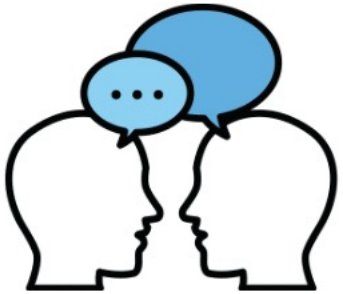
Laura Soffiantini
25/10/2022
Gand

"Yet there is *pattern* in't!"

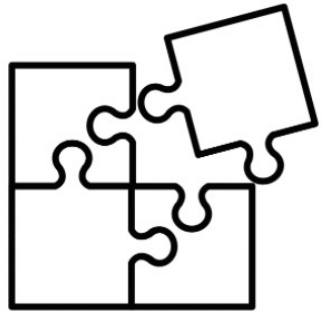
Studying formulaic expressions in Latin funerary epigraphy

Part 1. What is a formulaic sequence?



Half of the human production of speech (oral or written) is **formulaic**.

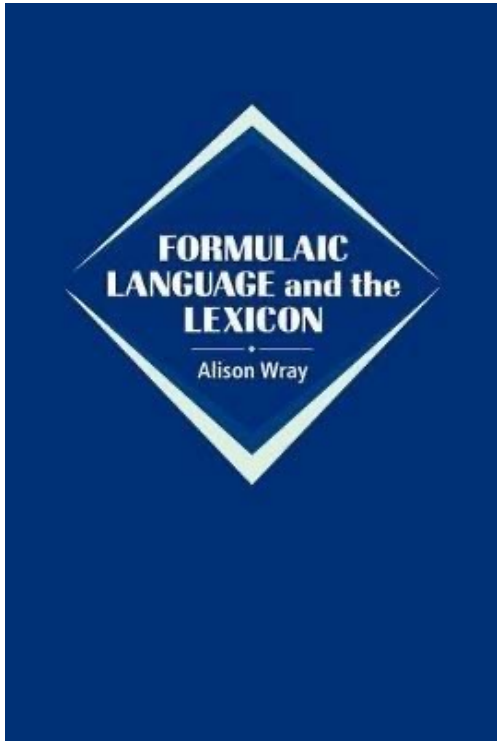
The ability to recall and fluently use formulae determine the **effectivity** of our communication.



The speakers appear to renounce the great freedom that the language offers and opts for **a single format**.

- idioms (*over the top of my head*)
- phrasal verbs (*pick up*)
- collocation (*strong coffee)*
- lexical bundles (*and as a result of*)
- conversational formulae (*'Are you OK?'*)

- idioms (*over the top of my head*)
- phrasal verbs (*pick up*)
- collocation (*strong coffee)*
- lexical bundles (*and as a result of*)
- conversational formulae (*'Are you OK?'*)



A. Wray, Formulaic language and the lexicon. 2002. Cambridge University Press.

and as a result of

"Multi-word sequence
in a (semi-)fixed form."

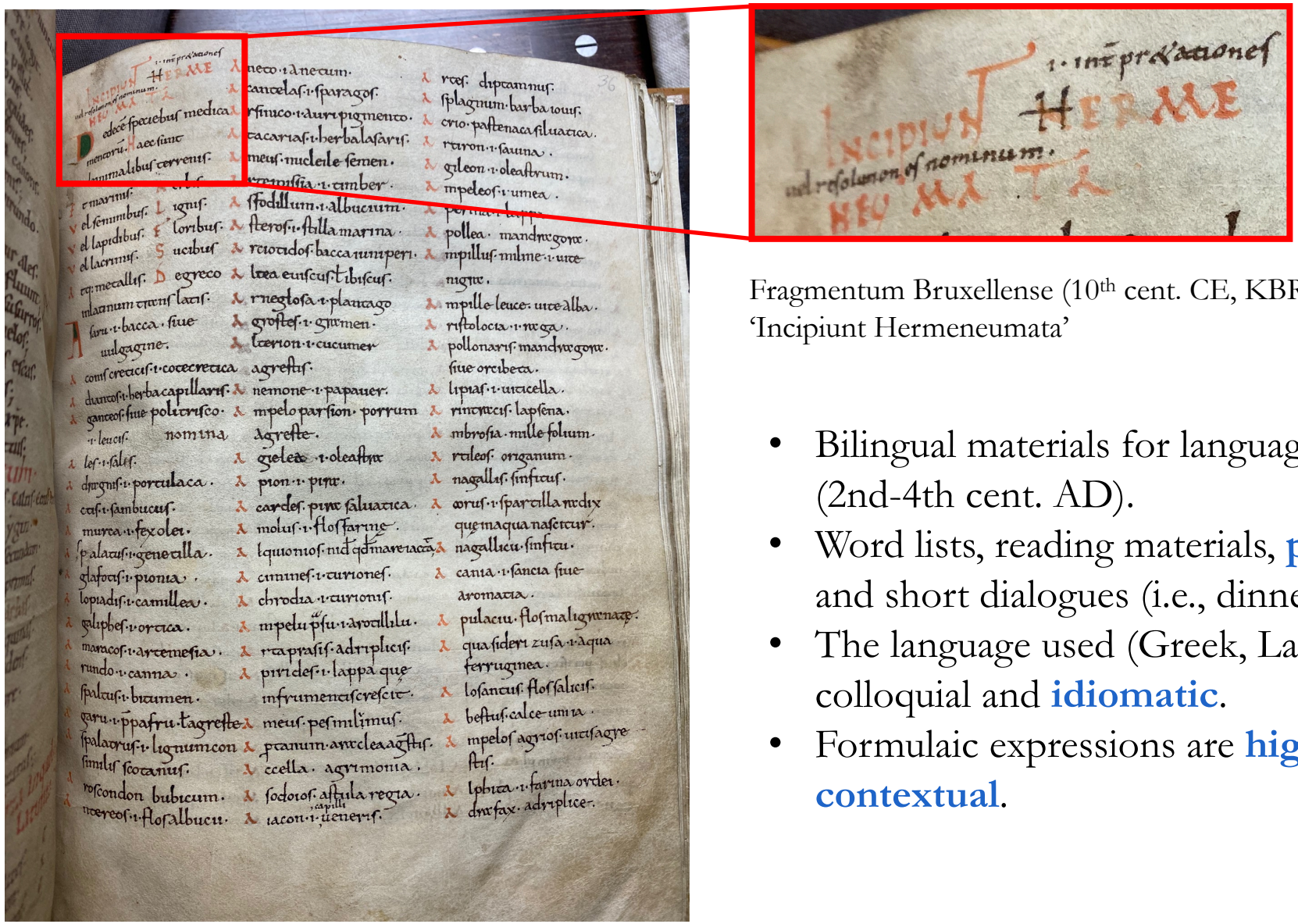
Formulaic Language (FL) and **Second Language Acquisition (SLA)**

Schmitt, Norbert. "Norbert Schmitt's essential bookshelf: Formulaic language." *Language Teaching*, 2022, 1-12.

Siyanova-Chanturia, Anna, Pellicer-Sánchez, Ana (eds). *Understanding formulaic language: A second language acquisition perspective*. 2019.

Wolter, Brendt, Yamashita, Junko. "Processing collocations in a second language: A case of first language activation?". *Applied Psycholinguistics*, 36 n. 5 (2015), 1193–1221.

Wray Alison. "Formulaic language." *Language Teaching*, 46 n. 3 (2013), 316-334.



Fragmentum Bruxellense (10th cent. CE, KBR): 'Incipiunt Hermeneumata'

- Bilingual materials for language learning (2nd-4th cent. AD).
- Word lists, reading materials, **phrasebooks** and short dialogues (i.e., dinner party).
- The language used (Greek, Latin) is colloquial and **idiomatic**.
- Formulaic expressions are **highly contextual**.

Several applications of FL studies (SLA, neuro-linguistics, automatic machine translation).

Several applications of FL studies (SLA, neuro-linguistics, automatic machine translation).

Novel methodologies to study larger corpora of formulaic texts.

Several applications of FL studies (SLA, neuro-linguistics, automatic machine translation).

Novel methodologies to study larger corpora of formulaic texts.

A broader discussion on formulaicity in Literary and Historical Studies.

Part 2. Why is FL relevant in epigraphy?

Military diploma. Provenience unknown, 114-117 AD.

Epigraphy is the study of texts (=inscriptions) written directly on durable materials (stone, pottery, metal).

Epigraphy aims at interpreting the meaning of the inscriptions, collocating them in their historical context.

S. Panciera, "What is an inscription?," *ZPE* (2012).

Orphic gold tablet. Thessaly, 4th cent. BC.

Inscriptions vary in material, form, language, and content.

Inscriptions are evidence of the thought, society and history of past civilizations.

Funerary inscription. Nikaia, 550-40 BC.

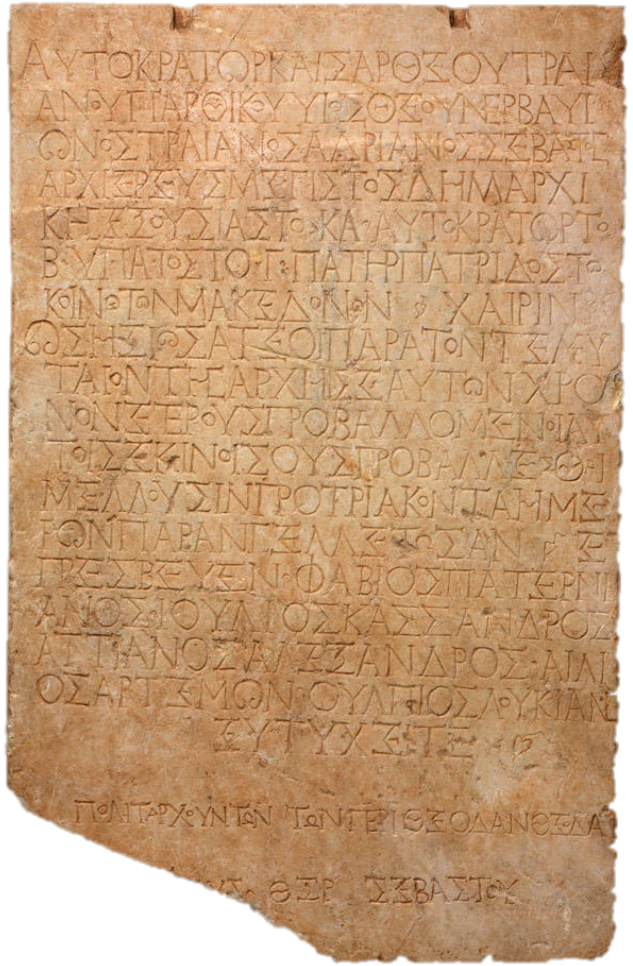
What is missing?

Funerary inscription. Nikaia, 550-40 BC.

Restoration: reconstruct what is missing.

Inscriptions of the same kind tend to share a standardized set of formulae.

Letter of the emperor Hadrian. Chalkidike (?) 137 AD.



IG X, 593 (cf. SEG 37.593)

Αὐτοκράτωρ Καῖσαρ Θεοῦ Τραι-
ανοῦ Παρθικοῦ υἱὸς, Θεοῦ Νέρβα υἱ-
ωνὸς, Τραιανὸς Ἁδριανὸς Σεβαστὸς,
ἀρχιερεὺς μέγιστος, δημαρχι-
κῆς ἐξουσίας τὸ κα', αὐτοκράτωρ τὸ
β', ὕπατος τὸ γ', πατὴρ πατρίδος,
τῷ Κοινῷ τῶν Μακεδόνων χαίρ(ε)ιν·
Ὡς ἠξιώσατε, οἱ παρὰ τὸν τελευ-
ταῖον τῆς ἄρχης ἑαυτῶν χρό-
νον ἑτέρους προβαλλόμενοι, αὐ-
τοῖς ἐκ(ε)ίνοις οὓς προβάλλεσθαι
μέλλουσιν πρὸ τριάκοντα ἡμέ-
ρων παρανγελλέτωσαν. Ἐ-
πρέσβευεν (sic) Φάβιος Πατερνι-
ανός, Ἰούλιος Κάσσανδρος,
Ἀττιανὸς Ἀλέξανδρος, Αἴλι-
ος Ἀρτέμων, Οὔλπιος Λουκιανός·
εὐτυχεῖτε·
πολιταρχούντων τῶν περὶ Θεοδᾶν Θεοδᾶ
[ἔτο]υς θξρ' σεβαστοῦ

Opening formula (name of the sender)

Greetings and name of the recipient

Closing formula ('Be fine!')

Date

Fragment (architectural?). Seleucia, 1st-2nd cent. AD.

- The language of inscriptions is **highly formulaic**.
- Identifying FL is a **crucial step** to reconstruct damaged texts.
- The study of FL in epigraphy relies on the access of **vast repositories** of texts.

Part 3. The project: Latin funerary formulae

Research question

How formulae are used in Latin funerary inscriptions? Are there some patterns in it?

Research goal

Understand the global properties of FL in Latin tomb texts.

Research question

How formulae are used in Latin funerary inscriptions? Are there some patterns in it?

Research goal

Understand the global properties of FL in Latin tomb texts.

Latin funerary inscriptions.

- Latin texts written on tombstone to **commemorate** a deceased person.
- Most **copious** inscriptions from the Roman Empire.

Research question

How formulae are used in Latin funerary inscriptions? Are there some patterns in it?

Research goal

Understand the global properties of FL in Latin tomb texts.

Latin funerary inscriptions.

- Latin texts written on tombstone to **commemorate** a deceased person.
- Most **copious** inscriptions from the Roman Empire.

Latin funerary formulae.

- Funerary texts are **highly standardized**.
- Examples of formulae: RIP, sit tibi terra levis', 'Dis Manibus', 'hic situs est'.

Research question

How formulae are used in Latin funerary inscriptions? Are there some patterns in it?

Research goal

Understand the global properties of FL in Latin tomb texts.

Latin funerary inscriptions.

- Latin texts written on tombstone to **commemorate** a deceased person.
- Most **copious** inscriptions from the Roman Empire.

Latin funerary formulae.

- Funerary texts are **highly standardized**.
- Examples of formulae: RIP, sit tibi terra levis', 'Dis Manibus', 'hic situs est'.

Methodologies.

- Testing two methods of **formulaic extraction** in epigraphic texts.
- Perform **Semantic Network Analysis** to the extracted formulae.

Identify FL
in Latin funerary texts
(OBJ1)

FL extraction

1. Rule-based method

Relying on a manually-annotated set of funerary formulae (RegEx).

2. Statistics-based method

Automatic extraction based the frequency-related definition of (Nltk).

Identify FL
in Latin funerary texts
(OBJ1)

FL extraction

1. Rule-based method

Relying on a manually-annotated set of funerary formulae (RegEx).

2. Statistics-based method

Automatic extraction based the frequency-related definition of (Nltk).

Explore the
combinations of
formulae in the texts
(OBJ2)

Semantic Network Analysis

Generate a **network** of co-occurrences of the formulae in the text.

Perform SNA and interpret the network metrics (density, degree, clustering coefficient) (NetworkX).

Identify FL
in Latin funerary texts
(OBJ1)

FL extraction

1. Rule-based method
Relying on a manually-annotated set of funerary formulae (RegEx).

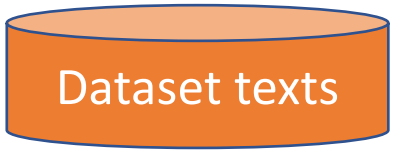
2. Statistics-based method
Automatic extraction based the frequency-related definition of (Nltk).

Explore the
combinations of
formulae in the texts
(OBJ2)

Semantic Network Analysis

Generate a **network** of co-occurrences of the formulae in the text.

Perform SNA and interpret the network metrics (density, degree, clustering coefficient) (NetworkX).

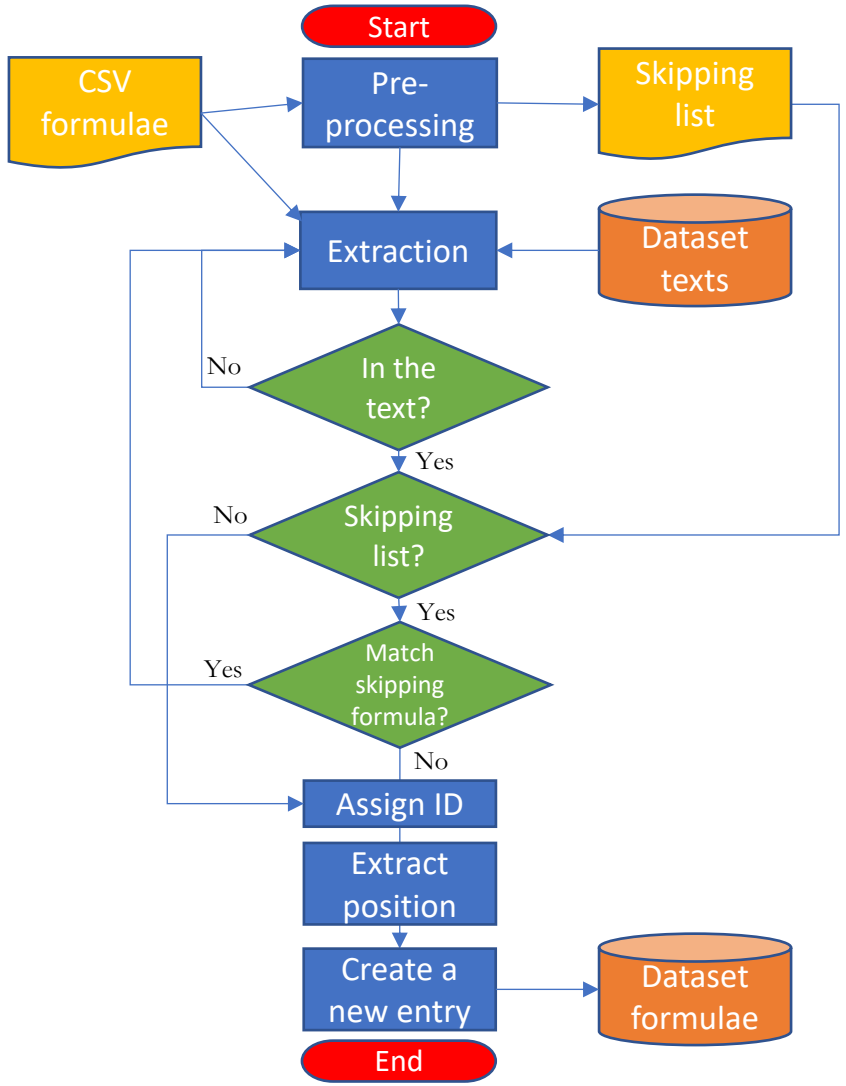


Total texts: 99,894
Language: Latin
Source: Clauss-Slaby Epigraphic database
ID: Trismegistos stable identifier

	Documents	Words	W/D
Rome	31,923	508,829	16
Latium	35,970	574,680	16
Numidia	13,888	172,049	12
Africa	10,565	136,331	13
Hispania	7316	105,104	14
Gallia	6244	79,173	13
Total	99,894	1,468,042	14.7

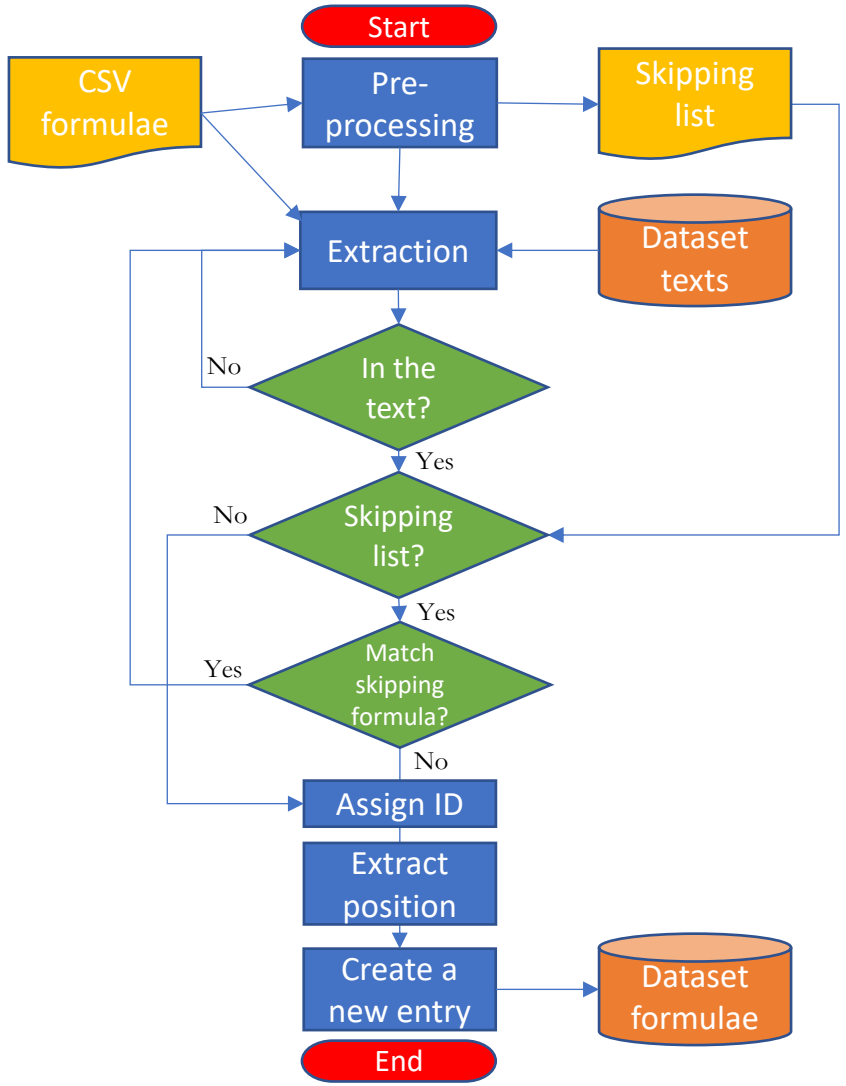
Provenience of the texts sorted by the size of the corpus.

FL extraction: manual annotation



New database of formulae: 323,058 entries. Each occurrence was assigned a unique numerical ID and associated with the position in the text and the ID of the corresponding text.

FL extraction: manual annotation

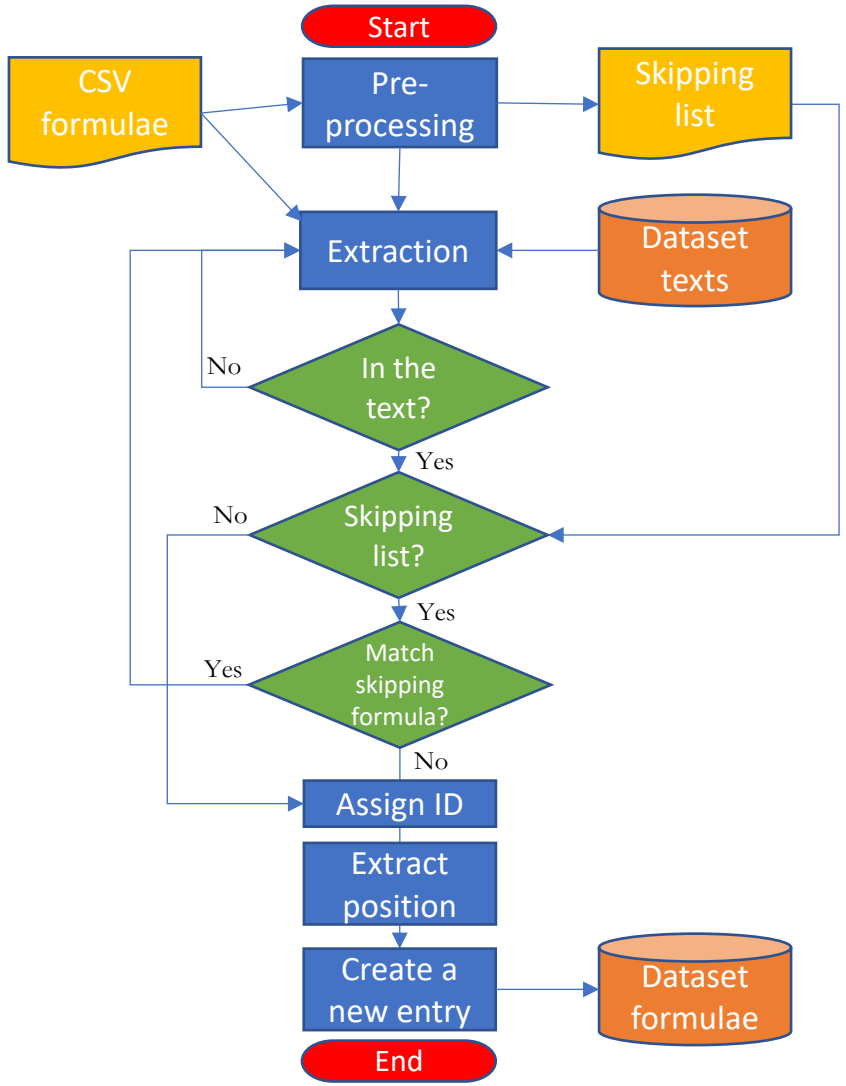


New database of formulae: 323,058 entries. Each occurrence was assigned a unique numerical ID and associated with the position in the text and the ID of the corresponding text.

Formula	Formula
1. Dis Manibus	14. hic sita sunt
2. Manibus	15. sit tibi terra levis
3. hic situs est	16. sit vobis terra levis
4. hic siti sunt	17. tibi terra levis
5. hic sita est	18. terra tibi levis sit
6. sita est hic	19. terra levis
7. situs est hic	20. sit terra levis
8. sita hic	21. tibi terra levis sit
9. hic sita	22. sit tibi terra
10. hic situs	23. vixit
11. ossa hic sita sunt	24. hic sita est
12. hic est situs	25. levis terra tibi sit
13. ossa	26. fecit

Sample of the 'CSV formulae'.

FL extraction: manual annotation

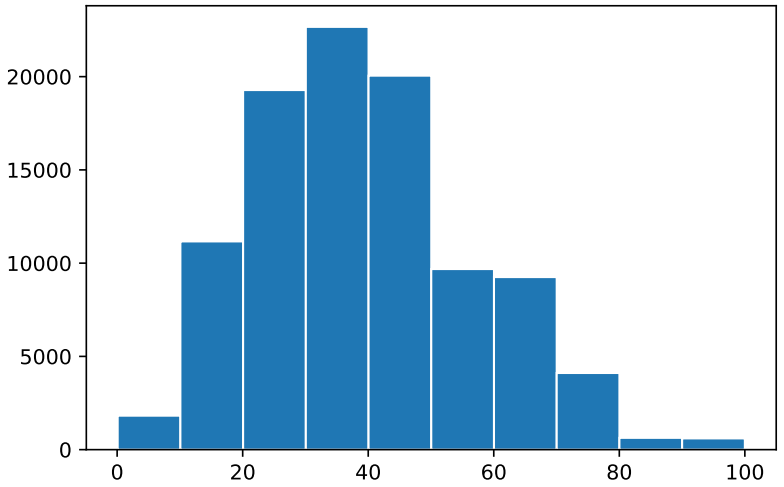


New database of formulae: 323,058 entries. Each occurrence was assigned a unique numerical ID and associated with the position in the text and the ID of the corresponding text.

Formula	Formula
1. Dis Manibus	14. hic sita sunt
2. Manibus	15. sit tibi terra levis
3. hic situs est	16. sit vobis terra levis
4. hic siti sunt	17. tibi terra levis
5. hic sita est	18. terra tibi levis sit
6. sita est hic	19. terra levis
7. situs est hic	20. sit terra levis
8. sita hic	21. tibi terra levis sit
9. hic sita	22. sit tibi terra
10. hic situs	23. vixit
11. ossa hic sita sunt	24. hic sita est
12. hic est situs	25. levis terra tibi sit
13. ossa	26. fecit

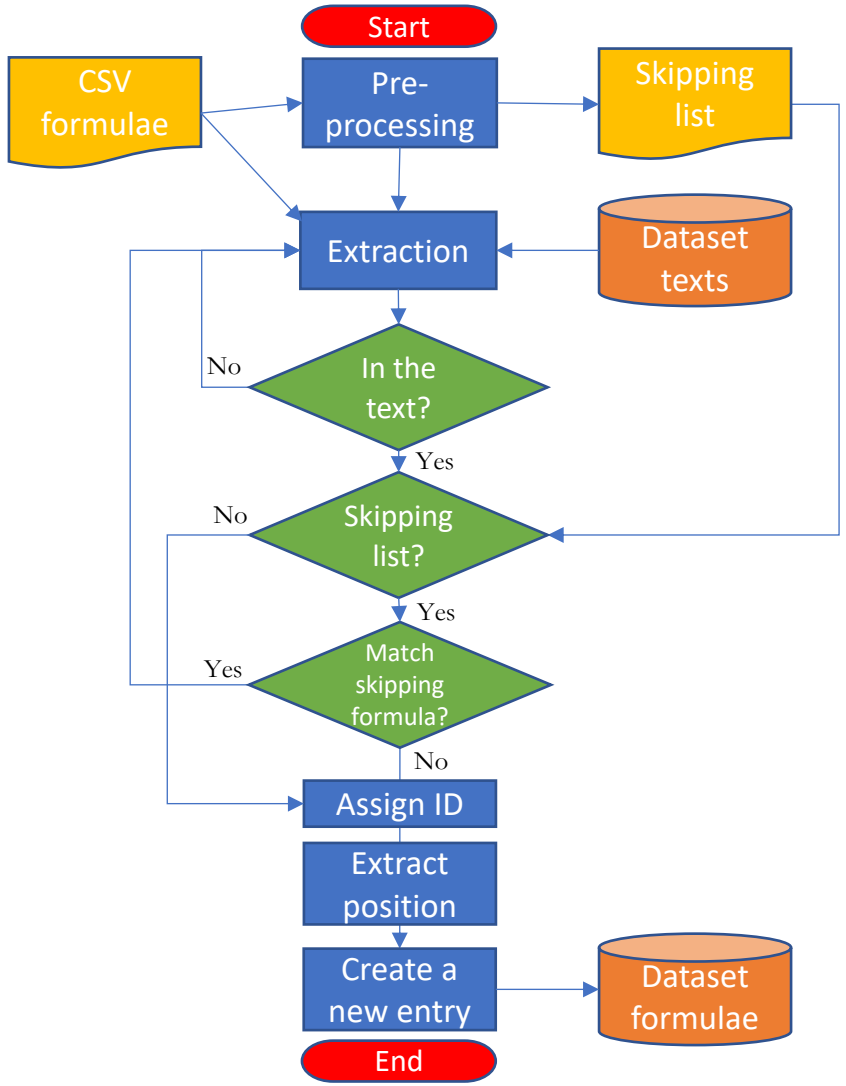
Sample of the 'CSV formulae'.

99% of texts were matched during the extraction. On average, 3.25 formulae were detected in each text.



The histogram shows the size of groups of texts (y-axis) with the same percentage (x-axis). On average, we extracted **40.38%** of each text.

FL extraction: manual annotation

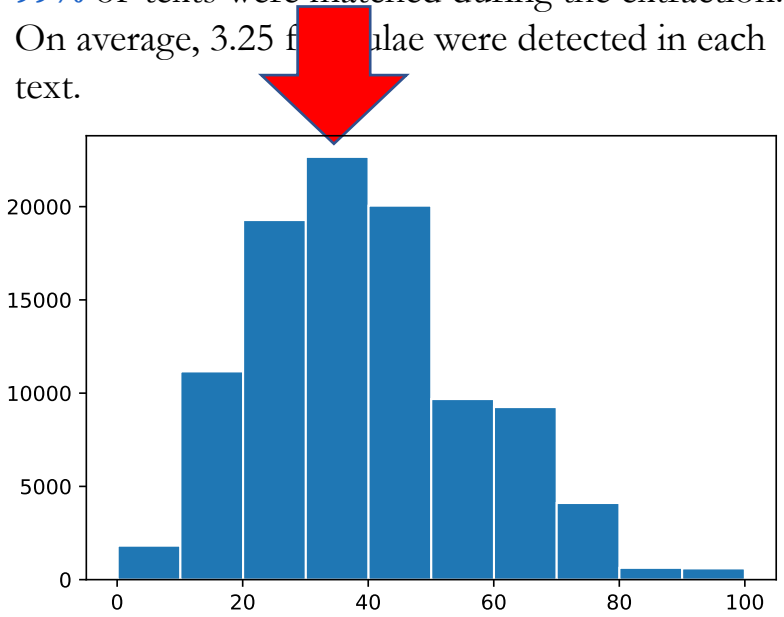


New database of formulae: 323,058 entries. Each occurrence was assigned a unique numerical ID and associated with the position in the text and the ID of the corresponding text.

Formula	Formula
1. Dis Manibus	14. hic sita sunt
2. Manibus	15. sit tibi terra levis
3. hic situs est	16. sit vobis terra levis
4. hic siti sunt	17. tibi terra levis
5. hic sita est	18. terra tibi levis sit
6. sita est hic	19. terra levis
7. situs est hic	20. sit terra levis
8. sita hic	21. tibi terra levis sit
9. hic sita	22. sit tibi terra
10. hic situs	23. vixit
11. ossa hic sita sunt	24. hic sita est
12. hic est situs	25. levis terra tibi sit
13. ossa	26. fecit

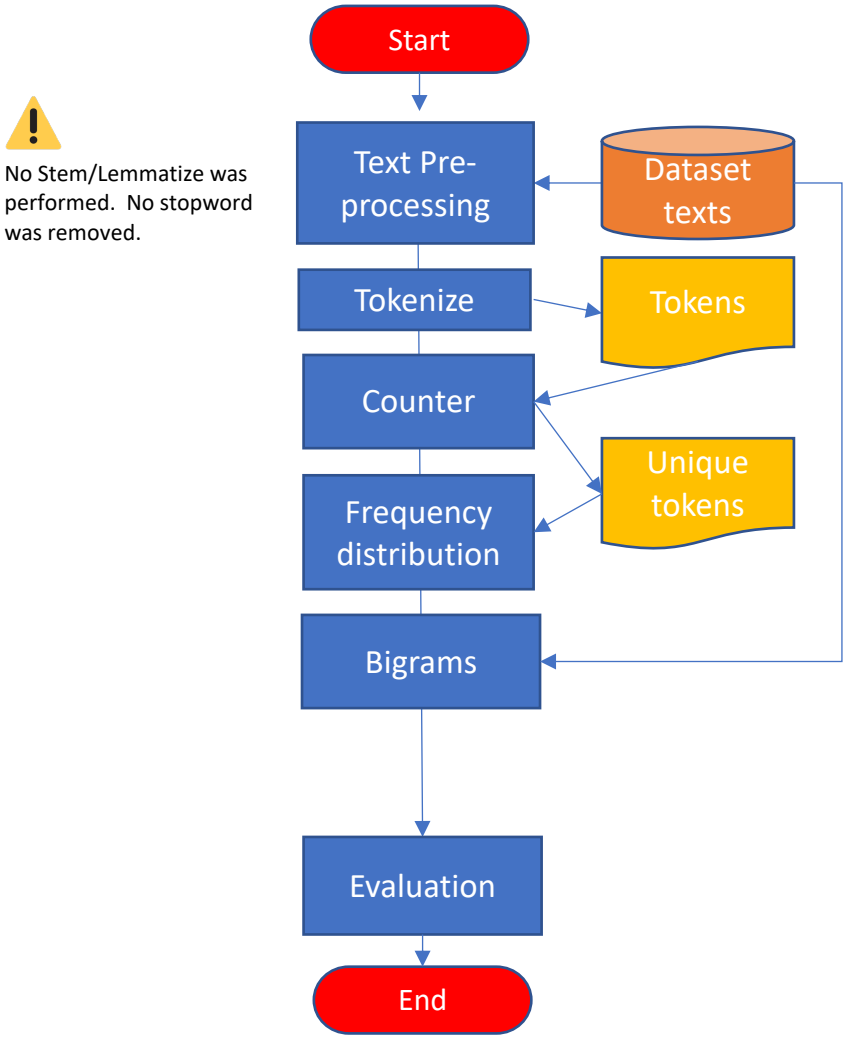
Sample of the 'CSV formulae'.

99% of texts were matched during the extraction. On average, 3.25 f[illegible]ulae were detected in each text.



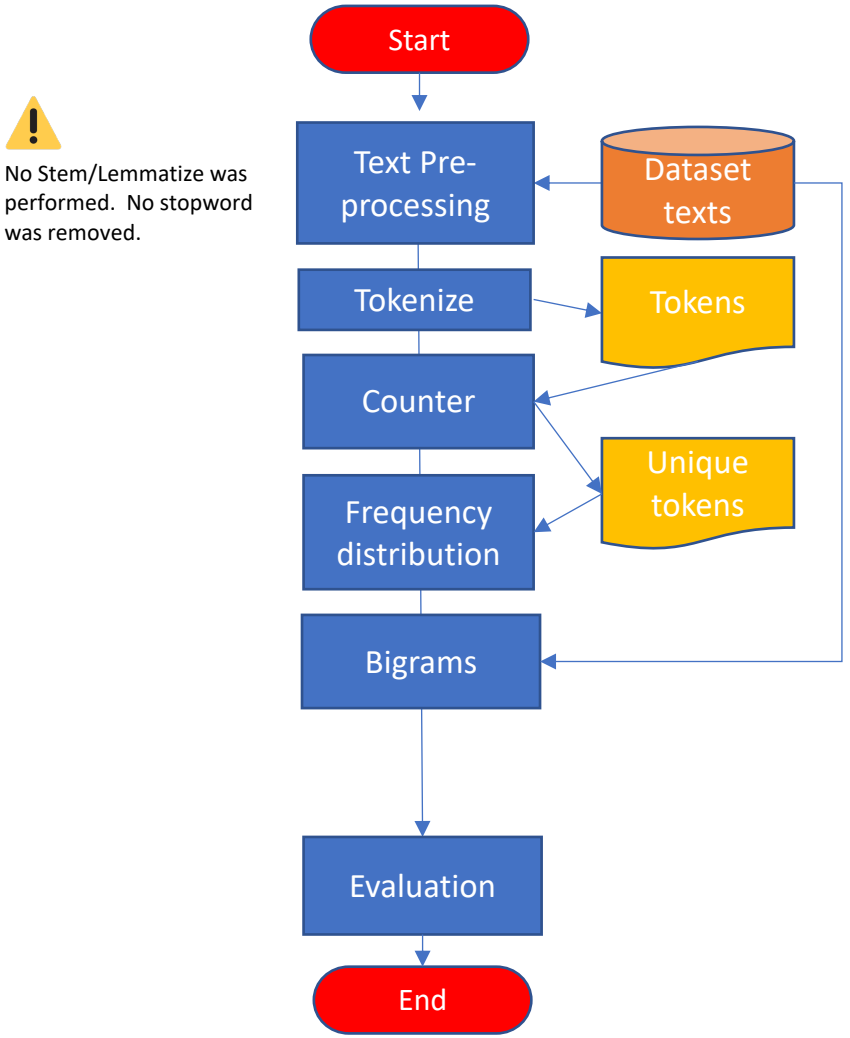
The histogram shows the size of groups of texts (y-axis) with the same percentage (x-axis). On average, we extracted **40.38%** of each text.

FL extraction: frequency-based definition of 'formula'



Total words: 1,468,042; **Vocabulary size**: 87,362 unique tokens; **Lexical diversity**: 16.8 (tot words/vocabulary size).

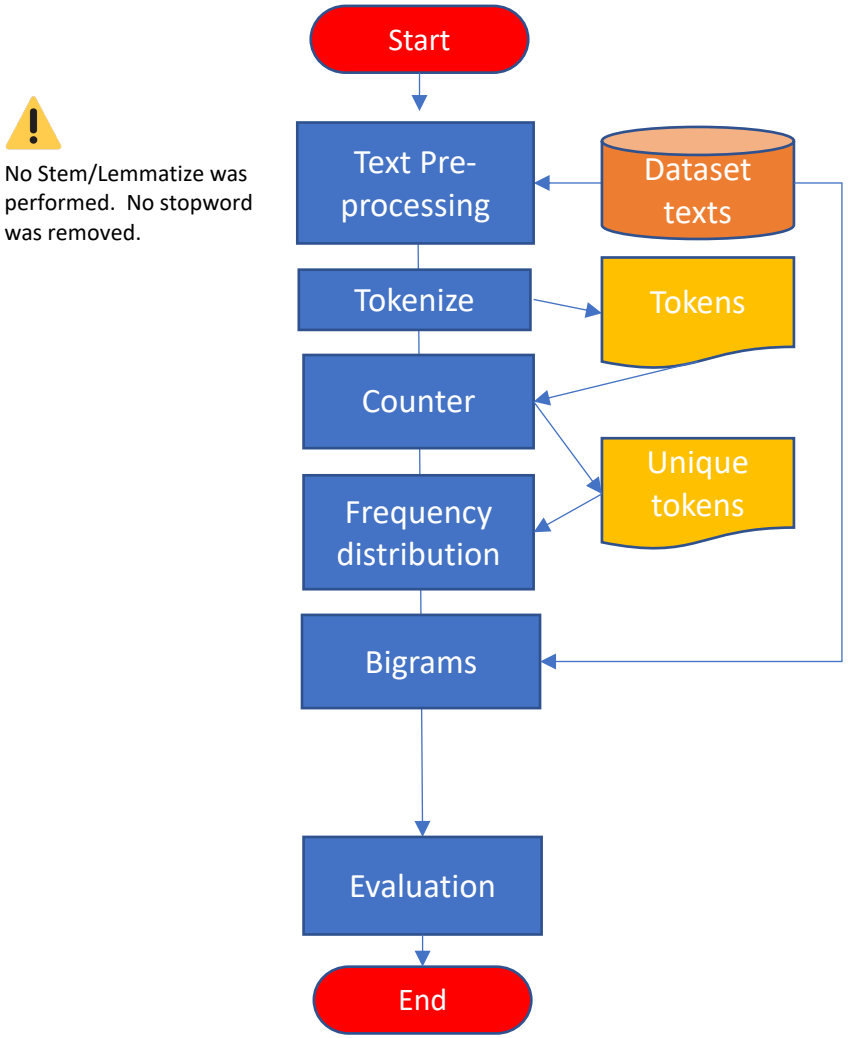
FL extraction: frequency-based definition of 'formula'



Total words: 1,468,042; **Vocabulary size**: 87,362 unique tokens; **Lexical diversity**: 16.8 (tot words/vocabulary size).

Formula	Frequency
manibus	63,219
dis	62,958
vixit	62,592
annos	46,151
et	36,327
hic	24,503
bene	22,912
est	21,030
merenti	20,659
sacrum	20,268
fecit	20,268
annis	17,418
situs	13,211
coniugi	12,661
in	12,338
qui	10,205
menses	10,016
filio	8262
sita	8090
annorum	7919

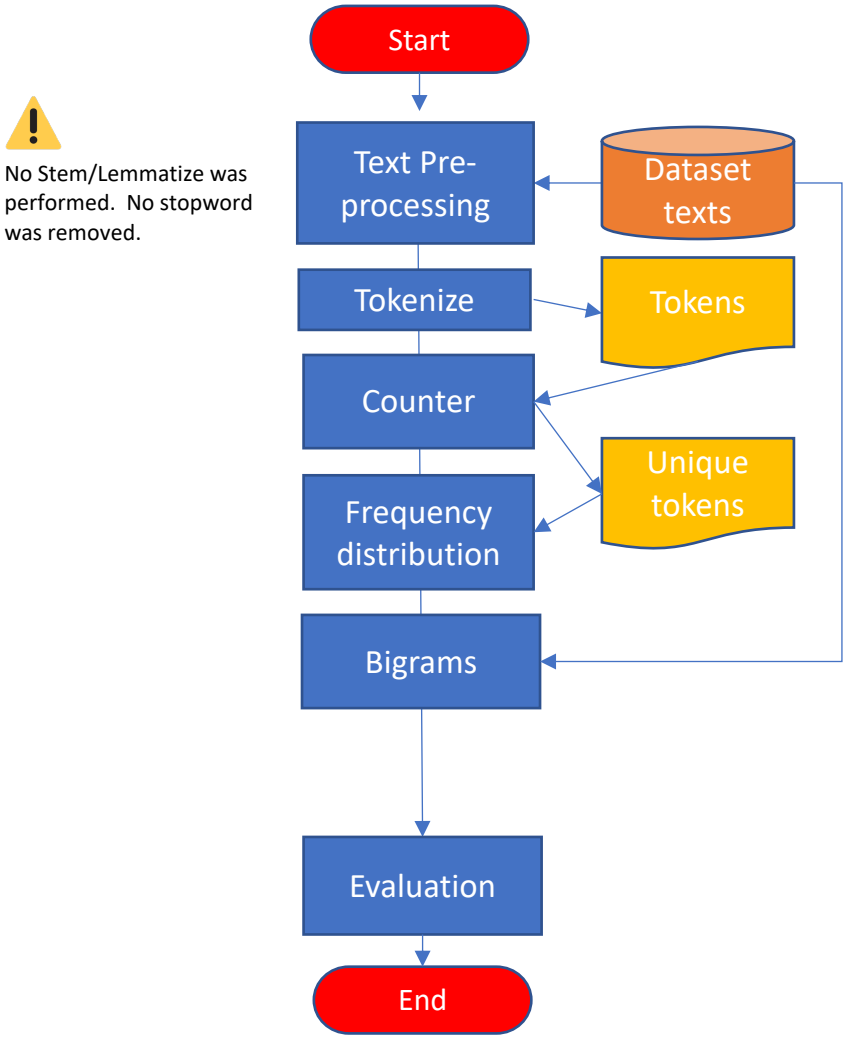
FL extraction: frequency-based definition of 'formula'



Total words: 1,468,042; **Vocabulary size**: 87,362 unique tokens; **Lexical diversity**: 16.8 (tot words/vocabulary size).

Formula	Frequency
manibus	63,219
dis	62,958
vixit	62,592
annos	46,151
et	36,327
hic	24,503
bene	22,912
est	21,030
merenti	20,659
sacrum	20,268
fecit	20,268
annis	17,418
situs	13,211
coniugi	12,661
in	12,338
qui	10,205
menses	10,016
filio	8262
sita	8090
annorum	7919

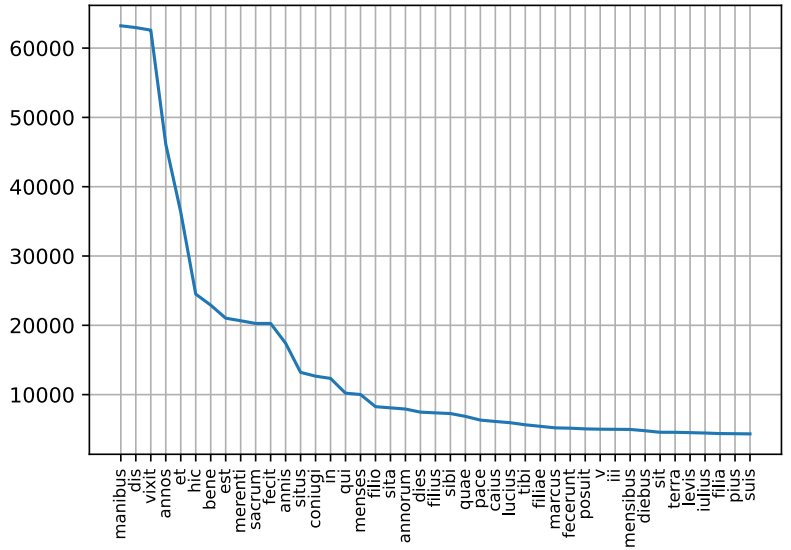
FL extraction: frequency-based definition of 'formula'



No Stem/Lemmatize was performed. No stopword was removed.

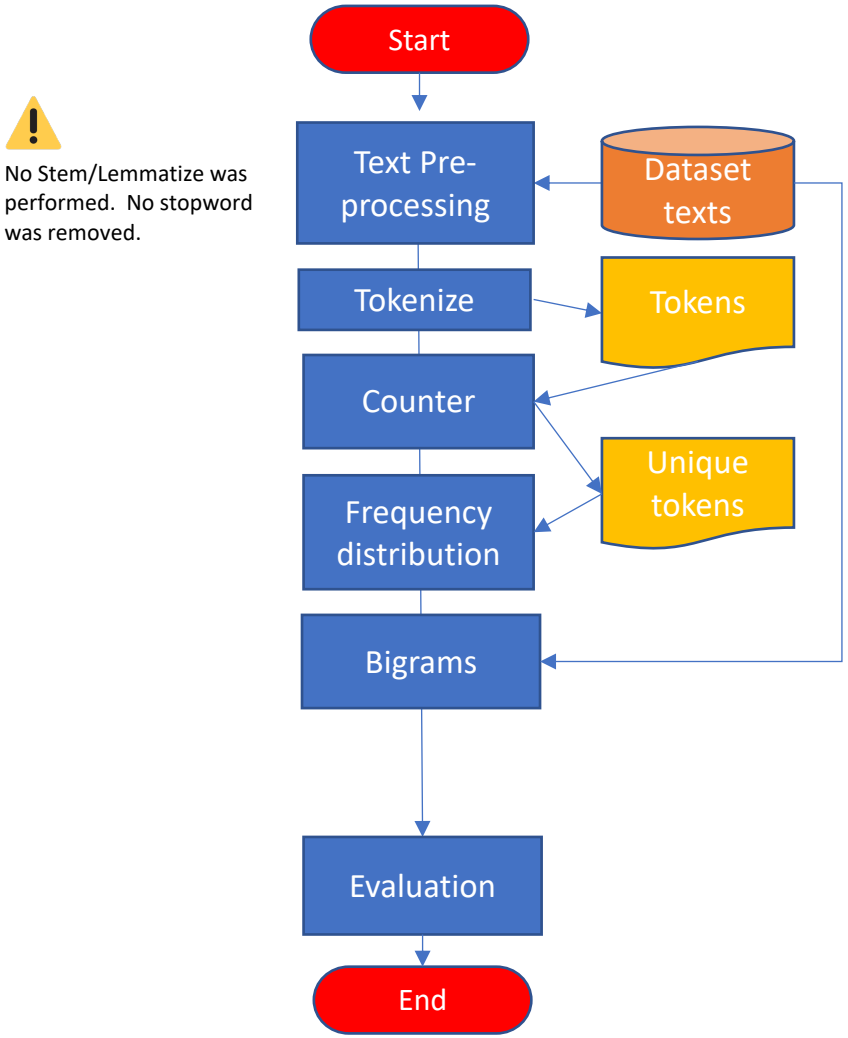
Total words: 1,468,042; **Vocabulary size**: 87,362 unique tokens; **Lexical diversity**: 16.8 (tot words/vocabulary size).

Formula	Frequency
manibus	63,219
dis	62,958
vixit	62,592
annos	46,151
et	36,327
hic	24,503
bene	22,912
est	21,030
merenti	20,659
sacrum	20,268
fecit	20,268
annis	17,418
situs	13,211
coniugi	12,661
in	12,338
qui	10,205
menses	10,016
filio	8262
sita	8090
annorum	7919



Plot of the frequency distribution of a sample of tokens (frequency>4000). **0.004%** of all the words have a cumulative frequency of 16% of all the occurrences in the corpus.

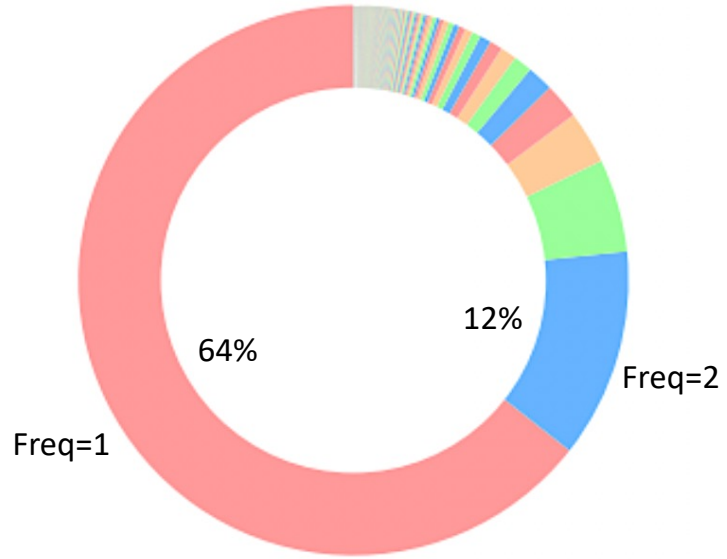
FL extraction: frequency-based definition of 'formula'



No Stem/Lemmatize was performed. No stopword was removed.

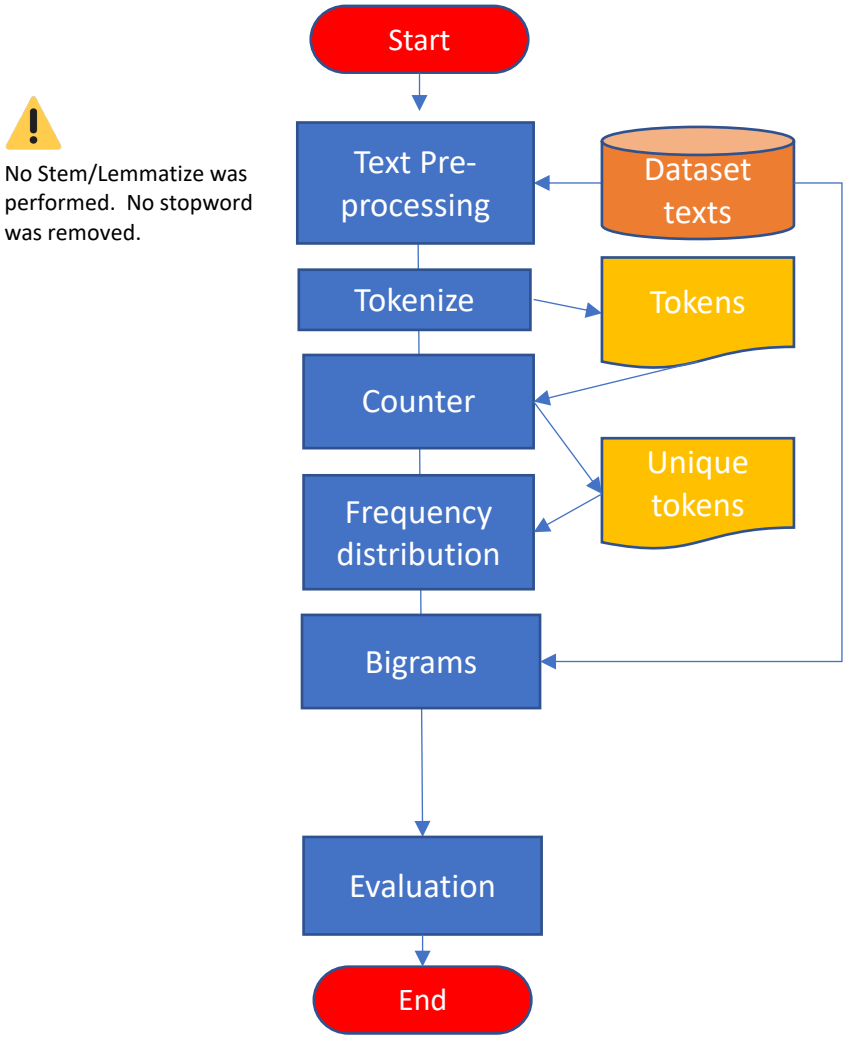
Total words: 1,468,042; **Vocabulary size**: 87,362 unique tokens; **Lexical diversity**: 16.8 (tot words/vocabulary size).

Formula	Frequency
manibus	63,219
dis	62,958
vixit	62,592
annos	46,151
et	36,327
hic	24,503
bene	22,912
est	21,030
merenti	20,659
sacrum	20,268
fecit	20,268
annis	17,418
situs	13,211
coniugi	12,661
in	12,338
qui	10,205
menses	10,016
filio	8262
sita	8090
annorum	7919



Donut chart of the frequency distribution. Most of the tokens (56,251=64%) have frequency equal to 1.

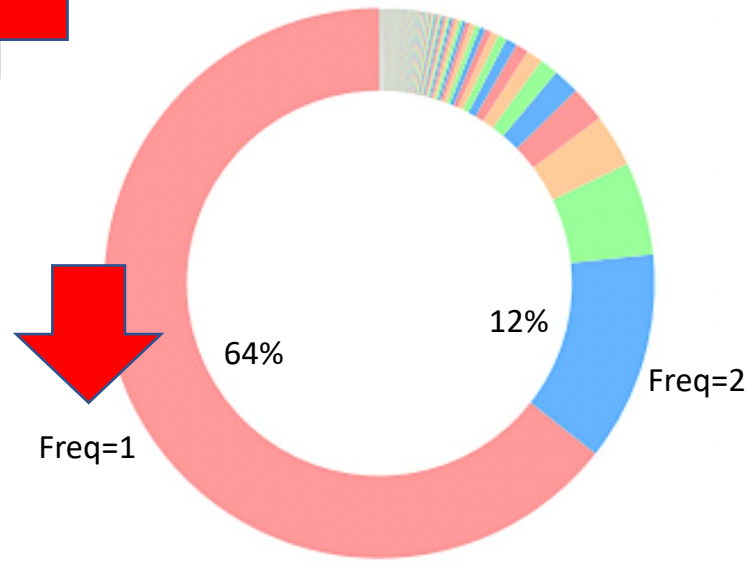
FL extraction: frequency-based definition of 'formula'



No Stem/Lemmatize was performed. No stopword was removed.

Total words: 1,468,042; **Vocabulary size**: 87,362 unique tokens; **Lexical diversity**: 16.8 (tot words/vocabulary size).

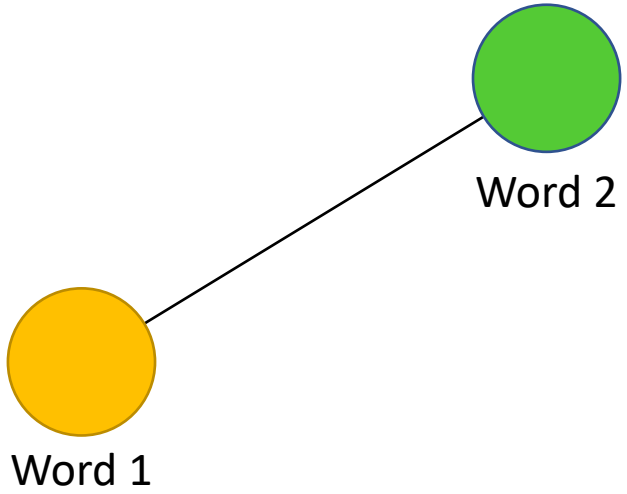
Formula	Frequency
manibus	63,219
dis	62,958
vixit	62,592
annos	46,151
et	36,327
hic	24,503
bene	22,912
est	21,030
merenti	20,659
sacrum	20,268
fecit	20,268
annis	17,418
situs	13,211
coniugi	12,661
in	12,338
qui	10,205
menses	10,016
filio	8262
sita	8090
annorum	7919



Donut chart of the frequency distribution. Most of the tokens (56,251=64%) have frequency equal to 1.

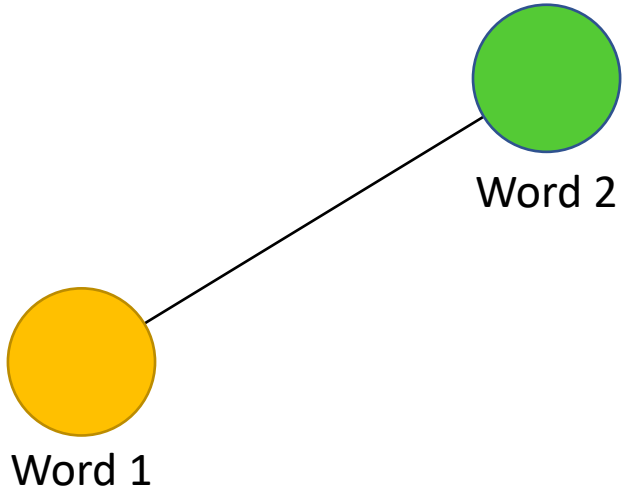
Semantic Network Analysis (SNA)

A **semantic network** represents relations between words. **Nodes** represent words and **edges** represent the relations between them.

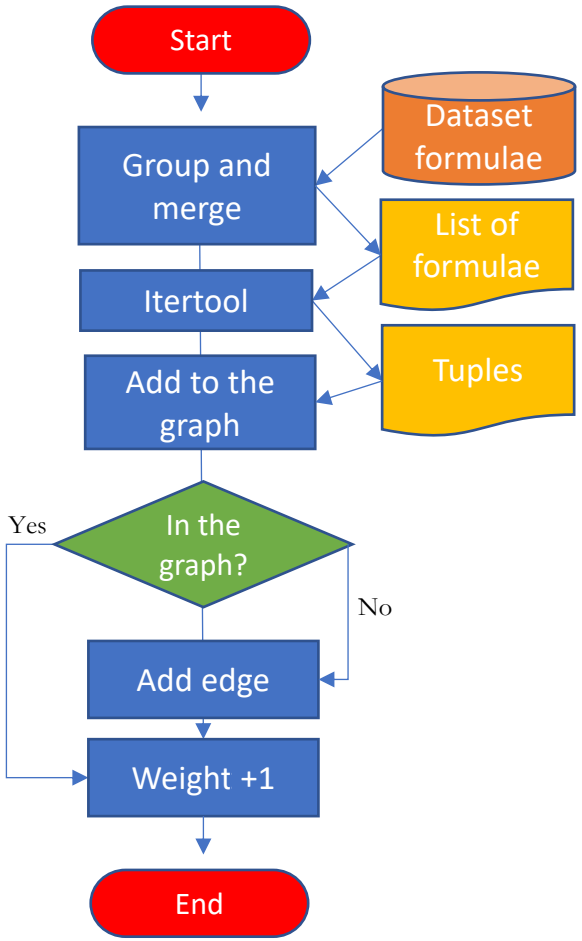


Semantic Network Analysis (SNA)

Clearly defining the kind of relation that we are looking at is a fundamental step (i.e., co-occurrence in the same text-window).

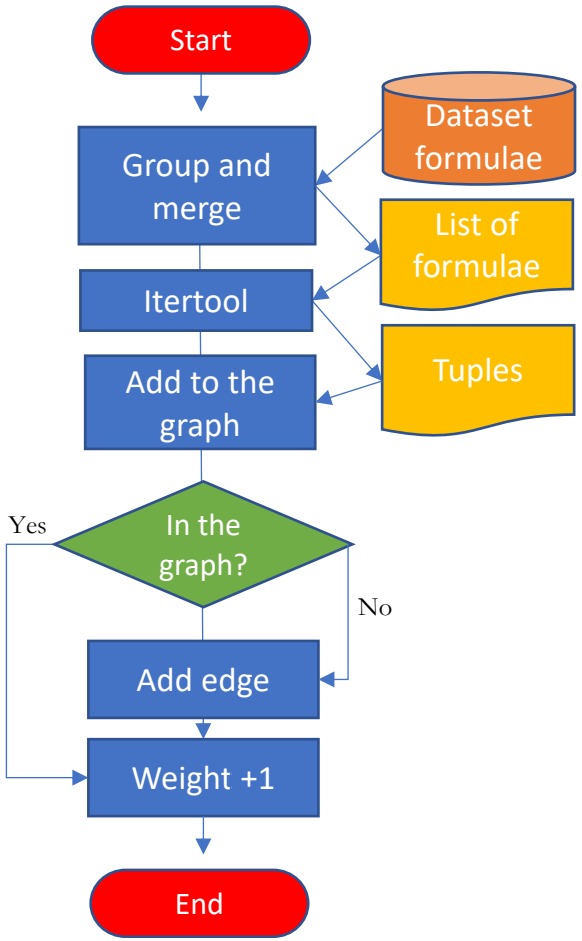


Semantic Network Analysis (SNA)



Undirected and **weighted** network. **Nodes**: 168; **Edges**: 3,866; **Density**: 0.275; **Sum of links**: 536,055; **Average degree**: 23; **Average clustering coefficient**: 0.826.

Semantic Network Analysis (SNA)

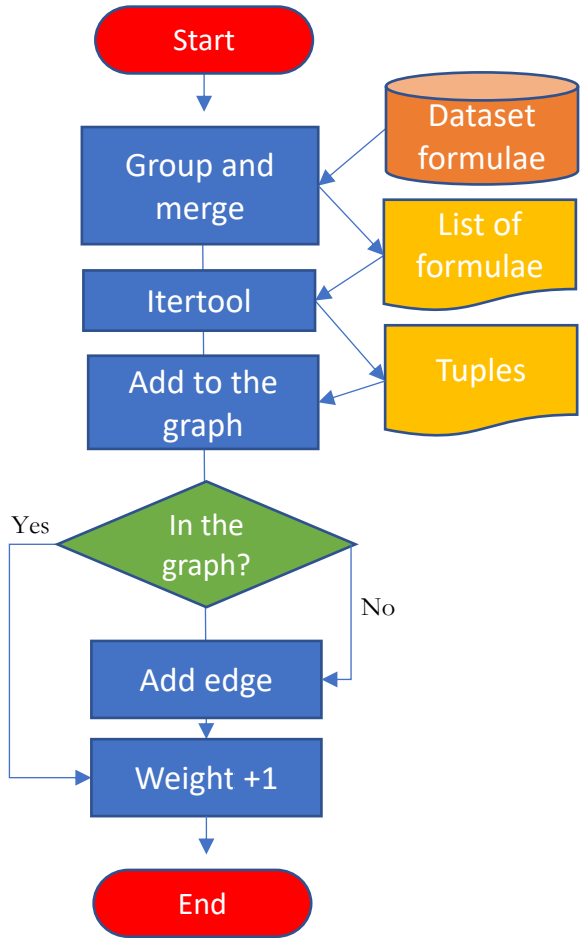


Undirected and **weighted** network. **Nodes**: 168; **Edges**: 3,866; **Density**: 0.275; **Sum of links**: 536,055; **Average degree**: 23; **Average clustering coefficient**: 0.826.

Formula	Degree	Normalized degree
vixit	150	0.8
annos	148	0.8
Dis Manibus	147	0.8
fecit	143	0.8
annis	138	0.8
bene merenti	129	0.7
Dis Manibus sacrum	126	0.7
sibi	125	0.7
fecerunt	124	0.7
annorum	120	0.7

The centrality degree is the number of edges touching a node The degree is normalized by dividing by the maximum possible degree in the graph.

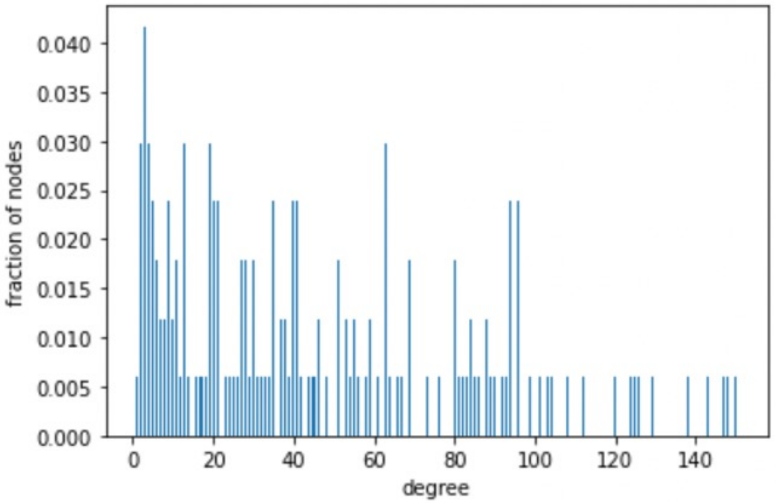
Semantic Network Analysis (SNA)



Undirected and **weighted** network. **Nodes**: 168; **Edges**: 3,866; **Density**: 0.275; **Sum of links**: 536,055; **Average degree**: 23; **Average clustering coefficient**: 0.826.

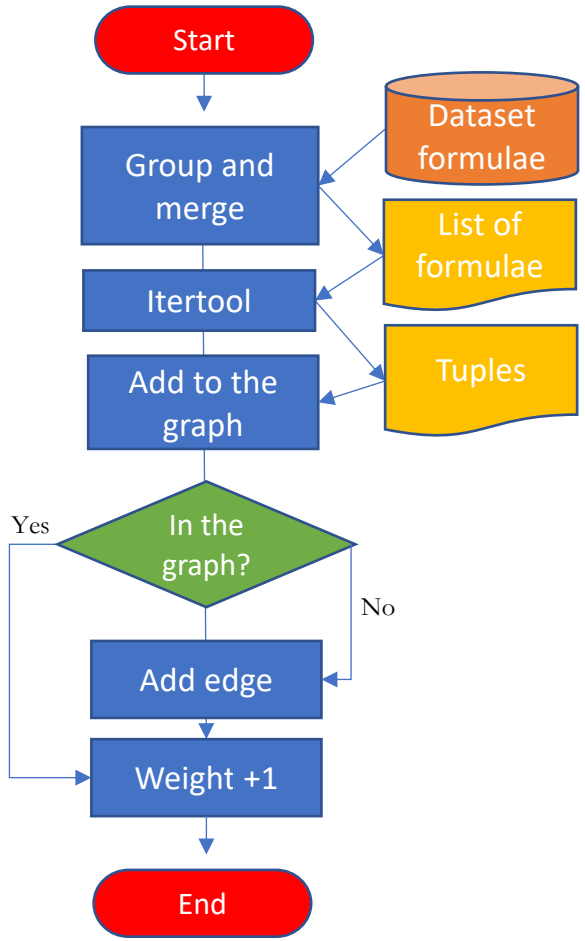
Formula	Degree	Normalized degree
vixit	150	0.8
annos	148	0.8
Dis Manibus	147	0.8
fecit	143	0.8
annis	138	0.8
bene merenti	129	0.7
Dis Manibus sacrum	126	0.7
sibi	125	0.7
fecerunt	124	0.7
annorum	120	0.7

The centrality degree is the number of edges touching a node The degree is normalized by dividing by the maximum possible degree in the graph.



The plot shows the degree distribution in the network. By counting how many nodes have each degree, the degree distribution of a degree (k) is the fraction of the nodes with degree k and the total number of nodes.

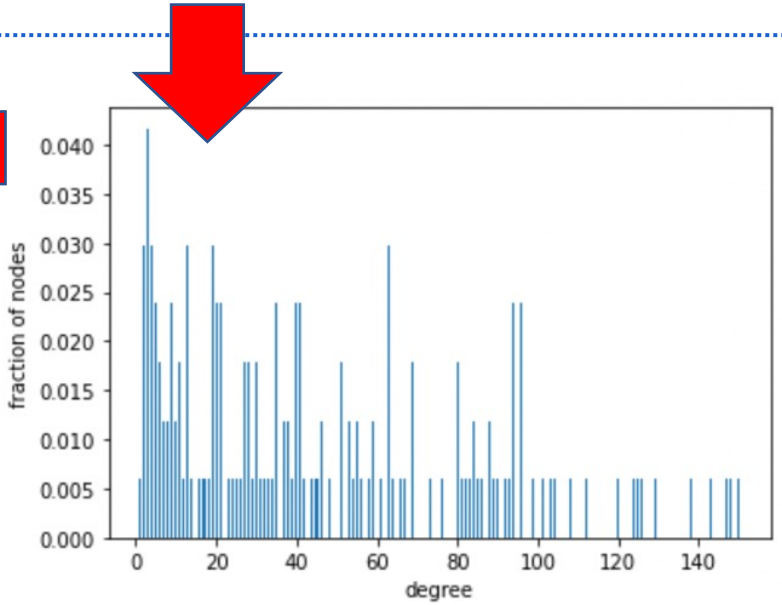
Semantic Network Analysis (SNA)



Undirected and **weighted** network. **Nodes**: 168; **Edges**: 3,866; **Density**: 0.275; **Sum of links**: 536,055; **Average degree**: 23; **Average clustering coefficient**: 0.826.

Formula	Degree	Normalized degree
vixit	150	0.8
annos	148	0.8
Dis Manibus	147	0.8
fecit	143	0.8
annis	138	0.8
bene merenti	129	0.7
Dis Manibus sacrum	126	0.7
sibi	125	0.7
fecerunt	124	0.7
annorum	120	0.7

The centrality degree is the number of edges touching a node The degree is normalized by dividing by the maximum possible degree in the graph.



The plot shows the degree distribution in the network. By counting how many nodes have each degree, the degree distribution of a degree (k) is the fraction of the nodes with degree k and the total number of nodes.

'She did it (for herself)"

CIL III 7408 (Thracia, undated):
[D]is Manibus // L(ucio) Titovio L(uci) l(iberto) Diadu/meno Flavia Vera / coniugi **bene merenti** / et **sibi** et suis **viva fecit**.

'viva': frequency: 691; **degree centrality**: 93; **closeness centrality**: 0.68.

Ego-network

fecit	0.25	annos	0.18	annorum	0.01
Dis Manibus	0.02	vixit	0.01	de suo	0.009
sibi	0.02	Dis Manibus sacrum	0.01	posuit	0.009
bene merenti	0.01	hic situs est	0.006	sit tibi terra levis	0.005
annis	0.01	hic sita est	0.006	ossa	0.003
fecerunt	0.01	militavit	0.005	faciendum curaverunt	0.002
dulcissimo	0.008	pia	0.004	defuncto	0.001

By community detection, 3 clusters were identified. The table shows the nodes with the highest betweenness centrality in each community.

Bibliography

Edmondson, J. "Family relations in Roman Lusitania: Social change in a Roman province?" In M. George (eds.), *The Roman Family in the Empire: Rome, Italy, and Beyond*, 183–229. Oxford. 2015.
Mueller, I. "Single Women in the Roman Funerary Inscriptions." *ZPE* 175 (2010), 295–303.
Saller, R., Shaw, B. "Tombstones and Roman Family Relations in the Principate: Civilians, Soldiers and Slaves." *JRS* 74 (1984), 124–56.

'She did it (for herself)"

CIL III 7408 (Thracia, undated):
[D]is Manibus // L(ucio) Titovio L(uci) l(iberto) Diadu/meno Flavia Vera / coniugi **bene merenti** / et **sibi** et suis **viva** **fecit**.

'viva': frequency: 691; **degree centrality**: 93; **closeness centrality**: 0.68.

Ego-network

fecit	0.25	annos	0.18	annorum	0.01
Dis Manibus	0.02	vixit	0.01	de suo	0.009
sibi	0.02	Dis Manibus sacrum	0.01	posuit	0.009
bene merenti	0.01	hic situs est	0.006	sit tibi terra levis	0.005
annis	0.01	hic sita est	0.006	ossa	0.003
fecerunt	0.01	militavit	0.005	faciendum curaverunt	0.002
dulcissimo	0.008	pia	0.004	defuncto	0.001

By community detection, 3 clusters were identified. The table shows the nodes with the highest betweenness centrality in each community.

Bibliography

Edmondson, J. "Family relations in Roman Lusitania: Social change in a Roman province?" In M. George (eds.), *The Roman Family in the Empire: Rome, Italy, and Beyond*, 183–229. Oxford. 2015.
Mueller, I. "Single Women in the Roman Funerary Inscriptions." *ZPE* 175 (2010), 295–303.
Saller, R., Shaw, B. "Tombstones and Roman Family Relations in the Principate: Civilians, Soldiers and Slaves." *JRS* 74 (1984), 124–56.

Thank you for the attention!