

‘Tecnologica cosa’: Modeling Storyteller Personalities in Boccaccio’s *Decameron*

A. Feder Cooper* and Maria Antoniak and Christopher De Sa and David Mimno
Bowers College of Computing and Information Science, Cornell University

Marilyn Migiel
Department of Romance Studies, Cornell University

Abstract

We explore Boccaccio’s *Decameron* to see how digital humanities tools can be used for tasks that have limited data in a language no longer in contemporary use: medieval Italian. We focus our analysis on the question: Do the different storytellers in the text exhibit distinct personalities? To answer this question, we curate and release a dataset based on the authoritative edition of the text. We use supervised classification methods to predict storytellers based on the stories they tell, confirming the difficulty of the task, and demonstrate that topic modeling can extract thematic storyteller “profiles.”

1 Introduction

The *Decameron* is a masterpiece of medieval Italian literature. Completed by 1353, the text is often referred to as “l’umana commedia” (The Human Comedy),¹ a name meant to strike a contrast in subject matter (and a parity in importance) with Dante’s well-known “divina commedia” (Divine Comedy) (Branca, 1975). In a structure similar to Geoffrey Chaucer’s *The Canterbury Tales*, it is a collection of 100 stories (*novelle*) woven together in the context of a frame tale: an honorable brigade (*brigata*) composed of 7 women and 3 men who have fled the ravages of plague in Florence to the relative seclusion of the Tuscan countryside.² The 100 *novelle* are told by the 10 *brigata* members over 10 days, with each day assigned a theme spanning matters of love, wit, and trickery.

While scholarship for *The Canterbury Tales* has engaged with both the stories and the storytellers (Kittredge, 1915; Lawton, 1985; Ginsberg,

2015), storyteller identity has received relatively less attention in the *Decameron*. Instead, literary research has tended to address themes and stories (Migiel, 2004, 2015; Marcus, 1979). Treatment of storyteller identity has thus far been sparse, perhaps due to storyteller personalities seeming generally³ difficult to distinguish at a high level via close reading.⁴

We therefore ask: **Do the members of the *brigata* exhibit distinct storytelling personalities?** We emphasize that this is not a question of authorship, as the text is attributed to Boccaccio alone, but rather one of thematic and stylistic differences among the fictional characters he depicts. To approach this question, we use computational tools to elicit patterns from the text—patterns that may have thus far remained elusive to scholars and could help constitute unique storyteller identities.

This case study highlights several challenges for digital humanities research. While the *Decameron* is a popular and well-studied text, it is written in medieval Italian, for which there are few language modeling resources; this forces us to rely on language-agnostic methods like classification and topic modeling (Section 4). Moreover, while some digitized resources do exist (Brown University Italian Studies Department; Branca, 2003), the text required multiple rounds of curation to be used for a computational study. To facilitate future digital *Decameron* scholarship, we release our user-friendly digital version (Section 3).⁵ In order to build a training corpus for this domain, similar curation

³“Generally” should be taken very generously; we do not intend to eclipse or elide the small yet rich corpus of scholarship that has either directly (Marafioti, 2001; Grossi, 1991) or indirectly (Richardson, 1978) discussed storytellers.

⁴This is debatable, but perhaps true in comparison to *The Canterbury Tales*. For over 100 years scholars have investigated pilgrim personalities. See Kittredge (1915), at p. 155, “The Pilgrims do not exist for the sake of the stories, but *vice versa*. ... [T]he stories are merely long speeches expressing, directly or indirectly, the characters of the several persons.”

⁵<https://github.com/pasta41/decameron>

*Corresponding author; afc78@cornell.edu

¹The first words of the Proem are “Umana cosa,” which roughly translate to “It is human” or “human quality” (Boccaccio, 1995); they immediately underscore the secular focus.

²The description of the chaos inflicted by plague in Florence has led to renewed international interest in the text (Findlen, 2020; Marcus, 2020; Prime et al., 2020).

will be necessary for other digitized medieval Italian texts, including additional works by Boccaccio and authors such as Dante and Petrarch.

Taken together, our classification and topic modeling results support existing humanist scholarship concerning storyteller identity and suggest new questions for further inquiry. More broadly, our work here serves as preliminary evidence that such tools can be useful for highly specialized academic digital humanities work—in non-English and non-standard (e.g., bygone language variant) domains.

2 Related Work

As one progresses through the 10 days of the *Decameron*, different members of the *brigata* seem to develop distinct storytelling personalities. Dioneo frequently tells bawdy tales, pushing the bounds of decorum. Emilia seems like she does not quite fit in with the rest of the group, and in fact may be (though we cannot be certain) an actual political outsider—the sole Ghibelline in the group of Guelphs (Richardson, 1978). Lauretta can perhaps be cast as a “bearer of bad news,” according to Marafioti (2001). Notwithstanding such stand-alone examples, scholarship has not addressed whether each of the 10 *brigata* members have clearly identifiable storytelling personalities. As there is no literary consensus, we apply computational tools to extract patterns that might be difficult for human readers to elicit.

Prior work in digital humanities has studied a variety of narrative questions across corpora of multiple texts. For example, research in cultural analytics has compared narrative structure (Chambers and Jurafsky, 2009; Pichotta and Mooney, 2016; Goyal et al., 2010), character arcs and relationships (Bamman et al., 2013; Iyyer et al., 2016), and authorship attribution (Hoover, 2004). Authorship can also be modeled as a latent factor in topic models (Rosen-Zvi et al., 2004). We do not explicitly model authors (or in our case, narrators) but instead rely on a simpler model to extract cross-cutting, interpretative themes. Moreover, unlike studies of focalization (Genette, 1983), we make no attempt to model the perspective or views of a character, but rather simply ask if characters are in any way distinguishable.

In computational studies that similarly focus on sections of a single work, Wang and Iyyer (2019) compare sections of Italo Calvino’s *Invisible Cities* and Brooke et al. (2015) investigate distinguish-

ing narrative voices in T.S. Eliot’s *The Waste Land*. Wang and Iyyer (2019) circumvent data size limitations by relying on large, pretrained contextual models to cluster the cities and compare thematic patterns; Brooke et al. (2015) rely on preexisting tools that elicit English-language features, including parts of speech and verb tense. Such pretrained models and featurization tools, while available for modern Italian (Polignano et al., 2019), are unavailable for the quite different medieval Italian of the *Decameron* (Salvi and Renzi, 2010; Dardano, 2012). We instead use language-agnostic computational tools for our experiments, which come with the added benefit of interpretability (Section 4).

3 Curating a *Decameron* Dataset

We constructed a `json` dataset of the *Decameron* from an XML version hosted online by the Sapienza University of Rome (Branca, 2003). This digitized version is based on Vittore Branca’s authoritative text (Boccaccio, 2014), and was published online in 2003 in the (at the time standard) TEI P4 format (Text Encoding Initiative, 2002). TEI P4 contains a variety of metadata that interrupt contiguous portions of Boccaccio’s text, which does not make the format amenable to commonly-used tools. We therefore spent considerable time simplifying this format to be more easily manipulable for modern data analysis. We manually and repeatedly verified that our curation process retained the integrity of the text.⁶ Where appropriate, we added metadata to annotate *novelle*, such as the *novella* storyteller, which was absent in the existing online version. Unlike the TEI P4 format, we avoid placing these metadata within the text of individual *novelle*, and provide scripts for those that wish to remove these metadata for their analyses.

We release this dataset publicly. Our hope is that our online version of Branca’s text will be more accessible to scholars of medieval Italian interested in engaging with digital tools, as the simplicity of our format should lower the barrier to entry for both computational and humanist scholars interested in the *Decameron*.

4 Case Study: Constructing Storyteller Profiles in Boccaccio’s *Decameron*

We use the problem of *Decameron* storyteller identity as a case study for exploring the challenges

⁶We document the process in our repository README.

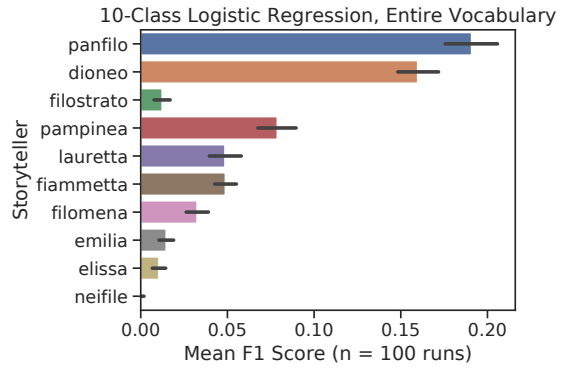
and opportunities for using digital humanities techniques in specialized literary domains. In particular, we investigate how such tools can be useful for 1) a small corpus containing a single text 2) modeling language that is no longer in contemporary use. While the question we ask is specific to our chosen domain—the *Decameron* and medieval Italian—we believe that the lessons we can derive are applicable to other scholarly digital humanities tasks with these same defining elements.

4.1 Problem Formulation

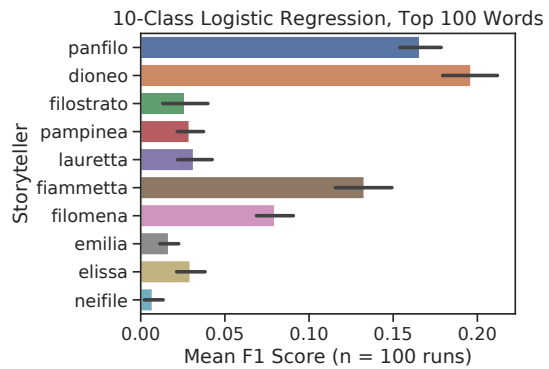
Wang and Iyyer (2019) were able to use pretrained contextual models like BERT (Devlin et al., 2019) for modeling a small, single-text English corpus, analogous tools are not available for studying the *Decameron*. While there is a modern Italian version of BERT (ALBERTO, trained on Twitter data (Polignano et al., 2019)), medieval Italian orthography and morphology are sufficiently different to contraindicate its use (Salvi and Renzi, 2010; Dardano, 2012). Moreover, such pretrained contextual ML models are difficult to interpret, and our goal is to assist humanist scholars in close-reading analysis. Learning about storyteller identity is not just about classification; we already know authoritatively who told which story. Rather, we would like to explain *how* our models distinguish among storytellers. Based on this goal, and the constraints we highlight in Section 4, we choose two language-agnostic, interpretable approaches: logistic regression to try to classify storytellers based on their *novelle* and topic modeling to model storytellers as distributions of lexical themes.

4.2 Modeling Storytellers using Logistic Regression

We first attempt to see if the storytellers can be identified from the *novelle* they tell. We train a logistic regression model for this classification task. For our training data, we divide each *novella* into 100-word chunks (converted to TF-IDF vectors) with the corresponding storyteller as the label. This results in a 10-class logistic regression problem, using an 80/20 train/test split where we ensure that we have equal representation of each storyteller in both sets (i.e., each storyteller tells 10 *novelle*, comprising the 100 *novelle* total; we randomly sample 8 *novelle* for each storyteller in train; the remaining 2 for each in test). We train our model 100 times, with variation coming from randomly sampling the *novelle*. Results are shown in Figure 1a.



(a)



(b)

Figure 1: Mean F1 scores for classifying *novelle* by narrator in 10-class logistic regression. For both experiments, $n = 100$. Using the *Decameron*'s whole vocabulary (1a), the model can identify Panfilo and Dioneo better than random. When we restrict the text to only contain instances of the 100 most frequent words (1b), the model is additionally able to identify Fiammetta.

Since there are 10 storytellers, in order to classify better-than-random, F1 scores would need to be > 0.1 . There are only two storytellers, who are both men, that pass this threshold consistently: Panfilo and Dioneo. It is perhaps unsurprising that this is true for Dioneo; he alone among the *brigata* has the special privilege of deviating from the Day's storytelling theme—a privilege he typically exercises to talk about sex. It is however less clear to us why Panfilo stands out in our results, which suggests a potential direction for future research.

We then re-ran this experiment, pre-processing the *Decameron* to only contain the 100 most frequently used words in the vocabulary. In addition to Panfilo and Dioneo, this model is also able to identify Fiammetta, one of the seven women, better than random (Figure 1b). This, too, suggests lines of further investigation, as it is unclear why Fiammetta

	Panfilo	Nefile	Filomena	Dioneo	Fiammetta	Emilia	Filostrato	Lauretta	Elissa	Pampinea
Highest	gentili compagni bocca figliuol vicina	veder onestà gentil credeva vicini	amico parenti ciascun cautamente dico	famigliare conoscere primieramente signor pose	meco cuore amava nell'animo venendo	basciò vivo accidente maravigliò buone	fama caldo cavallo oimè malvagia	gentile messere talvolta diè belle	figliuolo cavalli liberamente figliuoli veggendosi	freddo torre fante amante reina
Lowest	freddo buone figliuolo aperto pianamente	peccato disidero occhi cavaliere troppo	famigliare sentito porta tavola corte	valente morte corte madre cavaliere	cavallo tornare signor signore figliuoli	vedi uom famiglia partito cara	re speranza gentili figliuoli amor	incontanente animo medesima vedendo fante	dormire allato mille ciascun giovani	veramente ricco tavola bisogno morto

Table 1: The words with highest and lowest PMI for each storyteller (higher scores indicate stronger associations).

is more identifiable than the other women.⁷

We probe our classification results by extracting the words with highest and lowest pointwise mutual information (PMI) for each *brigata* member. This metric uncovers lexical associations with each narrator in comparison to all the narrators. Given a word w and a narrator n , $PMI(w; n) = \log \frac{p(w|n)}{p(w)}$. To improve interpretability, we remove words that occur fewer than five times for each narrator, and we manually remove stopwords.⁸ Table 1 shows that despite our low classifier performance, lexical differences between the storytellers are interpretable. Nefile, whom our classifier completely misses, scores high for words that signify honorability (e.g., *onestà*), while low for words that conote the opposite (e.g., *peccato*). Filostrato, who reveals his personal heartbreak, scores low for words concerning love and hope (e.g., *amor*, *speranza*).⁹

4.3 Modeling Storytellers using Topic Distributions

Since we were not able to generally distinguish storytellers via classification, as a second experiment we use latent Dirichlet allocation (LDA) (Blei et al., 2003) to model each *novella* as a distribution of topics. We group the results by storyteller to see if the distributions of *novella* topics are distinguishable for each of the 10 members of the *brigata*. In other words, we can view per-storyteller topic distributions as storyteller “profiles”—patterns that may indicate unique thematic features of particular *brigata* members.

⁷We repeated these experiments using one-versus-rest logistic regression to test if each storyteller is distinguishable compared to the other 9. The results were comparable to those presented in Figure 1.

⁸[è, che, la, quale, e, di, fu, le, per, col, aveva, avere, ha, il, lo, gli, i, de, in, cìd, ho]

⁹Notably, PMI scores do not incorporate semantics. As a result, two words that have the same semantic meaning but different morphology can have very different scores. For example, Panfilo’s *novelle* have high PMI for *figliuol* but low PMI for *figliuolo*—words that both mean “son” (in the medieval variant of the modern “figliolo”).

To perform this analysis, we used a Python wrapper for MALLET (Antoniak, 2021; McCallum, 2002). We used this framework because its implementation of LDA uses Gibbs sampling (Geman and Geman, 1984), an exact MCMC sampling method that has popularly been observed to have better performance for small datasets than inexact, variational inference-based implementations. We train our model with $k = 20$ topics¹⁰ and allow hyperparameter optimization. Before training, we lowercase the text and process each *novella* to create documents of 200 words each; we remove a custom list of common Italian stop words, and if the resulting document is fewer than 20 words long, we do not use it for training. This creates a training corpus of 1,203 documents.

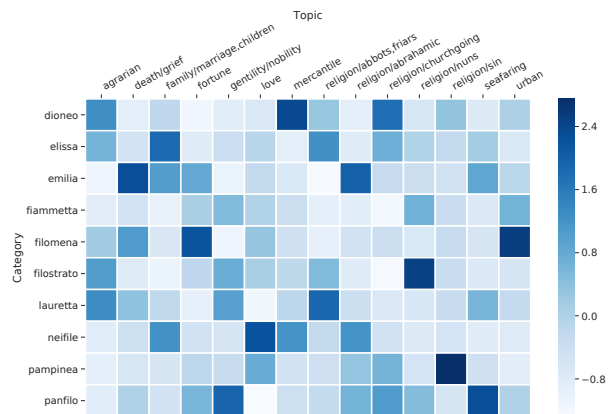


Figure 2: Each storyteller according to their underlying distribution of topics. Columns are normalized to highlight differences across topics.

We manually validated the quality of the resulting topics to see if they were semantically meaningful, and we were able to determine some clear themes. For example, one topic’s top words include *nave* (ship), *mare* (sea), *isola* (island), *barca*

¹⁰We tried different k and found that 20 resulted in the most interpretable, overarching topics for our small dataset.

(boat), and *vento* (wind), to which we assigned overall topic designation *seafaring*. Of the 20 topics, 14 had very clear semantic themes, while the remaining 6 were more illusive. Therefore, to achieve a clearer picture of the variation over storytellers, we removed these 6 topics in our plots. We then validated the remaining topics at the *novella* level, plotting a heat map in which each row is a *novella* topic distribution. This heat map enabled us to spot-check if particular *novelle* had reasonable topic distributions.¹¹

Our storyteller-topic results are summarized in Figure 2, which overall indicates that there are thematic differences between the individual storytellers. If the storytellers were truly indistinguishable, it is unlikely that we would observe variation in the topic signatures. Of particular note are cells in the heatmap that show a uniquely highly-weighted presence for a topic that is relatively absent for each of the other nine narrators. To call attention to three examples, we can see this for *religion/sin* for Pampinea, *mercantile* for Dioneo, and *seafaring* for Panfilo.

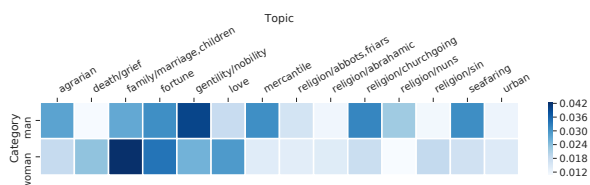


Figure 3: Topic distributions by storyteller gender.

To see another view of these results, we performed a similar analysis, in which we grouped topic distributions more coarsely—by storyteller gender instead of individual storyteller (Figure 3). Two interesting observations for these results are that the men discuss *mercantile* themes considerably more than the women, and the women discuss *love* more than the men. Perhaps the *mercantile* results are unsurprising, given men’s unrestricted ability to participate in economic endeavors—a privilege underscored in the Author’s Proem (Boccaccio, 2014). However, the result concerning *love* is somewhat surprising. The Author’s stated purpose in the Proem is to relieve the suffering of women in love, and the three men are said to be in love with three of the women of the *brigata*, so it may seem unusual for words

¹¹This heatmap is available at <https://github.com/pasta41/decameron>.

associated with love to be more strongly collocated with women than with men.

5 Conclusion and Future Work

While our work has focused on a specific question—whether the members of the *Decameron*’s *brigata* exhibit distinct storytelling personalities—we have illustrated broader lessons for small-text, specialized-language digital humanities scholarship. A central tension for work in low-resource domains is whether to focus on building tools and resources to mimic large, English-language resources or to instead work around the lack of resources by relying on methods that do not require much training data. While we have taken the latter path in this paper, we see ample opportunities for both developing models and annotating larger datasets for this domain (Bai et al., 2021). For example, while medieval Italian is syntactically quite different from modern Italian, some linguistic studies on specific texts indicate significant lexical overlap.¹² Based on this observation, future work could modify existing Italian contextual models for high-fidelity use on medieval Italian works.

For scholars of the *Decameron*, our highlighted results indicate areas for further inquiry. For example, a close-reading analysis of the *novelle* could explain when and why the women storytellers talk about more “male” topics (e.g., *mercantile* themes) and would complement our topic modeling results. More broadly, our release of a simplified format of the digitized text will facilitate future digital humanities research related to Boccaccio’s *Decameron*.

References

- Maria Antoniak. 2021. [little-mallet-wrapper](#).
- Fan Bai, Alan Ritter, and Wei Xu. 2021. Pre-train or annotate? domain adaptation with a constrained budget. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics.
- David Bamman, Brendan O’Connor, and Noah A. Smith. 2013. [Learning latent personas of film characters](#). In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 352–361, Sofia, Bulgaria. Association for Computational Linguistics.

¹²De Mauro (2001) points out that if we examine the fundamental vocabulary of Italian (i.e., the most common 2000 words) we find that 92% of them are words that Dante used in his *Divina Commedia*.

- David M. Blei, Andrew Y. Ng, and Michael I. Jordan. 2003. [Latent Dirichlet Allocation](#). *J. Mach. Learn. Res.*, 3:993–1022.
- Giovanni Boccaccio. 1995. *Decameron*, 2 edition. Penguin Books, London, England.
- Giovanni Boccaccio. 2014. *Decameron*, edited by Vittore Branca. Einaudi, Torino.
- Vittore Branca. 1975. *Boccaccio medievale*, 4 edition. G. C. Sansoni, Firenze.
- Vittore Branca. 2003. [The Decameron](#). Digitized by the Sapienza University of Rome, Biblioteca italiana Project.
- Julian Brooke, Adam Hammond, and Graeme Hirst. 2015. [Distinguishing Voices in The Waste Land using Computational Stylistics](#). In *Linguistic Issues in Language Technology, Volume 12, 2015 - Literature Lifts up Computational Linguistics*. CSLI Publications.
- Brown University Italian Studies Department. [Decameron Web](#).
- Nathanael Chambers and Dan Jurafsky. 2009. [Unsupervised learning of narrative schemas and their participants](#). In *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP*, pages 602–610, Suntec, Singapore. Association for Computational Linguistics.
- Maurizio Dardano, editor. 2012. *Sintassi dell'italiano antico*. Carocci, Roma. Vol. 1, La prosa del Duecento e del Trecento.
- Tullio De Mauro. 2001. Dante, il gendarme e l'articolo 3 della Costituzione. In *Dante, il gendarme e la bolletta. La comunicazione pubblica in Italia e la nuova bolletta*, pages 3–11. Laterza, Bari.
- Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. [BERT: Pre-training of deep bidirectional transformers for language understanding](#). In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.
- Paula Findlen. 2020. [What Would Boccaccio Say About COVID-19?](#) *The Boston Review*.
- Stuart Geman and Donald Geman. 1984. [Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images](#). *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-6(6):721–741.
- G rard Genette. 1983. *Narrative Discourse: An Essay in Method*. Cornell University Press, Ithaca, New York. Translation (Jane E. Lewin) of Discours du r cit, a portion of the 3rd vol. of the author's Figures, essais.
- Warren Ginsberg. 2015. *Tellers, tales, and translation in Chaucer's Canterbury Tales*. Oxford University Press, Oxford, United Kingdom; New York, NY.
- Amit Goyal, Ellen Riloff, and Hal Daum  III. 2010. [Automatically producing plot unit representations for narrative text](#). In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, pages 77–86, Cambridge, MA. Association for Computational Linguistics.
- Paolo Grossi. 1991. Per una rivalutazione dei narratori del Decameron: Filomena e la novella di Lisabetta (Decameron IV, 5). *Critica letteraria*, 19:145–57.
- David L Hoover. 2004. Testing Burrows's delta. *Literary and Linguistic Computing*, 19(4):453–475.
- Mohit Iyyer, Anupam Guha, Snigdha Chaturvedi, Jordan Boyd-Graber, and Hal Daum  III. 2016. [Feuding families and former Friends: Unsupervised learning for dynamic fictional relationships](#). In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1534–1544, San Diego, California. Association for Computational Linguistics.
- George Lyman Kittredge. 1915. *Chaucer and His Poetry*. Harvard University Press, Cambridge, Massachusetts.
- David Lawton. 1985. *Chaucer's Narrators*. D.S. Brewer, Suffolk, UK.
- Martin Marafioti. 2001. [Boccaccio's Lauretta: The Brigata's Bearer of Bad News](#). *Italian Culture*, 19(2):7–18.
- Millicent Marcus. 1979. *An Allegory of Form: Literary Self-consciousness in the Decameron*. Anma Libri.
- Millicent Marcus. 2020. [Reading the 'Decameron' Through the Lens of COVID-19: The Fallacy of Literary Distancing](#). *The Yale Review*.
- Andrew Kachites McCallum. 2002. [MALLET: A Machine Learning for Language Toolkit](#).
- Marilyn Migiel. 2004. *A Rhetoric of the Decameron*. The University of Toronto Press.
- Marilyn Migiel. 2015. *The Ethical Dimension of the 'Decameron'*. The University of Toronto Press.
- Karl Pichotta and Raymond Mooney. 2016. [Learning statistical scripts with lstm recurrent neural networks](#). *Proceedings of the AAAI Conference on Artificial Intelligence*, 30(1).

- Marco Polignano, Pierpaolo Basile, Marco de Gemmis, Giovanni Semeraro, and Valerio Basile. 2019. [AIBERTO: Italian BERT Language Understanding Model for NLP Challenging Tasks Based on Tweets](#). In *Proceedings of the Sixth Italian Conference on Computational Linguistics (CLiC-it 2019)*, volume 2481. CEUR.
- Kelly Prime, Mike Benoist, Tommy Orange, and Edgewise Danticat. 2020. [The Sunday Read: 'The Decameron Project'](#).
- Brian Richardson. 1978. The 'Ghibelline' Narrator in the Decameron". *Italian Studies*, 33:20 – 28.
- Michal Rosen-Zvi, Thomas Griffiths, Mark Steyvers, and Padhraic Smyth. 2004. The Author-Topic Model for Authors and Documents. In *Proceedings of the 20th Conference on Uncertainty in Artificial Intelligence, UAI '04*, page 487–494, Arlington, Virginia, USA. AUAI Press.
- Giampaolo Salvi and Lorenzo Renzi. 2010. *Grammatica dell'italiano antico*. Il Mulino, Bologna.
- Text Encoding Initiative. 2002. [The XML Version of the TEI Guidelines: <TEI.2>](#).
- Shufan Wang and Mohit Iyyer. 2019. [Casting Light on Invisible Cities: Computationally Engaging with Literary Criticism](#). In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 1291–1297, Minneapolis, Minnesota. Association for Computational Linguistics.