

A Coh-Metrix Analysis of Language Varieties between the Journal Articles of Chinese and American Scientists

Danmin Ye¹

¹School of Foreign Languages & Cultures, Nanjing Normal University, Nanjing, China

Correspondence: Danmin Ye, School of Foreign Languages & Cultures, Nanjing Normal University, Nanjing, China. Tel: 86-25-8359-8581. E-mail: demiyedm@gmail.com

Received: June 3, 2013 Accepted: June 22, 2013 Online Published: July 17, 2013

doi:10.5539/ijel.v3n4p63

URL: <http://dx.doi.org/10.5539/ijel.v3n4p63>

Abstract

This study presents the systematic language varieties and discourse characteristics that are indicative of the academic writings of Chinese and American scientists. We conduct a Contrastive Corpus Analysis using the computational tool, Coh-Metrix, to identify indicative linguistic features in Chinese science journal abstracts as compared to American science abstracts. The results suggest that Chinese scientists tend to employ different linguistic features from their American counterparts. Specifically, the science abstracts written by Chinese scientists are at a greater level of cohesion, more syntactically difficult, but less abstract in the structure of lexicon compared with those written by American scientists. We conclude that the results may account for the interpretation of Chinese academic writings of English as non-prototypical or Outsiders as opposed to the prototypical model of Insiders in terms of discourse style. This study sheds light on language varieties and methodology that may be helpful to English as a Second Language Learners as well as materials developers in non-English-speaking countries such as China.

Keywords: language varieties, science abstract writing, Contrastive Corpus Analysis, Coh-Metrix

1. Introduction

Early in 1990s, Swales has described English as “Tyrannosaurus rex”: “English as a powerful carnivore gobbling up the other denizens of the academic linguistic grazing grounds” (Swales, 1997). Although seemingly exaggerated, it reflects the preponderance of English as the international language of scholarly publication. English has indeed taken on an increasingly predominant role in the academic world for at least two decades when scientists and researchers seek to publish their findings in the world-class international journals. Since the credo of “Publish or perish” is ubiquitously followed in academia, the English writing of journal articles has drawn greater attention of scientists in different countries especially non-English-speaking countries in that they are often thought at a disadvantage as compared to their English-speaking counterparts (Baldauf & Jernudd, 1983; Gibbs, 1995; Wood, 1997a). As a large proportion of these non-native-English-speaking-researchers, Chinese scientists often find it frustrating to have their research articles resubmitted or rejected for language problems in this specialized area of science journals (Yu & Liang, 2006). Of course, there are many books that are helpful with basic academic writing in a second or foreign language (Tang, 2012). But for the level of scientific journal texts, as McCarthy and colleagues demonstrate, relatively little research has compared the texts of non-native-English-speaking scientists (or *Outsiders* as they are referred to in Min and McCarthy, 2012 in review) to those written by their native-English-speaking counterparts (or *Insiders* as they are referred to in Min and McCarthy, 2012 in review) so as to identify linguistic varieties using computational tools (McCarthy et al. & McNamara, 2009). Recently, however, interdisciplinary methods have been applied to the area of genre analysis, especially with the development of natural language processing (NLP) tools and techniques (Crossley & Louwse et al, 2007; Graesser & McNamara, 2011). These advances have made it possible to computationally analyze and compare different corpora (e.g., Chinese and American corpora in this study) to identify distinct language varieties.

To facilitate the written productions of the *Outsiders*, McCarthy et al. analyzed English articles written by Japanese, British, and American scientists (McCarthy et al. & McNamara, 2009). Their study provided evidence that Japanese authors used significantly more locational and temporal items, high frequency words, high familiarity words, cardinal numbers and higher syntactical similarity in sentence structures as compared to their

American counterparts. Building on this study, Duncan and Hall analyzed the journal articles of three groups of scientists: Americans; Koreans publishing articles in Korea; and Koreans publishing articles in America (Duncan & Hall, 2009). They found that the journal articles of Koreans publishing-in-Korea were the most distinct, and therefore, the least prototypical as compared to the other two groups. Recently Ye and Wang compared the journal texts of Chinese and American scientists using the *Gramulator* (Ye & Wang, 2013). Their findings suggested that Chinese scientists tend to use different *register phenotype* of the agent, the tense, and two different types of reporting verbs. All these studies provide us more confidence to identify the language varieties between the academic abstracts written by Chinese scientists and their American counterparts using computational tools. In the current study, we seek to address the following primary research questions:

1. *Do Chinese scientists employ distinct language varieties in academic science abstracts writings in comparison to a prototypical model from American scientists?*
2. *If so, how different do they use these non-standard language varieties compared with their American counterparts?*
3. *Do the findings of this study support McCarthy and colleagues' (2009) results concerning the comparison among Japanese, American, and British scientists?*

Hypotheses:

1. Chinese scientists employ distinct language varieties in comparison to a prototypical model from American scientists in their academic science abstracts writings.
2. The English writing of Chinese scientists characterize them as the Outsiders in comparison to the native speakers as Insiders.
3. The results of this study support the findings of McCarthy and colleagues' (2009) regarding Japanese, British, and American scientists.

2. Method

2.1 Contrastive Corpus Analysis

The origin of Contrastive Corpus Analysis (Cobb, 2003) can be dated back to the Brown corpus (Kučera & Nelson, 1967) in that its first collection of texts (500) enabled numerous studies, among which the most famous presumably is Biber's (Biber & Reppen, 1998), to understand text types as much by where they overlapped as where they did not. The principle of CCA is that any discourse unit (e.g., text-type, register, genre, variety, or section of text) is best understood, and perhaps only understandable, within the context of its contrast to some other discourse unit (McCarthy, Watanabe & Lamkin, 2012). CCA differs from traditional corpus analyses because it emphasizes on what two (or more) correlative corpora can reveal when their commonalities are excluded by computational and statistical techniques. In the field of second language learning (SLL), Cobb describes CCA as the comparison of two corpora through which what is present and what is not present can be derived. Thus, in two corpora that are highly related but differ minimally (e.g., scientific writing in English by Chinese scientists vs. scientific writing in English by American scientists), the linguistic features that are characteristic of one corpus, but non-characteristic of the sister corpus, is what is *indicative* of the text type.

2.2 Tool: Coh-Metrix

According to Graesser et al., recent developments in information technology and computer-based discourse analysis have "made it possible to computationally investigate various measures of text and language comprehension and supersede surface components of language to explore deeper, more global attributes of language" (Graesser & McNamara, 2011). Among these developments the computational tool Coh-Metrix, capable of measuring textual cohesion and difficulty at various discourse and conceptual analysis levels, is particularly employed in this study to distinguish the language varieties between academic journal abstracts written by Chinese and American scientists. The mechanism of Coh-Metrix is to analyze discourse by integrating "lexicons, pattern classifiers, part-of-speech taggers, syntactic parsers, shallow semantic interpreters, and other components that have been developed in the field of computational linguistics" (Jurafsky & Martin, 2002). Combining these measures into several cohesion metrics, the tool can analyze discourses on various dimensions of textual cohesion "including co-referential cohesion, causal cohesion, and density of connectives, latent semantic analysis metrics, and syntactic complexity" (Graesser & McNamara, 2011). Besides, Coh-Metrix also investigates "several lexical metrics such as word frequency, concreteness, polysemy, word meaningfulness, hyponymy, word age-of-acquisition scores, word imaginability, and word familiarity measures" (Graesser & McNamara, 2011). Taken together, the computational tool Coh-Metrix is capable of measuring deeper level linguistic features of texts that are related to text processing and reading comprehension.

In the current study, we empirically tested the hypothesis that language varieties exist between academic abstracts written by Chinese and American scientists. For this purpose, a corpus of Chinese and American scientific abstracts is collected and the findings from Coh-Metrix analysis are used as a statistical foundation for further classification. In general, this study combined computational, corpus, and statistical approaches to examine whether language varieties can be used to distinguish Chinese texts from American texts and what useful information about different preference of employing linguistic features can be gained from these findings.

2.3 The Corpus

Our corpus comprises 672 abstracts taken from 31 science journals published in either China or the United States respectively. The contents cover three genres: the so called “hard sciences” of biology, chemistry, and physics. These journals, all with high impact factors rank in the top five in each area of the three subjects. And all of the articles are published in the last five years (i.e., from 2007 to 2012). In addition, the journals are compiled as parallel as possible to ensure the comparability. For example, we have *Chinese journal of Inorganic Chemistry* in the Chinese corpus in parallel with *Inorganic Chemistry* in the American corpus. From these texts, two individual sub-corpora were compiled: (1) Chinese scientists in China (CSC) and (2) American scientists in America (ASA). The Chinese English corpus comprises Chinese scientists’ abstracts ($n = 335$), published exclusively in 15 different Chinese journals. The American English corpus (the assumed prototypical model) comprises U.S. scientists’ abstracts ($n = 337$), published exclusively in 16 U.S. journals.

To ensure the original nationalities of the authors, the model of McCarthy and colleagues was followed (McCarthy et al. & McNamara, 2009) (see also Min & McCarthy, 2012 in review and Duncan & Hall, 2009). This model has two major criteria: (1) the first author (generally the person who writes most of the paper or leads the projects in the field of science) and the last author (generally the supervisor) should be from universities or institutes within the same country (e.g., in this study, the first and the last authors of the Chinese English and the American English corpora should be from the Chinese and the American universities or institutes respectively). (2) The names of the primary and final authors must be ‘typical’ of the country of the classification. That is, the primary and final authors in the Chinese and American corpora represent the typical names for Chinese and Americans respectively. Of course, this model cannot always ensure the authenticity of the authors’ nationalities, but these criteria of classification are effective in determining the language backgrounds of the writers. For the classification of Chinese authors, it is not hard to do the task because the first author of this study is Chinese and the names are always written in Chinese characters in Chinese science journals. For the classification of the American authors, we ensure that both the first and the last authors are working in American universities or institutes, which means that they are American-based authors.

Following McCarthy and colleagues (McCarthy et al. & McNamara, 2009), the current study focuses on the abstracts of journal articles written by Chinese and American scientists to assess the language variety of each corpus. As a unique section of the discourse structure, science abstracts are the first reviewed and most frequently read part of the journal articles. They are also representative of the entire research, always available on internet, and easy to collect. These points make science abstracts a reasonable point of departure for the current study.

2.4 Data Analysis

To test the hypothesis that language varieties exist between academic abstracts written by Chinese and American scientists, we conducted a discriminant functional analysis. Considering the upgrade and development of Coh-Metrix, we selected a different set of measures from McCarthy and colleagues’ research (McCarthy et al. & McNamara, 2009). In the present study we reduced the large number of the Coh-Metrix measures to a more manageable set. According to Graesser et al, the 53 Coh-Metrix measures were grouped into those related most to words, sentences, and connections between sentences (Graesser & McNamara, 2011). Graesser et al conducted a Principal Component Analysis (PCA) on the TASA (Touchstone Applied Science Association) corpus, yielding five principle components that explained an impressive 67.3% of the variability among 37,520 texts in the corpus (Graesser & McNamara, 2011). These are: *narrativity*, *referential cohesion*, *syntactic simplicity*, *word concreteness*, and *situation model cohesion* (corresponding to *causal cohesion*, *verb cohesion*, *logical cohesion*, *temporal cohesion*). The Principle Components map quite favorably onto the levels of the multilevel theoretical framework articulated by Graesser and McNamara (Graesser & McNamara, 2011).

To select the variables from the five chosen PCs, we follow the common practice of corpus data investigation (Biber, 1993; McCarthy & Boonthum, 2012; McNamara, Graesser & Louwse, 2012). Specifically, we randomly divided the texts of each corpus into two-thirds training-set data (Chinese: 224 texts; American: 226 texts) and one-third test-set data (the remaining 111 texts for each corpus). We divided the training-set to identify

which of the variables contained in the chosen Coh-Metrix indices best classified the Chinese from the American scientific abstracts. We then used these selected variables to predict the Chinese from the American scientific abstracts in the training-set by setting up a model of discriminant function analysis, a common method to distinguish text types in genre analysis that is employed in many previous researches (e.g., Crossley, McCarthy & McNamara, 2007; Crossley & McNamara, 2009; McCarthy et al. & McNamara, 2009). The abstracts in the test-set data were later processed through the discriminant function analysis model.

3. Results

3.1 Results

A repeated measure of MANOVA was performed to investigate the potential effect of the different writers (e.g., Chinese and American scientists) on the five Coh-Metrix measures. Descriptive statistics on these measures are shown in Table 1.

Table 1. Descriptive Statistics M (SD) of Measures by countries in the training-set

Measures	Country			
	China		U.S.	
Narrativity	-1.770	(.685)	-1.826	(.571)
Referential Cohesion	.437	(1.215)	-.144	(1.151)
Syntactic Simplicity	-1.404	(1.155)	-.838	(.788)
Word concreteness	.078	(1.354)	-.233	(1.029)
Situation Model Cohesion	-.563	(1.188)	-.677	(1.124)

Post-hoc tests with Bonferroni correction were conducted to identify significant ($p < .05$ for all analyses unless specified otherwise) differences across countries. Such a test highlights the degree to which the relative genres of writing differ and the direction of those differences (see Table 2).

Table 2. Bonferroni *Post-hoc* Analysis Showing Direction of Differences between Chinese and American writers in the training-set

Dependent Variable	Mean Difference (Chinese-American)
Narrativity (ZNar)	.055
Referential Cohesion (ZRef)	.581**
Syntactic Simplicity (ZSyn)	-.566**
Word concreteness (ZConc)	.311*
Situation Model Cohesion (ZSit)	.114

Note: ** $p < .001$; * $p < .05$

The results are discussed in the following section as to the information they provide about the differences between the Chinese and American writers.

Narrativity (ZNar)

Narrativity describes the extent to which the text conveys a story, a procedure, or a sequence of episodes of actions and events with animate beings. Informational texts on unfamiliar topics, typically appearing in print, are at the opposite end of the continuum (Graesser & McNamara, 2011). The negative value of both Chinese and American abstracts demonstrates that the scientific abstracts have the lowest frequency in terms of *narrativity*. The very slight mean difference between the Chinese and American writers indicates that both of the scientists in two countries tend to invariably employ an informational way to present their findings in abstracts. This result

also confirms the consistency of our research and the PC *ZNar*. Precisely, the scientific abstracts generally have a clear distinction from those story-like genres characterized by higher frequency of narrativity and greater ease of comprehension.

Referential Cohesion (ZRef)

Referential cohesion captures the extent to which explicit content words and ideas in the text are connected with each other as the text unfolds. The noun phrases are important for providing co-reference and bridging the explicit clauses and sentences in the textbase. A referential cohesion gap occurs when a sentence has few if any words that overlap with previous sentences (Graesser & McNamara, 2011). The pairwise comparison between Chinese and American scientists in referential cohesion suggests that abstracts written by Chinese scientists are at a significantly higher level of cohesion as compared with their American counterparts. That is, more noun phrase overlaps are employed to achieve significantly higher cohesion in abstracts written by Chinese writers. The result is similar to those of McCarthy et al and Hinkel which also found greater cohesion in NNS texts (Hinkel, 2001; McCarthy et al. & McNamara, 2009).

Syntactic Simplicity (ZSyn)

Scores are higher when sentences have fewer words and simpler, more familiar syntactic structures. At the opposite end of the continuum are structurally embedded sentences that require the reader to hold many words and ideas in working memory (Graesser & McNamara, 2011). The finding of our current study reveals that Chinese scientists employ significantly more difficult syntax when compared with American scientists. They tend to produce more complex syntactical structures in their writing of scientific abstracts. In respect to genre, American writers usually use simpler syntax on informational texts because readers have lower prior knowledge. In other words, simpler syntax is usually employed to compensate for the more unfamiliar and challenging subjects in informational texts (Graesser & McNamara, 2011). Our findings suggest that Chinese scientists are obviously less aware of this trade-off in terms of genre. They use difficult grammatical structures to present ideas for challenging subjects, which characterizes their writing as non-prototypical (or Outsiders) as compared with their American counterparts (or Insiders). This result is contrary to the findings of McCarthy et al (McCarthy et al. & McNamara, 2009) in that Japanese scientists tend to avoid using difficult grammatical structures. It also suggests that Chinese writers have a better control of syntactical knowledge than Japanese writers.

Word concreteness (ZConc)

Scores are higher when a higher percentage of content words are concrete, are meaningful, and evoke mental images—as opposed to being abstract (Graesser & McNamara, 2011). The result generated from word concreteness demonstrates that Chinese writers appear to use significantly more concrete words when compared with American writers. It also indicates that Americans employ significantly more words that are hierarchically connected and are conceptually more abstract than Chinese. Therefore we can conclude that American scientists are not as concrete in their writings as Chinese scientists in terms of lexicon. Instead they may use higher frequency of word hypernymy and greater lexical connections between words.

Situation Model Cohesion (ZSit)

Scores are higher to the extent that clauses and sentences in the text are linked with causal and intentional (goal-oriented) connectives (Graesser & McNamara, 2011). The similarly low scores of situation model cohesion in both Chinese and American scientific abstracts show that informational texts generally are low in deep cohesion because of the difficult and less familiar topics involved in them.

3.2 Accuracy of the Model

We conducted a series of discriminant function analysis to test the accuracy of the findings. A discriminant analysis is a statistical procedure that is capable of predicting group membership (in this case Chinese and American language categories) using a series of independent variables (in this case the five measures of PC scores). The training-set is used to generate our discriminant function which acts as the algorithm that predicts group membership. Then the discriminant function analysis model generated from the training-set is applied to the test-set to predict group membership of the texts. If the results of the discriminant analysis are statistically significant, we could conclude from the findings that they support the predictions of the analysis that different language varieties exist between scientific abstracts written by Chinese and American scientists and that those differences can be used to identify texts from different language categories.

In this analysis, results demonstrate that the discriminant function successfully differentiate the two language groups. Specifically, using the combination of the 5 variables in the training-set correctly predicted 286 texts out of the entire 450 texts ($df = 2, n = 450$) $\chi^2 = 44.691, p < .001$). And in the test-set 163 texts are allocated correctly

out of the entire 222 texts ($df = 5, n = 222$) $\chi^2 = 75.612, p = .001$). As is typical of the discriminant function analysis, the estimation of the accuracy of the results is reported in terms of *recall* and *precision*. *Recall* is the number of *hits* (correct predictions) divided by the number of *hits* + *misses* (true items), while *precision* is the number of *hits* (correct predictions) over *false alarms* (incorrect predictions). This is important because if an algorithm predicts everything to be a member of a single group, the recall will be 100% but the precision will score poorly. Reporting both *recall* and *precision* values helps a better understanding of the accuracy of the model. In this study the accuracy of the model for predicting Chinese-English texts in the test-set was 71% (recall = 71.17%, precision = 70.23%). The accuracy of the model for predicting American-English texts in the test-set was 76% (recall = 75.68%, precision = 76.12%). These results reveal that there are significant differences between scientific abstracts written by Chinese and American scientists.

4. Discussion

In this study, we assess whether Chinese scientists employ distinct language varieties in academic science abstracts writings in comparison to a prototypical model from American scientists. By using the computational tool, the Coh-Metrix, we discuss and assess the language varieties from the perspectives of cohesion, syntax, and lexicon. Collectively our results indicate that the Chinese scientists tend to use more formal style and non-standard varieties of English by employing more difficult syntax and more referential cohesion while avoid using abstract words. These findings also characterize the Chinese-English writing of scientific abstracts as the Outsiders when compared with those of Insiders. That is, in Chinese-English texts, the higher level of cohesion often reflects Chinese scientists' attempts to construct a unified idea flow within the constraints of a limited lexical proficiency. They also use more complicated grammatical structures to present ideas while unknowing the common trade-off between syntax and genre.

This study addressed the three primary research questions: 1) *Do Chinese scientists employ distinct language varieties in academic science abstracts writings in comparison to a prototypical model from American scientists?* 2) *If so, how different do they use these non-standard language varieties compared with their American counterparts?* 3) *Do the findings of this study support McCarthy and colleagues' (2009) results concerning the comparison among Japanese, American, and British scientists?* Addressing the first question, our response is that Chinese scientists appear to employ distinctive language varieties in academic scientific abstracts writings which characterize them as non-prototypical or Outsiders as compared to their American counterparts. To answer our second question, our response is that Chinese scientists use more formal style and non-standard varieties of English by employing more referential cohesion and more difficult syntax while avoid using abstract words. Addressing the last question, we find that this study supports McCarthy and colleagues' findings of Japanese scientists (McCarthy et al. & McNamara, 2009) in that they both employ non-standard varieties of English in academic writing of science abstracts. But our findings differ from theirs for the Chinese scientists appear to employ more difficult syntax in their science abstract writings. It also suggests that Chinese writers have a better control of syntactical knowledge than Japanese writers.

Although our study provided these findings, future research needs to address the breadth of "standard" and "non-standard varieties". Future experiments also need to be done to discuss what changes should be made regarding linguistic features in the specific context. Moreover, future researches also need to address whether changes made in terms of the cohesion, syntax, and lexicon in Chinese-English scientific abstracts has a positive effect on reviewers and the subsequent success of publication.

Acknowledgements

This research was supported by the Project of Jiangsu Graduate Students Scientific Research and Innovation (CXZZ11_0864), and by the Priority Academic Program Development of Jiangsu Higher Education Institutions (20110101). The authors also acknowledge Nia Dowell, John Myers, and Dr. Art Graesser at the Institute for Intelligent Systems, University of Memphis, United States of America, for their help during the writing process of this paper.

References

- Baldauf, R. B., & Jernudd, B. H. (1983). Language of publications as a variable in scientific communication. *Australian Review of Applied Linguistics*, 6, 97-108.
- Biber D. (1993). *Register Variation and Corpus Design, Computational Linguistics*. Cambridge: Cambridge University Press.
- Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus linguistics: Investigating language structure and use*. Cambridge, UK: Cambridge University Press. <http://dx.doi.org/10.1017/CBO9780511804489>

- Cobb, T. (2003). Analyzing Late Interlanguage with Learner Corpora: Québec Replications of Three European Studies. *The Canadian Modern Language Review/La Revue canadienne des langues vivantes*, 59(3), 393-423. <http://dx.doi.org/10.3138/cmlr.59.3.393>
- Crossley, S. A., & Louwse, M. et al. (2007). A linguistic analysis of simplified and authentic texts. *Modern Language Journal*, 91, 15-30. <http://dx.doi.org/10.1111/j.1540-4781.2007.00507.x>
- Crossley, S. A., & McNamara, D. S. (2009). Computational Assessment of Lexical Differences in L1 and L2 Writing. *Journal of Second Language Writing*, 18, 119-135. <http://dx.doi.org/10.1016/j.jslw.2009.02.002>
- Crossley, S. A., McCarthy, P. M., & McNamara, D. S. (2007). Discriminating between second language learning text-types. In D. Wilson, & G. Sutcliffe (Eds.), *Proceedings of the 20th International Florida Artificial Intelligence Research Society Conference* (pp. 205-210). Menlo Park, California: The AAAI Press
- Duncan, B., & Hall, C. (2009). A coh-matrix analysis of variation among biomedical abstracts. *Proceedings of the twenty-second International Florida Artificial Intelligence Research Society Conference* (pp. 237-242). Menlo Park, California: The AAAI Press.
- Gibbs, W. (1995). Trends in scientific communication: Lost science in the third world. *Scientific American*, August, 76-83.
- Graesser, A. C., & McNamara, D. S. (2011). Computational analyses of multilevel discourse comprehension. *Topics in Cognitive Science*, 2, 371-398. <http://dx.doi.org/10.1111/j.1756-8765.2010.01081.x>
- Guignon, C. B. (1998). Existentialism. In E. Craig (Ed.), *Routledge encyclopedia of philosophy* (Vol. 3, pp. 493-502). London, England: Routledge.
- Hinkel, E. (2001). Matters of Cohesion in L2 Academic Texts. *Applied Language Learning*, 12(2), 111-132.
- Jurafsky, D. S., & Martin, J. H. (2002). *Speech and language processing*. Englewood, NJ: Prentice Hall.
- Kubrick, S. (Director). (1980). *The Shining* [Motion picture]. United States: Warner Brothers.
- Kučera, H., & Nelson, F. (1967). *Computational Analysis of Present-day American English*. Brown University Press
- McCarthy, P. M., & Boonthum, C. (Eds.). (2012). *Applied natural language processing and content analysis: Identification, investigation, and resolution*. Hershey, PA: IGI Global
- McCarthy, P. M., Hall, C., Duran, N. D., Doiuchi, M., Duncan, B., Fujiwara, Y., & McNamara, D. S. (2009). A Coh-Matrix Analysis of Discourse Variation in the Texts of Japanese, American, and British Scientists. *The ESPecialist*, 30, 141-173
- McCarthy, P. M., Watanabe, S., & Lamkin, T. A. (2012). The Gramulator: A Tool to Identify Differential Linguistic Features of Correlative Text Types. *Applied Natural Language Processing: Identification, Investigation and Resolution*. IGI Global.
- McLuhan, M. (1970a). *Culture is our business*. New York, NY: McGraw-Hill.
- McNamara, D. S., Graesser, A. C., & Louwse, M. M. (2012). Sources of text difficulty: Across genres and grades. In J. P. Sabatini, E. Albro, & T. O'Reilly (Eds.), *Measuring up: Advances in how we assess reading ability* (pp. 89-116). Plymouth, UK: Rowman & Littlefield Education
- Min, H. C., & McCarthy, P. M. (in review). Insiders and Outsiders: A Gramulator Analysis of the Journal Text of American and Korean Scientists. *TESOL Quarterly*.
- Swales, J. (1997). English language papers and authors' first language: preliminary explorations. *Scientometrics*, 8, 91-101. <http://dx.doi.org/10.1007/BF02025223>
- Tang, R. (2012). *Academic writing in a second or foreign language: Issues and challenges facing ESL/EFL academic writers in higher education contexts* (pp. 38-61). London: Continuum.
- Wood, A. (1997 a). International scientific English: Some thoughts on science, language and ownership. *Science Tribune*, April. Retrieved from <http://www.tribunes.com/tribune/art97/wooda.htm>
- Ye, D. M., & Wang, D. Z. (2013). Language Varieties of the Abstracts in Journal Articles Written by Chinese and American Scholars: A Contrastive Corpus Analysis Using the Gramulator. *International Journal of Applied Linguistics & English Literature*, 2(2), 12-25. <http://dx.doi.org/10.7575/aiac.ijalel.v.2n.2p.15>
- Yu, L., & Liang, Y. G. (2006). A Study of the Writing Model of English Scientific Papers. *Foreign Language Education*, 27(1), 34-37.

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).