

Linguistic Stereotyping, Gendered Representation, and the Shifting Portrayal of Women in Hindi Cinema: A Sociolinguistic Corpus Study

Manjri Suman

Assistant Professor

M.A., M.Phil., PhD in Linguistics

English and Modern European Languages.

Gautam Buddha University

Greater Noida, Uttar Pradesh, India

manjri@gbu.ac.in

Riya Raj

Assistant Professor

M.A., M.Phil., PhD in Linguistics

English and Modern European Languages.

Gautam Buddha University

Greater Noida, Uttar Pradesh, India

Srijana Jaiswal

Assistant Professor

M.A.in English

English and Modern European Languages.

Institution Name: Gautam Buddha University

Greater Noida, Utter Pradesh, India

Abstract

Hindi popular cinema in its studio era as well as in its OTT era has been an enduring medium through which speaker types have been socially encoded. The paper uses a bounded corpus of 2025 Hindi language feature films (40 theatrically released and two long-form web series) to explore how the language and paralinguistic decisions create and produce stereotypes of regions, religion, class, occupation, and gender, and how these decisions have changed in the OTT era. The unit of analysis is the character-scene, which is an individual character in an identifiable scene; utterances, accent features, lexical selection, address form and code-mixing are coded as embedded units. Operationalization of six stereotype categories is done by explicit indicator sets and not by impressionistic labels. Gendered language is distinguished from sexist language as the former is grammatical or referential, the latter evaluative and derogatory, placing women in an inferior, subordinate, or consumable position. Dialogue in Hindi–Urdu language is consistently romanised, and each cited example includes a romanised original, an English translation, identification of the speaker, and context for the scene. The paper shows that, while the stereotype grammar of Bihari-as-comic, Christian-as-westernized, English-speaking-woman-as-morally-suspect has been somewhat replaced, it has not been eliminated from the OTT corpus and that gendered-language markings continue to be used even though overt sexist framings are absent.

Keywords: Hindi Cinema, Sociolinguistics, Linguistic Stereotyping, Gendered Language, Sexist Language, OTT, Film Corpus

“Indian cinema is one of the oldest, most vibrant, and diverse forms of storytelling in the world. It has equally captivated generations of Indians and non-Indians with its enchanting stories, colorful characters, and captivating visuals” (Sahai et. al, 2024,). Cinema is considered as one of the most powerful art forms and a highly engaging medium of entertainment for

people. It is kind of a pause from the realities of daily lives and that takes them into a different dreamy world. In other words, we can say that for some, cinema is a break from the reality of a painful life, while for others, it is a source of comfort. When we watch a film, we do not merely observe; we participate, feel and inhabit it, and sometimes act as a dead spectator with complete emotional surrender. Cinema creates an illusion of continuous motion, by telling stories, evoking emotions or conveying an idea understating its cultural, philosophical and artistic significance.

Across roughly a century of production, it has taught audiences how a “good” woman speaks, what a Bihari character sounds like, what a Christian home looks like, and which languages travel with virtue and which with vice. Two working assumptions run through this paper. The first is Morin Edgar’s (1956) claim that we accept cinematic characters with such intimacy that, in the moment of viewing, our own awareness partly dissolves into theirs. The second is Christian Metz’s (1974) argument that film communicates through codes, systems of signs whose meanings are learned, so that a Bihari accent, an English exclamation, or a sleeveless blouse arrives on screen already loaded with connotation.

A previous paper on the changing portrayal of women in the OTT era (Suman, 2023) sit behind this study. The present paper attempts to fill the gaps of the previous version and also add a wider spectrum to the categories of stereotyping with substantive arguments.

Literature Review

Cinema, Language, and the Semiotics of Connotation

Metz (1974) argued that cinema is a signifying practice in which meaning is produced not only at the denotative level of what is shown but, more decisively, at the connotative level of what is culturally read into what is shown. A Bihari phonological substitution such as [f] → [p^h] is denotatively a phonetic feature. Connotatively, decades of repetition have made it a code for provinciality and comic incompetence. Bazin’s (1967) view that film has an “ontological

bond” with reality is useful as a foil: even if the image records the world, the frame, the accent, and the cut carry ideological weight.

Sociolinguistics has independently arrived at a similar conclusion. Labov’s (1966) New York City study showed that variation is systematic and stratified, and that non-standard varieties carry overt stigma while standard varieties carry overt prestige. Applied to Hindi cinema, the “standard” has historically been the register of educated, upper-caste, upper-middle-class North Indian speakers of Khari Boli Hindi with a modest English admixture. Everything else—Bhojpuri, Maithili, Awadhi, various South Indian Englishes, Bombay Hindi—has been available for marking.

Gender, the Gaze, and Feminist Film Theory

Mulvey (1975) proposed that mainstream cinema is organised by an asymmetric visual economy in which the active look is coded as masculine and the object of the look as feminine. Women, in her formulation, connote “to-be-looked-at-ness.” Spivak (1988) extends this by asking whether the subaltern woman can speak within representational regimes that have already spoken for her. The director the films were mostly men till the late 90s. Women almost never got a chance to tell their own stories or even represent themselves on screen. Thus, the female characters seldom had a say in their portrayal rather their character in the plot was woven alongside the male protagonist. Writing on film noir, Harvey (1978) notes that the family in film functions as a legitimising device that positions women’s value within a kinship economy, so that a woman’s standing in the household stands in for her standing in society. Bordo (1986) further argues that the historical identification of rationality with masculine modes of distance has produced a cultural crisis whose resolution requires the articulation of a “feminine” mode across epistemology, science, and politics. Roberge’s (1992) applies these lines of thought to Indian cinema, and Sibal (2018) offers a specifically Indian account of stereotyping. There is a persistent good-woman / vamp dichotomy in Bollywood (Sibal, 2018).

The current work uses these frameworks to analyse the wide range of data that are part of this study. It builds on some of these researches but shifts the emphasis from thematic critique to sociolinguistic coding.

Method

This section addresses, in order, the four methodological points on which the present paper is explicit: (1) a bounded film corpus with reported sampling and inclusion criteria; (2) a stated unit of analysis; (3) operational definitions for each stereotype category; and (4) a working distinction between gendered language and sexist language.

The Film Corpus

Population and Sampling Frame

The population of interest is Hindi-language popular feature films with theatrical or major-OTT release, plus long-form Hindi-language web series with a comparable production footprint. The sampling frame was Bollywood Hungama's all-India theatrical release database supplemented by BookMyShow archives for the theatrical years, and the JustWatch India catalogue combined with platform release calendars (Netflix India, Amazon Prime Video India, Disney+ Hotstar, SonyLIV, ZEE5) for the OTT years.

Sampling Period

The sampling window runs from 1 January 1965 to 31 December 2025, sixty-one calendar years. The 1965 start captures Khandan, which is a foundational text for the "wife-as-worshipper" frame. The 2025 endpoint reflects the OTT-era shift documented in Suman (2023) and includes films released up to the corpus lock date of 31 December 2025.

Selection Procedure

Films were selected through purposive stratified sampling. Two strata were defined: (a) commercially successful mainstream releases and (b) critically discussed or industry-recognised films that entered public commentary on gender or community representation.

Within each stratum, films were then filtered against the inclusion criteria below. Two long-form web series (Panchayat and Gullak) have been retained because prior work identifies them as inflection points in the OTT-era shift.

Genre Distribution

Family drama and romance dominate the pre-2010 sub-corpus, reflecting Bollywood's commercial spine in that period. The post-2010 sub-corpus is more mixed, with realist drama, biographical film, thriller, and long-form serialised drama all represented. Item songs, treated as embedded units within their host films, were coded separately where relevant.

Commercial Reach

Each film was among the top-thirty grossers of that year. OTT films and series were included if they figured in the Top 10 India lists on their respective platforms in the four-week window after release, using platform-published data where available.

Language Variety

Primary language of the film is Hindi–Urdu, with heavy code-mixing with English typical of urban Bollywood dialogue, in some of the films. Films dubbed into Hindi from South Indian originals were included only where the Hindi dub, and not the original, is the object of analysis (relevant for regional stereotyping in dub tracks). Films whose primary dialogue language is a non-Hindi Indian language were excluded.

Release Year

Distribution across decades is uneven by design, reflecting the argument that stereotype grammars intensify or shift by decade. The corpus contains 6 films from 1965–1989, 14 from 1990–1999, 10 from 2000–2014, and 10 from 2015–2025, plus 2 web series (2020–2024). This is 40 films and 2 series in total ($n = 42$ texts).

Inclusion Criteria

A text was included if it (i) had a Hindi-language primary dialogue track, (ii) was released between 1 January 1965 and 31 December 2025, (iii) met the commercial-reach threshold above or entered documented public commentary on gender or community representation, and (iv) contained at least one identifiable character-scene involving either a gendered speech act, a regionally, religiously, class-, or occupationally marked speech act, or an English/Hindi code-choice with an evaluative frame.

Exclusion Criteria

Excluded were (i) parallel-cinema releases with negligible commercial reach (ii) short films under thirty minutes, (iii) documentaries and non-fiction, (iv) films whose primary dialogue language is not Hindi–Urdu, and (v) remakes where the analytical target overlaps entirely with an already-included original.

Unit of Analysis

The unit of analysis is the character-scene: a single named or clearly identified character within an identifiable scene, treated as the atom of coding. Within each character-scene, five embedded units are coded where present: (a) individual utterances, romanized and translated as described below; (b) accent features, coded as phonological, prosodic, or paralinguistic; (c) lexical choices, including code-mixed items and marked address forms (e.g., dikra, anna, akka, saab, bhaiya); and (d) narrative function of the character (protagonist, antagonist, comic relief, mentor, love interest, foil). A film with N coded character-scenes contributes N observations, not one. This aligns the study with utterance-level sociolinguistic work rather than with holistic film-thematic criticism.

Operationalising the Stereotype Categories

Six categories are operationalised. Each is anchored to observable indicators so that coding does not depend on the analysts' general impression of the film.

Regional Stereotyping

A character-scene is coded as regionally stereotyped if it exhibits at least two of the following indicators, of which at least one must be linguistic: (a) marked phonological substitution associated with a regional variety (e.g., [f] → [p^h], sibilant merger for Bihari; retroflex-heavy consonantal profile and [ə] → [a] for stereotyped South Indian English); (b) formulaic lexical items indexing region (ayyo, anna, akka for South Indian; ee, hamaar, ba for Bhojpuri-inflected Hindi); (c) exaggerated prosodic contour beyond normal regional variation; (d) visual co-marking (dress, food, prop) that reinforces the linguistic marking; (e) narrative role of the regionally marked character as comic relief, minor antagonist, or non-central role.

An illustration of the format: in *Chennai Express* (2013), Meenamma, played by Deepika Padukone, says *nako, ayyo aisa mat kaho* (“no, ayyo, don’t say it like that”) to Rahul during their escape sequence in the interior compartment. The feature under analysis is the formulaic South-Indian English discourse marker *ayyo* transposed into Hindi dialogue as a portable identity tag.

Religious Stereotyping

A character-scene is coded as religiously stereotyped if it exhibits at least two of: (a) marked religious address terms or exclamations used non-diegetically for identification (Jesus, Ya Allah, Waheguru placed in dialogue where a neutral exclamation would serve); (b) marked lexical or grammatical features associated with a community-inflected variety (Urdu-heavy diction with *shayari* performance for Muslim characters; Hindi–English mixing marked for Christian characters with post-verbal *karega*, *aayela*, etc.); (c) visual co-marking (skull cap, cross, turban, *sudreh*) coordinated with linguistic marking; (d) narrative role as sidekick, comic figure, or antagonist coded through the religious identity; (e) family or community background presented as culturally out-of-step with a Hindu default.

Gendered Stereotyping

A character-scene is coded as gendered-stereotyped if it exhibits at least two of: (a) role assignment restricted to domestic sphere or defined through kin relation to a male character; (b) evaluative framing of the character's worth through modesty, silence, or self-erasure; (c) dress, exercise, English-use, or nightlife participation coded as evidence of moral fall; (d) address forms and pronoun choices that position the woman as recipient of the male gaze (see section on gendered vs. sexist language); (e) narrative resolution that returns an autonomous woman to marriage, family, or death.

Caste-adjacent Stereotyping

Caste is rarely spoken plainly in mainstream Hindi cinema, so this category is operationalised through proxies: (a) surname or occupational marker used to fix community identity (Pandit-ji, Thakur, Yadav, sweeper, night-soil worker); (b) skin-tone lighting and colour grading co-varying with narrative role; (c) speech register (unmarked Khari Boli Hindi vs. marked rural or working-class register); (d) narrative arc that places the character in a service, comic, or victim role tied to the proxy. A character-scene is coded as caste-adjacent stereotyped if at least two of these are present.

Class Stereotyping

Coded through: (a) English fluency treated as an evaluative marker rather than a background fact; (b) consumption of Western clothing, alcohol, or nightlife as narrative evidence of class positioning; (c) domestic labour represented without agency; (d) speech register aligned with the class marking. At least two indicators, of which at least one linguistic, are required.

Gendered Language vs. Sexist Language

The two are related but not identical, and this paper treats them separately.

Gendered language refers to structural and referential marking of gender in Hindi–Urdu discourse: grammatical gender agreement on verbs, adjectives, and postpositions; gendered address forms and honorifics (beta, bahu, ji, saahiba); pronoun choice and the pragmatics of second-person forms (tu, tum, aap); and lexical items that carry gendered load without being evaluatively derogatory (pativrata, ghar ki lakshmi). Gendered language is not, by itself, sexist. It becomes analytically interesting for this paper when its distribution across character-scenes co-varies with the other stereotype indicators.

Sexist language refers to evaluatively loaded usage that positions women as inferior, subordinate, consumable, or as objects of moral surveillance. It includes (a) derogatory vocabulary and slurs (besharam, kulta); (b) sexualising figuration in dialogue and lyrics (tandoori murgi, patakha, bomb ka dhamaka); (c) evaluative predications where female intelligence is treated as an exception (aaj pehli baar kisi aurat ne samjhdaari ki baat ki hai, from Ready, 2011); and (d) speech acts of policing (aurat ki tarah rehna seekhiye, from Bulandi, 2000). Separating the two allows the paper to say something more precise than “the language is patriarchal.” Some scenes are gendered without being sexist; some are sexist even when their grammatical gender marking is unremarkable.

Analysis

Gendered and Sexist Language in Practice

Applied to the corpus, the distinction above yields three patterns. First, the pre-2000 sub-corpus shows dense co-occurrence of gendered language and sexist framing. In Khandan (1965), the song tumhi mere mandir, tumhi meri puja, tum hi devta ho (“you are my temple, you are my worship, you are my god”), performed by the wife character, uses honorific plural forms addressed to the husband while the reverse is not required. The gendered language here is grammatical; the sexist frame is the deification of the husband as a normative feminine posture.

Second, the Sooraj Barjatya cycle of the 1990s formalises this into a household aesthetic. The *Hum Saath Saath Hain* (1999) line, *jahaan khaane ke table par hansi thitholi hoti ho, jahaan ki ladies aur bachchiyaan humein pyaar se khaana khilaayein, wahi ghar ghar hai* (“the house where there is laughter at the dining table, where the ladies and daughters lovingly serve us food, that is a real home”), spoken by a senior male family member at a family gathering, defines the “home” through female service. The grammatical marking is unremarkable; the sexist load is in the evaluative predicate.

Third, the 2011 *Ready* line, *aaj pehli baar kisi aurat ne samjhdaari ki baat ki hai* (“today, for the first time, a woman has said something sensible”), spoken by Salman Khan’s character in response to the female lead Asin’s remark, is sexist without being especially gendered in its grammar. Female intelligence is framed as an exception to a norm.

The 2000 film *Bulandi* contains a paradigm case: the line *aurat paida hui hain to aurat ki tarah rehna seekhiye* (“if you are born a woman, learn to live like one”), addressed to Raveena Tandon’s character, is a policing speech act delivered under the sign of tradition. What is analytically important is that when the female lead responds by slapping her interlocutor, the reprimand comes not from a man but from Rekha’s character: *aurat hoke mard pe haath uthati ho?* (“being a woman, you raise your hand on a man?”). Patriarchal norms are enforced through women too.

Item songs concentrate the sexist register. *Fevicol Se* (*Dabangg 2*, 2012) contains main *to tandoori murgi hoon yaar, gatka le saiyaan alcohol se* (“I’m a tandoori chicken, my love, gulp me down with alcohol”), performed by the female lead in a public performance frame within the film. The figuration equates the female body with a consumable dish. *Munni Badnam Hui* (*Dabangg*, 2010) and *Beedi Jalaile* (*Omkara*, 2006) show similar structures where the female body is a resource offered up to a masculine consumer position.

Regional Stereotyping

Bihari

Bihari characters in the corpus cluster around a small set of features: substitution of [f] with [p^h] (phive for five in *Hum Hain Rahi Pyaar Ke*, 1993); sibilant merging; Bhojpuri-marked lexis (ee hamaar ba); and a narrative role that is comic, minor-antagonistic, or, in dubbed South Indian material, villainous. The PK (2014) song title *Love ij bhaste* of time trades explicitly on the [f]/[p^h] and [z]/[j] substitutions. In *Super 30* (2019), Hrithik Roshan's portrayal of Anand Kumar, the mathematician from Patna, was widely criticised for over-marking the Bihari accent, which is exactly the criticism the coding scheme predicts: the linguistic marker is doing narrative work beyond what the biographical role requires. Moreover, the complexion of Bihari characters portrayed in Hindi Cinema is also darkened. Be it Hrithik Roshan playing, Anand Kumar of *Super 30*, or Alia Bhatt's character in *Udta Punjab*.

Another regional stereotyping of Bihar is found in the Hindi dubbing South Indian movies. In many a films, the negative, rustic, villainous character speaks in Bihari-UP dialect or accent reaffirming the false general perception that these people are uneducated and goons.

South Indian

Coding across the corpus flags the same three recurring devices: ayyo as a portable exclamation, kinship terms *anna* and *akka* used as address, and English pronounced with retroflex-heavy consonants and rhoticised endings. *Chennai Express* (2013) does this with Meenamma, who repeats *nako* and *ayyo* across scenes; *Hum Hain Rahi Pyaar Ke* (1993) does it with the Iyer family. In both, the linguistic feature is the vehicle for a narrative role.

Religious Stereotyping

Muslim characters

The pre-2000 corpus tends to place Muslim characters in one of two roles: the loyal Mamuji sidekick or the antagonist. The linguistic marking layers Urdu-heavy diction and

shayari performance onto visual co-markers (skull cap, kohl, embroidered kurta). Post-2010, the corpus records a partial shift: Kabir Khan in *Chak De India* (2007), Rizwan Khan in *My Name Is Khan* (2010), Danish Ali in *Wazir* (2016), Dr. Jahangir Khan in *Dear Zindagi* (2016), Murad Ahmed in *Gully Boy* (2019), and Sehmat in *Raazi* (2018) are protagonists whose religious identity is present but not narratively totalizing and they are not reduced to their religious identity.

Christian characters

The Christian speech caricature is grammatically distinctive. Post-verbal *karega*, *aayela*, and *vaaste* appear in dialogue that is otherwise Hindi (main God se pray karega, “I will pray to God”; hum tumhare vaaste chicken banake laya, “I made chicken for you”). The scheme codes these as marked grammatical features co-occurring with religious address (Jesus tumhara help karega, “Jesus will help you”). Coordinated visual and dietary co-markers (alcohol, non-vegetarian food, Western dress) complete the caricature. *Cocktail* (2012) and *Grand Masti* (2013) add sexualisation as a further layer, placing Christian women in the hyper-sexual counterpart position to the traditional Hindu female lead.

Sikh and Parsi characters

Sikh characters in *Mohabbatein* (2000), *Good Newz* (2019), and *Jo Bole So Nihaal* (2005), and Parsi characters in *Munna Bhai M.B.B.S.* (2003), *Ishq* (1997), and *Dhamaal* (2007), are coded through exaggerated accent, flamboyant dress, and comic-relief narrative role. The Parsi *dikra* address form is analytically parallel to the South Indian *ayyo*: a portable identity tag.

Class, Caste-adjacent, and Occupational Stereotyping

Class marking most often runs through English fluency and Western consumption. In the pre-2010 corpus, the character who drinks juice at breakfast, exercises, and speaks in English is coded as morally suspect if female (Rang, 1993; Beta, 1992; Judaai, 1997) or as

arrogant if male. Caste-adjacent stereotyping is largely mediated through occupational proxy and skin-tone lighting rather than explicit naming, which is why the coding scheme reads it through those proxies. Occupational stereotypes—the brusque Hindi-heavy police officer, the English-clipped doctor, the didactic teacher—recur across decades.

Aversion Against English

The historical treatment of English in Hindi cinema is a case where language ideology becomes character coding. In *Maine Pyar Kiya* (1989), the antagonist's English exclamations mark him as arrogant, while the heroine's abstention from English marks her as virtuous. *Raja Hindustani* (1996) uses the stepmother's English fluency in the same way. The character believes herself to be modern but to the audience she is a vamp. *Bulandi* (2000) turns the English/Hindi choice into a direct evaluative axis. And in *Kabhi Khushi Gham* (2001), we see that Rohan and his friends mock Puja and looks down upon the likes of puja just for speaking Hindi. *Cocktail* (2012) modernises the same code without abandoning it: Veronica, the English-speaking, drinking, partying character, is contrasted with Meera, the Hindi-speaking traditional woman, and the male lead chooses Meera for marriage. Although in a different context, Om Puri's reported remark, *angrezi koi hauwwa nahi; yeh bhi Punjabi ya Hindi ki tarah ek bhasha hi hai* ("English is not a big deal; it is just a language, like Punjabi or Hindi") (Shekhar, 2016, p. 33), reads as an internal industry protest against this coding.

Changing Patterns in the OTT Era

The films and web series released in the OTT era (2018-25) exhibit a partial shift but not a complete break. In the coded data there are four shifts. First, the binary of good-woman / vamp relaxes. Films like *Kapoor & Sons* (2016), *Dear Zindagi* (2016), *Lipstick Under My Burkha* (2017), *Thappad* (2020), *Darlings* (2022), and *Laapata Ladies* (2024) have women characters in ethically ambiguous situations whose speech is not used as a binary of either proof

or evidence of good or bad. Alia Bhatt's role in *Kapoor & Sons* drinks up, wears short clothes, has kissed with another man and none of these things are employed to make her a bad character. Second, the older mother figure, who appears to have lost her children and is dressed in white, the censorious hue of the era, is not the one to disappear, but characters such as Shefali Shah's in *Once Again* (2018) whose life goes on after their children grow up, whose romance in her later years isn't the scandal but a logical path. Thirdly, the working woman does not come with the old "moral charge". *The Family Man* (2019-2021) is a show that frames the professional relationship between two women as a personal relationship, a violation of which the husband is scolded by his colleague. *Panchayat* (2020–) depicts a rural woman as the head of the village, a position she may hold in name but not necessarily in fact, and a husband who is capable of influencing whether his daughter marries in a certain manner or not. The girl does not like the boy for some reason - he's "weird," and Gullak takes her feelings as gospel. Fourth, regional and religious stereotyping diffuses the source of writing and direction within the community involved. It's not a setting for laughter, it's the Purvanchali worlds of *Panchayat* and *Gullak*. In *Raazi* (2018), *Gully Boy* (2019) and *My Name Is Khan* (2010), Muslims are not defined by their faith alone as the leading characters of these films. The older Sikh character is displaced in serious roles in *Rocket Singh: Salesman of the Year* (2009), *Singh Is Kinng* (2008) and *Kesari* (2019). The changes are substantial yet partial. Major commercial success in the OTT-era window, *Kabir Singh* (2019) and *Animal* (2023) have both the old stereotype grammar, and sexist framing of it, on a stronger level than most films in the 90s. Item songs continue to figure women as consumables. Even today, in the mainstream cinema, the Bihari type of characters is still more often and more deeply marked for laughter, than for giving them their interior life. This paper does not uncritically celebrate the downstream effect of the lack of theatrical cuts with the release of foul language on OTT platforms, as young viewers have reported knowing by heart the abuse repertoire from *Mirzapur* (2018–).

Discussion

Two things follow from the analysis. First, the empirical case for reading Hindi cinema as a system of connotative codes, in Metz's sense, is strong. The distribution of accent features, address forms, code-mixed items, and lyric figuration across the corpus is not random. It aligns with narrative role in ways that are stable across decades even where surface content changes. The point is not that any single film is guilty; it is that the grammar of coding is inherited and reproduced. Second, some post-2015 films have started dropping the sexist language and reduced over-pronounced gender roles. Thus, the OTT-era corpus does divert from being straightforwardly sexist.

Limitations are worth naming. The corpus is purposive rather than probabilistic, and inter-coder reliability was not tested with an independent second coder, though the operational definitions above are written so that such testing is possible. Digital-only releases are under-represented for the earliest years for want of records. Item songs are treated as embedded units within their host films rather than as separate texts, which understates their weight. The paper also does not attempt causal claims about audience reception; the connotative reading is inferred from repeated coding within the texts themselves, not from viewer response data.

Conclusion

Hindi cinema has, for most of its history, run a fairly efficient sorting device. A Bihari accent produced laughter. A Christian character produced a westernised interior. An English-speaking woman produced moral suspicion. A wife who spoke back produced a corrective from an older woman. These are not accidents of individual films; they are the working parts of a coded language that Metz's framework and Labov's make visible in different ways.

The OTT-era shift is a partial dismantling of this device. It has not been dismantled by OTT alone; the paper does not claim that. It has been dismantled by a combination of new writers, new platforms, an audience that has changed at least somewhat, and a set of feminist

and subaltern arguments that have been available for decades and are, finally, showing up in commissioned work. The good-woman / vamp binary is looser than it was. Working women appear on screen. Widowed mothers are permitted to live. Regional worlds are, sometimes, permitted to be the default rather than the marked case.

What has not shifted is the underlying structure of coding itself. The industry still speaks in codes; the codes have simply been partially replaced. That is why a paper of this kind matters. Bounding the corpus, stating the unit of analysis, operationalising the categories, and distinguishing gendered from sexist language are not methodological ornaments. They are the minimum conditions under which the argument can be tested by someone else against a different corpus, and either confirmed or corrected.

Conflict of Interest: The corresponding author, on behalf of all the authors, declares that there are no conflicts of interest regarding the publication of this article.

Copyright: © 2026 Manjri Suman, Riya Raj, Srijana Jaiswal authors retain the copyright of their original work while granting the journal the right of first publication.

License: This work is licensed under the Creative Commons Attribution 4.0 International (CC BY 4.0) License. This license permits others to copy, distribute, remix, adapt, and build upon the work, including for commercial purposes, provided appropriate credit is given to the original authors. The authors are also free to archive and share the published work in institutional repositories, personal websites, academic social networking platforms, and other non-commercial or commercial platforms in accordance with the terms of the CC BY 4.0 License.

References

- Bazin, A. (1967). The ontology of the photographic image. In *What is cinema?* (Vol. 1, pp. 9–16; H. Gray, Trans.). University of California Press.
- Bordo, S. (1986). The Cartesian masculinization of thought. *Signs*, 11(3), 439–456.
- Bordwell, D. (1985). *Narration in the fiction film*. University of Wisconsin Press.
- Green, L. (2002). *African American English: A linguistic introduction*. Cambridge University Press.
- Harvey, S. (1978). Woman's place: The absent family of film noir. In E. A. Kaplan (Ed.), *Women in film noir* (pp. 22–34). British Film Institute.
- Labov, W. (1966). *The social stratification of English in New York City*. Center for Applied Linguistics.
- Metz, C. (1974). *Film language: A semiotics of the cinema* (M. Taylor, Trans.). Oxford University Press.
- Morin, E. (1956). *Le cinéma, ou l'homme imaginaire* (L. Mortimer, Trans., 2005). University of Minnesota Press.
- Mulvey, L. (1975). Visual pleasure and narrative cinema. *Screen*, 16(3), 6–18.
- Roberge, G. (1992). *The subject of cinema studies: Film theory and interpretation of films*. Ajanta Publications.
- Sahai, V., et al. (2024). Exploring Indian cinema: Languages, cultural diversity, and cinematic art. *International Journal of Innovative Research in Technology*, 11(2).
- Shekhar, M. (2016). *Cinema ka safar*. Rajkamal Prakashan.
- Sibal, V. (2018). Stereotyping women in Indian cinema. *Scholarly Research Journal for Interdisciplinary Studies*, 5(43).
- Spivak, G. C. (1988). Can the subaltern speak? In C. Nelson & L. Grossberg (Eds.), *Marxism and the interpretation of culture* (pp. 271–313). University of Illinois Press.

Suman, M. (2023). Changes in the portrayal of women in the OTT era and the shifting scales of language. *Streekal*, 5, 79–83.

Tecucianu, D. (2014). The ‘language’ of cinema: Film ‘grammar’ and visual literacy. *International Journal of Communication Research*, 4(3), 251–256.