Information Visualization Analysis of Academic Development of CSSCI Papers on China English from 2001-2015

Chuanming Yang1,*, Zhiqiang Zhang2, Lixin Liao1

1Northeast Agricultural University, Harbin 150030, Heilongjiang, China
2Anhui Jianzhu University, Hefei 230601, Anhui, China

*Corresponding author(E-mail: yangchuanming1980@163.com)

Abstract
In this paper, we build the database of related literature on China English that compiled by CSSCI (Chinese Social Sciences Citation Index) database from 2001 to 2015. And we use CiteSpace V, the analysis software of information visualization, to analyze inter-citation and the knowledge maps of 181 papers published in the database. What’s more, we respectively plot maps that could reflect the general development trend of China English and the analysis of key-node literatures, nodes journals, the research hotspots and the development trend, which aim to reveal the domestic dynamic development of studies on China English.

Key words: Information visualization, CiteSpace, China English, Cited literature

1. INTRODUCTION
As we all know, as an important carrier of information, academics put more emphasis on how to mine information that behind data quickly and efficiently. As the time for age of big data is coming, the total amount of information has increased greatly and the traditional statistical graphs confined to two-dimensional space have been unable to carry out multi-dimensional, multi-source, polymorphic, large-scale and non-spatial data information. While the emergence of data visualization is a good solution to this problem, because it can help researchers to analyze, understand, memorize data and discover important information hidden. From the perspective of coding and decoding of data, the graphical data is not only the rendering effect of image aesthetics, but also the corresponding decoding process. It can be said that the validity of data visualization is the unity of the validity of the data coding and the image decoding.

On the basis of computer graphics, data visualization cannot only extract, transform, map, highly abstract and integrate the values of information objects, but also map the multi-structure data into concise and recognizable visual symbols, which results in more concrete, understandable and memorable themes. Therefore, data visualization is highly generic, which can help researchers and readers quickly understand complex and multi-source data. Meanwhile, data visualization expresses the content and semantic process of information objects by means of plane figures, interactive animation and virtual simulation images, which enriches the connotation of data images, and effectively excavates, displays and understands the inherent characteristics, relationships and rules that hidden in a large amount of information.

Figure 1 Data visualization

2. DATA VISUALIZATION AND CITESPAC

2.1 Data visualization
Visualization is divided into four categories, including scientific computing visualization, information visualization, knowledge visualization and data visualization. Among them, the weather charts broadcasted on TV and the CT scans done in the hospital are typical cases of scientific computing visualization whose function is to visualize the results of calculations or experiments in daily life. As to information visualization, it is mainly used in visual design, human-computer interaction and so on, which transforms abstract, unstructured and non-geometric data into an understandable form. While there is another kind of schema that refers to knowledge visualization, such as concept maps, knowledge semantic diagrams and causality diagrams. Knowledge visualization means that people use graphic method to represent knowledge, form external manifestations that directly effects on human senses, and then promote the effective dissemination of knowledge. Finally, data visualization is a theory, method and technique that using computer graphics and image processing technology to switch data into graphics or images displayed on the screen and process interactively. In conclusion, data visualization is similar to scientific computing visualization. However, the scope of data visualization is far greater than that generated by scientific computing. Data visualization also includes other types. Data visualization has much richer forms than the other three visualizations. It can be said that data visualization is the integrated application of the above three kinds of visualization.

Figure 2 Data and knowledge

2.2 CiteSpace

CiteSpace is the leading international visualization application developed by Dr. Chen Chaomei at the School of Information Science and Technology at Drexel University, USA. It is suitable for multi-time, time-sharing, and dynamic analysis of complex networks. Java-based CiteSpace program is mainly used for scientific literature data measurement and analysis, identification and display of new trends and developments in scientific development, which has an accurate, convenient and efficient feature in the analysis and visualization of the common network. The CiteSpace menu includes files, items, data, network, visualization and other columns; the display box contains the basic data and software operating results report; and parameter settings include time cutting, term source, term type, node type, map refinement and visualization type. The function buttons in CiteSpace software are mainly Keyword, Cited Author, Cited Journal, Cited Reference and so on. At the same time, there is a variety of ways that maps are displayed in CiteSpace, such as cluster, timeline, and timezone. And researchers, medical scientists, science policy researchers and medical librarians are the main users. CiteSpace integrates information visualization methods, bibliometric methods and data integration methods on the basis of "co-citation" and "cited literature" to map and establish relationships among nodes. In general, this is the software that people use CiteSpace to identify and display the new trends and developments of scientific development so as to find research progress and current research frontier in a subject or a field.
3. OVERVIEW OF STUDIES ON CHINA ENGLISH BEFORE THE 21ST CENTURY

Since English came into China, as a language phenomenon, it has been the concern of the English education sector. In half a century, the usage scope and users of English in the world has been expanding and increasing. It is estimated that there would be a total of two billion people in the world to know English to a certain extent by the end of the twentieth century (Todd 1995: 41). With the rapid internationalization of English, it would inevitably and extensively lead to the nativization of English. What’s more, a variety of regional “English” continued to develop, which played increasingly important roles in the local domestic or external communication.

In the process of nativization of English in China, “China English” became the object of concern and study to many scholars because of its distinctive features and reasons for the development. “China English” could be traced back to the 1960s, when the study of “China English” was often corrected by some customary errors of learners. In the early 1980s, Chuangui Ge put forward “China English” as a serious concept for the first time. Ge pointed out the difference between Chinese English and China English in 1980. And then he indicated that “each country has the actual situation. As for my country, when speaking or writing English there always have something unique to express, both in old or new China.” However, Ge only proposed the concept of “China English” in the article without details. During this period, some scholars began to realize that the “China English” problem was far more complicated than the explanation of habitual errors, and then they pointed out that some “China English” phenomena were inevitable in the process of English translation. In the 1980s, the Chinese scholars did a certain bedding work in studying the formation mechanism of China English when they discussed the error analysis and English-Chinese comparison. While entering into the 1990s, more and more scholars realized that “China English” was an indispensable existence in the world English system, and more and more people were involved in the exploration and research of this problem, which resulted in a large number of research reports published and a variety of new technologies and new ideas applied. Wenzhong Li (1993) suggested that “China English” should be defined as: “to express the unique aspects of Chinese social and cultural fields with the core of Standard English, to communicate in English by means of transliteration, translation and semantic reproduction without the interference from mother tongue, finally to generate vocabularies, sentences and discourses with Chinese characteristics.” And he also advocated calling “Chinese English” for China English. Meanwhile, Yajun Jiang (1995) emphasized the vocabulary and syntax in the paper, who also pointed out characteristics in the phonetic level of China English.

In the 21st century, "China English” has no longer just the research topic to Chinese scholars; on the contrary, it broke through the scope of academic research of language research and became the social problem that gained global attention. Therefore, aiming at the research status of "China English” since the 21st century, we would have a study of data visualization analysis by means of CiteSpace V.

4. A VISUAL ANALYSIS OF THE PROGRESS IN "CHINA ENGLISH" USED BY CITESPACE V

4.1 Selection of materials
We filter data from CSSCI ("Chinese Social Science Citation Index") database, which is developed from the China Social Science Research and Evaluation Center of Nanjing University. CSSCI database follows the rules of bibliometrics, and takes the method of quantitative and qualitative evaluation to select academic and editorial journals from more than 2700 kinds of Chinese humanities and social sciences as the source journals. At present, the database includes more than 500 academic journals in 25 categories that cover economics, management, law, linguistics, pedagogy and so on.

The specific operations of data download include that firstly choosing the advanced search, and the search strategy is: key words = "China English" AND publication time = "2001-2015" AND paper type = "thesis", and then we obtain 181 results.

4.2 Data and discussion

(1) Overall trends and distribution characteristics: Statistics show that there published a total of 181 papers in CSSCI journals with a fluctuant rising trend in general during 15 years. In particular, there had two peaks respectively happened in 2010 and 2013 with 25 literatures; while the lowest point appeared in 2003 with only one literature. And after 2013, it presented to a small fall till in 2015 with 18 papers published. (see Figure 4)

![Figure 4 Trends in China English Research of CSSCI (2001-2015)](image)

(2) Analysis of key-node literatures: The key-node literatures can be carried out by generating a knowledge network of co-autographed authors, which is based on collecting and coordinating data. We use CiteSpaceV software to analyze and process the collated data and construct the co-citation network on China English in order to respectively study the literature in the domestic research field. After selecting the folder that contains the analyzed data, we set the time range from 2001 to 2015 and Years Per Slice is 1, then set the node type to "cited journal" with Top N% per slice. Furthermore, we sort the nodes of each time period by the cited frequency in descending order, and only retain the top 20%. Finally, we click Run to generate a knowledge map with co-citation authors, and Figure 5 is the generated map that based on CSSCI journals, which shows the situation of domestic co-citation authors.

![Figure 5 Key-node Literatures on China English in CSSCI (2001-2015)](image)
As shown in Figure 5, the lines represent the times, and their colors from red to blue are on behalf of the cited times that from short to long, and outer color of each node represents co-citation times of the node in different years. In this figure, the nodes represent the authors’ names and the larger the nodes are, the more co-citation times papers are cited; the lines that connecting with different nodes simultaneously means that the authors are shared, and the coarser the lines are, the more co-citation times and closer relationship the two authors have. In fact, the co-citation frequency of the author can show his centrality degree in the field, and the higher the closer to center, which also means that he plays an important role in the research field and lays a certain foundation for other researchers. Therefore, the literatures written by such kind of authors can be regarded as the key-node ones in the field.

<table>
<thead>
<tr>
<th>Frequency</th>
<th>Centrality</th>
<th>Year</th>
<th>Cited Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>0</td>
<td>2003</td>
<td>Yajun Jiang, FOREIGN LANGUAGE EDUCATION</td>
</tr>
<tr>
<td>5</td>
<td>0.01</td>
<td>2001</td>
<td>Ruiqing Du, FOREIGN LANGUAGE TEACHING AND RESEARCH</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>2003</td>
<td>Qiufang Wen, FOREIGN LANGUAGE LEARNING THEORY AND PRACTICE</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>2002</td>
<td>Narang Yang, JOURNAL OF XIDIAN UNIVERSITY (SOCIAL SCIENCE EDITION)</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>2008</td>
<td>Maocheng Liang, FOREIGN LANGUAGE TEACHING AND RESEARCH</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>2008</td>
<td>AMBRIDGE B, COGNITION</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>2004</td>
<td>JUFFS A, TRANSACTIONS OF THE PHILOLOGICAL SOCIETY</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>2005</td>
<td>JUFFS A, SECOND LANGUAGE RESEARCH</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>2002</td>
<td>Xiaohua Yan, CHINESE SCIENCE &amp; TECHNOLOGY TRANSLATORS JOURNAL</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>2001</td>
<td>Ruiqing Du, FOREIGN LANGUAGE TEACHING AND RESEARCH</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>2006</td>
<td>Shichu Gui, FOREIGN LANGUAGE WORLD</td>
</tr>
</tbody>
</table>

As mentioned in Table 1, the top five key-node literatures are: Foreign Language Education by Yajun Jiang (2003), Foreign Language Teaching and Research by Ruiqing Du (2003), Foreign Language Learning Theory and Practice by Qiufang Wen (2003), Journal of Xidian University (Social Science Edition) by Narang Yang (2002), and Foreign Language Teaching and Research by Maocheng Liang (2008).

We further analyze the historical evolution of the key-node reference (see Figure 6), and focus on the line charts of the above five papers. The result shows that five literatures refer to five writers, who all have the greatest impact on China English studies, and their theoretical views are most widely used. It can be said that these important theories constitute the theoretical basis of China English, which provides important reference and inspiration for the construction of research system of China English in China.

![Figure 6 The historical line Graph of Cited Key-node Papers on China English (2001-2015)](image)
(3) The analysis of nodes journals: We analyze the source journals of literature visually so as to provide a clearer understanding on the important ones of China English. Through the research and positioning of these important journals, researchers in the field can effectively search and manage more professional and advanced research literatures to gain more direct reference. In this paper, the key nodes of journals are analyzed and studied by using the visualization map of literatures. In the process of constructing knowledge map, nodes with magenta iris represent a high centrality which can be regarded as critical ones, while nodes with larger apertures represent journals with more literatures on China English which can be seen as high frequency journals.

After importing ready-processed data into the software, we construct knowledge network of a total of cited journals. Next we select the folder that contains analyzed data, set the time range from 2001 to 2015 and Years Per Slice is 1, then set the node type to "cited journal" with Top N% per slice. Furthermore, we sort the nodes of each time period by the cited frequency in descending order, and only retain the top 30%. Finally we generate the following map:

![Figure 7 The Nodes Journals on China English in CSSCI (2001-2015)](image)

Among them, the key-node journals play important roles in the study of China English, which mainly including Foreign Language Teaching and Research, Modern Foreign Languages, Applied Linguistics, Foreign Language Education, Language Learning, Foreign Languages and Their Teaching, Journal of PLA University of Foreign Languages and so on. The prunus apertures of key nodes in these magazines represent that they have a high degree of centrality and large cited numbers of literature related to China English, which also means a certain influence in the field. Although there has a relatively small number of papers published in the above journals, they contain some influential literature on China English, so these journals are the key-node ones of domestic research. Because of the high comprehensive influencing factors and compound influencing factors, along with the frontier contents, these high frequency journals have great influence in domestic language circles.

(4) The analysis of research hotspots and development trend: After selecting the folder that the data is being analyzed, we set the time range from 2001 to 2015, and Years Per Slice is 1, then set the node type to "Term" AND "Keyword" with Top N% per slice. Furthermore, we sort the nodes of each time period by the cited frequency in descending order and only retain the top 35% to generate Figure 8 finally. In this figure, The larger the nodes are, the more frequently the frequency of the keywords appear; the color and thickness of the inner circles indicate the frequency of occurrence during different time periods; the connections among nodes refer to a co-occurrence relationship, and the thickness means intensity, while the color corresponds to the time at which the node first co-exists; moreover, the change in color from blue to red shows the time from early to recent.

Figure 8 shows that in the past 15 years, there included 31 keywords of research hotspots that related to China English in CSSCI journals, and we selected the top five highest frequency keywords: China English, English Teaching, English Learners in China, Corpus and Second Language Acquisition.

At the same time, we should not only focus on hot words of the study, but also comb from the overall and macroscopic aspects so as to grasp the development and evolution of a discipline or a knowledge field. CiteSpaceV divides the large network of China English into several regions by clustering the co-citation networks, and each region represents a research field. The division of clusters depends on the co-citation relations among the literatures; obviously, a cluster means a kind of co-citation relation. What’s more, each
cluster has its own clear subject. From Figure 8, we can see that China English and English Learners in China are the two common clustering groups, which connected by the English Teaching in the early years. But with the rapid development of the application of corpus technology in the language research in recent years, Corpus has become a new connection between China English and English Learners in China. In Figure 8, the line between China English and Corpus is dark yellow, while between Corpus and English Learners in China is orange. This means that in the era of big data China English promotes the development of corpus and research, which ultimately feeds back to the English Learners in China.

![Figure 8](image)

Figure 8 Research Hotspots on China English in CSSCI (2001-2015)

5. CONCLUSIONS

In this paper, we use the CiteSpace V to analyze the citation data of studies on China English in CSSCI journals (2001-2015), draw the scientific knowledge map on China English studies in nearly fifteen years to clarify the most influential core literature and the key turning point in the process of its evolution, and finally sort out the overall development of China English research. At the same time, we clarify the key words of China English research to a certain extent, and furthermore, we explore and define the new hotspots and development trend of research on China English in recent years, which is based on the latest research that combines studies of China English with corpus.

This paper only makes a preliminary analysis of the knowledge structure of China English by means of co-citation analysis and keywords co-occurrence analysis. There are still many problems that would be worthy of further study.

Acknowledgements

This study was financially supported by 2015 Research Projects on Philosophy and Social Sciences of Heilongjiang Province, Multiple Needs Analysis of General Education in College English (15YYB07); 2016 Research Projects on Educational Reform of the Academic Degree and Postgraduate Education of Heilongjiang Province, the Development of Postgraduates' English Education from the Perspective of Supply-side Reform(JGXM_HLJ_2016074); 2017 SIPT Projects of NEAU, Study on building Heilongjiang International Image and C-E Translation of Publicity Material to Build the International Image(201710224266).

REFERENCES