

Ancient Texts NLP & Classification Project

Insights from Counts, Frequencies, and Weights

All code at https://github.com/izsolnay/Ancient_NLP

➤ ISSUE / PROBLEM

Can modern tools provide insights into ancient text corpora?

Objective

Develop a reliable machine learning model which can classify texts into genres

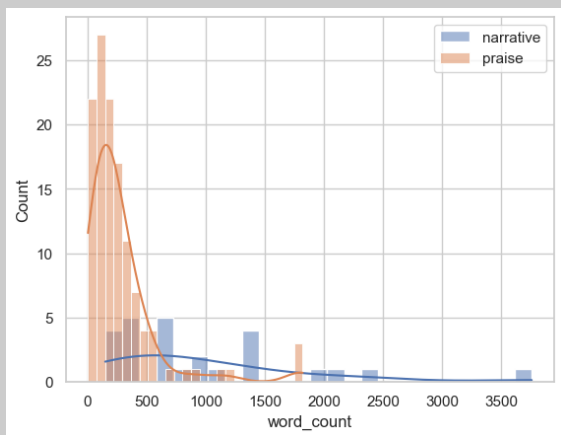
Steps for this stage

- Prepared text for word count, frequency, and weight analyses
- Introduced new features 'lemmatized', 'word_count', 'token_count', and 'split_tokens'
- Generated visuals
- Extracted top bi- and tri-grams per text
- Investigated top bi- and tri-grams by 'B_category', 'God', and 'Person'

➤ IMPACT

Text lengths vary considerably

- The genres 'City laments' and 'Narratives featuring heroes' had highest word counts with median counts above 1500
- Most genres had on average well below 1000 words
- Most gods featured in texts averaging 100 to 400 words



To address

- Grossly varying text lengths
- Previous insight that some genres have only 1 or 2 texts attributed to them

➤ RESPONSE

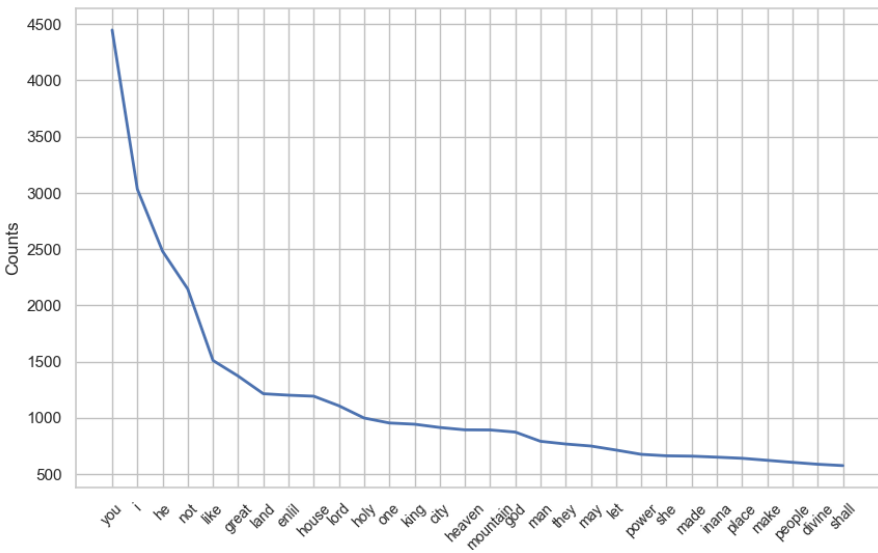
Sample of top words by count
(excluding pronouns)

- Narrative genres: 'like' and 'Inana'
- Praise texts: 'great', 'holy', 'may', and 'land'
- The goddess Inana: 'Inana' and 'heaven'
- The god Ninurta: 'Ninurta', 'Lord', and 'Enlil'

- The moon as Nanna: 'king' and 'great'
- The moon as Suen: 'heaven', 'divine', 'power', 'light', and 'earth'
- The moon as Nanna-Suen: 'house', and 'Enlil'

➤ KEY INSIGHTS

Top words by count for all texts were 'you', 'I', 'he', 'not', and 'like'



Top bi-grams

tfidf_top_word	count	percentage
divine power	12	3.1%
slave girl	8	2.1%
may he	5	1.3%

Top tri-grams

tfidf_top_word	count	percentage
No significant terms	76	19.9%
great mountain enlil	6	1.6%
life may he	5	1.3%
great divine power	4	1.0%

Notable top term in bi- and tri-grams: 'divine power'

- In 36 top tri-grams and 12 top bi-grams of all texts
- In 9 'Royal praise poetry and hymns to deities on behalf of rulers'
- In 5 trigrams of texts featuring person Ishme-Dagan
- In 6 trigrams of texts featuring the goddess Ninisina