

Context Matters in Translation: A Corpus-Based Study of Cultural Variation in Arabic Shakespeare Translations

Fatima Al-Abdullah¹, Eid Mohamed¹, Tariq Yousef²

¹Department of English Literature and Linguistics, Qatar University

²Department of Mathematics and Computer Science, University of Southern Denmark

This paper presents an ongoing research project that develops a context-aware computational framework for analyzing Arabic translations and adaptations of Shakespeare as culturally embedded language data. While translation is often modeled as semantic transfer, we treat it as a process of cultural and ideological transformation, where meaning is reshaped across historical, political, and literary contexts (Venuti, 2017; Spivak, 2021). The project begins with Arabic translations of *Hamlet* as a pilot corpus and is designed to expand to additional Shakespeare plays, enabling broader comparative analysis of Arabic Shakespeare reception (Litvin, 2011; Hennessey and Litvin, 2022; Mohamed, 2024).

The project follows a two-stage design combining manual annotation, corpus construction, and computational analysis. The current phase, which began in January 2026, focuses on building a parallel corpus aligned at the scene and passage level. The corpus is enriched with a multi-layer annotation schema capturing culturally significant transformation phenomena: lexical modulation, ideological amplification or attenuation, register variation, and constructions of authority and voice. These categories are informed by postcolonial translation and adaptation studies and operationalized as structured labels suitable for corpus analysis and downstream NLP tasks.

The annotation process is conducted by trained students from an English Literature program. Annotators participated in workshops introducing the theoretical background, annotation categories, and practical procedures. Detailed annotation guidelines were developed to support consistency and reliability, and the process is supervised by experts in digital humanities and NLP. This setup addresses a central challenge in culturally interpretive annotation: transforming qualitative literary judgments into reproducible and computationally usable data. From an NLP point of view, the corpus opens up

several directions for analysis, including classification of translation strategies, tracking changes over time, clustering similar adaptation patterns, and applying methods such as topic modeling, sentiment analysis, and network analysis to capture relationships between texts, translators, and contexts (Moretti, 2013; Jänicke et al., 2015). Beyond providing a new Arabic literary translation resource, the project also offers a practical way of modeling how translation varies across cultural and historical settings. Looking at the pilot material, we already see clear and consistent differences across translations in how political critique, religious framing, gendered language, and rhetorical authority are expressed or toned down (Hanna, 2016). Instead of treating this variation as noise, we take it as a meaningful signal of underlying adaptation strategies shaped by historical and ideological context. This perspective is particularly important for Arabic NLP (Yousef et al., 2023), where datasets that capture this kind of cultural and historical depth are still largely missing (Obeid et al., 2020).

In ongoing work, we are extending the annotated corpus and evaluating inter-annotator agreement to ensure the robustness of the annotation scheme. The resulting dataset will be used to train and evaluate classification models for predicting adaptation strategies, providing a first benchmark for culturally informed translation analysis in Arabic, and will be released together with annotation guidelines to support reproducibility. More broadly, this effort bridges NLP and digital humanities by integrating close reading, structured annotation, and computational cultural analytics. By treating translated literary texts as historically situated artifacts rather than decontextualized strings, the project advances a context-sensitive paradigm for Arabic NLP and computational literary studies, and invites further discussion on annotation design, theoretical operationalization, and the role of cultural context in NLP pipelines.

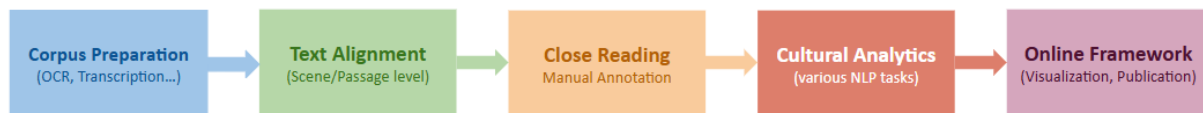


Figure 1: Project workflow from corpus preparation to NLP-based cultural analytics. The pipeline integrates text alignment, manual annotation, and computational analysis, culminating in an online framework for visualization and dissemination.

References

- Sameh Hanna. 2016. *Bourdieu in translation studies: The socio-cultural dynamics of Shakespeare translation in Egypt*. Routledge.
- Katherine Hennessey and Margaret Litvin. 2022. *Shakespeare and the Arab World*. Berghahn Books.
- Stefan Jänicke, Greta Franzini, Muhammad Faisal Cheema, Gerik Scheuermann, and 1 others. 2015. On close and distant reading in digital humanities: A survey and future challenges. *EuroVis (STARs)*, 2015:83–103.
- Margaret Litvin. 2011. *Hamlet's Arab journey: Shakespeare's prince and Nasser's ghost*. Princeton University Press.
- Eid Mohamed. 2024. The potential and limits of arabic digital humanities. *Journal of Cultural Analytics*.
- Franco Moretti. 2013. *Distant reading*. Verso Books.
- Ossama Obeid, Nasser Zalmout, Salam Khalifa, Dima Taji, Mai Oudah, Bashar Alhafni, Go Inoue, Fadhl Eryani, Alexander Erdmann, and Nizar Habash. 2020. Camel tools: An open source python toolkit for arabic natural language processing. In *Proceedings of the twelfth language resources and evaluation conference*, pages 7022–7032.
- Gayatri Chakravorty Spivak. 2021. The politics of translation. In *The translation studies reader*, pages 320–338. Routledge.
- Lawrence Venuti. 2017. *The translator's invisibility: A history of translation*. Routledge.
- Tariq Yousef, Lisa Mischer, Hamid Reza Hakimi, and Maxim Romanov. 2023. [Enhancing state-of-the-art NLP models for classical Arabic](#). In *Proceedings of the Ancient Language Processing Workshop*, pages 160–169, Varna, Bulgaria. INCOMA Ltd., Shoumen, Bulgaria.