

# A Bibliometric Analysis of Research Articles on Learner Corpus of English Writing (2012-2021)

Weilu Wang, Jianhong Zhao, Zhiling Wu\*

Foreign Languages College Inner Mongolia University, Hohhot, China

Email: \*imuwzl@163.com

**How to cite this paper:** Wang, W. L., Zhao, J. H., & Wu, Z. L. (2022). A Bibliometric Analysis of Research Articles on Learner Corpus of English Writing (2012-2021). *Open Journal of Modern Linguistics*, 12, 588-604.

<https://doi.org/10.4236/ojml.2022.125044>

**Received:** August 25, 2022

**Accepted:** October 11, 2022

**Published:** October 14, 2022

Copyright © 2022 by author(s) and  
Scientific Research Publishing Inc.

This work is licensed under the Creative  
Commons Attribution International  
License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## Abstract

Ever since the birth of the first computer-read corpus in the 1950s, corpora have been widely used in areas like language research, language teaching, and dictionary and textbook compiling. Given the fact that they pose the potential of reflecting the overall and actual use of learner language and understanding the real difficulties of a certain group of learners, learner corpora, a special form of corpora arose great interest from researchers. With the Citespace program, this paper makes a bibliometric analysis of research articles from the WOS core collection published in the last 10 years (2012-2021). This paper also presents visualizations of the overall research schema, research findings, and prospects.

## Keywords

Learner Corpus, English Writing, Collaboration Network, Clustering, Co-Citation

## 1. Introduction

Language, as a uniquely human ability, is an eternal topic for researchers. The very essential base for any language research could be nothing but language data, and there are three main ways to collect this kind of data: 1) intuition and introspection of researchers for the coming-up of examples of language use; 2) samples taken and surveys conducted for the extraction of language data; 3) questionnaires or inquiries induced for getting insights into language use. As an authentic and massive source of actual language use, a corpus is a databank of written and/or spoken language material processed and stored using computers to provide the most explicit and reliable sources ever for language study. A corpus can be hopefully and convincingly used to describe a language or to verify hypotheses about it. Since the 1980s, corpora have flourished in linguistic re-

search and are widely used in the study of languages and language teaching. In the same vein, questions like what the language of a learner is really like, in what way it varies from one to another, and how it evolves and develops over time, the answers to these questions can be definitely found in a learner corpus. A learner corpus is composed of various written and/or spoken languages by learners, and is pervasive in language teaching and research. Over the years, research on learner corpus has been evolving and upgrading in terms of focus, methodology and perspective.

This is a thriving subject of interest. As more learner corpora are built, corpus-based learner language research is on the rise, with foreign language teaching and second language acquisition as its main research areas (Díaz-Negrillo & Thompson, 2013). Researchers apply learner corpora to aspects of foreign language teaching, including textbook writing, learner dictionary compilation, teaching strategy application, and language testing (Capel, 2010; Gilquin, 2007). Learner corpora are widely used in the study of second language acquisition, and researchers believe that learner language used by a certain group of learners is an important source for the study of second language acquisition (Ellis & Barkhuizen, 2005). Learner corpora are also widely used for comparative studies of learner language with a native corpus (Granger, 2012) or studies on the characteristics of actual language use. In addition, learner corpora are also used for natural language processing or as writing aids, such as automatic recognition and modification of English writing errors (Leacock et al., 2014; Lu, 2011). There still remain numerous big questions to be answered comprehensively and urgently, such as (1) in what way is research in this field developing over time? (2) what are the most fundamental research papers and who are the most influential researchers pushing research in this field forward? (3) what does the future hold for study in this promising area?

Over the years, there has been abundant literature on learner corpus research, especially handbook-type of guides or introductions on the design, typology, methodology, analysis, annotation, and tools to use (Pravec, 2002; Tono, 2003; Granger, 2012; Granger et al., 2015). Research work of this type provides detailed and accessible instruction on the actual realization and use of learner corpora. Granger (2004) made a thorough review on the research of learner corpora in the years, and took a retrospective look at the research accomplished and considered the prospects for future research in both second language acquisition studies and foreign language teaching. The applications soon become the focus of academic attention, especially in second language acquisition and foreign language teaching (Granger et al., 2002; McEnery et al., 2019). Research work of these provides a comprehensive overview of CLC research and its developments and introduces corpus-based approaches to learner language, and learner-corpora applications to teaching. Learner corpus research has been evolving and developing over the past decades, but there is little thorough critical review conducted to summarize and analyze the latest achievements and prevailing trends.

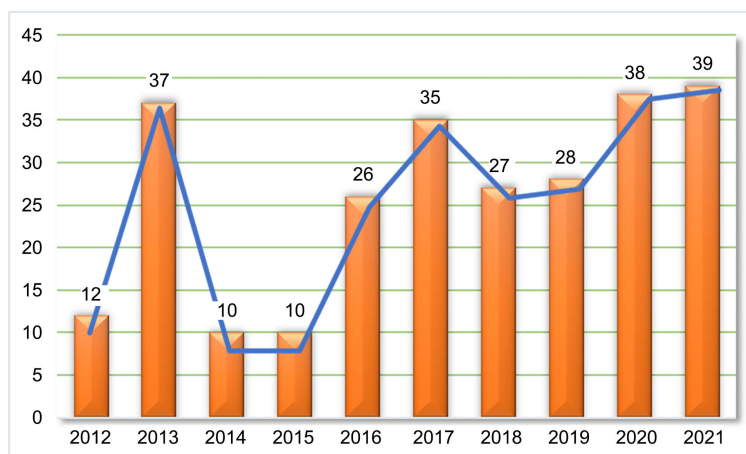
With Citespace, this study is focused on research papers on English learner corpora in WOS core collection over the past 10 years (2012-2021) as the data source for a bibliometric analysis to synthesize research in the past decade, summarize human knowledge accumulated during the period, and clarify future research directions in the field of learner corpora study. Based on the type of data collected and types of analysis available with Citespace (v.5.7. R1), this paper focuses on statistics on publications for the statistical dynamics in research papers published, collaboration network for visualization of collaboration among researchers, institutions and countries, the clustering and evolution of keywords in research papers, and co-citation network for the most fundamental research papers.

## 2. Statistics on Publications

### 2.1. Annual Publications

**Figure 1** shows statistics of the annual publication in WOS core collection in the past decade, that is 2012-2021. In this collection, a total of 187 academic research articles were obtained by a topic search with “English learners corpus OR English learner corpora” as the searching cluster. The range of the year to be searched was limited to the ten years from 2012 to 2021 and the dataset was set as the WOS core collection. The type of literature was restricted to research articles only and research topics were confined to linguistics and educational research. To ensure all the search results are definitely related to the subject matter of this research, clusters like “learner of Russian/French/Spanish/Korean/Chinese and oral/spoken/translation corpus/corpora” are excluded.

As shown in the figure, a general overall upward trend can be noticed, which indicates increasing attention to learner corpus research. During this period, the number of yearly publications is expanding from about 13 in 2012 to 39 in 2021, a sharp threefold growth. Unsurprisingly, the rising attention on learner corpus research coincides with an increasing number of the building of new learner corpora, as a learner corpus may provide new insights and novel data sources for learner language research.



**Figure 1.** Annual Publications 2012-2021.

## 2.2. Publications by Highly-Prolific Authors

**Figure 2** shows the highly-prolific authors for learner corpus research in the WOS core collection, who has published 2 - 4 articles in the past decade, of which the three authors with the most articles published are Sylviane Granger, Kristopher Kyle, and Akira Murakami. And authors with three research articles published are D Joseph Cunningham, Ben Naismith, Minchang Sung, Cassi L Liardet, Fanny Meunier, Kevin W H Tai, and Magali Paquot. Sylviane Granger is one of the leading researchers and founding figures of English learner corpus study, whose research concerns the theoretical basis, methodological innovation, and practical application of learner corpus in language acquisition, research, and teaching. During the decade, her work included two encyclopedic introductory handbooks: *Learner English on Computer* in 2014 and *The Cambridge Handbook of Learner Corpus Research* in 2015. Her numerous research articles covered a wide variety of research topics, including learner corpus design and analysis, learner language features, and distinctions between expert language and learner language. Her latest work was focused on learner corpus for cross linguistic studies, translation studies and data-driven learning, which points to the future direction of learner corpus study. Kristopher Kyle mainly focuses his study on automatic learner language evaluation and assessment including lexical diversity/richness/sophistication, syntactic sophistication, cohesion, sentiment, automatic scoring and feedback. He also gave special attention to longitudinal learner corpus, disciplinary differences in learner academic language, and comparisons of spoken and written language, which point out the possible research directions for learner corpus. Akira Murakami primarily conducted research on L1 influence and individual variation on the acquisition of English, sophisticated statistical modeling for L2 development, task type on complexity and accuracy of learner English, and developmental trajectories of L2 writing strategy.

## 2.3. Publications by Highly-Prolific Institutions

**Figure 3** shows the highly-prolific institutions in the WOS core collection, among which there are Catholic University Louvain, Georgia State University, Guangdong University of Foreign Studies, University California Santa Barbara, University Kansas, Seoul National University, University Cambridge, Georgetown University, most of these research institutes have corpus research centers or institutes with dedicated research teams. At Catholic University Louvain, there is the Research Institute for Language and Communication (ILC). The work of the Linguistics Research Cluster (PLIN) is divided into five areas, including automatic language processing and modern language acquisition, learning and teaching. Among the publications of the institute, there is one called *Corpora and Language in Use*, which aims at publishing research monographs and conference proceedings in the area of corpus linguistics and language in use. At Georgia State University, there is the Language Research Center with “Bio-behavioral Foundations & Development of Cognitive Competence” as one of the many research topics for which language and learning are among the many re-

search subjects. At Guangdong University of Foreign Studies, there is the Center for Linguistics and Applied Linguistics, and learner language and language acquisition are among the research topics of the center.

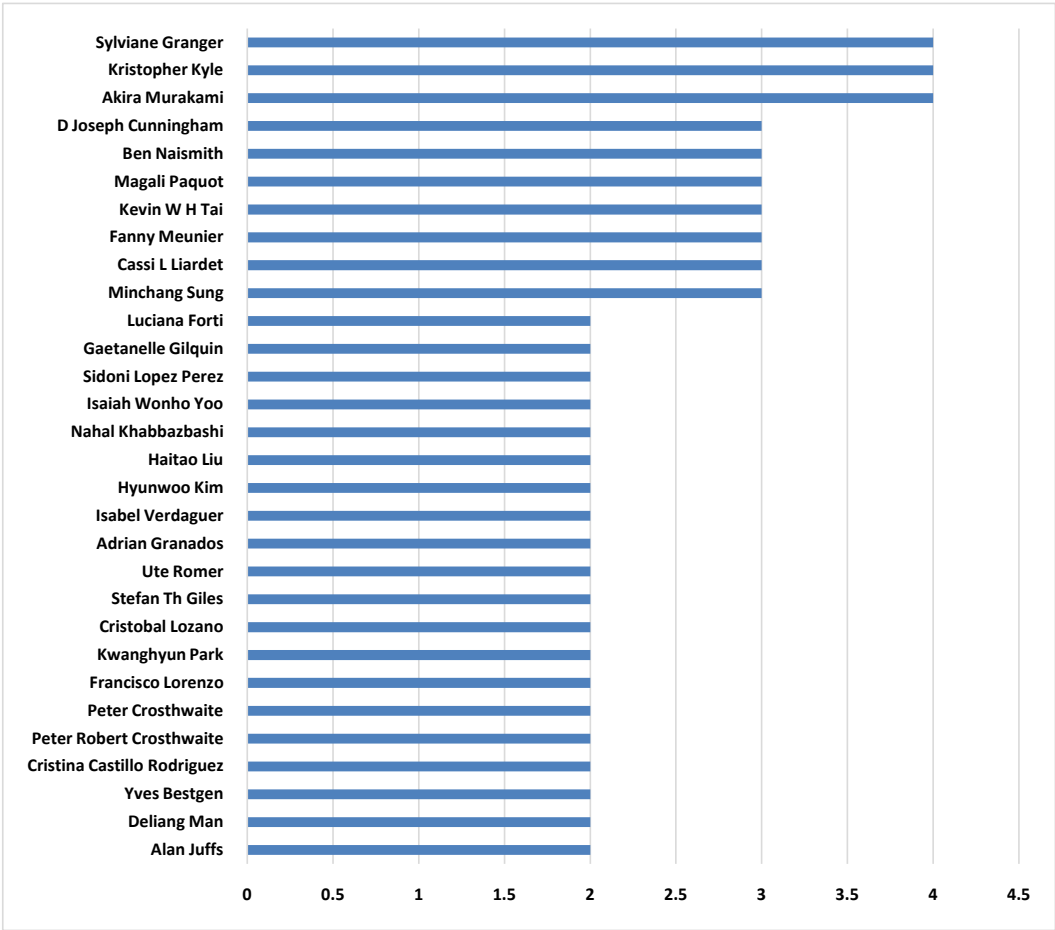


Figure 2. Number of Publications by Highly-prolific Authors.

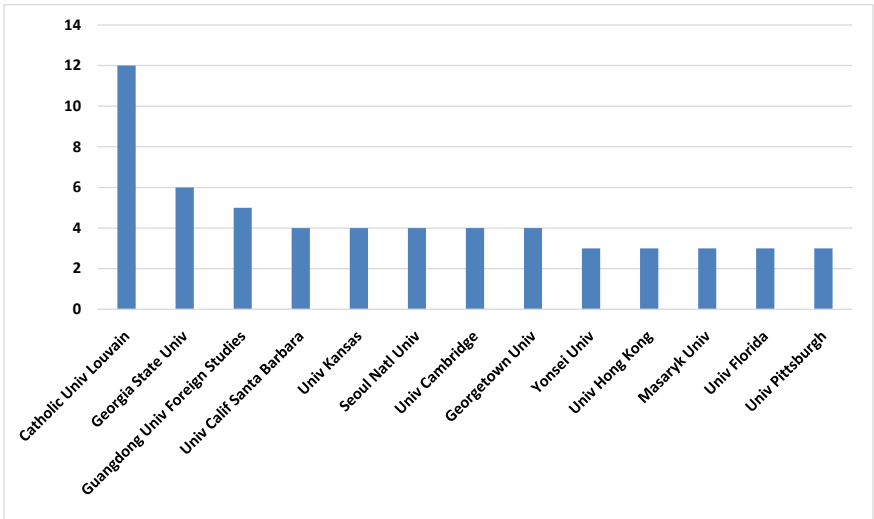


Figure 3. Number of publications by highly-prolific institutions.

### 3. Collaboration Network

#### 3.1. Author Collaboration

**Figure 4** shows the network of author collaboration in the WOS core collection. As shown in the figure, there is a certain simple cooperative relationship between authors (Density = 0.0056, Weighted Mean Silhouette  $S = 0.3739$ ), but this collaboration is mainly based on geographical and academic relations. The reason is that humanities and social science research are mostly carried out in small-scale research teams, while geographical availability or academic relations are more likely to form small-scale research teams, and international collaboration is rare.

Judging from centrality, the first in line is Fanny Meunier, whose centrality value is 5; the second place is taken by Kristopher Kyle, Akira Murakami, and Isabel Verdaguer with a centrality value of 4; next come Nina Vyatkina, Sylviane Granger, Hyunwoo Kim, Peter Crosthwaite, Fredrik Markowitz and Hubert Naets with a centrality value of 3. The above authors are in an important position in the cooperative network and play a major role in the collaboration. In the case of Fanny Meunier, her collaborators includes Gaëtanelle Gilquin, Sylviane Granger, Magali Paquot, Kristel Van Goethem, Isa Hendrikx, of whom the first four are all faculty of UC Louvain or members of research institute there. The last one Isa Hendrikx is a PHD student of UC Louvain. They worked together as research collaborators because they were working together in the same institution. Approachability and availability are the determining factors shaping collaboration in learner corpus research.

#### 3.2. Institution Collaboration

**Figure 5** shows the institution collaboration network in the WOS core collection. Judging from centrality, the first is Catholic University Louvain with a centrality value of 5; next come Georgia State University, University Lancaster, National Research University, Radboud University Nijmegen with a centrality value of 3; followed by University of California Santa Barbara, with a centrality value of 2. These institutions are prominent in the network of collaboration. The network here is not really well-formed (Density = 0.0052, weighted mean Silhouette  $S = 0.3739$ ), only sparse collaboration networking can be noticed. And geographical approachability is the main driving force for institutional collaboration.

#### 3.3. International Collaboration

**Figure 6** shows the international collaboration network of the English learner corpus research papers in the WOS core collection. The largest node in the figure is the USA, whose contribution value is 41; the second is Spain, 25, next comes England, 18. The fourth is PRC, 14; and the fifth is South Korea and Belgium, 13; followed by Germany, 9; Taiwan (China), 8; Czech Republic and Italy,



6. In terms of centrality, from high to low: Spain (centrality = 10), USA (centrality = 9), Germany (centrality = 9), Netherlands (centrality = 6), England (centrality = 5), PRC (centrality = 5), Belgium (centrality = 5), Sweden (centrality = 5), Norway (centrality = 5), Malaysia (centrality = 5). The above countries (regions) play an important role in relevant international collaboration; they are the driving forces pushing ahead learner corpus research and collaboration. The sparse network lines indicate that further efforts are needed to promote international collaboration in this research field.



Figure 4. Author collaboration network.

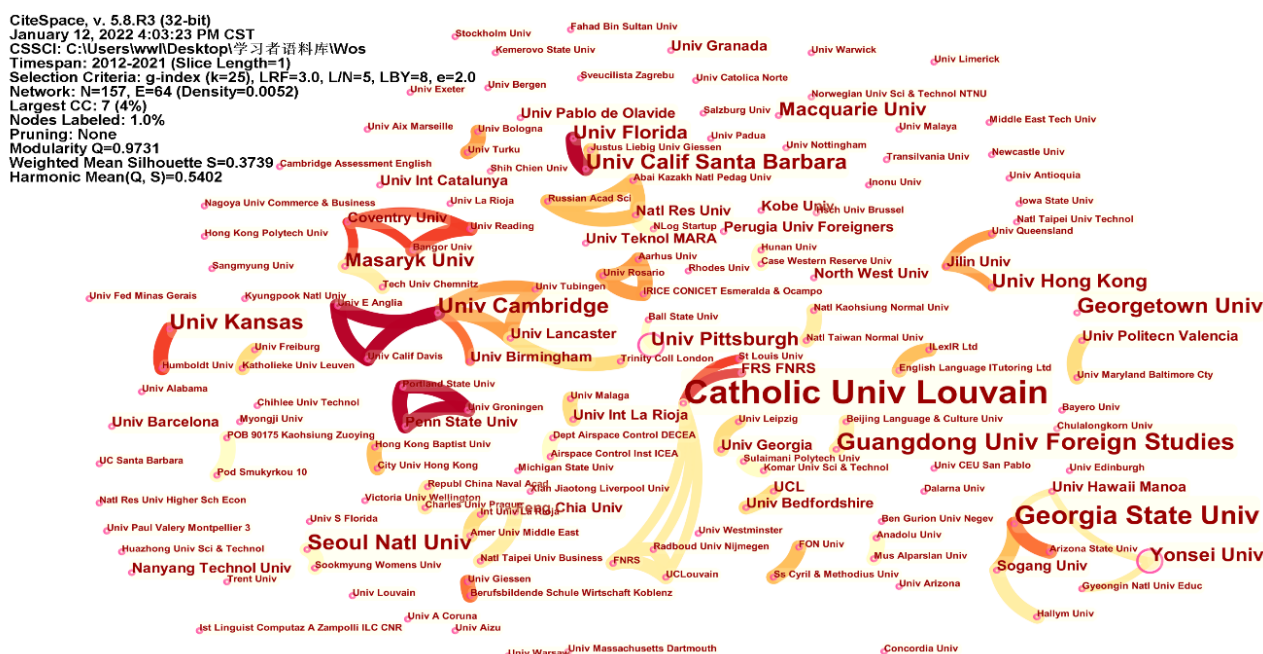


Figure 5. Institution collaboration Network.

CiteSpace, v. 5.8.R3 (32-bit)  
 January 11, 2022 10:05:20 AM CST  
 WoS: C:\Users\wwl\Desktop\学习者语料库\Wos  
 Timespan: 2012-2021 (Slice Length=1)  
 Selection Criteria: g-index (k=25), LRF=3.0, L/N=10, LBY=5, e=1.0  
 Network: N=259, E=492 (Density=0.0147)  
 Largest CC: 183 (70%)  
 Nodes Labeled: 1.0%  
 Pruning: None  
 Modularity Q=0.9555  
 Weighted Mean Silhouette S=1  
 Harmonic Mean(Q, S)=0.9772

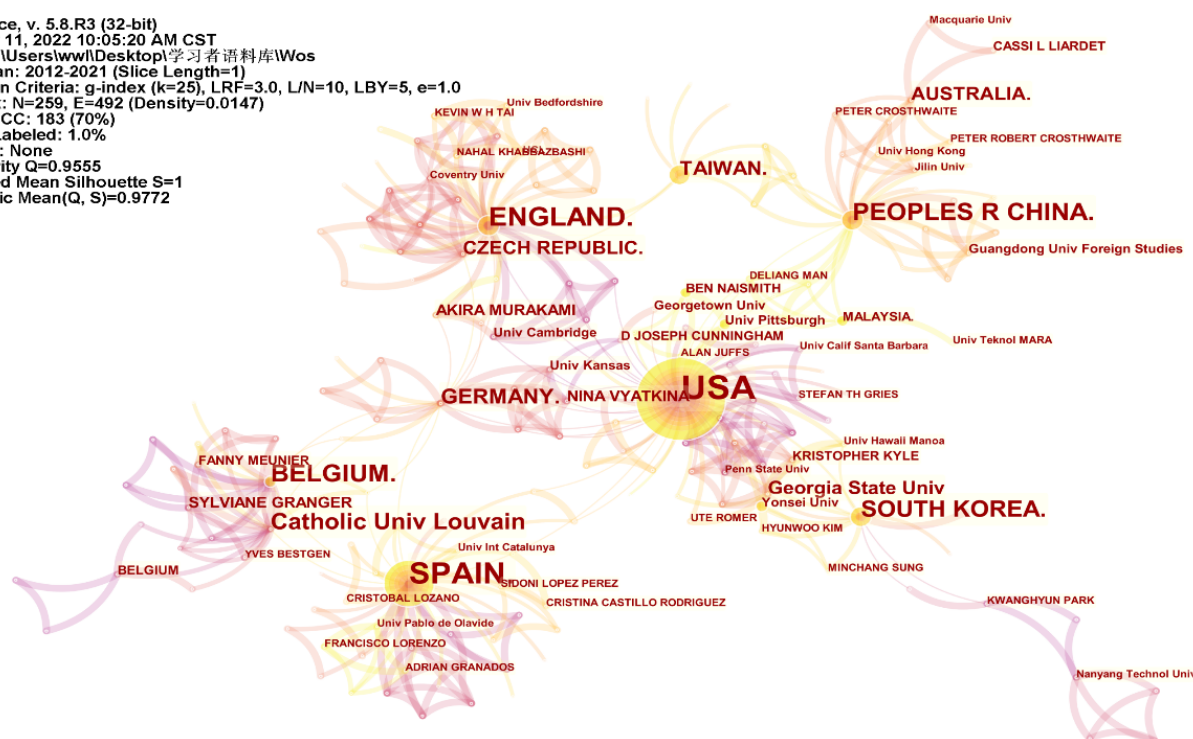


Figure 6. International collaboration network.

## 4. Keywords

### 4.1. Keyword Clustering

From Figure 7, It can be seen that keywords of learner corpus research papers in the WOS core collection generate 8 clusters, namely: association measure, attitude, developmental complexity, cognitive linguistics, historical literacy, construction learning, vocabulary explanations, and corpus use, which shows that in the past decade, the study of learner corpus has mainly focused on cognition, emotion, constructivism, learner language development, and the application of learner corpus. Cluster #0 indicates the very first focus of learner corpus research, that is learner language studied from the perspective of phraseology, including keywords like collocation, and statistical coefficients such as association strength. Cluster #1 shows the second prevailing focus of learner corpus research: the development of learner language and measurements of learner language to predict writing quality, including keywords like accuracy, fluency, complexity, and sophistication. Cluster #2 demonstrates the third biggest interest of learner corpus research: influencing factors that shape the features of learner language, including the learning environment, learning materials like textbooks used, and the learning process like instructions given and feedback.

### 4.2. Keyword Timeline

Figure 8 shows a timeline map of keyword clustering, as shown in the figure, Cluster #0, association measure, that is, the strength of association of collocations, which was initially used to study the relationship of elements in a lexical



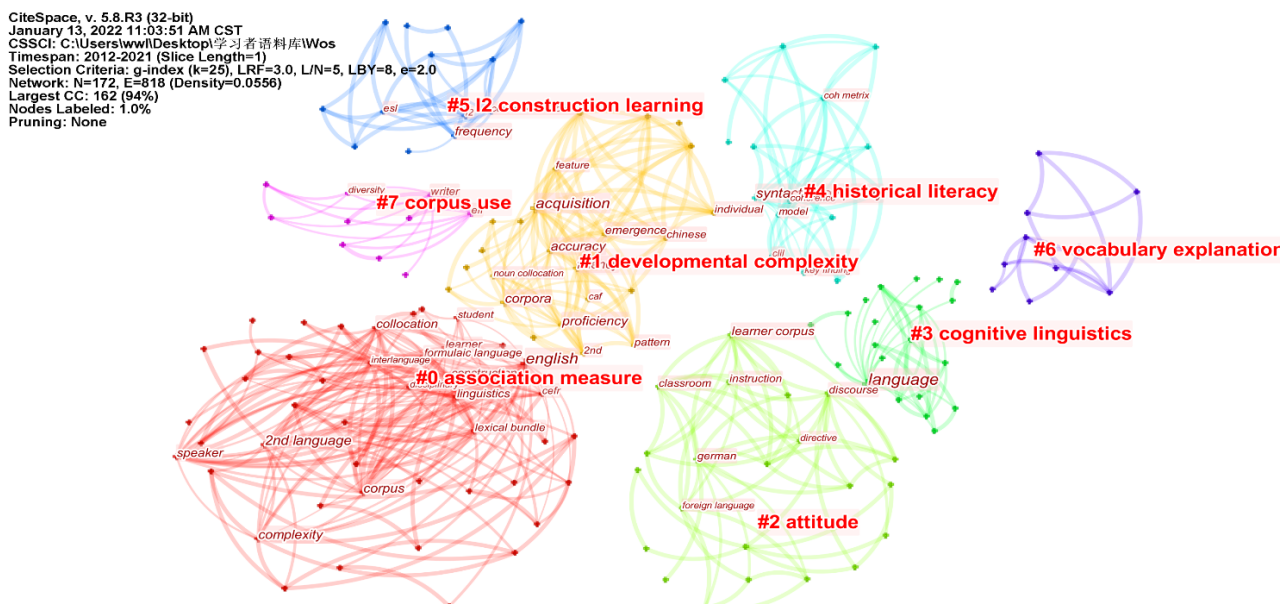


Figure 7. Keywords clustering.

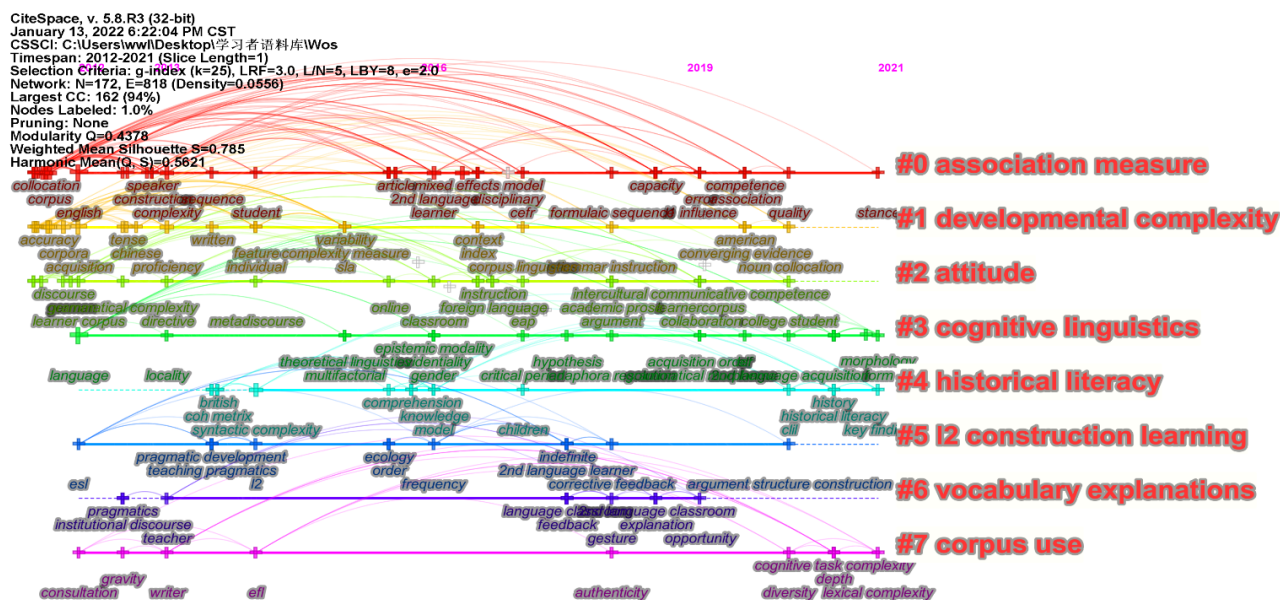


Figure 8. Keyword Timeline.

bundle or a cluster of words in writing and later moved to such linguistic aspects in spoken corpora. Since 2016, it has been applied to the study of individual differences in interlanguage and has recently been used for the analysis of identity and emotion, etc. Cluster #1: developmental complexity, is one of the important indicators of language description, and is first used in syntax, the vocabulary used, and discourse. Initial attempts focused on individual linguistic elements, but later research moves the interest to the complexity of phrases, noun phrases especially. Cluster #2, attitude, is first used for discourse analysis in classroom writing, then move to online communicative discourses, and then to English for

academic purposes. Cluster #3, cognitive linguistics, is used first for modal verbs, followed by grammar and morphology. Cluster #4, historical literacy, indicates that researchers have a special interest in summarizing the literature on learner language research. Cluster #5, construction learning, firstly concentrated on verb constructs, then there is the study of metastructure constructs. Cluster #6, vocabulary explanations have the initial research focus on discourse analysis; then there is classroom conversation analysis and teacher feedback research. Cluster #7, corpus use, initially focused on the use of online corpus resources, followed by the study of the writing process, then digital-driven learning (DDL).

### 4.3. Keyword Time Zone

**Figure 9** shows the time zone map of the keywords of the research papers on English learner corpus in the WOS core collection, which indicates the evolution of keywords of the research papers by year. In 2012, researchers focused their studies on features of language use mainly such as accuracy, complexity, fluency, etc., and paid special attention to the microstructure of learner languages such as strings, phrases, and collocations. In 2013, the research was especially focused on grammatical structure, discourse, etc.; in 2014, the study moved to the use of natural language processing tools, pragmatic development, and individual differences. In 2015, research was primarily conducted from the perspective of complex system theory. The same theoretical focus continued in 2016, with the focus moving to gender differences. Studies in 2017 were focused on longitudinal data and incorporated curriculum, disciplines, etc. The 2018 studies mainly focused on formulaic language, teacher feedback, and English for academic purposes. Studies in 2019 mainly concerned the influence of mother tongue, intercultural communication, order of acquisition, etc. In 2020, the research focused on qualifiers, meta-constructs, and noun collocations; The 2021 study covered morphology, the stance of a speaker, longitudinal corpus, etc. Therefore, it is expected that the future research directions in this field will be complex systems theoretical perspectives, longitudinal corpus, cognitive and psychological perspectives, and cultural approaches to learner language.

## 5. Co-Citations

### 5.1. Co-Citation Network of Literature

**Figure 10** shows the co-citations of the English learner corpus research papers in the WOS core collection. The top-first mostly quoted work is *The Cambridge Handbook of Lerner Corpus Research* edited by Sylviane Granger et al. (2015), 14 citations in total, which provides a rather comprehensive analysis of the past, the present state, and the future trends of the subject matter. In second place were works by Karin Aijmer (2009) and Bestgen Yves & Granger (2014), with a total citation of 9. Next comes the work by Scott Jarvis (2013), with a total citation of 8. Then there are the works by Sylviane Granger (2014), Yu-Hua Chen & Baker (2010), and Batia Laufer & Waldman (2011), with a total citation of 7, fol-

lowed by the works by Florence Myles (2015), Sylviane Granger et al. (2009), and Annelie Ädel & Erman (2012), with a total citation of 6. The literature above forms the intellectual basis for nearly a decade of learner corpus research.

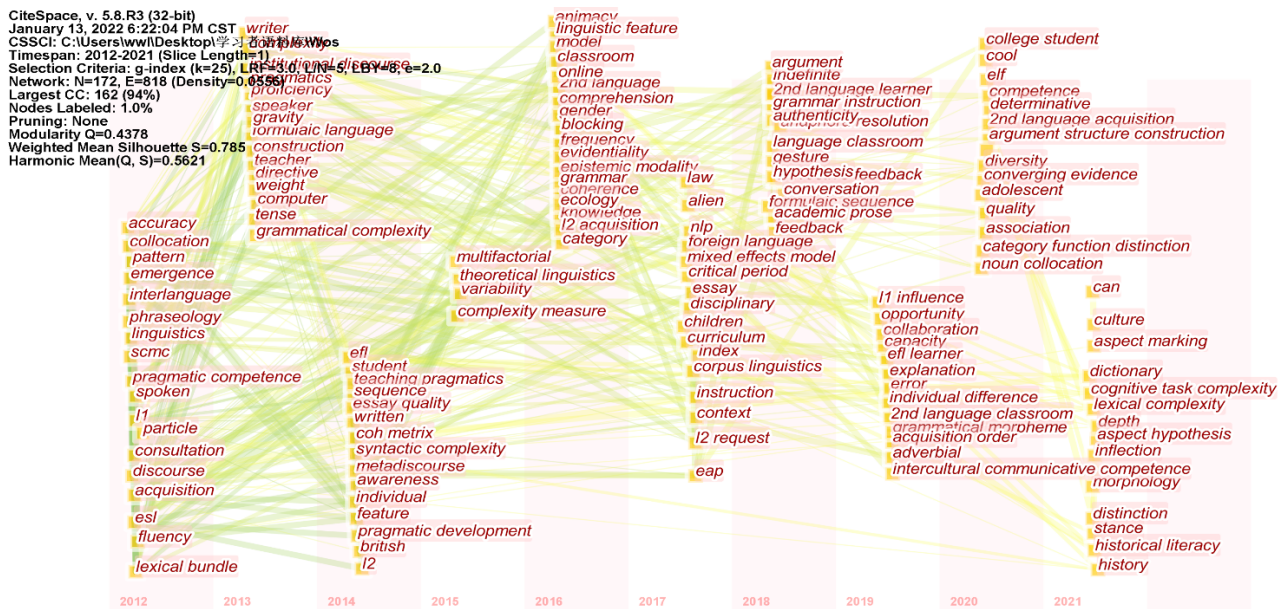


Figure 9. Keyword Time Zone.

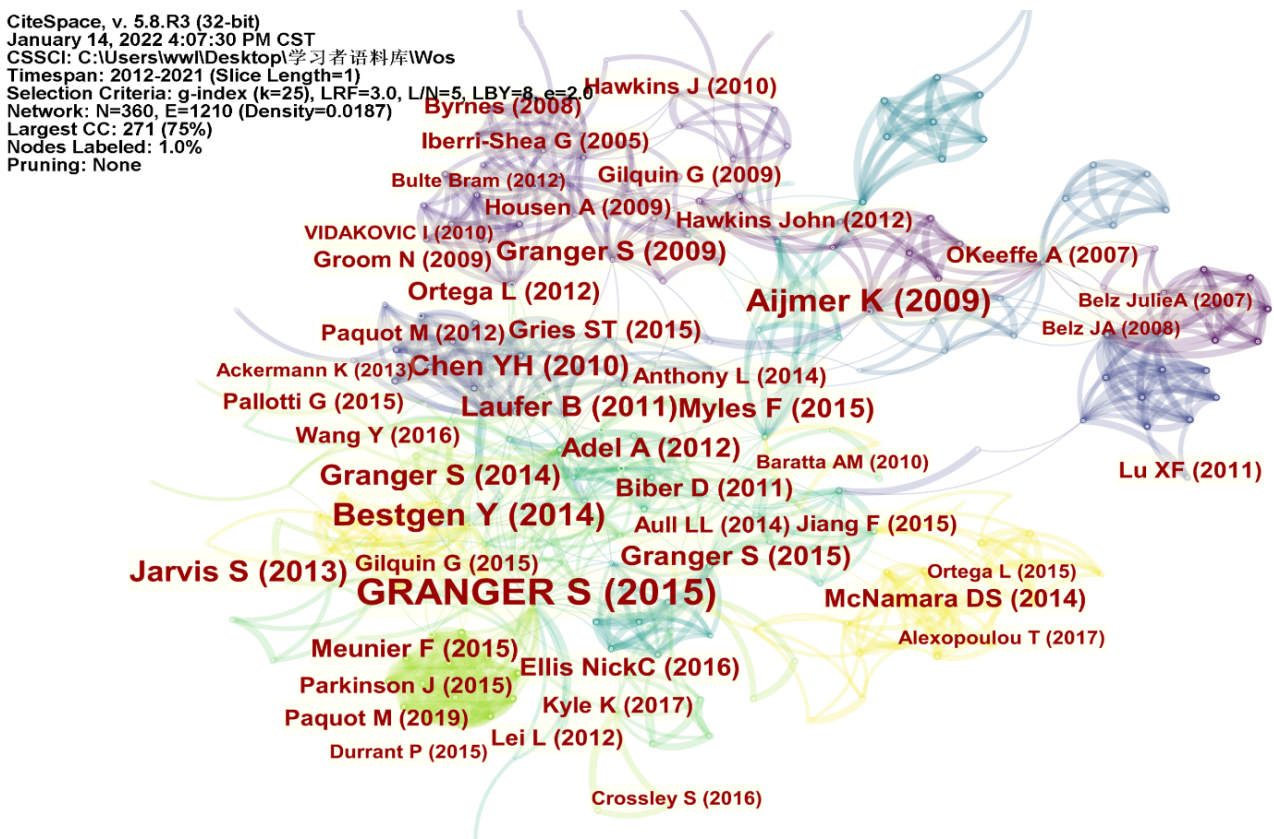


Figure 10. Co-citation network of literature.



In terms of centrality, the top research work is done by Sylviane Granger (2015), the centrality value is 42. Then there's the one by Bestgen Yves & Granger (2014) and Annelie Ädel & Erman (2012), with a centrality value of 30. The third one is by Douglas Biber (2011), 23 and the ones by Yu-Hua Chen & Baker (2010), Sylviane Granger (2014) and Belz & Vyatkina (2008), 22, Florence Myles (2015), 20, Anne O'Keeffe et al. (2007) and Philip Durrant et al. (2021), 19. These works play a significant part in forming the network, serving as the bridges for studies.

Figure 11 shows the bursts of highly quoted works on English learner corpus research. As can be seen that the largest burst is the work by Karin Aijmer (2009) from 2012 to 2015, with a burst value of 3.50; followed by that by Sylviane Granger et al. (2009) from 2012 to 2014, 3.27; the third is the research by Yu-Hua Chen and Baker (2010) from 2014 to 2017, 2.82; and the fourth is the one by Sylviane Granger (2015) from 2017-2021, 2.68. The results further indicate that the research focus in the period starting from 2012 to 2014 is turning to the application of learner corpus in language teaching and acquisition, and the building up of international learner corpus. And for the period starting from 2014 to 2017, the major subject of research moved to English for academic purposes. From 2017 onwards to 2021, researchers began to rethink the widely practiced methodology in learner corpus research: contrastive interlanguage analysis, and propose to involve contrastive study among learners of different cultural backgrounds and proficiency levels rather than sole attention to the contrast of learner language use and that of native speakers.

## 5.2. Co-Citation Network of Authors

Figure 12 shows the co-citation network of authors for learner corpus research in the WOS core collection. The author with the highest total citation is Sylviane Granger, with a total citation of 96, followed by Douglas Biber (83), Gaëtanelle Gilquin (47), and Ken Hyland (41), Stefan Th. Gries (35), Rod Ellis (34), Nick C Ellis (31), John Sinclair (31), Susan Hunston (30), Nadja Nesselhauf (26). These authors have made fundamental contributions to the study of corpus for English learners for nearly a decade.

In terms of centrality, Douglas Biber and Nick C Ellis are at the top of the list, with a centrality value of 63; followed by Sylviane Granger (51), Rod Ellis and Magali Paquot (50), Susan Hunston (49), Gaëtanelle Gilquin (47), Fanny Meunier (46), Batia Laufer (45) and Philip Durrant (44). These figures were playing a significant role in the network.

## 5.3. Co-Citation Network of Journals

Figure 13 shows the co-citation network of journals for English learner corpus research in the WOS core collection. The first place is taken by *Applied Linguists*, with a total citation of 108. Then there's *Culture Learning* (106), *TESOL Quarterly* (79), *International Journal of Corpus Linguistics and Thesis* (73), *Modern Lan-*

*guage Journal* (61), *Journal of Second Language Writing* (60), *System* (58), *Studies in Second Language Acquisition* (56) and *English for Academic Purposes* (51). These are the most influential journal for English learner corpus research.

From centrality, the top one is the *Annual Review of Applied Linguistics*, with a centrality of 72. And there are others following: *Modern Language Journal*

References	Year	Strength	Begin	End	2012 - 2021
Aijmer K, 2009, CORPORA LANGUAGE TEA, V0, P13, DOI DOI 10.1075/SCL.33.04GRA, DOI	2009	3.5	2012	2015	■■■■■
Granger S, 2009, INT CORPUS LEARNER E, V0, P0	2009	2.68	2012	2014	■■■■■
Chen YH, 2010, LANG LEARN TECHNOL, V14, P30	2010	2.82	2014	2017	■■■■■
GRANGER S, 2015, INT J LEARNER CORPUS, V1, P7, DOI DOI 10.1075/IJLCR.1.1.01GRA, DOI	2015	3.27	2017	2021	■■■■■

Figure 11. Bursts of highly-quoted literature.



Figure 12. Co-citation network of authors.

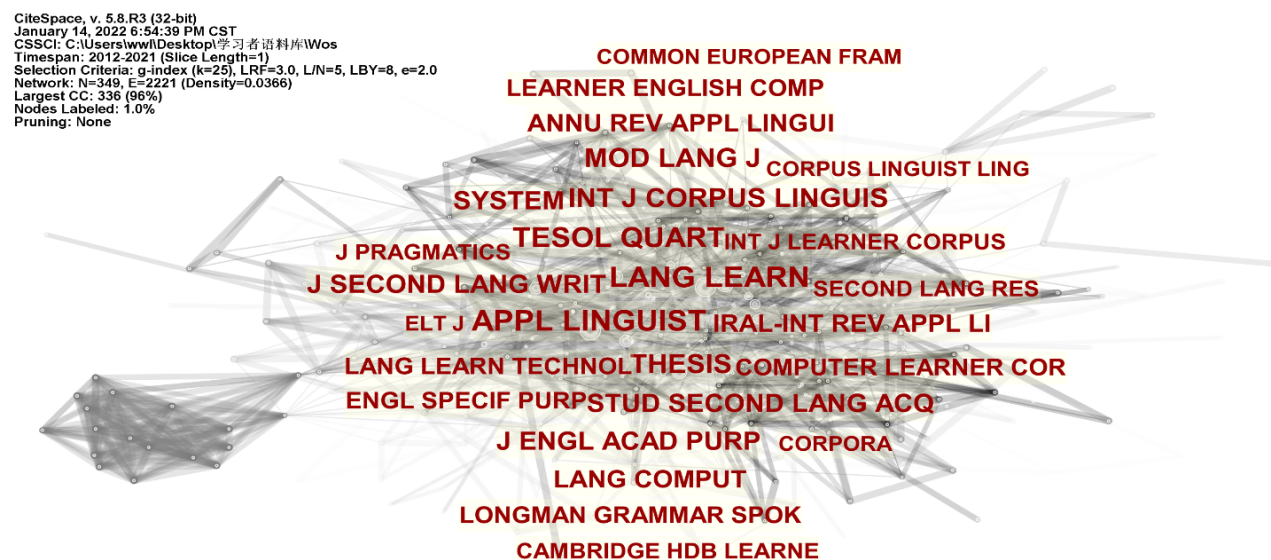


Figure 13. Co-citation network of journals.

(61), *The Canadian Modern Language Review* (57), *System* (53), *Studies in Second Language Acquisition* (52), *English for Specific Purposes* (51), *International Review of Applied Linguistics* (49), *Learner English Composition*, *Corpus Linguistics*, and *Linguistic Theory, Language Learning and Technology* (48).

## 6. Conclusion

In the past ten years, the academic community has paid continuous attention to the study of learner corpora of English writing, and the number of articles published has steadily increased. Researchers, academic institutions, and countries (regions) have formed a certain network of collaboration, but the network density is rather small (density < 0.006), and the cluster average contour value is small (Silhouette < 0.5), which shows that we have a reason to strengthen international and domestic collaboration in the study of English learner Corpus.

The most influential researchers whose research forms the knowledge base for the subject matter in the past decade include Sylviane Granger, Karin Aijmer, Yves Bestgen, Scott Jarvis, Yu-Hua Chen, Batia Laufer, Florence Myles, and Annelie Ädel. Sylviane Granger of the University of Louvain has made major contributions to the construction of The International Corpus of English (ICE), improved and expanded the research method of comparative interlanguage analysis (CIA), edited and published the first collection of learner corpus applications and research papers (Granger, 2014, 2015). Karin Aijmer edited and published a collection of essays on corpus and foreign language teaching (Aijmer, 2009). Yves Bestgen conducted an empirical study on learners' ability to use phrases in English writing (Bestgen & Granger, 2014). Scott Jarvis explores the issue of vocabulary diversity in learner language (Jarvis, 2013). Yu-Hua Chen studied strings of words in academic English writing (Chen & Baker, 2010). Batia Laufer conducted a study of noun-verb collocations in learner corpus (Laufer & Waldman, 2011). Florence Myles explored and summarized the connection between learner corpus and second language acquisition (Myles, 2015). Annelie Ädel studied commonly used phrases in academic English writing for learners and native speakers, extending the vocabulary study of learner corpus from general English to academic English (Ädel & Erman, 2012).

In the past decade, the research on the learner corpus of English writing has mainly focused on features of language use such as accuracy, complexity, and fluency, the microstructure of learner language such as word classes, strings of words, phrases, collocations, etc. and the macrostructure of learner language, including grammatical usages, syntactical structures, and discourse characteristics, and comprehensive aspects of learner language such as individual differences, the influence of curriculum and discipline and feedback, and the influence of mother tongue. The future research interests in this field are complex-system theoretical perspective, longitudinal corpus, cognitive analysis, psychological experiment, cultural studies, and other related research perspectives, and the focus of research is turning to English for academic purposes, English for specif-



ic purposes, oral English and translational English.

The findings of this research provide insights for researchers in the field to facilitate their perception of the possible future research focus, and a shortcut for a newcomer in learner corpus study to know the most influential researchers and the most fundamental research works, which saves a lot of time and efforts for a starter. The results of this paper, especially the part for the possible focus of future research, have profound implications for learner corpus research of other languages or translanguage comparisons. It brings about the need to deem the research of any field as an evolving process, and highlight specific markings for research during a certain period.

## Acknowledgements

The present research is funded by the Innovative Practice Base for Developmental Integration of Information Technology with Foreign Languages Teaching and Research, and the National Social Science Fund of China (17BYY042, A Typological Study of the Function-order Interactions of the Modifiers of English and Chinese).

## Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

## References

- Ädel, A., & Erman, B. (2012). Recurrent Word Combinations in Academic Writing by Native and Non-Native Speakers of English: A Lexical Bundles Approach. *English for Specific Purposes*, 31, 81-92. <https://doi.org/10.1016/j.esp.2011.08.004>
- Aijmer, K. (Ed.). (2009). *Corpora and Language Teaching* (Vol. 33). John Benjamins Publishing. <https://doi.org/10.1075/scl.33>
- Belz, J. A., & Vyatkina, N. (2008). The Pedagogical Mediation of a Developmental Learner Corpus for Classroom-Based Language Instruction. *Language Learning and Technology*, 12, 33-52.
- Bestgen, Y., & Granger, S. (2014). Quantifying the Development of Phraseological Competence in L2 English Writing: An Automated Approach. *Journal of Second Language Writing*, 26, 28-41. <https://doi.org/10.1016/j.jslw.2014.09.004>
- Biber, D. (2011). Corpus Linguistics and the Study of Literature: Back to the Future? *Scientific Study of Literature*, 1, 15-23. <https://doi.org/10.1075/ssol.1.1.02bib>
- Capel, A. (2010). A1-B2 Vocabulary: Insights and Issues Arising from the English Profile Wordlists Project. *English Profile Journal*, 1, e3. <https://doi.org/10.1017/S2041536210000048>
- Chen, Y. H., & Baker, P. (2010). Lexical Bundles in L1 and L2 Academic Writing. *Language Learning & Technology*, 14, 30-49.
- Díaz-Negrillo, A., & Thompson, P. (2013). Learner Corpora. In: A. Díaz-Negrillo, N. Ballier, & P. Thompson (Eds.), *Automatic Treatment and Analysis of Learner Corpus Data* (p. 9-30). John Benjamins Publishing. <https://doi.org/10.1075/scl.59.03dia>
- Durrant, P., Brenchley, M., & McCallum, L. (2021). *Understanding Development and*

- Proficiency in Writing: Quantitative Corpus Linguistic Approaches*. Cambridge University Press. <https://doi.org/10.1017/9781108770101>
- Ellis, R., & Barkhuizen, G. P. (2005). *Analysing Learner Language*. Oxford University Press.
- Gilquin, G. (2007). To Err Is Not All: What Corpus and Elicitation Can Reveal about the Use of Collocations by Learners. *Zeitschrift für Anglistik und Amerikanistik*, 55, 273-291. <https://doi.org/10.1515/zaa.2007.55.3.273>
- Granger, S. (2004). Computer Learner Corpus Research: Current Status and Future Prospects. *Applied Corpus Linguistics*, 52, 123-145. [https://doi.org/10.1163/9789004333772\\_008](https://doi.org/10.1163/9789004333772_008)
- Granger, S. (2012). Learner Corpora. In C. A. Chapelle (Ed.), *The Encyclopedia of Applied Linguistics* (pp. 1-8). John Wiley & Sons, Inc. <https://doi.org/10.1002/9781405198431.wbeal0669>
- Granger, S. (2014). *Learner English on Computer*. Routledge. <https://doi.org/10.4324/9781315841342>
- Granger, S. (2015). Contrastive Interlanguage Analysis: A Reappraisal. *International Journal of Learner Corpus Research*, 1, 7-24. <https://doi.org/10.1075/ijlcr.1.1.01gra>
- Granger, S., Dagneaux, E., Meunier, F., & Paquot, M. (Eds.). (2009). *International Corpus of Learner English* (Vol. 2). Presses Universitaires de Louvain.
- Granger, S., Gilquin, G., & Meunier, F. (Eds.). (2015). *The Cambridge Handbook of Learner Corpus Research*. Cambridge University Press. <https://doi.org/10.1017/CBO9781139649414>
- Granger, S., Hung, J., & Petch-Tyson, S. (Eds.). (2002). *Computer Learner Corpora, Second Language Acquisition, and Foreign Language Teaching* (Vol. 6). John Benjamins Publishing. <https://doi.org/10.1075/llt.6>
- Jarvis, S. (2013). Capturing the Diversity in Lexical Diversity. *Language Learning*, 63, 87-106. <https://doi.org/10.1111/j.1467-9922.2012.00739.x>
- Laufer, B., & Waldman, T. (2011). Verb-Noun Collocations in Second Language Writing: A Corpus Analysis of Learners' English. *Language learning*, 61, 647-672. <https://doi.org/10.1111/j.1467-9922.2010.00621.x>
- Leacock, C., Chodorow, M., Gamon, M., & Tetreault, J. (2014). *Automated Grammatical Error Detection for Language Learners*. Springer. <https://doi.org/10.1007/978-3-031-02153-4>
- Lu, X. (2011). A Corpus-Based Evaluation of Syntactic Complexity Measures as Indices of College-Level ESL Writers' Language Development. *TESOL Quarterly*, 45, 36-62.
- McEnery, T., Brezina, V., Gablasova, D., & Banerjee, J. (2019). Corpus Linguistics, Learner Corpora, and SLA: Employing Technology to Analyze Language Use. *Annual Review of Applied Linguistics*, 39, 74-92. <https://doi.org/10.1017/S0267190519000096>
- Myles, F. (2015). Second Language Acquisition Theory and Learner Corpus Research. In S. Granger, G. Gilquin, & F. Meunier, Eds., *The Cambridge Handbook of Learner Corpus Research* (pp. 309-332). Cambridge University Press. <https://doi.org/10.1017/CBO9781139649414.014>
- O'keeffe, A., McCarthy, M., & Carter, R. (2007). *From Corpus to Classroom: Language Use and Language Teaching*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511497650>
- Pravec, N. A. (2002). Survey of Learner Corpora. *ICAME Journal*, No. 26, 81-114.
- Tono, Y. (2003, March). Learner Corpora: Design, Development and Applications. In D.

Archer, P. Rayson, A. Wilson, & A. M. nMcEnery, Eds., *Proceedings of the Corpus Linguistics 2003 Conference* (pp. 800-809). UCREL Technical Paper Number 16. UCREL, Lancaster University.