: The State of the Art on Research Articles, Short Communications and Poster Presentations." *BIBLID*, Universidad de Las Palmas de Gran Canaria: Departamento de Filología Moderna, 269-286.
- Dahl, Trine. 2004. "Textual Metadiscourse in Research Articles: A Marker of National Culture or of Academic Discipline?" *Journal of Pragmatics*, 36(10): 1807-1825.
- Dougherty, Dale and Arnold Robbins. 1997. *Sed & Awk*. O'Reilly Media, Inc.

- Frankenberg-Garcia, Ana. 2009. "Are Translations Longer than Source Texts." A Corpus-Based Study of Explicitation In: Beeby, A., Rodríguez P. and Sánchez-Gijón, P.(eds.) *Corpus Use and Learning to Translate (CULT): An Introduction*. Amsterdam and Philadelphia: John Benjamins: 47-58.
- Gesuato, Sara. 2009. "Encoding of Information in Titles: Academic Practices across Four Genres in Linguistics." In Taylor, C. (Ed.) *Ecolingua: The Role of E-Corpora in Translation and Language Learning*, 125-157. Trieste (Italy): EUT.
- Ha, Le Quan, et al. 2002. "Extension of Zipf's Law to Words and Phrases." *Proceedings of the 19th International Conference on Computational Linguistics*, 1. Association for Computational Linguistics. 1-6.
- Haggan, Madeline. 2004. "Research Paper Titles in Literature, Linguistics and Science: Dimensions of Attraction." *Journal of pragmatics*, 36(2): 293-317.
- Hartley, James. 2005. "To Attract or to Inform: What Are Titles For?" *Journal of Technical Writing and Communication*, 35(2): 203-13.
- Hartley, James. 2007. "Colonic Titles." *Journal of the European Medical Writers Association*, 16(4): 147-149.
- Jalilifar, Alireza. 2010. "Writing Titles in Applied Linguistics: A Comparative Study of Theses and Research Articles." *Taiwan International ESP Journal*, 2(1): 29-54.
- Jamali, Hamid R. and Mahsa Nikzad. 2011. "Article Title Type and Its Relation with the Number of Downloads and Citations." *Scientometrics*, 88(2): 653-661.
- Lewis, Grant and James Hartley. 2005. "What's in a Title? Numbers of Words and the Presence of Colons." *Scientometrics*, 63(2): 341-356.
- Mendez, David I., M. Ángeles Alcaraz and Françoise Salager-Meyer. 2014. "Titles in English-Medium Astrophysics Research Articles." *Scientometrics*, 98(3): 2331-2351.
- Nakamaru, Sarah. 2010. "Lexical Issues In Writing Center Tutorials With International and US-Educated Multilingual Writers." *Journal of Second Language Writing*, 19(2): 95-113.
- National Information Standards Organization (NISO). 2004. *Understanding Metadata*. <http://www.niso.org/publications/press/UnderstandingMetadata.pdf>
- Peat, Jennifer, Elliott, Elizabeth, Louise Baur and Victoria Keena. 2002. *Scientific Writing: Easy When You Know How*. BMJ books.

- Rath, Alex. 2010. "Temporal Function of Research Process Nominal Groups in the English Titles of." *English for Specific Purposes World*, 31(10): 1-27.
- Romer, Ute and Ann Arbor. 2009. "English in Academia: Does Nativeness Matter." *Anglistik: International Journal of English Studies*, 20(2): 89-100.
- Sewell, William H. Jr. 2005. "The Concept (s) of Culture." *Practicing history: New Directions in Historical Writing after the Linguistic Turn*, 76-95.
- Soler, Viviana. 2007. "Writing Titles in Science: An Exploratory Study." *English for Specific Purposes*, 26(1): 90-102.
- Subotic, Sinisa and Bhaskar Mukherjee. 2014. "Short and Amusing: The Relationship between Title Characteristics, Downloads and Citations in Psychology Articles." *Journal of Information Science*, 40(1): 115-124.
- Sureka, Ashish and Atul Goyal. 2010. "Insights on Transferability of Dialog-Act Cue-Phrases Across Communication Domains, Modality and Semantically Similar Dialog-Acts." 8th International Conference on Natural Language Processing (ICON). 28-37.
- Thanopoulos, Aristomenis, Nikos Fakotakis and George Kokkinakis. 2002. "Comparative Evaluation of Collocation Extraction Metrics." *LREC*, 2.
- Vinay, Jean-Paul and Darbelnet, Jean. 1958. *Comparative Stylistics of French and English. A Methodology for Translation*. Amsterdam, Philadelphia: John Benjamins.
- Wang, Yan and Yongquan Bai. 2007. "A Corpus-Based Syntactic Study of Medical Research Article Titles." *System*, 35(3): 388-399.
- Wong, Sze-Meng Jojo and Mark Dras. 2009. "Contrastive Analysis and Native Language Identification." Proceedings of the Australasian Language Technology Association Workshop. 53-61.