

‘A Multidimensional and Digital Humanistic Analysis of Style in Amy Tan’s Novels’: A Research Critique

Zelei Lian 2306368

word count (excluding references): 2138

Introduction

In recent decades there has been a proliferation of mixed-method approaches towards the digital stylistic research, juxtaposing empirical and hermeneutic traditions. With easier-to-access analysis instruments and growing digital literacy in programming languages such as *Python* and *R*, researchers are increasingly empowered to combine quantitative hypothesis testing and qualitative text analysis, opening up an interdisciplinary and mutually-enriching research panorama. The research article under review, published in September 2023 in a respectable, peer-reviewed academic journal, is one of the most recent examples to apply this multi-methodological approach and merits close inspection. The two authors of this research article have demonstrated, with varying degrees of success, how this approach can facilitate deeper understanding of a literary phenomenon, albeit not without caveats. This paper briefly 1) examines its methodology, and its hypotheses followed from it; 2) critiques its research questions, process, and hypotheses testing; 3) discusses its results against the mixed-method paradigm; 4) and offers alternative approaches to its research aim.

The Methodology

Amy Tan is a text-book-famous American author of Chinese heritage, who is best known for her works exploring the inner worlds of second-generation Chinese immigrants in the United States. Academic research on her works abounds in the hermeneutic tradition, especially her portrayal of cultural and ethnic tensions. But studies on her unique linguistic style is scarce, let alone a comprehensively quantitative investigation. It is to address this research gap that the authors of the critiqued article, Professor Junwu Tian and Shunyue Liu from Beihang University teamed up and ran their tests. They start by asking how can we measure Amy Tan’s stylistic variations in her writing, though they already have their answer in hand, and that is Douglas Biber’s multidimensional analysis (MDA) model.

In his 1988 seminal work *Variation across Speech and Writing*, Douglas Biber first proposed his now famous 6-dimension factor analysis to discern patterns of linguistic variation across a range of registers:

1. Involved vs. Informational Production;
2. Narrative vs. Non-Narrative Concerns;

3. Explicit vs. Situation-Dependent Reference;
4. Overt Expression of Persuasion;
5. Abstract vs. Non-Abstract Information;
6. On-Line Informational Elaboration (Biber, 1995, p.122).

This 6-dimension approach is a synthesis of his factor analysis, which takes into account a total of 67 distinct linguistic features that co-occur frequently in texts, whose normed rate of occurrence he computed based on the London-Lund Corpus of Spoken English (LLC) and the Lancaster-Oslo/Bergen Corpus (LOB). Initially designed to only measure differences between English speech and writing, the MDA was gradually applied over the past few decades to other registers including English literary texts, due to its ingenious methodological innovation of factor analysis.

From a methodological point of view, the article under review is strictly an “applied” piece of Multidimensional analysis model of mix-method investigation of Amy Tan’s linguistic style, in the sense that it strictly adheres to all six-dimension of the 1988 version, using the exactly same linguistic features, or tags, making up its research backbone. This quantitative approach is complemented by its qualitative interpreting of the data, combining the advantages of data statistics and textual analysis.

This approach certainly flags concerns, which I believe the authors do not fully address. Most importantly, the original six-dimension was, as the authors acknowledged, designed with not the literary texts in mind. Though its applicability in literary texts has been generally confirmed by later case studies, readers cannot be blamed for expecting a justification for using the exact criteria three decades old. For example, should not the criteria, and its associated linguistic taggers, be updated and adapted so that they could better suit the specific needs of literary quantitative investigation? The strict methodological adherence, while in itself watertight in its research design, misses opportunities to engage with more difficult questions and other nuances that the investigated texts have to offer. As a useful contrast, Hermann (2017) while with a similar research outline, strengthens its arguments by deploying a new set of specifically designed linguistic taggers for the corpus of Kafka’s works.

That said, this piece is not without its innovations. First, after the MDA is conducted, it continues to seek to establish correlation in some of the six-dimensions surveyed. Interesting hypotheses are brought forward and duly tested. Second, unlike previous MDA studies in literacy texts in the 2010s, this article is blessed with the timely release of Multidimensional Analysis Tagger (MAT) in 2019, which, replicating the same linguistic taggers, proves “a powerful tool to generate a grammatically annotated version of the corpus and provides visualizing the Dimensions features of the input text” (Nini, 2019, p. 68). The result is an accurate depiction of the stylistic various across Amy Tan’s writings.

The Process

The research process of this article starts by putting forward the two research questions:

1. What stylistic variations do Amy Tan's novels possess along each of the six dimensions in Biber's model?
2. Is there a statistically significant correlation between any two dimensions when it comes to Amy Tan's novels? (Tian and Liu, 2023, p. 1282).

To answer question one, the authors set up the corpora, run the text through the Multidimensional Analysis Tagger program, and interpret the data with extracts from the texts. The first step is to set up the corpora. "Style" in its analysis is "the way in which language is used in a given context, and for a given purpose" (Leech and Short, 2007, p. 13). The authors quite arbitrarily decide the corpus body to be all six novels of Amy Tan but excluding para-texts. This arbitrary choice is open to two questions that the authors do not address adequately. For one thing, if it is Amy Tan's literary style that they are investigating, one would wonder why the large body of her short stories and children's books are excluded. The sampling of her novels only seems to require reasonable justification. Second, the authors do not entirely abide by their own criteria of excluding para-texts from the novels. There is at least one case of para-text slipping into the corpus, however, with the authors explaining that it is "included in our corpus as an important para-text, even though it is not in the body part of this novel" (Tian and Liu, 2023, p. 1294). One cannot help but ponder the uncomfortable implication of objectivity: By including one piece of para-text because it happens to lend support to their view but excluding others, it seems to have violated the validity of the data statistics, if not amounting to outright data cherry-picking. Readers can be comforted by the fact that these are rare exceptions and should not in theory jeopardize the article major findings.

The research tool is the aforementioned MAT, whose procedure is relatively straightforward. With the corpus set, the MAT returns a tagged version of the original text, using Biber's original 67 linguistic features. Each of Amy Tan's novel is run through the MAT successively for each of the six dimensions, yielding the standardized scores for each linguistic feature. At this stage, the authors have an overall picture of the standardized scores with mean scores for each of her novels, with the mean score of general fiction as a reference point. This then allows the authors the chance to dive into the qualitative reading of Amy Tan's novels, by interpreting individual scores and their differences using extensive extracts from these novels.

The answer to questions two is a logical extension from the answering of research questions one, because only at that stage do the authors have the data findings to ascertain whether there is indeed correlation between any of the dimensions. They raise two hypotheses:

1. There is a statistically significant negative correlation between dimensions 1 and 3.

2. There is a statistically significant positive correlation between dimensions 3 and 5. (Tian and Liu, 2023, p. 1292).

In order to test these two purely statistical hypotheses, the authors run the original data set using the software *R*, calculating the Pearson correlation coefficient (*r*) by *cor()* function, a common way to discern correlation between two variables. After that a *Rcorr()* function is run to calculate the p-value, determining whether the correlations are indeed statistically significant. With the p-values standing at 0.082 and 0.012 respectively, the authors conclude that the first hypothesis is unreliable but the second valid.

It merits attention that these two hypotheses are pure numerical. For them to provide insight, if any, into the linguistic style of Amy Tan's novels, it would normally require rigorous contextualization within her literary world, and thorough discussion of the statistical results, which are notably wanting. After providing the two p-values the authors move straight to announce their verdict. Careful readers might expect more detailed discussions of the results, because it seems to be abusing the use p-values, by announcing a hypothesis to be true or false simply by being on one side or the other side of p-value 0.05. Valuable nuances may be lost without adequate contextualizing and further interpretation. It is my humble opinion that the hypotheses testing process can be more robust and valid by taking into consideration of their relevance in the individual novels, instead of coming to an abrupt end.

The Results

The results section makes up the bulk of this research article, and this is where it contributes the most to our understanding of Amy Tan's unique linguistic style as is shown in her novels. By combining digital methodologies with hermeneutic insights, this section has confirmed some previous literary opinions, refuting some others with evidence, while suggesting new directions for further exploration.

Table 2. Dimension statistics for each novel

Novel	Dimension 1	Dimension 2	Dimension 3	Dimension 4	Dimension 5	Dimension 6
<i>The Joy Luck Club</i> (1989)	0.7	3.81	-0.51	-0.68	-1.45	-0.78
<i>The Kitchen God's Wife</i> (1991)	4.33	5.03	-1.61	0.12	-1.75	-0.26
<i>The Hundred Secret Senses</i> (1995)	-3.04	2.36	-0.56	-2.44	-1.61	-1
<i>The Bonesetter's Daughter</i> (2001)	2.64	5.4	-0.22	0.32	-0.98	-0.83
<i>Saving Fish From Drowning</i> (2005)	-7.74	4.74	0.94	-0.96	-0.16	-0.86
<i>The Valley of Amazement</i> (2013)	-0.23	4.61	0.43	1.35	-0.97	-0.95
General Fiction	-0.8	5.9	-3.1	0.9	-2.5	-1.6

Table 3. Dimension statistics for Amy Tan's novels

Dimension	Mean	Minimum value	Maximum value	Range	Standard deviation
Dimension 1	-0.56	-7.74	4.33	12.07	4.33
Dimension 2	4.33	2.36	5.4	3.04	1.1
Dimension 3	-0.09	-1.61	0.94	2.55	0.92
Dimension 4	-0.38	-2.44	1.35	3.79	1.3
Dimension 5	-1.15	-1.75	-0.16	1.59	0.58
Dimension 6	-0.78	-1	-0.26	0.74	0.27

(Tian and Liu, 2023, p. 1284).

The results, reproduced in two tables from the article, present individual and overall dimension statistics for Amy Tan's novels. For example, in dimension 1: involved vs. informational, a high score is associated with abundance of linguistic features such as use of private verbs, contractions, and THAT deletions, reflecting a more interactional tone than a low score across the spectrum, meaning the information presented is dense. In this dimension, we will find that the novel *Saving Fish from Drowning* scores the lowest at -7.74, while *The Kitchen God's Wife*, with a 4.33 score, stands at the other extreme. The mean score for general fiction is -0.8. These sets of diverse data naturally propel us to ask the question why. We can draw the easy conclusion that all Amy Tan's novels score higher in dimension 1 compared with general fiction, meaning her style is more 'involved' than an average novel. This conclusion is nothing new for literary critics because Amy Tan is famous for her portrayal of intimate mother-daughter relationship. But what this quantitative approach provides now is statistical confirmation. It actually does more than that. By showing the range of differences among all her novels, we quickly uncover the internal incongruousness, because while many critics assume a consistent degree of 'involvement' in Amy Tan's works, the data indicates otherwise with some novels scoring significantly higher than others. Qualitative analysis is then required to provide the context and explanation. This is exactly what the authors of the article have done. They follow the data 'deviation/deviance' to the text itself. By diving into its textual world, they provide what they believe to be the rationale of a certain phenomenon. In this case, for example, *Saving Fish from Drowning* scores the lowest, because it uses a larger number of nouns, prepositional phrase, and attributive adjectives in creating an "Oriental effect" in its vivid depiction of local customs. This in turn, by the interpretation of the authors, is the deliberate intention of Amy Tan to "earn trust from readers as a reliable guide acquainted with Oriental ethnography" (Tian and Liu, 2023, p. 1286). A portion of the original text is hauled to be analyzed to corroborate their assessment.

The same procedure is repeated six times, covering along all six dimensions. Along the statistical journey, the authors have managed to discuss all major differences in any one of the six given dimensions, yielding new insights hidden from literary critics. It is strongest when major differences are found and analyzed, but weak in some cases when no significant differences are registered.

Conclusion

The research article under review is commendable in applying Multidimensional analysis model to analyze linguistic style of Amy Tan in a methodologically rigorous way, having completed an empirical cycle, linking data, hypothesis to the theory. Some of its findings will no doubt be useful for literary studies on Amy Tan, and point to new lines of academic enquires. However, as is shown in previous sections, it could also benefit from a larger sample size (all literary works of Amy Tan instead of only six novels), stricter selection criteria for corpus (either include all para-texts or none), and an adapted set of new linguistic taggers more suitable for literary texts.

References

Biber, D. (1995) *Variation across speech and writing*. Cambridge: Cambridge University Press.

Herrmann, J. B. (2017) 'In a test bed with Kafka. Introducing a mixed-method approach to digital stylistics', *Digital Humanities Quarterly*, 11(4). Available at: <https://www.proquest.com/docview/2555434536?pq-origsite=gscholar&fromopenview=true>

Leech, G. N. and Short, M. (2007) *Style in fiction. A linguistic introduction to English fictional prose*. 2nd edn. New York: Pearson Longman.

Nini, A. (2019) 'The multi-dimensional analysis tagger', in Sardinha, T. B. and Pinto, M. V. (eds) *Multidimensional analysis: research methods and current issues*. London: Bloomsbury Academic, pp. 67-94.

Tian, J. and Liu, S. (2023) 'A multidimensional and digital humanistic analysis of style in Amy Tan's novels', *Digital Scholarship in the Humanities*, 38(3), pp. 1281-1295. Available at: <https://doi.org/10.1093/lhc/fqac096>