Hindis: Gender Switches (Or Not?)

Thakur Anjeev Prakash

Thakur Anjeev Prakash University of Delhi

How to Cite this Article:

Prakash, T. A. (2026). Hindis: Gender Switches (Or Not?). International Journal of Creative and Open Research in Engineering and Management, 02(04). https://doi.org/10.55041/ijcope.v2i4.807

https://doi.org/10.55041/ijcope.v2i4.807

Abstract

Hindi is one of the largest languages in India, spoken over a wide area from the western borders of India to the eastern most parts of India, namely Arunachal Pradesh. Though, Hindi as a mother tongue, in the “Hindi Belt” is not uniform. There are multiple varieties of the language spoken in different parts of India. This paper studies the contrast between two varieties, the Standard variety and the variety spoken in the regions of eastern Uttar Pradesh and large parts of Bihar. These two varieties contrast in terms of markedness in gender agreement. This paper seeks to study how a native speaker of one variety switches to another variety (or not) when given a cue of gender (un)markedness using a 2x2x2 self-paced reading SPR experiment.

Introduction

As the largest spoken language spoken in India over a large area, Hindi tends to assimilate itself with the regional languages and form a new variety. While Modern Standard Hindi (MSH) remains the only variety used for official and educational communication, other varieties developed while being in contact with regional languages like Bihari languages form the colloquial Hindi spoken in the region. The variety of Hindi spoken there develops features of the regional language and forms the major spoken language there over MSH.

Hoernle (1880) classifies the variety spoken in eastern Uttar Pradesh and Bihar as Eastern Hindi. Eastern Hindi (EH) does not have overt gender markings for inanimate subjects (see 2) unlike the Standard or the Western variety (MSH) (see 1). Though EH speakers use no gender marking while speaking, they are in constant touch with the western variety in the form of literature, movies and medium of instruction in school. This paper aims to study how the native speakers of Eastern Hindi and Western Hindi process gender agreement while they are faced with either variety.

1. mer-ii kiitaab ach-ii hai MSH variety
my-F book.F good-F AUX
'My book is good.'

2. meraa kiitaab achaa hai EH variety
my book good AUX 'My book is good.'

Alok et. al. (2013) classifies Bihari Hindi or Eastern Hindi (EH) as a mixed language as they study the morpho-syntactic contact based evidences to find that EH variety spoken in and around the city of Patna is a mixed language with Noun Phrases (NP) borrowed from Magahi retaining its original features from the language and the Verb Phrase (VP) using the structure of MSH. Kumar et. al. (2021) also study the Eh variety spoken in Magahi region and conclude the EH to be a mixed language with VP-NP distinction.

Though the speakers of EH lack gender markedness in their variety, they switch between the two varieties in between sentences and in discourse. They use agreeing verbs or adjectives in one clause or sentence and not in the following clause or sentence for the same NP. This paper aims to find the acceptability of grammatical and ungrammatical agreeing verbs or adjectives when they are given MSH cues early on in the sentence in an SPR experiment by the native speakers of EH. The experiment also studies the acceptability of the same by the MSH speakers in comparison to their EH peers.

Hypothesis

Since the EH speakers have constant exposure to the MSH variety, it can be fairly concluded that they would have no difficulty in processing the agreeing final verb or adjective for both the agreement systems where there is an overt markedness of gender agreement or lack thereof in short and long interspersed sentences. Whereas, the MSH speakers will have a longer processing time when faced with lack of overt gender markedness in the final agreeing verb or adjective in both short and long interspersed sentences.

Literature Review

Hindi is a macro-language covering a wide geographic area in India, consisting of multiple regional varieties that differ significantly in their grammatical structures. August Friedrich Rudolf Hoernle (1880) provided an early systematic classification of these languages. He divided them into Western Hindi and Eastern Hindi. Modern Standard Hindi (MSH) is based on the Khari Boli dialect, which is a Western variety. MSH enforces an overt two-way grammatical gender system. All nouns, including inanimate objects, are assigned either a masculine or feminine gender. MSH requires strict morphosyntactic agreement, meaning verbs, adjectives, and certain postpositions must agree with the head noun in gender, number, and person.Eastern Hindi varieties are spoken in eastern Uttar Pradesh and Bihar, and they include languages such as Bhojpuri and Magahi. These varieties do not have overt gender markings for inanimate subjects. For speakers of Eastern Hindi, the gender assignments of MSH are arbitrary. They do not natively produce these agreement markers in their everyday speech. However, Eastern Hindi speakers are exposed to MSH continuously through literature,

media, and the education system.

Continuous contact between regional Bihari languages and MSH has resulted in structural changes. Alok, Lahiri, and Kumar (2013) demonstrate that the variety of Hindi spoken in Bihar is a mixed language. Their linguistic analysis shows that Bihari Hindi combines elements from both MSH and local languages like Magahi. Specifically, the Noun Phrase (NP) in Bihari Hindi is borrowed from Magahi, which means the noun morphology lacks noun classes, number agreement, and ergativity.In contrast, the Verb Phrase (VP) in this variety uses the structure of Western Hindi. Kumar, Lahiri, and Alok (2021) provide further empirical evidence classifying the Eastern Hindi spoken around Patna as a conventionalized mixed language. This hybrid structure gives speakers a flexible internal grammar. They are accustomed to switching between agreement models during discourse. While this flexibility aids general comprehension, it complicates the processing of strict MSH gender agreement rules.

The interaction between MSH and Eastern Hindi involves sociolinguistic hierarchies. In northern India, schools treat MSH as the standard and prestige variety. LaDousa (2010) describes how the school system establishes language ideologies that elevate MSH while marginalizing regional mother tongues. Because MSH is the medium of instruction, it is strongly associated with education and correctness.This prestige influences how speakers evaluate language. When Eastern Hindi speakers read or hear MSH, they apply a sociolinguistic filter. Even if a sentence contains an agreement violation, the presence of standard MSH vocabulary and feminine possessive markers prompts them to judge it as acceptable. This dynamic creates an acceptability paradox. There is a gap between their real-time cognitive detection of a grammatical error and their offline, conscious judgment of the sentence. The prestige of MSH overrides the cognitive signal of the morphosyntactic violation.

Psycholinguistic research measures how speakers process language in real time. The traditional assumption in linguistics is that readers build complete and accurate representations of sentences. However, Ferreira, Bailey, and Ferraro (2002) proposed the "Good-Enough" approach to language comprehension. Their research shows that language processing is sometimes partial. Readers frequently develop shallow or superficial representations, especially when faced with complex sentences. To save cognitive resources, the parsing mechanism uses simple heuristics to extract a meaning that is "good enough" for the task.This framework is directly applicable to the processing of long-distance dependencies. When a subject and an agreeing verb are separated by several clauses,readers must retrieve the noun's grammatical features from working memory. This retrieval process is prone to errors. Bhatia and Dillon (2022) studied agreement attraction in Hindi and found that speakers frequently make errors when a distractor noun interferes with the agreement between the subject and the verb. This demonstrates that the parser relies on memory retrieval mechanisms that can fail. For Eastern Hindi speakers reading MSH, maintaining an unfamiliar feminine gender feature across a long distance is cognitively taxing. Consequently, they may default to the unmarked masculine form because it is easier to process.

Current psycholinguistic studies on Hindi agreement largely focus on native speakers of the standard variety. They treat the language as uniform and do not account for the linguistic reality of the Hindi belt, where millions of speakers use mixed regional languages. There is a specific research gap regarding how bidialectal speakers—whose native variety lacks inanimate gender agreement—process long-distance gender dependencies in a prestige language. Most studies measure either real-time processing or offline acceptability, but few integrate both to observe the conflict between cognitive error detection and sociolinguistic prestige bias. This study addresses the gap by using a self-paced reading experiment to compare online processing and offline judgments of MSH gender agreement among native MSH and Eastern Hindi speakers.

Methodology

This paper devises a self-paced reading experiment where the MSH cues have been given early on in the sentence to prime the participants of the experiment to expect the MSH variety throughout the sentence. The NPs chosen as the subject in the sentences are overtly marked as feminine with an agreeing adjective leading the feminine subject for the benefit of EH speakers who do not have an intuitive knowledge of which noun is feminine e.g. merii kiitaab ‘my.F book’. Another design aspect of the critical sentences is that they have been interspersed with relative clauses and/or adverbial clauses to make the sentences long or short to have some distance between the subject of the sentence and the final

agreeing verb or adjective to check the recall of the gender of the subject NP when the participant reaches the final agreeing verb or adjective. The number of lexical items interspersed between the subject NP and the final VP or AdjP (adjectival phrase) is manipulated as long with relative clauses with adverbial phrases and short with only relative clauses to check the recall differences. Each sentence is followed by a polar question for the grammaticality judgement of the participants to check whether their online and offline judgments align.

A. MSH Long Grammatical Condition

3. merii kiitaab jisko shyaam-ne paDhaa hai bahut din pehle achii
my.F book.F RP Shyam-DAT read AUX many day before good.F

hai AUX
'My book which Shyam read many days ago is good.'

B. MSH Short Grammatical Condition

4. merii kiitaab jisko shyaam-ne paDhaa hai achii hai
my.F book.F RP Shyam-DAT read AUX good.F AUX
'My book which Shyam has read is good.'

C. MSH Long Ungrammatical Condition

5. *merii kiitaab jisko shyaam-ne paDhaa hai bahut din pehle my.F book.F RP
Shyam-DAT read AUX many day before
achaa hai
good.M AUX
'My book which Shyam read many days ago is good.'

D. MSH Short Ungrammatical Condition

6. *merii kiitaab jisko shyaam-ne paDhaa hai achaa hai
my.F book.F RP Shyam-DAT read AUX good.M AUX
'My book which Shyam has read is good.'

The conditions are named grammatical and ungrammatical on the basis of agreement system of MSH as the subject NP of the sentences are overtly gender marked making the subject NPs of MSH variety and thus the final verb or adjective not agreeing with the MSH NPs will be deemed ungrammatical.

The experiment uses the Latin Square design to counterbalance the four conditions equally among the participants. The experiment was conducted using the PCIbex software. A total of 20 male postgraduate students between the age of 20-25 who are residents of Jubilee Hall, University of Delhi participated voluntarily and consensually in the study. 10 of the participants were the native speakers of the MSH variety and the rest were native speakers of EH.

Participants were given a trial experiment in the beginning to acquaint themselves with the working of the software i.e the PCIbex SPR setup. Detailed instructions were given before the actual experiment. Instructions were also accessible on the screen before the start of the actual experiment. Consent was taken digitally through PCIbex and their rights including that of confidentiality was verbally communicated to them. A total of 36 sentences were given to each participant including 12 critical sentences and 24 filler sentences.

Analysis

The following analysis examines the online and offline processing data obtained from the self-paced reading (SPR) experiment, focusing on how native speakers of Eastern Hindi (EH) and Modern Standard Hindi (MSH) navigate gender agreement cues in varying syntactic environments. To ensure data integrity, raw reading times from PCIbex were filtered to exclude outliers, defined as responses below 100 ms or above 1500 ms, which might indicate mechanical errors or participant distraction. The final dataset comprised approximately 20 postgraduate students, categorized into groups based on their self-identified mother tongue.

The experimental sentences were segmented into five distinct processing regions to track the real-time cognitive load as participants progressed through the stimuli: the Noun Phrase (NP), the Relative Clause (RC), the Adverbial Phrase (AdvP), the Critical Region (CR), and the Spillover Region (SP).

1. Initial and Interspersed Regions (NP, RC, and AdvP)

In the initial NP region, which provides the Western Hindi gender cue (e.g., merii kitaab 'my.F book'), a significant main effect of Variety was observed ($F(1, 256) = 17.33$, $p < 0.001$). Eastern Hindi speakers exhibited consistently higher baseline reading times compared to MSH speakers across all early regions. This disparity suggests that EH speakers may incur a higher initial cognitive cost when processing MSH gender markers, likely due to the absence of overt gender markings for inanimate subjects in their native variety.

As participants moved into the RC and AdvP regions, the clauses interspersed between the subject and the final verb, the data revealed a significant three-way interaction between Variety, Length, and Grammaticality ($F = 8.06$, $p = 0.004$). For EH speakers, reading times in the RC region were significantly higher in ungrammatical conditions compared to grammatical ones, suggesting that the complexity of the relative clause interacts with the expectation of a standard MSH structure. In contrast, MSH speakers maintained a stable, faster reading pace through these regions, demonstrating a higher level of automaticity in navigating complex Modern Standard Hindi syntax.

2. Critical Region (CR): The Adjective/Verb

The Critical Region (CR) consists of the gender-agreeing adjective or verb that either matches or violates the gender of the initial subject. Statistical analysis through a three-way ANOVA revealed a significant interaction between Sentence Length and Grammaticality ($F(1, 256) = 4.55$, $p = 0.033$).

For MSH speakers, the Long Grammatical condition surprisingly showed a mean reading time (MRT) of 253.9 ms, which was higher than the Long Ungrammatical condition at 220.5 ms. This reinforces the unexpected trend noted in earlier observations where MSH native speakers occasionally process ungrammatical structures faster in online tasks, perhaps due to a simplified processing strategy in long-distance dependencies. For EH speakers, however, the CR reading times for ungrammatical sentences remained relatively high, particularly in the Short Ungrammatical condition (327.3 ms), indicating a clear detection of the morpho-syntactic violation despite their native variety's lack of such features.

3. Spillover Region (SP): The Final Auxiliary

The Spillover Region (the auxiliary hai) serves to capture delayed processing effects from the critical region. A highly significant main effect of Variety persisted here ($F(1, 254) = 15.83$, $p < 0.001$). A striking finding in this region was the MRT for EH speakers in the Long Grammatical condition, which peaked at 353.5 ms. This is significantly higher than their MRT in the Long Ungrammatical condition (278.4 ms). This suggests that for EH speakers, a correct Modern Standard Hindi agreement across a long distance requires more integrative effort than a simple ungrammatical form, which they may simply treat as a native-like unmarked form.

While online processing data shows EH speakers are sensitive to violations, their offline grammaticality judgments tell a different story. EH speakers reported an acceptability rate of 86.66% for all conditions, including ungrammatical ones. This discrepancy between real-time processing (where they slow down for errors) and final judgment (where they accept errors) may be attributed to the sociolinguistic prestige of Modern Standard Hindi.

As the Standard variety and the primary medium of instruction, EH speakers may perceive any sentence starting with an MSH cue as inherently correct or standard, overriding their own online detection of errors. Furthermore, the EH variety is often classified as a mixed language(Alok et al., 2013) which may lead to a more flexible internal grammar that accommodates both marked and unmarked gender agreement models.

In summary, the statistical data supports the hypothesis that EH and MSH speakers utilize different processing mechanisms for gender agreement. MSH speakers show high sensitivity to grammaticality in short sentences but exhibit unexpected deviations in longer, more complex structures. EH speakers, while showing evidence of online error detection, remain unfettered in their final acceptability judgments, likely influenced by the sociolinguistic status of Standard Hindi and their familiarity with MSH through literature and education.

Discussion

The results of this study present a compelling look into the cognitive and sociolinguistic mechanisms at play when speakers of a non-standard variety navigate the morpho-syntactic constraints of a prestige standard. By integrating the statistical evidence from the self-paced reading (SPR) experiment with the sociolinguistic context of the Eastern Hindi (EH) and Modern Standard Hindi (MSH) speakers, we can begin to reconcile the apparent contradictions between real-time processing and offline grammatical judgments.

The most striking finding in this study is the acceptability paradox observed in EH speakers. While the SPR data indicated that EH speakers were sensitive to gender violations, evidenced by increased

mean reading times (MRT) in critical and spillover regions for ungrammatical conditions, their offline acceptability ratings remained consistently high (86.66%) regardless of grammaticality.

This disconnect suggests that while the EH internal grammar detects a morpho-syntactic "clash" during online processing, the offline judgment is filtered through a sociolinguistic lens of prestige. As noted, MSH carries the tag of being the standard variety. For an EH speaker, Standard Hindi is the medium of instruction, the language of literature, and the voice of high-budget cinema. Consequently, when presented with a sentence that is clearly cued as Western Hindi (using feminine possessives like merii ‘my.F’), the speaker may adopt a "standard-is-always-correct" bias. This bias effectively overrides the cognitive signal of an error, leading to the high acceptability rates observed.

The statistical analysis revealed that EH speakers generally exhibit higher MRTs across all regions compared to WH speakers. This is particularly evident in the "Long Grammatical" condition at the spillover region, where EH MRT reached 353.52 ms. In contrast, the "Long Ungrammatical" condition for EH speakers was significantly faster at 278.39 ms.

This suggests that for EH speakers, the act of successfully maintaining a long-distance gender dependency in a non-native (or semi-native) variety is cognitively taxing. Because Eastern Hindi does not have overt gender markings for inanimate subjects, the requirement to hold a feminine gender feature across multiple relative and adverbial clauses requires active integrative effort. When the sentence is ungrammatical, the speaker may simply default to the unmarked model of their native variety, which is cognitively cheaper to process. This aligns with the author's observation that while EH speakers use no gender markings in speech, they are forced to engage with them in school and written forms.

The data for MSH speakers revealed an unexpected trend: in long sentences, they appeared to process ungrammatical conditions faster than grammatical ones at the critical region (220.48 ms vs 253.86 ms). This anomaly can be explained through the "Good-Enough" processing theory in psycholinguistics.

In complex sentences with multiple interspersed clauses, even native speakers may not build a perfect, fully specified syntactic representation. Instead, they use heuristics to gain a good enough understanding of the sentence. When the distance between the subject (kiitaab ‘book’) and the adjective (achii ‘good.F’) is long, the MSH speaker may lose the specific phi-features of the noun. The ungrammatical masculine form (achaa ‘good.M’) might be processed faster because it is the default or unmarked gender in Hindi, requiring less retrieval effort than the specific feminine feature

stored at the beginning of the sentence.

The classification of Eastern Hindi/Bihari varieties as mixed languages, using Magahi or other Bihari language noun phrases with Hindi verbal structures, is central to understanding these results. This linguistic hybridity creates a grammar that is inherently flexible. EH speakers are used to switching between agreement models in discourse.

This flexibility is a double-edged sword. In an educational context, it allows for high comprehension of the standard variety, but it may also explain why EH speakers struggle with consistent gender production in Modern Standard Hindi. The fact that the written form primes EH speakers toward the MSH variety while speech does not suggest that the Standard is treated as a separate, formal code rather than a natural extension of their native grammar.

Limitations and Future Research

The reliance on a digital SPR setup like PCIbex presents challenges, particularly regarding data aberrations and participant sincerity. Future studies should aim to include:

Production Tasks: To see if the tilt toward WH in judgment matches actual speech.

Eastern Hindi Cues: Testing sentences that start with EH-specific markers to see if the prestige filter is removed.

Broader Demographics: Moving beyond doctoral and postgraduate residents of a single hostel to ensure the findings are not specific to a highly academic subgroup.

Ultimately, this study highlights that gender agreement in Hindi is not merely a rule of grammar, but a site of complex interaction between cognitive retrieval and sociolinguistic identity. EH speakers do not lack the ability to detect gender errors; rather, their linguistic environment has conditioned them to accept these variations within the broader spectrum of Standard Hindi.

References

Alok, D., Lahiri, B., & Kumar, R. (2013). Bihari Hindi as a mixed language. In Proceedings of Language Contact in India: Historical, Typological and Sociolinguistic Perspectives (pp. 199–208).

Bhatia, S., & Dillon, B. (2022). Processing agreement in Hindi: When agreement feeds attraction. Journal of Memory and Language, 124, 104334. https://doi.org/10.1016/j.jml.2022.104334

Ferreira, F., Bailey, K. G. D., & Ferraro, V. (2002). Good-enough representations in language comprehension. Current Directions in Psychological Science, 11(1), 11–15. https://doi.org/10.1111/1467-8721.00158

Hoernle, A. F. R. (1880). A comparative grammar of the Gaudian languages: With special reference to the Eastern Hindi, accompanied by a language-map and a table of alphabets. Trübner & Company.

Kachru, Y. (2006). Hindi. John Benjamins Publishing Company. https://doi.org/10.1075/loall.12

Kumar, R., Lahiri, B., & Alok, D. (2021). Descriptive study of Eastern Hindi: A mixed language. OSF Preprints. https://doi.org/10.31235/osf.io/aw49z

LaDousa, C. (2010). On mother and other tongues: Sociolinguistics, schools, and language ideology in northern India. Language Sciences, 32(6), 602–614. https://doi.org/10.1016/j.langsci.2009.12.004

Zehr, J., & Schwarz, F. (2018). PennController for Internet Based Experiments (IBEX). https://doi.org/10.17605/OSF.10/MD832.