Comparing the Writing Styles of Multiple Disciplines: A Large-Scale Quantitative Analysis

Shuyi Dong, School of Information Management, Nanjing University, China (sydong@smail.nju.edu.cn)
Jin Mao, Center for Studies of Information Resources, Wuhan University, China (maojin@whu.edu.cn)
Lei Pei, School of Information Management, Nanjing University, China (plei@nju.edu.cn)

Introduction

"Academics do not act in a social vacuum and knowledge is not constructed outside particular communities of practice." (Hyland, 2006, p. 39). It serves as a conduit for sharing knowledge, fostering scholarly communities, and advancing the frontiers of science. In scholarly communication, the choice of linguistic resources and writing conventions varies across disciplines, resulting in distinct writing styles. These styles facilitate communication and encapsulate the cultural traditions and codes of conduct within each academic field.

Our study undertakes a large-scale quantitative analysis to delve into the evolution of writing styles across disciplines and the reasons behind their formation. We uncover the dynamic nature of scholarly language and the diversity of writing styles across academic domains:

1. How have writing styles within disciplines evolved over time?
2. To what extent do writing styles differ among disciplines?

Method

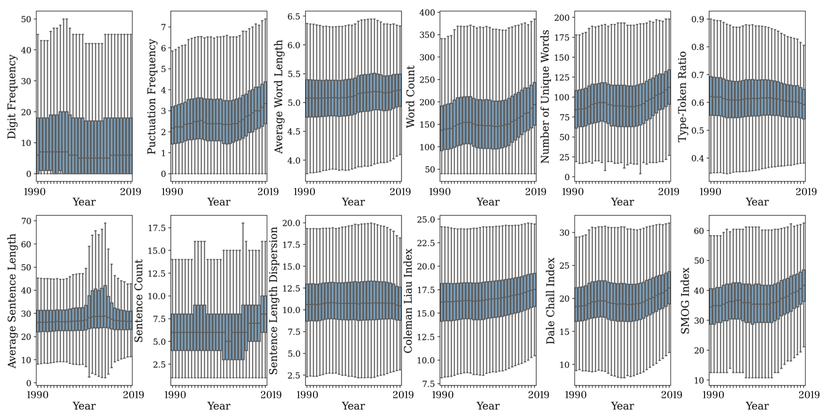
Data: 14 million abstracts from Microsoft Academic Graph (1990-2019) spanning 8 disciplines (Art, Philosophy, Sociology, Psychology, Biology, Chemistry, Computer Science, Mathematics). The first four disciplines falling under soft sciences and the last four under hard sciences.

Analysis: Quantitative analysis on linguistic features (symbolic, lexical, syntactic, readability). Employed t-tests and calculated Cohen's d effect size for significance.

Results

Trends in Scientific Writing Styles: Increasing Complexity and Informativeness

- Figure 1: Trends of linguistic features over 30 years.



Notable Shifts in Writing Styles Evident in Biology, Chemistry, Computer Science, and Psychology

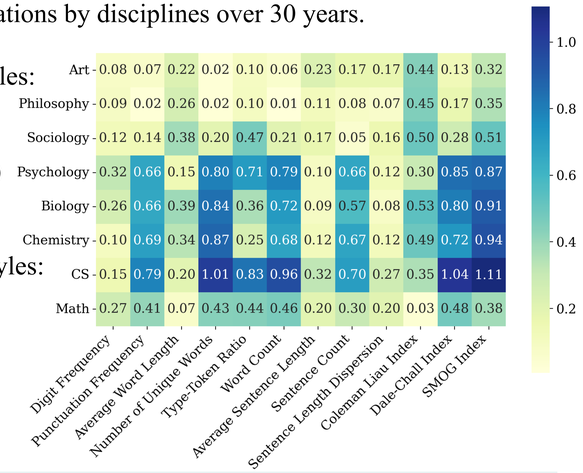
- Figure 2: Effect sizes of variations by disciplines over 30 years.

High Effect Sizes in Writing Styles:

- Biology (hard)
- Chemistry (hard)
- Computer Science (hard)
- Psychology (soft)

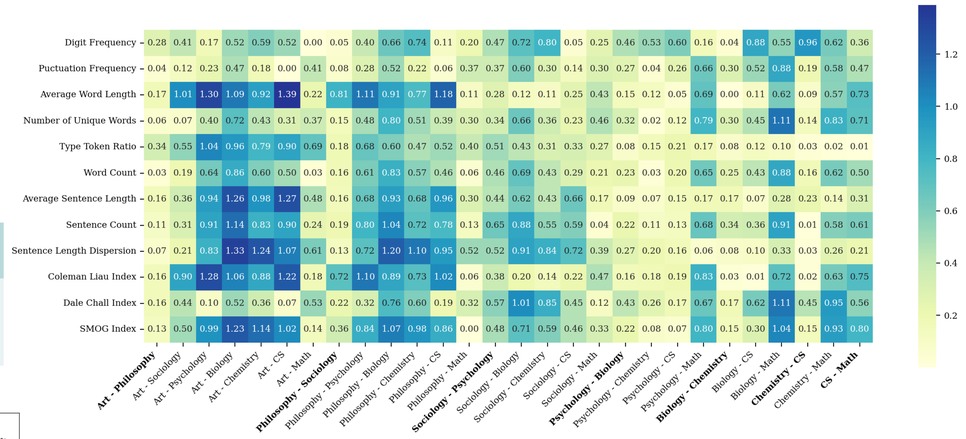
Small Effect Sizes in Writing Styles:

- Art (soft)
- Philosophy (soft)
- Sociology (soft)
- Mathematics (hard)



Distinctive Differences in Writing Styles Across Disciplines

- Figure 3: Effect sizes of pairwise comparisons of writing styles between disciplines.



Above:

- Most disciplines showed noticeable differences in writing styles.
- Fewer variations among neighboring disciplines
- Similar distinction patterns existed in Art- and Philosophy-related pairs.
- Consistent patterns of distinction emerge, epescially in Math-related pairs.

Left:

a. Symbols: More punctuation, more symbols in scientific papers over 30 years.
b. Words: Longer words, more words, more information. Fewer unique words.
c. Sentences: Stable length, more sentences recently. Consistent rhythm.
d. Readability: Harder to read, increasing complexity.

Conclusion & Future Plan

The results indicate a trend of increasing complexity and informativeness in scientific writing styles, with hard sciences experiencing greater changes. Disparities in writing styles were observed across disciplines, exhibiting varying degrees of differences in features.

Our study highlights the dynamic nature of scientific language and confirms that different disciplines employ distinct features to establish their unique linguistic identity. These insights contribute to a deeper understanding of disciplinary writing and discipline-oriented communication.

In our future research, we aim to explore deeply into disciplinary writing styles with:

- using more nuanced features
- employing explainable machine learning methods to classify and identify writing styles
- analyzing the contributions of features to uncover how linguistic traits encode scientific writing across disciplines

References

Ante, L. (2022). The relationship between readability and scientific impact: Evidence from emerging technology discourses. Journal of Informetrics, 16(1), 101252.
Abelson, P. H. (1980). Scientific communication. Science, 209(4452), 60–62.
Argamon, S., Dodick, J., & Chase, P. (2008). Language use reflects scientific methodology: A corpus-based study of peer-reviewed journal articles. Scientometrics, 75(2), 203–238.
Chen, B., Deng, D., Zhong, Z., & Zhang, C. (2020). Exploring linguistic characteristics of highly browsed and downloaded academic articles. Scientometrics, 122(3), 1769–1790.
Cohen, J. (1992). Statistical power analysis. Current Directions in Psychological Science, 1(3), 98–101.
Fanelli, D., & Glänzel, W. (2013). Bibliometric evidence for a hierarchy of the sciences. PLoS ONE, 8(6), e66938.
Hyland, K., & Bondi, M. (Eds.). (2006). Academic discourse across disciplines. Peter Lang.
Harrison, T., & Stephen, T. (1995). The electronic journal as the heart of an online scholarly community. Library Trends, 43(4), 592-608.
Jin, T., Duan, H., Lu, X., Ni, J., & Guo, K. (2021). Do research articles with more readable abstracts receive higher online attention? Evidence from Science. Scientometrics, 126(10), 8471–8490.
Juola, P. (2008). Authorship attribution. Foundations and Trends® in Information Retrieval, 1(3), 233-334.
Lu, C., Bu, Y., Dong, X., Wang, J., Ding, Y., Larivière, V., Sugimoto, C. R., Paul, L., & Zhang, C. (2019). Analyzing linguistic complexity and scientific impact. Journal of Informetrics, 13(3), 817–829.
Song, N., Chen, K., & Zhao, Y. (2023). Understanding writing styles of scientific papers in the IS-LS domain: Evidence from abstracts over the past three decades. Journal of Informetrics, 17(1), 101377.
Swales, J. (1990). Genre analysis: English in academic and research settings. Cambridge University Press.
Teich, E., Degaetano-Ortlieb, S., Fankhauser, P., Kermes, H., & Lapshinova-Koltunski, E. (2016). The linguistic construal of disciplinarity: A data-mining approach using register features. Journal of the Association for Information Science and Technology, 67(7), 1668–1678.
Wu, H., Zhang, Z., & Wu, Q. (2021). Exploring syntactic and semantic features for authorship attribution. Applied Soft Computing, 111, 107815.

Acknowledgement

This study was supported by Science Fund for Creative Research Group of the National Natural Science Foundation of China (No. 71921002).