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

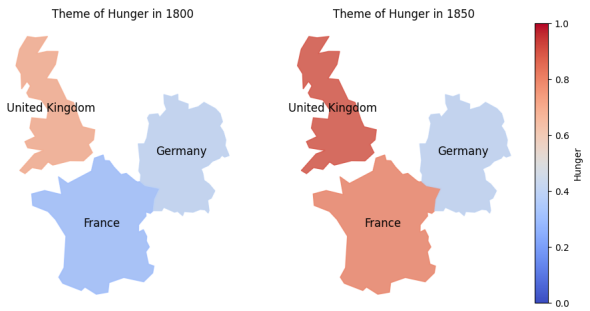

Fig. 1. Prototype mockup

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

1. GOAL

Our goal is to analyze and visualize the evolution of meta-patterns in poetry, such as thematic trends, grammatical structures, and stylistic conventions, across different regions and time periods. By leveraging classical NLP techniques and large language models, we aim to extract meaningful linguistic and conceptual features from poetry datasets and observe how these patterns propagate between cultures. This study will provide insights into how literary traditions influence one another, whether through direct cultural exchange, shared historical events, or independent parallel development.

In addition to historical insights, our work has applications in computational literary analysis, helping researchers and enthusiasts trace the origins and transformations of poetic traditions. The visualization component of our research will enable interactive exploration, allowing users to dynamically examine poetic influences over time and space.

2. SCOPE

Our final prototype will be a graphical interface that allows users to explore poetic influences and the diffusion of literary styles across time and geography. Users will be able to pan over a geographical map and observe a heatmap overlay illustrating the movement and evolution of meta-patterns across regions. A mockup is displayed in Figure 1.

3. DUTIES

Our plan is split up into three phases, each of which will be approached equally by each of our team members:

1. **Data gathering and annotation**. We begin by gathering poems that span large regions of time and space. We will find at least three nearby regions of space with long traditions of literature to pull from. For example, England, France, and Germany, or China, Korea, and Japan. These choices provide a minimum geographic granularity for free - country borders. Ideally, the granularity of regional data is much smaller, such as on the order of cities, to allow for more nuanced and educated analysis.

   Many existing datasets of poems tend not to include temporal data such as publication date. This can be ramified by selecting the midpoint of the author's lifespan, a commonly-available data point. Although this introduces a variance of ± 25 years to our temporal data, our target range is approximately 500 years (1500-2000 C.E.), so the difference remains statistically insignificant.

2. **Meta-pattern extraction**. We then utilize various NLP techniques to extract meta-patterns from the poems. Thematic information can be extracted through the use of transformer-based large language models or by analyzing sentence embeddings. Information such as grammatical structure, rhythym, and rhyme can be extracted using more traditional tokenization and part-of-speech tagging methods.

3. **Relational analysis**. Finally, we will begin to analyze the patterns of the resulting meta-patterns over time and space, potentially revealing the transfer of literary knowledge and practice between regions and languages over time. Because we will be comparing various regions, we will look for the regions with the most intermingling to present foremost in our demonstration.

4. SOURCES

We have selected these datasets to begin with:

1. DLK: Deutsches Lyrik Korpus containing 65,000 German poems by 254 authors
2. french_poetry containing 1,800 French poems by 56 authors
3. The Oupoco Database containing 4,800 French poems by 767 authors across 200 years
4. Twentieth-Century English Poetry containing 598 volumes of poetry by 283 poets over 100 years
5. Gutenberg Poetry Corpus containing roughly three million lines of poetry pulled from Project Gutenberg