



**HAL**  
open science

# Toward quantification of the evolution of modern Hindi: A chronological analysis of Hindi usage by established digital news publishers

Amit Dhawan

► **To cite this version:**

Amit Dhawan. Toward quantification of the evolution of modern Hindi: A chronological analysis of Hindi usage by established digital news publishers. 2024. hal-04669394

**HAL Id: hal-04669394**

**<https://hal.science/hal-04669394v1>**

Preprint submitted on 8 Aug 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - ShareAlike 4.0 International License

# Toward quantification of the evolution of modern Hindi

A chronological analysis of Hindi usage by established digital news publishers

Amit Raj Dhawan

amitrajdhawan@gmail.com

8 August 2024

Natural languages evolve with time, and they adopt new words and syntax. This work investigates the evolution of the Hindi language in the recent past by noting the words used by reputed Hindi news publishers over the years and comparing them to a list of words developed specifically for this task. 1.33 million news articles were analysed to obtain the number of new words per article over time. This indicates the rate of change, and demonstrates computationally how Hindi writing has expanded and changed significantly in a few years. We measure the influx of transliterated English words and Latin script words in Hindi text. This study demonstrates that the number of new words in a Hindi news article increased by 2 times from 2009 to 2024. Hindi alternatives exist for many of these new words because our analysis filtered many novel popular words (e.g. Internet, Facebook, Twitter, Tesla, and Meta). We attempt to understand the nature of changes in Hindi, their causes, and their implications. It is also discussed if the pace of this evolution is suitable across generations and regions. The transformation in published Hindi is compared with English, and examples indicating noticeable trends are presented.

## Contents

1	Introduction	1
2	Language development and standardization	2
3	Measuring Hindi evolution	3
3.1	Web scraping	6
3.2	Word list creation	6
4	$\beta$ -Hindi words	7
5	Latin words in Hindi text	10
6	Progression of Hindi and English in Indian publishing	11
7	A case for studying Hindi evolution	11
7.1	Classification of change in Hindi	13
7.2	Impact of change in Hindi	15
7.3	English yoga effect	15
7.4	Hindi writing and style	16
8	Controlling the change	17
9	Conclusion	18

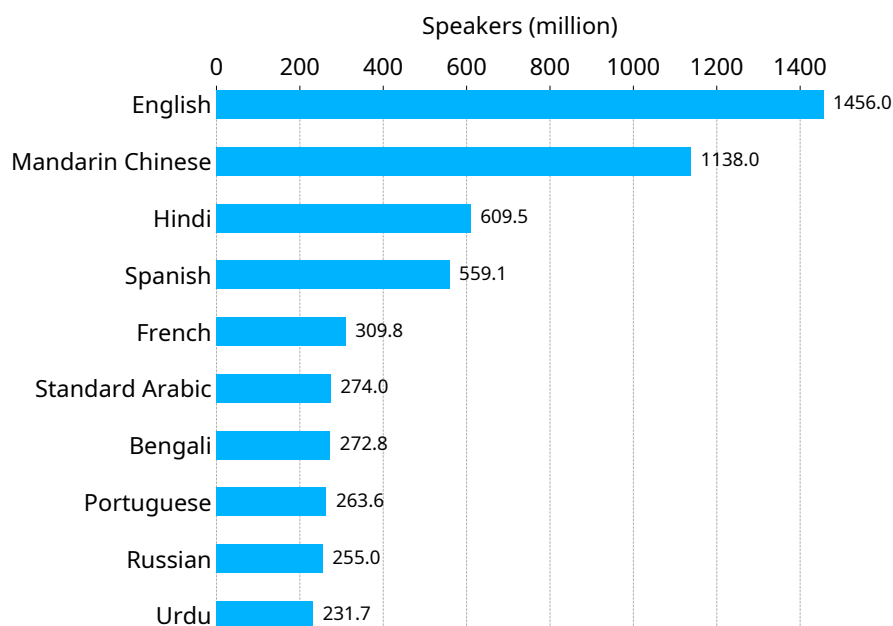
# 1 Introduction

Hindi language has undergone significant change in the past 20–30 years, which can be noted in day-to-day usage and on reputable digital publishing platforms. The fast rate of generation of new words like स्मार्टफोन (smartphone), एंड्रॉइड (Android), and पिक्सल (pixel) justifies the adoption of transliterations, but there is additional shift in Hindi usage. Words like एजुकेशन (education, which should be शिक्षा) and स्पेशल (special, which should be विशेष) are replacing the correct forms. Moreover, their plurals too are being absorbed without any modification as required by Hindi grammar. Examples include स्मार्टफोन्स (smartphones), शोज (shows), and फूड्स (foods). This contrasts with the Arabic and Persian words in Hindi like शीशा (*sheesha*, glass) and तरीका (*tareeka*, method), which are used according to Hindi grammar as शीशे (*sheeshe*, glasses) and तरीके (*tareekae*, methods). English has also borrowed words from Hindi such as ‘chutney’ and ‘jungle’, however their plurals follow English grammar. Therefore, English has ‘chutneys’, but not ‘chutniyaan’.

Code-mixing [1, 2, 3, 4] is the phenomenon of using two or more languages within short communication, which may happen the casual exchange between multilingual speakers. We observe the use of English words in Latin script within Hindi Devanagari text—an explicit form of *Hinglish* (code-mixing between Hindi and English), which may be easier to read and write for some Hindi speakers. Consequently, sentences like “पुलिस ने CCTV देखा” and “Women’s Asia Cup 2024 में भारत की जीत” are becoming increasingly common in Hindi newspapers. Interestingly, code-mixing is predominantly seen in Hindi news articles but not in English. The Times of India or Daily Pioneer are not known to write sentences such as “The देरी is causing frustration among the public and they are asking for न्याय”.

The myriad of new technology words and increased cultural connectivity are changing every major language (Figure 1) worldwide. But this transformation is not similar across different languages as most new words such as *Internet*, *smartphone*, *laptop*, and *clickbait* have an English origin or derivation. Therefore, the acquisition of these new words by English is natural, and to an extent, this is true for other Latin languages like French and Spanish. This is not the case for languages like Hindi, Chinese, and Arabic, where the integration of these words may seem unnatural due to lack of context. Translations of new technology words [5] remain sparsely used because these borrowed words, although alien to the language, are adopted before the translated versions. It is understandable that in today’s Internet-connected world, languages affect each other more than before. In this interaction, however, more widespread languages have a higher influence. Additionally, this work discusses other transformations that are unique to Hindi and are not common in languages like Chinese and French.

The language of personal communication carries the imprint of its usage but it is not normative. The analysis of messages and blog posts can reveal aspects of language usage but it is not indicative of a validated general writing style. Unlike social media messages and comments that often contain custom terms and non-standard usage, the text in news articles is well thought out and written. Therefore, to examine the general evolution of Hindi, this research examines the Hindi in these articles. This work discusses the role of standardisation in language dynamics, and compares Hindi standardisation with English. We describe our method used to compute the rate of transformation of Hindi and present the results. The increasing prevalence of *Hinglish* has garnered substantial research attention [6, 7, 8]. This work discusses the need to study and classify the transformation of Hindi, note its social, political, economic and psychological impact, and estimate its future. We attempt to elaborate on some of the causes behind the change in Hindi and ways to control them to enhance the language.



**Figure 1.** The most spoken languages in the world as a first or second language. There are 3 Indian languages in the top 10. Data source: [9].

## 2 Language development and standardisation

A new word is incorporated into a language when it is used frequently by many people. Its inclusion in dictionaries validates its acceptance. Incorporation into a dictionary or use by recognised publishing platforms like news websites, school books, and government websites signifies its inclusion. A school student will not miss points for using this word in an exam. English dictionaries and major English news websites have established procedures to validate the entry or use of words. Lexicographers at Oxford University Press research and maintain the Oxford English Corpus from which new words are meticulously selected and added to the Oxford English Dictionary [10]. The appearance of a word in the Oxford English Dictionary is a testimony to its acceptance into the English language.

Every popular world language has an organisation to oversee and regulate it. There is the French Academy [11] and French Alliance [12] for French, the Royal Spanish Academy [13] for Spanish, the Central Hindi Directorate [14] and the Department of Official Language (India) [15, 16] for Hindi. The impact of these organisations is varied and not dependent on the number of speakers. The French Alliance (Alliance française) has 834 centres in 133 countries [17], whereas the Central Hindi Directorate has only 5 centres in India [18] and none outside. We have yet to see the presence of any key Hindi language organisation on the world stage.

English news websites such as the BBC (British Broadcasting Corporation) [19] and The Guardian [20], and the Government of the United Kingdom [21] expect their authors to follow their style guidelines. The following text from BBC’s writing guide [19] demonstrates its dedication to conformity.

“

clear-cut ... hyphenated.

cloning ... Do not use phrases such as “embryo cloning” or “baby cloning”. It is not the baby or the embryo that is being cloned - rather, it is the adult human or the genetic material from an adult that is cloned to produce a baby or embryo.

Department for Education ... may be abbreviated at second reference to DfE (no gaps).

discharge ... People are not “released” from hospital - they are discharged (or sent home, allowed home etc).

”

Similarly, academic writing in English—reports, theses, articles, and books—generally follows popular style guides [22, 23, 24]. Despite the abundance of style guides [25], they agree on the main recommendations. In any case, this exhibits how seriously writing is taken in English. Modern guidelines for English writing insist on simplicity to increase comprehension [26, 27], whereas modern Hindi often includes unnecessary and incorrect words that decrease readability.

Articles published by English news websites [28, 29, 30] of India exhibit correctness and regulation of writing. Could it be claimed that Hindi news websites use the correct language and work within a well-defined framework? Unfortunately, even reputable Hindi news websites do not follow important style guidelines and often write incorrectly. For example, ‘l’ (*purnaviraam* or full stop in Hindi) has been replaced by ‘.’ (full stop of Latin), while Hindi textbooks and government publications still use the original (and correct) form. When, why and by whom was this modification implemented? Given the large number of Hindi speakers, its standardisation [31] needs expansion, improvement and implementation.

### 3 Measuring Hindi evolution

To our knowledge, this work is the first large scale study to measure the evolution of written Hindi on reputed publishing platforms. Although it chronologically analyses only a fraction of written Hindi, the analysed sources constitute the most reputable and widely read platforms, which make the conclusions indicative of an overall change. Moreover, on the one hand, the language style on these platforms depicts usage, and on the other hand, it influences usage. The English used by the BBC is considered correct and is followed confidently by students and writers. Similarly, esteemed Hindi news websites are normative sources for Hindi writing. This work analyses such publishing platforms as they strongly influence language usage.

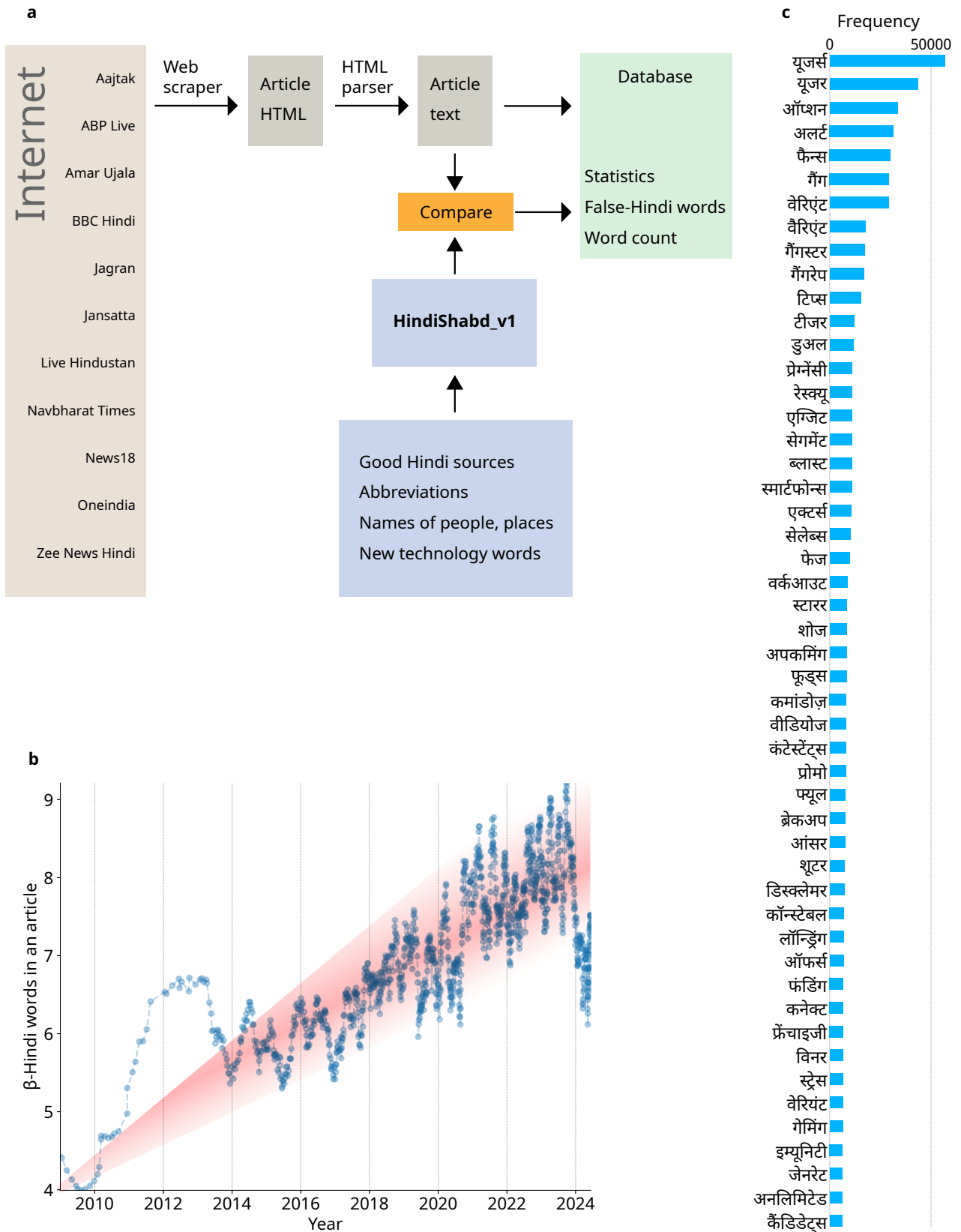
**Table 1.** Article sources

Total number articles from each website are rounded to a thousand, 1 K = 1000.

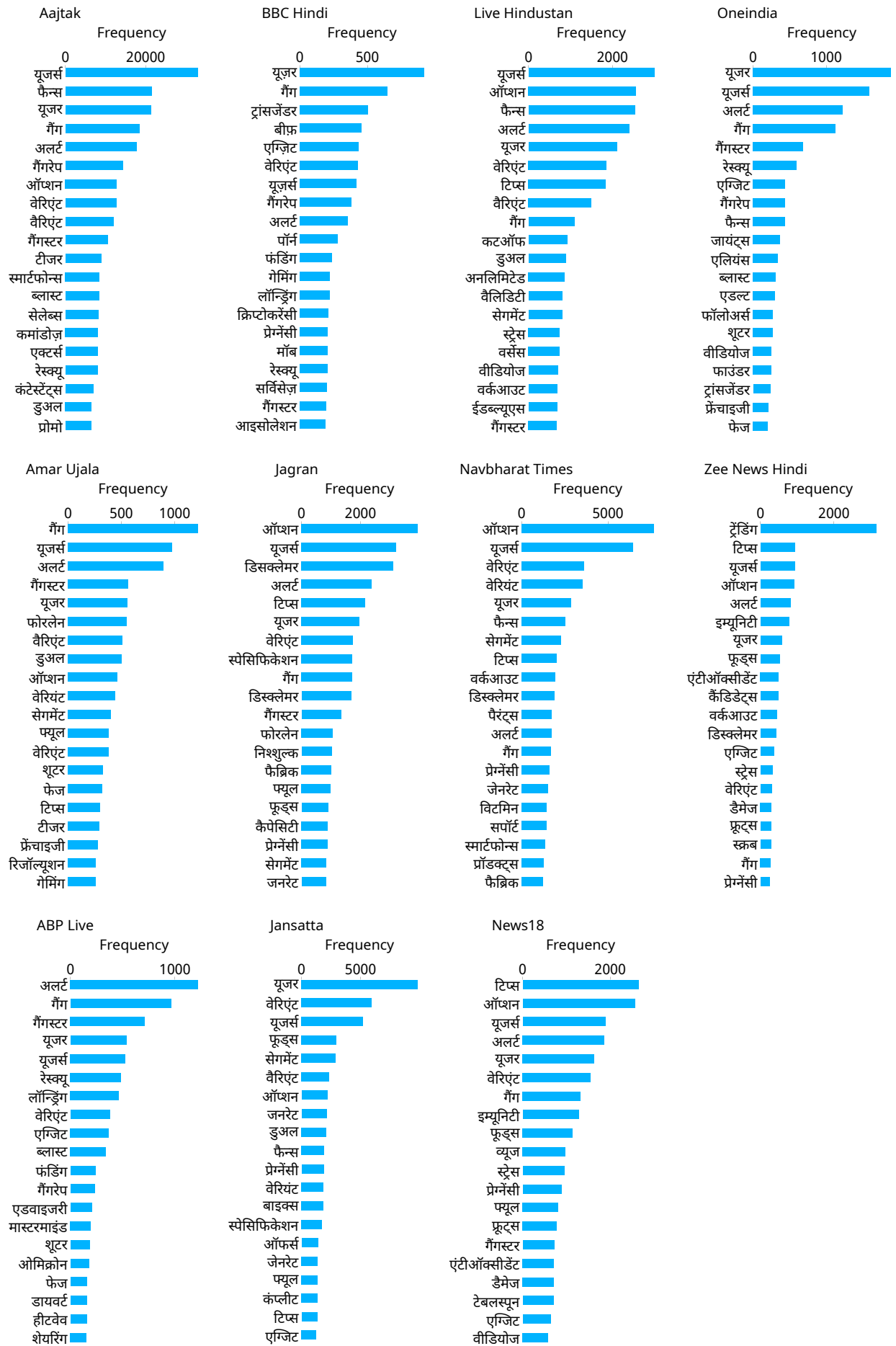
Website	Articles	Website	Articles	Website	Articles
Aajtak	597 K	Jagran	106 K	News18	65 K
Amar Ujala	49 K	Jansatta	77 K	Oneindia	40 K
ABP Live	25 K	Live Hindustan	92 K	Zee News Hindi	31 K
BBC Hindi	29 K	Navbharat Times	77 K		

**Table 2.** Main statistics

Total articles analysed	1.3 million
Total words analysed	540.7 million
Total Devanagari words analysed	520.1 million



**Figure 2. a** Process flow showing data accumulation, analysis, and storage. **b** Graph depicting the increasing number of  $\beta$ -Hindi words in an article from 2009 onwards with an overlay in red highlighting the trend. **c** Histogram of 50 most used  $\beta$ -Hindi words in 1.33 million articles with a total of 520.1 million Devanagari words.



**Figure 3.** Top 20  $\beta$ -Hindi words used by websites from 01/01/2009 to 10/06/2024.

### 3.1 Web scraping

News websites were scrapped using BeautifulSoup and Scrapy Python modules. Only Aajtak provided free access to its archives [32] and all its articles were analysed. Other websites were searched chronologically using Google using a simulated browser or API (Application Programming Interface) programmed in Python. The process is outlined in Figure 2a.

Article text (body, headline, section headings) and metadata (publication date, category) were parsed from its HTML (Hypertext Markup Language) or JSON (JavaScript Object Notation) code and were stored in a MongoDB® database. The article text was compared to the HindiShabd\_v1 [33] word list and the  $\beta$ -Hindi words (described below) in it were extracted and counted.

This work analysed 1.3 million Hindi articles published between January 2009 and June 2024 on 11 prominent Hindi news websites. Table 1 lists the websites and their article count, which totalled 1.3 million articles and 540.7 million words (Table 2).

### 3.2 Word list creation

The words from each article were compared to a list of Hindi words, and the rate of entry of new words and other statistics were measured. The main challenge was to create a list of Hindi words for meaningful comparison. The word list is called *HindiShabd* followed by version number. Here, we have used its first version, i.e. HindiShabd\_v1. The following states the requirements of this word list and how they were satisfied:

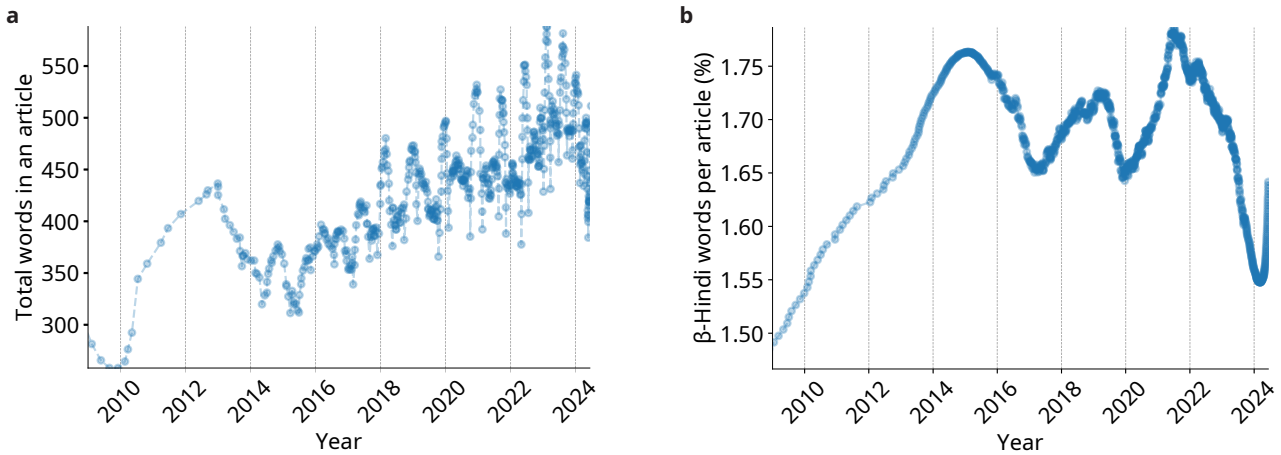
- **Correct Hindi words:** All words used in the 10 volumes of Hindi Shabdsagar (हिंदी शब्द सागर, 310,000 unique words) [34] published between 1965–1975 and in Wardha Hindi Shabdkosh (वर्धा हिंदी शब्दकोश, 68,400 unique words) [35] published in 2013 were included.
- **Other Hindi Sources:** Words from 12 digitally available Hindi NCERT (National Council of Educational Research and Training) textbooks, and some other credible books and sources [36, 37, 38] were added.
- **Novel technology and scientific words:** Words like इंटरनेट (Internet), स्मार्टफोन (smartphone), एंड्रॉइड (Android), ट्विटर (Twitter), व्हाट्सएप (Whatsapp), पिक्सल (pixel), ऐप (app) etc. were selected and added. However, it was attempted to not include their grammatically incorrect forms like फोन्स (phones) and साइकिल्स (cycles).
- **Proper nouns:** Names of many places, people, organisations, and companies—national and international—were included in the word list.
- **Orthographic variations:** These were incorporated to include diverse forms of Hindi usage. For instance, in the case of nuqta (नुक्ता ँ) words, both versions—with and without nuqta—are used, e.g. किस्मत (qismat) & किस्मत (kismat) and कागज़ (kaGaz) & कागज (kagaj). Thus both versions of such words were included. News websites use bindu (बिंदु ं) instead of chandrabinu (चंद्रबिंदु ँ), therefore variations of words obtained by replacing chandrabinu (ँ) with bindu (ं) were incorporated.
- **Abbreviations:** The letters in Hindi abbreviations should be separated by a period but most websites do not follow this rule. Therefore, abbreviations were not detected programmatically and were included from several sources.

In its final form, HindiShabd\_v1 contains about 630,000 unique Hindi words in Devanagari script that were encoded in UTF-8 format and stored in text format, which can be downloaded from [33]. This work defines  **$\beta$ -Hindi** words as those Devanagari words that are not in HindiShabd\_v1. It also



comprises commonly used words like हाउसबोट (houseboat), सेल्फी (selfie), and एयरबैग (airbag), which may be considered  $\beta$ -Hindi words. It was verified these additions do not change the trend as seen in Figure 2b, and only affect the measured value.

Open-access natural language processing (NLP) models for Hindi [39, 40] are still under development and prone to errors.<sup>1</sup> For example, many technology words and proper nouns are tagged as “unknown” or incorrectly tagged as proper nouns. Devanagari does not have capitalisation, which makes detecting proper nouns and abbreviations (which look like ordinary words due to the removal of period ‘.’) in Hindi more challenging than in English. Future extensions of this work will include trained NLP models to filter proper nouns and abbreviations, and estimate the evolution more accurately.



**Figure 4.** **a** Trend of average number of Devanagari words in an article. **b** Average percentage of  $\beta$ -Hindi words per article over the years.

## 4 $\beta$ -Hindi words

The results discussed here depict general trends despite their dependence on HindiShabd.v1. In other words, using another appropriate word list for comparison should not affect the trends, and will only change the quantified values.  $\beta$ -Hindi words from an article were extracted and counted. Since 2009, the average number of  $\beta$ -Hindi words in a Hindi article has doubled. In 2009–2010, an article had about 4  $\beta$ -Hindi words, which increased to about 9 in 2024 as shown in Figure 2b. This increase is more than 2 times.

The rise in  $\beta$ -Hindi words over the years (Figure 2b) is also due to an increase in the average number of words in an article. Figure 4a shows that the average number of words in an article increased from 300 in 2009 to about 500 in 2023. The average percentage of  $\beta$ -Hindi words per article or  $\beta$ -Hindi word density, defined as  $\left\langle \frac{\text{Number of } \beta\text{-Hindi words in an article}}{\text{Number of Devanagari words in an article}} \right\rangle$ , increased slightly over the years (Figure 4b). The relationship between unknown words in a text and its comprehension is nuanced. Under test conditions, research [41, 42] shows an approximately linear relationship. In field conditions, however, it has been reported that unknown words can discourage people from reading [43]. Consequently, for the same unknown word density, longer texts are less understood than shorter texts ones due to the discouragement factor.

<sup>1</sup>It is noteworthy that the first works on the classification of words as grammatical building blocks—Part-of-Speech (PoS)—originated in India in 6 BCE or 5 BCE in the works of Yaska and Panini [44].

As far as comprehension of online material is concerned, it seems that the trend of Figure 2b appropriately indicates the change in Hindi. It considers fast reading while browsing news articles. In 400–1000 word articles with the same unknown word density, arguably shorter articles are easier to understand than longer texts due to less cognitive load. The cumulative effect of interruptions due to unknown words can decrease comprehension. While unknown words in a text are opportunities to increase vocabulary, they inhibit understanding and decrease interest in reading.

For each of the 50 most used  $\beta$ -Hindi words in Figure 2c, proper Hindi alternatives exist (Table 3). The causes behind this progression demand probing. Figure 3 lists the 20 most used  $\beta$ -Hindi words by each website. There are alternative words in Hindi for most of these transliterated English words. Repeated use of  $\beta$ -Hindi words signals their acceptance into Hindi despite being unnecessary or grammatically incorrect. Many Hindi speakers will find the word फ्यूल (fuel) incorrect but its widespread use on normative platforms gives it acceptance. While HindiShabd.v1 contains the word स्मार्टफोन (smartphone) as it is a new technology word, we cannot justify the use of its grammatically incorrect plural स्मार्टफोन्स (smartphones). Moreover, transliteration from English into Hindi is not unique and there are no rules to oversee this. For example, for ‘smartphones’, we find variants such as स्मार्टफोन्स, स्मार्टफोन्ज़, स्मार्टफोन्स, and स्मार्टफोन्ज़. This arises due to disregarding the English pronunciation in transliteration. In this example, स्मार्टफोन्ज़ is phonetically the most accurate transliteration.

**Table 3.** Hindi alternatives for some  $\beta$ -Hindi words

$\beta$ -Hindi word	Hindi alternative
यूज़र	प्रयोक्ता
ऑप्शन	विकल्प
गैंगरेप	सामूहिक बलात्कार
प्रेग्नेसी	गर्भावस्था
अपकमिंग	आगामी
फ्यूल	ईंधन
आंसर	उत्तर
ऑफर्स	प्रस्ताव

Screen captures in Figure 5 from the analysed websites show the abundant use of  $\beta$ -Hindi and Latin words. These are highlighted in red colour, and include words like अपडेशन, नैचुरल, एजुकेशन, प्रेग्नेसी, सीक्रेट, एक्सपर्ट, and मेल (here it means ‘male’ but the original Hindi word means “act of meeting or mixing” or “similarity”). Furthermore, here these  $\beta$ -Hindi words are in headlines or description text. Again we note that correct Hindi words already exist for each of these words.

As most  $\beta$ -Hindi words are transliterations of English words, these constitute unknown words for non-English readers. Hu and Nation [42] suggest that readers should know 98% of the words to understand a text adequately. Their work measured the English comprehension of speakers whose second language was English. If we extend this finding to Hindi, the conclusions are alarming. Hindi readers who do not know English are not aware of at least 1.6% of words in a Hindi news article (Figure 4b). Adding unknown Hindi words to  $\beta$ -Hindi words would decrease the known words to less than 98%, which implies that Hindi text is not well understood by its native speakers who do not know English. How can a language flourish if a considerable proportion of its readers are limited in understanding its content? Using words from other languages in an elitist manner excludes many of its native speakers and reduces their engagement with Hindi. To keep up with modern developments, Hindi must evolve—the issue is the nature and pace of this evolution. By formulating and implementing the right policies, this evolution can become inclusive and harmonious, which enhances Hindi.

abp न्यूज़

होम न्यूज़ राज्य मनोरंजन खेल बिजनेस ऐदो लाइफस्टाइल जनरल नॉलेज टेक्नोलॉजी शिक्षा

**Bomb Blast Danapur: पटना के दानापुर में बम ब्लास्ट, खाना बना रही महिला समेत कई लोग जख्मी, आसपास के शीशे टूटे**

दानापुर धाना क्षेत्र के सुल्तानपुर में जनकधारी स्कूल के पीछे की यह घटना है, आवाज इतनी तेज थी कि बगल के अपार्टमेंट के शीशे भी टूट गए हैं, बम बलास्ट से मोहल्ले के लोग दहशत में हैं.

हिन्दुस्तान

फोटो वीडियो ई-पेपर शहर चुनें

होम राज्य देश बजट 2024 क्रिकेट शॉर्ट वीडियो मनोरंजन बिजनेस करियर विदेश वेब

माथ पर माग दाका, गलत हवा हर, राहुग में अक्सर सी इमीन दिल्ली नौर फावेली

**रातभर अनलिमिटेड डाटा FREE दे रही है यह कंपनी, किसी भी प्लान से करें रीचार्ज**

उत्ती डाटा खन होने की परेशानी से छुट्टी देते हुए टेलिकॉम कंपनी Vodafone Idea अपने यूजर्स के लिए V Hero Unlimited सेवा लेकर आई है। इसके साथ रात 12 बजे से सुबह 6 बजे तक अनलिमिटेड डाटा फ्री मिलता है।

आज तक

हिन्दी News / एड्युकेशन / एड्युकेशन न्यूज़

**IAS ने शेयर किए UPSC एग्जाम क्रैक करने के टिप्स, टॉपिक देखकर लोगों ने दिए ये कमेंट्स**

IAS Tips: आईएएस अदनीश शरण ने वीडियो शेयर करते हुए कैप्शन में लिखा कि यूपीएससी सिविल सेवा परीक्षा में कैसे फेल हों लेकिन इसमें आईएएस के टिप्स को अच्छे से समझाया गया है.

NBT नवभारत टाइम्स

ब्यूटी & स्किन लाइफस्टाइल भारत राज्य बिजनेस बजट खेल मनोरंजन घर्म दुनिया

ट्रेंडिंग कैंसर की दवा हुई इस्ती अनंत अंबानी राधिका मर्चेंट वैडिंग नीता अंबानी कार्दशियन सिस्टर्स

Hindi News / Lifestyle / Beauty & Skin / Althiya Shetty Kiara Advani To Alia Bhatt Vitamin E Capsule For Skin Care Is The

**सेलेब्स की सुंदरता का ओपन सीक्रेट है ये कैप्सूल, मेल और फीमेल सभी करते हैं जमकर यूज**

नवभारतटाइम्स.कॉम • 5 Jun 2021, 7:51 pm

Subscribe YouTube 10M

एक्ट्रेस की ग्लोइंग स्किन देखकर अक्सर यह सवाल मन में आता है कि आखिर ये ऐसा क्या करती हैं, जो इनकी चेहरा हर समय दमकता रहता है। क्या इसकी वजह इनके महंगे पालर ट्रिटमेंट्स हैं या फिर कुछ ...

अमर उजाला

होम जम्मू और कश्मीर जम्मू उधमपुर कठुआ पुंछ राजौरी श्रीनगर

Hindi News Photo Gallery Jammu And Kashmir Jammu News Jammu Kashmir Cloudburst Photo

**Kishtwar Cloudburst: किश्तवाड़ में बादल फटने से सात की मौत, 40 लोगों की खोज में लगी हैं टीमें**

NEWS 18 जीवन शैली

Trending Topics: Donald Trump Attacked

होम मनोरंजन फटाफट खबरें क्रिकेट मनी धर्म नॉलेज टेक देश प्रदेश लाइफ दुनिया क्राइम वीडियो

लेटेस्ट खबरें अजय गजब राशिकल फोटो हेल्थ वेब स्टोरीज ऑटो फूड करियर SadakSurakshaAbhiyan स्वच्छता और पाप्री

हिंदी समाचार / न्यूज / जीवन शैली / गर्मी में वर्कआउट करने का सही समय क्या है? एक्सपर्ट से जानें, किन बातों का रखें ख्याल

गर्मी में वर्कआउट करने का सही समय क्या है? एक्सपर्ट से जानें, किन बातों का रखें ख्याल

एक्सरसाइज करना हेल्दी रहने के लिए बहुत जरूरी है, लेकिन गर्मी के मौसम में वर्कआउट या हेवी एक्सरसाइज

BBC NEWS हिन्दी

होम पेज भारत विदेश मनोरंजन खेल विज्ञान-टेक्नॉलॉजी सोशल वीडियो पॉडकास्ट

**सात साल कोशिश, आईवीएफ और फिर नैचुरल प्रेग्नैसी से मां बनने की देबिना बनर्जी की कहानी**

oneindia

समाचार कारगिल इतिहास विशेष लेख फीचर देश-राज्य विदेश खेल कारोबार फोटो

चर्चा में बजट 2024 सामन मीट विवाद बंगलादेश में हिंसा मौसम अपडेट साप्ताहिक रणधियत

Latest Updates

दुनिया के सबसे ताकतवर पासपोर्ट

होम समाचार Entertainment

**आश्रम-3 की हॉट एक्ट्रेस का GYM में 'सेक्सी वर्कआउट', देखें वीडियो**

By Mukesh Pandey

Updated: Thursday, July 14, 2022, 20:05 (IST)

बॉलीवुड की बॉल्ड और हॉट एक्ट्रेस एक्ट्रेस ईशा गुप्ता इन दिनों अपने बॉल्डनेस से कहर दा रही हैं।

जागरण

होम ताज़ा ब्रेकिंग बजट 2024 राष्ट्रीय शेर बाजार दुनिया मनोरंजन क्रिकेट अद्ययाम

फोकस इतिहास नगर नदित अजय अकरी है पॉलिटिक्स जैवकुण्डर रॉफिकल क्या खरि

SHED NEWS JALANDHAR-CITY

**डिप्लोमा इन एलिमेंट्री एजुकेशन ईटीटी के 2021-23 के दाखिले शुरू, SCERT पंजाब ने जारी किया शेड्यूल; इस तारीख से शुरू होगी रजिस्ट्रेशन**

विद्यार्थियों को दाखिला प्रक्रिया में भाग लेने के बाद रोजाना विभाग की वेबसाइट से भी जुड़ा रहना होगा ताकि सिरि या परीक्षाओं को लेकर कोई भी अपडेटेशन आती है तो उन्हें उसकी जानकारी वेबसाइट के माध्यम से ही मिलेगी।

जनसत्ता

सामा खबर राहुग मनोरंजन खेल राज्य वेब स्टोरी वीडियो अजय लाइफस्टाइल हेल्थ टेक

3S Shorts IND vs ZIM 3S Special अजय का मौसम

Hindi News / Lifestyle / Know The 5th Month Pregnancy Precautions, It Is Necessary For Both Mother An

**Women Health: प्रेग्नैसी के पांचवें महीने में रखें इन 5 बातों का ध्यान, जच्चा-बच्चा दोनों के लिए जरूरी**

बच्चे की पोषण संबंधी आवश्यकताओं को पूरा करने के लिए प्रेग्नैसी में महिला को खान-पान से लेकर उठने बैठने तक के तरीके में बदलाव करना होगा।

DELHINCR हरियाणा

NEWS वेब स्टोरीज STATES LIVE-TV PHOTOS

वीडियो क्राइम NCR समाचार मुल्क लोकेशन हरियाणा विद्ययत गांवियबाद मुद्राग्राम जरा हके नौएडा फरीदबाद पांडकाल

Zee Delhi NCR Hariana Delhi NCR Hariana

**क्या है इंदौर का वेस्ट मैनेजमेंट मॉडल, जिसे दिल्ली में अपनाना चाहते हैं CM केजरीवाल**

इंदौर से सीख लेकर दिल्ली में लागू करने इंदौर का सच्छता मॉडल (Indore Model). दिल्ली सीएस केजरीवाल (Arvind Kejriwal) ने दिल्ली नगर निगम चुनावों (Delhi MCD Election 2022) से पहले बड़ा ऐलान किया है, आखिर क्यों पड़ रही है दिल्ली में इंदौर मॉडल को लागू करने की ज़रूरत, पढ़िए इस रिपोर्ट को.

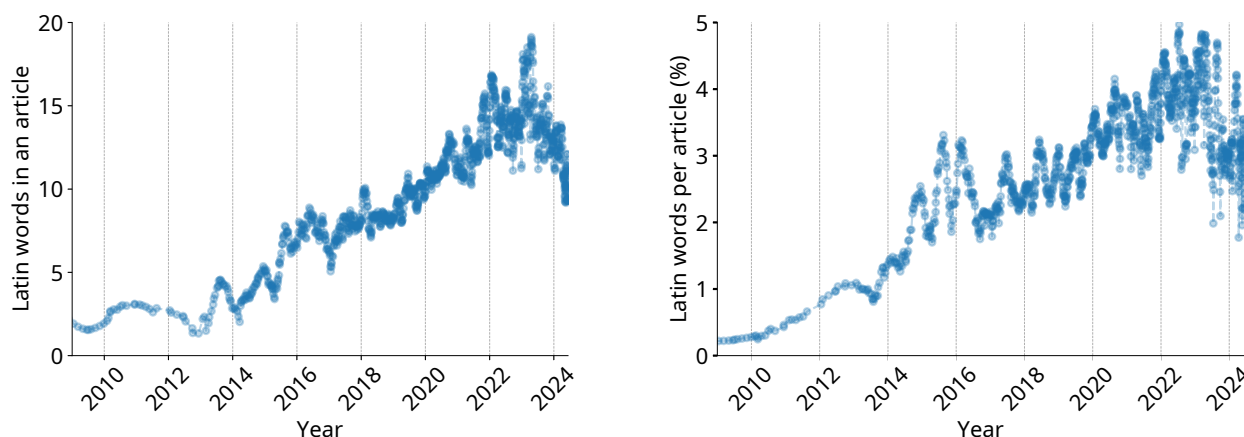
Figure 5. Screen captures from different websites with highlighted  $\beta$ -Hindi and Latin words.

The influx of new words revitalises a language, but unregulated and thoughtless inclusion of words can harm its integrity.  $\beta$ -Hindi words can be detrimental to:

- **Readability:** While reading enriches vocabulary, frequently encountering unknown words reduces reading speed, interest, and comprehension. Knowledge of words in a text is key

to effective understanding. Examples of such words: ओरिएंटेशन, एक्सक्लूसिव, and फैक्चुअल.

- **Accessibility and inclusivity:** Many readers, especially those who do not know or have limited knowledge of English, cannot understand  $\beta$ -Hindi words like प्योरिटी, कॉन्ट्रासेप्टिव, and इकोलॉजी. By limiting the accessibility of modern Hindi texts to English speakers, the world of evolving Hindi is failing to successfully include many of its native speakers.
- **Appeal:** Any text with incorrect grammar, style or words loses its appeal. While languages have changed with time, correct writing and grammar have always been appropriated and appreciated, esp. in publishing. Irrelevant words in Hindi reduce its appeal.



**Figure 6. a** The average number of Latin words per article over the years showing a 6 fold increase since 2009. **b** Average percentage of Latin words in an article, depicting a 25 times increase.

## 5 Latin words in Hindi text

English words in Latin script are ubiquitous in Hindi news articles. Hindi-English code-mixing in news articles was measured and is depicted in Figure 6. Between 2009–2024, the average number of Latin/English words in an article increased by 6 times, and the average Latin word density increased by 25 times (from 0.2% to 5%). This is indicated in the recent screen captures from several Hindi news websites in Figure 5. The Latin word density is defined as the average of

$\frac{\text{Number of Latin words in an article}}{\text{Total number of Latin and Devanagari words in an article}}$ . The drastic increase in English words in Hindi writing can be attributed to:

- **Expansion of English in India:** Almost all college and university education in India is in English, and it is growing as a medium of school education. The familiarity of readers with English is used as a permit by publishers to mix it with Hindi.
- **Style:** The influence of English is rising in India, more so after its independence in 1947. Most brand names are coded in Latin script—Indian or international—such as *TATA*, *Coca-Cola*, *Pepsi*, *Raymonds*, *Haldiram*, and *Thumsup*. Consumer products, business sign boards, and hoardings are early examples of Hindi-English code-mixing. Now this style is being incorporated into news articles as well as seen in the sentence below [45]:

“पहले यह समझना पड़ेगा कि वहां पर electrocution क्यों हुआ, इसके बाद ही आगे की कार्रवाई की जाएगी.”

- **Readability of Devanagari transliteration:** Transliterated versions of Latin script terms and abbreviations in Devanagari might be deemed difficult to read, which pushes the

usage of their original form. For example, DNA (डी.एन.ए. or डीएनए) and UNSC (यू.एन.एस.सी. or यूएनएससी).

- **Lack of Hindi coding:** In the absence of readily available Hindi options coded in Devanagari, new terms like *Follow*, *Like*, and *Subscribe* are used as presented. A trend once set is difficult to change. Many of these words appear on popular social media platforms like YouTube, Instagram, X (Twitter), and Facebook as text icons. Frequent exposure to words reduces eye fixation [46] and can enhance comprehension. Accordingly, the use of some words in their original form could facilitate reading and writing, although it compromises the integrity of Hindi.

## 6 Progression of Hindi and English in Indian publishing

An interesting dimension to explore in the evolution of languages of India is their comparative analysis. A detailed comparison of the type and rate of change between Hindi and English can help in estimating the future of languages in India, which has socio-economic, political, and psychological influences on its people.

The four news article texts in Figure 7 are from the Navbharat Times and Times of India, both of which are owned by The Times Group. From both news portals, we randomly selected and compared one article from 2002 and another from 2024. It can be said that this case reflects the change in Hindi quantified in Figures 2b and 2c, which contrasts with the absence of any significant change observed in English writing.

The Hindi article from 2002 uses no  $\beta$ -Hindi words, and is enriched with words like पनबिजली and समुचित (highlighted in green), whereas the 2024 article consists of many  $\beta$ -Hindi words (प्लेइंग, नाइट, फैंब्रिकेटेड, नॉमिनल) and uses English noun inflections (फ्लाईओवर्स). The only minor problems, highlighted in red, with the 2002 article text are the rare absence of space after punctuation and the use of bindu (ँ) instead of chandrabindu (ँ)—these patterns are also found in the 2024 article.

Both the 2002 and 2024 English articles respect writing style, which seems unchanged in 22 years. Proper nouns aside, only English words are used in both the articles. The 2024 English article does not contain transliterations of Hindi words, unlike the 2024 Hindi article, which is loaded with Latinagari or Hindi transliterations of English words. This shows that the entry of words is mostly one way—from English to Hindi, and not the other way. The 2002 article describes the abbreviations, which increases readability. The only issue with this text is the missing ‘the’ in the red highlighted positions. Although this is not representative of a trend, we note that the writing in the 2024 article (97% score) is better than the 2002 article (53% score). The score was generated by Grammarly [47], which employs artificial intelligence to analyse text.

## 7 A case for studying Hindi evolution

Language and culture are intertwined and this multi-dimensional relationship, which affects identity and policy, has attracted investigation [48, 49, 50]. Nationalism was at its peak as India celebrated independence in 1947. This benefitted Hindi, the most spoken language and a part of Indian identity. It was decades later that English started growing as a medium of education with the help of socio-political factors and the three-language policy [51]. Its prominence has been growing ever since as it has become associated with better quality of education and higher employment potential [52].

## विष्णु प्रयाग व श्रीनगर परियोजनाओं से उत्तरांचल को 12% बिजली मुफ्त

27 Nov 2002, 5:41 pm

**नई दिल्ली (भाषा):** विष्णु प्रयाग और श्रीनगर पनबिजली परियोजनाओं से उत्तरांचल को 12 प्रतिशत बिजली मुफ्त दी जाएगी। केंद्रीय बिजली मंत्री अनंत गीते ने बुधवार को राज्यसभा में पूरक प्रश्नों के उत्तर में कहा कि इन दोनों परियोजनाओं को लेकर उत्तर प्रदेश और उत्तरांचल के बीच चल रहे विवाद को जल्द ही सुलझा लिया जाएगा। मंत्री ने कहा कि इस सिलसिले में उन्होंने दोनों राज्य सरकारों से विचार-विमर्श किया है। इसके बाद उत्तर प्रदेश विष्णु प्रयाग परियोजना के मामले में सहमत हो गया है, जबकि श्रीनगर परियोजना पर राज्य मंत्रिमंडल द्वारा जल्द ही फैसला लिए जाने की उम्मीद है। उन्होंने कहा कि जिस समय ये दोनों परियोजनाएं शुरू हुई थीं, तब एक ही राज्य था। बाद में, उत्तर प्रदेश और उत्तरांचल अलग-अलग हो गए। परंपरा के अनुसार, परियोजना जिस राज्य में चल रही होती है, उसका लाभ उसे ही मिलता है। इस हिसाब से उत्तरांचल को 12 प्रतिशत बिजली मुफ्त मिलेगी। कांग्रेस की अलका क्षत्रिय ने सरकार से जानना चाहा था कि उत्तरांचल में संचालित इन दोनों पनबिजली परियोजनाओं में जो बाधा उत्पन्न हुई है, क्या उसके लिए उत्तर प्रदेश सरकार जिम्मेदार है। गीते ने बताया कि विष्णु प्रयाग और श्रीनगर पनबिजली परियोजनाएं निजी क्षेत्र और राज्य सरकार की साझेदारी में चल रही हैं। गीते ने बताया कि 280 मेगावाट वाली धौलीगंगा परियोजना 2005 तक पूरी कर ली जाएगी। मंत्री ने सदन को आश्वासन दिया कि उत्तरांचल सहित देश में पनबिजली क्षमता का पूरा इस्तेमाल किया जाएगा। उन्होंने बताया कि देश में जल विद्युत क्षमता के समुचित उपयोग और नई पनबिजली परियोजनाएं शुरू करने के लिए केंद्रीय बिजली प्राधिकार ने विभिन्न राज्यों में सर्वेक्षण किया है।

**Source:** <https://navbharattimes.indiatimes.com/-12-/article-show/29578214.cms>

## वाराणसी में फ्लाईओवर्स के नीचे कैरम, लूडो, बैडमिंटन खेलेंगे बच्चे, प्लेइंग जोन बनाने की तैयारी तेज

11 Jun 2024, 12:49 pm

**विकास पाठक, वाराणसी:** उत्तर प्रदेश के वाराणसी में फ्लाईओवर्स के नीचे खाली जमीन पर बच्चों के लिए प्लेइंग जोन बनाए जाएंगे। यहां बच्चे शतरंज, लूडो कैरम, टेबल टेनिस आदि खेल सकेंगे। नगर निगम और स्मार्ट सिटी ने संयुक्त रूप से प्लेइंग जोन का प्लान तैयार किया है। चुनाव खत्म होने के साथ ही इस प्रॉजेक्ट पर तेजी से काम शुरू करने की तैयारी है। नगर निगम के पीआरओ संदीप श्रीवास्तव ने बताया कि फ्लाईओवर के नीचे अतिक्रमण को रोकने के लिए कैंट-लहरातार फ्लाईओवर के नीचे नाइट बाजार विकसित किया गया है। यहां करीब सौ छोटी दुकानें आवंटित की गई हैं। कैंट रेलवे स्टेशन के सामने से फ्लाईओवर गुजरने से बाजार गुलजार है। इस क्रम में अब तीन और ककरमत्ता, चौकाघाट और फुलविरया फ्लाईओवर के नीचे की जगह पर इंडोर प्ले ग्राउंड बनेगा। यह काम ककरमत्ता फ्लाईओवर से शुरू होगा।

स्मार्ट सिटी के चीफ इंजीनियर अमरेंद्र तिवारी प्लेइंग जोन के लिए चिह्नित किए गए तीनों फ्लाईओवरों के नीचे की 100 मीटर लंबी और 8 फुट चौड़ी जगह को स्टील से फेब्रिकेटेड करते हुए टीन शीट और फाइबर शीट के साथ ग्लास लगाकर इंडोर प्ले जाने तैयार कराया जाएगा। इस प्रॉजेक्ट पर एक करोड़ रुपये खर्च होंगे। यहां पर बच्चों के लिए तमाम सुविधाओं के साथ कैरम, लूडो, बैडमिंटन, टेबल टेनिस आदि खेलने की व्यवस्था रहेगी। इसके संचालन का तरीका क्या होगा, इसे बनने के बाद तय किया जाएगा। खेल सुविधाओं के प्रयोग के लिए नॉमिनल रेट रखा जाएगा, जो अभी तय नहीं है।

**Source:** <https://navbharattimes.indiatimes.com/state/uttar-pradesh/varanasi/varanasi-playing-zones-to-be-built-under-three-flyovers-in-varanasi/articleshow/110896781.cms>

## APTransco played for harassing people

TNN / Sep 2, 2002, 02:54 IST

**HYDERABAD:** The District Development Review Committee (DDRC) meeting of Rangareddy district held here on Saturday came down heavily on the officials of the APTransco for harassing the poor people and farmers.

Interestingly, the attack was led by Telugu Desam Party (TDP) MLA K Pushpaleela who alleged that the Transco was doing so to defame the government.

She also criticised it for not providing proper fencing for transformers to protect people from getting electrocuted.

Chevella MLA Sabita Indra Reddy alleged that officials were inviting only Telugu Desam Party leaders for inauguration of transformers and other functions, ignoring the protocol.

Home minister T Devender Goud asked officials to take corrective measures and not to violate the protocol.

Deputy Speaker K Hareeswar Reddy expressed his unhappiness over the inordinate delay that had taken place in distribution of books from class I to IV.

Consequently, Goud directed the district education officer (DEO) to speed up the process. Union minister of state for railways Bandaru Dattatreya assured the meeting that the suburban train services would not be cut down even after commissioning of multi-modal transport system (MMTS). . . .

**Source:** <https://timesofindia.indiatimes.com/city/hyderabad/aptransco-played-for-harassing-people/articleshow/20911432.cms>

## 5 accused of killing 1993 Mumbai blasts convict arrested

TNN / Updated: Jun 10, 2024, 13:32 IST

**KOLHAPUR:** The Kolhapur court on Saturday awarded eight-day police custody to the five men accused of murdering a 1993 Bombay serial blast convict inside the Kalamba Central Jail on June 2.

Munna, alias Adil Khan alias Manojkumar Bhavarlal Gupta, who was convicted for helping Tiger Memon - the 1993 serial blast's key accused - collect a shipment of arms and ammunition brought for the blasts, was killed by five Kalamba jail inmates.

The five attacked were identified as Saurabh Sidh, Bablu alias Sandip Shankar Chavan, Pratik alias Pilya Patil, Raturaj alias Deja Inamdar and Dipak Netaji Khot. They had battered Munna with a concrete and iron cover of a drainage hole.

The prison department has initiated an internal inquiry and has decided to step up security, especially for the other four convicts of the 1993 Bombay serial bomb blast case.

The police had requested custody of the five accused against whom various cases are under hearing in different courts. After getting permission from the courts, police took the inmates from the jail and produced them before a local court. The cops wanted permission to interrogate them for Munna's murder. The court agreed and granted eight days of police custody. . . .

**Source:** <https://timesofindia.indiatimes.com/city/kolhapur/5-accused-of-1993-blast-convicts-killing-in-jail-sent-to-cop-custody/articleshow/110854128.cms>

**Figure 7.** Randomly selected articles from Navbharat Times and Times of India from 2002 and 2024 showing the change in Hindi and English publishing. Good terms and style are highlighted in green, and  $\beta$ -Hindi words, incorrect syntax, and missing words are highlighted in red.

The case of English in Indonesia and South Korea is quite different from India. In Indonesia, despite the presence of multiple languages at the time of its independence in 1945, Indonesian was declared the official language, driven by nationalist sentiments. The country's policy of 2006 to use English as a medium of instruction failed and was withdrawn in 2013 [53] despite the global relevance of English and Indonesia's large tourism industry. Similarly, South Korea's endeavour to teach science and mathematics in English in schools in 2008 faced challenges. The policy was abandoned after just six months [54]. Therefore, India's adoption of English cannot be attributed solely to natural evolution, especially given the divergent trends in other nations. Moreover, education is most effective in the native language [55]. The rich linguistic heritage and history of India, and the fascinating recent developments in Hindi provide an intriguing research opportunity.

Linguistics has explored the multifaceted impact of language, encompassing social, political, economic, and cultural dimensions, as well as its dynamic evolution. The advent of functional magnetic resonance imaging (fMRI) [56, 57] has allowed us to establish links between the brain and language using neuroimaging. We can observe how the brain comprehends native and non-native languages [58], how bilingualism can delay dementia [59], and how it perceives Devanagari [60]. It has been shown that human fetuses show sensitivity to language [61] and the cry of newborns carries the impression of their native language [62]. Indeed, our relationship with language begins before we are born and deepens as we grow. The power of language extends beyond mere communication; it shapes our perception, behaviour, and societal structure at microscopic and macroscopic levels. We have to deeply analyse and research the evolution of Hindi to find how the change in Hindi is changing us and our society.

## 7.1 Classification of change in Hindi

Languages change, some progress, some take different forms, and some die. Hindi is considered a flourishing language [63], but the nature and pace of its modification are concerning. Given the direct and transliterated influx of English words into Hindi, and the growth of English-speaking India, will English digest Hindi? In other words, could Hindi become so transformed that it is reduced to a form of English? This could start with the replacement of Devanagari with Latin script. Indeed, this has begun and is visible in casual digital communications like messages and comments. These are often in Latin script either as transliteration or directly in English. Moreover, most formal communication in non-governmental institutions like companies, hospitals, and schools is in English. These processes can be classified as:

- **Latinagari** in this work denotes the codification of Latin script words in Devanagari script. Latinagari words include  $\beta$ -Hindi, which are Devanagari transliterations of primarily English words. Example depicting Latinagari in boldface: सभी लोग इसके **यूज़र्ज़** हैं और **कॉन्टैक्ट** बनाएँगे, और कहीं **अपोलोजेटिक** नहीं होना।

Despite its increasing usage, Latinagari lacks standardisation like it was done to incorporate Persian and Arabic sounds into Hindi. Standardisation would address phonetic issues and avoid confusion arising due to multiple spellings of the same word, e.g. वेरिएंट & वैरिएंट, and यूज़र & यूज़र.

Latinagari standardisation will bring uniformity in Hindi words and their pronunciation. However, fine aspects of English pronunciation will have to be overlooked as Hindi currently does not have those sounds. For example, 'th' sound (/ð/) in the word 'the' does not exist in Hindi, therefore people use its closest approximation and pronounce 'the' (/ði/)

as र्. These are minor aspects that will find acceptance within English as Indian English becomes a standard like American English.

- **Romanization** is the codification of text from different writing systems in Latin script [64]. The absence of Hindi fonts in early mobile phones and short message text service of 1990s promoted Romanization [65], which is now everywhere. The increasing English familiarity of Indians is adding to it, and it can be argued that many people find it easier to communicate digitally using Romanization than in Devanagari.

Even today Romanized script is the most used in casual digital typing despite the availability of phonetic typing on all platforms such as Windows, Linux, Apple, and Android. It does not require extra effort from the user as it converts the letters typed in Latin to their phonetic equivalents in Devanagari. Romanization is further helped by the fact that it provides straightforward and fast code-mixing between Hindi and English. This allows writing English words in Hindi rapidly and without confusion, which often arises in Latinagari (e.g. यूज़ vs यूज). It can be claimed that Romanization is promoting the use of English words in Hindi, many of which constitute  $\beta$ -Hindi words. Example of Romanized Hindi with English words in boldface: 'Aap konsa **color like** karte ho? Kaala ya **pink**?'

Both Latinagari and Romanization are promoted by two factors: 1) The rise of English in Hindi-speaking India, and 2) The borrowing of English words by Hindi. How does Latinagari or Romanization affect the integrity and effectiveness of Hindi, and can these processes be tweaked to enhance the language? It is appreciated that 'enhancement' of Hindi is a subjective term but we have to start somewhere and learn as we proceed. The regular organisation and standardisation of the English language [66] is one of the main factors behind its successful growth. Rather than absolute adherence to linguistic rules, it is flexible adaptation of those rules that make a language evolve and thrive. This gives people the independence of expression while maintaining its integrity. With the nature and pace of changes in Hindi, the rise of *Hinglish*, an important question arises: Will English replace Hindi and become India's main language in the next decades?

The question raised above draws our attention to linguistic concepts of language death and linguistic obsolescence [67, 68, 69, 70]. Aitchison [71] describes language death as a drastic event when a language disappears suddenly. In this sense, Sanskrit is not a dead language but a transformed language that lives through Hindi, Bengali, Marathi and many other descendants. Similarly, Latin changed to become French, Italian, Spanish, and other derivations. However, the loss of knowledge in language transformations of this kind cannot be denied, which is especially true in the case of Sanskrit. Many ideas that are fundamental to Bharat (today's India) were coded in Sanskrit and mostly in poetical form. These include the Vedas, Upanishads, Gita, Manusmriti, Arthashastra, Lilavati, and Natyashastra. The loss of Sanskrit meant the loss of direct knowledge of many ideas that matured in centuries. Here, the complicated problem of translating poetry [72] adds to the challenge of translating metaphysical concepts with multilayered meanings. This explains why we often find completely misaligned and contrasting translations of Vedic texts. Therefore, Sanskrit's 'transformation' has been detrimental to Indian thought.

We need to study the progress of Hindi, analyse how it is changing, and understand the impact of this change to find whether it is merely changing or slowly dying. Language death is further classified as *suicide* and *murder* [71]. Other classifications of language death can be found [70, 68], and linguists have progressed in understanding language shift and linguistic obsolescence [67, 68, 69, 70]. The meanings of the terms are not completely settled amongst scholars and the processes



are not mutually exclusive. Nevertheless, research on these linguistic processes can help us analyse the evolution of Hindi. For instance, in most cases of death of a weaker language by a dominant language, it is noted that the weaker language declines its lexicon and massively borrows words from the dominant language [68]. This can be related to the increase in  $\beta$ -Hindi words (Figure 2b), Latin words (Figure 6), and Romanization in Hindi text.

Dorian's research [73] on the death of Gaelic by English found that in using noun plurals, the less competent Gaelic speakers either used Gaelic inflection, omitted it, or used English equivalents. We find a comparable pattern in Hindi. The English word 'cycle' was adopted in its Latinagari form साईकिल more than 100 years ago. Its plural, साईकिलें (*cyclein*) was formed according to Hindi grammar. But now we find another variant of its plural, साईकल्स (*cycles*), which is a transliteration of the English plural. This usage is gaining popularity and many words like स्मार्टफोन्स (*smartphones*), फैन्स (*fans*), टिप्स (*tips*), and एक्टर्स (*actors*) are frequently used as seen in Figure 2c. This highlights changes in Hindi that were observed in the dying Gaelic language. Certainly, it does not mean that Hindi is dying but such analysis can lead to fruitful findings about the language.

## 7.2 Impact of change in Hindi

The change in Hindi can be viewed as just a natural language transformation that does not fall within any of the language phenomena mentioned above like linguistic obsolescence. How can we classify the transformation of Hindi and predict its behaviour? This is a significant question because language change can have far-reaching socio-political repercussions.

The Vietnamese script underwent diachronic digraphia when the Latin-based Quoc-ngu [74] writing system replaced the Chinese script (chu Nom) in 1950. In India, the Hindustani language followed two distinct writing systems: the original Devanagari script that became Hindi and the acquired Perso-Arabic script that became Urdu. In India, however, the distinction between Hindi and Urdu was not limited to linguistic styles. These languages became associated with religious identities—Hindus identified with Hindi and Muslims with Urdu. The language-politics link was so strong that Urdu was transported from central India and made the national language of the Islamic Republic of Pakistan [75], where the most spoken languages were Bengali and Punjabi. Moreover, language played a role in dividing Pakistan and creating Bangladesh in 1971. This demonstrates the far-reaching impact of languages.

The Internet and social media platforms such as Facebook, TikTok, X, and Whatsapp have accelerated language transformation. In the context of Hindi, this transformation is further amplified due to the uninhibited borrowing of words and stylistic elements from English. While the younger generation readily adapts to the evolving forms of Hindi, the middle-aged and elderly feel left behind. Memory generally declines with age [76], and old people may struggle to remember familiar words [77]. Consequently, they find it challenging to adapt to the rapid language changes that introduce a plethora of new vocabulary and syntactical structures. Furthermore, the practice of replacing Hindi words with English transliterations (as seen in Table 3) places an additional burden on many older people, especially those who are not proficient in English. The current transformations in Hindi are creating a communication gap between generations, and we must explore ways to ensure that this linguistic change benefits everyone.

## 7.3 English yoga effect

Yoga, one of the six schools of Hindu philosophy [78], has gained worldwide popularity as a physical, mental, and spiritual practice. The word comes from Sanskrit (योगः) into Hindi as 'yog' (योग), and

made its way into English as 'yoga' (pronounced /jəʊ.gə./). Interestingly, inspired by the English pronunciation, many Hindi speakers say 'yoga' or योग when speaking Hindi. This incorrect usage is intriguing because Hindi speakers do not have to adopt a foreign pronunciation of their word, and if they want to use the original Sanskrit pronunciation, they can exactly say योगः as visarg (विसर्ग ः) exists in Hindi. Other examples of this 'English yoga effect' are Shiva, Rama, Krishna, Ramayana, and Mahabharata.

The inclusion of 'a' at the end in English is an attempt to use the Sanskrit form by compensating the visarg (absent in English) with the suffix 'a'. The fact that many Hindi speakers are using incorrect versions of their own words and that too in their own language is indicative of the influence of English over Hindi at a psychological level.

## 7.4 Hindi writing and style

The noticeable shift in Hindi writing style warrants thorough quantitative analysis. The readily visible modifications of Latinagari and Romanization in Hindi have been researched [6, 7, 79, 80], and many other linguistic aspects such as, sentiments [81, 82, 83], are also researched. Here, we discuss two cases of modification in Hindi writing style.

- **purnaviraam (।) replaced by full stop (.):** This incorrect writing style is used by most news websites despite clear instructions on punctuation in standardised Hindi [31]. Using any keyboard format, such as QWERTY or Dvorak, if one can type the entire text in Hindi, they would not be saving time by using a single key press for a full stop instead of a two-key combination for a purnaviraam. Does this not show disregard for established writing rules? Of course, a similar change cannot be found in English writing on any reputed news website on the Internet.
- **Omission of chandrabindu (ँ):** Anunaasik or chandrabindu (अनुनासिक, चंद्रबिंदु ँ) is used to express sound that is produced through mouth and nose. Anusvaar or bindu (अनुस्वार ं) is used in place of the fifth consonant when permitted, e.g. गङ्गा = गंगा, सञ्जय = संजय, सम्भव = संभव. Despite Hindi standardisation [31] recommending using chandrabindu, Hindi news articles use a bindu instead. This usage can be seen in articles published before 2005, but as it is not standardised, the Hindi grammar books of today still advise the use of chandrabindu. While this could be a change that may be adopted, its implementation without standardisation sets a bad example and confuses readers. We note a contrasting situation in English publishing. The spelling of words in English news articles follows accepted usage and dictionaries, e.g. we do not find the BBC writing 'culture' as 'kulture'.

The English and French languages are promoted with admiration. Spelling Bee and Dicos d'or competitions demonstrate this. In 2023, the world's largest dictation event was held in Paris [84], where 1,397 people used the desks arranged on the famous Champs-Élysées street to take a French dictation test. We do not see any equivalent for Hindi. On the contrary, flaunting the lack of Hindi speaking ability and mixing English (at least, 'you know' in short, long, high, and low versions) in it is admired by some. This is illustrated in the interviews of some Bollywood actors and actresses.

We are very particular about writing correct English, but improper Hindi does not seem to be an issue. The worldwide desire to write good English has led to the development of advanced English writing tools for digital publishing, Word® and Internet browsers. These check spelling, grammar, and syntax. Lately, many of these tools have implemented artificial intelligence support to augment comprehension and style. Hindi, on the other hand, lacks adequate spell-check support

even today. If there is enough demand for correct and convincing Hindi, appropriate tools will become available. An intriguing inquiry would be to find in detail why good Hindi is not viewed as good as good English.

## 8 Controlling the change

In this work, unless remarked, we used the terms ‘change’, ‘evolution’, and ‘transformation’ interchangeably. The evolution of Hindi can be enhanced to benefit its speakers while preserving its integrity. Achieving this would involve implementing effective policies and timely measures. Analysing our results, observing popular trends, and from our discussion, we believe that the following can potentially help in controlling the transformation of Hindi.

- **Accessible Hindi code:** Sometimes we find several transliterations for the same word, e.g. वेरिंट, वैरिंट for variant and स्मार्टफोन, स्मार्टफोन for smartphone. Many modern applications often use some words, e.g. ‘Option’, ‘Delete’, and ‘Paste’, which are used in their transliterated versions, e.g. ऑप्शन, डिलीट, and पेस्ट and not as translations, e.g. विकल्प, मिटाएँ, and चिपकाएँ. For many novel scientific terms, Hindi can offer meaningful translations. However, as this does not happen timely or is not well known, people start using transliterated versions, which then becomes the norm.

An organisation or institution—a government body or a consortium—that promptly provides suitable Hindi translations and transliterations will be a source of missing modern Hindi code. Hindi users, including authors, students, bloggers, digital content creators and artificial intelligence agents, would freely access the web portal of this organisation to find the correct terms. This will prevent the use of improper words that become popular due to the initial lack of appropriate alternatives.

- **Hindi writing tools:** Integrating advanced Hindi writing tools into popular computer applications will promote good writing and style. Moreover, these tools can use the institution described above to include novel words, revised instructions, and up-to-date suggestions. They can further improve from artificial intelligence-based text analysis.
- **Engagement:** The impact of an organisation or application depends on its engagement with its users, which includes individuals and entities. The engagement of users with the organisation and tools described above will require encouragement, especially in the initial stage. This can be achieved by events, seminars, text and video guides, and meetings with mainstream publishers that elucidate the necessity and potential of the task.
- **Promotion:** Hindi is not promoted as well as English or French, especially among children. There are no television programs or organisations like the British Council or French Alliance that popularise Hindi in India and abroad. Hindi language games, challenges, and spelling championships will also help. Promotion of good Hindi can have an avalanche effect in shaping the future of the language.
- **Analysis, inference, and adaptation:** The absence of extensive analysis, esp. data-based quantified research, inhibits our understanding of the Hindi evolution. We cannot solve a problem unless we understand it, and as the situation evolves, the solutions need updating using feedback. Careful observation of the multiple dimensions of Hindi transformation will lead to meaningful inferences about the phenomena, which will provide input for adapting language policies appropriately.

## 9 Conclusion

By comparing the words used by reputable Hindi news websites to HindiShabd.v1, we found that between January 2009 and June 2024, Hindi incorporated new words at a high rate. Most of these new words are transliterations of English words, and Hindi alternatives exist for many of them. Another noticeable trend is the widespread use of Latin words in Hindi text, which increased by twenty-five times from 2009 to 2024. The growth of English in India and the lack of Hindi coding seem to foster this transformation.

In the past 15–25 years, the Hindi used by news publishers has undergone notable change, some of which have been quantified in this work. Implications of this change include reduced readability (due to unknown words), and reduced comprehension for non-English Hindi speakers (many words are from English). The rapid and uncontrolled transformation of Hindi is fuelled by the lack of its up-to-date frequent standardisation and adherence to standardisation. News publishers have not limited themselves to using new forms that are required to write about novel technology, they have unreasonably adopted grammatically incorrect conventions in some instances. The availability of the latest Hindi code and state-of-the-art writing tools for Hindi will help the language improve and evolve.

The complex interlinked relationship between language, culture, identity, and socio-economic status suggests that language transformation is not limited to words—it extends far beyond and shapes our identities and thought. Therefore, the evolution of Hindi invites research to understand its nature, its causes, and most importantly, its implications.

## Bibliography

- [1] P. Muysken, *Bilingual Speech: A Typology of Code-Mixing*. Cambridge: Cambridge University Press, 2000.
- [2] S. Poplack, "Sometimes i'll start a sentence in spanish y termino en espanol: toward a typology of code-switching1," *Linguistics*, vol. 18, 1980.
- [3] A. K. Joshi, "Processing of sentences with intra-sentential code-switching," in *Coling 1982: Proceedings of the Ninth International Conference on Computational Linguistics*, 1982. [Online]. Available: <https://aclanthology.org/C82-1023>
- [4] S. Thara and P. Poornachandran, "Code-mixing: A brief survey," in *2018 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*. IEEE, Sep. 2018. [Online]. Available: <https://doi.org/10.1109/icacci.2018.8554413>
- [5] The Local France, "French government unveils 18 new official translations for tech vocab," 27 May 2020, Accessed: 20 July 2024. [Online]. Available: <https://www.thelocal.fr/20200527/french-government-unveils-18-new-official-translations-for-tech-vocab>
- [6] R. D. Parshad, S. Bhowmick, V. Chand, N. Kumari, and N. Sinha, "What is India speaking? exploring the "Hinglish" invasion," *Physica A: Statistical Mechanics and its Applications*, vol. 449, pp. 375–389, May 2016. [Online]. Available: <https://doi.org/10.1016/j.physa.2016.01.015>
- [7] N. Neema and J. K. Chawla, "The dialectics of hinglish: A perspective," *Applied Linguistics Papers*, vol. 2/2018, no. 25, pp. 37–51, Jun. 2018. [Online]. Available: <https://doi.org/10.32612/uw.25449354.2018.2.pp.37-51>
- [8] T. Bhatia and W. Ritchie, "Multilingual language mixing and creativity," *Languages*, vol. 1, no. 1, p. 6, May 2016. [Online]. Available: <https://doi.org/10.3390/languages1010006>
- [9] Ethnologue, "What are the top 200 most spoken languages?" 2024, Accessed: 20 July 2024. [Online]. Available: <https://www.ethnologue.com/insights/ethnologue200>
- [10] O. E. Dictionary, "How words enter the oed," 2024, Accessed: 20 July 2024. [Online]. Available: <https://www.oed.com/information/editorial-policy/how-words-enter-the-oed>
- [11] Wikipedia, "Académie Française," July 2024, Accessed: 20 July 2024. [Online]. Available: [https://en.wikipedia.org/wiki/Acad%C3%A7mie\\_Fran%C3%A7aise](https://en.wikipedia.org/wiki/Acad%C3%A7mie_Fran%C3%A7aise)
- [12] —, "Alliance française," July 2024, Accessed: 20 July 2024. [Online]. Available: [https://en.wikipedia.org/wiki/Alliance\\_fran%C3%A7aise](https://en.wikipedia.org/wiki/Alliance_fran%C3%A7aise)
- [13] —, "Royal Spanish Academy," July 2024, Accessed: 20 July 2024. [Online]. Available: [https://en.wikipedia.org/wiki/Royal\\_Spanish\\_Academy](https://en.wikipedia.org/wiki/Royal_Spanish_Academy)
- [14] "Central Hindi Directorate," 2024, Accessed: 20 July 2024. [Online]. Available: <https://www.chd.education.gov.in/>
- [15] Govt. of India, "Department of Official Language," 2024, Accessed: 20 July 2024. [Online]. Available: <https://rajbhasha.gov.in/>
- [16] —, "Central Hindi Training Institute," 2024, Accessed: 20 July 2024. [Online]. Available: <https://chti.rajbhasha.gov.in/>
- [17] Alliance Française, "Discover alliance française," Accessed: 28 July 2024. [Online]. Available: <https://af-france.fr/discover-alliance-francaise/>
- [18] "Central Hindi Directorate Offices," 2024, Accessed: 20 July 2024. [Online]. Available: <https://www.chd.education.gov.in/en/contact-us>
- [19] BBC, "A-z complete guide," 2024, Accessed: 20 July 2024. [Online]. Available: <https://www.bbc.co.uk/newsstyleguide/all>
- [20] T. Guardian, "Guardian and observer style guide," 2024, Accessed: 20 July 2024. [Online]. Available: <https://www.theguardian.com/guardian-observer-style-guide-a>
- [21] UK Govt., "Style Guide: A to Z," 28 June 2024, Accessed: 20 July 2024. [Online]. Available: <https://www.gov.uk/guidance/style-guide/a-to-z-of-gov-uk-style>

- [22] R. M. Ritter, *The Oxford Guide to Style*. Oxford, UK: Oxford University Press, 2002.
- [23] University of Chicago Press, *The Chicago Manual of Style*, 17th ed. Chicago: The University of Chicago Press, 2017.
- [24] P. Peters, *The Cambridge Guide to English Usage*. Cambridge: Cambridge University Press, 2004.
- [25] Wikipedia, "List of style guides," Accessed: 20 July 2024. [Online]. Available: [https://en.wikipedia.org/wiki/List\\_of\\_style\\_guides](https://en.wikipedia.org/wiki/List_of_style_guides)
- [26] The European Union, *English Style Guide: A handbook for authors and translators in the European Commission*. European Union, May 2024.
- [27] S. Vincent, "Sentence length: Why 25 words is our limit: 2013; Inside GOV.UK — insidegovuk.blog.gov.uk," 4 August 2014, Accessed: 25 July 2024. [Online]. Available: <https://insidegovuk.blog.gov.uk/2014/08/04/sentence-length-why-25-words-is-our-limit/>
- [28] "Times of India," Accessed: 20 July 2024. [Online]. Available: <https://timesofindia.indiatimes.com/>
- [29] "Hindustan Times," Accessed: 20 July 2024. [Online]. Available: <https://www.hindustantimes.com/>
- [30] "The Pioneer," Accessed: 20 July 2024. [Online]. Available: <https://www.dailypioneer.com/>
- [31] केंद्रीय हिंदी निदेशालय (Central Directorate of Hindi), "देवनागरी लिपि तथा हिंदी वर्तनी का मानकीकरण," 2019, Accessed: 20 July 2024. [Online]. Available: <https://www.chd.education.gov.in/sites/default/files/devanagarilipiandhindivartanikamankikaran.pdf>
- [32] "Aajtak archive," 2024, Accessed: 10 June 2024. [Online]. Available: <https://www.aajtak.in/archive>
- [33] A. R. Dhawan, "HindiShabd.v1: Hindi word list in Devanagari script," 2024. [Online]. Available: <https://doi.org/10.7910/DVN/8YSAIN>
- [34] S. Dasa, Ed., *Hindi Sabdasagara*. Kasi: Nagari Pracarini Sabha: Navina samskarana, 1965-1975, vol. 1-10.
- [35] R. P. Saksena, Ed., *Wardha Hindi Dictionary*. Wardha: MGAHV, 2013.
- [36] सरलीकरण विशेषज्ञ समिति, सरल प्रशासनिक शब्दावली. राजभाषा विभाग, दिल्ली: केंद्रीय अनुवाद ब्यूरो, 2014.
- [37] रामचन्द्र शुक्ल, "हिन्दी साहित्य का इतिहास," 2009. [Online]. Available: <https://www.hindisamay.com/alochana/shukl%20granthavalis/hindisahity%20itih%20-%20shukl%20index.htm>
- [38] "हिंदी कहानी," 2024, Accessed: 20 July 2024. [Online]. Available: <https://hindikahani.hindi-kavita.com>
- [39] NLTK Team, "Nltk 3.8.1," January 2023, Accessed: 10 June 2024. [Online]. Available: <https://www.nltk.org/index.html>
- [40] P. Qi, Y. Zhang, Y. Zhang, J. Bolton, and C. D. Manning, "Stanza: A Python natural language processing toolkit for many human languages," in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, 2020. [Online]. Available: <https://nlp.stanford.edu/pubs/qi2020stanza.pdf>
- [41] B. Kremmel, B. Indrarathne, J. Kormos, and S. Suzuki, "Unknown vocabulary density and reading comprehension: Replicating hu and nation (2000)," *Language Learning*, vol. 73, no. 4, pp. 1127–1163, Nov. 2023. [Online]. Available: <https://doi.org/10.1111/lang.12622>
- [42] M. Hu and P. Nation, "Unknown vocabulary density and reading comprehension," *Reading in a Foreign Language*, vol. 13, 2000. [Online]. Available: <https://doi.org/10.26686/wgtn.12560354.v1>
- [43] I. Rets, L. Astruc, T. Coughlan, and U. Stickler, "Approaches to simplifying academic texts in english: English teachers' views and practices," *English for Specific Purposes*, vol. 68, pp. 31–46, Oct. 2022. [Online]. Available: <https://doi.org/10.1016/j.esp.2022.06.001>
- [44] B. K. Matilal, *The Word and the World: India's Contribution to the Study of Language*. Delhi: Oxford University Press, 1992.
- [45] P. Jain, "UPSC छात्र की करेंट से मौत मामले में . . .," 24 July 2024, Accessed: 28 July 2024. [Online]. Available: <https://rb.gy/seutrth>
- [46] M. A. Just and P. A. Carpenter, "A theory of reading: From eye fixations to comprehension." *Psychological Review*, vol. 87, no. 4, pp. 329–354, 1980. [Online]. Available: <https://doi.org/10.1037/0033-295X.87.4.329>
- [47] Wikipedia, "Grammarly," 2024, Accessed: 2 August 2024. [Online]. Available: <https://en.wikipedia.org/wiki/Grammarly>
- [48] F. Sharifian, Ed., *The Routledge Handbook Of Language and Culture*, ser. Routledge handbooks in linguistics. London: Routledge, 2015.

- [49] J. Simpson, Ed., *The Routledge Handbook of Applied Linguistics*, ser. The Routledge Handbook of Applied Linguistics. London: Routledge, 2013.
- [50] P. Auer, Ed., *Code-switching in Conversation: Language, Interaction and Identity*. London: Routledge, 2013.
- [51] Govt. of India, "National policy on education 1986," Accessed: 28 July 2024. [Online]. Available: [https://www.education.gov.in/sites/upload\\_files/mhrd/files/upload\\_document/npe.pdf](https://www.education.gov.in/sites/upload_files/mhrd/files/upload_document/npe.pdf)
- [52] S. N. Sridhar and K. K. Sridhar, "English in india's multilingual ecology: Present-day use, users and usage," in *Language Studies in India: Cognition, Structure, Variation*, R. Kumar and O. Prakash, Eds. Singapore: Springer Singapore Pte. Limited, 2023.
- [53] S. W. Fitriati and E. Rata, "Language, globalisation, and national identity: A study of english-medium policy and practice in indonesia," *Journal of Language, Identity & Education*, vol. 20, no. 6, pp. 411–424, Jul. 2020.
- [54] S. Bax, "Researching english bilingual education in thailand, indonesia and south korea." [Online]. Available: [https://www.teachingenglish.org.uk/sites/teacheng/files/publication\\_2\\_-\\_researching\\_english\\_bilingual\\_education.pdf](https://www.teachingenglish.org.uk/sites/teacheng/files/publication_2_-_researching_english_bilingual_education.pdf)
- [55] J. Bernhofer and M. Tonin, "The effect of the language of instruction on academic performance," *Labour Economics*, vol. 78, p. 102218, Oct. 2022. [Online]. Available: <https://doi.org/10.1016/j.labeco.2022.102218>
- [56] R. B. Buxton, *Introduction to Functional Magnetic Resonance Imaging: Principles and Techniques*, 2nd ed. Cambridge: Cambridge Univ. Press, 2009.
- [57] G. H. Glover, "Overview of functional magnetic resonance imaging," *Neurosurgery Clinics of North America*, vol. 22, no. 2, pp. 133–139, Apr. 2011.
- [58] M. Hasegawa, P. A. Carpenter, and M. A. Just, "An fmri study of bilingual sentence comprehension and workload," *NeuroImage*, vol. 15, no. 3, pp. 647–660, Mar. 2002. [Online]. Available: <https://doi.org/10.1006/nimg.2001.1001>
- [59] S. Kim, S. G. Jeon, Y. Nam, H. s. Kim, D.-H. Yoo, and M. Moon, "Bilingualism for dementia: Neurological mechanisms associated with functional and structural changes in the brain," *Frontiers in Neuroscience*, vol. 13, Nov. 2019. [Online]. Available: <https://doi.org/10.3389/fnins.2019.01224>
- [60] N. C. Singh and C. Rao, "Reading in devanagari: Insights from functional neuroimaging," *Indian Journal of Radiology and Imaging*, vol. 24, no. 01, pp. 44–50, Feb. 2014. [Online]. Available: <https://doi.org/10.4103/0971-3026.130691>
- [61] A. J. DeCasper and M. J. Spence, "Prenatal maternal speech influences newborns' perception of speech sounds," *Infant Behavior and Development*, vol. 9, no. 2, pp. 133–150, Apr. 1986. [Online]. Available: [https://doi.org/10.1016/0163-6383\(86\)90025-1](https://doi.org/10.1016/0163-6383(86)90025-1)
- [62] B. Mampe, A. D. Friederici, A. Christophe, and K. Wermke, "Newborns' cry melody is shaped by their native language," *Current Biology*, vol. 19, no. 23, pp. 1994–1997, Dec. 2009. [Online]. Available: <https://doi.org/10.1016/j.cub.2009.09.064>
- [63] Ethnologue, "Hindi," Accessed: 28 July 2024. [Online]. Available: <https://www.ethnologue.com/language/hin/>
- [64] Wikipedia, "Romanization — Wikipedia, the free encyclopedia," 2024, Accessed: 28 July 2024. [Online]. Available: <http://en.wikipedia.org/w/index.php?title=Romanization&oldid=1230647528>
- [65] R. Mhaskar, "Romanagari an alternative for modern media writings," *Bulletin of the Deccan College Research Institute*, vol. 75, pp. 195–202, 7 2015. [Online]. Available: <https://www.jstor.org/stable/26264736>
- [66] J. Algeo, *The Origins and Development of the English Language*, 6th ed. Wadsworth Cengage Learning, 2010.
- [67] N. C. Dorian, "The fate of morphological complexity in language death: Evidence from East Sutherland Gaelic," *Language*, vol. 54, no. 3, p. 590, Sep. 1978. [Online]. Available: <https://doi.org/10.2307/412788>
- [68] W. Wolfram, "Language death and dying," in *The Handbook of Language Variation and Change*. Wiley, 2004, pp. 764–787. [Online]. Available: <https://doi.org/10.1002/9780470756591.ch29>
- [69] A. M. S. McMahon, *Understanding Language Change*. Cambridge: Cambridge Univ. Press, 2010.

- [70] L. Campbell and M. C. Muntzel, "The structural consequences of language death," in *Investigating obsolescence: Studies in language contraction and death*, N. Dorian, Ed. Cambridge: CUP, 1989, pp. 181–196.
- [71] J. Aitchison, *Language Change: Progress or Decay?*, 3rd ed. Cambridge University Press, 2001.
- [72] K. Risager, "Linguaculture: The language–culture nexus in transnational perspective," in *The Routledge Handbook Of Language and Culture*, ser. Routledge handbooks in linguistics, F. Sharifian, Ed. London: Routledge, 2015.
- [73] N. C. Dorian, *Language Death: The life cycle of a Scottish Gaelic dialect*. Philadelphia, Pa.: Univ. of Pennsylvania Press, 1981, bibliogr. S. 188 - 196.
- [74] Encyclopedia Britannica, "Quoc-ngu," 10 December 2014, Accessed: 28 July 2024. [Online]. Available: <https://www.britannica.com/topic/Quoc-ngu>
- [75] A. Ayres, *Speaking Like a State: Language and Nationalism in Pakistan*. Cambridge University Press, Jul. 2009. [Online]. Available: <https://doi.org/10.1017/CBO9780511596629>
- [76] F. I. Craik, "Memory changes in normal aging," *Current Directions in Psychological Science*, vol. 3, no. 5, pp. 155–158, Oct. 1994. [Online]. Available: <https://doi.org/10.1111/1467-8721.ep10770653>
- [77] D. M. Burke and M. A. Shafto, "Aging and language production," *Current Directions in Psychological Science*, vol. 13, no. 1, pp. 21–24, Feb. 2004.
- [78] Encyclopedia Britannica, "Yoga," 26 June 2024, Accessed: 31 July 2024. [Online]. Available: <https://www.britannica.com/topic/Yoga-philosophy>
- [79] A. Srivastava, K. Bali, and M. Choudhury, "Understanding script-mixing: A case study of Hindi-English bilingual twitter users," in *Proceedings of the 4th Workshop on Computational Approaches to Code Switching*, 2020, pp. 36–44. [Online]. Available: <https://aclanthology.org/2020.calcs-1.5.pdf>
- [80] A. Pratapa, G. Bhat, M. Choudhury, S. Sitaram, S. Dandapat, and K. Bali, "Language modeling for code-mixing: The role of linguistic theory based synthetic data," in *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 2018, pp. 1543–1553. [Online]. Available: [https://www.microsoft.com/en-us/research/uploads/prod/2018/05/language\\_modeling\\_cm.pdf](https://www.microsoft.com/en-us/research/uploads/prod/2018/05/language_modeling_cm.pdf)
- [81] N. Mittal, B. Agarwal, G. Chouhan, P. Pareek, and N. Bania, *Discourse Based Sentiment Analysis for Hindi Reviews*. Springer Berlin Heidelberg, 2013, pp. 720–725. [Online]. Available: [https://doi.org/10.1007/978-3-642-45062-4\\_102](https://doi.org/10.1007/978-3-642-45062-4_102)
- [82] R. Sharma and P. Bhattacharyya, "A sentiment analyzer for Hindi using Hindi senti lexicon," in *Proceedings of the 11th international conference on natural language processing*, 2014, pp. 150–155. [Online]. Available: <https://aclanthology.org/W14-5124.pdf>
- [83] A. Joshi, A. Prabhu, M. Shrivastava, and V. Varma, "Towards sub-word level compositions for sentiment analysis of Hindi-English code mixed text," in *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, Y. Matsumoto and R. Prasad, Eds. Osaka, Japan: The COLING 2016 Organizing Committee, Dec. 2016, pp. 2482–2491. [Online]. Available: <https://aclanthology.org/C16-1234>
- [84] BBC, "France: Paris Champs-Élysées hosts mass spelling contest," 5 June 2023, Accessed: 28 July 2024. [Online]. Available: <https://www.bbc.co.uk/news/world-europe-65809523>