



Analyzing Research Tendencies of ELT Researchers and Trajectory of English Language Teaching and Learning in the last Five Years

Erdal Ayan ^{a *}, Elif Demirel ^b

^a Department of Computer Education and Instructional Technology, Hacettepe University, Ankara, 06800, Turkey

^b Department of Applied English and Translation, Ufuk University, Ankara, 06836, Turkey

Abstract

Owing to new advances in language teaching methodologies and integration of high technology tools as well as web applications, much scientific research has been recently published on English language teaching (ELT) and learning (ELL). Yet it is still a significant question to investigate exactly what types of research topics are mostly studied among the researchers from different countries. The current research aimed at finding out the most frequent research contexts and topics in the last five years through analyzing research papers published in leading academic journals in the field, and compare tendencies of the researchers from different institutions and countries in terms of selecting their research context and topics, and to figure out the trajectory for future studies. In this study researchers used a corpus-based detection methodology composed of storing variable data in .txt files and analyzing variables over the concordancer. The corpus-based data from the variables were analyzed by means of a statistical software, known as JASP in order to clear out potential differences among the researchers. A short analysis of the data indicated that the researchers still focus on the key words such as explicit learning and knowledge, implicit learning and knowledge as well as age and bilingualism. It was also observed that meta-analysis is an important topic in the studies conducted lately. Further results of the study could be beneficial for all followers including researchers and learners inside and outside ELT and help people focus on less frequently studied contexts and topics.

© 2017 IJCI & the Authors. Published by *International Journal of Curriculum and Instruction (IJCI)*. This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (CC BY-NC-ND) (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Keywords: Research Tendencies; Trajectory of ELT; ELT/ELL; Corpus-based detection

1. Introduction

English language English language teaching is a promising field of study for many experts on language education and students from different fields of study all around the world. This is mostly because number of the students who want to take English language

* Corresponding author. Tel.: +90-364-777-0252
E-mail address: erdal_ayan@yahoo.com

education has dramatically increased since the 1950s. Therefore governments tend to take political and educational decisions to direct and innovate their educational systems as well as to support schools with English language education. Also decision makers and private businesses frequently encourage researchers to conduct research on improving new approaches, techniques and technologies for the purpose of meeting upcoming language needs of the new generation learners who are technology junkies. As a result, the researchers have noted novel ideas and developments in integration of language education in to instructional applications at schools in the last years. In accordance with that, there are many scientific research published over academic journals on English language teaching (ELT) and learning (ELL) in recent years. However, on the one hand, it is still a significant question to research that exactly what types of research topics are mostly studied among the researchers from different countries. What are the leading research groups on the world? Even though there are many studies to clarify mostly studied topics and trajectory of the researches on ELT by means of text/content mining methodology, and there are very few studies to compare research tendencies of the researchers. Despite the fact that there are many papers reviewing literature, the scope of such studies are mostly limited. On the other hand, a corpus based detection methodology, which may illuminate those research tendencies and trajectory, and come up with descriptive results in the field, is actually missing. In sum, the current research aims at finding out the most frequent research contexts and topics in the last five years through analyzing research papers published in leading academic journals in the field, and compare tendencies of the researchers from different institutions and countries in terms of selecting their research context and topics, and to figure out the trajectory for future studies. In this study, the researchers hypothesize that there may be different tendencies among the researchers in terms of their selecting research contexts and topics, which should be revealed for future researches. Researchers use a corpus-based detection methodology in this study, which is composed of storing variable data in .txt files and analyzing variables over the AntConc. Corpus-based detection method defines process of gathering textual data mentioned in the variables and analyzing them by means of AntConc. The corpus-based data from the variables are analyzed by means of a statistical software, known as JASP in order to clear out potential differences among the researchers. Potential results of the study could be beneficial for all followers including researchers and learners inside and outside the field of ELT and help people focus less frequently studied contexts and topics.

Results of the studies using text/content mining are beneficial for all followers including researchers and learners inside and outside the field of language education and help people focus less frequently studied topics and areas. In this regard, the current paper is going to begin with illustrating novel research studies and literature reviews on text mining and findings of different researchers on how to analyze journal data. Then, hypothesis and research questions will be clarified. Methodology is going to be mentioned

in the following section. After that, data analysis and results are going to be followed. At the end, discussion and conclusion parts will take place.

2. Literature Review

Integration of internet and web applications is of great importance for many experts on English language teaching and learning as well as readers from different fields of study all around the world. Google's Ngram viewer is one of the closest examples for such applications which simply illustrate where certain words have appeared. Nowadays, publishers of scientific journals are of huge digital contents for readers, and capabilities of the networks and computers are beneficial in terms of discovery of deeper knowledge for average users and researchers. Moreover, text/content mining of journal articles, newspapers and other printed media and building up corpora for specific purposes in ELT/ELL have been on the rise since the 1990s (Stubbs, 1996; Burnard & McEnery, 2000; Thompson, 2001; Stuart & Botella, 2009). For example, Buckmaster (2015) writes that texts that will be used in the courses should be analyzed both pedagogically and linguistically before teaching and small scale corpora may particularly help teachers detect needs of the learners. Bernardini (2004) and Tsui (2004) attract attention to the functions of corpora building in EFL/ESL courses in terms of teachers' novel positions as facilitators, corpus-based discovery learning, self-evaluation of students' needs, learner autonomy, sequencing course contents in curriculum, teacher-teacher and teacher-student interaction. According to their perspectives, having corpus based teaching methodologies may contribute a lot to active participation of the learners into teaching-learning processes and support the process of self-discovery. Also, Thompson (2001) investigates different types of citations used in PhD theses in scope of academic writing in social sciences, humanities and science texts. In accordance with that, there are many scientific research published on language and linguistics in recent years. According to SCImago Journal Rank Indicator, there are 581 journals published in 2015. Total number of documents published by these journals is 13.582 in 2015 and 45.573 in the last three years. H Index of the ranking indicator showed that mean of the number of the articles which received citations in the period is 12.21. Means and sums of the journal ranking (SJR), total references and total cites in three years were also mentioned in the following table (see Table-1).

Table 1. Descriptive statistics for various indicators

	H index	SJR	Total Refs.	Total Cites (3years)
Valid	581	581	581	581
Missing	1	1	1	1
Mean	12.21	0.3478	906.9	60.30
Median	5.000	0.1380	600.0	8.000
Mode	1.000	0.1010	0.000	0.000
Std. Deviation	20.02	0.4732	1259	185.7
Minimum	0.000	0.1000	0.000	0.000
Maximum	171.0	3.403	1.118e+4	2226
Sum	7093	202.1	5.269e+5	3.503e+4

The statistics by the same indicator also reveal that most of the journals were released in the European countries and the USA (See Table 2).

Table 2. Frequencies for country

	Frequency	Percent	Valid Percent
Belgium	20	3.4	3.4
Czech Republic	13	2.2	2.2
France	24	4.1	4.1
Germany	65	11.2	11.2
Italy	21	3.6	3.6
Netherlands	63	10.8	10.8
Poland	15	2.6	2.6
Spain	43	7.4	7.4
United Kingdom	126	21.7	21.7
United States	77	13.3	13.3
...
Total	581	100.0	100.0

Text/content mining is not an area-specific research methodology. Together with support from corpora, it brings oversight and discovery of real knowledge and relationship analysis of the data produced by the researchers in various institutions. Text/content mining technologies such as Mendeley and Citeseer as automated citation indexing systems are very popular among the publishers and readers, who want to take impressions before conducting their research and follow certain research groups during their research process (Giles, Bollacker, & Lawrence, 1998).

Text/content mining is closely related to information processing. It has been a research methodology including computer based automated tools, techniques, latent semantic

indexing, *n*-grams depicting researcher's interests, topic models and sentiment analysis frequently employed by many researchers for different purposes nowadays (Smit & Van Der Graaf, 2012; Soriano, Au, & Banks, 2013). For example, Smit and Van Der Graaf (2012) has listed following purposes; *a) to identify and select relevant information, b) to extract information from the content, c) to identify relationships within/between/across documents and between incidents or events for meta-analysis, d) to improve information retrieval/navigation to the content of the publishers, e) to improve meta-data and semantic tagging of people, places and organizations, f) to create new products and services* (36-41).

In academic research articles, researchers are usually asked to provide keywords which characterize their research area and specific interest of the article itself. These keywords act like a digital identity for the research article, making it easily accessible through database searches. These keywords reported by the original researchers are carefully selected keywords act as reliable indication of scientific concepts referenced in them (Wittaker, Courtial & Law, 1989). However, keyword analyses usually use text-mining techniques to detect keywords from texts since different from research articles other kinds of texts do not provide a readily available list of keywords. Keywords analyses based on text mining techniques derive keywords statistically and use the data for purposes of literary analysis and exploration of genre differences to name a few. The technique of statistically deriving keywords, however, is not free of criticism. For example Conway (2010) reports that frequent words characterize texts better than keywords or key-keywords which are derived using more computationally intensive methods. In this study, a statistical keyword extraction method was not necessary since the keywords had already been reported by the researchers themselves. The purpose of the present study was to detect research tendencies, reported keywords provide sufficient and reliable information on this issue since the reported keywords are usually frequently repeated keywords in a research article.

In text analysis or text/content mining, using keywords provides important information for carrying out in-depth analyses within texts or making comparisons across texts or corpora since keywords create intratextual or intertextual networks of meaning (Stuart & Botella, 2009). Berber Sardinha (1999) examines the benefits of carrying out key word analyses using concordancing tools, namely Wordsmith tools. He notes that keyword analysis can serve to compare stylistic and developmental characteristics across texts while it can help to detect internal topic boundaries, and to distinguish between local and global topics or major or minor topics.

More importantly, text/content mining allows researchers to visualize when certain topics have exactly been frequently studied and become popular research topics within years. Such a visualization and conceptualization especially over 3-D mappings of evolution of the research field may definitely come up with formative and decisive data in

terms of educational and instructional perspectives and purposes since those who are following the researchers in the field may simply decide what is missing in the field and focus more on these areas. Smit and Van Der Graaf (2012) have already foreseen that “content mining will expand into new areas, enabled by easier software tools and will develop further into automated information extraction from large sets of content” (44). Text/content mining through building corpora makes it very practical and time-saving to mine and find out relationships between the research topics and upcoming tendencies of the researchers in the field without wasting time by reading detailed or limited literature reviews. It also makes it very simple to analyze non-uniformly structured texts and text archives, and change them into meaningful statistical data.

There are similar advantages of text/content mining with corpus based methodology in that it provides researchers with overviews and detailed statistical analysis through vocabulary and word order carrying significant signals and reduces time to find relevant, informative and meaningful results dealing with the research field. This study navigates the topics and research areas over key words that researchers focus in the last five years in the field of English language teaching and learning.

3. Hypothesis and research questions

The current study is a significant descriptive research for both researchers, PhD and master’s degree seekers as well as teachers and learners of English. The readers of the research may benefit a lot from the data presented, since it provides a broad photo of information on studies conducted by the researchers studying on ELT and ELL in certain institutions and countries. Aforementioned studies and many other studies have focused on the frequencies of the key words and contextual relationships. These studies have come up with very unique results in terms of providing very informative data for the researchers and followers of the mentioned journals. However, it seems that the data and results from the research are missing and even disregarding certain information dealing with the most frequent key words in contexts and upcoming research topics. The researchers of the current study hypothesizes that there are different tendencies among the researchers in terms of their selecting research topics and keywords. In this sense, the following research questions were formulated;

1. What are the most frequently used keywords by the researchers in two journals between the years 2012-2016?
2. What are the countries and institutions that the research with high frequency keywords are conducted in the journals?

4. Methodology

The current study is a descriptive and comparative research based on qualitative and quantitative data collected by means of a citation manager, called Zotero. The data was analyzed over a corpus concordancer, named AntConc. Researchers have used a corpus-based detection methodology in this study, which is composed of storing variable data in .txt files and analyzing variables over the concordancer. The corpus-based data from the variables have been analyzed by means of a statistical software, JASP in order to clear out potential differences among the researchers in terms of the keywords they produced. The following parts are going to provide detailed step-by-step process of the research.

4.1. Selection of the journals

Nowadays there are many academic journals and publications released on English language teaching and learning under title of language and linguistics. However very few of them have average international scientific standards to study on. For the current study, the academic journals known as Modern Language Journal and Language Learning were selected as the scientific journals to conduct the research. The journals are two leading ones with high .pdf standards, which is extractable information including abstracts, names of the writers, journals, issues, publication dates, etc. Their citations and frequency of publishing were highly regarded appropriate for the study. Both journals are published four times a year, which means a volume composed of four issues and special issues in various months. Both of them are indexed in both Linguistics and Education and Educational Research by twenty-five indexing services. The most common keywords used to define the scope and content of both journals are “modern, language, journal, modern language journal, mlj, languages, teaching, foreign, ACTFL, ESL, linguistic, second, acquisition, translation, abstracts, surveys, news, research, French, German, Spanish, English, analysis, periodical, article, reviews, studies, book, technology, instruction, language, learning, research, education, linguistics, studies, journal, acquisition, psychology, cognitive, science, education, neuroscience, ethnography, sociolinguistics, sociology, semiotics, semantics, periodical, analysis”. The impact factor of the Modern Language Journal is 1.188 and ranking is 42/181 according to *ISI Journal citation Reports* in Linguistics in 2015. Journal of Language Learning is of 1.869 impact factor and ranking is 10/181 according to *ISI Journal citation Reports* in Linguistics in 2015. According to Google Scholar Metrics, its ranking is first in foreign language learning, second in language and linguistics and second in Humanities, Literature and Arts. SCImago Journal Ranking¹ has also indicated informative descriptive statistics dealing with the ranking, H index, total documents published both in 2015 and within three years, references and citations (See Table 3).

Table 3. Descriptive statistics for the journals

Country	Title	Rank	SJR	H index	Total Docs. (2015)	Total Docs. (3years)	Total Refs.	Total Cites (3years)
United Kingdom	Language Learning	7	2.473	62	50	186	2513	373
United States	Modern Language Journal	41	1.147	46	61	157	3477	229

4.2. Tools for Data Collection and Analysis

As mentioned earlier, Zotero was utilized in order collect data from the journal sites. The journal articles were downloaded from the journal database together with all necessary information (e.g. writers' names and surnames, details of the publication including date of the article, all keywords and abstract of the research accompanying volume and issue numbers, etc.). All the data was stored under Zotero's library on a personal computer according to the years of publications. Then the data was exported in to a .txt file with RefWorks Tagged format. That format enabled the researchers to detect numbers of the keywords and length of the abstracts list and index the data (See Figure 1).

```

1 |RT Journal Article
2 |T1 Orders and Sequences in the Acquisition of L2 Morphosyntax, 40 Years On: An Introduction to the Special Issue
3 |A1 Hulstijn, Jan H.
4 |A2 Ellis, Rod
5 |A3 Eskildsen, Søren W.
6 |T2 Language Learning
7 |AB This Special Issue of Language Learning provides a state-of-the-art overview of current perspectives on acquisition orders and sequences in second language acquisition. It contains examples of empirical studies, theoretical reflections of, and a dialogue between proponents of Dynamic Systems Theory, Learner Varieties, Processability Theory, Processing Determinism, Sociocultural Theory, and Usage-Based Linguistics. This introduction presents the context and rationale for the collection of contributions.
8 |FD March 1, 2015
9 |YR 2015
10 |DO 10.1111/lang.12097
11 |VO 65
12 |IS 1
13 |SP 1
14 |OP 5
15 |JO Language Learning
16 |LA en
17 |SN 1467-9922
18 |ST Orders and Sequences in the Acquisition of L2 Morphosyntax, 40 Years On
19 |UL http://onlinelibrary.wiley.com/doi/10.1111/lang.12097/abstract
20 |RD 2017/02/27/13:30:32
21 |K1 interlanguage development
22 |K2 orders and sequences
23 |K3 second language acquisition
24 |K4 0
25 |K5 0

```

Figure 1. Tagged article data

By means of that method 374 articles were downloaded and tagged in total. The tagged data has been later analyzed over AntConc in order to detect frequency of key words and

Clusters/N-Grams. The analysis came up with informative statistical results on the frequency of use of the keywords by the researchers (See Figure 2).

Rank	Freq	Range	Cluster
1	2	1	k1 adolescent
2	2	1	k1 classroom
3	1	1	k1 academic
4	1	1	k1 advanced
5	1	1	k1 affect
6	1	1	k1 affordance
7	1	1	k1 age
8	1	1	k1 attitudes
9	1	1	k1 authenticity
10	1	1	k1 bilingual
11	1	1	k1 communicative
12	1	1	k1 comprehension
13	1	1	k1 computer
14	1	1	k1 corrective
15	1	1	k1 culture
16	1	1	k1 decolonization
17	1	1	k1 discourse
18	1	1	k1 epistemic
19	1	1	k1 first
20	1	1	k1 globalization
21	1	1	k1 incidental
22	1	1	k1 meta
23	1	1	k1 motion
24	1	1	k1 psycholinguistics
25	1	1	k1 reading

Figure 2. Frequency of key words produced

4.3. Process of research

The study took three months and it was completed in three main steps; a) data collection, b) data synthesis/analysis, and c) reporting. The researchers began the study by defining the journals with high quality PDF standards (i.e. files were allowing to get titles, names of the authors and other variables like volume, issue numbers and publication years) and published regularly. Citation indexes covering the journals and the citation metrics were also regarded. The articles from the journals were downloaded by means Zotero. The articles were checked if they were written on ELT or ELL and the irrelevant articles were deleted from the files. Then the researchers built up a small scale corpus covering the articles from the journals. When all the articles were completed, the articles were extracted from Zotero's store as .txt files. The format of the extracted files were regulated as RefWorks Tagged, which provided automated tagging for the variables such as A1 for Author, AB for Abstract and K1 for Keywords. During the process, it was realized that all the authors did not provide keywords in the same numbers; some of them wrote only three and some others wrote more than six. Therefore, only five keywords and one author were defined in each article and if an article was missing keywords, then it was disregarded and deleted from the file. The files were purified according to the scope of purpose of the study. Due to the fact that automated tagging system of Zotero tagged all key words with K1 and Authors with A1, the researchers had

to re-tag all those variables. It was also realized that Zotero randomly tagged keywords, and the variables such as names of the institutions and countries were missing in the PDF files, which challenged the process a lot. Therefore it has to be mentioned that K1 is not always the first keyword defined by the authors. The extracted files were later transferred into AntConc and the keywords were sorted out by means of Concordance tool. The sorted keywords were copied into an excel file including other variables like name of the journal, year, author, title. The names of the institutions and countries were later added into the file after they were manually taken from the website of the journals. The most frequent keywords were defined by means of Clusters/N-Grams tools of AntConc and all the keywords were listed in the excel file. The excel file was later analyzed via JASP and contingency tables were formed. The keywords with high frequency were used to find out the authors, institutions and countries that the research conducted.

4.4. Limitations

The study is limited to the articles produced between the years 2012 and 2016 and retrieved from the journals mentioned. Special issues published in the journals were not downloaded but only regular issues were taken. Only the most frequent keywords were surveyed in annual bases. The researchers with different institutions and countries were described according to the frequency of the keywords. The articles which were irrelevant and inappropriate for the study were disregarded.

5. Findings

The articles downloaded from the journals were analyzed initially over AntConc and then JASP in terms of the numbers of the articles published in the last five years. The analyses provided descriptive results dealing with the keywords, authors, institutions and countries. First of all, it was found out that the average number of the published articles in Modern Language Journal over the last five years was 40 while the articles published in Language Learning was 35. The number of the purified articles for Modern Language Journal was 149 and for Language Learning it was 138 (See Figure 3).

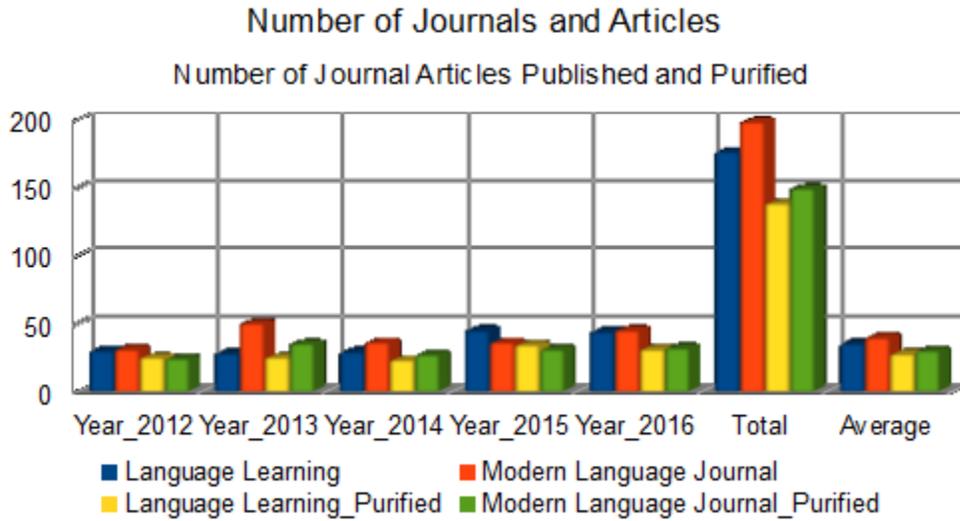


Figure 3. Number of journals and articles

Secondly, the compiled corpus provided informative data dealing with the keywords which were frequently used. Accordingly, it was revealed through Clusters/N-Grams tool that the keywords defined as K1 mostly included words such as classroom, bilingualism, acquisition, age, cognitive, bilingual and computer (See Figure 4).

File Global Settings Tool Preferences Help								
Corpus Files		Concordance	Concordance Plot	File View	Clusters/N-Grams	Collocates	Word List	Keyword List
LL_2012.txt		Total No. of Cluster Types 161		Total No. of Cluster Tokens 263				
LL_2013.txt		Rank	Freq	Range	Cluster			
LL_2014.txt		1	11	5	k1 classroom			
LL_2015.txt		2	8	4	k1 bilingualism			
LL_2016.txt		3	7	1	k1 0 k			
MLJ_2012.txt		4	6	3	k1 acquisition			
MLJ_2013.txt		5	6	5	k1 age			
MLJ_2014.txt		6	6	4	k1 cognitive			
MLJ_2015.txt		7	6	4	k1 l			
MLJ_2016.txt		8	6	4	k1 language			
		9	5	4	k1 bilingual			
		10	5	3	k1 chinese			
		11	5	4	k1 english			
		12	4	3	k1 computer			
		13	3	2	k1 academic			
		14	3	3	k1 accent			
		15	3	3	k1 advanced			
		16	3	3	k1 comprehension			
		17	3	3	k1 dynamic			
		18	3	2	k1 form			

Figure 4. Frequently used keywords

Sorting out the whole corpus by means of Concordancer tool indicated that the keywords frequently used were actually accompanied by various collocates. For instance,

K1 classroom was collocated by the words such as “interaction, research, discourse, intervention and -based research”, K1 acquisition was followed by the words like “sequences, order, learning/development” and K1 cognitive was coming before such words like “diagnosis modeling, abilities, fluency, conflict, and skills”. (See Figures 5, 6 and 7).

9922.2011.00630.x/abstract	RD 2017/02/27/08:45:58	K1 classroom interaction K2 humor K3 incidental fo
9922.2012.00716.x/abstract	RD 2017/02/27/08:49:52	K1 classroom research K2 error correction K3 impli
0.1111/lang.12020/abstract	RD 2017/02/27/09:51:46	K1 classroom research K2 corrective feedback K3 ES
4781.2013.01429.x/abstract	RD 2017/03/12/21:54:02	K1 classroom discourse K2 EFL K3 Japan K4
4781.2013.12035.x/abstract	RD 2017/03/12/22:06:37	K1 classroom intervention K2 factor analysis K3 le
4781.2013.12026.x/abstract	RD 2017/03/12/22:06:37	K1 classroom discourse K2 CLIL K3 discourse functi
4781.2013.12027.x/abstract	RD 2017/03/12/22:06:37	K1 classroom discourse K2 classroom ecology K3 gra
4781.2014.12061.x/abstract	RD 2017/03/16/20:15:24	K1 classroom interaction K2 English as a Lingua
0.1111/modl.12118/abstract	RD 2017/03/16/20:22:36	K1 classroom-based research K2 emergentism K3 fore
0.1111/modl.12193/abstract	RD 2017/03/16/20:28:44	K1 classroom-based research K2 goal setting K3
0.1111/modl.12273/abstract	RD 2017/03/17/19:29:46	K1 classroom-based research K2 English as a

Figure 5. Words collocated with “K1 classroom”

0.1111/lang.12089/abstract	RD 2017/02/27/13:30:32	K1 acquisition sequences K2 de-idealization K3
0.1111/lang.12095/abstract	RD 2017/02/27/13:30:32	K1 acquisition order K2 Processability Theory K3
0.1111/modl.12275/abstract	RD 2017/03/17/19:29:46	K1 acquisition/learning/development K2 child langu
0.1111/modl.12278/abstract	RD 2017/03/17/19:29:46	K1 acquisition K2 development K3 English as
0.1111/modl.12274/abstract	RD 2017/03/17/19:29:46	K1 acquisition/learning/development K2 English as
0.1111/modl.12344/abstract	RD 2017/03/17/19:53:54	K1 acquisition K2 corpus linguistics K3 input-

Figure 6. Words collocated with “K1 acquisition”

0.1111/lang.12016/abstract	RD 2017/02/27/09:53:37	K1 cognitive diagnosis modeling K2 English languag
0.1111/lang.12011/abstract	RD 2017/02/27/09:53:37	K1 cognitive abilities K2 high-level L2
0.1111/lang.12084/abstract	RD 2017/02/27/12:05:28	K1 cognitive fluency K2 pauses K3 second
0.1111/modl.12241/abstract	RD 2017/03/17/19:26:24	K1 cognitive conflict K2 complexity K3 engagement
0.1111/modl.12306/abstract	RD 2017/03/17/19:35:07	K1 cognitive K2 guidance K3 handover K4
0.1111/modl.12367/abstract	RD 2017/03/17/20:49:22	K1 cognitive skills K2 diagnostic assessment K3

Figure 7. Words collocated with “K1 cognitive”

In contrast to the K1 classroom, it was found that the keywords such as K1 bilingualism and K1 age (except for one case) were used without collocations (Figures 8 and 9).

0.1111/lang.12032/abstract	RD 2017/02/27/10:13:29	K1 bilingualism K2 dyslexia K3 naming K4 phonologi
0.1111/lang.12099/abstract	RD 2017/02/27/14:31:04	K1 bilingualism K2 language use K3 linguistic inte
0.1111/lang.12171/abstract	RD 2017/02/27/15:53:10	K1 bilingualism K2 cognitive control K3 cognitive
0.1111/lang.12189/abstract	RD 2017/02/27/15:53:10	K1 bilingualism K2 L2 monitoring K3 speech compreh
0.1111/lang.12191/abstract	RD 2017/02/27/15:53:10	K1 bilingualism K2 binding problems K3 cell assemb
0.1111/lang.12196/abstract	RD 2017/03/12/20:46:01	K1 bilingualism K2 cognition K3 interdisciplinary
0.1111/lang.12179/abstract	RD 2017/03/12/20:50:09	K1 bilingualism K2 biliteracy K3 case study K4
4781.2013.12034.x/abstract	RD 2017/03/12/22:06:37	K1 bilingualism K2 CANAL-FT K3 language aptitude

Figure 8. No collocated word with “K1 bilingualism”

9922.2012.00719.x/abstract	RD	2017/02/27/10:00:12	K1 age	K2 attitudes	K3 C-test
9922.2012.00731.x/abstract	RD	2017/02/27/09:57:20	K1 age	of L2 learning	K2 age-
0.1111/lang.12031/abstract	RD	2017/02/27/10:13:29	K1 age	K2 gender	K3 intercultural competence
0.1111/lang.12052/abstract	RD	2017/02/27/10:15:11	K1 age	K2 explicit knowledge	K3 priming
4781.2013.12044.x/abstract	RD	2017/03/12/22:17:09	K1 age	K2 classroom codeswitching	K3 first
4781.2014.12059.x/abstract	RD	2017/03/16/20:15:24	K1 age	K2 foreign accent	K3 L2

Figure 9. No collocated word with “K1 age”

Regarding all the keywords tagged, the most frequent ones were revealed through AntConc’s Clusters/N-Grams tool. The cluster size was fixed as min.2 – max.2 and it was found that some of the keywords were repeatedly used in different frequency levels. For example, the keyword classroom was used 11 times as K1 and 3 times as K2. The keywords such as “explicit, implicit, longitudinal, teacher, reading, foreign” were among the ones which were frequently used in different levels (See Table 4).

	Keyword	Frequency	Keyword	Frequency	Keyword	Frequency	Keyword	Frequency	Keyword	Frequency
1	k1 classroom	11	k2 language	13	k3 language	18	k4 language	17	k5 second	15
2	k1 bilingualism	8	k2 english	11	k3 second	10	k4 second	14	k5 reading	7
3	k1 acquisition	6	k2 second	7	k3 english	6	k4 motivation	5	k5 teacher	7
4	k1 age	6	k2 foreign	5	k3 foreign	5	k4 speech	5	k5 language	6
5	k1 cognitive	6	k2 explicit	4	k3 implicit	5	k4 task	5	k5 vocabulary	5
6	k1 language	6	k2 grammar	4	k3 reading	4	k4 mixed	4	k5 longitudinal	4
7	k1 bilingual	5	k2 lexical	4	k3 speech	4	k4 reading	4	k5 study	4
8	k1 chinese	5	k2 meta	4	k3 explicit	3	k4 statistical	4	k5 mixed	3
9	k1 english	5	k2 classroom	3	k3 incidental	3	k4 teacher	4	k5 social	3
10	k1 computer	4	k2 corpus	3	k3 individual	3	k4 foreign	3	k5 whorf	3
11	k1 academic	3	k2 corrective	3	k3 linguistic	3	k4 korean	3		
12	k1 accent	3	k2 cross	3	k3 memory	3	k4 linguistic	3		
13	k1 advanced	3	k2 efl	3	k3 research	3	k4 longitudinal	3		
14	k1 comprehension	3	k2 frequency	3	k3 teacher	3	k4 morphology	3		
15	k1 dynamic	3	k2 mixed	3						
16	k1 form	3	k2 reading	3						

Table 4. Frequency of keywords

Considering the first research question of the study, it was revealed that the production of the keywords changed from one year to another. 63 keywords were defined as the most frequent ones. And the rare keywords were excluded from the analysis. It became clear that some of the keywords were repeatedly used by the researchers. For example, K1 bilingualism is the most frequent one and densely used in 2016 and K1 age was mostly used in 2014 only. The analysis also showed that some keywords such as K1 heritage language and K1 linguistic relativity did not emerge before 2016.

Table 5. Year based keyword frequencies

KWIC1	Year					Total
	YR 2012	YR 2013	YR 2014	YR 2015	YR 2016	
K1 bilingualism	0	1	1	1	5	8
K1 age	1	1	3	0	0	5
K1 advanced learners	0	1	1	0	1	3
K1 accent	0	2	1	0	0	3
K1 classroom discourse	0	3	0	0	0	3
K1 classroom-based research	0	0	1	2	0	3
K1 acquisition	0	0	0	1	1	2
K1 acquisition/learning/development	0	0	0	2	0	2
K1 EFL	0	1	0	0	1	2
K1 conversation analysis	0	1	0	1	0	2
K1 elicited imitation	0	0	0	2	0	2
K1 heritage language	0	0	0	0	2	2
K1 interlanguage development	0	0	0	2	0	2
K1 language teacher cognition	0	0	0	2	0	2
K1 linguistic relativity	0	0	0	0	2	2
K1 meta-analysis	0	0	1	1	0	2
K1 adaptation	0	1	0	0	1	2
K1 aptitude	0	2	0	0	0	2
K1 associative learning	0	0	1	1	0	2
K1 classroom interaction	1	0	1	0	0	2
K1 classroom research	1	1	0	0	0	2
K1 communicative competence	0	0	1	0	1	2
K1 comprehensibility	0	0	0	1	1	2
K1 corrective feedback	0	0	1	0	1	2
K1 form-focused instruction	0	0	1	0	1	2
Total	3	14	13	16	17	63

The second research question was about the institutions and countries that these researches with high frequency keywords were conducted. The analysis indicated that the frequent keywords were produced in very different institutions. However Georgia State University, Ghent University, Indiana State University, Lancaster University, Michigan State University, University of Amsterdam and University of Alberta were the most leading universities in terms of the research institutions where these researches were completed (See Table 5).

Table 5. Frequencies for institutions

	Frequency	Percent	Valid Percent	Cumulative Percent
Georgia State University	3	4.8	4.8	14.3
Ghent University	2	3.2	3.2	17.5
Indiana University	2	3.2	3.2	20.6
Lancaster University	2	3.2	3.2	31.7
Michigan State University	2	3.2	3.2	39.7
University of Amsterdam	2	3.2	3.2	60.3
University of Alberta	2	3.2	3.2	100.0
Total	63	100.0	100.0	

The most frequent six keywords were defined in order to find the institutions where these keywords were produced. It was seen that K1 bilingualism was mostly produced in European universities such as Ghent University, Lancaster University and University College London. The second most frequent key word, K1 age was produced in the institutions located in Korea, New Zealand, Spain, the UK and the USA.

Table 6. Institutions according to frequent keywords

Institution	KWIC1						Total
	K1 accent	K1 advanced learners	K1 age	K1 bilingualism	K1 classroom discourse	K1 classroom-based research	
Aoyama Gakuin University	0	0	0	0	0	1	1
Bar-Ilan University	0	0	0	1	0	0	1
Brigham Young University	0	0	1	0	0	0	1
Georgia State University	0	1	0	0	0	1	2
Ghent University	0	0	0	2	0	0	2
Indiana University	0	0	0	0	1	0	1
James Madison University	1	0	0	0	0	0	1
Korea Military Academy	0	0	1	0	0	0	1
Laboratoire Parole et Langage	0	0	0	1	0	0	1
Lancaster University	0	0	0	1	0	0	1
Leiden University	0	0	0	1	0	0	1
Lund University	0	1	0	0	0	0	1
Universidad Complutense de	0	0	0	0	1	0	1

Institution	KWIC1						Total
	K1 accent	K1 advanced learners	K1 age	K1 bilingualism	K1 classroom discourse	K1 classroom-based research	
Madrid							
Universitat de Barcelona	0	0	1	0	0	0	1
University College London	0	0	0	1	0	0	1
University of Alberta	2	0	0	0	0	0	2
University of Leeds	0	0	1	0	0	0	1
University of Nottingham	0	0	0	0	0	1	1
University of Queensland	0	0	0	0	1	0	1
University of South Florida	0	0	0	1	0	0	1
Victoria University of Wellington	0	0	1	0	0	0	1
Virginia Polytechnic Institute	0	1	0	0	0	0	1
Total	3	3	5	8	3	3	25

The locations of the institutions were also diverse from each other and there were countries from different continents. However, the institutions located in the USA, UK and Canada were mainly leading. The institutions from European countries like Belgium, Holland, Sweden and Germany were followed.

Table 7. Frequencies for location of institution

	Frequency	Percent	Valid Percent	Cumulative Percent
Australia	3	4.8	4.8	4.8
Belgium	2	3.2	3.2	7.9
Canada	6	9.5	9.5	17.5
Denmark	1	1.6	1.6	19.0
Finland	1	1.6	1.6	20.6
France	1	1.6	1.6	22.2
Germany	1	1.6	1.6	23.8
Holland	4	6.3	6.3	30.2
Iran	1	1.6	1.6	31.7
Israel	2	3.2	3.2	34.9
Italy	1	1.6	1.6	36.5
Japan	1	1.6	1.6	38.1

	Frequency	Percent	Valid Percent	Cumulative Percent
Korea	1	1.6	1.6	39.7
New Zealand	2	3.2	3.2	42.9
Singapore	1	1.6	1.6	44.4
Spain	2	3.2	3.2	47.6
Sweden	2	3.2	3.2	50.8
Taiwan	1	1.6	1.6	52.4
UK	8	12.7	12.7	65.1
USA	22	34.9	34.9	100.0
Total	63	100.0	100.0	

6. Discussion and conclusion

All in all the current study described the frequent keywords used in the journal articles by the researchers from different institutions and countries. 287 articles from both journals were collected and compiled by means of Zotero, purified and analyzed through a corpus detection methodology over AntConc and a statistical analysis program. In accordance with the hypothesis and research questions, the most frequent keywords were defined in annual bases and the key words such as bilingualism, age, advanced learners, accent, classroom discourse, classroom-based research were the most frequent keywords produced by the researchers. Therefore, it can be said that mentioning that most researchers tended to research on “bilingualism, age, and classroom” in the journals mentioned. It is an interesting result to visualize that the keywords beginning with “classroom” is one of the leading ones since it may indicate that the researchers still mostly study on classroom applications for English language teaching and learning. The institutions and countries where these studies conducted were also indicated. Accordingly, the leading location of the universities was mainly in the USA but the European institutions were the places where the most frequent keyword, bilingualism, produced.

All the data analyzed provided very limited results in terms of the frequency distributions and therefore it seems that more data is needed to reach more relevant results and comments.

Acknowledgements

We hereby declare that this study is our unaided work and that this paper has not been submitted for any other degree or award before. This research was not financially supported by any Turkish institution. The paper is an open access study and can be accessed by all interested researchers, teachers and learners. We hereby thank Adriana Ferreira R.S. Silva, B.A. for her unique contributions, feedback and comments about the paper.

References

- Berber Sardinha, T. (1999) Using KeyWords in text analysis: Practical aspects. *DIRECT Papers* 42:1-8. São Paulo and Liverpool
- Bernardini, S. (2004) Corpora in the classroom: An overview and some reflections on future developments In J.McH. Sinclair (Ed.) *How to use corpora in language teaching* (pp.15-35). Amsterdam/Philadelphia: John Benjamins Publishing Company
- Buckmaster, R. (2015) *Text Analysis and Corpus Linguistics for English Language Teachers. English Ideas.*
- Burnard, L., & McEnery, T. (Eds.). (2000). Rethinking language pedagogy from a corpus perspective: *Papers from the third international conference on teaching and language corpora* (Lodz Studies in Language). Hamburg, Germany: Peter Lang.
- Conway, M. (2010). Mining a corpus of biographical texts using keywords. *Literary and Linguistic Computing*, 25(1), 23-35
- Giles, C. L., Bollacker, K. D., & Lawrence, S. (1998). Citeseer: An automatic citation indexing system. In I. Witten, R. Akscyn, & F. Shipman (Eds.), *Third ACM conference on digital libraries* (pp. 89–98). New York: ACM Press.
- Smit, E., & Van Der Graaf, M. (2012). Journal article mining: the scholarly publishers' perspective. *Learned Publishing*, 25(1), 35–46.
- Soriano, J., Au, T., & Banks, D. (2013). Text mining in computational advertising. *Statistical Analysis and Data Mining*, 6(4), 273–285.
- Stuart, K.& Botella, A. (2009). Corpus linguistics, network analysis and co-occurrence matrices. *International Journal of English Studies, Special Issue*, 1-28.
- Stubbs, M. (1996). *Text and corpus analysis: Computer-assisted studies of language and culture.* Massachusetts: Blackwell Publishers.
- Thompson, P. (2001) Looking at citations: Using corpora in English for academic purposes. *Language Learning & Technology*, 5(3), 91-105.
- Tsui, A. B.M (2004) What teachers have always wanted to know – and how corpora can help. In J. McH. Sinclair (Ed.), *How to use corpora in language teaching* (pp. 39-61). Amsterdam/Philadelphia: John Benjamins Publishing Company
- Wittaker, J., Courtial, J. P. & Law, J. (1989). Creativity and conformity in science: Titles, keywords and co-word analysis. *Social Studies of Science*, 19(3), 473-496.

Note:

- ¹ See SCImago Journal Ranking Portal <<http://scimagojr.com/index.php>>

Copyrights

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

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (**CC BY-NC-ND**) (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).