

Gender Differences in Academic Writing: A Cross-Linguistic Analysis of French, Spanish, and English - Individual Contribution

Proteeti Kushari
Trinity College Dublin
kusharip@tcd.ie

14/04/2025

1 Hypotheses

Involvement-Information Hypothesis: Women authors would demonstrate higher Involvement-Informational Ratio (IIR) than male authors throughout the three analyzed languages. The analysis incorporates Tannen's discourse style framework which demonstrates women prefer rapport-oriented dialogue caused by their quantity of involvement and information characteristics.

2 Metrics - Emotional Involvement

We used the method defined by Kedric et al Kedrick et al. (2022) to judge whether a given author uses emotion-signifying words such as personalized words such as pronouns or questions.

1. **Involvement Rate:** This metric seeks to measure the degree to which an author is personally engaged with the written content of their work. This is measured by the frequency of the use of personal pronouns, questions, and coordinating conjunctions.

$$IR = \frac{\text{Personal Pronouns} + \text{Questions} + \text{Coordination}}{\text{Total Words}} \quad (1)$$

2. **Informational Rate:** This computes the density of the factual content in a text which is indicated by the frequency of various confidence-indicating terms such as noun phrases, technical terms, and references.

$$InfR = \frac{\text{Noun Phrases} + \text{Technical Terms} + \text{References}}{\text{Total Words}} \quad (2)$$

3. **Involvement-Informational Ratio:** This metric serves to compute the balance between the above two metrics: an analysis of the trade-off between personal engagement and informational language in a sample of text. A high positive value (≥ 1) suggests a greater emotional tone, while a value between 0 and 1 suggest a more factual language.

$$IIR = \frac{IR}{InfR} \quad (3)$$

3 Results

We observed a marked difference in the writing styles of men and women as we evaluated our collated dataset over a variety of metrics that allowed us to validate the hypotheses posed earlier in this paper.

3.1 Emotional Involvement

As expected the IIR values are low, indicating a fact-driven, unemotional writing style, consistent with the academic domain that we have chosen.

From Figs. 1 and 2, we see that for women authors a marginally heightened IIR value is indicated for both subjects. For both English and Spanish, we observe a higher emotional involvement from women. This confirms our hypotheses, as we are able to identify a point of linguistic difference in gender-based academic literature, while validating that the trend holds up for Spanish.

The expected pattern is not observed for French papers, where the men seem marginally more involved. This points to the academic environment of France, which differs from the Anglosphere. The French literary style places emphasis on *esprit critique* (critical spirit), which encourages men to make empathetic statements. This is supported by the detailed analysis where we see that the male authors within the French corpus display a higher involvement rate (IR). On the other hand, working in a male-dominated field, women might be pressured to adopt a less emotional style so as to avoid scrutiny. This, combined with the previous observation is also a contributing factor in this deviance.

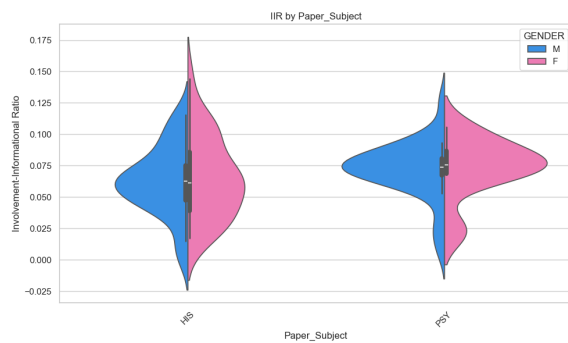


Figure 1: Frequency distribution of Involvement-Informational Ratio per Subject

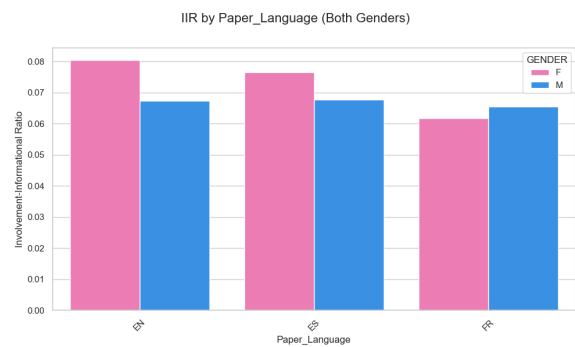


Figure 2: Involvement-Informational Ratio by Language

4 Conclusion

Gender-based language disparities in academic writing in English, French, and Spanish across the disciplines of History and Psychology are validated within this endeavour.

4.1 Discussion

Our initial hypotheses are primarily validated, as we are able to confirm, both qualitatively and quantitatively, a higher emotional involvement in women's writing style within the expanse of our data corpus. We notice some surprising results, especially for French, that can be theoretically explained by certain social phenomena, which will be further validated upon a more detailed analysis of the same.

4.2 Limitations

The dataset we have amassed is significant, and yet a larger corpus would undoubtedly render more reliable results. To analyse the French and Spanish papers we relied on spaCy's translation models, and a deeper knowledge of the linguistic nuances of these languages could prove more useful.

4.3 Future Work

Our research in the current scope is limited only to the study of papers within the domain of History and Psychology. The analysis of a wider range of academic disciplines will allow us to discover further

nuances within the purview of our initial research questions. Further, the study of a greater gender diversity, beyond the binary approach that we have undertaken, will undoubtedly reveal deeper insights and further enrich our understanding of the role of gender in academia.

5 Declarations

5.1 Conflict of Interest

I declare that I do not have any conflicts of interest with regards to the particular topic we have chosen for our research endeavours.

5.2 AI Statement

I declare that I have not used any generative AI tool to reproduce any part of my research endeavour or any part of our written project.

References

Kedrick, K., E. Levitskaya, and R. J. Funk (2022). Investigating writing style as a contributor to gender gaps in science and technology. arXiv preprint arXiv:2204.13805.

A Appendix

Table 1: Average of Involvement Metrics by Gender, Language, and Subject

Gender	Lang - Sub	IR	InfR	IIR
M	EN - HIS	0.0305	0.5117	0.0606
M	EN - PSY	0.0068	0.0534	0.1265
M	FR - HIS	0.0257	0.4657	0.0548
M	FR - PSY	0.0351	0.4914	0.0715
M	ES - HIS	0.0288	0.5120	0.0650
M	ES - PSY	0.0340	0.4845	0.0668
F	EN - HIS	0.0338	0.4328	0.0803
F	EN - PSY	0.0000	0.0000	0.1676
F	FR - HIS	0.0198	0.3984	0.0469
F	FR - PSY	0.0404	0.4871	0.0832
F	ES - HIS	0.0459	0.4108	0.1126
F	ES - PSY	0.0327	0.4945	0.0634