Journal of Education and Educational Research
ISSN: 2957-9465 | Vol. 1, No. 1, 2022

The Use of Corpus in English Writing: Scholarly Development and Implications

Xugui Gui [1], Ling Deng [2,*], Guilin Jiang [3]

[1] Department of College English, Zhejiang Yuexiu University, Shaoxing, 312000, China
[2] School of Hospitality Management, Zhejiang Yuexiu University, Shaoxing, 312000, China
[3]School of Education, Anqing Normal University, Anqing, 246133, China

[*] **Corresponding author**: Ling Deng (Email: 20151034@zyufl.edu.cn)

Abstract: Corpus linguistics, as a methodological paradigm, is one of the striking fields in recent scholarly advancement. Its use in theoretical and practical explorations is generally accepted as a productive and efficient approach to provide affirmative support and empirical data and therefore, is wildly seen in disciplines of languages, linguistics and education. English writing, one of the commonest English language skills and a key area of studies for applied linguistics, has also aroused many interests in its integration with corpus linguistics. The use of corpus in English writing, though having achieved rich products in scholarly studies, still lacks a systematic investigation or a full summary on what have been developed and inspired. To fill in this gap, this study is aimed: by a bibliometric-and-corpus analysis on the titles, abstracts, and keywords, etc. of the scholarly articles of corpus and English writing, to find out their concerns, conclusions and implications.

Keywords: Corpus; English Writing; Implications; Scholarly Development.

1. Introduction

Corpus linguistics, as a methodological paradigm, is one of the striking fields in recent scholarly advancement; especially in English linguistics and some of the interdisciplinary studies (see Meyer, 2004 [1]; Hardie & McEnery, 2012 [2]; and many others). Its empirical use of data collection and verification in theoretical and practical explorations is generally accepted as a productive way and also an efficient approach to provide support of affirmation and evaluation and therefore, is wildly seen in disciplines of languages, linguistics and education.

English writing, as one of the commonest skills in English language and also a key area of applied linguistic studies, has unsurprisingly shown great interests in its integration with corpus linguistics. The use of corpus in English writing, though having achieved rich products in scholarly studies, still lacks a systematic investigation or a full summary on what have been developed till this day and what have been inspired for tomorrow.

To fill in this gap, this study is aimed to summarize the concerns, conclusions and indications of existing literature of corpus and English writing. By an autonomous collection of all scholarly articles entitled with corpus and English writing from the google scholar, this study is designed to conduct a bibliometric and corpus analysis on their titles, abstracts, and keywords, etc. and to identify some clues of their purposes, methods, discussions, results and limits. The corpus-based analysis together with its detailed discussion is expected to offer some explicit information about current advances and also implied inspirations for interdisciplinary studies in fields of corpus linguistics and English writing.

2. Literature Review

2.1. Corpus and English Corpus Linguistics: The Name, Nature and Significance

Corpus with its plural form as corpora, is a word referring to “body” with its origin in Latin. According to the glossary (Baker, Hardie, and McEnery, 2006 [3]), corpus in linguistics means a ‘body’ of language, or more specifically, a collection of texts stored in an electronic database. So, corpus, to some extent, is a carefully selected sample of real language in living practice. The difference between a corpus and an archive, for example, lays largely in its representativeness, balance and comparativeness (Hardie & McEnery, 2012).

Due to the boom of computer science and technology, the collection of large amounts of real samples of language in use or a big size of corpus comes into practice. And eventually, it brings an end to the doubt of the availability of corpus use in empirical analysis or the capability of corpus test in language studies. Since the edition of Brown Corpus and LOB Corpus, for example, different types of corpora have been created and then widely applied into both quantitative and qualitative analyses.

Therefore, corpus linguistics finally comes into being as a scholarly enterprise concerned with the compilation and also the analysis of corpora. It, though named as a separate discipline, is actually a methodological paradigm in nature, which is generally accepted as an efficient and effective approach to describe language use in real practice (Kennedy, 1998 [4]; Baker, Hardie, and McEnery, 2006). After decades of efforts from those corpus linguists and from some of those “neo-Firthian” scholars like Sinclair and Stubbs in particular, corpus and corpus linguistics have gained a great popularity among almost all branches of language and linguistic studies, normally in the forms of so-called corpus-based studies and/or corpus-driven analyses (Tognini-Bonelli, 2001[5]).

According to Meyer (2004), Hardie and McEnery (2012), the birth, growing and revival of corpus linguistics has contributed largely to the development of academic studies, especially in fields of English linguistics and interdisciplinary areas. It enables the sample collection of empirical data, the statistic description of discourse, language and grammar, and the quantitative verification of theoretical hypotheses; all of which are of significance for studies of today or of future.

2.2. The Use of Corpus in Applied Linguistics: An English Linguistic Tradition

Just as Meyer (2004), Hardie & McEnery (2012) indicates, the use of corpus in intra-disciplinary studies of linguistics is long rooted in English linguistic tradition. Most of the leading scholars and research centers of corpus linguistics are coming from or located within territories of the UK and the US. Early in the 1960s, for example, researchers like Quirk and Firth and institutions like UCL and the Brown University have started their work in collecting samples of modern English.

Apparently, the use of corpus in applied linguistics is almost in line with the developmental history of English Corpus Linguistics, which, most probably, can be traced back to the early days of Randolph Quirk's work in UCL. Started in 1959, the Survey of English Usage (SEU) launched by Quirk is viewed as the first attempt to provide an ongoing collection of textual data and an exploration of Varieties of English, the purpose of which is mainly targeted at some grammatical analysis of present-day English (Hardie & McEnery, 2012).

Followed by Sinclair, Leech and many others, those efforts as Brown Corpus in written texts and grammars are diversified into studies in spoken discourse and lexical items just as LLC, LOB and COBUILD have demonstrated. And then the construction and studies of corpus in applied linguistics have expanded to multi-lingual version and multi-disciplinary comparison, just as shown by the LDC and the UPenn treebanks. So, studies of the use of corpus have developed into almost every field of applied linguistics like listening, speaking, reading, writing and translating.

Among those literature about the use of corpus in applied linguistics, however, most of the studies are focused on a specific issue as English teaching, English learning or second language acquisition with few of their attention paid on an overview of the historical development of a scholarly question. Fewer attempts can be found to give a bibliometric or corpus analysis on the concerns, contributions and gaps of articles and publications in a certain field of scholarly circles. This is especially true when compared to those studies of interdisciplinary features. So, take the use of corpus in English writing as an example, this study is designed to identify the scholarly status of intradisciplinary studies between corpus linguistics and applied linguistics in terms of English writing, to uncover its concerns and findings, its strengths and weaknesses, its past journeys as well as its future trends.

3. Research Design and Instruments

3.1. Research Design

In order to conduct a bibliometric and corpus analysis on the use of corpus in English writing, this study is designed mainly into 3 steps to achieve the preset goals. The 1st step is to collect the sample literature of existing papers or books on corpus linguistics and English writing, which is accomplished by a well-developed software of literature ranking and collection named Publish or Perish by Harzing. The 2nd step is to have a bibliometric analysis on the author(s), cites, source, year of publishing, and publishers of the current studies centered on corpus and English writing and to draw a general image on their developmental trends on the basis of the collected samples of articles. The 3rd step is to launch a corpus analysis on the titles, abstracts and key words of sampled papers or books to have a detailed description of their concerns, conclusions, and contributions for studies on corpus and English writing.

The detailed structure of this research is shown as below in Figure 1.

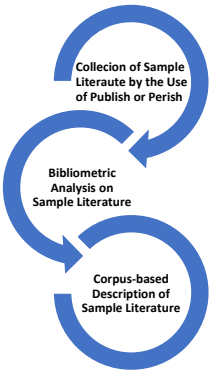


Figure 1. A Three-step Research Design

3.2. Research Instruments

Based on the research design, several tools are employed in the study for literature searching, sample collection, and data retrieval and analysis. Specifically, they are Publish or Perish 8, Microsoft Excel 2209 and LancsBox 6.0.

Publish or Perish 8 is first used in this study for literature searching and collection. It is a multi-systematic program or application initially designed by Anne-Wil Harzing's husband in 2006 for her retrieval and analysis of academic citations. According to Harzing (2016) [6], Publish or Perish 8 is capable to use a variety of data sources (normally the popular academic search engines or authoritative index of journals as google scholar, scopus, and web of science) to obtain the raw citations, then to put these citations into analysis and to present a range of literature details and citation metrics, including author(s), publisher(s), source(s), the year for publishing, the number of papers, total citations, and h-index, etc.. The searching results of literature is presented in the middle of the program window with their metrics of citation shown on the right top corner of the window and the output operation buttons left on the right bottom. The specific interface and function of Publish or Perish 8 is shown in Figure 2.

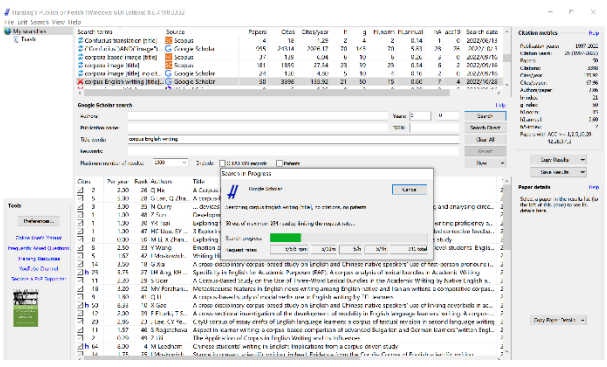


Figure 2. The Interface and Function of Publish or Perish 8

After the output and downloading of retrieval results from Publish or Perish 8, Microsoft Excel 2209 is then introduced as a tool for bibliometric analysis of the searching results. By the use of tools as data filter, data sort, and creative table in excel application, those articles featured as by the same author, from the same publisher, and in a same year, etc., will be clearly identified and vividly demonstrated. Therefore, the signs of developmental trends of scholarly studies on the use

of corpus in English writing can be traced back or forecasted.

When the bibliometric analysis of the sample literature is done, the retrieval results of target literature are expected to form a new corpus for a further investigation. Thus, a program software named LancsBox will be invited to have a detailed study on the collection of sample articles. LancsBox is one of the popular tools for corpus analysis, a new-generation software package for the analysis of language data and corpora. It is developed mainly by Brezina (leader), Platt (Developer), and McEnery (advisor) at the Lancaster University and LancsBox 6.0 is its newest version. There are many tools integrated in LancsBox 6.0, from the traditional use of KWIC concordance to the advanced Graph drawing tool of collocates, or from the professional N-gram analysis to a comprehensive Wizard of report writing. LancsBox 6.0 can also be used for the comparison of textual dispersion and statistic calculation of frequency and distribution. It is very helpful for the collection of data from the corpus of sample literature and to identify some specific details of the concerns, conclusions and contributions of the existing articles or books in corpus linguistics and English writing.

4. Date Collection

4.1. Collection of Sample Literature

Just as indicated above, Harzin's Publish or Perish 8 offers us a good option of resources for searching and locating of scholarly literature, namely some popular academic search engines as google scholar and some other authoritative lists of journals or indexes as scopus and web of science. In order to maximize the source of sample literature, to collect as large a collection of articles or books as possible; google scholar is chosen in this study to be used as the target source of sample corpus; the range and diversity of which will also be helpful for the guarantee of the balance and representativeness of those sampled literature.

Having decided google scholar as the source for searching and collecting sample literature, the retrieval window of google scholar in Publish or Perish 8 is expected to be opened and the search terms as "corpus" and "English writing" will be input into the title words column of this retrieval window or page. Then as Figure 3 shows, the results of searching action will come out as a list of 251 papers.

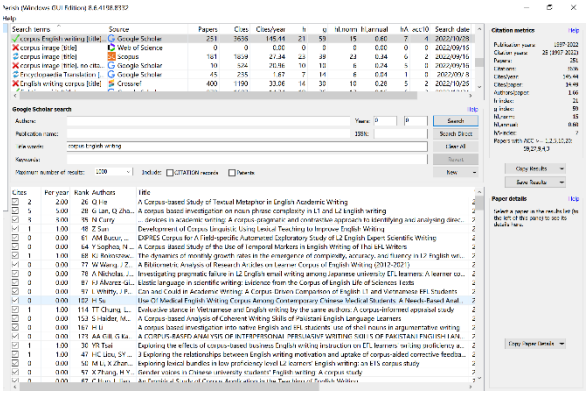


Figure 3. Searching Results of "Corpus" and "English writing" in Google Scholar (by Publish or Perish)

The searching results of 251 papers will be saved either as an extended report in a form of rtf file or a bibliographic collection in the form of csv file. By a click on the "save results" button of the Publish or Perish software, a detailed file of literature samples, including their titles, citations, publishers, abstracts and source links, etc., will be downloaded and kept for a further investigation.

4.2. Collection of Literature Metrics

The searching results of "corpus" and "English writing" conducted by Publish or Perish 8 in the database of google scholar will present a group of metrics of the general citations of sample literature and also some specific information of the literature metrics of every individual paper.

From Figure 4, we can see that quite a lot of metrics is shown with detailed information about each article sampled from the Google Scholar. Those metrics included in the csv file are cites, authors, titles, years, sources, publishers, article URLs, cites URLs, GS Ranks, Query dates, types, DOIs, ISSNs, author counts, abstracts, full text URLs, Related URLs, citation URLs, volumes, issues, start pages and end pages.

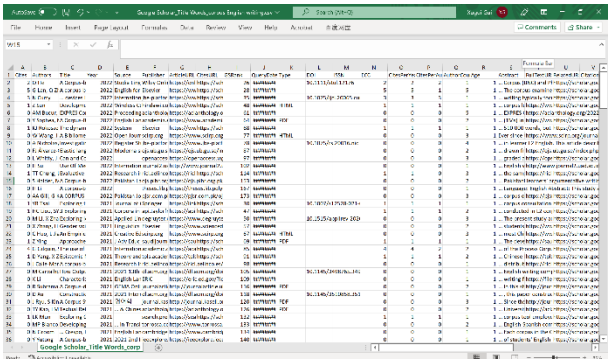


Figure 4. Literature Metrics included in the Retrieval Results

Just as what have been demonstrated in Figure 4, the metrics included in the retrieval results are normally a collection of raw date, which should be deduplicated and the carefully categorized for sake of a clear identification of those meaningful indicators. Those repeated publication of articles and books, for example, should be omitted from the list of sample literature and those metrics of sampled literature can be generally grouped into such categories as indicators of source (e.g., authors, sources, publishers, article URLS, DOIs, full text URLS), factors of influence (e.g., cites, cites URLS, GS Ranks, author counts, citation URLs), and features of literature (e.g., years, types, start pages and end pages).

When the null and duplicate figures have been removed from the raw data, the csv format of table is ready to be converted into an excel file and then available for a further data retrieval and calculation.

4.3. Collection of Corpus Data

After the collection of literature metrics, the original texts and figures included in the excel file can be also converted into a txt file and then used for a corpus analysis on the scholarly advances that were made by the sample literature, i.e., their concerns, conclusions, and also suggestions for future explorations.

In order to do this, the newest version of LancsBox is applied in the collection of data from the new built corpus. According to Baker & McEnery (2015) [7] and Wang & Pan (2020) [8], there are mainly 7 tools or sections in LancsBox which can be used individually or comprehensively to achieve the goals of item retrieval (e.g., lemmas, words, phrases, or structures), graph generating, textual comparison, listing of words, identification of N-gram collocates, textual reading and report writing. Correspondingly, they are named as KWIC, Graphcoll, Whelk, Words, Ngrams, Text and Wizard.

Normally, the tools used most often in corpus analysis are the so-called KWIC, Graphcoll, Words and Ngrams. They are also adopted in this study for purpose of identifying keywords, drawing graphs of collocation, generating wordlist, and finding 2-gram or 3-gram collocates of high-frequency. More specifically, the corpus is mainly constituted with texts of titles and abstracts, which is helpful for the clarification of key words or collocates from the sample literature of corpus and English writing. Then, by a KWIC concordance of those terms or lemmas as "corpus", "writing", "concern", "aim", "target", "discuss/discussion", "fail", "conclu*", etc., it is expected to find some clues about the achievements or gaps that our scholars have made or try to fill in in the fields of integrating corpus linguistics with English writing.

5. Data Analysis and Discussion

5.1. Bibliometric Analysis

Based on the collection of sample metrics, the bibliometric analysis in this study is mainly carried out from 3 aspects: the source of literature, the influence of literature and the issuing features of literature.

First, in terms of literature sources, those indicators as authors, sources, and publishers are all put into the calculation, filtering and sorting operation of the excel application. By a sorting of data by author(s), for example, those scholars as Moskowich (2008, 2011, 2012, 2013, 2014, 2019, 2020) and McIntyre (2003, 2004, 2011, 2012) are identified as two of the most productive authors in studying the use of corpus in English writing. Table 2 has listed some more details of the scholars who have been working in the field of corpus linguistics and English writing, a group of names that are very helpful for the reference of scholarly circles.

Similarly, the sorting of sources and publishers can also give more suggestion of literature reference and also clues of research sites for scholars working in some related fields. Data sorted by source (Table 3) shows that there are at least 33 conferences or forums ranging from 2010 to 2022 having laid their interests in processing papers on corpus linguistics and English writing. And from table 3, it can be revealed that among the global publishers, the *International Journal of Corpus Linguistics*, the *English Language Teaching*, the *Theory and Practice in Language Studies*, the Atlantis Press, John Benjamins Publishing Company, and the Springer are the top ranked journals and publishing houses that favored studies on corpus and English writing most.

Table 1. A List of the Most Productive Authors in Studying Corpus Linguistics and English Writing

Author	Year
I Moskowich	2015
I Moskowich	2013
I Moskowich	2011
I Moskowich, B Crespo	2014
I Moskowich, B Crespo, I Lareo, GC Rioboo	2012
I Moskowich, B Crespo, L Puente-Castelo, et al.	2019
I Moskowich, J Parapar	2008
I Moskowich, L Puente-Castelo, B Crespo, et al.	2020
D McIntyre, B Walker	2012
D McIntyre, B Walker	2011
D McIntyre, C Bellard-Thomson, ...	2003
D McIntyre, C Bellard-Thomson, J Heywood, et al.	2004
C Dayrell	2011
C Dayrell	2009
C Nan	2021
C Nan	2020
G Xia	2020
G Xia	2018
H He	2016
H He	2015
J Calle-Martín	2021
J Calle-Martín, J Romero-Barranco	2015
LH Ang, HA Rahim, KH Tan, etc.	2011
LH Ang, KH Tan	2018
M Leedham	2014
M Leedham	null
M Li	2015
M Li, X Zhang, BL Reynolds	2021
M Narita	2000
M Narita, K Kurokawa, T Utsuro	2002
M Nasseri	2021
M Nasseri	2016
ML Roca-Varela	2014
ML Roca-Varela	null
P Scheffler	null
P Scheffler	null
S Ge, J Zhang, X Chen	2016
S Ge, X Chen	2018
SA Joharry	2013
SAB Joharry	2016
X Li	2021
X Li	2021
Y Miyazaki, S Tanaka, Y Koyama	2014
Y Miyazaki, S Tanaka, Y Koyama	2011
Z Lili	2015
Z Lili	2015
Z Ruihua, G Libo, H Huaqing	2013
Z Ruihua, L Guo, H Huaqing	null

Table 2. Conferences or Forums Related to Corpus Linguistics & English Writing

Year	Source
2021	… & Chinese Language Processing, Volume 26 …
null	DANI MATE DEMARINA, RAZVOJNI ASPEKTI U …
2007	… and Change in English
2016	… and Natural Language Processing Based on …
2011	… Asian Journal of English …
2017	… Conference on Education, Language, Art and …
2016	… Conference on Education, Management Science and …
2016	… Conference on Electronics, Mechanics, Culture and …
2017	… Conference on Frontiers and Management Sciences …
2006	… Conference on Natural Language Processing (in …

2013	… Fifth International Conference of English …
2003	… for Computer Corpus …
2010	… IEEE International Conference on Progress in …
2021	… in Translation and Contrastive Research in …
2004	… of Modern English
2019	… of Qualitative Social …
2018	… of the Association-Institute for English …
2016	… on Education, Management, Computer and Society
2018	… Symposium on Emerging Technologies for Education
2014	… the tyranny of writing, Campus Walferdange, 26-28 …
2010	2010 5th International Conference on …
2010	2010 International Forum on …
2015	2015 2nd International Conference on Education …
2015	2015 International Conference on Social Science and …
2015	2015 Joint International Social Science, Education …
2016	2016 2nd Workshop on Advanced Research and …
2016	2016 International Conference on Economics, Social …
2020	2020 International Conference on Social Sciences and …
2021	2021 13th International Conference on …
2021	2021 2nd International Conference on Education …
2021	2021 International Conference on Aviation Safety and …
2021	2021 International Conference on Social Sciences and …
2013	Proceedings of TALN 2013 …
2022	Proceedings of the …
2017	Proceedings of the …

Table 3. Journals or Publishers that Favor Papers on Corpus and English Writing

Journal or Publisher	Papers/Books Published
International Journal of Corpus Linguistics	5
English Language Teaching	4
Theory and Practice in Language Studies	4
Overseas English	3
English for Academic Purposes	2
English Language &Linguistics	2
English Teaching	2
International Journal of Emerging Technologies in	2
Journal of English for Academic Purposes	2
LREC	2
Research in Corpus Linguistics	2
Revista de Lenguas para Fines Específicos	2
The Asian Journal of Applied Linguistics	2
TOKEN. A Journal of English Linguistics	2
Atlantis Press	13
John Benjamins Publishing Company	9
Springer	8
Elsevier	5
Taylor & Francis	4
Brill	4

Second, in terms of the influence of literature, those metrics as cites, Google Scholar Ranks, and author counts offer us some clues to locate and find the most influential papers or books that are referred most by the scholars. Table 5 has shown the literature of corpus and English writing studies that cited most by the scholarly circles. Those data included in this table can also give some extra information like the publish year and the GS rank about the literature.

Table 4. Top 20 Papers that Cited Most by Scholar Circles

Cites	Authors	Year	GS Rank
758	E Semino, M Short	2004	1
683	B Laufer, T Waldman	2011	7
631	D Biber	2009	6
279	K Bolton, G Nelson, J Hung	2002	2
90	J Horváth	2000	3
85	LH Ang, HA Rahim, KH Tan, ...	2011	11
73	P Pahta, I Taavitsainen	1998	8
64	M Leedham	2014	4
61	I Taavitsainen, P Pahta	1997	5
50	X Gao	2016	10
47	DJL Salazar	2011	15
46	D McIntyre, C Bellard-Thomson, J Heywood, ...	2004	13
34	M Callies, E Zaytseva	2013	22
33	I Moskowich, B Crespo, I Lareo, et al.	2012	31
31	A Gerbig	2010	17
31	M Wynne, M Short, E Semino	1998	16
28	D McIntyre, B Walker	2011	9
27	SO Ebeling, A Heuboeck	2007	19
25	I Moskowich, J Parapar	2008	12
23	J Sun, L Shang	2010	14

Third, in terms of literature features, such metrics as publishing years, literature types, and content pages can show us a clear image about the trend, the format and the size of sample literature. Take the publishing year of literature for example, the filtering and sorting of data by the year of publishing can show more details of the rapid development of the papers and books in exploring corpus and English writing.

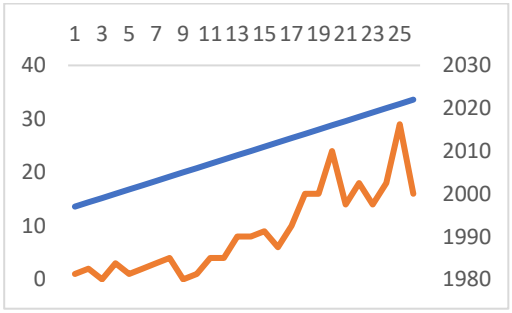
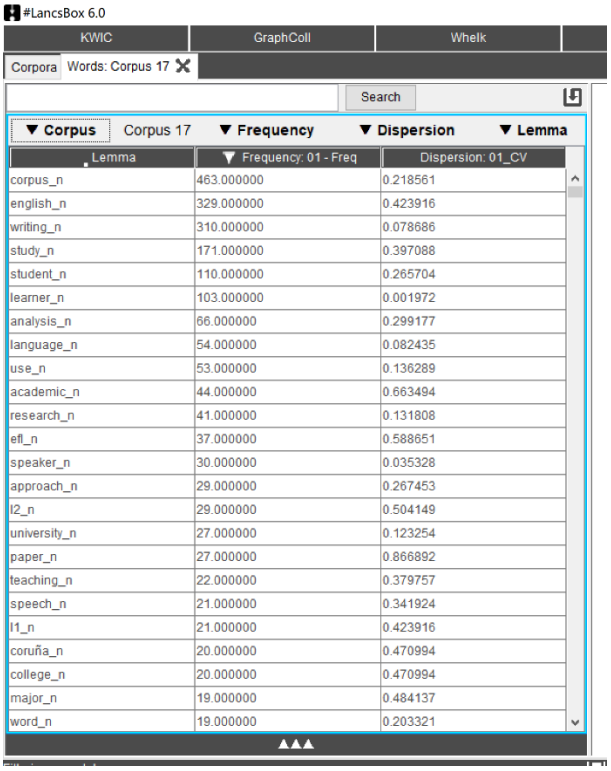


Fig 4. Number of Papers Published in Each Year

5.2. Corpus Analysis

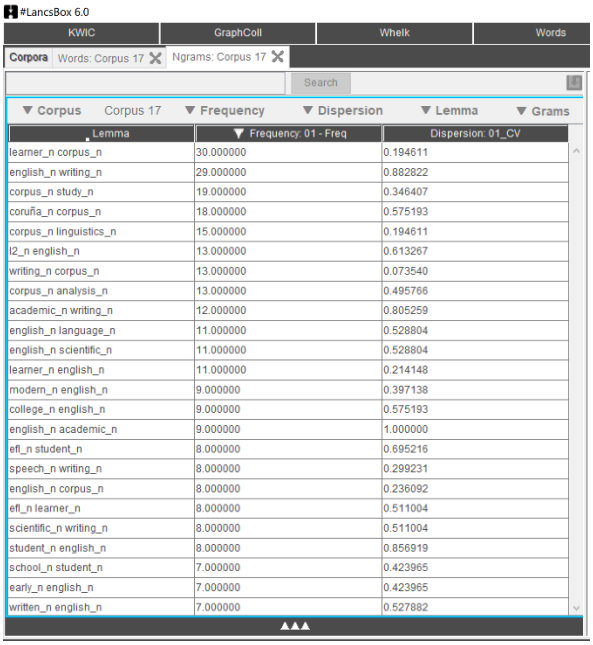
When to evaluate the development and implications of studies on corpus linguistics and English writing, the use of LancsBox in corpus analysis is mainly taken from two aspects: from analysis on titles to survey on abstracts, and also from analysis on concerns and contributions to review of gaps and shortcomings.

First, a wordlist survey of the corpus shows that corpus, English, writing, study, student, learner, analysis, language, use, academic and research are the top 11 nouns used in the titles and abstracts of the sample articles, which means that the study and analysis of students' and learners' language use is the dominant concern of those texts included in the corpus. And this can also be verified by the results of a retrieval of 2-gram collocates in the same corpus. Figure 5 and Figure 6 have demonstrated a high degree of matching or sameness among the words and collocations that are most frequently used in this corpus.



Lemma	Frequency: 01 - Freq	Dispersion: 01_CV
corpus_n	463.000000	0.218561
english_n	329.000000	0.423916
writing_n	310.000000	0.078686
study_n	171.000000	0.397088
student_n	110.000000	0.265704
learner_n	103.000000	0.001972
analysis_n	66.000000	0.299177
language_n	54.000000	0.082435
use_n	53.000000	0.136289
academic_n	44.000000	0.663494
research_n	41.000000	0.131808
efl_n	37.000000	0.588651
speaker_n	30.000000	0.035328
approach_n	29.000000	0.267453
l2_n	29.000000	0.504149
university_n	27.000000	0.123254
paper_n	27.000000	0.866892
teaching_n	22.000000	0.379757
speech_n	21.000000	0.341924
l1_n	21.000000	0.423916
coruña_n	20.000000	0.470994
college_n	20.000000	0.470994
major_n	19.000000	0.484137
word_n	19.000000	0.203321

Figure 5. A List of Nouns used in the Corpus



Lemma	Frequency: 01 - Freq	Dispersion: 01_CV
learner_n corpus_n	30.000000	0.194611
english_n writing_n	29.000000	0.882822
corpus_n study_n	19.000000	0.346407
coruña_n corpus_n	18.000000	0.575193
corpus_n linguistics_n	15.000000	0.194611
l2_n english_n	13.000000	0.613267
writing_n corpus_n	13.000000	0.073540
corpus_n analysis_n	13.000000	0.495766
academic_n writing_n	12.000000	0.805259
english_n language_n	11.000000	0.528804
english_n scientific_n	11.000000	0.528804
learner_n english_n	11.000000	0.214148
modern_n english_n	9.000000	0.397138
college_n english_n	9.000000	0.575193
english_n academic_n	9.000000	1.000000
efl_n student_n	8.000000	0.695216
speech_n writing_n	8.000000	0.299231
english_n corpus_n	8.000000	0.236092
efl_n learner_n	8.000000	0.511004
scientific_n writing_n	8.000000	0.511004
student_n english_n	8.000000	0.856919
school_n student_n	7.000000	0.423965
early_n english_n	7.000000	0.423965
written_n english_n	7.000000	0.527882

Figure 6. High Frequently used Noun Phrases in the Corpus: a 2-gram Retrieval

Next, a further investigation on the shared collocates of "corpus" and "writing" as shown in Figure 7 reveals that such words as student, learner, study, academic, EFL (English as Foreign Language), speech, L2 (the 2nd language), business, course are the most often used lemmas before or after the searching items in the sample corpus.

Similarly, in Figure 8, a retrieval of verbal collocates commonly share by "corpus" and "writing" will indicate that the use, improvement of speaking is also often concerned or compared by the studies of corpus and writing. Those verbs as "base", "employ", "find", "think", "develop" and "publish" are normally used to describe the methods and aims of those studies. While a more specific observation of the shared adjective collocates between "corpus" and "writing" will give us more information about the methodology and purpose of those sampled articles. A positional graph of the adjectives commonly shared by "corpus" and "writing" in Figure 9, for example, will prove that parallel, specialized or comparable corpus will usually be used in English writing studies. And some other commonly-shared collocates as "academic", "medical" and "online" has clearly indicated the genre of texts that mainly included in the studies.

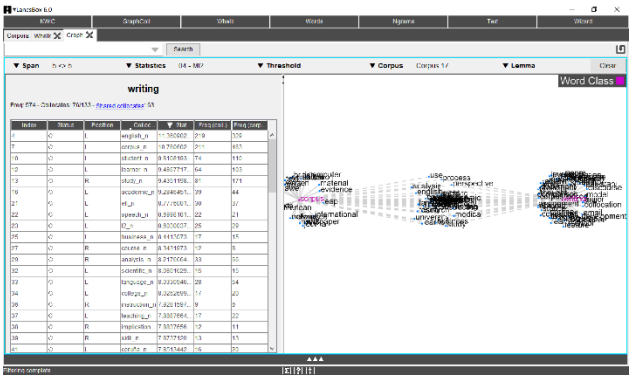


Figure 7. Nominal Collocates Commonly Shared by "corpus" and "writing"

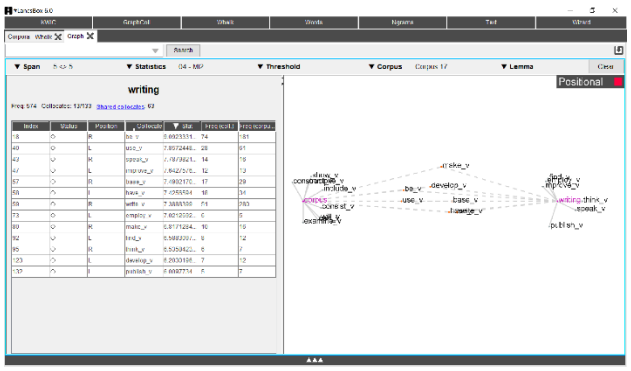


Figure 8. Verbal Collocates Commonly Shared by "corpus" and "writing"

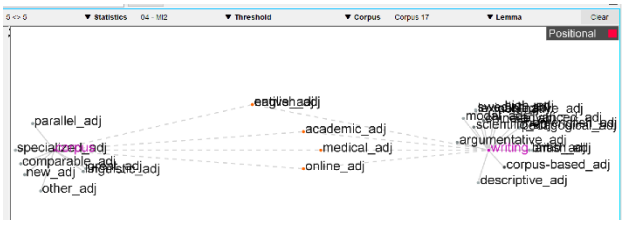


Figure 9. A Positional Graph of Adjective Collocates

Then, in order to find more implications from the corpus, a KWIC concordance of lemmas like "aim", "purpose", "method", "approach", "discuss", "result", "hope", "will" and "find", etc., are expected to be retrieved in the corpus; which is hopeful for the clarification of the objectives, methodology, conclusion, findings and limits of the sampled corpus. The results of KWIC concordance in Table 6, for example, reveals that most of the studies included in the corpus are aimed at the improvement of EFL students' writing proficiency or the L2 learners' writing skills in English; with some of them

Left	Node	Right
school testing requires… This is why I	aim	to explore the writing skills of this
of the corpus and various ways… The	aim	of this article is to compare the
in English and in French, with an	aim	to define implications for translation…
… The	aim	of the present study is to study
a corpus of student academic writing. The	aim	is **to…** Academic Writing in English (BAWE)
learners whose mother tongue is Arabic. The	aim	of this study is **to…** Este volumen
learners even at advanced levels. Therefore, the	aim	of this thesis is to investigate the
collected in Chinese Learner English Corpus. The	aim	to improve their English writing level and
of a new annotated learner corpus. The	aim	is to use this corpus to develop
English Scientific Writing is a project whose	aim	is to create a corpus for the…
… For this	purpose,	we compiled a learner corpus that consists
the corpus of Chinese English learners. The	purpose	of the study is to investigate the
graduate students' master's … Therefore, the	purpose	of this study was to find the
… in contextualized academic writing. As the	purpose	of this study is to compare the
This study intends to use a corpus-based	method	to explore a Malaysian English learner corpus…
English (Eng123) student ... A corpus linguistics	method	was used to interrogate the two corpora.
addressed this gap by employing a corpus-based	method	to investigate… L2 English writers with 11
…writing from the perspective of corpus-based	method.	Result of contrastive analysis of writing… in
… from the perspective of corpus-based	method.	Result of contrastive analysis of writing… can
This continuation task of L2-English writing	method	tries to work out how L2 learning
… Corpus-Based	Approach	in the teaching of writing would enhance
errors of Thai students using the corpus-based	approach.	The samples were two groups of upper…
… a corpus-driven	approach	to identify the most common multi-word patterns
…, the	approach	of corpus-based research on… English majors in
… The corpus-based	approach	to the linguistic analysis of…
In the case of genre, writing may	approach	the standards of… of the Coruña Corpus
… Since the corpus… the corpus stylistic	approach	is based on the corpus linguistic methods
… In line with this	approach,	We suggest a corpus-based implementation of several
L2 learners of English within a corpus-based	approach.	In so doing, the learner corpus… results
… a corpus-based	approach	to the analysis of English false friends
… corpus-based	approach	to enhance their English writing abilities. From
semantic prosody in the writing of… corpus-based	approach	in exploring the lemma CAUSE in Malaysian
… The empirical	approach	adopted here is grounded on Corpus… characteristic
… corpus	approach	to college English writing. The empirical study
… of the Corpus-Based	Approach	in Developing EFL Students' Writing Proficiency: The
Japanese university EFL learners: A learner corpus	approach	Error Analysis of Taiwanese University Students' …
of Arab Students of English: A Corpus-based	Approach	Reviewing Malaysian university English test report
writing: The move analysis of a corpus-based	approach	At close range: prefaces and other text
writing environment and learner corpus A corpus-driven	approach	to formulaic language in English: Multi-word patterns
thought in spoken British English: A corpus-based	approach	Investigating the presentation of speech, writing and
Arab L2 learners of English: A corpus-based	approach	Exploring lexical morphology across languages:
…: A corpus-pragmatic and contrastive	approach	to identifying and analysing direct and indirect
… of Students' Experiences with Corpus-based	Approach	on High School Students' English Paragraph Writing
… List (ACL)- A corpus-driven and expert-judged	approach.	Journal of English for Academic Purposes 12
… Writing by Chinese Speakers: A corpus-based	Approach	Chinese Students' writing in English:
Literature A study on a computer-based corpus	approach	to college English writing Application of Self-Constructed
… errors in English majors' writing a corpus-based	Approach	Actual Use of Corpus with Online Dictionaries
extensive picture of Chinese student writing today,	discusses	the findings of a corpus… English in
by corpus approaches to linguistic inquiry are	discussed	: …applying corpus-driven lexical analyses of second language
forms and image schemas in English writing, and	discusses	the possibility of using English corpus to…,
frequency of words appeared in their writing	result;	and concerning of Language Competence, this research
…	result	in a major challenge for English Language
… from the perspective of corpus-based method.	Result	of contrastive analysis of writing… in this
… from the perspective of corpus-based method.	Result	of contrastive analysis of writing… can employ
making it resemble spoken language as a	result.	Therefore, this paper examines the degree of
… through computer-aided error analysis in the	hope	of helping college learners improve their English…
description of Modern English scientific writing. I	hope	that both the Coruña Corpus and the…
in English writing for non-English majors. I	hope	we can further find out the regularity
to improve their English writing ability, meanwhile	hope	to…
though, few… of writing. Furthermore, this paper	will	further explore how English language teaching may
…	will	be carried out by analyzing a written
… In the sections that follow I	will	present the corpus material used for this
… corpus	will	serve as material for our research project
… I	will	then briefly review four corpus studies in
and Generation 1.5 writing, the resident corpus	will	now be…
in thought patterns between Chinese and English	will	be necessary in the acquisition of English
(in Biber… The corpus material and methodology	will	be…
influenced by their L1 use. The participants	will	be 40 sophomores who are taking the
own typical forms and distinctive functions. We	will	now therefore use the…
… the corpus-informed writing… Instead, we	will	focus on the description of Promociona-TÉ, a
corpus linguistic research on learner English and	will	have implications for the development of teaching
resources offered by the corpus.… The corpus	will	make great contribution to English writing if…
task of L2-English writing in flipped class. It	will	combine reading with writing, build creative use
the purpose of this study was to	find	the presence of SE in the students'
… speakers' academic writing in order to	find	out the potential trouble spots in SLA.
non-English majors, I hope we can further	find	out the regularity of the use of

Figure 10. A KWIC Concordance of "aim", "purpose", etc.

Focusing on the learning and teaching of English Writing in Special Purpose, especially on English for Scientific Writing or the English writing for Academic Purpose, while others paying their attention to writing of English for General Purpose. Most of the sampled articles devote their interests to college students or learners at advanced levels, including those majoring in English and those of non-English majors. And some of the studies are targeted at the writing ability and competence of high school student, graduate students, and even English native speakers.

In order to achieve their objectives, scholars have adopted quite a number of ways to collect their data. English testing reports, translation, contextualized academic writing, the case of genre, sematic prosody, documents of speech, lexical morphology, English paragraphs, writing errors, forms and image schemas in English writing, and frequency of words, etc., have all been used as the material for analysis. In terms of methodology, however, corpus linguistic approach is the predominant method for their scholarly research. Though comparison and contrast are also mentioned in some of the studies, corpus-based and corpus-driven methods are the most often used techniques in those studies, with corpus-based analysis taking its dominance.

According to Figure 10, most of the discussion and findings embodied in the sample texts of the corpus are centered at the English writing development of L2 learners, including the Arab students of English, the Chinese, the French learners of English, the students from Thailand, and the Malaysian English learners. By the employment of different kinds of English Learners Corpus, or different self-constructed or self-annotated corpus, the literature included in this study has offered us not only some empirical data of English writing teaching and learning, but also some advice for theoretical exploration and practical exercise of pedagogical activities.

The findings identified in the corpus, though, having shown their richness and diversity in many aspects, still lack the testification from a large scale, in a long term and with a hybrid or integrated pattern. Those papers exploring the use of corpus in English writing, for example, should be broaden in terms of genres and stylistics or should be deepened from aspects of pedagogical practice in addition to flipped class.

6. Conclusion

From all above, it can be concluded that studies on the use of corpus in English writing is an emerging field coming along with the development of corpus linguistics in the 1990s, which can be verified by its historical trend as implied in Figure 3. The collection of sample literature by Publish or Perish in the crossing area of corpus linguistics and English writing makes it possible to conduct a bibliometric analysis and also a corpus analysis on the achievements of integrated studies between corpus linguistics and English writing.

By the use of excel application and LancsBox software, the bibliometric and corpus analysis that made in this study is expected to give a general impression about the development of corpus linguistics in English writing studies and then provide some clues of those scholarly literature's implications. The discussions of this study are mainly based on the results of data collection; the details of which, however, still need to be continuously testified and then specified in the future.

References

[1] C. F. Meyer, English Corpus Linguistics: An Introduction. Cambridge University Press, New York, 2004.

[2] T. McEnery and A. Hardie, Corpus linguistics: Method, theory and practice. Cambridge University Press, 2012.

[3] P. Baker, Glossary of corpus linguistics. Edinburgh University Press, 2006.

[4] G. Kennedy, An introduction to corpus linguistics, London: Longman, 1998.

[5] E. Tognini-Bonelli, Corpus linguistics at work. Amsterdam/ Atlanta, GA: John Benjamins, 2001, pp. 1-236.

[6] A. W. Harzing, The Publish or Perish tutorial: 80 easy tips to get the best out of the Publish or Perish software. Tarma Software Research, 2016.

[7] Baker and McEnery, Eds. Corpora and Discourse Studies: Integrating Discourse and Corpora, Springer, 2015.

[8] L. J. Wang and F. Pan, “The Functions and Applications of #LancsBox, a New Generation Visual Corpus Software,” Contemporary Foreign Language Studies, vol, 05, pp. 77-90, 2020.