Tracking Meta-Patterns in Poetry across Time and Space

CHARLES AVERILL[1], SRISTI DAKSHIT[1], AND JOEL STEPHEN[1]

[1] University of Texas at Dallas

Compiled May 6, 2025

We present a method for tracking the evolution of high-level thematic and structural patterns in poetry across time and geography. While our original intent included English poetry, data limitations led us to focus on French and German poetic corpora. By extracting recurring meta-patterns, we identified common literary themes and traced their changing prominence over two centuries. Our results reveal both periodic fluctuations and plausible cultural exchanges between French and German literary traditions.

http://dx.doi.org/10.1364/ao.XX.XXXXXX

Our work is available on Github: github.com/CharlesAverill/poetry_metapatterns

1. UPDATED GOAL

Our updated goal remains aligned with our original intent: to analyze and visualize the spread and evolution of meta-patterns in poetry, such as theme, over time and geography. However, due to significant limitations in the availability of temporally and geographically annotated English poetry, we pivoted to focus our analysis on French and German poetry. With these richer datasets, we successfully extracted thematic patterns and mapped their prevalence over time. We then investigated evidence of inter-regional influence between France and Germany based on the temporal alignment and recurrence of themes.

2. METHODOLOGY AND IMPROVEMENTS OVER BASE MODELS

After collecting our datasets [1–3], we augmented them with inferred temporal and geographic metadata (based on author lifespans) to enable long-term historical analysis. This involved pulling author data from Wikipedia, filtering the text for geopolitical entities, and selecting the most-frequently-occurring entity as where the poem was likely written.

We developed a custom pipeline for thematic clustering using multilingual sentence embeddings from transformer models. This allows for extraction and comparison of themes in different languages without explicit translation, which could misrepresent the poems and therefore the themes they elicit. Common themes were pulled from the Poetry Foundation [4]

By measuring the relative prevalence of these themes over time in each country, we were able to construct normalized temporal frequency functions of key themes for each country. We plotted these functions onto a geographical map to visualize the rise and fall of certain themes over time. Our interactive heatmap tool, as shown in Figure 1, allows users to select a theme from a dropdown menu, move a time slider to view the change in prevalence of the theme in France and Germany side-by-side over time, and read short historical narratives explaining possible causes of thematic shifts. This tool serves both as a research aid and an educational platform for literary scholars.

Finally, we developed narrative explanations for each major theme based on external cultural and historical context. This creates interpretative bridges between the data and literary history.

3. FINDINGS

We extracted 15 dominant themes, including **nature**, **religion**, **war**, **morality**, and **death**. For each theme, we plotted its normalized prevalence in French and German poetry from 1800–2000.

Several themes showed a **temporal relationship** with world events, such as themes of power and freedom correlating with the rise and fall of oppressive regimes. Certain themes such as “nature” and “romantic love” showed **lagged influence**, peaking in one country and then in the other a few decades later. Themes like “religion” showed **mutual decline**, suggesting shared socio-political shifts, such as secularization and shifts away from monarchy. These patterns support the hypothesis that literary influence can flow between nations due to cultural exchange,

shared trauma, and broader philosophical trends.

4. TESTING

We demonstrated our project to 6 people external to our team and received several questions, comments, and concerns about our methodologies. Many were interested in continuations of the project. This resulted in several future directions the project could be taken, such as

- pulling several themes from each poem to enhance data collection,
- integrating numerical economic data to measure influence, such as inflation, GDP, or stock market behavior, and
- using the existing datasets to analyze non-geographic patterns in theme, such as the development of individual poets' uses of certain themes over time.

5. LESSONS LEARNED

Data quality and coverage is crucial. Many corpora omit important metadata, making historical analysis difficult. **Linguistic preprocessing is language-sensitive**. POS tagging, sentence splitting, and even thematic clustering must be tuned or adapted for the specific structure of each language. **Narrative framing is essential** for making sense of trends. Pure data visualization does not reveal cause without historical interpretation. **Collaboration** between technical and literary analysis led to more insightful, nuanced conclusions than either could provide alone.

Future work on this project falls into two categories, continued data collection and additional meta-pattern analysis. By gathering more data covering more time and geography, we can discover larger and more precise trends that correspond more closely with historical events and inter-cultural ideology shift. By analyzing additional meta-patterns, we could uncover more aspects of literature that reveal the ideologies and methodologies of poets during these time periods and in these cultures.

6. CONTRIBUTIONS

Charles Averill

- Proposed meta-pattern analysis concept
- Inferred missing metadata by extracting and parsing data from Wikipedia page text

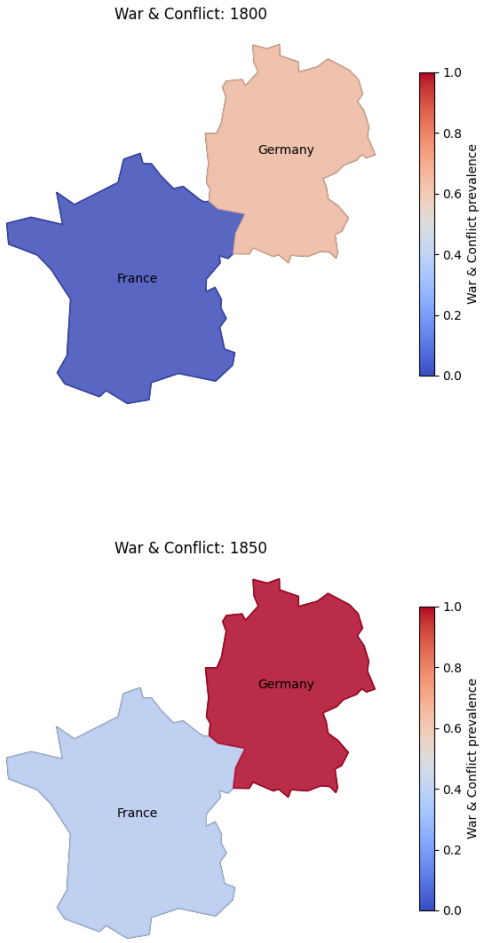

Fig. 1. Prevalence of "War & Conflict" themes over time in France and Germany, demo available at https://theme-atlas.vercal.app

- Preprocessed Oupoco and Gutenberg datasets
- Built theme analysis and visualization pipeline

Sristi Dakshit

- Extracted poem, poet, and location metadata from poetry datasets and corresponding Wikipedia pages for the DLK dataset
- Developed a theme generator using a local LLM (llama3)
- Built a standardized global theme mapping system using Sentence Transformers and cosine similarity for consistent cross-dataset theme assignment
- Generated historical narrative analyses of theme prevalence across time and countries using a

local LLM, contextualized with significant historical events.

Joel Stephen

- Created a global theme set by processing and clustering theme labels from multiple datasets, leveraging an LLM-based theme mapping pipeline to consolidate semantically similar themes into coherent groups
- Leveraged large language models, NLP techniques and web scraping to extract themes, author metadata, and temporal markers from poetry datasets, enabling semantically rich and context-aware visualizations
- Consolidated diverse poetry datasets into a unified, consistent format, facilitating thematic analysis across time and historical context.
- Designed and developed a dynamic Vue.js frontend application to visualize the evolution of poetic themes in France and Germany, helping users explore thematic trends and understand their historical context.

7. SELF-SCORING

A. Charles Averill

Points	Contribution
80	Significant exploration beyond baseline
30	Innovation or creativity: used stochastic approaches over Wikipedia text to approximate missing metadata
10	Complexity: discovered and preprocessed two large datasets, developed robust pipeline for interpretation of theme data, visualization, and trend analysis

B. Sristi Dakshit

Points	Contribution
80	Significant exploration beyond baseline
30	Creative approach in building a cross-dataset theme mapping system using multiple advanced NLP techniques
10	Complexity: Incorporating extraction techniques to fill out missing metadata; Developing a narrative generator to analyze the historical significance of poetry themes

C. Joel Stephen

Points	Contribution
80	Significant exploration beyond baseline
30	Creative approach in clustering theme labels using an LLM-based mapping pipeline to form a global theme set, and unifying diverse poetry datasets into a consistent format for temporal thematic analysis
10	Complexity: Built a detailed frontend visualization system that allows users to interactively explore fluctuations in poetic themes across France and Germany through 1700–1900s

REFERENCES

1. tnhaider, "DLK: Deutsches Lyrik Korpus (DLK) / German Poetry Corpus," (2019).
2. T. Poibeau, F. Mélanie-Becquet, C. Plancq, *et al.*, "Sonnets base of the Oupoco project," (2022).
3. aparrish, "gutenberg-poetry-corpus: A corpus of poetry from Project Gutenberg," (2018).
4. "The Poetry Foundation - Themes and Topics," (2024).