



Methodology

Our Budget Speeches Capture the Evolution of Government Policy

Centre for Social and Economic Progress (CSEP)
CSEP Research Foundation

6, Dr Jose P Rizal Marg, Chanakyapuri, New Delhi 110021, India

Ph: 011 2415 7600

CSEP Research Foundation

Regd. Off: Second floor, 6, Dr Jose P Rizal Marg
Chanakyapuri, New Delhi 110 021 India
T: +91 11 2415 7600 E: info@csep.org W: www.csep.org

Writer(s)

Ayush Khare and Shruti Gupta

CSEP

Know More Details: <https://csep.org/44c2KZ6>

We aim to arrive at keywords unique to budget speeches from each decade and infer the pattern in these keywords over successive decades. The keyword pattern reveals how government policy narrative has changed over the decades.

Methodology –

- We analyse the budget speeches from the last four decades, excluding the FY21 speech due to its close proximity to the COVID-19 pandemic. We combine speeches from FY91 & FY92, FY01 & FY02, FY11 & FY12, and FY22 & FY23. Separately, we also analyse trends in the latest FY24 budget speech. In the article, when we mention a decade's speech, we are referring to the two combined speeches from that decade.
- We start by eliminating some very common words like 'will', 'propose', 'lakhs', 'government', 'project', etc from the combined speeches.
- Using Python's Natural Language Toolkit, we generated a list of all the words in the text along with their frequencies.
- We list out all words in the combined speeches in descending order and keep the first 300 words from each decade while discarding the rest. There is a slight difference in speech lengths (~10%), which did not significantly alter our findings.
- The top 300-word list is further divided into 3 sections based on frequency - top 60 are high frequency, 61 to 140 are medium frequency, and 141 to 300 are low frequency. This division was made by comparing the frequencies across all decades and determining where the frequencies were similar.
- Next, we take the low and medium frequency words from each decade to identify unique words that occur in only one decade and not in the other three. For high frequency words, we identify words that appeared in at most 2 decades. Our goal was to eliminate common words that did not provide any significant information about that decade. We ended up with a list of high, medium, and low frequency relevant words for each decade.
- To spot common themes across all decades, we pick words that have appeared in at least three or all four decades in each frequency category. We use these common words to create the categories.
- We created eight categories or 'buckets' to classify themes that are discussed in the budget speeches: infrastructure, innovation, social security, human development, food & agriculture, demographics, national income, and fiscal & financial policy.
- Using this approach, we determined the keywords for each decade with their relative importance based on frequency. Our findings are based on keywords which showed a clear pattern across decades, such as those that went from high to low frequency across decades. We have disregarded any findings that were inconsistent.

Budget speeches in India are important markers of the central government's policy evolution. The annual budget speech outlines the government's fiscal policies, revenue and expenditure plans, and economic priorities for the upcoming year. These speeches provide valuable insights into the government's approach to addressing key economic and social issues, and how it plans to allocate resources to achieve its goals.

Over time, the central government's policies and priorities have evolved in response to changing economic and social conditions. By examining the themes and priorities outlined in the past budget speeches since 1990, this article attempts to track the narrative change in government policy over the last few decades.

We identify themes that remain common and trace the shift in focus areas. In sum, we find that the emphasis of government policy has moved towards building more infrastructure, embracing digital innovation and rapid urbanisation and providing better education and healthcare to its vast population.

Graphic

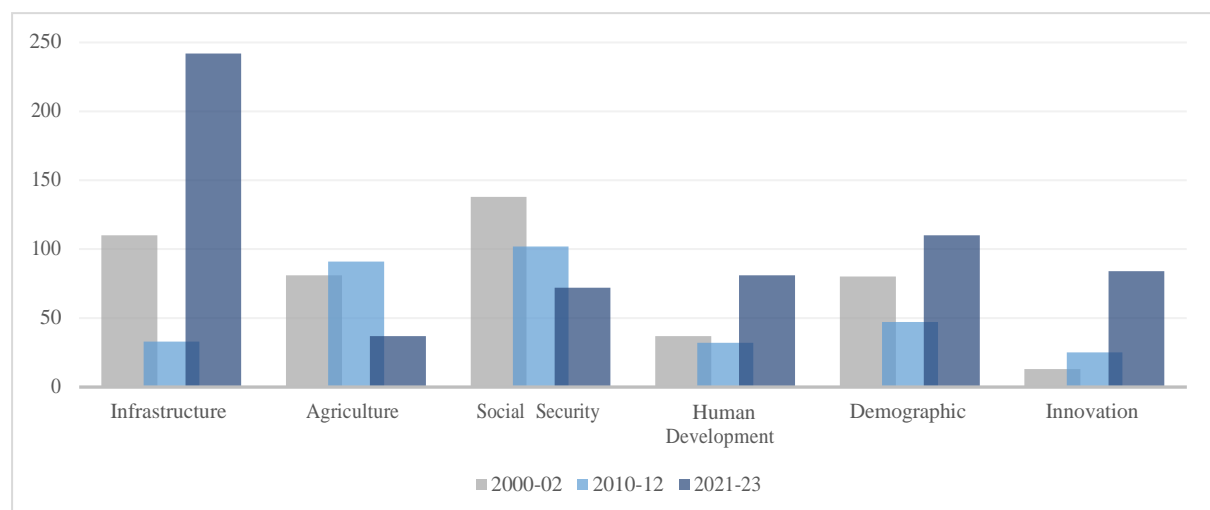


Figure: Relative importance of themes across budgets in early 2000s, 2010s and 2020s based on the usage of unique keywords in budget speeches. Y axis indicates the number of times keywords in each theme are mentioned in a budget.