

MATE SELECTION IN INDIA THROUGH A HISTORICAL LENS: EVIDENCE FROM TIMES OF INDIA'S MATRIMONIAL COLUMNS

*Aasli Abdi Nur*¹
*Aliakbar Akbaritabar*²
*Koyel Sarkar*³
*Nitin Kumar Bharti*⁴
*Luca Maria Pesando*⁴

Abstract

This project adopts a novel historical and computational perspective to investigate evolving patterns of mate selection and family formation in India using newspaper data from the Times of India's matrimonial columns. These data, spanning over three centuries from 1860 to 2011, provide a valuable source to explore long-term trends in the changing salience of attributes such as potential partner's ethnicity, gender, education, occupational status, etc., as well as gender differences in partner-seeking behavior. We do so by web scraping over 1.2 million classified ads from an archive of newspaper data. We apply text analysis techniques including, but not limited to, topic modeling to the 160,112 matrimonial columns identified within those advertisements. Our preliminary findings uncover important changes in gendered patterns of mate selection, with attributes such as physical appearance and skin complexion becoming more salient over time, particularly for women. All in all, the study contributes fresh insights to the discourse surrounding Indian marriages, shedding light on the evolving significance of core attributes influencing marriageability for both men and women.

Keywords: Mate selection; family; matrimonial ads; text-as-data; India.

Introduction

In India, as in other patriarchal societies, the institution of marriage has traditionally centered around unions arranged by families, wherein decisions regarding partners are heavily influenced, if not entirely determined, by family and kin's preferences (Afridi et al., 2023; Desai & Andrist, 2010). With India's social fabric woven tightly with kinship networks, joint-family systems, intergenerational co-residence, strong familial bonds, and frequent social interactions, these decisions typically take into account various factors such as ethnicity –including caste divisions, religious affiliation, *gotra*, *jaati*, etc.– socioeconomic status, family reputation, honor, education, employment, and physical attributes and appearance (Anukriti & Dasgupta, 2017; Banerjee et al., 2013; Dugar et al., 2012). In the context of Indian families, characterized by a patrilocal and patrilineal structure where women typically relocate to their husband's joint family residence, criteria guiding partner selection have historically weighed more heavily on women than on men. While preference for ethnic homogamy remains widespread, the specifics of desired attributes vary significantly by gender. For women, these attributes often center on domestic management skills and reproductive health, including proficiency in cooking,

¹ University of Oxford, Oxford, United Kingdom.

² Max Planck Institute for Demographic Research, Rostock, Germany.

³ Centre for Demographic Research, Autonomous University of Barcelona, Spain.

⁴ New York University Abu Dhabi, Abu Dhabi, United Arab Emirates.

good health, an education level supportive of family growth (rather than economic independence), and physical appearance. In contrast, men's criteria tend to focus less on physical attributes, instead prioritizing their ability to provide financially by focusing on education, income, and employment prospects.

Avenues for seeking partners have also evolved. In the past, individuals typically looked for partners within their social circles or relied on intermediaries, often local priests, who carefully matched eligible brides and grooms within the community based on their biodata and preferences. As time went by, partner's search transitioned to placing print advertisements on matrimonial columns in local newspapers and, even more recently, to matrimonial websites and dating applications. Today, the online matrimonial industry thrives as a multimillion-dollar enterprise, extending its reach not only within India but also overseas, as Indian expats increasingly explore potential partners in foreign lands. In urban India, various dating applications have gained traction, predominantly among youngsters for casual use, yet they have also proven effective in facilitating marriages (Chakraborty, 2012). While avenues for finding partners have expanded, limited scholarly research exists on how mate search, couple matching, and preferences for specific partners have evolved. It is therefore important to explore these areas to unravel the complexities surrounding traditional components of Indian marriages, i.e., marriages arranged by parents/kin, which still dominate marital arrangement in the country (Sarkar and Rizzi 2024), albeit in a more "hybridized" fashion whereby marital decisions are increasingly jointly negotiated between parents and children (Allendorf and Pandian 2016).

Objectives

This project adopts a novel historical and computational perspective to investigate evolving patterns of mate selection and family formation using newspaper data from the Times of India's matrimonial columns. These sources, first in paper format and now digitized, span over a period of three centuries, i.e., from 1860 to 2011, and provide a particularly valuable source to explore long-term trends in the salience of particular mate-selection attributes such as ethnicity, gender, education, occupational status, etc. We explore shifting trends across multiple dimensions:

1. *Ethnic preferences*, i.e., whether there has been any shift in partner's search preferences by religion, caste, caste-no-bar, *gotra*, *jaati*, etc.
2. *Gender* (men looking for women vs women looking for men), i.e., whether there is any evidence that gender differentials in search attributes have changed and/or narrowed over time.
3. *Ranking of attributes* of potential partners, i.e., which attribute ranks first, second, and third, and whether there has been any shift over time.
4. *"Who" is leading/initiating* the search, i.e., whether there is a shift in matrimonial advertisements placed by parents/families to potential candidates themselves.
5. *Conventionality/originality* of ads, i.e., whether ads have become more/less detailed in terms of requests and which ethnic group advertises the most/least conventional partner preferences.
6. *External shocks*, i.e., how/whether major historical events, such as wars, partitions, famines, etc. shaped partner-seeking behavior.

Data

We use archived data sourced from matrimonial columns published in the Times of India spanning from 1860 to 2011. Our database comprises over 1.2 million classified ads. Access to this extensive dataset is made possible through resources provided by the NYU Library and Clarivate’s ProQuest Text and Data Mining (TDM) Studio, rendering it one of the most distinctive and original sources for investigating shifts in mate selection practices in India.

Methods

To construct our database, we obtain the Extensible Markup Language (XML) documents of 1.2 million “classified ads” from the ProQuest TDM Studio that were published in Times of India. We use BeautifulSoup, Dask, and DuckDB libraries in Python (Rossum and Team 2018) to set up a parallelized data-processing pipeline to extract advertisements’ text from these XML documents. Subsequently, we use NLTK (Bird et al., 2009) and Spacy (Honnibal and Montani 2017) libraries in Python to conduct Natural Language Processing (NLP) on the advertisements’ text. Here, following Akbaritabar and Rubin (2024) and Castro Torres and Akbaritabar (2024), we adopt two strategies: 1) a “deductive” and top-down approach and 2) an “inductive” and bottom-up approach. Using our first deductive approach, we search for exact terms such as *jaati* and caste names, religions, city and state names, Indian languages, and ethnicity names, to name a few. In addition, we curate a list of covariates by manually checking some of the matrimonial columns such as occupation names, education levels, social class, geographical location (e.g., based in USA, based in India, willing to relocate, etc.), looks and appearance, skin tone, nature and behavior, and other terms (such as divorced, widow/er, etc.). We search the advertisements’ text for these lists and tag any occurrences of these terms in the ad text.

Our second, inductive text analysis strategy, uses a bottom-up approach to identify any noun-phrase-clauses used in the advertisements’ text. These noun-phrase-clauses are identified using pre-trained models based on social media, scientific and publicly available corpora of text in Spacy library. Examples of the identified noun-phrase-clauses include: “47 years Mumbai Hindu male”, “Mumbai based bride 34-43”, and “large Mumbai-based family”. Based on these two text analysis strategies, we score advertisements and exclude those classified ads in this newspaper that do not relate to matrimonial columns.

Next, we use Structural Topic Modeling (STM) (Roberts et al., 2014) to identify prevalent topics within the advertisement text. STM is a probabilistic topic modeling method that assumes the presence of topics in a document is influenced by both its content and the metadata, which can be introduced into the model as covariates (Lindstedt 2019). We use the *stm* package in R to read the text data and fit a topic model (Roberts et al., 2019; R Core Team, 2021). Once we fit the model, we read a sample of the top advertisements in each of the emerging topical groups to summarize the main themes and label them accordingly.

Preliminary Findings

We observe a dramatic increase in the number of matrimonial columns published in the second half of our observation period. We attribute this increase in large part to the rising levels of literacy and expanded access to media observed in India. The downturn in more recent years may be due to the growing availability of Internet platforms.

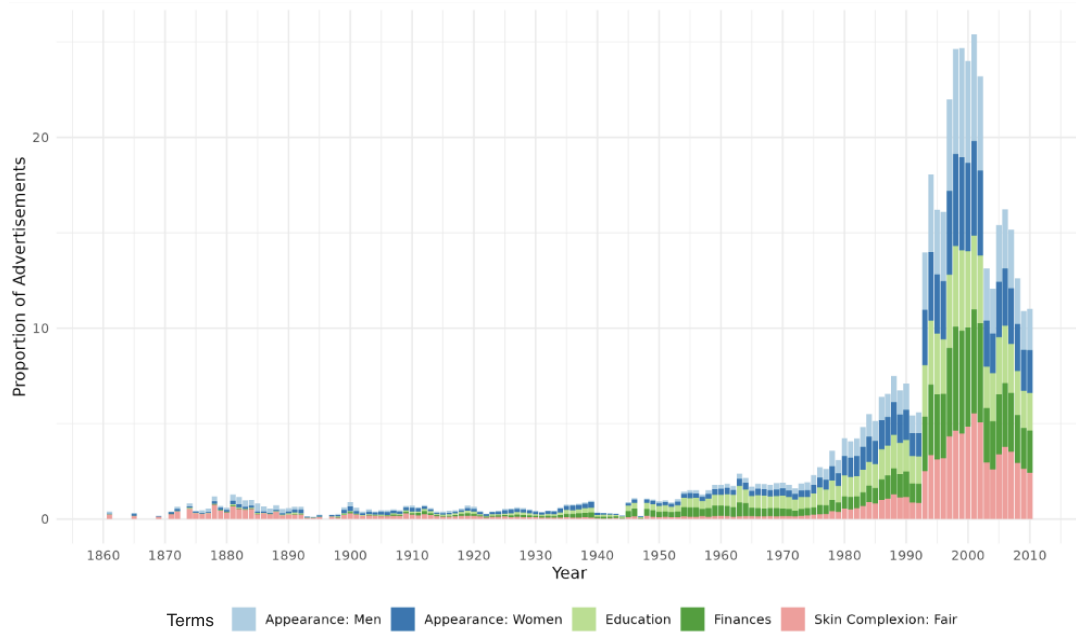


Figure 1. Trends in the use of selected terms over time

Using text analysis, we tokenize the words in matrimonial columns, allowing us to produce counts on the most frequently occurring terms. In Figure 1, we present trends based on selected terms related to physical appearance, education, and socioeconomic status. We find that women’s appearance, particularly their beauty, size, and complexion are frequently discussed and to a greater extent when compared to men’s appearance. Next, we produce a topic model to identify the most salient topics in the advertisements.

Topic	Topic Title	Top Words
Topic 4	Seeking women	girl, well, family, match, fair, alliance, settled, invited, boy, wanted
Topic 8	Women’s appearance	girl, bful, fair, seeking, well, caste, alliance, seeks, beautiful, slim
Topic 10	Educated men	boy, handsome, alliance, educated, family, graduate, well, settled, mba, invited
Topic 3	Established men	boy, settled, qlfd, well, caste, invited, seeking, bar, wanted, working

Table 1. Subset of matrimonial columns topics identified

In Table 1, we present a subset of topics emerging from the structural topic model. The results illustrate gender differentials in search attributes. They also suggest a potential ranking of attributes consistent with our findings in Figure 1, whereby the physical attributes of women may be prioritized compared to the socioeconomic attributes of men.

Discussion and Next Steps

We used a large corpus of matrimonial advertisements spanning over 160 years to provide a temporal perspective on change in mate selection practices in India. So far, our preliminary results show a focus on physical over socioeconomic attributes within matrimonial columns, particularly terms associated with the beauty, size, and complexion of women. In the next phase of our analysis, we will further refine our text analysis to categorize ads based on the person placing them, such as family members, parents, the bride or groom themselves. This will enable us to create a picture of mate-selection preferences and how they have changed over time, and it will also offer a dyadic view to

specific groups, castes, and geographical regions and whether their mate-selection preferences “towards specific other groups” have become more prevalent or strict over time. We will plot the topical prevalence of these results by covariates using linear regression to analyze trends in the expected topical prevalence over time.

References:

- Afridi, F., Arora, A., Dhar, D., & Mahajan, K. (2023). Women’s Work, Social Norms and the Marriage Market (SSRN Scholarly Paper No. 4361917).
- Akbaritabar, A., & Rubin, B. P. (2024). The evolution of plasticity in the neuroscientific literature during the second half of the twentieth century to the presence. *Journal of the History of the Neurosciences*, 0(0), 1–22.
- Allendorf, K., & Pandian, R. K. (2016). The Decline of Arranged Marriage? Marital Change and Continuity in India. *Population and Development Review*, 42(3), 435–464.
- Anukriti, S., & Dasgupta, S. (2017). Marriage markets in developing countries. *The Oxford Handbook of Women and the Economy*, 97-120.
- Banerjee, A., Duflo, E., Ghatak, M., & Lafortune, J. (2013). Marry for What? Caste and Mate Selection in Modern India. *American Economic Journal: Microeconomics*, 5(2), 33–72.
- Bird, S., Klein, E., & Loper, E. (2009). Natural Language Processing with Python. *O’Reilly Media*.
- Castro Torres, A. F., & Akbaritabar, A. (2024). The use of linear models in quantitative research. *Quantitative Science Studies*, 1–32. https://doi.org/10.1162/qss_a_00294
- Chakraborty, K. (2012). Virtual mate-seeking in the urban slums of Kolkata, India. *South Asian Popular Culture*, 10(2), 197–216.
- Desai, S., & Andrist, L. (2010). Gender scripts and age at marriage in India. *Demography*, 47(3): 667-687.
- Dugar, S., Bhattacharya, H., & Reiley, D. (2012). Can’t Buy Me Love? A Field Experiment Exploring The Trade-Off Between Income And Caste-Status In An Indian Matrimonial Market. *Economic Inquiry*, 50(2), 534–550.
- Honnibal, M., & Montani, I. (2017). spaCy 2: Natural language understanding with Bloom embeddings, convolutional neural networks and incremental parsing.
- Lindstedt, N. C. (2019). Structural topic modeling for social scientists: A brief case study with social movement studies literature, 2005–2017. *Social Currents*, 6(4), 307-318.
- R Core Team (2021). R: A language and environment for statistical computing. *R Foundation for Statistical Computing*, Vienna, Austria. URL <https://www.R-project.org/>.
- Roberts, M. E., Stewart, B. M., Tingley, D., Lucas, C., Leder-Luis, J., Gadarian, S. K., Albertson, B., & Rand, D. G. (2014). Structural topic models for open-ended survey responses. *American Journal of Political Science*, 58(4), 1064-1082.
- Roberts, M. E., Stewart, B. M., & Tingley, D. (2019). stm: R Package for Structural Topic Models. *Journal of Statistical Software*, 42.
- Rossum, G. V., & Team, P. D. (2018). The Python Language Reference: Release 3.6.4. 12th Media Services.
- Sarkar, K., & Rizzi, E. L. (2024). Self-Arranged Marriages in India: Change Amidst Sociocultural Underpinnings and Hanging Norms. *Marriage and Family Review*, 60(2), 76–108.