

Mapping Linguistic Innovation, Cultural Continuity, and Socio-Economic Disparity Across Indian States – A Culture Based Development Application

J. D. Prabhakar¹, Annie Tubadji²

¹ GITAM University, Hyderabad, India

² Swansea University, Swansea, United Kingdom

Received: 28 April 2026/Accepted: 10 September 2026

Abstract. This study explores how language evolution and social reality meet in India's varied states. Adopting the Culture Based Development lens, we quantify local culture as two sub-types, cultural heritage (CH) and living culture (LC), and the balance between them (cultural entropy– CE), where the latter is theoretically expected to explain socio-economic inequality between cultural groups in a place. Thus, the research explores how a culturally different balance across India is associated with socio-economic inequality across space. To achieve this, we use language as a proxy for cultural attitudes. Namely, by comparing the frequency of traditional cultural markers (“Krishna”) (approximating cultural heritage) to current linguistic innovations (“Influencer”) (approximating living culture) we check the statistical association of these linguistic variables with measures for English language dominance and caste-wise unemployment. The findings indicate that, while living-culture-related terms increase their frequency over time, they have not yet superseded traditional cultural moorings, i.e., the cultural entropy in Indian places is still dominated by the cultural heritage component. Hence, we still observe local inefficiencies and inequalities in terms of structural unemployment especially among Scheduled Castes (SC) and Scheduled Tribes (ST). The findings are explained from socio-linguistic and Myrdalian point of view.

Key words: Culture Based Development, Cultural Entropy, Krishna, Influencer, Unemployment, Linguistic Imperialism

1 Introduction

Language is a reflective mirror and therefore, we argue, a powerful proxy of culture, tradition, and economy throughout societies. Language variations are indicative of cultural and traditional differences (see [Gumperz 1982](#)). Additionally, language serves as a means of transferring knowledge systems created by one generation to the next ([Sapir 1921](#)). In the past, the native language or languages used across all spheres of society typically became high-resource languages, developing strong digital presence and widespread use among the population. Many languages vanished from the surface of the earth as a result of language hegemony, colonisation, and dominance ([Crystal 2000](#)). Over 230 languages have already vanished since 1950, and about 40% of the world's languages are endangered ([UNESCO 2010](#)). Because language loss contributes to the loss of culture and customs,

it has recently gained international attention. Alternatively,, when society's customs and culture have been conserved and developed over time, the cultural wealth is reflected in the wealth of local languages with abundant resources which make a crucial component of local cultural heritage. Culture-centric development allowed a culturally and linguistically diverse nation like India to flourish as a multilingual and multicultural nation. Yet, its development is uneven across different regions (see [Nayyar 2019](#)). This makes India a perfect case for studying how the local dynamics and composition of language relate to local cultural dynamics and its impact on local development.

It has been argued that a high-resource language is a sign of that country's growth and cultural enrichment. Individuals with abundant linguistic resources can be viewed as effective economic agents capable of actively and fully engaging in economic activity. According to a study done in the Indian state of Telangana, 94% of industrial economic activity in India is carried out in Indian languages, which boosts the country's GDP ([Prabhakar 2024](#)). The latter study contends that native tongues serve as a potential source of human capital for economic endeavours. In addition, [Blommaert \(2010\)](#) argues that language reflects broader social structures shaped by globalisation and power relations. Social mobility, inequality, and institutional resource access are frequently measured by linguistic practices. As a result, linguistic patterns can disclose underlying socioeconomic hierarchies in communities ([Blommaert 2010](#)). Language serves as a vital conduit for cultural identity and collective memory, also according to [Fishman \(1991\)](#). Languages pass down cultural values from generation to generation through family, religion, and local institutions. Thus, there is socio-linguistic argument and evidence for the relation between language, culture and development in general and in India in particular.

Drawing on the linguistic advocacy of [Uma Maheshwar Rao \(2017\)](#), the institutionalisation of Telugu as a primary medium of instruction can be seen as a key intervention to counter the systemic erosion of regionally grounded knowledge systems. The concentration of information within a single dominant language creates a linguistic monopoly, effectively marginalising other tongues in academic and professional contexts. This dominance poses a direct threat to linguistic diversity, fostering a cycle of endangerment where native speakers feel compelled to forgo their mother tongue for global economic mobility. By institutionalising Telugu within the educational framework, the state can elevate it into a high-resource language, safeguarding the unique cultural and intellectual systems of its speakers against the encroachment of a singular lingua franca. This aspect of the literature indicates that the cultural (institutional) channel of discrimination in the educational sector is why language reflects closely the culture-development relationship.

Theoretically and practically, this means a shift in perceiving mother tongue as a dynamic economic engine rather than a dormant cultural relic. Given that the bulk of industrial and local commerce already operates in native languages, aligning the education system with this reality bridges the 'comprehension gap' that currently stifles innovation. Utilising a native language (such as Telugu) as a foundational pillar democratises economic agency - empowering rural and semi-urban populations to transition from passive labor to active leadership. By scaling localised knowledge systems - from agricultural expertise to service sectors - this approach transforms linguistic integrity into a strategic asset for sustainable growth and collective mobility.

Importantly while various cultural and economic factors drive social inequality in India, caste-based stratification remains the definitive engine of systemic imbalance. This hierarchy is pervasive, transcending the rural-urban divide and manifesting clearly in rigid occupational structures and disproportionate unemployment rates. As [Deshpande \(2011\)](#) argues, a pervasive 'grammar of caste' continues to dictate economic outcomes despite the forces of modernisation. To address these persistent socio-economic disparities and dismantle entrenched discrimination, the Constitution of post-independence India institutionalised reservations in public employment. While initially envisioned as a decadal measure, these affirmative action policies have been periodically extended to foster enduring social equity - a necessity underscored by [Thorat, Newman \(2010\)](#) in their analysis of blocked economic mobility. Despite these institutional efforts, disparities in employment opportunities persist because discriminatory practices have become deeply embedded within the cultural fabric ([Deshpande 2011](#)). Members of the Scheduled Castes

(SC), Scheduled Tribes (ST), and other backward communities continue to be disproportionately affected across generations, often facing what [Ambedkar \(1936\)](#) identified as a systemic exclusion that goes beyond simple economics. These inequalities manifest in every sector, dictated by caste, religion, and socio-economic status. Addressing these issues in the contemporary era requires a rigorous, scientific research methodology to yield viable solutions. Such rigorous empirical work on language, culture and socio-economic development per group with different cultural power (i.e., experiencing different degree of cultural ‘taste for discrimination’ locally) is still very scarce.

This paper uses the wisdom from the above depicted literature to channel it through the lens of the Culture Based Development (CBD) paradigm ([Tubadji 2025a](#)), in order to secure an advanced study of language as a proxy for cultural impact on development, viewed through different degrees of cultural taste for discrimination towards different groups co-existing across the regions of India.

The CBD Paradigm postulates that the moral code of valuation, termed culture, consist of attitudes towards every item of reality in terms of how good and bad this item is for one’s survival. The attitudes are partially inherited from past generation (Cultural Heritage – CH) and partially acquired through own experience (Living Culture – LC). The balance between the amount of CH and LC reflects to what degree people will stick to old cultural morals and to what degree they will be open to novelty. CBD theoretically argues this balance (termed Cultural Entropy) predicts essentially the size of the taste for discrimination towards what is novel or different than what is familiar to the incumbent in a region ([Tubadji 2025a,b](#)). CBD has also argued that among other proxies for CH and LC, language can be used to quantify the positive valuation of items from reality that is proxied by the frequency of use of the word that signifies this item. If the word ‘bread’ has higher frequency than ‘rice’ locally, CBD argues this is a proxy for bread being more valuable to people than rice locally (see [Tubadji, Webber 2023](#), [Tubadji 2025a](#)).

This paper extends the Culture Based Development (CBD) perspective by positing that language serves as a mirror of societal structures and a form of “symbolic power” ([Bourdieu 1991](#)). As such, language as a proxy for culture can reflect, capture, and quantify the effects of culturally embedded discrimination experienced by groups occupying different positions within local cultural power hierarchies. To operationalise the CBD theoretical notions of cultural dynamics through measures for linguistic dynamics, this study adopts the following approach. We quantify linguistically the CH component of culture through the frequency of use of the word “Krishna” which is loaded with traditional cultural implications. We quantify LC with the frequency of use of the word “Influencer” – a word that has gained a lot of importance only lately with the high technological development and the new age of digital influencers’ activity. By scrutinising the frequency of these terms and which of the two prevails locally in use in language and by how much over the other, these terms and their use in contemporary discourse allows us to approximate the CBD notion of Cultural Entropy. Thus, once we have secured a linguistic approximation of the main CBD theoretical notions, this study seeks to investigate how language reflects the evolving power of cultural capital and how this, in turn, shapes social status and unemployment outcomes across different groups within a region.

The implications of this study are wider than just extending the CBD take on language as a proxy for culture. Developmental economists have long recognised culture as a fundamental force shaping economic behavior and institutional trajectories. Early scholarship established that cultural norms and social hierarchies are not merely peripheral but central to perpetuating structural inequality across generations ([Myrdal 1944](#)). Building on this foundation, contemporary research has increasingly integrated cultural variables into regional growth models. Culture Based Development research has shown that cultural discrimination relates to similar Myrdalian findings also today and in Europe (see [Tubadji et al. 2017](#)). Within this evolving development economics discourse, the CBD framework provides a rigorous, systematic methodology for analysing how cultural values dictate economic decision-making and drive regional development outcomes. According to earlier empirical studies ([Tubadji 2012](#)), economic performance is impacted by cultural attitudes ingrained in regional institutions. This argument is further supported by later studies that demonstrate how cultural configurations influence regional devel-

opment patterns throughout Europe (Tubadji 2013). But finding the multi-dimensional measurement that CBD requires in its unique quantitative approach is often challenging as cultural statistics is scant worldwide. Our study points out powerfully that language – whose footprint is freely available especially in the digital age, can be a readily available multi-dimensional proxy of culture (offering as many individual indicators as many words exist in the language). The CBD framework has offered to development economics a further advanced quantitative measurement based on the mathematics of Complexity Theory – the Cultural Entropy measure. The most recent theoretical and empirical CBD work models cultural bias through this complex measure, Cultural Entropy, that requires not only multi-dimensional measurement of culture but also precisely quantifying the balance between the parts (CH and LC) that build the local cultural capital (Tubadji 2025a,b). Our paper demonstrates that the linguistic proxies are able to quantify also Cultural Entropy. Thus, our paper suggests that language can be used as a proxy for the entities CH and LC and their balance to predict the degree of cultural discrimination between groups with different cultural power locally (such as castes) in all instances relevant for development economics, potentially entrepreneurship, innovation, education, and not just unemployment as the direct results of our study demonstrate. Further extension exploring the language-education-culture nexus is particularly relevant to shed more light on the power and meaning of the linguistic proxy in development economic contexts.

In particular, through our case study, important questions concerning the connection between linguistic innovation, cultural continuity, and socioeconomic inequality among Indian states are brought up by these divergent linguistic dynamics. On the one hand, collective memory and social identity are still shaped by traditional cultural symbols. However, emerging forms of social and economic transformation are reflected in new linguistic expressions linked to digital culture and contemporary economic roles. The coexistence of these linguistic patterns presents a special chance to conduct an empirical investigation of Cultural Entropy and to show its practical implication in policy-making relevant context.

Thus, in a nutshell, our study investigates the distribution of linguistic expressions representing cultural innovation and heritage across Indian states, as well as the relationship between these patterns and socioeconomic inequality. Put differently, we manage to investigate through a linguistic proxy the relationship between cultural dynamics and structural indicators, such as unemployment differences between castes. Using the English language, and particular linguistic indicators that references to traditional cultural symbols (like Krishna) and modern social roles (like influencer) we operationalise the full CBD theoretical and empirical toolkit. By applying the Culture Based Development framework through linguistic data, the study seeks to provide new insights into how language can be used and is powerful for studying cultural structures and their interaction with economic inequalities. We use the case of India's diverse socioeconomic landscape and demonstrate the power of language to capture quantitatively the local information about the differences in local cultural power between groups and to trace the socio-economic inequalities emerging from these differences.

The rest of the paper offers a focused literature review on the main streams of literature that support our reasoning and outline the gaps that specifically motivate our estimation strategy and related analysis. A description of our data and a descriptive analysis for the case of India. We offer this contribution as a pointer to the power of language to quantify cultural discrimination through the CBD paradigm and suggest it as a starting point for further research in this domain.

2 Literature Review

2.1 *Language, culture, and economic development*

Many interdisciplinary research studies in linguistics, economics, and sociology have tried to see economic development through their lens. From the astonishing works of developmental economists, a link was found between culture and economic development, arguing that economic development outcomes are influenced by the culture, institutions, and val-

ues of a society. Gunnar Myrdal's work (Myrdal 1944) highlighted that a country's economic development is deeply rooted in non-economic factors, such as culture, social norms, and institutions. His other work, 'An American Dilemma' (Myrdal 1944), showed the prevailing racial discrimination and how it was affecting the economic outcomes by limiting access to education, employment, and resources. His key concept, "circular cumulative causation," elucidates that social factors and economic factors mutually reinforce one another in a circular manner, which in turn influences development. He strongly supports the idea that cultural and social factors influence a country's economic growth and development. Myrdal has also a major work, relevant to our case study: 'Asian Drama' (Myrdal 1968). This contribution talks about institutions based in India, like the caste system and the beliefs associated with it, which have slowed economic progress and people's resistance to change. This confirms that the Myrdalian processes might be very related to our case study.

CBD is a framework that has long focused on Myrdalian processes and is strongly specialised on quantifying culture. Tubadji's recent CBD book (Tubadji 2025a) presents a synthesis of the fundamental principles of the CBD paradigm that postulate that culture is a multidimensional factor that shapes an individual's economic decision-making and through aggregation of these individual choices culture affects the aggregate economic system. CBD argues that people make economic decisions within a cultural framework that is shaped by evolving in time but generally sticky cultural beliefs and traditions. From this foundation, Culture Based Development (CBD) then proposes that the balance between evolution and stickiness in culture (termed Cultural Entropy) is the main source of the cultural bias studied by CBD.

CBD paradigm argues that culture is a measurable force that affects economic outcomes. Yet, crucially, to quantify culture and its evolution versus stickiness balance (i.e., cultural bias), the CBD highlights the importance of quantifying the complex nature of culture. Namely, according to CBD, the size of this cultural bias depends on the balance between what Tubadji (2012) distinguished as the adaptive side of culture, which is the living culture (LC), and the culturally sticky inherited side of culture, termed as cultural heritage (CH). Regions where culture is more balanced between CH and LC shape values that promote innovation and creativity, and are open to diversity, and therefore can experience better economic outcomes (Tubadji 2025a).

Tubadji's work on CBD in the regions of China (Tubadji 2023) showcases various quantitative approaches that CBD uses for extracting LC and CH measures from multi-dimensional measures of culture. Yet, notably, Tubadji (2025b) asserts that the first CBD law about the quantification of culture requires at least one mono-dimensional proxy for cultural heritage and one for living culture to be used – a condition Tubadji terms 'just identified culturally model'. CBD research has used monodimensional linguistic measures for culture in the past (see Tubadji, Webber 2023). Up to our knowledge, this paper is the first to use a multi-dimensional linguistic proxy for CH and LC to obtain the CBD Cultural Entropy measure. Therefore, this is a gap we would like to contribute towards here. This study wants to show that language can be used to quantify and express the statistical power of CH, LC and Cultural Entropy.

2.2 *Language as a proxy for cultural attitudes*

Culture (as values and attitudes) can be seen as inherently complex and heterogeneous, whereas language offers a practical and observable means of examining cultural attitudes, because every item from reality is clearly related to a linguistic unit, a word, that can be easily observed and counted. The latter is particularly fortunate, because culture and language are related hence language can be used as a proxy for culture. As Blommaert (2010) argues, language manifests widespread social structures formed by globalisation and power relations. In everyday use, linguistic practices often show differences in social mobility, inequity and access to institutional resources. Therefore, patterns of language use can provide meaningful insights into the socioeconomic hierarchies that exist within societies (Blommaert 2010). Similarly, Fishman (1991) emphasises the role of language as a carrier of cultural identity and collective memory, transmitted across generations through family, religion, and community institutions. In this sense, preserving language

becomes essential for sustaining cultural continuity and social cohesion.

According to [Fishman \(1991\)](#), language is deeply rooted in habitual interaction, which is an identity marker. He also integrates the concept of collective memory; the narratives that a group carries have a collective memory attached to them. It's not just a narrative or language a generation is carrying but also their identity, culture, and community wholeness (see [Fishman 1991](#)). Thus, language can be accepted as a proxy for cultural heritage (CH).

[Grin \(2003\)](#) conceptualises language as a form of human capital by highlighting the concept of 'linguistic externalities'. The benefits of a widely shared language help in international trade markets, diaspora linkages, connecting business networks, etc. He draws attention to the linguistic dominance in global markets and linguistic externalities entrenched in trust-gaining. Thus, linguistics has already agreed that language can be used to approximate cultural bias (which is Cultural Entropy in CBD terms).

Beyond linguistic studies, CBD itself has shown that language can express quantitatively the cultural weight. Language acts as a footprint in the global marketplace, directly indicating how economies culturally interact across borders. For instance, [Tubadji, Webber \(2023\)](#) demonstrate that the linguistic measures can predict the cultural ties between UK and China and can predict future patterns in bilateral trade. [Tubadji, Webber \(2023\)](#) reveals that when a shared language emerges, this actively associates with lowering of the invisible cultural 'transaction costs' of doing business¹. Thus, in line with linguistic literature motivation, our current study goes beyond [Tubadji, Webber \(2023\)](#) by approximating through language all CH, LC and Cultural Entropy in the CBD-recommended multi-dimensional manner.

2.3 Linguistic imperialism and English dominance

Linguistic imperialism discusses the dominance of one language over other languages across space. For instance, it can be applied as a term to describe that English dominates and is widespread in use worldwide which relates to power dynamics of history. The concept of 'linguistic imperialism' was introduced by [Phillipson \(1992\)](#) and provides a framework for understanding colonial dominance, which is tied to global power relations rather than merely a functional one. The institutionalisation of English in colonial spaces reveals structural inequalities among English-speaking members by marginalising local language speakers. It portrays English as inherently superior, thereby creating a linguistic hierarchy within global power relations. As [Kachru \(1994\)](#) classified it, worldwide, there are three different regions depending upon their usage of the English language: the inner circle, the outer circle, and the expanding circle. Where countries like India come under the outer circle, in which English is a functional language in administration, which is institutionalised in the local cultures. Kachru argues that even though English prevails as the outer circle in India, its spread was uneven in the region, explicitly overpowered in urban and elite spaces ([Kachru 1994](#)).

Relatedly, [Mohanty \(2018\)](#) argues that in the Indian context, by having English in education, administration, and economic life, it formed a hierarchy in the linguistic order by marginalizing regional and indigenous language speakers. His concept of 'subtractive multilingualism' talks about the devaluing and fading away of the mother tongue languages by the English language ([Mohanty 2018](#)). The superimposition of English as a medium of education creates a fear in native-language speakers by delimiting their participation in education and employment.

Based on the above literature, we can expect that language usage will be varying across the states (regions) of India, and that this will associate with different cultural discrimination to the same time of cultural groups (such as casts) across the different regions².

¹Ultimately, this may even indicate that language is far more than a passive cultural proxy, but itself a vital economic asset that facilitates real-world exchange.

²Further research on the various local linguistic groups and their share locally in comparison to using English is also very relevant as further research but is limited by data availability to be pursued in this study.

2.4 *Cultural continuity, linguistic innovation, and inequality*

Language can carry footprints from both inherited cultural markers and modern cultural innovations. And indeed, in living languages, there is often a variation in language reflecting the cultural shift in societies expressed in the prevailing use of the contemporary linguistic innovations over the linguistic expressions associated with traditional cultural markers. The traditional cultural markers are deeply rooted in society, like religion, beliefs, and shared narratives, and are carriers of collective memory which is locally fundamental for maintaining social solidarity. Thus cultural identities evolved in history and strongly laid their foundations around certain cultural markers. Frequency of use of language (words) referring to such inherited cultural markers – as religion for example – approximates the importance of the inherited factor in the time of use of the language. Linguistic innovations in recent times are evolving as a reflection of new market forces, new occupational identities, changing social trends, and shifting aspirations. The frequency of use of words referring to such new phenomena and entities (such as the new quasi-occupational term ‘Influencer’) can approximate the cultural and socio-economic innovation and reflects its importance in the minds of the local population in the time of use of language.

The language innovations are rapidly responding to socio-economic changes. Whereas traditional cultural expressions are slower to economic change and retain historical depth. This is consistent with how CBD described culture and its two components, CH and LC. CH is sticky, while LC is the adaptive innovative part of culture. Hence, one more consistency that allows using language as a proxy for culture.

In the Indian context, as discussed in [Myrdal \(1944\)](#), people’s responses to change vary depending on the cultural conditions embedded in their prevailing social institutions. The caste system and cultural hierarchies limit the marginalised groups from accessing economic opportunities. [Deshpande \(2011\)](#) showed that the scheduled castes (SC) and scheduled tribes (ST) in India are excluded from many economic opportunities, and sometimes they are segregated and hierarchically placed in an occupational ladder because of their pre-defined positioning in the cultural setting. So, when we talk about cultural continuity, it is not neutral with respect to Indian geography. Put differently, language and its balance between CH and LC may be statistically related in a different manner across different groups such as castes because of their different cultural power in society. This is one of the novel, Myrdalianly-relevant aspects, of our contribution.

2.5 *Towards a Cultural-Linguistic Approach to Regional Inequality*

Let us sum up here the main take and rationale of the current study. What do we propose?

Rather than treating culture as an abstract notion, the CBD approach ([Tubadji 2025a](#)) provides a systematic way to incorporate cultural dynamics into economic analysis by handling culture as a quantifiable complex dynamic entity. Our novel proposition here is that language can be used as a proxy that can fully quantify the complex dynamic entity culture according to the CBD advanced criteria for quantifying culture and the state of its entropy.

Countries like India, which comprise multilingual societies where language is intertwined with cultural variables, provide unique linguistic data for examining the cultural foundations of development. In such complex cultural context, language may actually prove one of the most precise proxies for the aggregate state of culture. In addition, it is well known that language determines access to opportunities by encouraging the individual to participate in education and economic markets ([Bourdieu 1984, 1991, Sen 2004, Suresh 2026](#)). Thus, a language-based measure is expected to be statistically powerful predictor for socio-economic development as we know from quantitative linguistics and CBD brings this to a new level of sophistication and precision of quantification through the Cultural Entropy measure. Through the here-proposed linguistic CBD approach, one can use language as a source for the multi-dimensional quantification of culture as the latter is recommended by CBD; researchers can probe to see the variability between socio-economic indicators, unemployment, and linguistic expressions in a way going beyond the standard quantitative linguistic manner by employing the CBD measurement

tools. It also helps in mapping very comparably the results throughout the regions because count of words and frequency of use of the words are very statistically normalised pieces of information. Thus, language can open a very fruitful window to quantification of culture in the CBD approach to allow well-identified cultural models and significantly help to avoid cultural-under-specification of studies of economic inequality across space.

3 Data

This study constructs a novel dataset derived from Google Search, capturing the search frequency of the terms “Krishna” and “Influencer” over comparable 24-hour periods across the regions of India. By systematically collecting search interest (i.e., frequency of search) for both terms within the same temporal window with geographic granularity, the dataset provides a previously unavailable, directly comparable linguistic footprint of how current regional linguistic corpora differ in their use of specific culturally loaded words across space.

The choice of the two focal terms is theoretically grounded in the Culture Based Development (CBD) paradigm and supported by the previously reviewed literature. Specifically:

“**Krishna**” serves as a proxy for cultural heritage (CH) in the local cultural capital, approximating language usage associated with tradition, historical identity, and inherited cultural narratives and religion (Weber 1905, on religion as culture).

“**Influencer**” serves as a proxy for living cultural (LC) in the local cultural capital, capturing contemporary language related to modernity, novel cultural lifestyle trends, and present-day cultural-economic entities (see also Bourdieu 1986, Trigg 2001, on form of life).

Together, these two words jointly approximate the dual components of cultural capital central to the CBD framework – CH & LC (see Tubadji 2012, 2013, 2023, 2025b). The simultaneous measurement of CH & LC satisfies the First Law of Culture Based Development for the quantification of culture (see Tubadji 2025a), which stipulates that any valid quantification of culture must include at least one proxy for both cultural heritage (CH) and living culture (LC). Next, we calculate the sum of the frequencies of both words used to measure CH and LC and find the corresponding percentage of use of the word ‘Krishna’ (approximating CH) and the percentage of use of the word ‘Influencer’ (approximating LC). Essentially, thus we calculate the probability of the word classified as LC or CH to be used locally. These percentages are used to substitute in the Shannon Entropy index formula which is the mathematical formula that CBD uses to calculate the level of Cultural Entropy (Tubadji 2025b,a). Namely, we calculate Cultural Entropy using the following linguistically adapted formula:

$$CE = H(p_i) = - \sum p_i \log(p_i) \quad (1)$$

where,

i : denotes two classes: Living Culture (LC) and Cultural Heritage (CH);

p_i : represents the probability of the word classified as LC or CH to be used in a region;

$\log(p_i)$: in this case is the logarithm with base 2 (because $i=2$) of the corresponding percentage of CH or LC;

H : stands for Shannon Entropy, measuring the variability of the proxies for LC and CH within the local linguistic corpus and thus approximating the dynamics of local cultural capital.

By aligning in the above way with the CBD theoretical requirements, our dataset offers a rare readily available comprehensive measure – an empirical solution for quantifying the distribution and dynamics of the complex dynamic entity cultural capital

across regions. It allows us to observe how different localities balance heritage-related and modern culture-related linguistic expressions, thus generating a culturally meaningful and methodologically coherent basis for further CBD analysis.

In addition to the linguistic indicators, our dataset incorporates socioeconomic outcome variables. Data on unemployment rates – both overall and disaggregated by caste groups, including *Scheduled Castes* (SC), *Scheduled Tribes* (ST), and *Other Backward Classes* (OBC), are included alongside with explanatory factor: the overall literacy rate (in percentage) at the locality level. All these socioeconomic indicators were obtained from Census of India ([Government of India 2011](#)) and align with our desired Myrdalian relevance for the language-culture-education-channel line of analysis.

Together, these linguistic and socioeconomic measures provide a relevant, multidimensional foundation for analysing the regional distribution of cultural capital within the CBD analytical framework through the proxy: language. It also enables a unique socio-linguistic analysis on regional level that is very rarely implemented in socio-linguistic studies that have not so far used the CBD measurement approach (CH, LC and Cultural Entropy). The main descriptive statistics of all available variables are provided in the Appendix (Table A.1).

4 Descriptive Analysis

4.1 Descriptive Analysis Overview

First, a descriptive analysis of the dataset is presented to provide a clear picture of the quantified cultural entities and the trends embedded within the data. Figure 1 illustrates how *Cultural Heritage* (CH) and *Living Culture* (LC) vary substantially across different geographies, and, importantly, it shows the *proportion* between the two components differs markedly from one locality to another.

The latter observation is further corroborated by Figure 2, which displays the precise balance between CH and LC as calculated using the CBD Cultural Entropy (CE) Index according to equation (1). The CE index reveals significant variation: certain states such as Manipur or Meghalaya exhibit a distinctly lower Cultural Entropy (around 0.6), indicating a more stagnant or one-sidedly-dominant cultural equilibrium, whereas others – most notably Delhi – score considerably higher, reflecting a more dynamically open cultural climate. Interestingly, the majority of localities fall on the higher side of the index, portraying India as an overall vibrant, culturally dynamic landscape, largely open to socio-economic change. Only a small cluster of geographic centres appears to experience comparatively higher cultural stickiness and reduced dynamism (i.e., low CE).

Third, as shown in Figure 3, when analysing *Cultural Entropy* in relation to inequality in experienced unemployment across the regions of India, careful attention is required. The visual evidence suggests that unemployment varies far more substantially across space than Cultural Entropy, even among regions with similarly high CE scores. For instance, regions such as Delhi and Goa, both characterised by high Cultural Entropy and thus a relatively dynamic cultural climate, nevertheless display substantial variation in the levels of unemployment experienced by their populations. Moreover, the figure highlights a pronounced inequality in unemployment across caste groups, with Scheduled Castes (SC), Scheduled Tribes (ST), and Other Backward Classes (OBC) experiencing different levels of joblessness. This inequality is not uniform but varies significantly from region to region, underscoring the importance of contextualising cultural indicators such as CE within the broader socio-economic landscape and potential confounding factors from culture on unemployment.

Therefore, as a fourth step, we examine key confounding factors that help explain regional differences in unemployment – most notably, literacy rates across caste groups, as shown in Figure 4. Cultural Entropy and literacy are confounded, and since they both play an essential role for unemployment, their Myrdalian interplay must be accounted for when interpreting unemployment disparities. As illustrated in Figure 4, the pattern strongly echoes the classic findings of *An American Dilemma* ([Myrdal 1944](#)), where labour-market discrimination and unequal outcomes were shown to be deeply intertwined

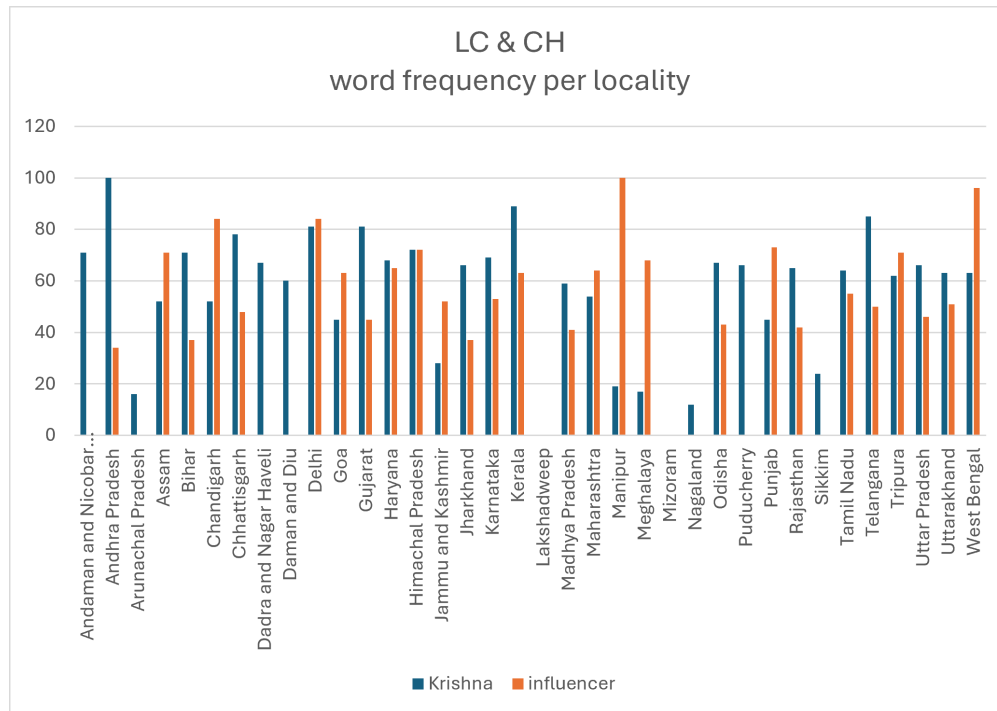


Figure 1: Quantifying Culture according to CBD First Law of Quantification of Culture

with educational inequalities. A similar dynamic is visible in India: inequality in unemployment between caste groups is closely linked to differences in literacy, which themselves are culturally embedded. The relationship however is not entirely intuitively as expected although highly consistent in our explorations – higher unemployment corresponds to higher literacy. Potentially, consistent with Myrdal’s discrimination interpretation, this means that regions that are more equitable to allow literacy for all face higher competition on the labour market and hence more unemployment results as more willing and able to work individuals are available on relatively same as size labour markets in the country.

5 Regression Analysis

Table 1 below presents the results of an OLS regression with robust standard errors. The analysis investigates two key relationships: first, the association between our linguistic measure of Cultural Entropy and regional literacy levels; and second, the relationship between literacy and unemployment across India’s localities. While one might normally expect higher illiteracy to correspond to higher employment, the empirical picture that emerges is considerably more complex. The results confirm that unemployment tends to be higher in regions with higher literacy, implying that literacy is negatively and significantly associated with local unemployment rates. This finding aligns with the impressions gleaned from the descriptive statistics, but it also brings to light a sobering feature of India’s labour-market landscape.

Crucially, although Cultural Entropy is clearly associated with more equitable access to literacy – as predicted by the Culture Based Development (CBD) paradigm and confirmed in Specification 2 of Table 1 – the counter-intuitive link between literacy and unemployment (Specification 1) complicates the expected protective role of Cultural Entropy against labour-market exclusion. In other words, even though regions with higher CE exhibit more balanced cultural environments that foster literacy equity, this does not automatically translate into more equitable or favourable employment outcomes. This intriguing tension calls for deeper investigation, and in the following second step of our empirical analysis we turn to a more granular analysis by examining unemployment patterns separately for each caste category represented in our dataset.

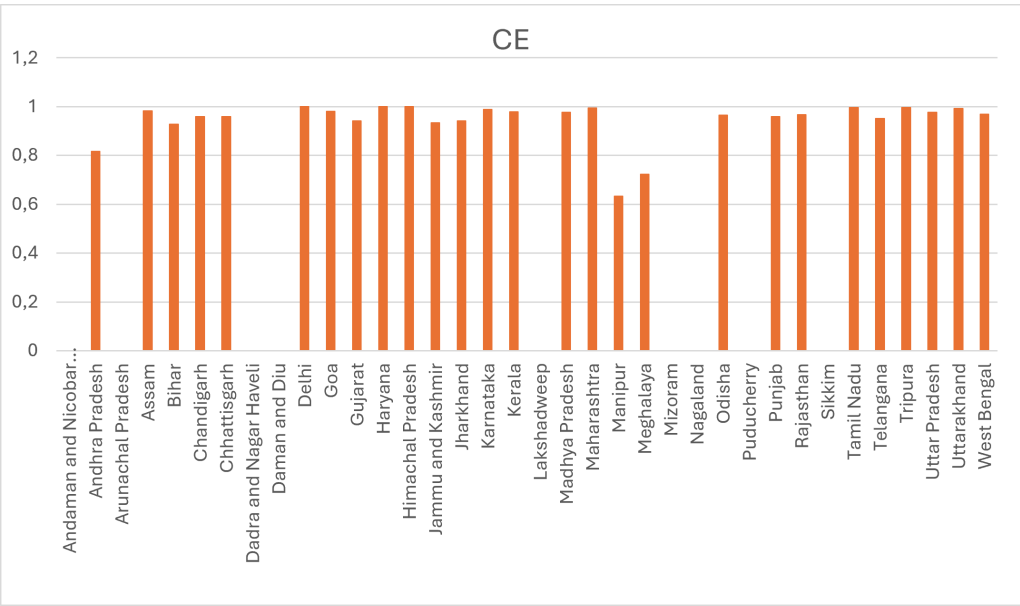


Figure 2: Variation of Cultural Entropy across Space

Table 1: Unemployment, Literacy and Cultural Entropy

Variables	Spec.1 unemployment rate	Spec.2 literacy
illiterate	-0.129* (-0.067)	
Cultural Entropy		30.905** (-14.341)
Constant	6.199*** (-1.376)	53.311*** (-13.435)
Observations	31	26
R-squared	0.178	0.151

Notes: Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

As seen from Table 2, if we look at the association between unemployment and literacy by caste, it becomes apparent that while on average (Specification 1) and for the Scheduled castes (Specification 2) and Other Backward Classes (Specification 4) the negative association between illiteracy and unemployment exists, this is not present as an effect for the Scheduled Tribes. In other words, on average, and for all but the Scheduled Tribes, we see higher literacy creating crowding out effects on the labour market. This is only not true for the Scheduled Tribes. Put differently, the unemployment among the Scheduled Tribes does not associate with their literacy levels in either direction in a statistically significant manner, hence, we can say they are likely to be discriminated not indirectly through access to education but directly on the labour market itself.

Thus, in a last step, as seen in Table 3, it is possible with the available data to importantly look at the relationship between unemployment and Cultural Entropy. Apparently because of the positive association (15% R-squared) between Cultural Entropy and literacy, clearly we see Cultural Entropy positively associated with the unemployment rate, since literacy is positively associated with unemployment. This confirms that indeed the effect from Cultural Entropy on unemployment is present but indirect and it passes through the confounding factor: literacy in a way as in the USA American Dilemma case, but also in a particular manner because in India higher literacy increases the competition on the market and more educated people remain unemployed.

It is very notable that while the literacy component was not present as an impact

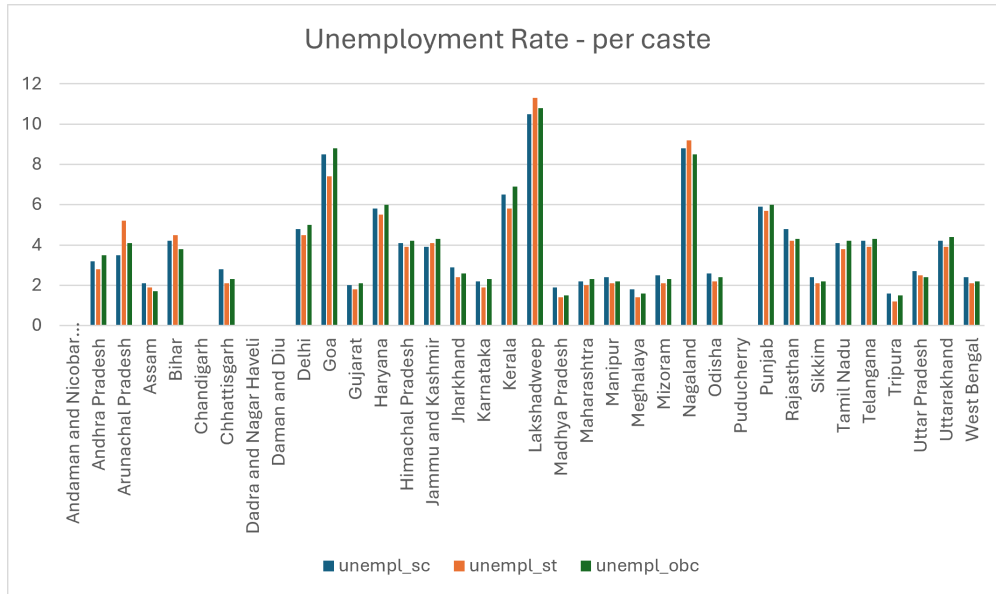


Figure 3: Variation of Unemployment across Space

Table 2: Unemployment by Caste & Illiteracy

Variables	Spec.1 unempl rate	Spec.2 unempl sc	Spec.3 unempl st	Spec.4 unempl obc
illiterate	-0.129* (-0.067)	-0.122** (-0.056)	-0.108 (-0.064)	-0.128** (-0.06)
Constant	6.199*** (-1.376)	5.963*** (-1.194)	5.511*** (-1.313)	6.026*** (-1.255)
Observations	31	31	31	31
R-squared	0.178	0.2	0.137	0.196

Notes: Robust standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

factor for unemployment of the Scheduled Tribes, Cultural Entropy is strongly significant and consistent as impact on the Scheduled Tribes' unemployment as well as on the rest of the population. This confirms that culture, while indirect in impact for some, is a very potent factor that affects all agents in a locality.

It also confirms the association between the Scheduled Tribes and unemployment potentially through the direct cultural channel, yet the confounding factors determine the direction of the effect due to the aggregate effects of culture on unemployment. In other words, cultural direct and indirect effects potentially influence each other in the locality, and this calibrates the final impact from culture for the different castes.

6 Discussion

6.1 Cultural Dynamics and Labour-Market Inequality in India

The results presented across our descriptive analyses and regressions collectively illustrate that culture exerts a measurable and meaningful influence on unemployment outcomes in India, although this influence operates in a nuanced and indirect manner across the different regions of the country. By using a *linguistic approximation* of culture – Cultural Entropy derived from regional Google Search behaviours – we were able to observe cultural variation at a level of geographical granularity that traditional published linguistic corpora simply cannot capture. This granularity reveals that Cultural Entropy is strongly associated with more equitable literacy outcomes, in line with the Culture Based Development paradigm. Yet, paradoxically, because literacy itself is positively correlated

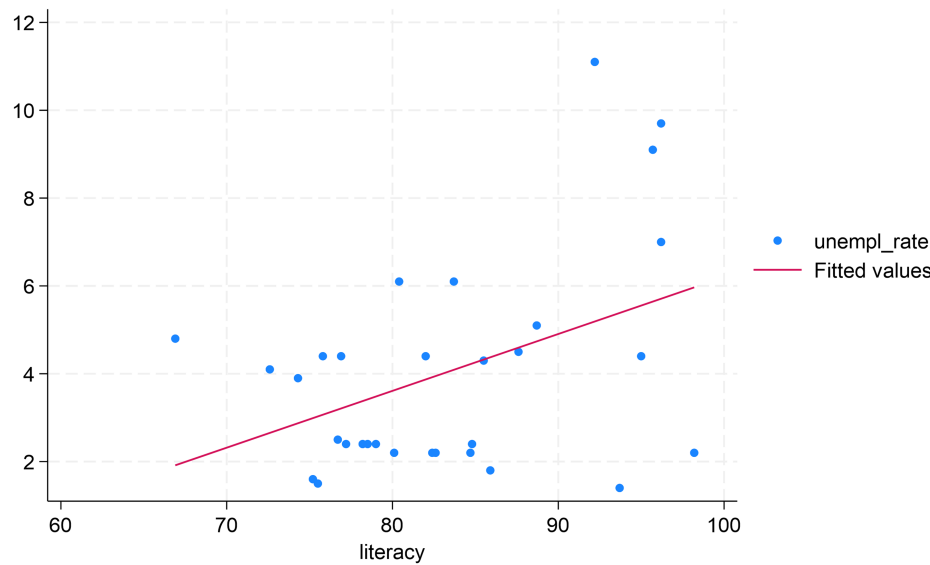


Figure 4: Confounding Factor for the Impact of Cultural Entropy on Unemployment

Table 3: Unemployment by Caste & Cultural Entropy

Variables	Spec.1 unempl rate	Spec.2 unempl sc	Spec.3 unempl st	Spec.4 unempl obc
Cultural Entropy	5.707** (-2.376)	5.148** (-1.944)	4.993** (-1.909)	5.560** (-2.118)
Constant	-1.703 (-2.018)	-1.252 (-1.615)	-1.444 (-1.619)	-1.68 (-1.763)
Observations	26	26	26	26
R-squared	0.064	0.071	0.072	0.069

Notes: Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

with unemployment across Indian regions, the otherwise “protective” cultural effect becomes intertwined with a labour-market reality in which greater access to literacy also increases the number of individuals competing in the formal labour market, thereby increasing the unemployment rates. This mechanism closely mirrors the logic described in Myrdal’s ‘An American Dilemma’, where structural inequalities in one sphere translate into self-fulfilling prophecy outcomes in another. India’s case shows a similar dynamic but through its own culturally specific pathway: higher Cultural Entropy enhances literacy equity, but this in turn intensifies labour-market competition, thus partially offsetting the inclusive effects of cultural balance on unemployment.

6.2 Cultural Channels, Caste Variation, and the Value of Linguistic Measures

The final step of the analysis, as summarised in Table 3, highlights that although literacy moderates much of the relationship between Cultural Entropy and aggregate unemployment, culture still exerts a direct and significant effect on unemployment patterns, particularly for Scheduled Tribes and, to varying degrees, other caste groups as well. Notably, literacy does not explain unemployment among Scheduled Tribes, yet Cultural Entropy remains a robust predictor of their labour-market outcomes, suggesting a strong direct cultural channel alongside the indirect, literacy-based channel observed in the broader population. This duality – indirect effects via literacy and direct effects via cultural structure – demonstrates that culture shapes unemployment both through the distribution of human capital and through deeper socio-cultural dynamics embedded in

the locality. Crucially, these insights are only observable because the linguistic measure of culture captures real-time, behavioural, region-specific cultural variation, something that standard published linguistic corpora cannot accomplish due to their static, aggregated nature. By grounding cultural measurement in contemporary linguistic behaviour, this approach unveils the subtle, layered ways in which cultural environments influence socio-economic inclusion on regional level, revealing India not merely as an economy with labour-market disparities, but as a culturally textured landscape where cultural conditions shape both opportunity structures and the lived realities of unemployment that vary on regional level.

7 Conclusion

The objective of this study was to investigate how cultural dynamics shape unemployment outcomes across India, using a novel linguistic approximation of culture grounded in the Culture Based Development (CBD) paradigm. By operationalising culture through Cultural Entropy, derived from region-specific Google Search behaviour, we sought to capture both *cultural heritage* and *living culture* on regional level in a way that traditional linguistic corpora cannot, thereby allowing us to trace cultural influences on socio-economic conditions with unprecedented linguistic regional precision.

To address this objective, this study constructed a dataset combining (i) linguistic indicators of Cultural Entropy, (ii) regional unemployment rates disaggregated by caste, and (iii) literacy and illiteracy rates as key confounding factors. Using descriptive analysis and OLS regression with robust standard errors, the study examined both the direct and indirect channels through which culture may influence unemployment. Our analysis revealed several noteworthy patterns. First, Cultural Entropy is strongly associated with more equitable literacy outcomes, validating the CBD prediction that more culturally dynamic environments foster broader human-capital inclusion. Second, literacy itself is positively associated with unemployment, suggesting that regions with higher literacy experience greater labour-market competition – a phenomenon that complicates the expected protective role of culture. Third, even after accounting for literacy, Cultural Entropy retains a direct and significant influence on unemployment for certain groups – particularly Scheduled Tribes – indicating that culture shapes economic participation not only through educational pathways but also through deeper socio-cultural structures embedded in localities.

These findings, however, must be interpreted in light of several limitations. The linguistic measure of culture, while innovative, relies on Google Search frequencies, which may be partially affected also by differences in internet access, digital literacy, or on-line behaviour across regions. Likewise, the cross-sectional nature of the data limits causal inference and prevents the capturing of temporal dynamics in cultural change or labour-market adjustment. Finally, some socioeconomic variables, such as underemployment or informal-sector participation, remain unavailable at the necessary resolution and therefore could not be incorporated into the analysis.

In summary, this study demonstrates that language is a potentially very statistically powerful quantification of culture, which is a complex, multi-layered force shaping unemployment outcomes in India, and that culturally grounded inequalities manifest both directly and through critical moderators such as literacy. Most importantly, by deploying a unique dynamic linguistic measure of culture at the regional level, this study reveals patterns that conventional datasets cannot detect. This linguistic CBD approach to measuring culture can open a new empirical window in the economic development study of the cultural foundations of socio-economic inclusion and provides fertile ground for future linguistic-measures-employing research into the cultural determinants of development. We hope that regional researchers will employ this linguistic CBD measurement approach to culture through constructing further datasets and employing relevant regional analysis spatial econometric techniques. Our findings promise that further research with our approach is likely to be very fruitful.

References

- Ambedkar BR (1936) *Annihilation of caste*. Blumoon Books. <https://www.mea.gov.in/-books-writings-of-ambedkar.htm>
- Blommaert J (2010) *The sociolinguistics of globalization*. Cambridge University Press, Cambridge, UK. [CrossRef](#)
- Bourdieu P (1984) *Distinction: A social critique of the judgement of taste*. Harvard University Press, Cambridge, MA. Original work published 1979
- Bourdieu P (1986) The forms of capital. In: Richardson J (ed), *Handbook of theory and research for the sociology of education*. Greenwood Press, Westport, CT, 241–258
- Bourdieu P (1991) *Language and symbolic power*. Harvard University Press, Cambridge, MA
- Crystal D (2000) *Language Death*. Cambridge University Press, Cambridge, UK
- Deshpande A (2011) *The grammar of caste: Economic discrimination in contemporary India*. Oxford University Press. [CrossRef](#)
- Fishman JA (1991) *Reversing language shift: Theoretical and empirical foundations of assistance to threatened languages*. Multilingual Matters, Bristol, UK
- Government of India (2011) *Census of India 2011: Language tables*. Office of the Registrar. General & Census Commissioner, Ministry of Home Affairs
- Grin F (2003) Language planning and economics. *Current Issues in Language Planning* 4: 1–66. [CrossRef](#)
- Gumperz JJ (1982) *Language and Social Identity*. Cambridge University Press, Cambridge, UK
- Kachru BB (1994) English in South Asia. In: Burchfield R (ed), *The Cambridge History of the English Language*. Cambridge University Press, Cambridge, UK, 497–553. [CrossRef](#)
- Mohanty AK (2018) *The Multilingual Reality*. Multilingual Matters. [CrossRef](#)
- Myrdal G (1944) *An American dilemma: The Negro problem and modern democracy*. Harper & Brothers, New York, NY
- Myrdal G (1968) *Asian Drama: An Inquiry into the Poverty of Nations*. Pantheon Books, New York
- Nayyar D (2019) *Asian transformations: An inquiry into the development of nations*. Oxford University Press, Oxford, UK. [CrossRef](#)
- Phillipson R (1992) *Linguistic imperialism*. Oxford University Press, Oxford, UK
- Prabhakar JD (2024) *Socio-economic profiling of Indian languages*. LINCOM GmbH, Munich, Germany
- Sapir E (1921) *Language: An Introduction to the Study of Speech*. Harcourt, Brace & Company, New York
- Sen A (2004) Elements of a theory of human rights. *Philosophy & Public Affairs* 32: 315–356. [CrossRef](#)
- Suresh G (2026) Bank access and cultural entropy across the regions of India. *REGION* 13: 153–170. [CrossRef](#)
- Thorat S, Newman KS (2010) *Blocked by caste: Economic discrimination in modern India*. Oxford University Press, Oxford, UK

- Trigg AB (2001) Veblen, Bourdieu, and conspicuous consumption. *Journal of economic issues* 35: 99–115. [CrossRef](#)
- Tubadji A (2012) Culture-based development: Empirical evidence for germany. *International Journal of Social Economics* 39: 690–703. [CrossRef](#)
- Tubadji A (2013) Culture-based development—culture and institutions: Economic development in the regions of europe. *International Journal of Society Systems Science* 5: 355–391. [CrossRef](#)
- Tubadji A (2023) Culture based development in the regions of china. *Journal of Economic Issues* 57: 8–35. [CrossRef](#)
- Tubadji A (2025a) Cultural entropy, innovation, and growth. *Politics & Policy* 53: 70050. [CrossRef](#)
- Tubadji A (2025b) *Culture based development: Modelling cultural bias in economic choice*. Edward Elgar Publishing. [CrossRef](#)
- Tubadji A, Gheasi M, Nijkamp P (2017) Immigrants’ socio-economic achievements and cultural diversity: Economic effects of individual and local cultural capital. *International Journal of Manpower* 38: 712–728. [CrossRef](#)
- Tubadji A, Webber D (2023) Tea for two: Language and bilateral trade with china. *Journal of Economic Issues* 57: 153–171. [CrossRef](#)
- Uma Maheshwar Rao G (2017) *Language in Chaos: In the Telugu States (in Telugu)*. Telugu Jati (Trust)
- UNESCO (2010) *Atlas of the world’s languages in danger* (3rd ed.). UNESCO Publishing
- Weber M (1905) Die protestantische ethik und der geist des kapitalismus. *Archiv für Sozialwissenschaft und Sozialpolitik*: 20–21

A Appendix: Descriptive Statistics of Main Variables

Table A.1: Descriptive Statistics of Main Variables

variable	N	Mean	SD	Min	Max
Krishna	34	58.7	21.8	12	100
Influencer	27	59.6	17.9	34	100
CH	34	58.7	21.8	12	100
LC	27	59.6	17.9	34	100
p_CH	27	0.51	0.1	0.16	0.75
p_LC	27	0.49	0.1	0.25	0.84
Cultural Entropy (CE)	27	0.94	0.1	0.63	1
unempl_rate	31	4.0	2.5	1.4	11.1
unempl_sc (Scheduled Caste)	31	3.9	2.2	1.6	10.5
unempl_st (Scheduled Tribe)	31	3.7	2.4	1.2	11.3
unempl_obe (Other Backward Castes)	31	3.9	2.3	1.5	10.8
literacy rate	31	83.3	8.1	66.9	98.2



© 2026 by the authors. Licensee: REGION – The Journal of ERSA, European Regional Science Association, Milan, Italy. This article is distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).