Gender and Power Dynamics in Latin Narratives

Marisa Hudspeth
mhudspeth@umass.edu

Samuel Kovaly
skovaly@umass.edu

Minhwa Lee
minhwalee@umass.edu

Chau Pham
ctpham@umass.edu

1 Introduction

Latin-specific language models are trained on Latin text, so they may learn the gender prejudice embedded in these ancient corpora. It is thus important to understand the extent to which the data is biased before using these models' outcomes for interpretation. In this project, we study gender discrimination in the Latin narrative poem *Metamorphoses* by Ovid. Specifically, we will use word embedding and topic models to study the differences in words and themes associated with each gender. [1] [2]

2 Related work

Many NLP methods which usually target English have been replicated for Latin. The Classical Language Toolkit (10) offers support for common NLP tasks like tokenization, POS tagging, morphology tagging, and NER for classical languages. Researchers have released several pretrained word embedding models, both non-contextual (21; 5) and contextual (2), which have been successfully used for intertextuality analysis (5). Another example of textual analysis done directly on the Latin text is Sprugnoli et al.'s lexicon for sentiment analysis in Latin, which they applied to the tragedy *Medea* by Seneca. The authors were able to track the overall sentiment across storytime as well as the sentiment of each character.

We aim to analyze Latin text through the lens of gender. An ongoing line of research has examined gender representation in English textual data, ranging from greeting card messages (22) to history textbooks (12) and literary novels (6). These works have leveraged Natural Language Processing (NLP) methods like word embeddings and topic modeling to study the differences in words and themes associated with genders. Findings from these works demonstrate that there is indeed a discrepancy in gender depiction, i.e. feminine characters are more often discussed in association with the home (11).

In this project, we specifically employ the use of connotation frames, first developed by Rashkin et al. for sentiment and later extended to power and agency by Sap et al.. Other researchers have also used connotation frames to examine gender, in the context of the #MeToo movement (7), birth stories (1), and Japanese light novels (9). In addition to connotation frames, we leverage Latent Dirichlet Allocation (LDA) developed by Blei et al. to perform topic modeling analysis. Previous research have examined topical diversity of gender tropes in films, televisions and literature (8), as well as topical differences among gender groups in GPT-3 generated stories (11).

We conduct our analyses on two English translations of the same Latin poem. Although there have been no computational analyses of gender between translations that we know of, there has been discourse within the Classics regarding the influence translators have on gender bias in their translations. Within the past few years, the first women translators of Homer's *Odyssey* and Ovid's *Metamorphoses* have discussed how previous translations often downplayed sexual violence and enforced more rigid gender roles than what is actually present in the Latin (North; Mendelsohn).

3 Background and hypotheses

3.1 Background

1. Behavior (B): Latin narratives may reflect the cultural discrimination against women; there

[1] We acknowledge that gender lies on a spectrum, and reducing it to a male-female binary is simplistic; however, our data limits a more complex understanding of gender.

[2] A link to our code can be found here: https://github.com/chtmp223/gender_power

may be differences in how women are portrayed by different translators

2. Environment (E): Latin narrative poem (two English translations)

3. System (S): Gender inference, word embeddings, and topic model.

4. Task (T): Examine the difference in terminology and topics associated with each gender.

3.2 Hypotheses or a research problem

1. What kinds of topic are associated with characters in the text? How do these topics differ among gender groups?

2. What is the distribution of power and agency among characters in the text? Are feminine characters described with more or less power than their masculine counterparts? What is the nature of the relationships among characters of different genders?

4 Data

We examine the Latin narrative poem *Metamorphoses* by Roman author Ovid.[3] This text has a few characteristics that make it suitable for our analyses. Ovid tells the history of the world from creation through the reign of Julius Caesar, weaving together hundreds of myths. Each myth is a contained story, with its own characters and plot independent of the others. This provides an opportunity for comparison across these myths. Several involve the rape or assault of female characters, a topic which has been widely studied within Classics (18; 23) but has not been analyzed from a computational perspective.

The text itself is comprised of about ”250 myths, 15 books, and 11,995 lines.”[4]

In our analyses, we compare two English translations of Ovid's poem, one translated in 2000 by A.S. Kline[5] and an older one translated in 1922 by Brookes More.[6]

[3]http://www.perseus.tufts.edu/hopper/text?doc=Perseus%3Atext%3A1999.02.0029

[4]https://en.wikipedia.org/wiki/Metamorphoses

[5]https://www.poetryintranslation.com/PITBR/Latin/Ovhome.php

[6]http://www.perseus.tufts.edu/hopper/text?doc=Perseus%3Atext%3A1999.02.0028%3Abook%3D1

5 Techniques and methods

5.1 Data Pre-processing

To pre-process Ovid's *Metamorphoses*, we apply standard coreference resolution to both translations using BookNLP (3). This replaces all pronouns with names of the characters they refer to so that we can detect which characters are mentioned in each sentence. BookNLP also determines the inferred gender of each character based on which pronouns a character is most commonly referred by. Additionally, from BookNLP's dependency parse we were able to extract verbs for each character in which they were the agent or semantic theme.

We found BookNLP to perform well at gender inference, but coreference was noisy. We could perform analysis for men and women as a whole for the full text, but any comparison between specific characters would be too inaccurate. BookNLP often identifies more characters than actually exist, meaning we would miss many character references. We decided to rerun the BookNLP pipeline on a few excerpts which involve rape, these being the stories of Alpheus and Arethusa, Apollo and Daphne, Pluto and Proserpina, Hermaphroditus and Salmacis, Jupiter and Callisto, Jupiter and Io, and Tereus and Philomela.[7] Then, after doing a close reading of each excerpt, we mapped the names of important characters to all of the BookNLP character ids that they should be labeled as. This allows us to capture more of the verbs and modifiers actually belonging to these characters.

5.2 Research Problem 1

To answer our first research question, we leveraged Latent Dirichlet Allocation (LDA) (4) for topic modeling analysis. Specifically, we used a Python wrapper [8] around the topic modeling functions of MALLET (13) with default parameters.

LDA models assign probabilities of topics to groups of text (documents). They also assign probabilities of tokens to topics. The myths in the *Metamorphoses* could be used as the documents for LDA; however, the text length of the myths can vary greatly. Instead, we assigned individual sentences as documents. This provided us with a

[7]Most stories we chose were analyzed by van der Merwe

[8]https://github.com/maria-antoniak/little-mallet-wrapper

larger number of documents that has similar size (approximately 25 tokens/document on average).

We ran the topic model with 5, 10, 15, 20, 30, 50, 100 and 200 topics. We saw that topics are most comprehensible and not too saturated at k=10. After closer examination, we also decided to remove stop words such as modal verbs (could, would, should, etc.) and common articles (the, a, an, etc.), which significantly improved the appearance of our topics.

5.3 Research Problem 2

For this question, POS tagging is required in order to get descriptor words (e.g., adjectives or verbs) for characters in the stories. We also needed to perform coreference resolution to extract subject-verb-object triples. To do so, we also used BookNLP (3) on two English translations. Note that we originally planned to do our analysis directly on the Latin text, but we decided to analyze the English translations instead as there is not a coreference resolution system for Latin available.

Once descriptor words are identified for each character, analysis can be applied to these words to search for trends. One such trend is determining the correlation of gender to themes such as power and agency. In English, researchers have used connotation frames to measure the power, agency, and sentiment of characters/people in the context of movies (19), #MeToo stories in news media (7), and birth stories posted to online forums (1).

6 Results of the analysis

6.1 Topic Modeling

We implemented topic modeling to understand the topic distributions for each gender. We ended up with a list of top 10 topics across the entire corpus in Table 1. If we look at the average probability of the topics, we can see that *family* and *body parts* are the most probable topics, which makes sense given that the unifying theme of *Metamorphosis* concerns various physical transformations. If we focus on the top topic terms, we can see that feminine words such as *daughter* appear in the *family* topic, while masculine words like *son* appear in the *kingdom and war* topic, suggesting an interesting difference in the topic distributions for feminine and masculine characters.

We then tried to determine which topics occur in association with a gender group more than in others. Following Lucy and Bamman 11, we examined the average probability of each topic occurring in documents featuring only feminine characters versus only masculine characters. To do so, we first determined the gender of characters in each document using BookNLP's gender inference. Then, for each topic *t*, we calculated $\Delta T(t) = P(t|female) - P(t|male)$, where $P(t|male)$ is the average probability of a topic occurring in documents with mostly masculine characters, and $P(t|female)$ is that with mostly feminine characters. We consider a sentence to feature mostly feminine characters if more than 75% of the characters in the document are feminine characters.

Looking at the result in rightmost column of Table 1, we can that topics like *nature* and *family* show up in documents featuring only feminine characters. On the other hand, topics like *ocean*, *kingdom and war* and *hunting* occur more in documents featuring only masculine characters. This interesting difference correspond to the traditional association between masculinity and fighting and expanding territory, which merits further analysis and validation.

6.2 Word Cloud Analysis

To answer the second research problem, we identified descriptor words for each character and analyzed these words by the gender of the characters by creating word clouds that show the types of these descriptor words used in each gender.

Figure 1 displays a set of word clouds that present the top 50 agency words for each gender that were present in the entire story. While showing similar words (e.g., *say, have, see*), female characters in both translations show a higher frequency of words that represent more agency that impose negative meaning (e.g., *cry, escape, leave, flee, tear, bear*) than male characters. This means that the story characters inferred as women took agency mostly when they were in hardship or life-threatening situations.

Figure 2 presents the word cloud of top 50 possessive words (i.e., words that show who or what something belong to) for each gender in the entire story. We observe that higher frequency of words related to human body parts were shown in female characters in both versions of the translation (e.g., *breast, body, hand, eye, arm, face*), while male characters were described in the story with heroic

Average probability	Top 10 topic terms	Topic label	$\Delta T(t)$
0.25206	gods said take power long many tell let time goddess	Divine communication	-0.002
0.14142	tree white changed wings branches hair form gold leaves water	Trees and transformation	0.008
0.14516	earth air sky fire light sun clouds stars chariot world	Nature	0.009
0.51117	father mother said love daughter words son see girl god	Family	0.021
0.36471	body arms back hand blood hair like hands face spear	Body parts	0.009
0.10436	sacred wine incense altar rites horns gods king temple altars	Sacred rituals	-0.015
0.15505	sea waters waves nymphs river water ocean wind deep land	Ocean	-0.003
0.16991	son city king war father walls achilles also troy people	Kingdom and war	-0.022
0.09292	wild fields fierce birds cattle blood bull sight woods bird	Hunting	-0.011
0.0637	times three old nine four age seven lived black years	Age	0.007

Table 1: Ten most prominent topics in Ovid's Metamorphosis

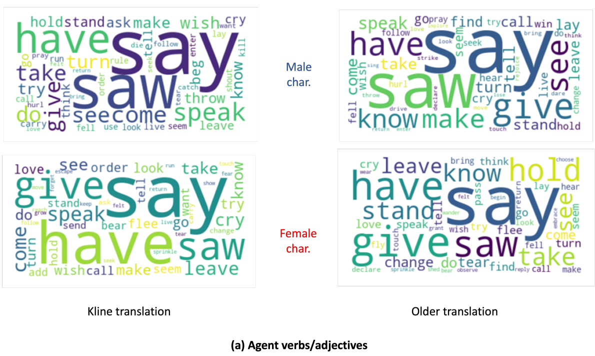


Figure 1: The word cloud for **agency verbs and adjectives** for each gender in the entire story. Left is the results from Kline's translation (modern) and right from older translation.

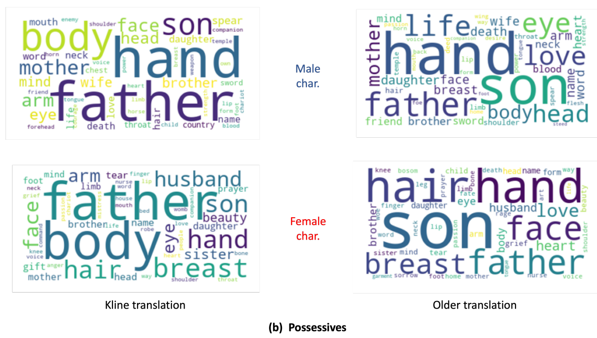
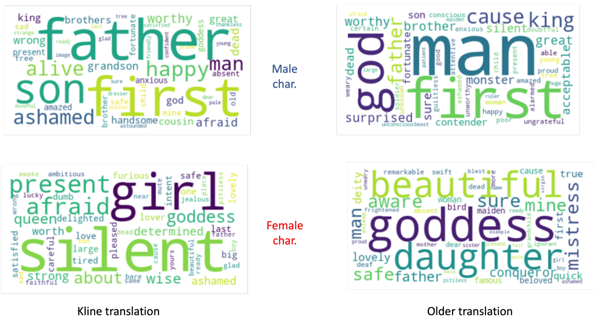


Figure 2: The word cloud for **possessive words** for each gender in the book. Left is the results from Kline's translation (modern) and right from older translation.

or more powerful words (e.g., *sword, blood, country*).

Analyzing the word cloud in Figure 3, we observe several modifiers that contain more agency and power in male characters (e.g., *first, god, king, great, man, happy, alive*) than female characters. However, Kline's translation of the Ovid's Metamorphosis, which is more modern version, used less high-agency and powerful modifiers when describing these male characters (e.g., *ashamed, afraid, anxious*) compared to the older translation. We can assume that modern translation of this Latin narratives relatively de-beautify the male characters who were mostly described as heroes or authorities in older translation.

Another insight from Figure 3 is that the older translation described female characters with higher frequency of modifiers that particularly represent the virtue of beauty and nobility, which were expected by women in traditionalist cultures (e.g., *queen, beautiful, deity, lovely, beloved*). Note that Kline translation also described the female characters in the entire story with some women-specific modifiers (e.g., queen, goddess), but it shows higher frequency of descriptor words that mean more power and agency (e.g., *determined, strong, wise*) or negative emotion (e.g., *silent, afraid, ashamed, dumb, furious*). We can infer that modern translation of this Latin narratives attempted to explain female characters in the story in a more gender-neural way.



(c) Modifiers

Figure 3: The word cloud for **modifier words** for each gender in the entire story. Left is the results from Kline's translation (modern) and right from older translation.

6.3 Power and Agency

We used Sap et al.'s connotation frames to compute power and agency scores for characters. Each verb in the lexicon can have a positive, neutral, or negative impact on the agency of the verb's subject; and can give power to the verb's agent, theme, both, or neither. 101 verb lemmas out of 510 that occurred in our selected stories were missing from this power lexicon, so two of our authors modified the lexicon to include these missing verbs. We attempted to find the nearest synonym already in the lexicon and copy its entries for power and agency.

Since BookNLP's gender inference was reliable, we used it on the full text, which gave us the inferred genders of each character. BookNLP gives us a list of each action/verb that characters took and each action that characters were the recipient of. For each character and each action, we looked up its connotation frame and computed a score. For power, we assigned a score of 1 to power_agent, 0 to power_equal and -1 to power_theme. For actions that characters were the recipient of, these indicator values were negated. For agency, we assigned 1 to agency_pos, 0 to agency_equal and -1 to agency_neg. For each character, these indicators were summed up into a total power or agency over the whole text. After this, we grouped each character by their inferred gender and took a sum over all characters.

In tables 2 and 3, we show the gender aggregated power and agency statistics for both the Kline and More translations. The characters column represents how many characters of each gender appeared in the text. Power mean and agency mean indicate the normalized power and agency scores. Normalization was required since the occurrence count for each gender was not balanced, which gives the majority gender a bias to have a larger power or agency sum.

First, we will note that for both translations, males have a significantly higher power and agency score than females. However, males appear significantly more than females (which is a known bias in literature), as seen in the characters column. Therefore we show the power mean and agency mean values, which are normalized to the total male and female characters. Interestingly, the Kline translation has a higher power mean for females than it does more males. This may be due it being a more modern translation with a more modern view of the text. However, the Kline agency mean, More power mean and More agency mean all have higher scores for males. The they/them characters had very little power and agency. This may be due to females being misclassified as they/them due to a lack of details about them provided in the text, or it may reflect non-human entities that served as obstacles for the main characters to use or overcome.

In extension to power, we also analyzed dominance, which is a more specific trait and thus has potential to have a wider margin between the genders. We used the NRC Valence, Arousal, and Dominance (NRC-VAD) Lexicon (15) where there is a mapping of words to their dominance score, ranging from 1 to -1 continuously. We mapped the modifiers and possessive nouns of each character in the text to it's dominance score as given by the lexicon. We then normalized it by the character count as we did before for power and agency. In tables 2 and 3 we show this result in the dominance mean column. As clearly seen, for both translations, males had more dominance than females, and females more than they/them by a factor of about 2. This is the most significant difference that we have seen thus far for any metric. It seems that male and females characters bot have some level of power and agency, however males are portrayed are much more dominant than females as this metric is not close. Again, the more recent Kline translation has a better score for females than the older More translation, which again

gender	characters	power mean	agency mean	dominance mean
he/him/his	**554**	0.448	**1.01**	**0.418**
she/her	356	**0.511**	0.874	0.231
they/them/their	401	0.282	0.476	0.103

Table 2: Gender statistics for Kline Translation (2000).

gender	characters	power mean	agency mean	dominance mean
he/him/his	**594**	**0.579**	**0.983**	**0.504**
she/her	363	0.441	0.801	0.212
they/them/their	397	0.164	0.355	0.097

Table 3: Gender statistics for More Translation (1922).

shows insight into the translation differences.

Figure 4: Agency and Power scores for Tereus and Philomela story

As a case study, we also examined the power and agency of characters from a single story. We chose the story of Tereus and Philomela because its BookNLP output was the most accurate of all the excerpts we collected. We estimated BookNLP's accuracy through a close reading of the HTML file BookNLP outputs, which allowed us to read the story with all the coreferences highlighted. To further improve accuracy, we then manually edited incorrect coreferences for the key characters. To summarize this story: Tereus, King of Thrace, kidnaps and rapes his wife Procne's sister Philomela. Procne and Philomela enact revenge by killing Tereus' and Procne's son, Ithys, and feeding him to Tereus. Our results are shown in 4. As expected, Philomela and Itys have very little power in both translations. Procne's agency remains about the same, but her power increases in the modern translation. Philomela's agency increases but her power decreases slightly. Tereus, like Procne, has similar agency but increased power in the Kline translation. This could be a result of Kline not "sugarcoating" his translation. Kline actually uses the verb "rape" where More uses "seize."

Because we expect the power and agency of the characters to change across the story, we then divided the text into four main segments (introduction, kidnapping, escape, revenge) and computed power and agency scores for each. Our results can be seen in figure 5. Note that Tereus does not have agency or power scores during the escape segment because he does not appear during that part of the story. There are clear trends for power: Tereus starts as the most powerful, but by the end of the story Procne and Philomela have surpassed him. He and Procne are more powerful in the modern Kline translation, but Philomela is generally less powerful (in the last three segments).

7 Contributions of group members

All members participated in writing the report, with their particular work as described below:

- Marisa: generated and cleaned up BookNLP output, did gender/power/agency analysis for selected stories and characters
- Sam: did gender/power/agency/dominance analysis for full text, aggregated summary statistics based on each gender.

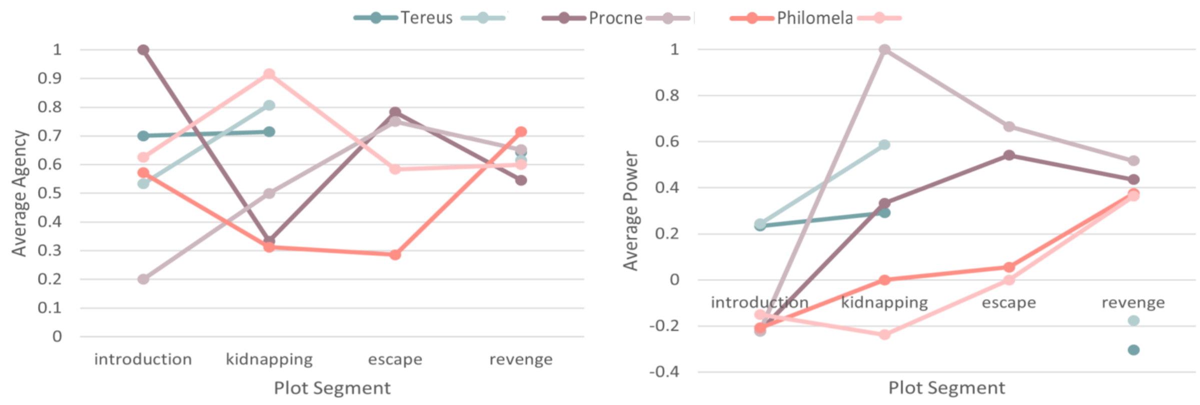


Figure 5: Agency scores (left) and power scores (right) across four segments of the Tereus and Philomela story. Dark shades represent the More translation; light shades represent the Kline translation.

- Minhwa: collected data from the web, conducted word cloud analysis
- Chau: topic modeling code and analysis

8 Conclusion

Our analysis of power and agency was mixed. Although we sometimes saw differences between translations, we cannot say there is a clear trend, with one translation portraying men with more/less power/agency than women. Our case study of the Tereus and Philomela story suggests that there is often more nuance, as characters' power can change over the course of a story.

Our results from topic modeling analysis are aligned with our expectation of topical differences between gender groups: While feminine characters are discussed in the context of family and nature, masculine characters are associated with activities such as hunting, war, and ocean exploration. Future work in this space includes confirming the validity of gender-topic correlation with significance testing.

References

[1] Antoniak, M., Mimno, D., and Levy, K. E. C. (2019). Narrative paths and negotiation of power in birth stories. *Proceedings of the ACM on Human-Computer Interaction*, 3:1 – 27.

[2] Bamman, D. and Burns, P. J. (2020). Latin BERT: A contextual language model for classical philology. *CoRR*, abs/2009.10053.

[3] Bamman, D., Underwood, T., and Smith, N. A. (2014). A Bayesian mixed effects model of literary character. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 370–379, Baltimore, Maryland. Association for Computational Linguistics.

[4] Blei, D. M., Ng, A. Y., and Jordan, M. I. (2003). Latent dirichlet allocation. *J. Mach. Learn. Res.*, 3(null):993–1022.

[5] Burns, P. J., Brofos, J. A., Li, K., Chaudhuri, P., and Dexter, J. P. (2021). Profiling of intertextuality in Latin literature using word embeddings. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 4900–4907, Online. Association for Computational Linguistics.

[6] Chaturvedi, S., Srivastava, S., Daumé, H., and Dyer, C. (2016). Modeling evolving relationships between characters in literary novels. In *AAAI*.

[7] Field, A., Bhat, G., and Tsvetkov, Y. (2019). Contextual affective analysis: A case study of people portrayals in online metoo stories. *Proceedings of the International AAAI Conference on Web and Social Media*, 13(01):158–169.

[8] Gala, D., Khursheed, M. O., Lerner, H., O'Connor, B., and Iyyer, M. (2020). Analyzing gender bias within narrative tropes. In *Proceedings of the Fourth Workshop on Natural Language Processing and Computational Social Science*, pages 212–217, Online. Association for Computational Linguistics.

[9] Gong, X., Lin, Y., Ding, Y., and Klein, L. (2022). Gender and power in japanese light novels. *Proceedings http://ceur-ws. org ISSN*, 1613:0073.

[10] Johnson, K. P., Burns, P. J., Stewart, J., Cook, T., Besnier, C., and Mattingly, W. J. B. (2021). The Classical Language Toolkit: An NLP framework for pre-modern languages. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing: System Demonstrations*, pages 20–29, Online. Association for Computational Linguistics.

[11] Lucy, L. and Bamman, D. (2021). Gender and representation bias in GPT-3 generated stories. In *Proceedings of the Third Workshop on Narrative Understanding*, pages 48–55, Virtual. Association for Computational Linguistics.

[12] Lucy, L., Demszky, D., Bromley, P., and Jurafsky, D. (2020). Content analysis of textbooks via natural language processing: Findings on gender, race, and eth-

nicity in texas u.s. history textbooks. *AERA Open*, 6(3):2332858420940312.

[13] McCallum, A. K. (2002). Mallet: A machine learning for language toolkit. http://mallet.cs.umass.edu.

[Mendelsohn] Mendelsohn, D. Should ovid's metamorphoses have a trigger warning? *The New Yorker*.

[15] Mohammad, S. M. (2018). Obtaining reliable human ratings of valence, arousal, and dominance for 20,000 english words. In *Proceedings of The Annual Conference of the Association for Computational Linguistics (ACL)*, Melbourne, Australia.

[North] North, A. Historically, men translated the odyssey. here's what happened when a woman took the job. *Vox*.

[17] Rashkin, H., Singh, S., and Choi, Y. (2016). Connotation frames: A data-driven investigation. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 311–321, Berlin, Germany. Association for Computational Linguistics.

[18] Richlin, A. (1992). *Pornography and Representation in Greece and Rome*, chapter Reading Ovid's Rapes. Oxford University Press.

[19] Sap, M., Prasettio, M. C., Holtzman, A., Rashkin, H., and Choi, Y. (2017). Connotation frames of power and agency in modern films. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 2329–2334, Copenhagen, Denmark. Association for Computational Linguistics.

[20] Sprugnoli, R., Passarotti, M., Corbetta, D., and Peverelli, A. (2020). Odi et amo. creating, evaluating and extending sentiment lexicons for latin. In *Proceedings of the 12th Language Resources and Evaluation Conference*, pages 3078–3086.

[21] Sprugnoli, R., Passarotti, M., and Moretti, G. (2019). Vir is to moderatus as mulier is to intemperans-lemma embeddings for latin. In *CLiC-it*.

[22] Sun, J., Wu, T., Jiang, Y., Awalegaonkar, R., Lin, X. V., and Yang, D. (2022). Pretty princess vs. successful leader: Gender roles in greeting card messages. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, CHI '22, New York, NY, USA. Association for Computing Machinery.

[23] van der Merwe, C. (2020). Reading ovid in the metoo era: a feminist reception of rape scenes in the metamorphoses. Master's thesis.