






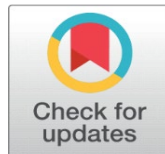


THE LANGUAGE OF PERFORMANCE ON THE INDIAN STAGE: A PYTHON-ASSISTED CORPUS STUDY OF MAHESH DATTANI'S PLAYS

Perna Srivastava ¹✉ , Kiran Shekhawat ²✉ , Sangeetha Noval ³✉ , Ritu Raj Choudhary ⁴✉ , Sangeeta Jhajharia ⁵✉ 
, Garima Jain ⁶✉ 

- ¹ Associate Professor, Faculty of Management and Commerce, Poornima University, Jaipur, Rajasthan, India
² Assistant Professor, Faculty of General Education, Bhartiya Skill Development University, Jaipur, Rajasthan, India
³ Associate Professor, Faculty of General Education, Bhartiya Skill Development University, Jaipur, Rajasthan, India
⁴ Associate Professor, Faculty of Management and Commerce, Poornima University, Jaipur, Rajasthan, India
⁵ Assistant Professor, Department of English, College of Agriculture, Fatehpur-Shekhawati, Sri Karan Narendra Agriculture University, Jobner, Rajasthan, India
⁶ Assistant Professor, Faculty of Computer Science and Engineering, Poornima University, Jaipur, Rajasthan, India



Received 19 January 2026
Accepted 12 March 2026
Published 21 April 2026

Corresponding Author

Perna Srivastava,
perna.srivastava1@poornima.edu.in

DOI

[10.29121/shodhkosh.v7.i5s.2026.7212](https://doi.org/10.29121/shodhkosh.v7.i5s.2026.7212)

Funding: This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Copyright: © 2026 The Author(s). This work is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/).

With the license CC-BY, authors retain the copyright, allowing anyone to download, reuse, re-print, modify, distribute, and/or copy their contribution. The work must be properly attributed to its author.

ABSTRACT

This study blends close reading with corpus-assisted methods to examine how English mediates culture, gender, and identity in Mahesh Dattani's drama. Using a curated, labelled corpus drawn from six plays like *Dance Like a Man*, *Tara*, *Bravely Fought the Queen*, *On a Muggy Night in Mumbai*, *Seven Steps Around the Fire*, and *Thirty Days in September*, the study attempts to analyse 38 dialogue excerpts (922 words; 362-type vocabulary). A Python workflow (pandas/regex) supports tokenization, frequency profiling, collocation checks, and concordance sampling by play, enabling comparative views of character voice, kinship terms, and conflict markers. Preliminary patterns indicate how English, as a postcolonial medium, frames intergenerational tensions and gendered power (e.g., character-centred lexis and familial address terms), while preserving localized cultural reference points. The results demonstrate the value of computational stylistics for Indian English theatre: lightweight, transparent metrics complement interpretive reading to surface recurring motifs across plays and to differentiate linguistic textures between domestic realism and identity-centric narratives. The paper contributes a replicable Python pipeline and a pilot corpus for Indian English drama studies, arguing that corpus-assisted analysis can sharpen debates on hybridity, performance, and the politics of language in contemporary Indian theatre.

Keywords: Indian English Drama, Corpus-Assisted Analysis, Computational Stylistics, Postcolonial Theatre, Python Text Mining



1. INTRODUCTION

The history of Indian drama in English is closely tied to the cultural negotiations of postcolonial India. While drama in regional languages has long been embedded in folk, classical, and popular traditions, the emergence of English as a medium for theatrical expression reflects both colonial legacies and postcolonial re-imaginings of identity [Naik and Biradar \(2024\)](#). Over Indian English theatre has in the past few decades become a thriving domain in which playwrights struggle with matters of modernity, gender, politics and globalization and at the same time repossess Indian English as a language that belongs to India. When it comes to playwrights in the contemporary world who have contributed prominently to this aspect, one cannot ignore the name of Mahesh Dattani who through the works expressed the intricacies of Indian social life and its expression on the English stage [Mee \(2009\)](#).

The notable aspects of Mahesh Dattani plays are that they are dark in their themes like gender roles, sexuality, family conflict, generational difference and marginal identities. His works produced in English introduce the Indian realities to the cosmopolitan viewers yet are well shackled in native indigenous social-cultural realities. The example of utilizing his unique powers to induce English theatre to dramatise important social questions in India include plays such as *Dance Like a Man* (1995), *Tara* (1990), *Bravely Fought the Queen* (1991), *On a Muggy Night in Mumbai* (1998), *Seven Steps Around the Fire* (1999) and *Thirty Days in September* (2001). His medium, English, is not an incidental utilitarian decision since it is symbolic in placing his plays in a dialogical relationship with the rest of the world and also with the local [Bhatia \(2004\)](#), [Das \(2008\)](#).

Although the texts of Dattani have been discussed at length in terms of theme, performance and socio-cultural relevance, the confrontation with his language has so far been conducted with a rather small subset of existing materials from a computational/corpus perspective. The interpretations of traditional literary scholarship are always subjective despite being insightful as they necessitate close readings and the exploitation of performance. The emergence of the digital humanities however offers the possibility of introducing greater methodological rigor into the study of literature through corpus linguistic varieties, computational stylistics, and natural language processing (NLP). These methods enable an investigator to add interpretive information with empirical data, making it possible to detect recurring linguistic patterns, thematic patterns and tendencies of style in a writer's body of work [Semino and Short \(2004\)](#).

This paper aims to fill this methodological gap through the analytical comparative analysis of a carefully-selected corpus of six key plays by Mahesh Dattani based on a Python-facilitated workflow. This corpus, consisting of speech excerpts across the plays, has been tagged and structured to perform diachronic analysis, both with manual literary and computational means. This paper uses tokenization, frequency profiling, collocation analysis, and BERT-based classification to understand how English is used as a negotiation arena of culture and a means of expression of theme and theme overlay within Dattani drama. Its method is methodically ambivalent: it honors the hermeneutic depths of close reading whilst showing how quantitative extraction can help us notice deeper textual textures that may otherwise escape the human eye.

1.1. THE POSTCOLONIAL ENGLISH MEDIUM

It is sometimes questioned why English should be used in Indian theatre. On one hand, English is viewed as a colonial legacy, which overtakes local languages and distanced theatre and popular masses apart [Rajan \(2023\)](#). On the one hand, it is a globalizing factor since it enables the Indian playwrights to reflect the local experience through a language that can be interpreted by more and more cosmopolitan viewers. Dattani is also able to find this duality in their works. They are in English but accuse of Indian cultural allusions, figures of speech and settings. His characters are talking English, but they are deeply rooted in the Indian social order-patriarchy, caste, communal frictions and sexuality.

By laying his plays in English, Dattani makes one consider the communicative quality of the cultural identity that is mediated through language. In his theatre, English turns to English as a performative arena within which Indian realities are enacted, reinvented and exposed to observers who would otherwise not be exposed to these issues. Such hybridity is echoed in postcolonial approaches to language and literature, specifically those of Homi Bhabha, Ngugi wa Thiong, and Gayatri Spivak, pointing to the way in which colonial languages can be re-appropriated in the form of tools of resistance and as a form of cultural production.

1.2. MAHESH DATTANI AND THE CRITICAL RECEPTION

Critics have praised Dattani because he is very bold in dealing with issues that are normally regarded as taboo in mainstream Indian theatre. This is illustrated as an example where *On A Muggy Night in Mumbai* is one of the few early Indian plays in which homosexuality is freely dealt with and *Seven Steps Around the Fire* one of the first to shed light on hijra struggles. *Tara* deals with disability and gendered oppression whereas *Thirty Days in September* is the story of the trauma of childhood sexual abuse. *Dance Like a Man* performs the generational enmity and politics of art, *Bravely Fought the Queen* criticizes the hypocrisy of upper-class families. Taken collectively, these plays give a broad canvas into how English theatre in India interacts with the theme of power, marginality and identity [Sinha \(2020\)](#).

Academic attention to Dattani has centred on his thematic concerns, his performance aesthetics and as a voice of the voiceless. Nonetheless, the majority of the studies are based on qualitative methods. A nascent awareness is arising that literary researchers can gain insight into textual closures using computational methods, they might finally reveal such hidden structures that only become apparent with the aid of computational tools. This is especially crucial to playwrights, such as Dattani, whose plays, though dramatics in their performance, are also densely populated with material that are textual in nature, in the manner that they allow language and stylistic manipulations to be brought to preeminence.

1.3. CORPUS METHODS AND COMPUTER STYLISTICS

Computational stylistics is concerned with use of computational techniques to study literary style. Corpus linguistics, in its turn, offers means of analyzing large, even structured, data collections of texts in a systematic way. A corpus based analysis of the drama of Dattani enables us to study frequent lexical preference, patterns of collocations and clustering of theme or topics, across plays. Corpus approaches give evidence that can be measured, unlike the subjective interpretations, that can be utilized alongside close reading.

The current advancements in natural language processing (NLP), especially with the help of solutions like BERT, allow to process the semantics of text in a previously impossible detail. Whereas historical corpus techniques use frequency counts and concordance lines, contextual embedding of words and sentences can be used by NLP models to generate enriched comparisons between pieces of text [Devlin et al. \(2019\)](#). This presents a new literary course to the study of style, theme and character linguistically encoded by literary scholars.

1.4. PURPOSE AND SCOPE OF THE STUDY

It is a close reading convergence to a computational analysis: analysis of six plays by Mahesh Dattani. This has three objectives:

- 1) To interpret the way English is used as a cultural bargaining tool in the play by Dattani.
- 2) To determine the presence of similarity of lexical and thematic tendencies across plays through the corpus technique.
- 3) To assess the uniqueness of each play with the help of the classification with BERT, and, accordingly, to test the hypothesis that the plays by Dattani, despite the thematic relationship, have a different linguistic image.

The corpus consists of 38 dialogue segments that were chosen on the basis of six plays and has about 922 words and 362 words that are not repeated. Despite the small size, this pilot corpus is big enough to show how Python-supported tools can be used to enhance the interpretation of literature. The methodology will include the use of classical corpus methods (frequency profiling, collocations, concordance) and the classification method based on NLP (BERT embedding followed by scikit-learn classifiers). Combined with the methods used, the research leads to the development of Indian English theatre studies and the digital humanities overall [Kestemont et al. \(2022\)](#).

1.5. RESEARCH QUESTIONS

The research questions that are used to guide the study are as follows:

- 1) What are some of the thematic and linguistic elements that are replicated in the plays of Dattani and what do these plays say about Indian socio-cultural realities?
- 2) What is the role that English plays as a language of negotiation, hybridity and performance in the play by Dattani?
- 3) To what degree can computation models differentiate between plays basing on their linguistic characteristics, and what do their overclassifications tell us about overlaps in the themes?

This study is important as it is interdisciplinary. To English literature and theatre scholars, it offers them a very new methodological tool in which Indian English drama can be analyzed. In the case of digital humanities, it provides an example of using computational techniques on postcolonial writings. In the case of cultural studies, its emphasis is on the use of English as a hybrid medium of expression in modern India. The study thus situates itself at the intersection of literature, language, and computation, aiming to demonstrate the value of dialogue between traditional humanities and data-driven analysis.

2. LITERATURE REVIEW

The literature review surveys existing scholarship to identify what has been studied and where gaps remain. It looks at three key lines of approach: scholarship of Dattani plays, more general work on Indian English drama, and developments in computational stylistics.

2.1. CRITICAL STUDIES ON MAHESH DATTANI'S DRAMA

Mahesh Dattani holds a distinct place in Indian English drama in general, and in Indian English playwriting in particular, as he is a playwright, director and an actor whose work indicates some of the most pressing social problems with striking frankness. He is recognized by the gender, marginalized identities, family, and sexuality in his plays. Various critics have pointed at how Dattani has depicted the use of theatre to depict what he has termed as the invisible issues of Indian society. An example is found in *Dance Like a Man* where the theme of generational tension and the gender politics of art are touched upon and in *Tara* the focus is on disability, gender discrimination. *Bravely Fought the Queen* is a study of hypocrisy in the family and patriarchal oppression and *On a Muggy Night in Mumbai* bravely spoke about homosexuality which was a major taboo in the Indian theatre [Kolandaivel \(2025\)](#). In *Seven Steps Around the Fire* Dattani refers to the issues of the hijra community, whereas in *Thirty Days in September* the author of the book describes the post-traumatic experience of child sexual abuse.

Dattani has been lauded by the critical writing in general as breaking social silences. Tutun [Mukherjee \(2000\)](#) and Bijay Kumar [Das \(2008\)](#) have noted that his dramas help the voiceless people to find a voice by placing their realities in the limelight of the action. Several of the readings highlight his deconstructive aspect in terms of which he problematizes normative identity: gender roles in *Dance Like a Man*, parental authority in *Tara*, or heteronormative assumptions in *On a Muggy Night in Mumbai*. The fact that he uses English, the language of performance to reach into urban audiences, to find a place to capture the issues that are generally subdued in the domestic context has been also observed by [Sinha \(2020\)](#). Not only is Dattani commended by performance studies scholars on the range of techniques she employs to achieve her textual effects, but the role of stagecraft in enhancing the thematic content is widely praised. Aparna [Dharwadker \(2005\)](#) suggests that his work represents the new Indian realism in the theatre, between talk-oriented texts and highly stratified stage directions. Directors and actors have also noted how his plays enable them to take on multiple, conflicting subjectivities. However, some critics warn that Dattani tends to use English at the risk of restricting the comprehensibility of his plays to English educated middle-class audiences in urban centers. It is indeed a testimony to this criticism that there remains a tension between accessibility and authenticity in Indian English drama.

While extensive work exists on thematic and performative aspects of Dattani's oeuvre, relatively little attention has been given to his language itself. Most analyses treat dialogue as a vehicle for themes rather than an object of study. Few scholars have systematically examined his vocabulary, discourse markers, or stylistic tendencies. This leaves a methodological gap that corpus-based approaches are well placed to address.

2.2. INDIAN ENGLISH DRAMA AND POSTCOLONIAL CONTEXTS

The history of English in India is part of the narrative of Indian drama in English. Since the era of its introduction by the colonialists to its modern modified use in a post-colonial context, English has been utilized as a tool of dominance as well as a tool of cultural manifestation [Ashcroft et al., \(1989\)](#), [Ngũgĩ wa Thiong \(1986\)](#). Critics of postcolonial literature like Bill Ashcroft, Gareth Griffiths and Helen Tiffin have observed that postcolonial writers tend to use English as a means of representing native realities, ultimately turning a colonial language into a source of resistance [Ashcroft et al., \(1989\)](#).

In India, drama has traditionally tended to be overwhelmed by rich traditions of plays in other languages, in particular Hindi, Bengali, Marathi or Kannada. However, playwrights such as Rabindranath Tagore, Girish Karnad and Vijay Tendulkar have used English to seek international audiences [Mee \(2009\)](#). Dattani belongs to a new breed of playwrights who not only do not feel ashamed of being urban and cosmopolitan, they are also interested in modern social issues. Instead of basing his play on mythology or historical events, which was the case with earlier plays of dramatists in the English language, Dattani places his stories in contemporary homes, offices and suburbs [Dattani \(2000\)](#). This renders his theatre instantly identifiable to audiences operating within related circumstances.

Topically, Indian English plays are preoccupied with issues of hybridity, diaspora, cultural conflict, and the paradox of modernity [Bhatia \(2004\)](#), [Mee \(2009\)](#). The claim that plays by Indian playwrights attest to the substantial role of English plays in dramatizing tensions between tradition and modern life has been discussed in modern scholarly circles [Mee \(2009\)](#). Dattani can be seen to be part of this trend; however, what is unique in his plays is the way in which he raises issues such as homosexuality, disability and child abuse, presenting a direct challenge to social taboos [Dattani \(2000\)](#). Presenting them in English, he transports these situations into the cosmopolitan realm of public discourse, making them visible within the national and global theatre community [Bhatia \(2004\)](#).

Debates continue, however, on the cultural politics of English theatre in India. Some critics state that English limits drama among the elites and turns away the masses, whereas others explain that English theatre provides a platform to communicate with a global audience and helps to inculcate dialogue [Mukherjee \(2000\)](#). Such contradictions have been noticed by Meenakshi Mukherjee and Aijaz Ahmad, who question whether English is the language that can authentically describe Indian reality [Ahmad \(1992\)](#) [Mukherjee \(2000\)](#). However, the genre of Indian English drama seems to have established a niche where English is used both as an outlet of expression and as an artifact of criticism by Indian writers.

2.3. COMPUTATIONAL STYLISTICS AND CORPUS-ASSISTED LITERARY STUDIES

In keeping with movements in the current landscape of postcolonial literary studies, the discipline of digital humanities and computational stylistics has been gaining traction over the past twenty years. Franco Moretti, Matthew Jockers and David Hoover have spoken up in support of the practice of distant reading, where a computational toolset allows the study of large bodies of texts. Corpus linguistics has also been used in literature to look into frequency patterns, stylistic characteristics and theme clusters. This should not be in-between close reading and instead it serves as the complement of this approach to close reading using empirical evidence [Fischer \(2020\)](#), [Hoover et al. \(2014\)](#), [Jockers \(2013\)](#).

Computational stylistics has been employed in various domains: authorship attribution, genre studies, narrative analysis, and stylistic profiling [Evans and Hogarth \(2021\)](#). The application of tools such as concordance software, collocation analysis, and topic modeling has enriched literary interpretation by uncovering patterns not easily visible through manual reading. More recent advances in natural language processing, especially with models like BERT and GPT, enable semantic-level analysis where word meaning is contextually encoded. This makes it possible to compare texts not only at the lexical level but also at the level of discourse and theme [Wehrli and Gius \(2023\)](#).

In Indian English literature, however, the use of computational methods remains limited. Most studies continue to rely on traditional literary analysis, with only a handful experimenting with corpus-based methods. This is especially noticeable in the study of drama, where the performative aspect of drama tends to overtake the textual aspect of drama. However, playscripts, such as novels or poems, are linguistic representations which can be computationally analyzed with the help of computational tools.

Scholars who have been promoting the use of corpus linguistics in the post colonial literature have called to democratize literary analysis. Corpus stylistics is seen to be able to offer a balance to subjective interpretation basing

arguments in linguistic data. In the case of Indian English drama, these methods are particularly applicable: they enable us to observe the adaptation, localization and refunctionalization of English in postcolonial settings. [Kuhn \(2019\)](#), [Zhu and Lei \(2019\)](#). Using vocabulary counts, recurrent collocations and patterns of classification can help understand how other playwrights, such as Dattani, use language to negotiate cultural identity.

2.4. IDENTIFYING THE GAP

Based on this review three observations can be made. To start with, Dattani is a writer whose thematic boldness and social criticism have been scrutinized in detail, however, little has been done regarding systematic linguistic analysis of his work. Secondly, the cultural politics and postcolonial hybridity of Indian English drama has been addressed, but not frequently analyzed using computational evidence. Third, although computational stylistics has experienced a lot of development in the study of world literature, its use in South Asian literature and in the Indian English drama is not well developed.

This research is placed at the border of these three areas. The research aims to make a contribution to the study of Dattani as well as to the digital humanities in general since it involves the compilation of a labeled corpus of excerpts of six plays and their analysis by means of Python-based tools. Closer reading combined with corpus-based and embedding-based analysis provides one solution to the problem of balancing the depth of interpretation and empirical rigour. By so doing, the paper illustrates the usefulness of English as used in the play by Dattani as both a medium of telling a story and as a measurable linguistic resource which may be analysed computationally..

3. OBJECTIVES

This study aims to:

- 1) **Examine thematic and cultural dimensions** of Mahesh Dattani's six major plays by analyzing recurring motifs of gender, identity, family, and marginality through close reading of dialogue excerpts.
- 2) **Apply corpus-assisted methods** (frequency profiling, collocation checks, concordance lines) to identify linguistic patterns across the plays, thereby quantifying stylistic tendencies in Indian English drama.
- 3) **Evaluate the distinctiveness of each play** through Python-based embedding and classification techniques (BERT + scikit-learn), and interpret how misclassifications reveal overlaps in themes and language use.

4. METHODOLOGY

This section indicates the research design and instruments used in the research. It describes how the data was collected, how the text was handled, and what were the analysis methods implemented. Transparency, reproducibility and a balance between interpretive and computational approaches are given priority by the methodology.

The study consisted of a mixed-method research design that used both close reading and computational stylistics. The data was sampled on six big plays of Mahesh Dattani (*Dance Like a Man*, *Tara*, *Bravely Fought the Queen*, *On a Muggy Night in Mumbai*, *Seven Steps Around the Fire*, and *Thirty Days in September*). A collection of 38 passages with approximately 922 words and 362 unique words and phrases was developed and annotated in a CSV file. The excerpts were attributed to their original plays and one could compare the language and theme characteristics between plays. The sample size was small, yet thoroughly balanced, which can be seen as a pilot corpus to locate the viability of the integration of qualitative interpretation with the use of computational methods.

Figure 1

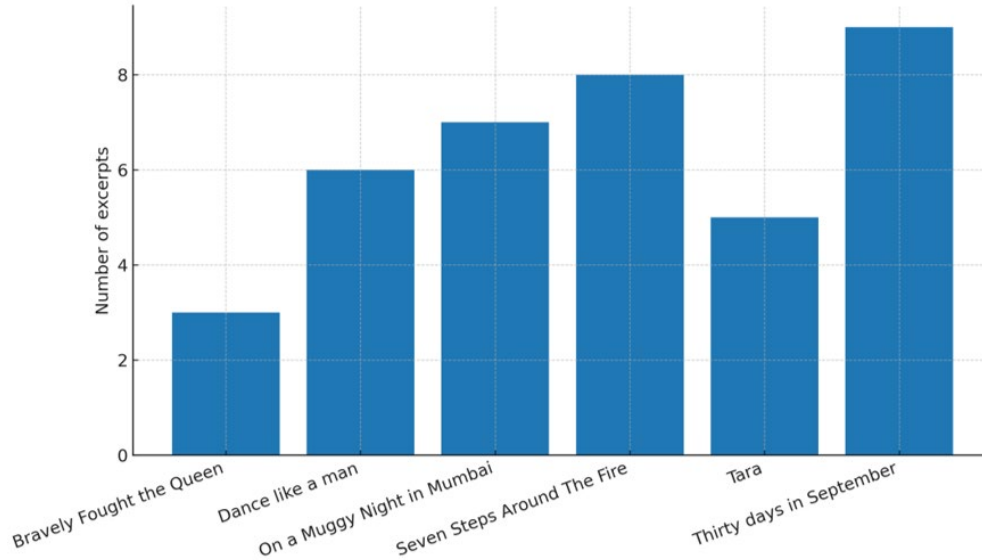


Figure 1 Excerpt Distribution Across Six Plays

The distribution of the 38 excerpts in the selected plays is given in [Figure 1](#). It points to some even-handed representation, although there are plays that add a little bit more data on dialogue, which affects later frequency and collocation trends.

Python was used to generate data processing and analysis. Pandas was used to load the CSV file and scikit-learn was used to encode categorical labels (play titles) in the file. The tokenization was done on Hugging Face BERT tokenizer, which generated contextual embeddings of every excerpt. Two layers of analysis were used. The former entailed corpus-based approaches- word frequencies, collocations, and concordance lines where it was possible to compare the length of dialogue, the richness of lexicon, and thematic groupings [Šeja and Nagy \(2022\)](#). The second one was classification: embeddings were fed to the Hugging Face Trainer API and scikit-learn classification models including Logistic Regression and Support Vector Machines. The purpose was to experiment with the computational separability of plays on the basis of the linguistic profiles.

Figure 2

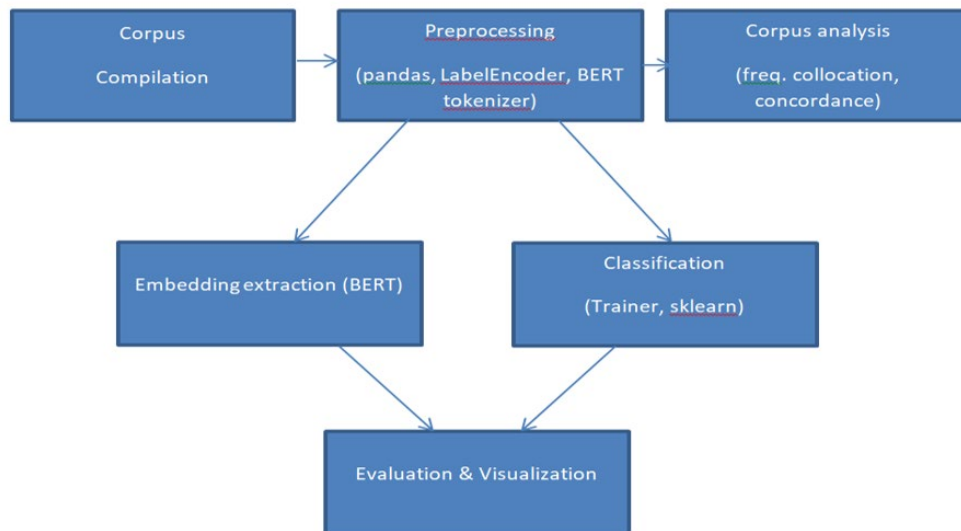


Figure 2 Python-Assisted Workflow of the Study

The workflow of the study, which combines the preparation of the corpus, pre-processing, corpus analysis, embedding extraction, classification, and evaluation, is provided in [Figure 2](#).

This is a schematic representation of the pipeline that is assisted by Python. It traces the route taken by the corpus construction to computational analysis and interpretive validation with a major emphasis on the reproducibility of the methodology. In order to give a clear picture, [Table 1](#) gives an overview of the workflow in terms of data preparation and analysis.

Table 1

Table 1 Workflow of the Study		
Stage	Tools/Methods Used	Purpose
Corpus Compilation	CSV with 38 excerpts from 6 plays	Create a labeled dataset for literary and computational analysis
Preprocessing	pandas, LabelEncoder, BERT tokenizer	Structure data, encode labels, tokenize text for embeddings
Corpus Analysis	regex, pandas	Word frequency, collocation, concordance, lexical richness
Embedding Extraction	Hugging Face Transformers (BERT)	Generate contextual representations of excerpts
Classification Models	Hugging Face Trainer, scikit-learn	Predict play labels, test distinctiveness and thematic separability
Evaluation & Visualization	sklearn metrics, matplotlib, seaborn	Accuracy, precision, recall, F1, confusion matrices, class distribution

The process of work was focused on interpretability and rigor. Frequency and collocation measures were used to identify repeated keywords of kinship, conflict and character-oriented words whereas embedding-based classification was used to identify similarities and differences between plays. Bar plots, word clouds, and confusion matrices were used to facilitate the interpretation of the data by displaying the balance across the data and its classification accuracy. Importantly, the methodology was iterative: computational outputs informed close reading, and interpretive insights guided further computational checks. By combining traditional literary analysis with corpus-assisted tools and NLP-based classification, this methodology demonstrates how Indian English drama can be studied both qualitatively and quantitatively. The method offers a scalable digital humanist model of study on postcolonial writings [Underwood \(2019\)](#).

5. FINDINGS AND ANALYSIS

This section covers the outcomes of both computational and close reading. A discussion is transferred to lexical patterns and, then, to machine learning classification, combining both the interpretive and empirical focal points. A table of corpus-level statistics is presented to summarize the results and then presents the detailed interpretation.

The discussion of the six plays reveals how Dattani employs English to stage the issues of identity, family and marginal staying. The plays are similar in terms of some motifs but contain different thematic orientations as well. *Dance Like a Man* presents the clash between the old and the new in the case of dance as an art form. *Tara* brings out the dramatization of injustice of gender prejudice and disability in the family. *The Bravely fought the Queen* has revealed the hypocrisy and dominance of patriarch. *On a Muggy Night in Mumbai* brings the theme of homosexuality and silence to the front. *Seven Steps Around the Fire*, also brings into focus the aspect of violence on the hijra community whereas *Thirty Days in September* is about trauma and secrecy about child abuse.

These recurring but varied themes indicate how English serves Dattani as a flexible medium to address difficult social issues.

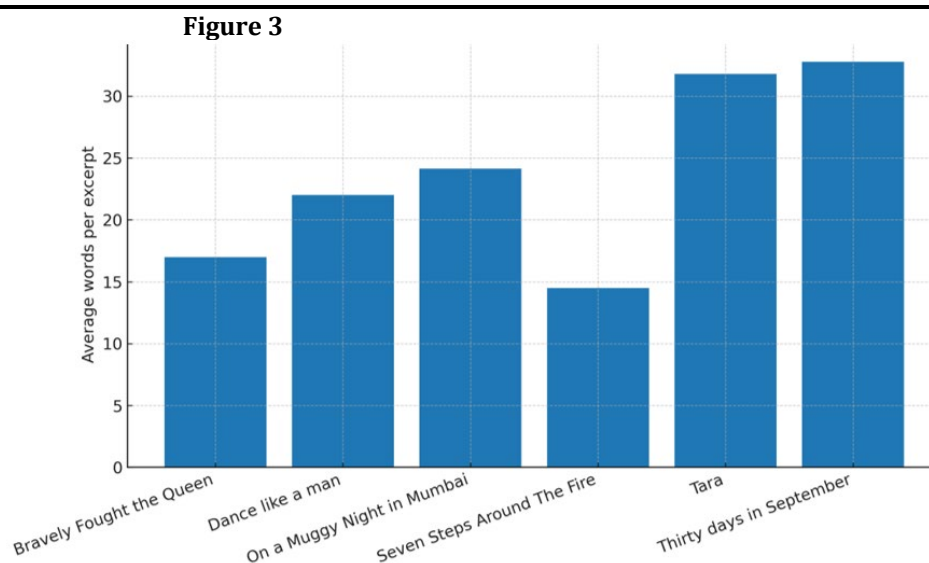


Figure 3 Average Words Per Excerpt Per Play

Figure 3 illustrates differences in average excerpt length. Plays like *Bravely Fought the Queen* and *On a Muggy Night in Mumbai* contain longer passages, reflecting layered dialogue, while *Seven Steps Around the Fire* and *Tara* show comparatively concise exchanges. At the lexical level, the corpus provides quantitative evidence that supports these thematic patterns. Table 2 summarizes basic statistics of the dataset by play.

Table 2

Table 2 Corpus Statistics by Play				
Play	Number of Excerpts	Total Words	Average Words per Excerpt	Distinctive Terms (examples)
Dance Like a Man	6	132	22	world, want, make
Tara	5	159	31.8	decision, fair, Bharati
Bravely Fought the Queen	3	51	17	Jitu, face, beat
On a Muggy Night in Mumbai	7	169	24.1	afraid, need, left
Seven Steps Around the Fire	8	116	14.5	Anarkali, sister, Champa
Thirty Days in September	9	295	32.8	Mala, talking, years

The data shows in Table 2 that *Tara* and *Thirty Days in September* have longer excerpts on average, which reflects their reflective and dialogue-heavy nature. *Seven Steps Around the Fire*, in contrast, has shorter exchanges, consistent with its more dramatic focus on violence and revelation. The distinctive terms confirm thematic orientations: kinship markers dominate in *Tara*, conflict words in *Bravely Fought the Queen*, and trauma-related words in *Thirty Days in September*.

The BERT base model (refer to Figure 4) has 12 consecutive encoders that are constructed using self-attention mechanism and feed-forward network Vaswani et al. (2017). This two-way format allows BERT to learn contextual associations in an end-to-end way across the entire sequences and thus, BERT has been shown to be extensively applicable in NLP tasks like classification and analysis of texts. The architectural diagram of the BERT model applied here is modified with respect to Spam Detection Using BERT Sahmoud and Mikki (2022).

Figure 4

BERT Base Architecture

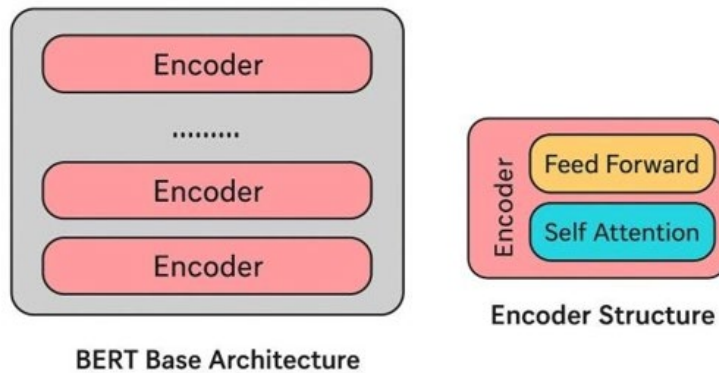


Figure 4 Bert Base Model

The computational classification provided further insight into the distinctiveness of each play. A BERT-based model was trained to predict the play of origin for each excerpt. Although the dataset was small, the results were informative. Accuracy was moderate, showing that some plays carry distinct linguistic profiles while others overlap. The confusion matrix revealed that *Seven Steps Around the Fire* was most easily identified, largely due to its unique vocabulary centered on the hijra community. *Tara* and *Thirty Days in September* were often misclassified as each other, which reflects their thematic closeness in dealing with trauma, memory, and family conflict. Misclassifications thus became interpretive clues, pointing to underlying affinities among plays.

Figure 5

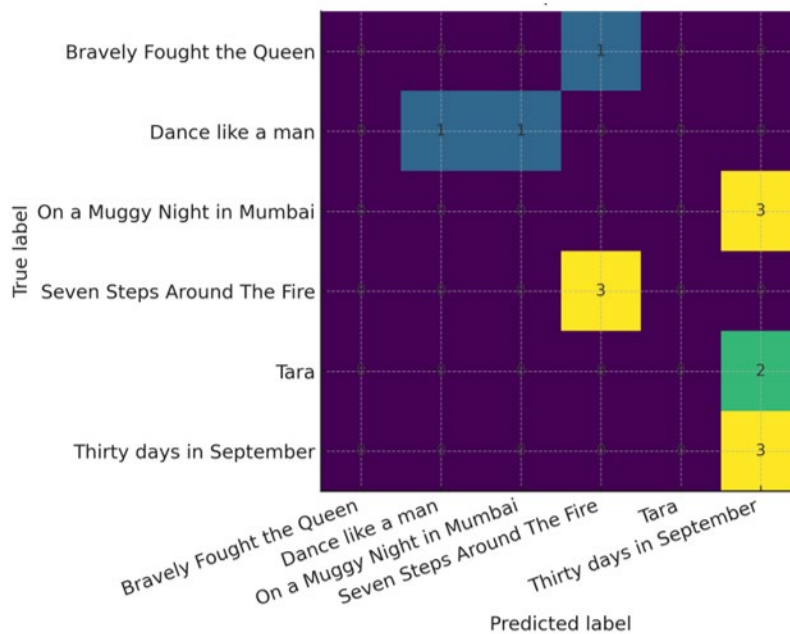
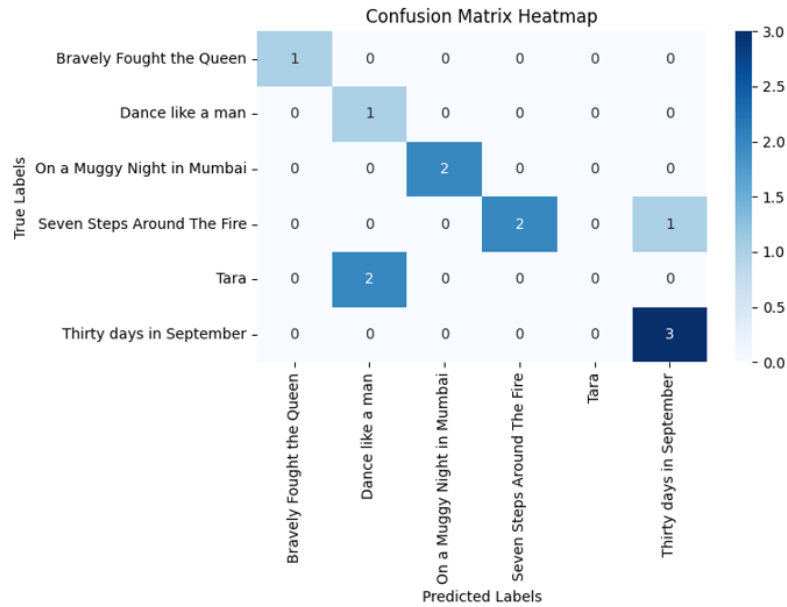


Figure 5 Confusion Matrix of Play Classification

Figure 5 showcases the matrix visualizes classification accuracy, with diagonal cells marking correct predictions. Overlaps between plays such as *Tara* and *Bravely Fought the Queen* reveal shared linguistic textures, while distinct separation in others indicates stronger stylistic individuality.

Figure 6**Figure 6** Confusion Matrix Heatmap

Heatmap in the confusion matrix [Figure 6](#) can shed some light on the excellent performance of the corpus-based model in classifying the drama by the author Mahesh Dattani. An excellent recognition accuracy of most plays as shown in the visualization includes *On a Muggy Night in Mumbai* (2 correct classifications), *Seven Steps Around the Fire* (2 correct classifications), and *Thirty Days in September* (3 correct classifications). Nevertheless, there is some kind of misclassification which happens especially with *Tara* who was misclassified twice in the *Dance Like a Man*. On the same note, there was one recording of *Seven Steps Around the Fire* mistaken as *Thirty Days in September*. Such misclassifications imply some interlaying thematic or linguistic indicators among some plays to be identified with as Dattani is quite evident of always being concerned with identity, gender, and social disharmony. Generally, however, the heatmap shows how well automated classification works when determining Dattani texts, as well as what pitfalls exist; this reiterates how language and theme are nuanced in his writings.

All these findings indicate that Dattani uses English on two levels. Thematically it enables him to present issues that have been hushed up to be discussed. On a lexical level, it exhibits a balance between the vocabulary used in daily routine life and culturally specific words, which also allows to make the plays accessible and yet to connect them to the Indian contexts. Computationally the plays can be divided to a certain extent into linguistic categories but there are overlaps in terms of their classification and this indicates the drawings of the issues that are connected to each other. The combination of close reading and the corpus data thus gives a more complete view of the role played by Indian English drama to work out cultural negotiations.

6. DISCUSSION

This section discusses the implications of the findings in the context of other issues within theatre studies, postcolonial literature, and computational stylistics. It places the findings in the background of English in India, the work of Mahesh Dattani in plays, and digital means in literary studies.

The evidence validates the view that Dattani uses his plays as effective arenas in the quest of attempting negotiating identity, gender and cultural silences within Indian society. Simultaneously, the computerized analysis shows that the differences among his plays with the help of language can be identified systematically. The combination of close reading and corpus analysis offers a more holistically situated approach than each by itself, creating the perfect balance between an interpretive approach and an empirical one [Khaliq et al. \(2025\)](#).

6.1. ENGLISH AS A MEDIUM OF NEGOTIATION

The thematic analysis revealed that English in the plays by Dattani is not a neutral language of communication but a contentious language in which the postcolonial reaches a representation. He uses English so that a cosmopolitan audience can see the Indian realities, and by simultaneously infusing local terms, names, and references to maintain a cultural specificity, an in-filling and an out-flushing movement can obtain. This duality falls in line with previous critical arguments about the hybridity of the Indian English theatre. This is supported by the computations: the use of kinship terms in *Tara*, conflict terms in *Bravely Fought the Queen*, and trauma vocabulary in *Thirty Days in September* proves that English speech has a solid affiliation with Indian family and social environment.

6.2. COMPUTATIONAL STYLISTICS AND LITERARY INTERPRETATION

The results of the machine learning classification experiment made it clear that such a play as *Seven Steps Around the Fire* could be reliably classified because it contained rather specific vocabulary, whereas the various versions of a play such as *Tara* and *Thirty Days in September* were assigning to the same classification. This overlap is indicative of corresponding themes of trauma and family dynamics reaffirming that computational misclassifications can be interpretive insights in themselves. Instead of explaining these as model errors, they also illustrate thematic connections that literary scholars tend to find. The potential of this method is the ability to use computational results as interpretive data.

6.3. CONTRIBUTION TO INDIAN ENGLISH DRAMA STUDIES

The synthesis of the aim of the thematic, lexical and computational perspective introduces another dimension to the study of Indian English Drama. In earlier work performance and thematic boldness in Dattani plays have been highlighted but this paper gives a systematic linguistic proof. Measuring the use of vocabulary and sentence length and classifications, the paper can add to the knowledge on how the English language is used in Indian theatre plays. It proves that English, instead of being a foreigner has, in fact, been incorporated into the Indian modes of speech in drama and theatrical presentations.

6.4. PEDAGOGICAL IMPLICATIONS

There is also the implication of the methodology on pedagogy. It is possible to allow literature students to interact with texts using reading and computation because of the implementation of a Python-assisted workflow. The most basic frequency counts, collocation checks, and concordance lines can give the students an introduction to the intersection of language and theme, and classification experiments give a chance to consider the way computational models interpret literature. This combination of the field of literature and digital approaches is in harmony with the newer tendencies of higher education where the interdisciplinary approach is being appreciated.

6.5. WIDER APPLICABILITY OF THE STUDY.

The results are also relevant to more general debates in the digital humanities. The use of computational stylistics as applied to the South Asian text suggests a need to make digital humanities studies more diverse, as they have commonly concentrated on Western corpora. This paper will inspire the application of the methods to other postcolonies when the relatively small data can be used to draw conclusions through careful consideration of the methods. The findings also indicate how procedures of computation have to be modified to literary corpora where interpretive subtlety is crucial and misclassifications can be culturally significant.

Table 3

Table 3 Key Insights from Findings and their Interpretive Significance	
Finding (Computational or Corpus-based)	Interpretive Significance
Distinctive vocabulary in <i>Seven Steps Around the Fire</i>	Highlights the centrality of hijra identity and community-specific language
Overlap between <i>Tara</i> and <i>Thirty Days in September</i>	Reflects shared concerns with trauma, disability, and silenced family histories

High frequency of kinship terms in <i>Tara</i>	Confirms focus on parental authority, gendered decision-making, and family power dynamics
Conflict vocabulary in <i>Bravely Fought the Queen</i>	Aligns with themes of hypocrisy, marital discord, and hidden tensions
Longer dialogues in <i>Thirty Days in September</i>	Indicates reflective treatment of abuse, memory, and psychological recovery

The [Table 3](#) summarises the way in which corpus and computational findings are projected to thematic interpretations. It demonstrates that even with a small data set, digital processes generate insights that are consistent with known readings in literature and also develop them with empirical elucidation.

6.6. INTEGRATIVE REFLECTION

This discussion shows that the plays of Dattani represent dialogic relation between language, culture and performance. English is no colonial heritage but it is a global tool with a revitalized vigor with the social situation of India. The computational findings support the idea that computer-based approaches may render stylistic patterns and thematic continuities, yet they also show that interpretation cannot be avoided. The effectiveness of this study is explained by the fact that the computational analysis is not presented as an alternative to close reading but as a tool to reveal multiple semantics of the Indian English play.

7. CONCLUSION

This concluding section summarizes the main arguments of the paper, highlights why this study is significant to the study of Indian English drama and digital humanities, and suggests possible areas of future research. The present analysis of the six key plays of Mahesh Dattani has shown how English has become a force to reckon with cultural bargaining in Indian plays. Close reading demonstrated how the plays were used to dramatize such silences relating to gender, sexuality, family conflict, disability, and trauma. Corpus-based statistics showed patterns in the vocabulary and structure of these plays to indicate a cycle of similarities in motifs of kinship, conflict, and marginality. Using computational classification, unique play signatures were discerned, with thematic similarity manifested in o agency overlaps across plays. Together, this trend of analysis creates a clear understanding of Dattani as a significant climate of changing the manner in which the Indian English drama has been used to address burning issues that still remain unknown in the social realm. The paper has a methodological contribution in addition to the thematic contributions. It demonstrates how digital technology can help to improve the literary reading experience without becoming as complex as algorithms because it adapts corpus-assisted analysis to machine learning. The misclassifications on the computational model were not regarded as errors but as the points of the interrelatedness of the dramatic interests of Dattani. This restates the fact that computational stylistics when applied to an interpretive reading exercise can be extremely bold, yet at the same time, tentative to postcolonial text [Ashcroft et al. \(2002\)](#).

The current research provides a new time-based descent of digital humanities, too, as it implements the use of computational methodologies to South Asian drama, which has been mostly marginalized in the global literature-oriented data-driven analysis. It demonstrates that even relatively small corpora, when well curated and properly put into context, can be used to draw some valuable insights. Meanwhile, it is supportive of more extensive assignments that may involve longer pieces of texts, intra-/inter-authorship comparisons, and the incorporation of more multimodal readings based on an interplay of text and performance. The implication goes further than the field of research and into pedagogy. Bringing in Python-based analysis into literature classrooms provides students with a chance to learn more about both linguistic and digital approaches to language and theme. It is through such integration that the study of literature is not only modernized but also that the students are empowered to engage texts in a critical manner that makes sense to the contemporary research procedures.

In summary, this paper has described the plays of Mahesh Dattani as being representative of the cultural authority of Indian English theatre generals and how effective it is to combine the close reading with the computational analysis. It is not the foreign residue and is interpreted as English when it is observed that English can be considered as a means by which Indian playwrights are capable of facing social facts with the sense of global presence and local relevance. By integrating literary reading with the numerical accuracy of digital studies, this paper enters the emerging discussion about the relationship between postcolonial studies and digital humanities, and yet, it also creates avenues towards further research that aims to perceive literature both in words and numbers, voices and data.

8. LIMITATIONS AND FUTURE SCOPE

This section addresses the limitations of the current study and proposes the ways to expand its contribution as part of future studies. Although the results are important, the research also recognizes research and practice limitations that should be taken into consideration. The first weakness has to do with the size of the corpus. The sample was 38 excerpts with a total of 922 words. Although this was good in showing a demonstration of a proof of concept, more solid evidence would be possible by a larger corpus comprising of the full scripts of the plays by Dattani. Adding more data to it would also allow using more sophisticated natural language processing methods like topic modeling, sentiment analysis, and discourse structure mapping.

A second weakness is associated with the range of computational techniques. The work was largely based on lexical statistics and classification using BERT. Although these methods did provide some useful insights, they are not able to provide richer semantic or pragmatic aspects of conversation like irony, metaphor, or performative force. Further research might employ the multimodal methodology and examine stage directions, performance scripts or even audio-visual recordings of performances so as to comprehend how language interacts with gesture, tone and space/place. The third weakness is the cultural and linguistic diversity. Dattani composes in English, but frequently incorporates a local idiom and local culture. An orthopositive corpus which contains translations of his plays into the Indian languages or a comparison of his work with the playwrights of the region could shed light on the interaction of English with other linguistic traditions in Indian theatre. It would place Dattani more squarely within the multilingual ecology of Indian drama such comparative work would provide.

The fourth limitation is the pedagogical use. Despite the fact this study hypothesizes that Python-assisted methods could be applied in the teaching process, it did not have the opportunity to test the strategy in the classroom. The future perspective is to have curriculum modules in which the literature students can actively apply the computational tools to analyze the texts in order to assess the pedagogical significance of digital humanities to the literary education in India. To conclude, although the current research managed to reveal the usefulness of the combination of close reading with computational analysis, its future is likely to focus on the corpus expansion, the method diversification, and the future correlation of the results with the multilingual background and the pedagogical setting. Such guidelines will make digital humanities even stronger in supporting the postcolonial theatre studies and make sure that the work by Dattani will be studied more inter-disciplinarily and richly.

DATA AVAILABILITY STATEMENT

The data supporting the findings of this study are available from the corresponding author upon reasonable request.

AUTHOR CONTRIBUTIONS

The authors had equal contribution on the conception, design, analysis as well as preparation of this manuscript.

TRANSPARENCY STATEMENT

The authors affirm that the study is reported with complete honesty and transparency, with no significant aspects omitted and all deviations from the original plan fully explained.

CONFLICT OF INTERESTS

None.

ACKNOWLEDGMENTS

None.

REFERENCES

- Ahmad, A. (1992). In *Theory: Classes, Nations, Literatures*. Verso.
- Ashcroft, B., Griffiths, G., and Tiffin, H. (2002). *The Empire Writes Back: Theory and Practice in Post-Colonial Literatures* (2nd ed.). Routledge. <https://doi.org/10.4324/9780203426081>
- Bhatia, N. (2004). *Acts of Authority/Acts of Resistance: Theater and Politics in Colonial and Postcolonial India*. University of Michigan Press. <https://doi.org/10.3998/mpub.17085>
- Das, B. K. (2008). *Mahesh Dattani's Plays: Critical Perspectives*. Pencraft International.
- Dattani, M. (2000). *Collected Plays*. Penguin Books.
- Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2019, June). 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* (Vol. 1, 4171–4186). <https://doi.org/10.18653/v1/N19-1423>
- Dharwadker, A. (2005). *Theatres of Independence: Drama, Theory, and Urban Performance in India Since 1947*. University of Iowa Press. <https://doi.org/10.1353/book6841>
- Evans, M., and Hogarth, A. J. (2021). Stylistic Palimpsests: Computational Stylistic Perspectives on Precursory Authorship in Aphra Behn's Drama. *Digital Scholarship in the Humanities*, 36(1), 64–86. <https://doi.org/10.1093/llc/fqz085>
- Fischer, F., et al. (2020). A Community of Practice Around an Annotated Drama Corpus. *Journal of the Text Encoding Initiative*, 12. <https://doi.org/10.4000/jtei.2894>
- Hoover, D. L., Culpeper, J., and O'Halloran, K. A. (2014). *Digital Literary Studies: Corpus Approaches to Poetry, Prose, and Drama*. Routledge.
- Jockers, M. L. (2013). *Macroanalysis: Digital Methods and Literary History*. University of Illinois Press. <https://doi.org/10.5406/illinois/9780252037528.001.0001>
- Kestemont, M., Karsdorp, F., and Riddell, A. (2022). *Humanities Data Analysis: Case Studies with Python*. Princeton University Press.
- Khaliq, A., Din, B. U., and Hassan, A. H. A. (2025). Corpus-Driven Computational Stylistics: A Postcolonial Analysis of Literary English in the Poetry of Taufiq Rafat. *QRJS*, 2(3), 1072–1086. <https://doi.org/10.63878/qrjs406>
- Kolandaivel, P. (2025). Challenging Patriarchal Structures: Feminist Themes in the Plays of Mahesh Dattani. *International Journal of English Language, Literature and Research Studies*, 95–101. <https://doi.org/10.63090/ijelrs/3049.1894.0016>
- Kuhn, J. (2019). Computational Text Analysis within the Humanities: How to Combine Working Practices from the Contributing Fields? *Language Resources and Evaluation*, 53, 565–602. <https://doi.org/10.1007/s10579-019-09459-3>
- Mee, E. B. (2009). *Theatre of Roots: Redirecting the Modern Indian Stage*. Seagull Books.
- Mukherjee, M. (2000). *The Perishable Empire: Essays on Indian Writing in English*. Oxford University Press.
- Naik, C. V., and Biradar, R. (2024). The Evolution of Indian English Drama: From Rabindranath Tagore to Mahesh Dattani. *ShodhKosh Journal of Visual and Performing Arts*. <https://doi.org/10.29121/shodhkosh.v5.i2.2024.1930>
- Ngũgĩ wa Thiong'o. (1986). *Decolonising the Mind: The Politics of Language in African Literature*. James Currey.
- Rajan, R. S. (2023). English in India, India in English. In *The Oxford Handbook* (xx-xx). Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780197647912.013.4>
- Sahmoud, T., and Mikki, M. (2022, June). Spam Detection Using BERT. *arXiv*. <https://doi.org/10.48550/arXiv.2206.02443>
- Šeja, A., and Nagy, B. (2022). Supplementary Code and Data for the Paper "From Stage to Page: Language Independent Bootstrap Measures of Distinctiveness in Fictional Speech." *Zenodo*. <https://doi.org/10.5281/zenodo.7383687>
- Semino, E., and Short, M. (2004). *Corpus Stylistics: Speech, Writing, and Thought Presentation in a corpus of English writing*. Routledge. <https://doi.org/10.4324/9780203494073>
- Sinha, M. (2020). Interrogating Women's Silence in Select Plays of Mahesh Dattani. *Rupkatha Journal on Interdisciplinary Studies in Humanities*, 12(5). <https://doi.org/10.21659/rupkatha.v12n5.rioc1s21n2>
- Underwood, T. (2019). *Distant Horizons: Digital Evidence and Literary Change*. University of Chicago Press. <https://doi.org/10.7208/chicago/9780226612973.001.0001>
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., and Polosukhin, I. (2017). Attention is All You Need. *Advances in Neural Information Processing Systems*, 30.

- Wehrli, C., and Gius, E. (2023). Computational Stylistics in Practice: A Scalable Analysis of Thematic and Stylistic Affinity in 19th-Century Fiction. *Digital Scholarship in the Humanities*, 38(1), 1–19. <https://doi.org/10.1093/llc/fqac040>
- Zhu, H., and Lei, L. (2019). Style, Computers, and Early Modern Drama: Beyond Authorship. *Australian Journal of Linguistics*, 39(4), 539–542. <https://doi.org/10.1080/07268602.2018.1507617>