

The Passive Voice in Abstracts in Research Articles: A Diachronic Variation across Disciplines

Fan Zhang, Rongping Cao*

School of Foreign Languages, Beijing Forestry University, Beijing, 100083, China

* Corresponding author: Rongping Cao (Email: caorongping@bjfu.edu.cn)

Abstract: The passive voice has been commonly used in academic writing for decades. This research aims to investigate the diachronic changes in the use of passive voice in abstracts across four disciplines in natural sciences and social sciences over the past fifteen years. The study corpus comprises 180 abstracts in four disciplines, ranging from 2008-2012, 2013-2017, and 2018-2022. Both tag tools and statistical tools were utilized to reveal the changes in the use of passive voice. The results indicated a slight decline in the use of passive voice in the 180 abstracts over the years, with an increase in natural sciences and a decline in social sciences. Additionally, in the past five years, there has been a discernible difference between the use of passive voice in natural sciences and social sciences, with natural science abstracts containing more passive voice. This research offers insights into changes in passive voice use in abstract writing and contributes to English for Academic Purposes (EAP) instruction.

Keywords: Abstract, Corpus-based, Disciplinary Variation, Passive Voice.

1. Introduction

The passive voice in academic writing is always considered as a means to achieve formality, in accordance with abstract vocabulary and complex sentence structure [1]. Due to the formal requirements of academic writing, it seems that the passive voice has been widely applied throughout the texts. The use of passive voice has long been investigated since Barber's article [2], which found that 28% of the non-modal verb forms were in the passive voice in the sample texts. Despite studies indicating the significance of passive voice in academic field, the research on the passive voice in academic writing has shown quite different results [3,4]. Some studies proposed the rise of the passive voice in academic writing [5], whereas some other studies demonstrated that the use of passive voice had witnessed a decline in academic writing in various fields, including both hard sciences and soft sciences [6,7]. The academic divergence on the use of passive voice shows the necessity of verifying the passive voice in academic writing.

Research on the use of passive voice often focuses on the comparison of different disciplines or the diachronic changes of the same discipline. Scholars demonstrated the decline of the passive voice in academic writing in natural sciences, while the social sciences received relatively less attention [8]. Nevertheless, studies showed that academic writing in social sciences show more passive voice than natural sciences [9]. Moreover, although studies also explored the diachronic changes of passive voice in the same disciplines, the evolutionary trajectory of passive voice is still uncertain [10]. Additionally, previous studies on passive voice mainly focuses on the whole academic texts or the extractions from the general corpora, yet few research investigate the passive voice in abstract sections.

Accordingly, this research aims to explore the diachronic variations in the use of the passive voice within the abstract sections of natural sciences and social sciences. Additionally, it endeavors to identify the disparities among different disciplines during the same time periods. By doing so, on one hand, the study attempts to furnish evidence regarding the

changes in the utilization of the passive voice in research article (RA) abstracts. On the other hand, it intends to offer valuable insights into the teaching and learning of English for Academic Purposes (EAP) within the context of English as a Foreign Language (EFL) education.

2. Literature Review

2.1. Passive Voice in Academic Writing

The use of passive voice has gradually increased in academic writing since the nineteenth century, driven by the pursuit of efficient communication and strong argument [8,11]. As a crucial tool for packaging information, the passive voice allows for the presentation of abstract information in an objective tone by focusing on the work rather than the agent responsible [3,4]. Consequently, it has become a primary means of achieving an impersonal and factual style. Studies also witnessed the rise of the passive voice in academic writings. For instance, Banks demonstrated that the use of finite verbs in the passive voice had increased from 21.2% in 1700 to 32.8% in 1980 [5].

However, the rise of the passive voice seems to be altered in recent years. Leech showed a decline in passive voice in all registers, particularly in academic prose [6]. Similarly, Seoane and Loureiro-Porto suggested that passives in hard science and medicine articles fell in both British scientific English and American scientific English [9]. The decline of passives has also occurred in other fields, for instance, in biology, engineering, and sociology, etc. [5,7,8]. Additionally, domestic scholars have explored the differences of the use of the passives between Chinese researchers and native English researchers, and found that the voice in academic papers of international journals has developed from passives to active voice, whereas Chinese researchers and learners often apply passives in research articles [12,13]. Moreover, studies also suggest the different changes between various disciplines and specific disciplines within the same discipline. For instance, Hundt showed that the actives dominate the text in natural sciences and technology while passives prevail in social sciences and humanities [9].

Despite the previous studies have showed the diachronic changes of passives within various disciplines, the dataset they chosen were limited. Some studies extracted academic texts from general corpora such as Brown, which cannot represent academic research comprehensively [10]. Meanwhile, some datasets were inadequate due to the small size or the single source of articles. Moreover, scarce attention was given to the passives in the abstract section of research articles, although the abstract always plays a critical role in an article.

2.2. Voice in Abstract

Abstract is a fundamental element of research articles, which is even considered as a distinct sub-genre of academic writing [14]. The abstract is the first part of most academic texts and is supposed to invite readers to continue reading [15]. Meanwhile, the quality of the abstracts is important as they appear in the abstracting and indexing of publishers. For many years, researchers' attention has been drawn to abstracts in general, and various works have been investigated different linguistic, stylistic and structural characteristics of this section [13,16]. For instance, Yoneoka and Ota found that high-quality abstracts in articles contain longer words, shorter sentences, a larger proportion of noun phrases and smaller proportion of verb phrases than low-quality abstracts [17]. Hyland and Tse investigated the frequencies as well as forms and functions of 'evaluative that' in the research articles, MA dissertations and PhD thesis abstracts written by English L2 writers [18].

Moreover, there are a few studies investigated the features of voice in abstract section. For instance, Wang and Wang compared the passive voice application in MA and PhD theses and international journals based on the self-compiled corpus, and found that L2 learners use more *get+V-ed* than the experts in international journals [11]. Scholars also indicated that abstracts in international journals tend to apply the active voice rather than the passives in recent years [19]. The multi-dimensional comparative analysis of abstracts of doctoral dissertations between Chinese and foreign countries also suggested that native speakers tend to use first-person pronouns to express their views, while learners tend to use a large number of nouns to enhance the informative content of the abstract, and tend to use 'we' to refer to the author [20]. Nevertheless, the description of the passive voices of the abstract section in different disciplines and timespans is an under-investigated area [21].

In line with these problems, this research intends to verify whether the previous assumptions about the passives in academic writing applies to RA abstracts in different disciplines in the last fifteen years based on RA abstract published in authoritative international journals of four disciplines in natural sciences and social sciences over the 2008-2022 period. Research questions are to be addressed: 1) How does the occurrence of passives vary across the three periods in the natural and social sciences RA abstracts, respectively? 2) How does the occurrence of passives vary across the disciplines in RA abstracts in the three periods, respectively?

3. Method

3.1. Corpus Description

To achieve research objectives, this thesis collected 180 RA abstracts. These include 90 abstract texts the natural sciences

and 90 abstract texts for the social sciences. These abstracts have been drawn from empirical research articles published from 2008 to 2022 (five years per period) across four disciplines. The period of 2008-2022 is split into three timespans: 2008-2012, 2013-2017, and 2018-2022. The journals included in the corpus are international peer-reviewed journals indexed in either the SCI (Science Citation Index) or SSCI (Social Sciences Citation Index) databases. These two databases provide a wide selection of journals from various disciplines which are highly respected in their fields. Additionally, the impact factors of each journal in a particular field were also considered, starting from the journal with the highest impact factor in that field. In the meantime, research articles in the journal were checked. Ultimately, 12 journals from four disciplines were incorporated into the corpus. The final corpus comprises four disciplines, including cell biology and engineering from the natural sciences, and linguistics and education from the social sciences. The abstracts were extracted and saved as .txt files. Each text was stored independently, titled with information code such as knowledge field, journal, and publishing year. For example, A_Cellular101 indicates that the text belonged to the natural science field and was published between 2008 and 2012 from Cellular and Molecular Life Sciences. Finally, there were texts to be included in the corpus. The abstract corpus of academic writing (ACAW) contains 180 texts and 34,356-word tokens. Table 1 provides detailed information about the corpus.

Table 1. Corpus Description of ACAW

Academic field	Disciplines	2008-2012	2013-2017	2018-2022
Natural sciences	Cell biology	15	15	15
	Engineering	15	15	15
Social sciences	Linguistics	15	15	15
	Education	15	15	15
Total (texts)		60	60	60
Total (words)		11,564	11,268	11,524

3.2. Definition of the Passive Voice

The basic form of the passive voice, i.e., *be+V-ed* is adopted to identify passives. *Be* includes *am, is, are, was, were, be,* and *been*. *V* is restricted to only finite verbs in the form of both long passives (containing agents) and short passives (omiting agents). Other unusual forms such as non-finite passives, e.g., those headed by present participle verbs, are not considered as they usually function as modifiers where the agent of the action does not matter. Passives which occur in subordinate clauses are also counted. Copula verbs such as *be, become, seem, appear, remain,* and *stay*, and another set of verbs including *concern, bear, suppose,* and *situate* were excluded, as they never or rarely passivize [9].

3.3. Procedure

The purpose of this study is to investigate the diachronic changes of passive voice in abstract sections across academic disciplines.

Having extracted the 180 abstract sections in Web of Science Core Collection, SpaCy in Python 3.7 was used to tag the part of speech of words in all abstracts. First, the sentences in each abstract text were splited. Then, the words and voice of each sentence in all abstracts were tagged. Aftetwards, based on the prior coding, the passives were manually checked through close examination of the context.

After the coding, the number of sentences and the frequency of passives in each abstract were computed through Regular Expression in Python 3.7 and recorded in an Excel document for subsequent analysis.

The software SPSS Statistics 26.0 was used for statistical comparison of the data across disciplines and timespans. One-way ANOVA analyses was conducted to determine if observed differences were statistically significant.

4. Results

To allow comparison across timespans and disciplines, the distribution of passives in abstracts is investigated by

counting the proportion of abstracts where passives occur. The extent of passives in each abstract is operationalized as the mean occurrence of passives per sentence (MPPS).

4.1. The Diachronic Changes of the Passive Voice

A total of 530 passives were identified in 1642 sentences in 180 abstracts. Descriptive statistics of MPSS in all abstracts are presented in Table 2. As shown by the mean value and standard deviation, although the number of passive voice sentences declined, there is no significant difference in three periods.

Table 2. Descriptive statistics of MPPS in all abstracts

Year	N	Mean	Std.	95% Confidence interval for mean		Min.	Max.
				Low	Upper		
2008-2012	60	0.3250	0.1809	0.2782	0.3717	0.0000	0.7778
2013-2017	60	0.2839	0.1832	0.2366	0.3313	0.0000	0.8000
2018-2022	60	0.3172	0.1969	0.2664	0.3681	0.0000	0.0800

Nevertheless, the statistics show differences between natural sciences and social sciences. As Table 3 showed, in the field of natural science, the minimum proportion of passive voice usage between 2008-2012, 2013-2017, and 2018-2022 was 0.00, while the maximum proportion was 0.78, 0.8, and 0.8 respectively. This indicates a slight increase in the maximum proportion over time. Additionally, the mean proportion of passive voice usage was 0.32 for 2008-2012, 0.33 for 2013-2017, and 0.39 for 2018-2022. In summary, there appears to be a trend of increasing passive voice usage in natural science over time.

Table 3. Descriptive statistics of MPPS in natural sciences

	N	Mean	Std. deviation	Min.	Max.
2008-2012	30	0.32	0.16	0.00	0.78
2013-2017	30	0.33	0.20	0.00	0.80
2018-2022	30	0.39	0.20	0.00	0.80

Meanwhile, Table 4 shows the MPPS in social sciences. In the field of social science, the minimum proportion of passive voice usage between 2008-2012, 2013-2017, and 2018-2022 was 0.00, while the maximum proportion was 0.78, 0.59, and 0.50 respectively. This indicates a trend of decreasing maximum proportion of passive voice usage over time. Additionally, the mean proportion of passive voice usage was 0.33 for 2008-2012, 0.23 for 2013-2017, and 0.24 for 2018-2022. In summary, there has been a significant decrease in passive voice usage in social sciences from 2008-2012 to 2013-2017, and a slight increase from 2013-2017 to 2018-2022, but overall, there has been a general decline in the proportion of passive voice usage.

Table 4. Descriptive statistics of MPPS in social sciences

	N	Mean	Std. deviation	Min.	Max.
2008-2012	30	0.33	0.20	0.00	0.78
2013-2017	30	0.23	0.16	0.00	0.59
2018-2022	30	0.24	0.17	0.00	0.50

To sum up, while the use of passive voice in the overall summary has remained relatively consistent over time, there have been slight changes in natural sciences and social sciences across the three periods. Although previous studies have suggested a decline in the use of passive voice in natural

sciences, the findings of this study suggest a slight upward trend in recent years. In contrast, passive voice usage in social sciences has continued to decline, which is consistent with previous research.

4.2. The Differences in the Use of Passive Voice Across Disciplines

Academic writing in natural science and social science is known to be quite distinct, and this is particularly evident in the writing of abstracts. Therefore, in addition to exploring the changes in passive voice usage over time, this study also seeks to analyze differences in the use of passive voice across different disciplines within the same time period. Based on the results of the difference test for passive voice usage across different disciplines in 2008-2012 in Table 5, the average proportion of passive voice usage in Cell Biology, Engineering, Linguistics, and Education disciplines was 0.30, 0.34, 0.33, and 0.33, respectively. The F statistic for the difference in passive voice usage across different disciplines in 2008-2012 was 0.168, which did not reach the significance level of 5%. Therefore, it is concluded that there was no significant difference in passive voice usage across different disciplines in 2008-2012.

Table 5. The passive voice in different disciplines in 2008-2012

	N	Mean	Std. deviation	F	p
Cell biology	15	0.30	0.129	0.168	0.918
Engineering	15	0.34	0.184		
Linguistics	15	0.33	0.209		
Education	15	0.33	0.206		

Table 6 shows the changes of the passive voice in four disciplines in 2013-2017. The results showed that the average proportion of passive voice usage in Cell Biology, Engineering, Linguistics, and Education was found to be 0.30, 0.36, 0.21, and 0.26, respectively, in 2013-2017. The F statistic for the difference in passive voice usage across different disciplines in this period was 1.971, but it did not reach the significance level of 5%. Therefore, it is concluded that there was no significant difference in passive voice usage across different disciplines in 2013-2017.

Table 6. The passive voice in different disciplines in 2013-2017

	N	Mean	Std. deviation	F	p
Cell biology	15	0.30	0.162	1.971	0.129
Engineering	15	0.36	0.228		
Linguistics	15	0.21	0.152		
Education	15	0.26	0.163		

The statistics in Table 7 indicate that the average proportion of passive voice usage in Cell Biology, Engineering, Linguistics, and Education was 0.37, 0.42, 0.27, and 0.22,

Table 7. The passive voice in different disciplines in 2018-2022

	N	Mean	Std. deviation	F	p	LSD
Cell biology	15	0.37	0.191	3.615	0.019	1>4,2>3,2>4
Engineering	15	0.42	0.211			
Linguistics	15	0.27	0.158			
Education	15	0.22	0.176			

Overall, this study reveals that there were few differences in the use of passive voice between natural science and social science subjects in the first ten years. The real difference began in the last five years, with a significant increase in the use of passive voice in natural science subjects and a decrease in social science subjects. These findings suggest that the use of passive voice in academic writing may be subject to changes over time and across disciplines. Furthermore, these results differ from previous studies and call for further investigation into the reasons behind such changes.

5. Discussion

From a diachronic perspective, the occurrence of passive voice has slightly declined in academic writing, reflecting a trend towards plain language in recent years. Also, this phenomenon is a response to the preference of active voice. This trend is a response to the preference for active voice advocated in academic community. However, current data suggests that the use of passive voice has not changed significantly, and the standardized format of abstracts may be one reason.

Different from the previous results focusing on the whole research texts, the occurrence of passive voice in natural sciences abstract witnessed an increase, whereas the passive voice in social sciences declined in the former two period and slightly increased in the recent five years. Meanwhile, abstracts in natural sciences have more passive voice than it in social sciences, which is opposite to the previous studies. The reasons behind could be the genre differences between natural sciences and social sciences, as well as the author intention for various communication purposes. For instance, research in natural sciences tends to indicate the standard procedure, emphasizing on data rather than human factors, which can be seen in (1), where *is promoted* is a passive voice emphasizing the radio tomographic imaging methodology. In contrast, abstracts in social sciences may indicate the subjectivity and agency of researchers and participants.

(1) Radio (NOUN) tomographic (NOUN) imaging (PROPN) ((PUNCT) RTI (PROPN)) (PUNCT) *is* (AUX-passive) *promoted* (VERB) as (ADP) a (DET) simple (ADJ) yet (ADV) flexible (ADJ) way (NOUN) of (ADP) facilitating (VERB) the (DET) required (VERB) position (NOUN) - (PUNCT) based (VERB) stigmery (NOUN) . (PUNCT)

Furthermore, the use of passive voice in abstract sections may be influenced by the author's intentions. For instance,

respectively. The F statistic for the difference in passive voice usage across different disciplines in this period was 3.615, reaching a significant level of 5%. Therefore, it is concluded that there was a significant difference in passive voice usage across different disciplines in 2018-2022. Post-hoc comparisons revealed that in this period, the proportion of passive voice in Cell Biology was significantly higher than that in Education, and the proportion of passive voice in Engineering was significantly higher than that in Linguistics and Education.

authors may opt for active voice when speculating on their own work and passive voice when referring to the work of others [4]. Additionally, authors may use passive voice to emphasize certain elements. Therefore, further research is needed to explore the reasons behind the changes in passive voice usage in specific verbs and clauses. It is important to note that the slight variations in passive voice usage may be due to limited timespans and corpus size, which should be addressed in future studies.

6. Conclusion

The investigation of diachronic change in the use of passives and the changes across disciplines has complemented the previous studies from the aspect of abstract genre. The results show that abstract writing appears to be different in natural sciences and social sciences, with the former applies a slightly increasing passive voice and an obvious decrease in the latter abstracts. Meanwhile, the differences between natural sciences and social sciences have emerged in the recent five years, which is quite different from previous studies showing that natural sciences witnessed a more obvious decline in the use of passives than social sciences.

The above results depict a trajectory of passive voice in RA abstracts of both natural sciences and social sciences in three periods. They reflect the emergence of informational, reader-friendly style in abstract writing, which could give implications to academic writing and EAP instruction as well as learning. However, this research has its limitations. The details of passive clauses and the reasons for both passive voice and active voice have not been studied in this research. The use of passive voice is always related to certain intentions, which could be analyzed and summarized with pragmatic theoretical frameworks in future studies. Moreover, no significant difference in results may also be due to limited timespans and corpus size. Therefore, research with large timespans is needed to verify the former results.

Acknowledgments

We thank the professors who teach corpus linguistics courses, giving us advice on this paper. And we'd like to give thanks to the friends, who provided companionship and motivation during this journey, even during late-night hours.

References

- [1] C. Coffin, M. Curry, S. Goodman, A. Hewings, T. Lillis, and J. Swann, *Teaching academic writing: A toolkit for higher education*, London: Routledge, 2003.
- [2] C.L. Barber, "Some measurable characteristics of modern scientific prose," in *Contributions to English syntax and philology*, ed. F. Behre, Stockholm: Almqvist & Wiksell, 1962, pp. 21–43.
- [3] D. Biber, *Variation across speech and writing*. Cambridge University Press, 1988.
- [4] Y. Cao and R. Xiao. Cao, "A multi-dimensional contrastive study of English abstracts by native and nonnative writers," *Corpora*, vol. 8, no. 2, pp. 209-234, 2013.
- [5] D. Banks, "The extent to which the passive voice is used in the scientific journal article, 1985-2015," *Functional Linguistics*, vol. 4, no. 1, pp. 12, 2017.
- [6] G. Leech, M. Hundt, C. Mair, and N. Smith, *Change in contemporary English: A grammatical study*, Cambridge University Press, 2009.
- [7] K. Hyland, and F. Jiang, "Is academic writing becoming more informal?" *English for Specific Purposes*, vol. 45, pp. 40–51, 2017.
- [8] A. Leong, "The passive voice in scientific writing through the ages: A diachronic study," *Text & Talk*, vol. 40, no. 4, pp. 467–489. 2020.
- [9] M. Hundt, G. Schneider, and E. Seoane, "The use of the be-passive in academic Englishes: Local versus global usage in an international language", *Corpora*, vol. 11, no. 1, pp. 29–61, 2016.
- [10] Z. Li, "Is academic writing less passivized? Corpus-based evidence from research article abstracts in applied linguistics over the past three decades (1990–2019)", *Scientometrics*, vol. 127, no. 10, pp. 5773–5792, 2022.
- [11] A. G. Gross, J. E. Harmon, and M. Reidy, *Communicating science: The scientific article from the seventeenth century to the present*. Oxford University Press, 2002.
- [12] L. Wang, and N. Wang, "Colloquialisation in EFL learners' theses in China," *Modern Foreign Languages*, no. 2, pp. 275-286, 2017.
- [13] Y. M. Ju, "A comparative study of rhetorical persuasion of English research article abstracts by Chinese and English writers," *Foreign Languages Bimonthly*, no. 01, pp. 85-92, 2020.
- [14] M. Nasserri, and P. Thompson, "Lexical density and diversity in dissertation abstracts: Revisiting English L1 vs. L2 text differences," *Assessing Writing*, vol. 41, pp. 100511, 2021.
- [15] J. Bitchener, *Writing an applied linguistics thesis or dissertation: A guide to presenting empirical research*, Basingstoke: Palgrave MacMillan, 2010.
- [16] X. Hu and Y. Huang, "A contrastive study on the use of common high-frequency lexical bundles in different moves of English abstracts in Chinese and English engineering journals," *Foreign Language Learning Theory and Practice*, no. 04, pp. 31-36, 2017.
- [17] D. Yoneoka, and E. Ota, "Evaluating association between linguistic characteristics of abstracts and risk of bias: Case of Japanese randomized controlled trials," *PLOS ONE*, vol. 12, no. 3, Article e0173526, 2017.
- [18] K. Hyland, and P. Tse, "Hooking the reader: A corpus study of evaluative that in abstracts," *English for Specific Purposes*, vol. 24, pp. 123–139, 2005.
- [19] Y. H. Liu and Y. Zhang, "A contrastive study of the abstracts in international journal articles by Chinese and English-speaking scholars", *Foreign Language World*, no. 05, pp. 20-27, 2016.
- [20] Q. Xie and Y. Y. Ma, "A multi-dimensional contrastive analysis of English abstracts in dissertations by Chinese and Native Ph.D. students," *Foreign Language Research*, no. 04, pp. 39-44, 2021.
- [21] F. Jiang, "EAP research over the past 40 years: Topics and development," *Foreign Language Teaching and Research*, no. 03, pp. 413-424, 2022.