

# A Bibliometric Analysis of Linguistics Publications in the Web of Science

Mohammed Ali Mohsen<sup>1</sup>, Hui-Zhen Fu<sup>2</sup>, Yuh-Shan Ho<sup>3\*</sup>

<sup>1</sup>Najran University, College of Arts and Sciences, P.O. Box. 1988, Najran, SAUDI ARABIA.

<sup>2</sup>Information Resources Management, School of Public Affairs, Zhejiang University, Hangzhou 310058, Zhejiang, People's Republic of CHINA.

<sup>3</sup>Trend Research Centre, Asia University, No. 500, Lioufeng Road, Wufeng, Taichung County 41354, TAIWAN.

## ABSTRACT

This article investigates the world's publications in the category of linguistics based on Thomson Reuters' Social Science Citation Index (SSCI) from 2005-2014. Research achievements were recorded and analyzed, including publication year, characteristics of journals, productivity of authors, institutions, and countries, publication years, citation life of highly cited articles. Results found that while the number of publications dramatically increased, the average number of authors, cited references, and count page in proportion to the number of publications per year remained quietly stable. The USA was the most productive country in all the bibliometric indicators; total of publications single country articles, national and internationally collaborative article; of which University of Illinois in USA ranked the first institution all these indicators. M. On slow from the University of Melbourne was the most productive author. Journal of Memory and Language was the most productive venue of publication; among which, Baayen's et al. (2008) article was seen as the most top article. Implications of the study findings to researchers and research policy makers are highlighted at the end of the article.

**Keywords:** Linguistics, SSCI, Scientometrics, Web of Science Core Collection, Citations.

## INTRODUCTION

Web of Sciences database is almost the best standard index to indicate the research performance across different disciplines. Hence, it becomes crucial to assess the research performance in a category to see productivity of academics, countries, institutions that have been done so far.<sup>[1,2]</sup> A bibliometric analysis is an efficient way to quantify the quality of published work for organization, authors, and countries by analysing the data obtained from several indices and converting it into numerical Figures. Research has analyzed the research outcome from SSCI across different

fields such as business, business finance, economics, management,<sup>[3]</sup> nursing,<sup>[4]</sup> social work,<sup>[5]</sup> health care sciences and services,<sup>[6]</sup> psychiatry,<sup>[7]</sup> and information science and library science.<sup>[8]</sup> Bibliometric analyses have also been carried out on research from A and HCI to evaluate the quality of published work in various disciplines, for example dance<sup>[9]</sup> and religions.<sup>[10]</sup> Linguistics category, like other categories, has a long history, dating back to the work of von Humboldt and Saussure in 19<sup>th</sup> century and to the Bloomfield in the early 20<sup>th</sup> century.<sup>[11]</sup> Yet, this work had not been documented in research papers. The first leading periodical in linguistics category that indexed in SSCI is *The Modern Language Journal* which was launched in 1916, starting with 29 documents; 12 articles and 17 book reviews, and had gained 11% of world's publications till the year 2014. One year later, *The International Journal of American Linguistics* had been created, and the *Language Journal* has appeared in the mid of 1920s, covering 193 documents up to the year 1930. In the subsequent decades, (i.e, 1930, 1940s, and 1950s), 12,949 documents were published in 14 journals. Publications had been tremendously increased in the 1960s to have 98% of the publications appeared in the period between 1930-1960. These dramatic changes could be ascribed to the work of Noam Chomsky with his well-known theory "Innatist

\*Address for correspondence:

**Yuh-Shan Ho**

Trend Research Centre, Asia University, No. 500,  
Lioufeng Road, Wufeng, Taichung County 41354,  
TAIWAN. Phone no: 886 4 2332 3456  
E-mail Id: ysho@asia.edu.tw

### Access this article online

Official Publication of	
	Website: www.jscores.org
	DOI: 10.5530/jscores.6.2.16

Theory” as a reaction to behaviorists’ views of language acquisition proposed by Skinner in the beginning of 20<sup>th</sup> century. Publications had been increased over the decades in parallel with increase in the venue of publications to gain its height rate in the first decade of the new millennium; 34,130 records published in 188 journals, covering several sub-areas of linguistics such as applied linguistics, theoretical linguistics, psycholinguistics, neurolinguistics, computational linguistics, and translation. It is of paramount importance to analyze the productivity of Linguistics category across countries, authors, and institutions to see the world’s contributions in the last ten years to see the productivity over territories, authors, institutions, and venue of publication.

We claim that there is no single bibliometric study that investigated thoroughly the world’s publications in the field of linguistics. There was a bibliometric study in this regard<sup>[11]</sup> but it focused solely on a comparison of linguistics’ publications between SSCI and A and HCI. There was also an attempt made by Mohsen<sup>[12]</sup> to examine the attitudes of researchers towards publishing in the Web of Science but it was descriptive and subjective-based. However, Arik<sup>[11]</sup> did not analyze other factors such as citation per publications, the productivity across nations, authors, areas of research, organizations. This bibliometric study aims to investigate thoroughly world’s publications in Linguistics category in SSCI from 2004–2014 to see how this category had been researched and what are the research trends run in the last ten years period of examination.

## METHODOLOGY

The analysis provided in this study is based on the Social Sciences Citation Index (SSCI) database of the Web of Science Core Collection of Thomson Reuters (updated on February 05<sup>th</sup>, 2016). According to Journal Citation Reports (JCR) of 2014, there are 3,154 journals with citation references across 56 scientific disciplines in SSCI edition. 172 of those journals are listed in the category of linguistics. A total of 2,261,354 documents from 2005 to 2014 were found in SSCI. Results were refined by selecting the Web of Science category of linguistics (50,351 documents). Therefore, 2.2% of the total documents were published in the Web of Science belong to category of linguistics, including articles, book reviews, editorial materials, proceedings papers, reviews, meeting abstracts, biographical items, corrections, letters, bibliography, news items, book chapters, software reviews, reprints, books, and hardware reviews. The total articles retrieved for further analysis were 33,479. We chose articles in this bibliometric analysis because they contain description of complete researches and results.<sup>[13]</sup> Data about those articles and the total annual citations for each article were downloaded. All results were analyzed using Microsoft Excel 2013.<sup>[14]</sup>

The total number of citations from Web of Science Core Collection since publication to the end of 2014 was referred to as the  $TC_{2014}$ <sup>[15,16]</sup>. The advantage of  $TC_{year}$  is that it is an invariant parameter, thus ensuring repeatability, in comparison with the index of citation from Web of Science Core Collection which has been updated from time to time.<sup>[17]</sup> Citations per publication ( $CPP_{2104}$ ) is  $TC_{2014}/\text{total publication}$ .<sup>[18,19]</sup>

## RESULTS AND DISCUSSION

We analyze the data in terms of publication year, productivity of authors, countries and institutions, sourcing titles, citations per publication. All acronyms are explained beneath the tables.

### Publication Year

Several characteristics have been launched for the journals’ articles for the timespan set for this study<sup>[20]</sup>. Table 1 shows that number of publications had been dramatically increased. Number of publications in 2014 was almost four times of number of publications in 2005. Number of authors, references cited, and page count had been increased tremendously. However, the average number of each indicators in proportion to the number of publications per year remained quite stable. For example, 261% increase in the number of authors, but the average number of authors remained quite the same; from 2.2 in 2005 to 2.1 in 2014 with average of 2.0. Similarly, the average number of references cited in relation to total number of publications per year had not been changed significantly, reaching 45 in 2005 and 49 in 2014 with whole average of 46 for the total periods. The average number of pages remained also the same in proportion to the number of publication, obtaining 21 pages in 2005 and 20 pages in 2014. There was a steep in the average of citations per publication ( $CPP_{2104}$ ) over the ten years period, reaching the peak in 2005 with average of 17 and went down to 0.25 in 2014.

### Characteristics of the Venue of Publications

Only the top 15 journals were selected to be analysed for their quantity and quality. To analyse the journals’ quantity, number of publications was counted, and the percentage was calculated out of the total 195 journals indexed in linguistics category in the period between 2005 and 2014. Additionally, journal’s quality was assessed by their impact factor in Journal Citation Report (JCR) 2014. Results indicated that *Journal of Pragmatics* was ranked the first in number of publications (1,349 articles, 4.0% of 33,479 articles). Concerning journal’s quality, *Journal of Memory and Language* obtained the highest rank in terms of the impact factor ( $IF_{2014} = 4.23$ ), while *Dynamics of Language: An*

**Table 1: Characteristics of journal articles in Web of Science category of linguistics, 2005–2014.**

Year	TP	AU	AU/TP	NR	NR/TP	PG	PG/TP	TC2014	CPP2104
2005	1,257	2,722	2.2	56,056	45	26,471	21	21,506	17
2006	1,451	3,049	2.1	65,412	45	29,790	21	22,234	15
2007	2,026	3,908	1.9	86,616	43	42,311	21	21,540	11
2008	3,159	5,781	1.8	128,382	41	63,601	20	25,240	8.0
2009	3,611	6,649	1.8	148,923	41	71,932	20	21,933	6.1
2010	3,997	7,494	1.9	179,509	45	78,943	20	17,928	4.5
2011	4,257	8,129	1.9	196,800	46	83,791	20	13,227	3.1
2012	4,329	8,470	2.0	206,568	48	86,924	20	8,580	2.0
2013	4,623	9,417	2.0	224,920	49	91,394	20	4,585	1.0
2014	4,769	9,831	2.1	232,888	49	95,000	20	1,208	0.25
Total	33,479	65,450		1,526,074		670,157		157,981	
Average			2.0		46		20		4.7

TP: total number of articles; AU: number of authors; NR: number of cited references; PG: number of pages; TC<sub>2014</sub>: total citations since publication to the end of 2014; CPP<sub>2104</sub>: TC<sub>2014</sub>/TP

**Table 2: Top 15 published journals in Web of Science Category of Linguistics.**

Journals/Books	TP (%)	IF <sub>2014</sub> (rank)	CPP <sub>2104</sub> (rank)
Journal of Pragmatics	1,349 (4.0)	0.831 (57)	4.8 (43)
Journal of Speech Language and Hearing Research	1,211 (3.6)	2.070 (5)	11 (5)
Lingua	940 (2.8)	0.647 (72)	4.3 (46)
Brain and Language	858 (2.6)	3.215 (2)	13 (3)
Journal of Memory and Language	657 (2.0)	4.237 (1)	20 (2)
Clinical Linguistics and Phonetics	617 (1.8)	0.575 (82)	4.1 (49)
International Journal of Language and Communication Disorders	461 (1.4)	1.471 (22)	7.1 (28)
Journal of Child Language	446 (1.3)	1.598 (14)	7.3 (26)
Language and Cognitive Processes	439 (1.3)	2.134 (4)	10 (9)
Language Sciences	416 (1.2)	0.826 (58)	2.4 (85)
System	401 (1.2)	0.721 (62)	3 (75)
Hispania-A Journal Devoted to the Teaching of Spanish and Portuguese	395 (1.2)	0.054 (161)	0.57 (162)
Journal of Phonetics	382 (1.1)	1.598 (14)	8.1 (21)
Modern Language Journal	370 (1.1)	0.942 (51)	7.9 (24)
Foreign Language Annals	370 (1.1)	0.875 (55)	3.5 (62)

TP: total number of articles; IF<sub>2014</sub>: impact factor in 2014; CPP<sub>2104</sub>: TC<sub>2014</sub>/TP

*Introduction* (book series) got the highest rank in CPP<sub>2104</sub>, gaining 20 Table 2.

### Institutions and Countries

Publications over institutions and countries were analysed and compared to see the most productive institution and country in the category of linguistics. To avoid ambigu-

ity of territories' names, articles originating from England, Scotland, Northern Ireland, and Wales were reclassified as being from the United Kingdom (UK). Excluding 546 articles (1.6% of 33,479 articles) without any author address information on the Thomson Reuters Web of Science, the remaining were 32,933 articles originated from 137 countries. Among those articles with author affiliations, 28,105 (85%) were independent publications by 116

countries, while 4,828 (15%) were internationally collaborative publications from 120 countries. In recent years, Ho and co-workers proposed several indicators to examine the publication performance of countries and institutions, including total publications (*TP*), independent publications (*IP*), collaborative publications (*CP*), first authored publications (*FP*), corresponding authored publications (*RP*), and single authored publications (*SP*).<sup>[21]</sup>

Table 3 shows characteristics of top 18 contributing countries that published over 400 articles in the category of linguistics. Concerning the productivity over countries, USA obtained the first place in all the indicators; *TP* (11,738 articles), *SP* (6,660; 57%), *ICP* (2,062; 18%), and *NCP* (3,016; 26%), followed by the United Kingdom in all the indicators; *TP* (4,521 articles), *SP*(2,454; 54%), *ICP* (1,417; 31%), and *NCP* (650; 14%). Canada was ranked in the third top countries in total number of publications (1,965 articles), but got the fourth place in *SP* (1,020; 52%), *ICP* (651; 33%), and *NCP* (294; 15%), while Spain gained the third rank in *SP* (1,181; 64%) and the fourth place in *TP* (1,482 articles). Germany achieved the third place in internationally collaborative research (697; 39%)

while Australia was ranked in the third place in nationally collaborative research (310; 18%).

A ratio (*S:N:I*) of percentage of single institute articles in a country (*%SP*): percentage of nationally collaborative articles in a country (*%NCP*): percentage of internationally collaborative articles in a country (*%ICP*) might be used to describe institutions or countries publication characteristics.<sup>[22]</sup> USA's ratio of *S:N : I* is 57: 26: 18 in the Web of Science category of linguistics. It revealed that USA's research in linguistics field had more national teamwork or collaboration among different institutions than international collaborations among different countries. USA was the only one that had more national than international collaborations Table 3. South Africa was more inclined or able to conduct research independently with *S :N : I* = 70 : 7.0 : 23. Thirteen countries in Table 3 had at least 50 % institutional independent articles of their total publications. Germany and Spain published similar articles; however *S :N : I* shows that Germany (49 : 12 : 39) published much more internationally collaborative articles while Spain (64 : 15 : 20) published much more single institute articles. Average *S :N : I* for 18 countries in Table 3 was 54

**Table 3: Characteristics of top 18 contributing countries (TP > 400).**

Country	Rank (TP)	Rank (SP)	Rank (NCP)	Rank (ICP)	%SP	%NCP	%ICP
USA	1 (11,738)	1 (6,660)	1 (3,016)	1 (2,062)	57	26	18
UK	2 (4,521)	2 (2,454)	2 (650)	2 (1,417)	54	14	31
Canada	3 (1,965)	4 (1,020)	4 (294)	4 (651)	52	15	33
Spain	4 (1,842)	3 (1,181)	5 (284)	8 (377)	64	15	20
Germany	5 (1,810)	6 (887)	7 (226)	3 (697)	49	12	39
Australia	6 (1,739)	5 (899)	3 (310)	6 (530)	52	18	30
Netherlands	7 (1,387)	8 (603)	8 (214)	5 (570)	43	15	41
China	8 (1,257)	7 (708)	9 (172)	8 (377)	56	14	30
France	9 (1,054)	10 (434)	6 (237)	7 (383)	41	22	36
Japan	10 (793)	11 (417)	10 (145)	12 (231)	53	18	29
Belgium	11 (777)	13 (373)	11 (132)	10 (272)	48	17	35
Italy	12 (703)	14 (341)	12 (130)	11 (232)	49	18	33
South Africa	13 (678)	9 (473)	18 (50)	14 (155)	70	7.4	23
Taiwan	14 (592)	12 (395)	15 (85)	19 (112)	67	14	19
Sweden	15 (545)	15 (293)	13 (92)	13 (160)	54	17	29
Israel	16 (472)	16 (278)	16 (77)	18 (117)	59	16	25
New Zealand	17 (448)	17 (258)	21 (35)	14 (155)	58	7.8	35
Finland	18 (417)	18 (223)	14 (86)	20 (108)	53	21	26

TP: total articles; SP: single institute articles; NCP: nationally collaborative articles; ICP: internationally collaborative articles; %SP: percentage of single institute articles in a country; %NCP: percentage of nationally collaborative articles in a country; %ICP: percentage of internationally collaborative articles in a country.

**Table 4: Top 20 most productive institutes during 2005-2014.**

Institute	Rank (TP)	Rank (SP)	Rank (NCP)	Rank (ICP)	%SP	%NCP	%ICP
University of Illinois, USA	1 (386)	1 (167)	1 (153)	17 (66)	43	40	17
University of Toronto, Canada	2 (340)	2 (151)	13 (80)	4 (109)	44	24	32
University of Edinburgh, UK	3 (334)	5 (124)	28 (68)	2 (142)	37	20	43
UCL, UK	4 (314)	4 (125)	8 (92)	6 (97)	40	29	31
University of Wisconsin, USA	5 (308)	10 (116)	2 (147)	40 (45)	38	48	15
Radboud University Nijmegen, Netherlands	6 (294)	38 (83)	22 (71)	3 (140)	28	24	48
Penn State University, USA	7 (272)	14 (112)	5 (101)	22 (59)	41	37	22
University of Manchester, UK	8 (270)	7 (118)	30 (67)	9 (85)	44	25	31
University of Ghent, Belgium	9 (267)	9 (117)	50 (51)	5 (99)	44	19	37
Ohio State University, USA	10 (265)	7 (118)	3 (112)	60 (35)	45	42	13
University of Maryland, USA	11 (264)	15 (109)	4 (102)	34 (53)	41	39	20
University of Queensland, Australia	12 (257)	32 (87)	6 (96)	12 (74)	34	37	29
Indiana University, USA	13 (255)	3 (133)	9 (87)	60 (35)	52	34	14
University of Hong Kong, China	14 (250)	17 (106)	47 (53)	7 (91)	42	21	36
Max Planck Institute for Psycholinguistics, Netherlands	15 (241)	91 (50)	56 (46)	1 (145)	21	19	60
University of Arizona, USA	16 (236)	13 (114)	15 (79)	43 (43)	48	33	18
University of London, UK	17 (235)	10 (116)	34 (62)	29 (57)	49	26	24
University of Sydney, Australia	18 (224)	35 (85)	20 (74)	19 (65)	38	33	29
University of Amsterdam, Netherlands	19 (221)	15 (109)	41 (57)	30 (55)	49	26	25
University of Alberta, Canada	20 (218)	20 (99)	68 (42)	10 (77)	45	19	35

TP: total articles; SP: single institute articles; NCP: nationally collaborative articles; ICP: internationally collaborative articles; %SP: percentage of single institute articles in a country; %NCP: percentage of nationally collaborative articles in a country; %ICP: percentage of internationally collaborative articles in a country.

: 16 : 30. However, average S : N : I for 137 countries in linguistics field was found to be 50 : 8.0 : 40. “SNI” shows three important rates related with collaboration which was comprehensively and visually higher than just one traditional collaboration rate at a time for measurement and comparison.

Of all the 32,933 articles with author addresses in Web of Science, 21,295 (65%) were institutional independent articles and 11,638 (35%) were collaborations by two or more institutes. Among the top 20 institutes, seven were in the USA, four in the UK, three in Netherlands, two in Australia, two in Canada, and one each in Belgium and China Table 4. University of Illinois got the first rank in the total of articles (TP = 386), in SP (167; 42%), and in NCP (153; 40%). University of Toronto in Canada achieved the second rank in both TP and SP. Max Planck Institute for Psycholinguistics in Netherlands obtained the first rank in ICP (145; 60%). A pattern of S:N : I = 41: 30: 29 was found for all institutions in Table 4 and 31: 44: 25 for all institu-

tions in in linguistics field. University of Illinois (43%) and University of Toronto (44%) published almost the same percentage of single institute articles in in linguistics field. University of Illinois had more national collaborations while University of Toronto had more international collaborations. In the USA, Indiana University (52 :34: 14) had the highest percentage of independent articles while University of Wisconsin (38: 48: 15) had the highest percentage of nationally collaborative articles.

### Citation Life Cycles of Highly Cited Articles

The citation history of papers gives more details of the impact character of articles<sup>[16]</sup>. In last decade, citation life cycles of highly cited articles were discussed in number of research topics<sup>[17,21]</sup>. High citation of a published document indicates the more visible of that cited document. The higher the citation, the more value a document will likely get. Figure 1 shows the highly cited articles in linguistics category. As it shown in Table 5 and Fig. 1, the

**Table 5: Ten most frequently cited articles in Web of Science category of linguistics.**

Rank (TC <sub>2014</sub> )	Rank (C <sub>0</sub> )	Rank (C <sub>2014</sub> )	Rank (TCPY)	Article title	References
1 (1,172)	51 (5)	1 (338)	1 (167)	Mixed-effects modeling with crossed random effects for subjects and items	Baayen et al. (2008)
2 (506)	1287 (1)	3 (160)	3 (72)	Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models	Jaeger (2008)
3 (288)	174 (3)	15 (33)	4 (32)	Evaluating WordNet-based measures of lexical semantic relatedness	Budanitsky and Hirst (2006)
4 (248)	174 (3)	23 (27)	6 (25)	Improved picture naming in chronic aphasia after TMS to part of right Broca's area: An open-protocol study	Naeser et al. (2005)
5 (226)	4553 (0)	5 (45)	6 (25)	From usage to grammar: The mind's response to repetition	Bybee (2006)
6 (221)	1287 (1)	6 (44)	10 (22)	The Proposition Bank: An annotated corpus of semantic roles	Palmer et al. (2005)
7 (198)	1 (26)	2 (172)	2 (99)	Random effects structure for confirmatory hypothesis testing: Keep it maximal	Barr et al. (2013)
8 (177)	4553 (0)	28 (25)	17 (18)	The independence of combinatory semantic processing: Evidence from event-related potentials	Kim and Osterhout (2005)
9 (169)	4553 (0)	9 (41)	8 (24)	Variation and the indexical field	Eckert (2008)
10 (157)	4553 (0)	24 (26)	20 (17)	Cortical interactions underlying the production of speech sounds	Guenther (2006)

TC<sub>2014</sub>: number of citations since publication to the end of 2014; C<sub>2014</sub>: number of citations in 2014; C<sub>0</sub>: number of citations in publication year; TCPY: TC<sub>2014</sub> per number of years

article entitled “Mixed-effects modelling with crossed random effects for subjects and items” published in *Journal of Memory and Language* by<sup>[23]</sup> Baayen et al. got the first rank in almost all the citations parameters; total citations since publication to the end of 2014 (TC<sub>2014</sub> = 1,172); total citations in 2014 (C<sub>2014</sub> = 338); and total citations per year (TCPY=167). Baayen et al.<sup>[23]</sup> introduced mixed-effects models for the analysis of repeated measurement data with subjects and items as crossed random effects. R.H. Baayen<sup>[23]</sup> as both first and corresponding authors from University of Alberta in Canada published an article with a “distinguished pattern”<sup>[24,25]</sup> in linguistics field Figure 1. A steep slope could be found with the distinguished patterns of citations per year<sup>[24]</sup>. Article entitled “Random effects structure for confirmatory hypothesis testing: Keep it maximal”<sup>[26]</sup>, published in *Journal of Memory and Language*, got the first rank in the total citation of the publication year (C<sub>0</sub> = 26) and the sec position in the total citation in 2014 (C<sub>2014</sub> = 172). This paper stated that researchers using Linear mixed-effects models (LMEMs) for confirmatory hypothesis testing should minimally adhere to the standards that have been in place for many decades<sup>[26]</sup>. Furthermore, Jaeger’s<sup>[27]</sup> article entitled “Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models” obtained sec and third ranks in TC<sub>2014</sub> and C<sub>2014</sub> respectively. This paper identified several serious problems with the widespread use of ANOVAs for the analysis of categorical outcome variables such as forced-choice variables, question-answer accuracy, choice in production, et cetera<sup>[27]</sup>. The third rank in TC<sub>2014</sub> was for Budanitsky and Hirst’s<sup>[28]</sup> article entitled “Evaluating WordNet-based measures of lexical semantic relatedness” which was published in *Computational Linguistics Journal*. Budanitsky and Hirst’s<sup>[28]</sup> article also obtained the fourth place in total citation per year (TCPY=32). Table 5 also shows that, Naeser’s et al.<sup>[29]</sup> arti-

cle entitled “Improved picture naming in chronic aphasia after TMS to part of right Broca’s area: An open-protocol study”, which was published in *Brain and Language Journal*, got the fourth rank in TC<sub>2014</sub>. Most of the top cited articles had low C<sub>0</sub>, which meant the citations received at the beginning after publication did not influence the total citations. It is worth-mentioning that the journals which got high citations, and thus obtained high impact factors, are intersected with other categories such as neuroscience, audiology and speech language pathology, psychology, computer science, and artificial linguistics.

**Authors and Publication Characteristics of Authors**

Several bibliometric indicators are used to assess authors’ productivity. These indicators include (1) total number of publications, (2) single author publication, (3) first author articles, (4) corresponding author articles, and (5) Y-index.<sup>[8]</sup> Order of authors is crucial to recognize who has the most substantial role in doing and reporting the research and to reveal the magnitude of authors’ contribution. It is widely known that the first author has the major responsibility in the project’s leadership and primary role for publication while the sec author is to contribute in article’s analysis or writing.<sup>[30]</sup> The leaders of scientific groups appear in the last place (or in the first place) of the authors’ lists of multi-author papers more often than other scientists (group-members).<sup>[31]</sup>

Ho<sup>[32,33]</sup> proposed Y-index<sup>[27]</sup> to evaluate publication characteristics of authors and institutions for highly cited articles. The Y-index is related to numbers of first author publications (FP) and corresponding author publications (RP), defined as:

$$j = FP - RP \tag{1}$$

$$h = \tan^{-1} \left( \frac{RP}{FP} \right) \quad (2)$$

The higher the  $j$  an author has, the more potential the author must take the first order among the list of authors or to act as a corresponding author. Being the first or corresponding author means to play the major role in leading the publication. The value of  $h$  indicates publication characteristic constant that differentiates the nature of the leadership role. When  $h > 0.7854$ , it indicates more corresponding author publications and when  $h < 0.7854$ , means more first author publications. When  $h = 0$ ,  $j =$  the number of first author publications and when  $h = \pi/2$ ,  $j =$  the number of corresponding author publications. There is a limitation for analysis of publication characteristics of authors and institutions by Y-index. Only articles with both first author and corresponding author information could be considered. A total of 33,244 articles by 34,308 authors in linguistics field were analysed. Figure 2 displays the distribution of the top 32 authors with  $j \geq 25$ , ( $j \cos h$  and  $j \sin h$  are chosen as the x and y coordinate axes). Each dot represents one value that could be one author or many authors. The author who had the most potential to publish articles in linguistics field was M. Onslow ( $j = 43$ ), followed by R.A. van Compernelle ( $j = 42$ ), and L.B. Leonard and S. Montrul with  $j = 40$  respectively. Publication characteristics constant,  $h$ , could help to obtain the different proportion of corresponding author articles to first author articles. The advantage of the Y-index is that, when  $j$  of authors is the same, publication characteristics of authors can be indicated by  $h$ . For example, the  $j$  of L.B. Leonard and S. Montrul were both the same of 40 Figure 2. However  $h$  of Leonard was 0.9343 but  $h$  of Barrera was 0.7854. Leonard had greater proportion of corresponding author articles to first author articles than Montrul. Within these 32 authors, T.A. Hall was the only author with  $h < 0.7854$  ( $h = 0.7454$ ). Hall had more first author articles than corresponding author articles, indicating that the top productive authors contributing to linguistics field were more likely to be designated as the corresponding authors. Of the 32 authors, 17 had more corresponding author articles than the first author articles. The remaining 14 authors were just on the boundary line ( $h = 0.7854$ ), having equal number of authorship characteristics; be first author or corresponding author. Figure 1 shows they had different publication potential. These authors probably contributed more to the initial conception and supervision of study.<sup>[34]</sup> A potential bias in analysis of authorship might occur when different authors have the same name or authors used different names over time in their articles.<sup>[20]</sup>

### Keywords Analysis

In recent years, Ho and co-workers proposed an indicator to infer research trends and its main foci.<sup>[24,35]</sup> Distribution

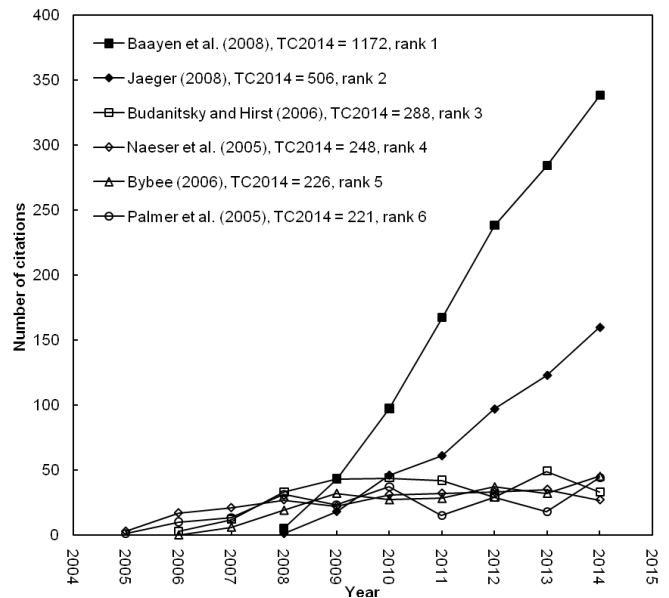


Figure 1: Citation lives of the top six publications with the TC2014 ≥ 220.

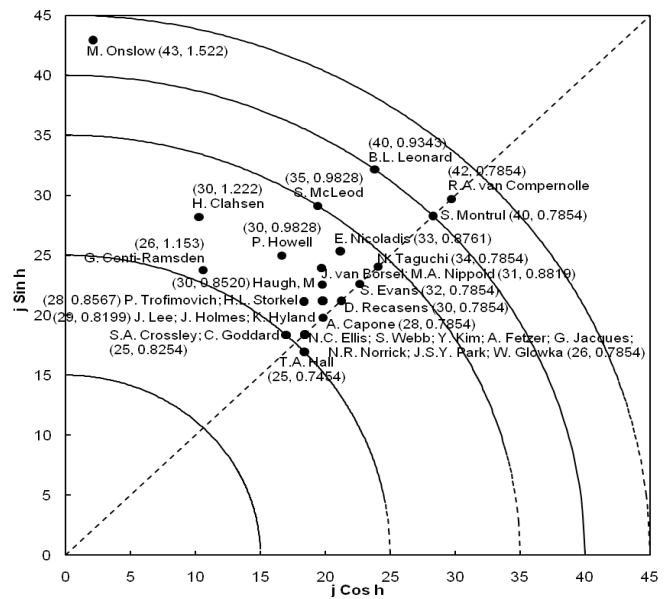


Figure 2: Top 32 authors with Y-index (j ≥ 25).

of title words, words in abstract, authors' keywords, and Key Words Plus over different periods of time have been thoroughly examined to elucidate research focus.<sup>[36,37]</sup> Key Words Plus in the Web of Science database supplied additional search terms extracted from the titles of articles cited by authors in their bibliographies and footnotes.<sup>[38]</sup> Word clusters has been further applied to determine the research foci and to characterize article lives in subsequent years.<sup>[24,35]</sup> Among 33,479 articles in linguistics field indexed in SSCI, 31,894 articles (95%) had recorded information of abstract, 24,248 articles (72%) had recorded information of author keywords while 22,748 articles (68%) with Key Words Plus information were analyzed.

### Words in titles

One way of measuring the research tendency is to analyze words in titles. Authors select words in titles carefully as to appear thought-provoking to readers to draw their attention to the published document to detect the research focus, the variable investigated, and research environment. Words in titles are commonly analyzed in bibliometric studies to see the inferences of the scientific literature and the highly demanding research areas.<sup>[35]</sup> In this study, we analyzed only the content words appeared in research titles while grammatical words such as prepositions, articles, conjunctions were discarded.<sup>[39]</sup> Eleven words appeared in titles more than 1,000 articles included “language” (5,938 articles; 18% of 33,479 articles), “English” (2,978; 8.9%), “children” (2,202; 6.6%), “speech” (1,860; 5.6%), “learning” (1,489; 4.4%), “case” (1,349; 4.0%), “effects” (1,223; 3.7%), “Spanish” (1,159; 3.5%), “discourse” (1,034; 3.1%), “evidence” (1,029; 3.1%), and “word” (1,014; 3.0%). The word “language” had been mostly used in research titles in the period 2005–2014. Further, the average of using the word “language” over the years were almost the same.

The second most used word was “English” which indicated that research had mostly investigated English language as it is widely spoken and written worldwide, followed by “Spanish” which was also most repeated in research titles, gaining the 8<sup>th</sup> rank. Words such as “learning”, “speech”, and “children” showed that most research trends in the target 10 years period was focusing speech pathology as well as language acquisition area

### Author keywords

Author keywords analysis may reveal the most areas of interest to researchers. Recently, the managing submissions systems used by journals ask authors to place their keywords, which would be later used by journals’ editors to assign reviewers for new submissions reach journals. Author keywords had been used for analysis in recent bibliometric studies<sup>[36]</sup> to track the research hotspots and major directions of scientific research.<sup>[16]</sup> Author keywords appeared in the articles were calculated and ranked by total 10-year and 2-year sub-period. In total 49,499 authors

**Table 6: The 20 most frequently used author keywords.**

Author keywords	TP	05-14 R (%)	05-06 R (%)	07-08 R (%)	09-10 R (%)	11-12 R (%)	13-14 R (%)
Language	466	1 (1.9)	1 (2.3)	1 (1.9)	3 (1.8)	3 (1.8)	3 (1.7)
Bilingualism	433	2 (1.8)	9 (1.7)	3 (1.5)	8 (1.3)	8 (1.3)	1 (2.1)
Identity	413	3 (1.7)	20 (1.1)	2 (1.7)	1 (1.9)	1 (1.9)	2 (2.0)
Conversation analysis	386	4 (1.6)	15 (1.3)	13 (1.2)	4 (1.7)	4 (1.7)	9 (1.3)
Spanish	370	5 (1.5)	53 (0.66)	6 (1.4)	2 (1.8)	2 (1.8)	6 (1.6)
Aphasia	364	6 (1.5)	2 (2.2)	4 (1.5)	10 (1.2)	10 (1.2)	5 (1.6)
Children	362	7 (1.5)	3 (2.0)	7 (1.4)	16 (1.1)	16 (1.1)	10 (1.2)
Stuttering	346	8 (1.4)	3 (2.0)	4 (1.5)	5 (1.6)	5 (1.6)	7 (1.5)
Syntax	340	9 (1.4)	9 (1.7)	16 (1.2)	9 (1.2)	9 (1.2)	4 (1.7)
English	317	10 (1.3)	16 (1.2)	10 (1.3)	6 (1.4)	6 (1.4)	11 (1.2)
Assessment	304	11 (1.3)	6 (1.9)	18 (1.0)	12 (1.2)	12 (1.2)	15 (1.1)
Phonology	302	12 (1.2)	7 (1.8)	10 (1.3)	6 (1.4)	6 (1.4)	17 (1.0)
Discourse	290	13 (1.2)	11 (1.6)	13 (1.2)	12 (1.2)	12 (1.2)	12 (1.2)
Prosody	282	14 (1.2)	8 (1.7)	17 (1.0)	15 (1.1)	15 (1.1)	14 (1.1)
Pragmatics	280	15 (1.2)	20 (1.1)	10 (1.3)	10 (1.2)	10 (1.2)	8 (1.3)
Specific language impairment	260	16 (1.1)	3 (2.0)	27 (0.79)	19 (1.0)	19 (1.0)	21 (1.0)
Narrative	259	17 (1.1)	16 (1.2)	8 (1.4)	20 (1.0)	20 (1.0)	13 (1.2)
Reading	257	18 (1.1)	13 (1.4)	23 (0.9)	27 (0.91)	27 (0.91)	17 (1.0)
Gender	254	19 (1.0)	22 (1.1)	19 (1.0)	12 (1.2)	12 (1.2)	21 (1.0)
Speech perception	251	20 (1.0)	11 (1.6)	22 (0.93)	30 (0.89)	30 (0.89)	25 (0.91)

TP: total articles; R: rank



words were used. Table 6 shows the most frequently used 20 words.

The top 20 author keywords showed that most of these words are mostly shared by different disciplines such as “speech pathology”, “applied linguistics”, and “discourse analysis”. Two words that were exclusively used by the “speech pathology” area; (1) “aphasia” which was ranked as the top seventh author keywords and “stuttering” which took the ninth position in the author keywords list Table 6. It has been argued that the authors’ keywords analysis may have some limitations to generalize the research findings in terms of research trends due to use of synonymous words, use of acronyms, and spelling variations. To avoid this problem, we checked the words retrieved from the Web of Science to see if the indicators were available in our set of data. We found two synonymous words in the data set; “sec language” and “L2”.

### Key Words Plus

*Key Words Plus* examined in bibliometric analysis to show the novelty in research<sup>[36]</sup>. According to Garfield<sup>[38]</sup>, *Key Words Plus* could be extracted from the titles of cited documents in authors’ biodata placed either in their bibliographies or footnotes in the Web of Science database, leading to augment the total words in titles and author keywords indexing. *Key Words Plus* were calculated for the 10 years with a total of two-year sub-period. Like the author keywords, results of the *Key Words Plus* showed that top frequent *Key Words Plus* were used by multiple areas of linguistics such as “applied linguistics”, “speech pathology”, “discourse analysis”, and “theoretical linguistics”. For example, the words “language”, “English”, “acquisition”, “children”, and “comprehension” were used much frequently in the targeted articles. The word “language” was placed in the first rank of the top frequent word in *Key Words Plus* during the 10-year period, followed by the word “English” as the sec rank of the most frequent *Key Words Plus*. The word “acquisition” was in the third rank which was exclusively related to the area of first or sec language acquisition (i.e., a sub-area of applied linguistics). All together 15,481 *Key Words Plus* were found in 22,748 articles. Only six *Key Words Plus* can be found in more than 1,000 articles including language (3,006 articles; 13% of 22,748 articles), English (2,331; 10%), acquisition (1,768; 7.8%), speech (1,587; 7.0%), children (1,403; 6.2%), and comprehension (1,025; 4.5%).

### CONCLUSION

An overview of the world’s research in the category of linguistics during 2005–2014 was presented in this bibliometric study. Several issues have been investigated to provide insights into the research tendencies of the world’s

publications in linguistics during the last ten years period. We observed a steady increase of publications in the category of linguistics, but the average of publication was quite stable in relation to the average of authors, cited references and count number. The USA, UK, Canada, Spain, and Germany were the most productive countries in the linguistics publications. However, the USA took the leading position in total of publication along with highly citations per years, followed by the UK. Highly cited venues of publications along with highly cited articles were those that studied the area of speech pathology which intersected with other areas such as psychology and neuroscience. Research trends as shown by analysis of words in titles, author keywords and keywords plus indicate that the area of speech pathology was the most researched area in the linguistics category. These study findings are of implications to the research policy makers to increase the publications in the category of linguistics particularly the Asian countries where the research in this area is still scarce. There is a myriad of languages in Asian countries which worth investigation by linguistics researchers to enhance the quantity and quality of research.

### REFERENCES

1. Rojas-Sola JI, and Aguilera-Garcia AI. Global bibliometric analysis of the ‘mining and mineral processing’ subject category from the Web of Science (1997-2012). *Mineral Processing and Extractive Metallurgy Review*. 2015;36(6):349-69.
2. Long X, Huang JZ, Ho YS. A historical review of classic articles in surgery field. *American Journal of Surgery*. 2014;208(5):841-9.
3. Benavides-Velasco CA, Quintana-García C, Guzmán-Parra VF. Trends in family business research. *Small Business Economics*. 2013;40(1):41-57.
4. Wong ELY, Tam WWS, Wong FCY and Cheung AWL. Citation classics in nursing journals: The top 50 most frequently cited articles from 1956 to 2011. *Nursing Research*. 2013;62(5):344-51.
5. Ho YS. Classic articles on social work field in Social Science Citation Index: A bibliometric analysis. *Scientometrics*. 2014;98(1):137-55.
6. Hsu YHE, Ho YS. Highly cited articles in health care sciences and services field in Science Citation Index Expanded: A bibliometric analysis for 1958-2012. *Methods of Information in Medicine*. 2012;53(6):446-58.
7. Igoumenou A, Ebmeier K, Roberts N, Fazel S. Geographic trends of scientific output and citation practices in psychiatry. *Bio Med Central Psychiatry* 14 Article Number. 2014;14(1):332.
8. Ivanović D, Ho YS. Highly cited articles in the Information Science and Library Science category in Social Science Citation Index: A bibliometric analysis. *Journal of Librarianship and Information Science*. 2016;48(1):36-46.
9. Ho HC, Ho YS. Publications in dance field in Arts & Humanities Citation Index: A bibliometric analysis. *Scientometrics*. 2015;105(2):1031-1040.
10. Engler S. Bibliometrics and the study of religion\\$. *Religion*. 2014;44(2):193-219.
11. Arik E. An experimental study of Turkish vowel harmony. *Poznan Studies in Contemporary Linguistics*. 2015;51: 359-374.
12. Mohsen M. Contributions of Saudi Institutions in Applied Linguistics’ Journals Indexed in SSCI: Perspectives from

- Academics and Journals' Editors. *International Journal of Applied Linguistics and English Literature*. 2016;5(4):102-9.
13. Ho YS, Satoh H, Lin SY. Japanese lung cancer research trends and performance in Science Citation Index. *Internal Medicine*. 2010;49(20):2219-28.
  14. Li Z, Ho YS. Use of citation per publication as an indicator to evaluate contingent valuation research. *Scientometrics*. 2008;75(1):97-110.
  15. Chuang KY, Wang MH, Ho YS. High-impact papers presented in the subject category of water resources in the Essential Science Indicators database of the Institute for Scientific Information. *Scientometrics*. 2011;87(3):551-62.
  16. Wang MH, Fu HZ, Ho YS. Comparison of universities' scientific performance using bibliometric indicators. *Malaysian Journal of Library and Information Science*. 2011;16(2):1-19.
  17. Fu HZ, Wang MH, Ho YS. The most frequently cited adsorption research articles in the Science Citation Index (Expanded). *Journal of Colloid and Interface Science*. 2012;379(1):148-56.
  18. Fu HZ, Ho YS. Independent research of China in Science Citation Index Expanded during 1980-2011. *Journal of Informetrics*. 2013;7(1):210-22.
  19. Ho YS. A bibliometric analysis of highly cited articles in materials science. *Current Science*. 2014;107(9):1565-72.
  20. Ho YS. Bibliometric analysis of adsorption technology in environmental science. *Journal of Environmental Protection Science*. 2007;1(1):1-11.
  21. Chiu WT, Ho YS. Bibliometric analysis of homeopathy research during the period of 1991 to 2003. *Scientometrics*. 2005;63(1):3-23.
  22. Fu HZ, Long X, Ho YS. China's research in chemical engineering journals in Science Citation Index Expanded: A bibliometric analysis. *Scientometrics*. 2014;98(1):119-36.
  23. Baayen RH, Davidson DJ, Bates DM. Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*. 2008;59(4):390-412.
  24. Lin CL, Ho YS. A bibliometric analysis of publications on pluripotent stem cell research. *Cell Journal*. 2015;17(1):59-70.
  25. Ho YS, Hartley J. Classic articles in psychology in the Science Citation Index Expanded: A bibliometric analysis. *British Journal of Psychology (in press)*. 2016;107(4):768-80.
  26. Barr DJ, Levy R, Scheepers C, Tily HJ. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*. 2013;68:255-278.
  27. Jaeger TF. Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*. 2008;59(4):434-46.
  28. Budanitsky A, Hirst G. Evaluating WordNet-based measures of lexical semantic relatedness. *Computational Linguistics*. 2006;32(1):13-47.
  29. Naeser MA, Martin PI, Nicholas M, Baker EH, Seekins H, Kobayashi M, *et al.* Improved picture naming in chronic aphasia after TMS to part of right Broca's area: An open-protocol study. *Brain and Language*. 2005;93(1):95-105.
  30. Osborne JW, Holland A. What is authorship and what should it be? A survey of prominent guidelines for determining authorship in scientific publications. *Practical Assessment Research and Evaluation*. 2009;14(15):1-19.
  31. Kosmulski M. The order in the lists of authors in multi-author papers revisited. *Journal of Informetric*. 2012;6(4):639-44.
  32. Ho YS. Top-cited articles in chemical engineering in Science Citation Index Expanded: A bibliometric analysis. *Chinese Journal of Chemical Engineering*. 2012;20(3):478-88.
  33. Ho YS. The top-cited research works in the Science Citation Index Expanded. *Scientometrics*. 2013;94(3):1297-312.
  34. Wren JD, Kozak KZ, Johnson KR, Deakyne SJ, Schilling LM, Dellavalle RP. The write position: A survey of perceived contributions to papers based on by-line position and number of authors. *EMBO Reports*. 2007;8(11):988-91.
  35. Mao N, Wang MH, Ho YS. A bibliometric study of the trend in articles related to risk assessment published in Science Citation Index. *Human and Ecological Risk Assessment*. 2010;16(4):801-24.
  36. Li LL, Ding GH, Feng N, Wang MH, Ho YS. Global stem cell research trend: Bibliometric analysis as a tool for mapping of trends from 1991 to 2006. *Scientometrics*. 2009;80(1):39-58.
  37. Zhang GF, Xie SD, Ho YS. A bibliometric analysis of world volatile organic compounds research trends. *Scientometrics*. 2010;83(2):477-92.
  38. Garfield E. Keywords plus-ISIS breakthrough retrieval method. 1. Expanding your searching power on current-contents on diskette. *Current Contents*. 1990;32:5-9.
  39. Tanaka H, Ho YS. Global trends and performances of desalination research. *Desalination and Water Treatment*. 2011;25(1-3):1-12.

**How to cite this article:** Mohsen MA, Fu HZ, Ho YS. A Bibliometric Analysis of Linguistics Publications in the Web of Science. *J Scientometric Res*. 2017;6(2):109-18.