Proceedings of the First Workshop on Natural Language Processing for Political Sciences (PoliticalNLP), pages 72–78
Marseille, 24 June 2022

What Does the Indian Parliament Discuss? An Exploratory Analysis of the Question Hour in the Lok Sabha

Suman Adhya, Debarshi Kumar Sanyal
Indian Association for the Cultivation of Science
Jadavpur, Kolkata – 700032, India
adhyasuman30@gmail.com, debarshisanyal@gmail.com

Abstract

The TCPD-IPD dataset is a collection of questions and answers discussed in the Lower House of the Parliament of India during the Question Hour between 1999 and 2019. Although it is difficult to analyze such a huge collection manually, modern text analysis tools can provide a powerful means to navigate it. In this paper, we perform an exploratory analysis of the dataset. In particular, we present insightful corpus-level statistics and a detailed analysis of three subsets of the dataset. In the latter analysis, the focus is on understanding the temporal evolution of topics using a dynamic topic model. We observe that the parliamentary conversation indeed mirrors the political and socio-economic tensions of each period.

Keywords: Parliament of India, dynamic topic model, latent Dirichlet allocation, TCPD-IPD, political data

1. Introduction

The Parliament of India is the highest legislative body of India. The members of its Lower House or the Lok Sabha are directly elected by the people while its Upper House comprises representatives elected by the members of all State Legislative Assemblies. Although parliamentary proceedings are immensely useful to a political scientist, they are too large to be manually analyzed. This motivates the use of algorithmic tools to explore them. In this paper, we analyze the TCPD-IPD dataset (Trivedi Centre for Political Data, 2019) of around 298K pairs of questions and answers (QA) in English discussed in the Lok Sabha during the Question Hour – the first hour of every business day of the Parliament – from 1999 to 2019 spanning four Lok Sabha terms (13th term: 1999-2004, 14th: 2004-09, 15th: 2009-14, 16th: 2014-19). During the Question Hour, any Member of Parliament in the Lok Sabha (abbreviated: MP) may ask any question to the ministers related to the administrative activity of the government, and thus, hold it accountable for its actions (Sanyal, 2016; Tripathi and Kumar, 2021). Question time is also an integral part in many other parliamentary democracies like those of Canada, Australia and UK (Martin and Rozenberg, 2014).

Technical specification of the TCPD-IPD dataset appears in (Bhogale, 2019). But the dataset has not been explored, except in (Sen et al., 2019) where the authors aim to identify, for a few chosen themes, whether the questions asked by MPs echo the trend in mass media and social media. Topic modeling has been used to analyze the parliamentary proceedings of various countries, see, e.g., (Greene and Cross, 2017; Gkoumas et al., 2018; Ishima, 2020). In this paper, we study TCPD-IPD using the following pipeline. First, a static topic model of the entire dataset is built and the top topics identified. Then a subset of the dataset is selected for further analysis as follows: (a) A dynamic topic model is built on it; (b) The temporal evolution of topics and words in a topic are plotted; (c) The top-ranking documents at a given time in each of these plots are analyzed. We obtain interesting insights from this analysis. This demonstrates the effectiveness of our technique. Our specific contributions are:

1. We describe important high-level statistical features of the dataset and bias in the participation of MPs (Sec. 2.).

2. We identify the top topics in the entire dataset (Sec. 3.).

3. We make a more nuanced study of the QA pertaining to three specific ministries – Finance, Railways, and Health and Family Welfare — by building a dynamic topic model in each case (Sec. 4.).

4. For each ministry mentioned above, we identify words that showed significant variation in their probability in a topic over time (Sec. 4.) and the major events to which they relate. Thus, word choreography in a dynamic topic model is used to reconstruct events in political history. We hope the insights and lessons from the past will help inform future responses to critical national issues.

2. High-Level Features of TCPD-IPD

We enumerate below some interesting insights we obtained from the dataset.

2.1. Ministry-wise data distribution

TCPD-IPD contains questions related to 85 different ministries. Fig. 1 shows the data distribution for the top ten ministries (comprising almost 50% of the full dataset). Clearly, *Finance*, *Railways* and *Health and Family Welfare* are the top three ministries.

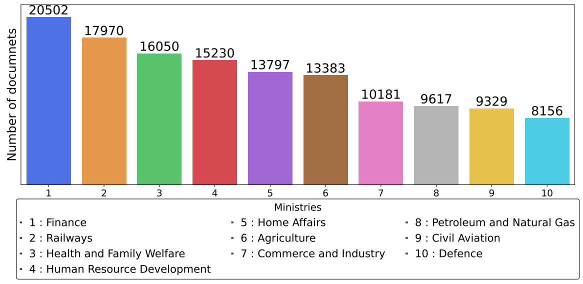


Figure 1: Ministry-wise data distribution.

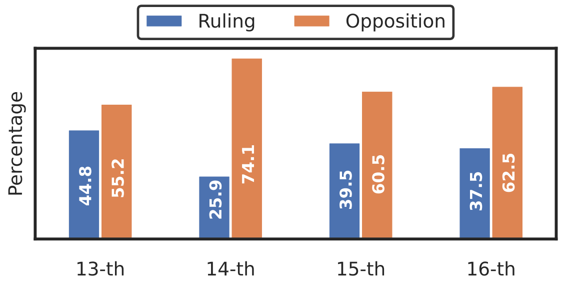


Figure 2: Participation of ruling alliance vis-a-vis that of the opposition.

2.2. Term-wise data distribution

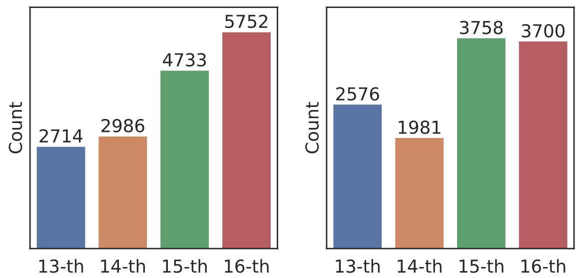
Among a total of 298,292 questions, the number of questions asked in each term of the Lok Sabha is as follows: 13th: 73,531; 14th: 66,371; 15th: 79,401; 16th: 78,989. We found out that over the span covered in this dataset, Lok Sabha always met for more than 50 sittings in a year, except in 2004 (48 sittings) and 2008 (46 sittings). This correlates with the lowest number of questions asked in 2004 (9398 questions) and 2008 (9851 questions). Two of the three sessions in 2004 and all sessions in 2008 are included in the 14th Lok Sabha. Although the fewer sittings are highlighted in many news reports (Anuja, 25 Nov 2020), we did not find mention of its impact on the Question Hour, that we clearly identified above.

2.3. Participation of MPs

Unlike the previous Lok Sabha terms, in the 16th the ruling alliance asked more questions than the opposition. In that term, they had a historic 65% share of the House. Since the numeric strength of the ruling alliance is, by rule, higher than that of the opposition, we normalize them to get an idea of participation from each side *had the number of representatives from either side been equal*. We believe this will afford a fairer comparison between their participation. Let R_n and O_n be the number of members of the ruling alliance and the opposition, respectively, and R_q and O_q be the number of number of questions asked by the ruling alliance and the opposition, respectively. Then $R_{pp} = \left(R_q \times \frac{O_n}{R_n+O_n}\right), O_{pp} = \left(O_q \times \frac{R_n}{R_n+O_n}\right)$. As seen in Figure 2, in the transformed space, the opposition still asks more questions than the ruling alliance, matching the expectations from a healthy democracy.

2.4. Gender and caste bias in participation

Over the four Lok Sabha terms covered by the dataset, 91.6% questions were asked by men while 8.4% questions were raised by women. The average gender ratio of men to women was $8.3{:}1$ over the same four terms. Thus the the skewed gender ratio correlates with the distribution of questions. It is noteworthy here that the Women's Reservation Bill proposing the reservation of one-third of the seats in the Lok Sabha for women has been pending since 2010 (Marwah, 2019), thus, allowing the bias to continue. As regards the caste distribution, 80.6% questions were raised by MPs from the general caste while the rest come from the reserved categories. Note that 24.03% of Lok Sabha seats are reserved for the reserved categories while the rest belong to the general caste. Figure 3 shows the number of questions on gender and caste-related issues asked in the Parliament; Appendix 6.5. lists the keywords we used for this analysis.



(a) *Gender*-related. (b) *Caste*-related.

Figure 3: *Gender* and *caste* related discussions in each Lok Sabha term.

3. Topic Model for TCPD-IPD

We used topic modeling to get a thematic view of the TCPD-IPD dataset. Researchers have observed that Latent Dirichlet Allocation (LDA) ((Blei et al., 2003)), which employs Gibbs sampling for inference, often produces superior topics than those from modern variational inference-based topic models; see, e.g., (Blei et al., 2017; Lisena et al., 2020). This motivated us to use LDA instead of neural topic models. We pre-processed the entire TCPD-IPD corpus, used LDA to extract 50 topics, and manually labeled them (See Appendix 6.1.-6.3.). We filtered out a few noisy, heterogeneous topics and among the rest plotted the top ten topics in Fig. 4. Clearly, there is a huge emphasis on growth in economy and science, at the state and national levels.

4. Ministry-wise Analysis

We have selected three ministries – Finance, Railways, and Health and Family Welfare – for further analysis.

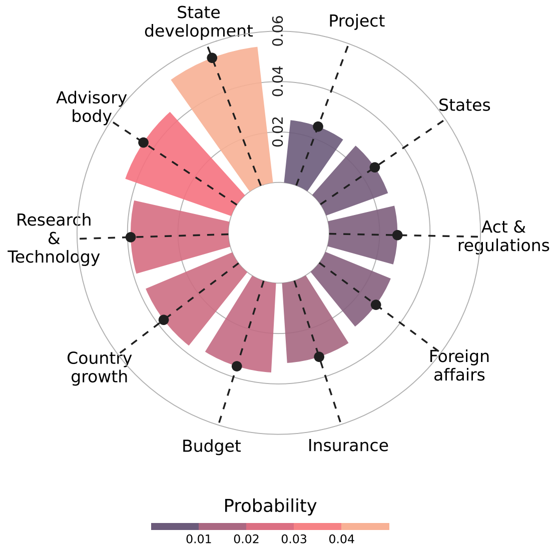


Figure 4: Hot topics discussed in the Indian parliamentary QA sessions.

We set the topic count to 20 for all the three ministries. To model the temporal variation in topics, we had to make a choice between the classical Dynamic Topic Model ((Blei and Lafferty, 2006)) which is essentially an adaptation of LDA for sequential data (hereafter, called LDAseq) and the more modern Dynamic Embedded Topic Model (D-ETM) ((Dieng et al., 2019)), both developed by David Blei and collaborators. In our experiments, we found LDAseq produced better topics (see Appendix 6.4.). Similar observation is reported made in (Dieng et al., 2019).

Using LDAseq, we obtained the temporal evolution of the probabilities of five topics for each ministry, as shown in Fig. 5. In Finance, the peaks in 'agricultural loan' in 2008 and 2014 relate to the debt waivers announced at that time, though farmers' loans remain a perennially important topic. The focus on 'rural development' slowly reduces while other topics like 'banking', 'economic growth' (more questions asked when GDP change is unexpectedly high or low) and 'pension schemes' remain more stable. In Railways, the early 2000's witnessed many new government projects and associated parliamentary QA, but the focus gradually shifted to 'infrastructure development' and 'passenger amenity', where questions veered around the increasing private participation. The announcement of many projects on rail safety in the last Lok Sabha term is indicated by increased presence of the topic 'railway safety'. The steep price hike in passenger and freight fare in 2014 sparked intense deliberation in the Parliament. In the Ministry of Health and Family Welfare, discussions on women and child care and rural medical infrastructure increased after the National Health Mission was launched in 2005. The focus on medical research in the early part of the decade led to the establishment of many premier medical institutions throughout the country but gradually the interest waned.

The *rare* words in a topic were often more informative and captured specific events or issues. So in the following sub-sections, we choose a few representative topics from each ministry and show the temporal evolution of the probability of the rare words in the selected topics. In each plot, we also annotate one of the dominant words with example questions asked during the Question Hour.

4.1. Finance

We have selected two topics, *'banking'* and *'pension reforms'*, and plotted the probability of a few selected tokens in them as a function of time in Figure 6. In the topic 'banking' shown in Figure 6a, we find that the word 'credit_card' peaks in 2007. Our analysis shows that most of the questions around this time are related to the growing credit card frauds in India and the sudden rise in credit card interest rates by some banks, coinciding with reports in mass media. Another aspect in credit card-related discussion is *Kisan Credit Card* (KCC) – a low-interest credit card for farmers, which was introduced in 1998 and significantly improved in 2004. The steady rise in discussion on debit cards correlates with the increasing adoption of debit cards (that avoided the debt trap of credit cards) in India. With demonetization and increased government emphasis on end-to-end digital – as opposed to cash – transactions, terms like 'atm', 'digital_transaction', and 'cyber_security' gain prominence while the popularity of more traditional mediums like 'cheque' reduces.

Figure 6b shows the topic on pension reforms. In late 2003, Government of India notified that it was abolishing the then existing government-funded pension system for all its new employees and that they would come under the National Pension System (NPS) to be administered through the new Interim Pension Fund Regulatory and Development Authority (PFRDA). NPS enabled subscribers to make planned savings for post-retirement income. NPS was extended to all Indian citizens in 2009. Being a monumental change, NPS provoked a number of questions that peaked around 2011; MPs wanted to know the details of the scheme, including its performance, implementation challenges, extension to unorganized sectors, and even its security. The government introduced the Voluntary Retirement Scheme (VRS) in nationalized banks in early 2000's to reduce the financial load on the public exchequer. Thousands of employees across various organizations accepted VRS in 2000-2001. Given the high unemployment rate in the country, VRS generated a lot of panic among people and pointed questions in the Parliament on the the government's future plans about its workforce. The other visible terms 'apy' 'dbt' and 'jan_dhan' refer to recent financial inclusion programs.

4.2. Railways

Here, we highlight only one topic, namely, 'infrastructure development' which is displayed in Fig. 7. Ob-

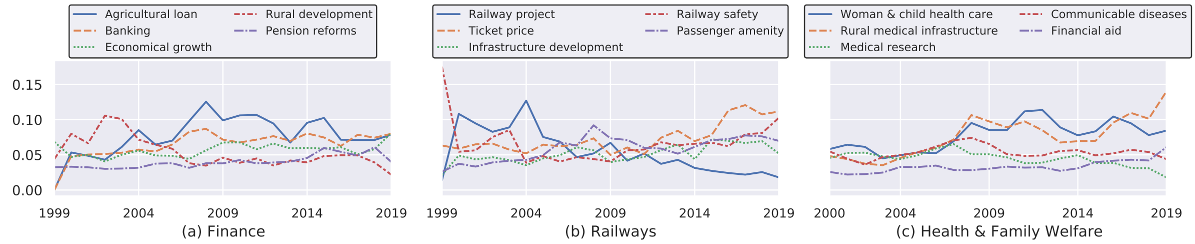
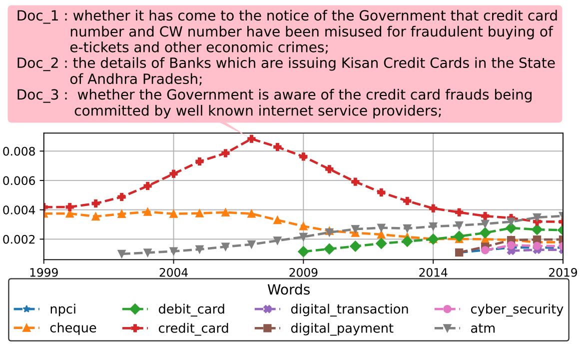
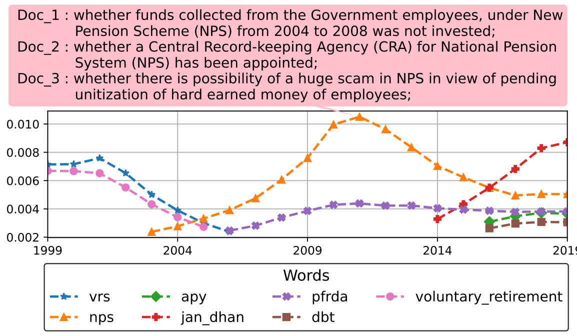


Figure 5: Temporal evolution of topics in 3 subsets of TCPD-IPD using LDAseq.



(a) Topic: *Banking*



(b) Topic: *Pension Reforms*

Figure 6: Selected topics obtained by running LDAseq on the Finance subset of TCPD-IPD.

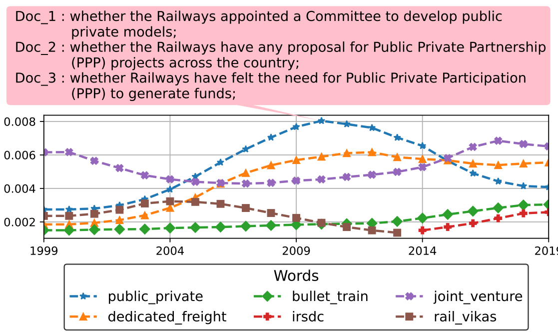


Figure 7: Topic *Infrastructure Development* in Railways.

serve the peak of the term 'public_private' around 2010 when the Railways ministry introduced a new model for public-private partnership to modernize the Indian Railways. Indeed, there has always been questions and panic in the Lok Sabha on the public-private models as privatization of the economy could increase fares, job loss, and casualization of labor (Makhija, 2006; Reddy, 2019). A related term 'joint_venture' was a part of many discussions. While early uses of it (in 2000's) focused on joint ventures of Indian Railways with other public sector companies, the recent focus (since 2015) has been on the increasing role of private players. Indeed similar exchanges occurred between the MPs and the Civil Aviation ministry on the privatization of airlines. Terms like 'rail_vikas' and 'irsdc' refer to companies owned by Indian Railways and entrusted with maintenance of Railways. The rise in freight volumes led to the ideation of Dedicated Freight Corridors (DFC) in 2005 and generated a number of questions ('dedicated_freight') related to their cost, progress, and expansion. Discussions on the introduction of bullet trains have been present for a long time but they gathered momentum when a vision document was tabled by the government in December, 2009 and the construction of the Mumbai-Ahmedabad high-speed rail corridor started in 2017.

4.3. Health and Family Welfare

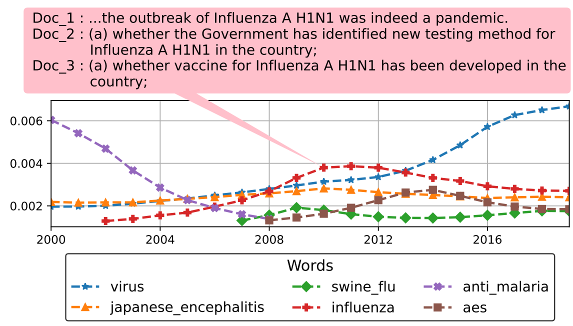


Figure 8: Topic *Communicable Diseases* in Health and Family Welfare.

Here, we have chosen the topic 'Communicable Diseases'. Fig. 8 shows the evolution of word probabilities in this topic. Clearly, we find an increasing focus on the word 'virus' because India has been repeatedly hit by the Swine influenza virus (such as H1N1), including in 2009 when swine-flu turned into a pandemic and in 2014-15 (Kshatriya et al., 2018). MPs in the Lok Sabha enquired about the number of cases, test adminstration, government interventions, vaccination drives, the role of WHO, and the effort to develop indigenous vaccines. Another major disease in India has been malaria but a response in Lok Sabha informs us that malarial death reduced steadily over the years, and that is attested to by the steady decline on its focus in the Parliament. India recorded thousands of deaths due to Acute Encephalitis Syndrome (AES) in 2008-14 (Ghosh and Basu, 2016) and Japanese Encephalitis in 2005-11 (Adhya et al., 2013). During these unfortunate occurrences, the Lok Sabha witnessed intense discussions on the diseases.

5. Conclusion and Future Work

We identified the salient features of TCPD-IPD and then illustrated the temporal evolution of topics in three important subsets of the data. In future, we will attempt to automatically detect topical change points, annotate them with trigger events (e.g., VRS announcement), auto-summarize the top documents containing a specific topic or word at a given time, and motivate investigative reporting or research on the impact of the most sensitive topics discussed in the Parliament (e.g., the effect of VRS on mid-age employees). We hope our study will help construct more probing parliamentary questions and formulate better national policies.

6. Appendix

6.1. Data preprocessing for topic modeling

We have removed the punctuation from the dataset, then lowercased and lemmatized the words. We have also removed the stopwords, and filtered out the remaining words having document frequency lower than 0.001 and higher than 0.95. We have only kept the words that have at least 3 characters. Finally, we removed the documents with less than 3 words in them. After preprocessing, we created bigrams to better capture word co-occurrence statistics.

6.2. Configuration for topic models

We used the following hyperparameters to run LDA, LDAseq and D-ETM.

1. **LDA** (Blei et al., 2003): We use Gensim's implementation of LDA model[1]. To enable reproducibility, we use a fixed random seed, i.e., set the `random_state` = 2021. We set `passes` to 20 and use the default values for the remaining hyperparameters.

2. **LDAseq** (Blei and Lafferty, 2006): We use the Gensim implementation[2]. We set the `random_state` value as 2021 and `passes` as 20. For the rest of the hyperparameters, we use the defaults.

3. **D-ETM** (Dieng et al., 2019): We use the original implementation[3] with `batch_size` = 64 and `epochs` = 100, and keep the default values for the rest of the hyperparameters.

6.3. Topics in TCPD-IPD

We used LDA to extract 50 topics from the entire TCPD-IPD dataset as it achieved the highest coherence score (see Fig 9) when topic count was varied from 25 to 200 in steps of 25. The topics with their manual

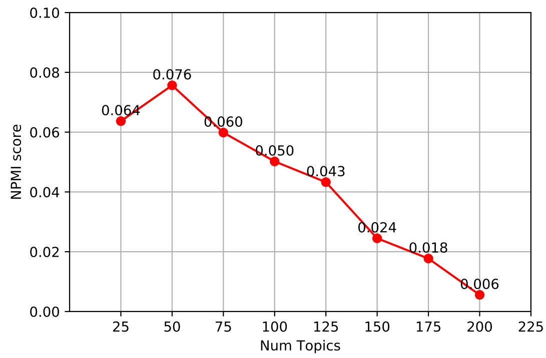


Figure 9: NPMI values for different number of topics in LDA over the full TCPD-IPD dataset.

labels are shown in Table 1. Note that we could not run LDAseq because it was taking too much time. It did not terminate even after running for 72 hours. We could not run D-ETM as it experienced posterior collapse, which could not be resolved even after changing the hyperparameters.

6.4. Topics in subsets of TCPD-IPD

We have carried out a pilot study on dynamic topic modeling with data subsets for the three ministries (Finance, Railways, and Health and Family Welfare). We set the topic count to 20. We have divided each data subset into year-wise slices, partitioned each slice into train:validation:test as 8:1:1, and run LDAseq and D-ETM on them. Then we calculated the coherence (NPMI), diversity, and topic quality ($= NPMI \times Diversity$) for each model by averaging them over the time slices. Table 2 shows that LDAseq performs best and hence we choose it to extract topics from the ministry-specific datasets. The topics that appear in Fig. 5 in the main text have been manually labeled by looking at their top five words. Table 3 shows these topic compositions.

6.5. Keyword extraction for gender and caste-related discussions

To generate keywords for the analysis of gender and caste-related discussions, we have applied SkipGram,

[1] https://radimrehurek.com/gensim/models/ldamodel.html

[2] https://radimrehurek.com/gensim/models/ldaseqmodel.html

[3] https://github.com/adjidieng/DETM

Top-5 words in topic	Manual label
country, foreign, agreement, international, sign	*Foreign affairs*
standard, quality, safety, use, pollution	*Pollution*
urban, city, housing, delhi, construction	*Urban development*
security, state, police, home, affair	*Law & order*
road, highway, national, construction, state	*Road construction*
bank, loan, rbi, credit, account	*Banking*
committee, state, review, recommendation, report	*Advisory body*
state, andhra_pradesh, uttar_pradesh, maharashtra, tamil_nadu	*States*
fund, crore, release, year, state	*Budget*
gas, oil, milk, production, natural_gas	*Petroleum*
air, airport, defence, civil_aviation, airline	*Civil aviation*
post, employee, central, office, officer	*Central Employee*
export, import, product, trade, textile	*Market overview*
project, complete, work, cost, sanction	*Project*
case, court, high_court, person, disability	*Judicial system*
solar, renewable_energy, energy, system, power	*Renewable energy*
act, state, provision, section, rule	*Act & regulations*
service, information, provide, telecom, network	*Information technology*
payment, tax, pay, revenue, amount	*Tax*
power, plant, capacity, supply, state	*Electricity plant*
sport, fertilizer, ltd, limit, corporation	*Fertilizer plant*
madam, action, complaint, case, report	*Grievance*
water, river, state, resource, drinking_water	*Water resources*
education, school, university, student, human_resource	*HR in Education*
scheme, state, development, provide, implement	*State development*
coal, mine, production, mineral, mining	*Coal mining*
rural, district, area, village, functional	*Rural issues*
delhi, mumbai, city, gujarat, chennai	*Metropolitan areas*
health, state, family_welfare, drug, medical	*Health services*
answer, lok_sabha, reply, statement, lay	*Parliament*
farmer, agriculture, crop, agricultural, production	*Agriculture*
china, bangladesh, island, nepal, disaster	*Natural disaster*
increase, year, rate, country, reduce	*Country growth*
steel, port, connectivity, capacity, major	*Ports shipping*
tribal, schedule, minority, scholarship, tribe	*Tribal affairs*
food, price, consumer, state, foodgrain	*Food price*
sector, private, policy, public, investment	*Private sector*
woman, child, employment, worker, labour	*Woman & child labour*
ngo, bihar, society, organisation, organization	*NGO*
tourism, culture, site, tourist, development	*Tourism*
industry, development, infrastructure, unit, scheme	*Industry*
company, issue, guideline, application, insurance	*Insurance*
research, technology, centre, training, national	*Research & Technology*
railway, train, station, passenger, rail	*Railway*
land, forest, area, environment_forest, state	*Wildlife conservation*
state, proposal, set, chhattisgarh, propose	Miscellaneous
vehicle, procurement, website, award, contract	Miscellaneous
year, wise, state, last_three, number	Miscellaneous
due, pleased, loss, reason, affect	Miscellaneous
thereto, reaction, chaudhary, manoj, true	Miscellaneous

Table 1: Topics in the entire TCPD-IPD dataset.

Dataset	Coherence		Diversity		Topic Quality	
	LDAseq	D-ETM	LDAseq	D-ETM	LDAseq	D-ETM
Finance	0.088	0.078	0.652	0.496	0.057	0.039
Railways	0.129	0.094	0.686	0.558	0.088	0.052
Health	0.103	0.096	0.617	0.571	0.064	0.055

Table 2: Topic quality analysis.

which is one of the models used in the neural network-based word2vec algorithm to generate word embeddings (Mikolov et al., 2013). We have run SkipGram on the entire TCPD-IPD dataset and taken the top twenty neighbors (based on cosine similarity of the generated word vectors) of each of the keywords 'gender' and 'caste'. The words are shown in Table 4. Then, we have counted the documents that contain those words. We have ignored the words shown in italics in the table as they introduced many irrelevant documents into the count.

6.6. Explanation of certain terms

Words in the main text that are difficult to understand outside the Indian context are explained below.

1. 'npci': National Payments Corporation of India, created by the Reserve Bank of India under the

Year	Top-5 words in topic	Manual label
	Ministry of Finance	
1999	rate, growth, cent, increase, year	*Economical growth*
2009	rate, cent, growth, increase, year	
2019	growth, cent, rate, economy, sector	
1999	bank, rbi, reserve, issue, guideline	*Banking*
2009	bank, rbi, issue, guideline, reserve	
2019	bank, rbi, fraud, issue, transaction	
1999	project, state, world, development, bank	*Rural development*
2009	project, state, development, infrastructure, rural	
2019	project, development, state, infrastructure, fund	
1999	bank, loan, credit, nabard, state	*Agricultural loan*
2009	bank, loan, credit, farmer, year	
2019	loan, bank, farmer, credit, scheme	
1999	scheme, employee, pension, interest, deposit	*Pension reforms*
2009	scheme, pension, fund, deposit, interest	
2019	scheme, account, pension, state, pradhan_mantri	
	Ministry of Railways	
1999	work, complete, progress, project, line	*Railway project*
2009	work, complete, section, line, gauge_conversion	
2019	work, section, complete, line, gauge_conversion	
1999	freight, traffic, passenger, good, ticket	*Ticket price*
2009	ticket, passenger, freight, increase, scheme	
2019	passenger, ticket, fare, freight, train	
1999	system, safety, track, committee, report	*Railway safety*
2009	system, track, safety, committee, signal	
2019	system, track, safety, train, committee	
1999	project, corporation, rail, development, company	*Infrastructure development*
2009	project, development, corridor, rail, identify	
2019	development, project, corridor, rail, high_speed	
1999	station, facility, provide, platform, provision	*Passenger amenity*
2009	station, facility, provide, platform, work	
2019	station, provide, facility, platform, scheme	
	Ministry of Health and Family Welfare	
2000	research, institute, study, council, develop	*Medical research*
2009	research, study, clinical_trial, institute, council	
2019	research, medical, study, clinical_trial, council	
2000	child, population, programme, national, reproductive	*Woman & child healthcare*
2009	child, programme, national, care, woman	
2019	child, care, woman, national, provide	
2000	disease, malaria, control, case, death	*Communicable diseases*
2009	disease, control, case, report, malaria	
2019	patient, treatment, provide, free, scheme	
2000	centre, care, area, service, rural	*Rural Medical Infrastructure*
2009	patient, treatment, provide, free, hospital	
2019	national, public, healthcare, provide, include	
2000	patient, treatment, provide, hospital, free	*Financial aid*
2009	project, crore, fund, expenditure, cost	
2019	project, fund, crore, cost, completion	

Table 3: Temporal topics for each ministry in TCPD-IPD dataset.

Word	Top 20 neighbors
gender	**gender**, **gender_equality**, **gender_disparity**, **women**, **woman**, **gender_sensitivity**, **gender_sensitization**, **gender_gap**, **gender_parity**, **girl**, **gender_sensitive**, **female_literacy**, **sex**, **male_female**, **disparity**, **child_sex**, **child**, **girl_child**, **literacy_rate**, **sex_selective**
caste	**caste**, **tribe**, **scs**, **obcs**, **obc**, *schedule*, **caste_tribe**, **scheduled_tribe**, **dalit**, **social_justice**, **scs_sts**, **scheduled_caste**, *vijay_sampla*, *empowerment_napoleon*, *belong*, *subbulakshmi_jagadeesan*, **atrocity**, **minority**, *pal_gurjar*, *empowerment_smt*

Table 4: Top 20 neighbors (using `Word2Vec`) for each keyword.

Ministry of Finance, to enable digital payments and settlement systems in India.

2. 'nabard': National Bank for Agriculture and Rural Development, operating under the Ministry of Finance. It regulates the institutions that supply financial help to the rural society.
3. Kisan Credit Card (KCC): Farmers' Credit Card.
4. 'apy': Atal Pension Yojana. 'Yojana' means scheme.
5. 'dbt': Direct Benefit Transfer.
6. 'jan_dhan': Pradhan Mantri Jan Dhan Yojana. Translates to 'Prime Minister's People's Wealth Scheme'.
7. 'rail_vikas': Railway Vikas Nigam Limited is owned by the Ministry of Railways involved in building rail infrastructure.
8. 'irsdc': Indian Railway Station Development Corporation.
9. 'scs', 'obc', 'obcs', 'scs_sts', 'scheduled_tribe', 'scheduled_caste': These words denote historically disadvantaged communities in India. Scheduled Castes, Scheduled Tribes, and Other Backward Classes are abbreviated as SC, ST, OBC, respectively.

7. Bibliographical References

Adhya, D., Dutta, K., and Basu, A. (2013). Japanese Encephalitis in India: risk of an epidemic in the National Capital Region. *International Health*, 5(3):166–168.

Anuja, G. V. (25 Nov 2020). Parliament may see historically low number of sittings this year. *Mint*.

Bhogale, S. (2019). TPCD-IPD: TCPD Indian Parliament codebook (question hour). *Trivedi Centre for Political Data, Ashoka University*.

Blei, D. M. and Lafferty, J. D. (2006). Dynamic topic models. In *Proceedings of the 23rd International Conference on Machine Learning*, pages 113–120.

Blei, D. M., Ng, A. Y., and Jordan, M. I. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research*, 3:993–1022.

Blei, D. M., Kucukelbir, A., and McAuliffe, J. D. (2017). Variational inference: A review for statisticians. *Journal of the American Statistical Association*, 112(518):859–877.

Dieng, A. B., Ruiz, F. J. R., and Blei, D. M. (2019). The dynamic embedded topic model. *CoRR*, abs/1907.05545.

Ghosh, S. and Basu, A. (2016). Acute encephalitis syndrome in India: the changing scenario. *Annals of Neurosciences*, 23(3):131.

Gkoumas, D., Pontiki, M., Papanikolaou, K., and Papageorgiou, H. (2018). Exploring the political agenda of the Greek Parliament plenary sessions. In *Proceedings of the LREC 2018 Workshop ParlaCLARIN: Creating and Using Parliamentary Corpora*.

Greene, D. and Cross, J. P. (2017). Exploring the political agenda of the European Parliament using a dynamic topic modeling approach. *Political Analysis*, 25(1):77–94.

Ishima, H. (2020). How electoral reform alters legislative speech: Evidence from the parliament of Victoria, Australia 1992–2017. *Electoral Studies*, 67:102192.

Kshatriya, R., Khara, N., Ganjiwale, J., Lote, S., Patel, S., and Paliwal, R. (2018). Lessons learnt from the Indian H1N1 (swine flu) epidemic: Predictors of outcome based on epidemiological and clinical profile. *Journal of Family Medicine and Primary Care*, 7(6):1506.

Lisena, P., Harrando, I., Kandakji, O., and Troncy, R. (2020). ToModAPI: A topic modeling API to train, use and compare topic models. In *Proceedings of Second Workshop for NLP Open Source Software (NLP-OSS)*, pages 132–140.

Makhija, A. K. (2006). Privatisation in India. *Economic and Political Weekly*, pages 1947–1951.

Martin, S. and Rozenberg, O. (2014). *The roles and function of parliamentary questions*. Routledge.

Marwah, V. (2019). Gender, caste and indian feminism: The case of the women's reservation bill. In *Women's and Gender Studies in India*, pages 151–163. Routledge India.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., and Dean, J. (2013). Distributed representations of words and phrases and their compositionality. *Advances in Neural Information Processing Systems*, 26.

Reddy, V. Y. (2019). Privatisation in India. *Privatisation in Developing Countries*, pages 178–198.

Sanyal, K. (2016). Regulating the regulators: Role of Parliament. *Economic and Political Weekly*, 51(13):16–19.

Sen, A., Ghatak, D., Kumar, K., Khanuja, G., Bansal, D., Gupta, M., Rekha, K., Bhogale, S., Trivedi, P., and Seth, A. (2019). Studying the discourse on economic policies in India using mass media, social media, and the parliamentary question hour data. In *Proceedings of the 2nd ACM SIGCAS Conference on Computing and Sustainable Societies*, pages 234–247.

Tripathi, V. and Kumar, R. (2021). Parliament amidst pandemic: Situating the opposition. *Economic and Political Weekly*, 56(33):16–19.

8. Language Resource References

Trivedi Centre for Political Data, Ashoka University. (2019). *TPCD-IPD: TCPD Indian Parliament Dataset (Question Hour) 1.0*.