

Visualization Analysis of Domestic language processing research measurement Based on CiteSpace III

Qin Xu ^{a,*}, Lili Yan ^a

School of Foreign Language, Henan University of Technology, Zhengzhou 450001, China

^a1803691705@qq.com

Abstract: This paper uses the information measurement visualization tool CiteSpace III and uses CSSCI publications related to language processing during from 1998 to 2017 period in the CNKI journals database as data sources to carry out the analysis from research scholars, research institutions, hotspots research direction and research trends. It aims to sort out the development context and trends of the language processing field in China, and provides Chinese scholars with innovative thinking and multiple inspirations for language processing related research.

Keywords: CiteSpace III; language processing; research fronts and hotspots.

1. INTRODUCTION

Language processing and language cognitive processing are the result of linguistic principles and psychological mechanisms (David W. Carroll, 2004). Language processing mainly explores the relationship between language and mind and brain, and its purpose is to understand the neurological and psychological mechanisms of understanding, generation, acquisition, and learning of human language (Cui Gang, 2015). In view of the previous research results, this paper uses CiteSpace III to sort and analyze the literature related to language processing in the past ten years, which provides scientific inspiration for the development of this field.

2. RESEARCH DESIGN

2.1 Data Sources

The research data originated from the CNKI (China Knowledge Network) database. Through searching the CSSCI literature related to “linguistic processing” from 1998 to 2017, 558 actual and valid documents were finally obtained. The data retrieval time is January 6, 2018.

2.2 Research tools and methods

CiteSpace III is a bibliometric tool developed and tested by Prof. Chen Chaomei that detects and displays research progress and cutting-edge hotspots. By analyzing the time-dual relationship between research frontiers and basic knowledge, the paper analyzes the dynamic

evolution path and frontier change trend of knowledge (Chen, 2006). This paper uses CiteSpaceIII to extract data including key words, authors, institutions, and source journals etc. from the 558 documents and conducts citation analysis.

3. ANALYSIS OF DOMESTIC LANGUAGE PROCESSING DEVELOPMENT

3.1 Research Achievements in Domestic Language Processing from 1998 to 2017

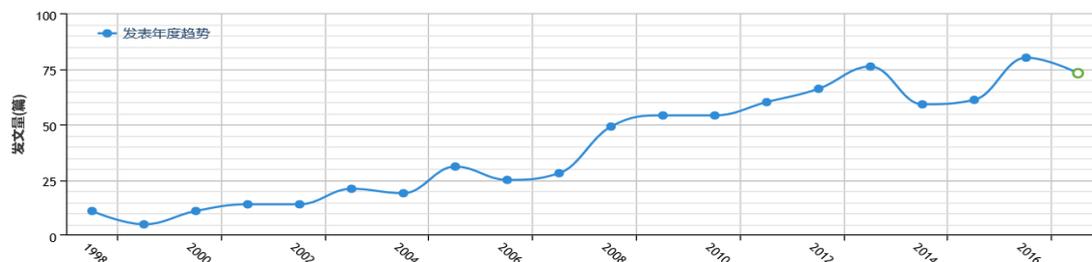


Figure 1 Distribution of domestic language processing research results

Figure 1 shows that the overall domestic research results of language processing between 1998 and 2017 show an upward trend. Although the number of documents declined slightly from 2014 to 2015, it still maintains an average of 60 papers per year. This shows that China's language processing related fields research is constantly being paid attention by researchers.

3.2 Overview of Subjects and Journals in Domestic Language Processing Research

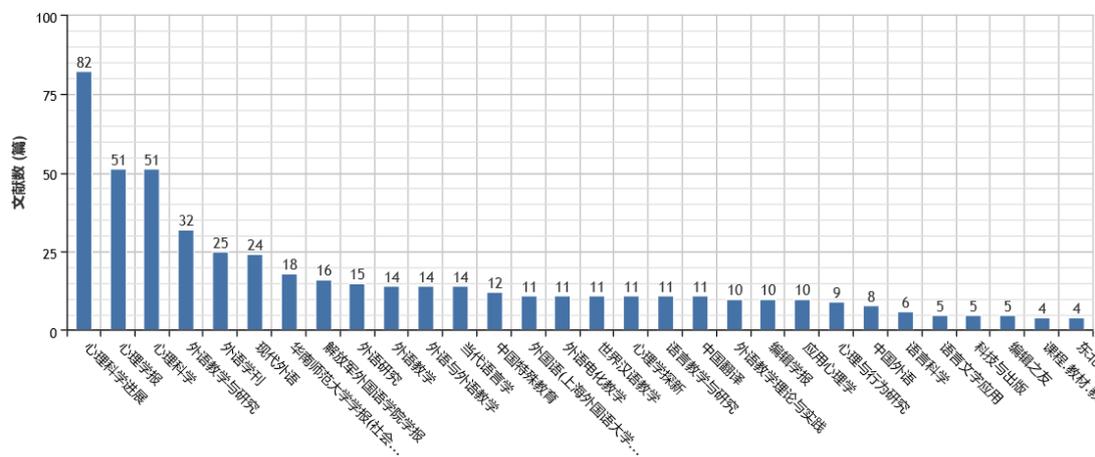


Figure 2 Distribution of Domestic Language Processing Research Journals

In the CSSCI core journals, the research issuing numbers for language processing from high to low is: 1) Psychology periodicals. 2) Foreign language journals. 3) University journals. 4) Editing, translation, science and technology journals. Figure 2 further shows the importance of psychological journals in the study of language processing. However, the gap between "Psychological Science Progress" (67 articles) and "Psychology Science" (31 articles) is almost double, which also explains the research results of the language processing field are unevenly distributed. In addition, the volume of foreign language core journals has also reached nearly 27%, and it also occupies an important position in language processing research.

3.3 Major research institutions

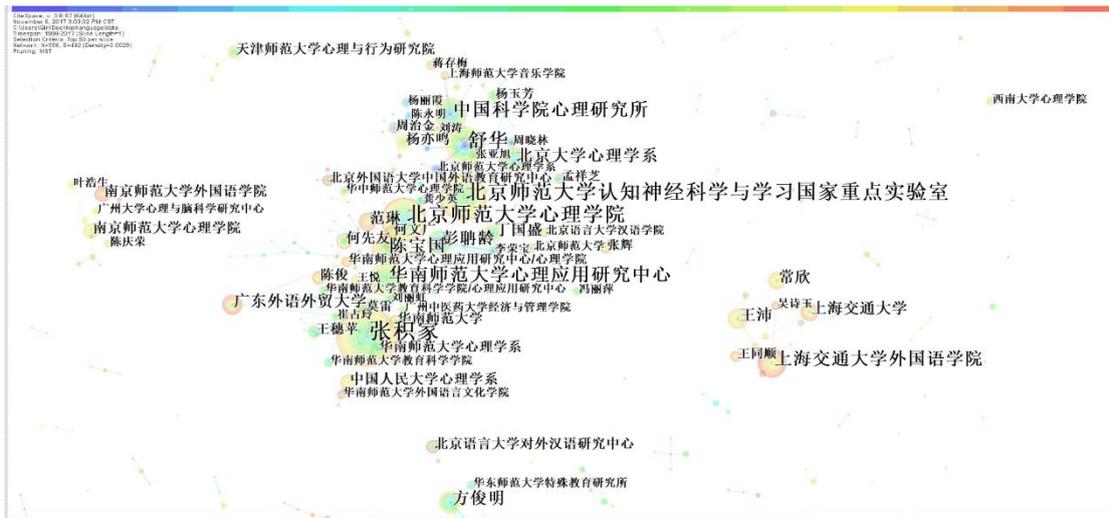


Figure 3 Distribution of domestic language processing researchers and research institutions

Figure 3 is got through using CiteSpace III. According to the clustering network maps, it can present four scientific research positions formed due to the close cooperation between institutions or researchers. The best is Beijing Normal University, the second is South China Normal University, and the third is Shanghai Jiaotong University. The fourth is Nanjing Normal University. The network layout of nodes in the map is relatively loose, indicating that there is less author cooperation and institutional cooperation on language processing research in China. Future researchers and research institutions can enhance collaboration and achieve greater breakthroughs.

3.4 Research fronts and research hotspots



Figure 4 Key words co-occurrence map of domestic language processing research

The keyword is a high refining of the article’s core content and theme. Analysis of changes in the frequency of keyword occurrences can lead to research hotspots in various periods (Chen Chaomei, Chen Yue, Hou Jianhua, 2009). If the same keywords repeatedly occur in the

different article, they can be considered as research hotspots (Liu Xia, Xu Jiabin, 2014). Each node in the map represents a key word. The size of each node represents the frequency of occurrence of the keyword. The year ring of different colors in the node indicates the year of reference is different. The colder the color, the earlier the year and the warmer the color, the earlier the year nearer. The color of the line between the nodes also indicates the year that was first referenced. It can be seen from Figure 4 that the “ERP” node is the largest node in the map and has a strong co-occurrence relationship with other cluster nodes; in turn, it is “working memory”, “event-related potential”, “Cognition”, “language processing”, “N400”, “eye movement”, “syntactic processing”, “second language acquisition”, and “relative clauses”. All of these keywords have higher word frequencies, indicating that these keywords are often hot topics in the field of language processing, and are highly favored by researchers.

Table 1 List of domestic language processing keywords (in order of centrality)

序号	中心性	频数	年份	关键词
1	0.23	28	2003	ERP/事件相关电位
2	0.22	25	2001	工作记忆
3	0.19	22	2007	二语习得
4	0.18	15	2007	认知
5	0.15	14	2008	句子加工
6	0.13	6	1999	信息加工
7	0.12	9	2008	句法加工哦给
8	0.11	11	2008	隐喻
9	0.09	14	2008	关系从句
10	0.08	8	2004	神经机制
10	0.08	10	2012	具身认知
12	0.07	10	2000	语义启动
13	0.06	9	2008	双语者
13	0.06	13	2012	眼动
15	0.05	5	2004	语音加工
15	0.05	9	1998	阅读
15	0.05	5	2000	图式
18	0.04	12	2013	生命性
19	0.03	10	2001	词汇加工
20	0.02	7	2016	心理表征
20	0.02	13	2013	语块
20	0.02	6	2002	发展性阅读障碍

According to Figure 4, this paper classifies and summarizes keywords based on frequency and centrality indicators (Table 1). The key node size is directly proportional to the citation frequency. Comprehensive Figure 4 and Table 1 show that ERP is the largest node and the highest cited frequency. The highest central value is also ERP, which proves that the research on EEG using ERP technology is the research frontier and classic paradigm in this field.

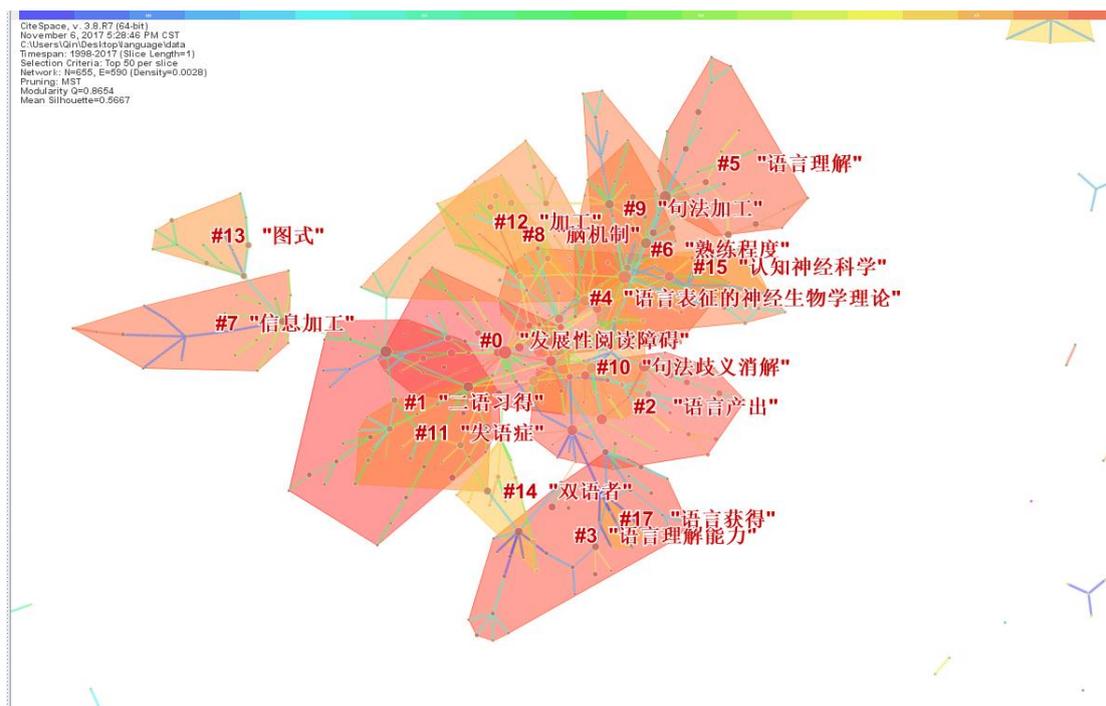


Figure 5 Clustering Atlas of Domestic Language Processing Keywords

The paper uses CiteSpace III to carry out cluster analysis for keyword network (Figure 5). Where the clustering value of Modularity (Q value) is in [0, 1], and the larger the Q value, the better the network clustering effect; the Silhouette (S value) is used to measure the network homogeneity index. The closer the value is to 1, the higher the homogeneity of the network (Li Jie, Chen Chaomei, 2016). In Figure 5, the Q value of 0.7982 shows that the clustering effect of the map is good, the internal nodes of the cluster are more closely linked, and the key points of the cluster nodes are strong. The S value is 0.3791, indicating that the clustering result is reasonable and has a reference value.

The use of the LLR (Log Likelihood algorithm) for naming cluster labels represents the major areas of domestic language processing research from 2007 to 2016. The higher the cluster number, the larger the cluster size and the greater the number of nodes involved. As can be categorized from Figure 5, ERP/fMRI, second language acquisition, working memory, language comprehension processing, verbs, and bilinguals are all hot topics in the field of language processing in the past decade. This is illustrated in Figure 4 and Table 1. The conclusions are basically consistent.

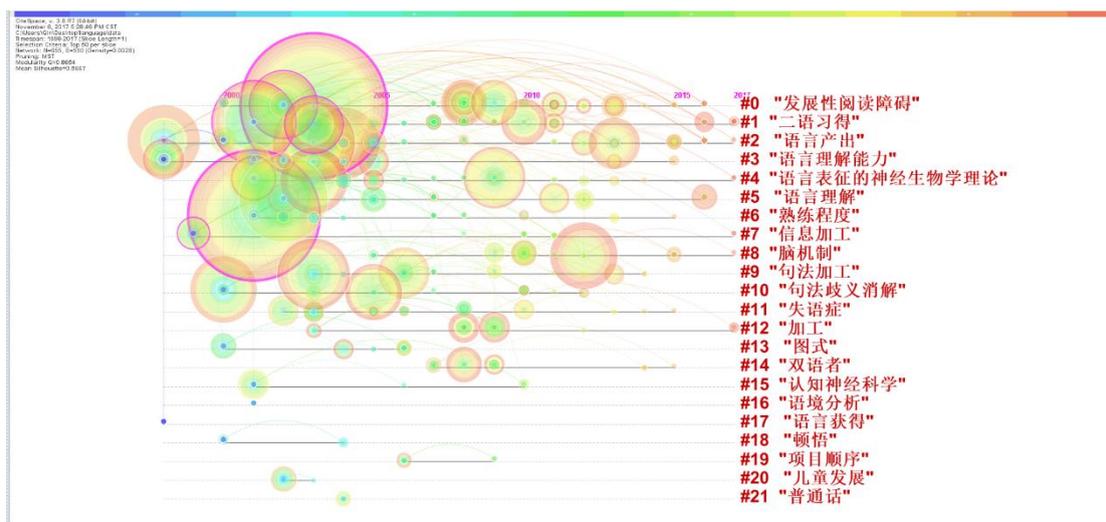


Figure 6 Timeline of the Keyword Web in Domestic Language Processing Research

Using CiteSpace III to visualize keyword clustering through the Time Line view, the evolution of psycholinguistic research can be observed more visually and intuitively. In Figure 6, the keywords clustered on the right side are sorted in order of the size of the cluster. The smaller the number, the larger the cluster size of the keyword and vice versa. The solid line in the figure indicates that the research on the clustering keyword is in an increasing state in the corresponding year; and the dashed line indicates that the research on the clustering keyword is in a decreasing state in the corresponding year.

According to Figure 6, the clustering keywords such as "ERP", "Language Acquisition", "Working Memory", "Language Understanding" and "Language Processing" have not been attenuated and are valued by researchers.

The related research is also rising. Some emerging keywords appear late, so the scale of aggregation is relatively small, and some have not even been clearly expressed in the map, but they have shown strong development momentum, leading the development of language processing in the forefront. For example, "eye movement", "fMRI", "brain mechanism", etc.

4. CONCLUSION

This paper uses the information measurement visualization tool CiteSpace III and uses CSSCI publications related to language processing during from 1998 to 2017 period in the CNKI journals database as data sources to carry out the analysis from research scholars, research institutions, hotspots research direction and research trends.

(1) First, research in the field of language processing in the country is on a steady upward trend and involves a wide range, among which psychology and linguistics are the mainstream disciplines of language processing research; in addition, research results in language processing are unevenly distributed. They are mainly distributed in the psychology journal.

(2) There are many research institutes and researchers. The State Key Laboratory of Cognitive Neuroscience and Learning at Beijing Normal University, the Center for Psychological

Applications at South China Normal University, the School of Foreign Languages at Shanghai Jiaotong University, and the College of Psychology at Nanjing Normal University are the four research gathering places of language processing fields. But various research institutions have been decentralized and have not formed a close network of relationships. In the future, it is necessary to strengthen cooperation and promote development.

(3) Thirdly, the research in the field of language processing presents a diversity of perspectives, a wide range of fields, and advanced methods. However, the slower pace of growth and the use of advanced cognitive neuroscience techniques have contributed to the high level of development of these language processing research paradigms, which will benefit our country's language processing research field is in line with the international frontier.

ACKNOWLEDGMENTS

Research project on humanities and social sciences of Henan provincial department of education, 2018-zzjh-113, *The study on the cognitive mechanism of vocabulary acquisition for college English learners.*

Henan provincial philosophy and social science planning project in 2017, 2017BYY004, *The neurolinguistic research on the acquisition of vocabulary with high level English learners in henan province.*

REFERENCES

- [1] Chen, Chaomei. 2006. CiteSpace II: Detecting and visualizing emerging trends and transient patterns in scientific literature [J]. *Journal of the American Society for Information Science and Technology* 57(3): 359-377.
- [2] David W.Carroll. *Psychology of Language*[M]. Thomson Learning, 2004, 35-36.
- [3] Gazzanig, M. S. *Cognitive Neuroscience (in Chinese)* [M]. Translated by Shen Zheng. 1998, Shanghai: Shanghai Education Press: 3-15.
- [4] Kutas, M. & S. Hillyard. Brain potentials during reading reflect word expectancy and semantic association[J]. *Nature*, 1984, 307 (5947) : 161-163.
- [5] Stillings, N. A., Weisler, S. E., Chase. C. H., Feinstein, M. H., Garfield, J. L., & Rissland, E. L. *Cognitive science: An introduction*[J]. Cambridge, MA: MIT Press, 1995.
- [6] Cai Yanling, A CiteSpace Mapping- Knowledge- Domain Analysis of Psycholinguistics in china [J]. *Journal of Henan University of Technology(Social Science Edition)*,2016
- [7] Chen Chaomei,Chen Rui,Hou Jianhua,Liang Yongxia. CiteSpace II :Detecting and Visualizing Emerging Trends and Transient Patterns in Scientific Literature[J]. *Journal of the China Society for Scientific and Technical Information*,2009.03
- [8] Cui Gang. *Neurolinguistics*[M]. Tsinghua University Press, 2015, 54-56
- [9] LiJie,Chen Chaomei. CiteSpace: Scientific text mining and visualization[M]. Beijing: Capital University of Economics and Business Press,2016
- [10]Liu Xia, Xu Jiajin, Liu Lei. An overview of corpus linguistics research in China(1998-

- 2013):A CiteSpace based analysis[J]. *Corpus Linguistics*,2014
- [11]Zhang Hui, An ERP Study on Syntactic Processing of Second language Learners[J]. *Journal of PLA University of Foreign Languages*, 2014, (1): 88-89.