Topic Modelling of Pompeian Wall Inscriptions

This is the final project for Computational Literacy course autumn 2024. In this project I am exploring the themes that are represented on Pompeii's walls using computational methods, specifically Voyant Tools and Palladio. The Pompeian wall inscriptions provide a unique lens into the lives and cultural expressions of ancient Pompeii, documenting personal, social, and public interactions in a way that literary texts of the time could not. My goal is to identify prevalent themes within this dataset and understand how they reflect the daily lives of Pompeii's inhabitants.

Research Questions

1. Can topic modelling be used to discover and identify prevalent themes in wall inscriptions found in Pompeii?
2. What are the main topics represented in Pompeian wall inscriptions, and how do they reflect the daily lives and cultural expressions of Pompeii's inhabitants?

Background

Pompeii was a Roman city that was preserved by the eruption of Mount Vesuvius in 79 CE. It offers a unique glimpse into daily life of its inhabitants through its inscriptions. These writings, which include both monumental and graffiti inscriptions, cover a wide range of topics, from politics and love to humour and philosophical reflections. Roughly 10 000 inscriptions have been found in Pompeii, many of which are wall inscriptions, known as *inscriptiones parietariae*. These inscriptions were not official records but rather personal expressions, often scratched onto walls by everyday people.

Among these, electoral inscriptions are also prominent. Used by candidates and their supporters, these short inscriptions aimed to promote candidacies and political views, often found in public spaces. These writings shed light on the political engagement of Pompeii's citizens, providing a glimpse into the public aspect of Roman elections, where personal and political messages were conveyed on the walls for widespread visibility.

Data

The dataset for this project consists of inscriptions from Volume IV of the *Corpus Inscriptionum Latinarum* (*CIL*), which documents inscriptions from Pompeii, Herculaneum, and Stabiae. This volume provides a comprehensive collection of Latin inscriptions from these regions, including both public and private texts. The inscriptions were collected from the *Epigraphik-Datenbank Clauss/Slaby*, an online database that records almost all Latin inscriptions, allowing access to the texts as plain text files.

For this project, I focused on inscriptions from Pompeii and specifically selected wall inscriptions. These include inscriptions made on stone (*lapis*) or plaster (*tectorium*), which were the most common materials for wall inscriptions in the city. The initial dataset contained over 15,000 inscriptions, which were processed to focus on those that met these criteria.

Methods

The data extraction from the *Epigraphik-Datenbank Clauss/Slaby* resulted in a total of 15,895 inscriptions. These inscriptions were copied into a Notepad file and saved as edcs_raw_data.txt. My original plan was to use the *Corpus Inscriptionum Latinarum* (*CIL*), particularly Volume IV, but the inscriptions on the *CIL* website were difficult to extract due to their dispersion across multiple pages and subcategories (https://objects.auxiliary.idai.world/Tei-Viewer/cgi-bin/teiviewer.php?manifest=BOOK-ZID881596).

Next, I cleaned the raw data to remove metadata and irrelevant information. I decided to focus on inscriptions from *CIL* IV, as these are considered the most reliable in terms of provenance. I filtered out any inscriptions that were not listed in this volume. Additionally, I only included inscriptions from wall materials (stone or plaster) and excluded those made on other materials, such as wood, metal, pottery etc. For inscriptions where the material was not specified, I included those listed under *Inscriptiones pictae* or *Inscriptiones graphio exaratae* in *CIL*, which refer specifically to wall inscriptions. I also removed editorial elements, such as brackets and parentheses indicating missing letters or abbreviations (e.g., *c(onsul)* or *co(n)s(ul)*), to produce clean, plain text. I used a Python script to execute this cleaning process.

The next step was lemmatization. I used the *Collatinus* application (https://outils.biblissima.fr/en/collatinus/) to lemmatize the cleaned text, which I then exported as lemmatized_text.txt. Since I wanted to focus on the lemmas (the root forms of the Latin words), I wrote another Python script to remove unrelated morphological details and French translations, resulting in a file containing approximately 54,000 words.

To prepare the data for analysis, I uploaded the lemmatized text into Voyant Tools for inspection. Upon reviewing the output, I noticed that many of the most frequent words were proper nouns. As these proper nouns were not relevant to the thematic analysis, I decided to remove them. To do so, I compiled a list of Roman nomina and cognomina from Wikipedia (https://en.wikipedia.org/wiki/List_of_Roman_nomina and https://en.wikipedia.org/wiki/List_of_Roman_cognomina), which I enhanced by adding a few names not found in these lists. This list was saved as roman_names.txt, and I used a Python script to remove these proper nouns from the lemmatized text. After this removal, the dataset contained approximately 37,500 words.

Once the proper nouns were removed, I further refined the corpus using Voyant Tools, which automatically removed Latin stop words. The cleaned corpus consisted of 3,510 unique words, which was still too large for meaningful thematic analysis. To focus the analysis, I selected the 100 most common words, resulting in a list of the most frequent terms in the corpus. I manually categorized these words into themes. For words that were ambiguous or uncertain, I used Voyant's Contexts and Collocates tools to identify co-occurring words, which helped to refine the thematic categories. Some general words, such as verbs, could have gone under any category, so I left them outside the analysis.

Finally, I used Palladio (https://hdlab.stanford.edu/palladio/) and Excel to create visual representations of the relationships between these categories and their occurrences in the dataset. These visualizations provided insights into the thematic structure of the inscriptions, helping to highlight patterns and connections across the dataset.

Results and Analysis

The analysis of the 100 most common words in Pompeian wall inscriptions revealed several major themes that reflect the daily life, political climate, and social interactions of the city's inhabitants. A significant portion of the inscriptions focused on electoral announcements, with words like *aedilis* 'municipal magistrate', *duumvir* 'magistrates', and *rogo* 'I ask' appearing frequently. These inscriptions, accounting for 5 384 occurrences, highlight how electoral campaigns were embedded in the public space of Pompeii.

Another prominent theme was salutations and wishes, with 1 431 occurrences. Words such as *vale* 'goodbye', *salve* 'hello', and *feliciter* 'happily' reflect the social interactions and greetings exchanged among Pompeii's citizens, emphasizing the importance of personal exchanges in everyday life.

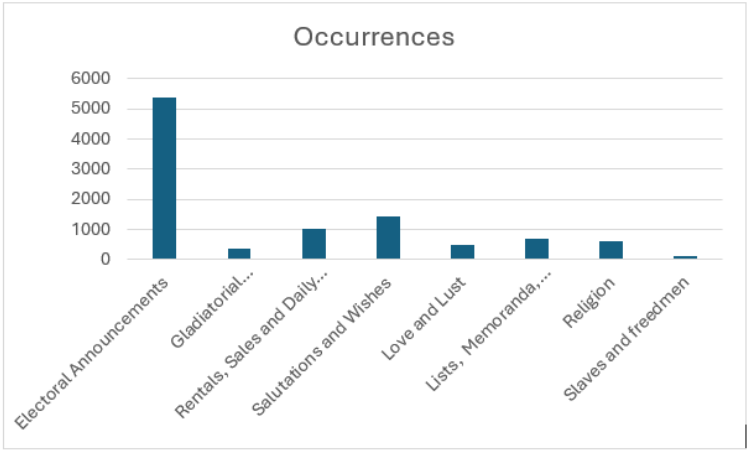
Below is a table listing the most common words found in the inscriptions, categorized by their thematic significance. This table offers a more detailed view of the words analysed and provides context for the themes discussed.

Category (8 of 8 rows displayed)	Occurrences (8 of 8 rows displayed)	Words
Rentals, Sales and Daily Life	1004	garum, vinum, pondo, scomber, semis, edo, as, denarius, vinus, asso, pasco, liquamen
Salutations and Wishes	1431	vale, valeo, salve, feliciter, saluto, bene, fortuno, venio, ago
Gladiatorial Announcements and Entertainment	375	gladiator, venatio, vis, pugno, vinco, pugna
Love and Lust	476	venus, amo, futuo, fello, cinaedus, pario
Electoral Announcements	5384	aedilis, duumvir, publica, publicum, oro, reus, jus, jure, rogo, magnum, vir, bonum, juvenis, ver, filius, vicinus, amplio, quinquennalis, verum, consul, procuro
Religion	618	lucus, sacer, sacrum, aedis, aedes, celo, sodalis
Lists, Memoranda, Notices	674	kalendae, idus, januarius, annus, nonus, nona, nonae, december
Slaves and Freedmen	98	servus, libertus

In addition to personal exchanges, the commerce and daily life of Pompeii's inhabitants were often represented in inscriptions. Terms like *garum* 'fish sauce' and *vinum* 'wine' appeared 1 004 times, suggesting that economic activities, particularly trade in food and goods, were central to public life. The theme of love and lust also emerged, with words like *Venus* 'goddess of love' and *amo* 'I love' appearing 476 times, providing insight into the personal and romantic relationships of Pompeians.

The most common words found in the Pompeian wall inscriptions. The size of each word indicates its frequency in the dataset, with larger words appearing more frequently.

The remaining themes included lists, memoranda, and notices, with 674 occurrences, reflecting practical matters such as tracking time and recording events. Timekeeping terms, like *kalendae* 'calendar' and *idūs* 'the Ides', appeared across many different categories, especially in relation to electoral announcements and public life. Religion was also represented, with terms like *sacer* 'sacred' and *aedes* 'temple' appearing 618 times, reflecting the religious life in Pompeii. The slaves and freedmen category, with 98 occurrences, reflects the social hierarchy, marking the presence and status of slaves and freed individuals in Pompeian society.



The frequency of different thematic categories in the Pompeian wall inscriptions.

Bias and Limitations

There are many issues and biases related to this dataset. The inscriptions documented in CIL represent only a small fraction of all the inscriptions from the area. Many inscriptions have been lost or destroyed or remain undiscovered. Furthermore, some inscriptions survive solely through the documentation provided in this centuries-old documentation, which is full of human errors. As the original texts have perished, there is no way to check the accuracy of the dataset or even its authenticity in some cases. In addition, the handwritten inscriptions are very hard to read and interpret, as they are in cursive style and their grammar and orthography diverge from the literary norms most Latinists are accustomed to. This will also make it difficult for machine processing.

The corpus used in this project is very large, and there may be topics that are not represented in the analysis. However, by focusing on the 100 most occurring words, I believe the goal of identifying general themes has been achieved. I am not focusing on highly specific topics, and the selection of the most frequent words should adequately cover broader themes.

During the lemmatization process, it is possible that some words were counted twice, such as "vale" and "valeo", which are different forms of the same word. However, this should not significantly impact the results, as the corpus is large enough to ensure that general themes are still identifiable.

While I made efforts to remove proper names, some may have been missed. This is unlikely to affect the overall outcome, as the focus is on large topics rather than specific individuals.

Additionally, the raw data went through multiple Python scripts, and it is possible that some information was lost in the process. Despite this, I believe the final result is reliable and represents the core themes present in the inscriptions.

Conclusion

This project shows how computational tools, like Voyant Tools, can be used to explore themes in Pompeian wall inscriptions. The analysis of the most common words in the corpus revealed key themes that align well with previous studies on Pompeii's inscriptions. While many current studies focus on very specific topics, like love and sex, this project shows that although these

themes do appear in the corpus, they aren't as common. Instead, themes related to public life and daily activities, including electoral announcements and commercial life, dominate the inscriptions. While there are some biases in the data, the results still provide useful insights into the cultural and social dynamics of Pompeii and support traditional methods of studying these inscriptions.

References

Epigraphik-Datenbank Clauss/Slaby. (n.d.). Epigraphik-Datenbank Clauss/Slaby. Retrieved from https://db.edcs.eu/epigr/epimap.php

Collatinus. (n.d.). Collatinus: A Latin lemmatizer and morphological analyzer. Retrieved from https://outils.biblissima.fr/en/collatinus/

Wikipedia. (n.d.). List of Roman nomina. Retrieved from https://en.wikipedia.org/wiki/List_of_Roman_nomina

Wikipedia. (n.d.). List of Roman cognomina. Retrieved from https://en.wikipedia.org/wiki/List_of_Roman_cognomina