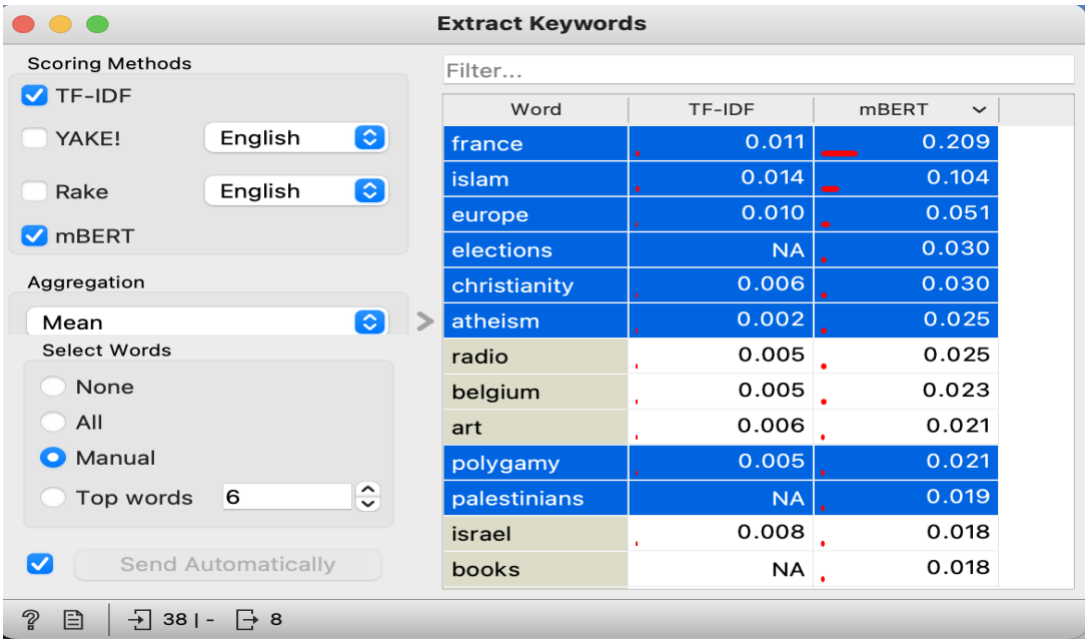
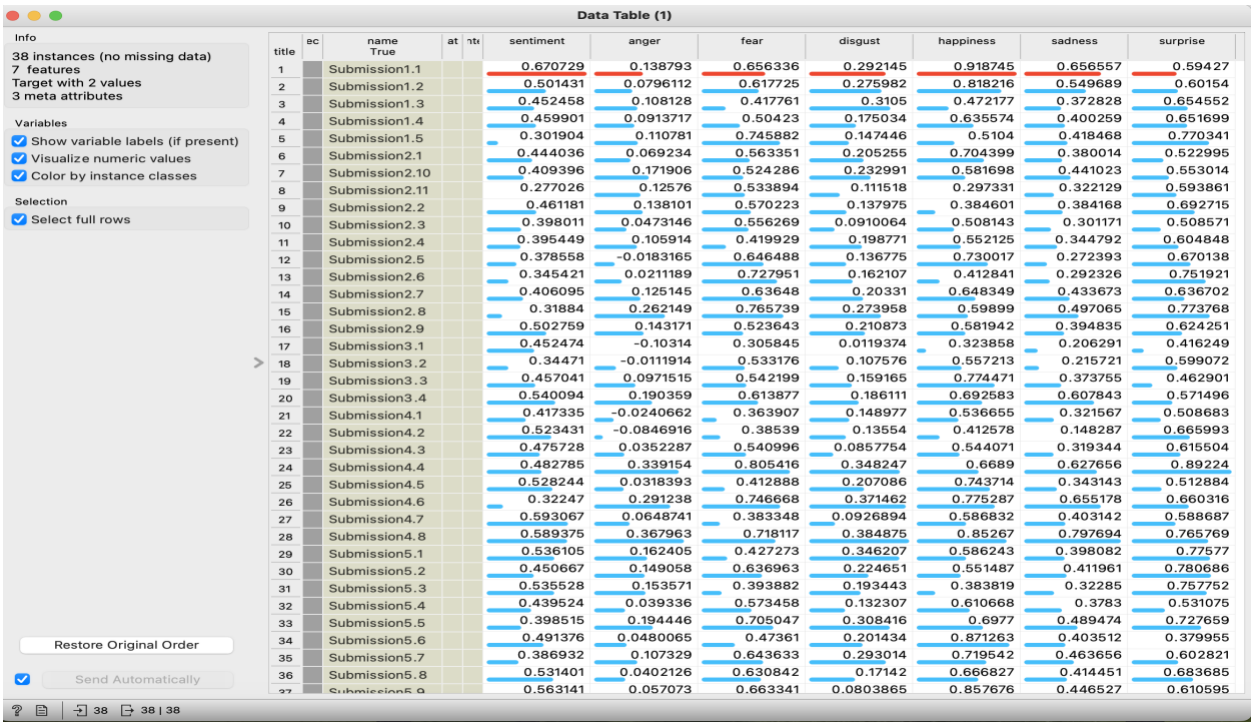
Rhodes Doctoral Fellowship in the Computational Humanities: Completion Report

Initial Research: As a comparative literature scholar working on refugee and migration discourses, I became interested in the increasing popularity of literature that expressed anxieties about migrants from the global South perceived as replacing White populations ("The Great Replacement Theory"). These discourses about the decline of 'Western' values and way of life were based on the cultural construct of 'the West' which was a rhetorical invention going back to the end of the 19th century (Christopher Gogwilt, *The Rhetorical Invention of the West,* 9). While close readings of dystopian novels such as Jean Raspail's *The Camp of Saints* (1973) demonstrated that these texts represented non-White migratory populations in racist terms, what were the rhetorical patterns and structural elements that could provide empirical support for my thesis?

In this project, I employed NLP techniques like topic modeling (LDA) and Multidimensional Scaling (MDS) to identify and visualize the overall topic distribution in the corpus, and sentiment analysis to quantitatively determine affective trends in these texts. For sentiment analysis, apart from tools like VADER (Valence Aware Dictionary and sEntiment Reasoner) and FLAIR, I used SentiArt which is a compact, linguistically grounded sentiment-analysis tool that combines distributional semantics with emotion-annotated lexica to produce fine-grained, interpretable affect scores for texts. By applying these techniques on Raspail's novel and Michel Houellebecq's *Submission* (2015), I discovered that while the plot in these stories was action-driven, the primary perspective in these novels was subjective. Even when describing apocalyptic events, the semantic heft of the narration rested in the White subjects' inner reaction to the situation. The texts primarily described the decline of 'Western' values; the dystopia emerged as a *consequence* of that primary Fall. These initial findings suggested that the genre of invasion novel was not qualitatively different from the narratives depicting the decline of the West; instead, the two represented different sides of the same coin.

I presented this research and findings at the Open Scholarship in the Humanities event organized by the Duke University Libraries and the Digital Humanities and Summer Institute Conference (DHSI) at the University of Victoria (Canada) in June 2024. Figure 1 demonstrates the keywords from the novel *Submission* extracted using different models such as TF-IDF and mBERT. Figure 2 displays the sentiment analysis score of different sections of the novel in *Orange*.





Pedagogical Project: Building on this research, I proposed a pedagogical project on this theme for the Rhodes Fellowship in the Computational Humanities. I designed and taught an interdisciplinary course titled “Enemy at the Gates: Reading Dystopian Literature” in Fall 2024. The COVID-19 pandemic had ignited popular and literary imagination about widespread chaos and destruction. The problems of social and economic inequalities, environmental catastrophes, wars, and migration exacerbated contemporary anxieties about the present and the future. In this course, I wanted the class to collectively interrogate whether the novelistic imagination of foreign invasions and dystopian scenarios was merely harmless amusement or weaponized

nostalgia. How did the rhetoric of decline and dystopia affect our attitudes towards "outsiders"- refugees, foreigners, strangers? What were the various tropes of dystopias as presented in these fictional works? We set out to research these questions through the use of distant reading and computational methods, supplementing traditional close readings.

FALL 2024

ENG 290S

CROSS LISTED **LIT/ICS 290**
(ALP, CZ)
W/F 4.40-5.55

ENEMY AT THE GATES: READING DYSTOPIAN LITERATURE

Novels imagining a bleak future for the US or the West are not plain entertainment. **Each fantasy of the end of the West, like the collapse of the US economy or an invasion from the East, either questions or privileges the existing world order.**

In this course, we will read novels that imagine stories of foreign invasions and dystopian scenarios. Are such novels harmless amusement or weaponized nostalgia? **How does imagining a bleak future affect our attitudes towards "outsiders"- refugees, foreigners, strangers?** We will answer these questions by closely reading novels like Lionel Shriver's *The Mandibles,* Michel Houellebecq's *Submission*, and films like *The End of America* (2008). We will also read works that complicate our understanding of insiders and outsiders like Toni Morrison's *Paradise* (Nobel prize 1993) and J.M. Coetzee's *Waiting for the Barbarians* (Nobel prize 2003). Our readings will be grounded in the theoretical fields of postcolonial, race, and migration studies.

Our methods will combine close reading of these texts with computational methods to analyze popular discourse about immigrants and the construct of 'the West'. What can we discover using computer software that can read thousands of pages at once? During the semester we will build up our tools from scratch and practice new techniques like sentiment analysis, topic modeling, and Multi-dimensional Scaling (MDS) in class. The course will culminate in a final project where we bring together our interpretations of these texts and results produced using **Natural language processing (NLP)** methods.

Class assessment will include 1. Weekly reading responses and peer comments (250 words) 2. One mid-term paper (4-5 pages), and 3. a final digital project (w/reflection) that students will present in class. **No exams! No coding experience required!**

nitin.luthra@duke.edu

The semester-long class introduced students to core methods in text analysis, topic modeling, and data visualization using various digital and computational tools such as *Voyant*, *Orange Data Mining*, *ArcGIS Storymaps*, and *Python*. The class combined weekly readings and seminar discussion of literary texts and theory with hands-on lab sessions and two projects in which students used computational techniques to identify patterns and connections in novels that might not be easily apparent in close readings. In their research, students delved into questions of morality, justice, economy, colonialism, violence, and subjectivity – all through computationally analyzing the textual elements.

Students produced several illuminating discoveries in their projects. For instance, tools such as collocation, concordance, and sentiment analysis helped in identifying Sinophobic language in novels like the Australian writer Heather Rose's *Bruny* (2019). The language in the novel associated the word "Chinese" with a government that is authoritarian (as opposed to representative). Moreover, the text identified Chinese laborers less in humanized terms as "people" and more as "workers" without subjectivity or names.

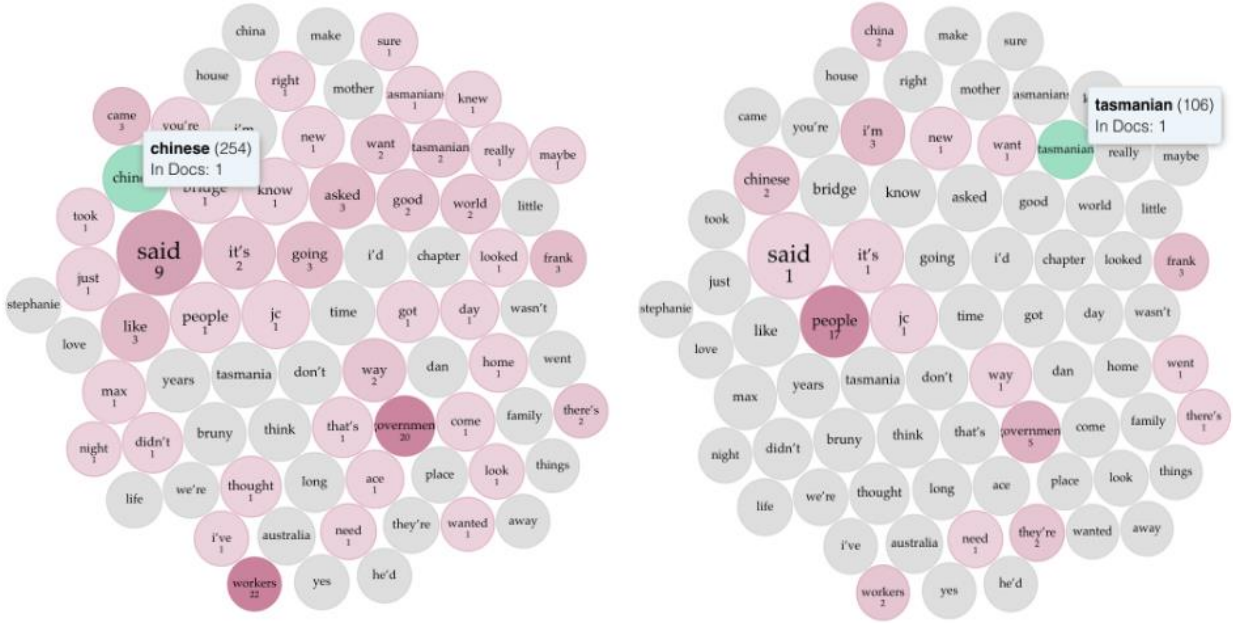
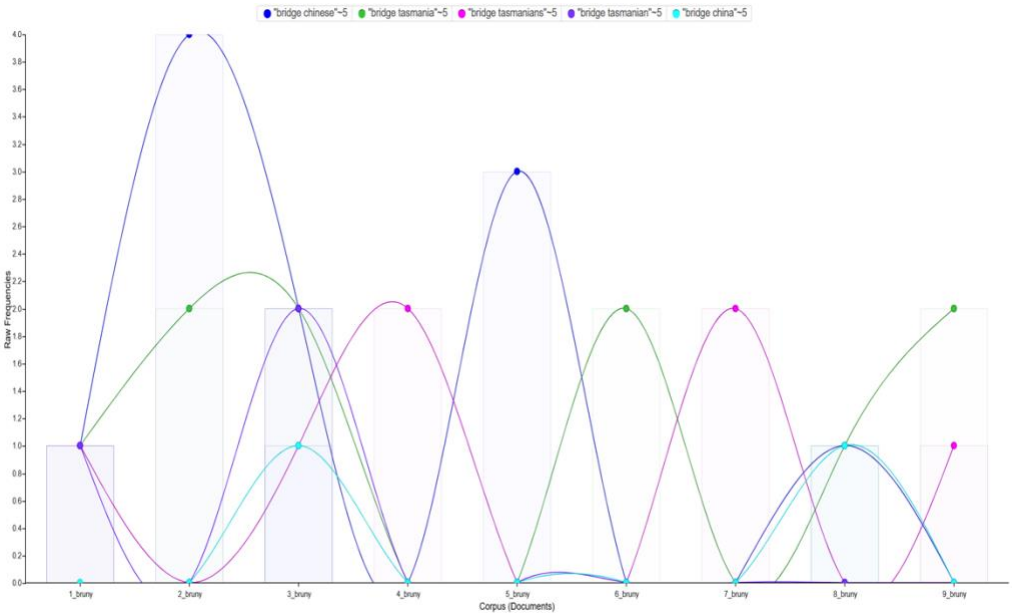


Figure 3 demonstrates the use of the "Termsberry" tool in Voyant that visualizes the co-occurrence of high-frequency terms to identify bi-grams involving the words Tasmanian and Chinese. These results were corroborated using the concordance widget in *Orange*. These implicit biases in the language of the text, likely not made consciously by the authors, stood exposed once the chapters and sentences were broken into word tokens and analyzed computationally.

Sophia Berg, an English major, in her project, "Contemporary Orientalism in Heather Rose's *Bruny,*" wrote: "Both the Orange and Voyant tools supported my intuitions from close reading with empirical data. The high consolidation of "Chinese" with dehumanizing words and "Tasmanian" with humanizing words, along with the increased fear and anger sentiment

paralleling increased Chinese representation in the novel, affirms the hypothesis that *Bruny*'s central political message is really one of racial othering." Berg created a storymap combining computational tools with historical-philosophical investigation for her final project.

Another student, Cadence Kahn, used topic modeling, time series, and collocation together to argue that *Bruny* employs tropes from the 'Chinese Invasion' genre to reflect and potentially reinforce contemporary immigration anxieties. Kahn tracked the changes in sentiment towards the bridge, a key motif in the novel, through its collocation with the words 'Chinese' and 'Tasmanian' to read the changing trends and sentiment surrounding the bridge in the novel. The stacked bar graph in Figure 4 visualizes the chapters of the novel where the bridge is more closely aligned with China or Tasmania. The areas where the words have equal frequencies in the context of the bridge suggest heightened tension or conflict. For her final project, Kahn analyzed the novels *Bruny* and *The Mandibles*: *A Family* as representative of anxieties about foreign influence and its perceived threat to democracy. She presented her work in this storymap.



In order to chart overall affective shifts over hundreds of pages, students combined Named Entity Recognition (NER) techniques with powerful language models like BERT. In a study of the "emotional and relational transformations" in Michel Houellebecq's novel *Submission*, Rohil Kanaparti, a CS major, used BERT to study how the language in the novel mirrored social and personal transformations. He used sentiment analysis and NER in Python to quantify and chart the changes in the protagonist's tone in the novel. The sentiment analysis showed that the protagonist's emotional tone transitions from negative to positive by the end of the novel, reflecting his acceptance of the new socio-political order. Figure 5 visualizes the shift from negative to positive sentiment near the end of the novel. The protagonist's sentiments towards other key characters in the text mirrored this shift in his attitudes. Kanaparti's final

project video and code can be found here.

While distant reading can be useful to identify patterns in a single novel spanning hundreds of pages, it really shines when analyzing text at scale. Near the end of the semester, the class experimented with using computational tools to distant read multiple novels simultaneously, something that would not be possible with traditional close reading methods in the span of a few weeks. Could students use their close reading of the novels and literary criticism discussed in class to identify generic patterns and larger trends? Judy Chen, majoring in Electrical and Computer Engineering (ECE) and Computer Science (CS), analyzed several dystopian novels spanning thousands of pages to identify their structural and rhetorical strategies. Topic modeling using Latent Dirichlet Allocation (LDA) helped identify key features of dystopian novels such as historical framing, dehumanizing and violent language, fear-driven plot structure (conspiracy, economic failure, invasion, and white-male protagonists bemoaning erosion of individual liberties) leading to doomsday scenarios (death, military, weapons, government), all of which aligned these novels with jingoist and right-wing discourses.

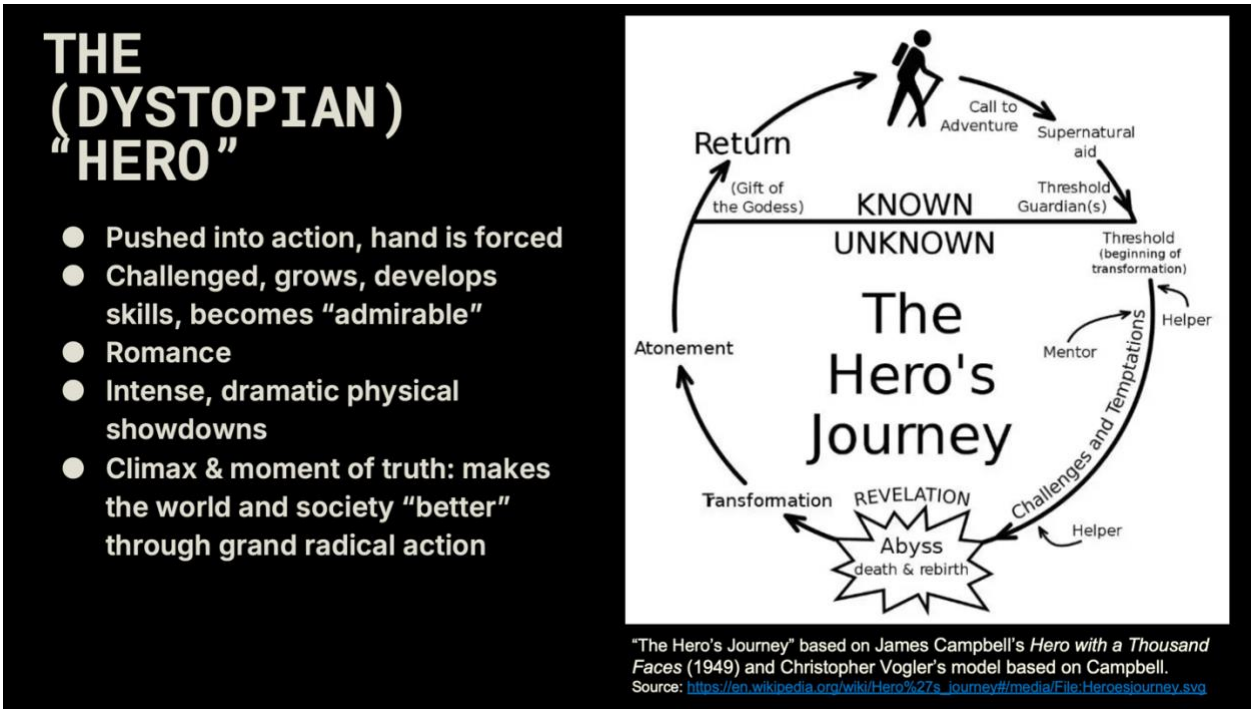


Figure 6 synthesizes all of these aspects in delineating the traits of the "dystopian" hero modifying the model of traditional hero's myth pattern pictured in James Campbell's *Hero with a Thousand Faces* (1949) and Christopher Vogler's model based on Campbell. Another student, Tatum McKinnis, majoring in Computer Science and English, summaries the key features of dystopia through her analysis of the themes and sentiment in J.M. Coetzee's novel *Waiting for the Barbarians* in the following equation: totalitarian regime + othering + archetypal protagonist + paranoia = dystopia. Chen and McKinnis presented their final project in the YouTube videos here and here.

Through different projects, the class collaborated on reading and sharing several computational methods to quantify the traits and moods in several dystopian novels. The NLP techniques we discussed facilitated the research of linguistic and structural features, tracking the evolution of specific motifs, and analyzing affective trends across several fictional texts. The visualizations of research findings enriched the class's understanding of how dystopian literature reflects and critiques contemporary social anxieties. By leveraging these computational tools to supplement their close reading of literary and historical texts, the class engaged with dystopian narratives in a nuanced and comprehensive manner, revealing the intricate ways in which these texts engage with themes of colonialism, power, government, and migration. The class was featured in this article published on the department website. The materials and student projects from the class can be accessed here.

Next Steps: The NLP methods and lessons from this course have made their way into the final chapter of my dissertation , "Allegories of Hospitality: Postcolonial Literature of Migration and Displacement," which reads literary texts from the postcolonial spaces through the trope of hospitality. In this chapter, I will develop and study a corpus of digital narratives of forcibly

displaced subjects from Asia and Africa. My research will involve analyzing the rhetoric of these texts through close and distant reading techniques I developed and refined for my dystopian literature class.

I am grateful to the Rhodes Information Initiative (iiD) at Duke for awarding me the Rhodes doctoral fellowship in the computational humanities, which allowed me to design and teach this course. I look forward to developing my skills in computational methods and will continue to report future research stemming from this work.