

Classification of parliamentary speeches in the Japanese House of Representatives

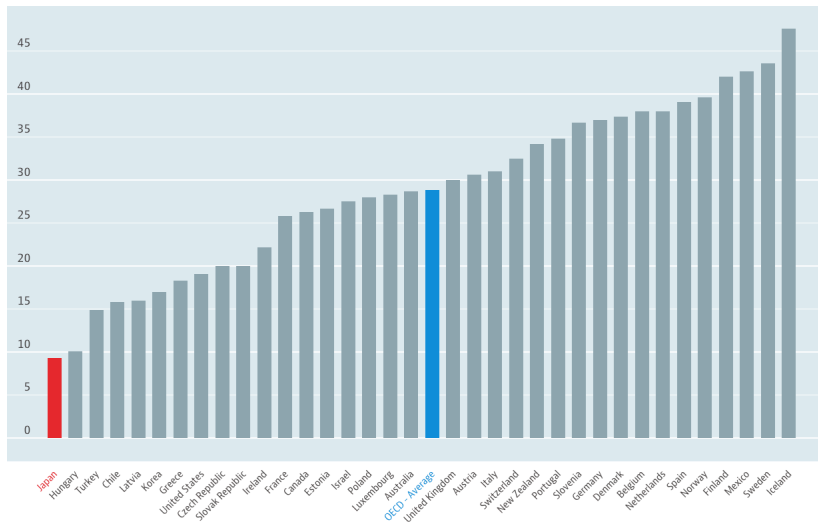
Gender and Party

September 27, 2019 (Kobe University)

Akitaka Matsuo

Motivation

- Analyze parliamentary speech in Japan's Lower House



Women in politics (Taylor-Robinson, 2014; Wängnerud, 2009)

- ▶ Descriptive representation
 - ▶ Proportion of female representatives matters
- ▶ Substantive representation
 - ▶ Female representatives promote women's interests
 - ▶ Bill sponsorship (Devlin and Elgie, 2008; MacDonald and O'Brien, 2011)
 - ▶ Ministerial appointment (Krook and O'Brien, 2012)
 - ▶ Legislative speech? : Female representatives speak less but speak more "feminine" policy areas (Taylor-Robinson, 2014; Bäck, Debus and Müller, 2014; Bäck and Debus, 2019; Catalano, 2009)

There is research on "when" female representatives speak, but no research on "what" they speak.

Feminine and Masculine Policy Areas

- ▶ Krook and O'Brien (2012): Categorize cabinet portfolio as "feminine" and "masculine".
 - ▶ Feminine policy area with the connections with the private sphere of the home and family, such as education and health care.
- ▶ Bäck and Debus (2019): Using this categorization, shows that female representatives speak more in feminine policy issues.

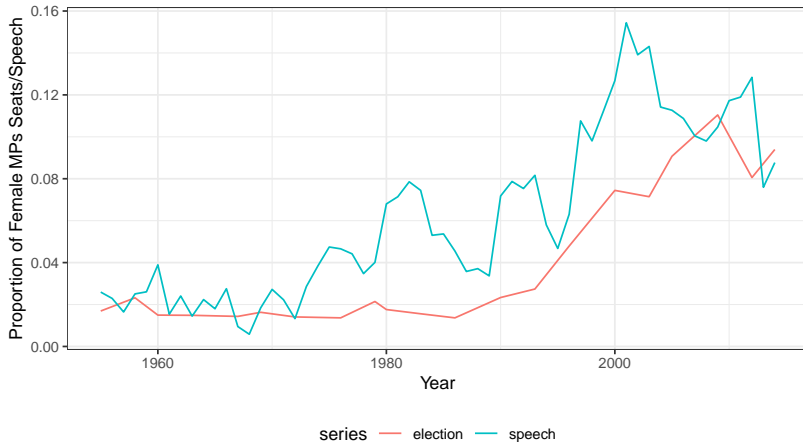
Question: Do female representatives' speech are different from male counterparts?

1. Is it possible to detect the gender of speaker from the speech text?
2. If 1. is true, which words/features are distinctive to female representatives' speech?
3. Any variations across policy fields?

Committee Speeches in the House of Representatives

- ▶ Period: 1955-2014
- ▶ Downloaded using Kokkai Kaigiroku API using `kaigiroku` package
- ▶ Using only rank-and-file member speeches
- ▶ Gender record from Reed and Smith data
 - ▶ Name matching
 - ▶ Yomi matching (using MeCab + NEologd-dictionary)

Proportion of Seats and Speeches of Female Representatives



Proportion of Female Members' Speech by Committee

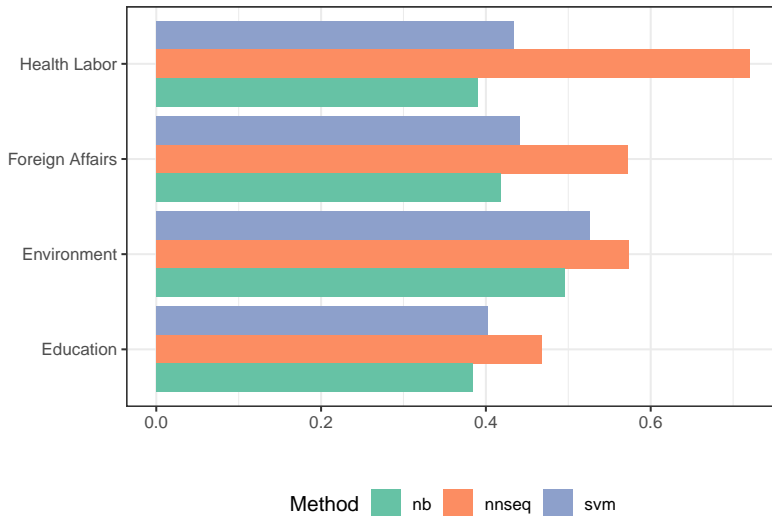
Committee Name	N	Prop	Feminine	Masculine
Environment	24404	0.215		
Foreign Affairs	107337	0.136		x
Education and Science	87356	0.114	x	
Health (Social) and Labor	252989	0.112	x	
Judicial Affairs	103398	0.056		
Security	16361	0.053		x
Agriculture and Fishery	127886	0.038		x
Budget Committee	179041	0.034		
Transportation	83964	0.032		x
...				

- ▶ Use top 4 committees in female representation
- ▶ Run analysis for each of the four Committees
- ▶ Two types of analysis
 1. Classification
 2. Document vector estimation

Is it possible to detect the gender of speaker from the speech text?

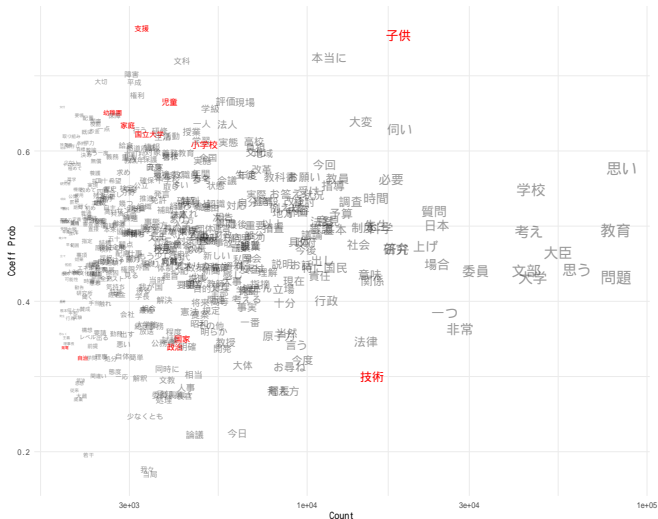
- ▶ Using the word as features, classify speakers gender
- ▶ Try several classification algorithm:
 1. Naive Bayes Classifier
 2. Linear Support Vector Machine
 3. (Shallow) Neural Net

Classification Performance: F1-score

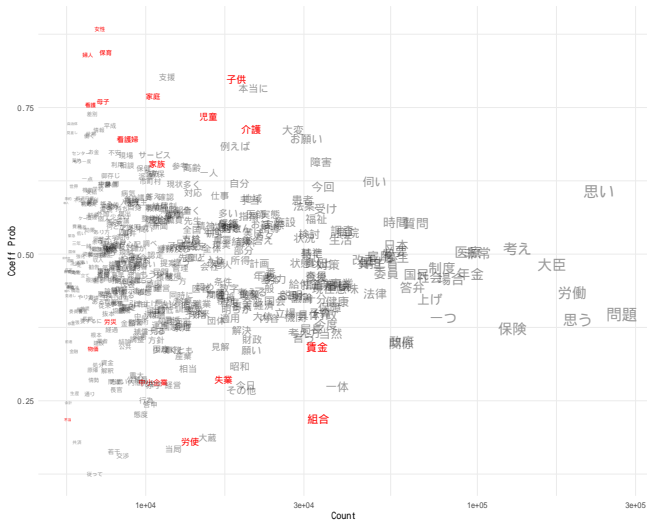


- ▶ Detection of gender is possible, and the performance is more or less the same across committees.
- ▶ What comes next is to explore which words/features are characteristics to female representatives' speech using coefficients from Naive Bayes.

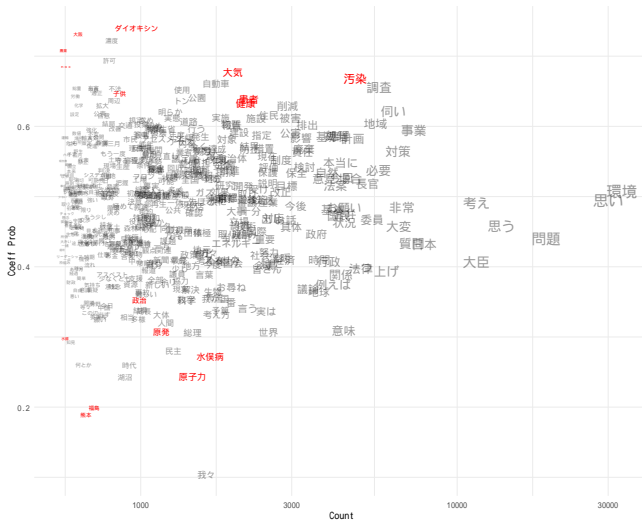
Naive Bayes Coefficients: Education



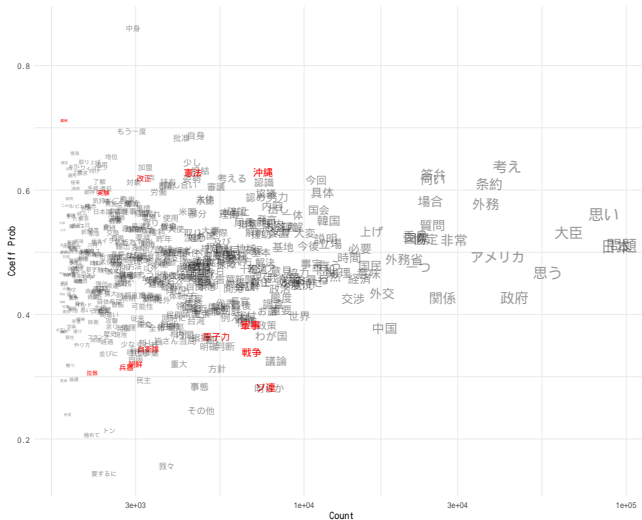
Naive Bayes Coefficients: Health and Labor



Naive Bayes Coefficients: Environment



Naive Bayes Coefficients: Foreign Affairs



Word embedding

- ▶ Convert each word into a dense vector (e.g. a 50 dimensional vector)
- ▶ Not a typical bag-of-words method
- ▶ Use the contextual information for estimation (e.g. sequences of words (word2vec), or word cooccurrence with narrow window (Glove))
- ▶ **Example usage of word vectors:**
 - ▶ Finding synonyms (e.g. for dictionary development)
 - ▶ Word arithmetic
 - ▶ $\text{King} - \text{Man} + \text{Woman} = ?$ (Mikolov et al., 2013)
 - ▶ "Man is to Computer Programmer as Woman is to Homemaker" (Bolukbasi et al., 2016)
 - ▶ Short text comparison (e.g. Similarity of SNS messages)

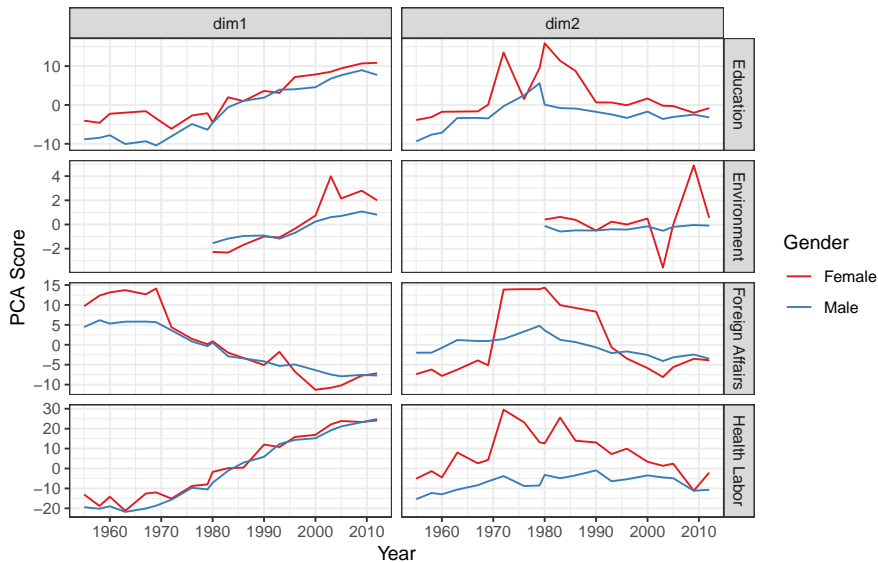
Analysis (2): Gender Embedding

- ▶ Analogues to **Party Embedding** (Rheault and Cochrane, 2019)
 - ▶ Use **doc2vec** (Le and Mikolov, 2014)
 - ▶ Use document dummy variable as input along with words
 - ▶ Words and documents are in the same vector space so the direct comparison between them is possible.
 - ▶ Use party * year as document variable
 - ▶ After estimating doc2vec model, conduct dimension reduction for party year vectors with principal component analysis (PCA).
 - ▶ Find the words related to the PCA dimensions.

In this research:

- ▶ Gender * (Election Year) as document level input
- ▶ Follow party embedding procedures
- ▶ Software: **doc2vec** in **gensim**
- ▶ Setting: 300 dim, 10 words window, 30 iterations

Analysis (2): Gender * Year output: PCA outcome



Analysis (2): Gender High PCA score terms (Dim 2)

Health and Labor

Dir	Terms
Pos	納得, 国民, 救済, 責任, 改善, 反対, 修正, 今日, 生活, 期待, 仕事, 厚生省, 与党, 賛成, 民主党, 議員, 実現, 協議, 図る, 政府, 実行, 実施, 国会, 法律, 昨年, 子供, 今回, 約束, 内閣, 遺族
Neg	意見, 分の, 方式, 数字, 高い, ページ, 基づく, 力所, 制度, 一手, 科学, 答弁, 階層, 項目, 幾ら, 種類, 諸国, 状況, ケース, 実績, 段階, 万人, 低い, 産業, 来る, 常用, 以下, 結論, 根本, お話

Education

Dir	Terms
Pos	授業, 終わる, 取り組む, 施設, 教える, 教師, 長い, 起こる, テスト, 続ける, 悪い, センター, 試験, 始める, 実際, 願う, 小学校, 選択, 入れる, 食べる, 参る, 移る, 職務, 給与, 全部, 時間, 起る, クラス, 民間, 挙げる
Neg	委員会, 意味, 認識, 承知, 質疑, 反対, 趣旨, 失礼, 質問, 態度, 経過, 報告, 無理, 表明, 一致, 国会, 解釈, 精神, 法的, 大変, 憲法, 附帯決議, 見解, 新聞, 目的, 公的, 要請, 答弁, 明確, 賛成

Analysis (2): High PCA score terms (Dim 2)

Foreign Affairs

Dir	Terms
Pos	出る, 協議, 可能性, 会談, 計画, 政策, 攻撃, 研究, 起こる, 段階, 立つ, 持ち込む, 対話, 見通し, 危険, 書く, 戦略, 来る, 変わる, 至る, 余地, 大きい, 能力, 構想, 始める, 疑惑, 余り, 準備, 起きる, 協定
Neg	認める, 理解, 保障, 賠償, 一番, 補償, フランス, 原則, 日本国, 独立, 確保, 予算, 払う, 保護, 返還, 行使, 一体, 放棄, 民間, 及び, 貿易, 平等, 防止, 決定, 期間, 行為, 禁止, 税金, 理由, 規定

Environment

Dir	Terms
Pos	中心, 明確, 検出, 発言, 立つ, 散布, 廃棄物, 使う, 計画, 立場, 発生, 挙げる, 産廃, 連携, 隠す, 焼却, 申請, 投棄, 参加, 見る, 図る, 認定, 不法投棄, 任せる, 通告, 残す, 収集, 派遣, 回収, 仕事
Neg	書く, 部分, 一点, 実効, 循環型社会, 理解, 省庁, 国民, 危惧, 分野, 一つ, 思う, 予算, 専門家, 条文, 結論, 状況, 部門, 基本法, 一般, 継続, 程度, 流れ, 人間, ある意味, 統計, 国内, 批判, 懸念, 技術

Takeaway

- ▶ Separation of speeches by female representatives from male representatives is possible, but depends.
- ▶ For the committees on “feminine” area, the factors that contribute to the separation are clear.

What to do next

- ▶ Incorporation of electoral contexts.
- ▶ More intensive use of word vectors.
 - ▶ Difference when participating the same debates?
 - ▶ Sentiment detection?

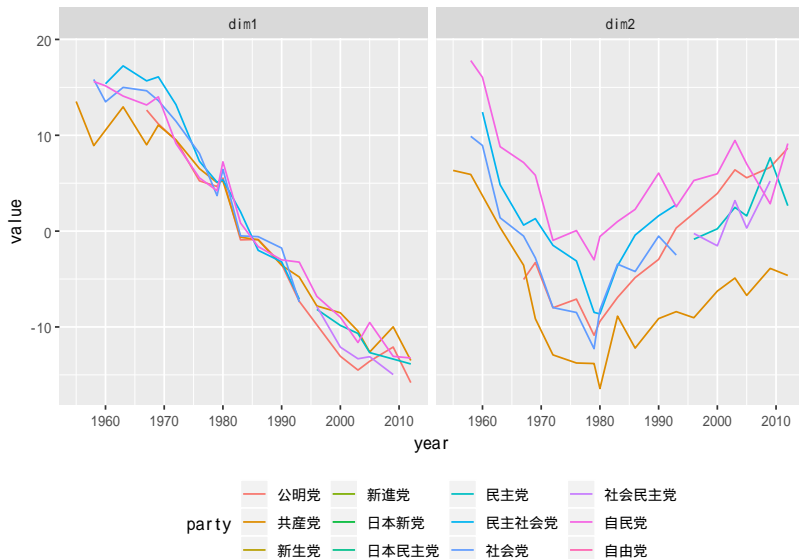
Reference I

- Bäck, Hanna and Marc Debus. 2019. "When Do Women Speak? A Comparative Analysis of the Role of Gender in Legislative Debates." *Political Studies* 67(3):576–596.
- Bäck, Hanna, Marc Debus and Jochen Müller. 2014. "Who Takes the Parliamentary Floor? The Role of Gender in Speech-making in the Swedish Riksdag." *Political Research Quarterly* 67(3):504–518.
- Bolukbasi, Tolga, Kai-Wei Chang, James Zou, Venkatesh Saligrama and Adam Kalai. 2016. "Man is to Computer Programmer as Woman is to Homemaker? Debiasing Word Embeddings." *30th Conference on Neural Information Processing Systems* (NIPS 2016):1–9.
- Catalano, Ana. 2009. "Women Acting for Women? An Analysis of Gender and Debate Participation in the British House of Commons 2005–2007." *Politics and Gender* 5(1):45–68.
- Devlin, Claire and Robert Elgie. 2008. "The effect of increased women's representation in parliament: The case of Rwanda." *Parliamentary Affairs* 61(2):237–254.
- Krook, Mona Lena and Diana Z. O'Brien. 2012. "All the president's men? the appointment of female cabinet ministers worldwide." *Journal of Politics* 74(3):840–855.
- Le, Quoc V. and Tomas Mikolov. 2014. "Distributed Representations of Sentences and Documents." 32.
- MacDonald, Jason A. and Erin E. O'Brien. 2011. "Quasi-experimental design, constituency, and advancing women's interests: Reexamining the influence of gender on substantive representation." *Political Research Quarterly* 64(2):472–486.
- Mikolov, Tomas, Kai Chen, Greg Corrado and Jeffrey Dean. 2013. "Efficient Estimation of Word Representations in Vector Space." pp. 1–12.
- Rheault, Ludovic and Christopher Cochrane. 2019. "Word Embeddings for the Analysis of Ideological Placement in Parliamentary Corpora." *Political Analysis* pp. 1–22.
- Taylor-Robinson, Michelle M. 2014. Gender and Legislatures. In *The Oxford Handbook of Legislative Studies*, ed. Shane Martin, Thomas Saalfeld and Kaare W. Strøm. Oxford University Press.
- Wängnerud, Lena. 2009. "Women in Parliaments: Descriptive and Substantive Representation." *Annual Review of Political Science* 12(1):51–69.

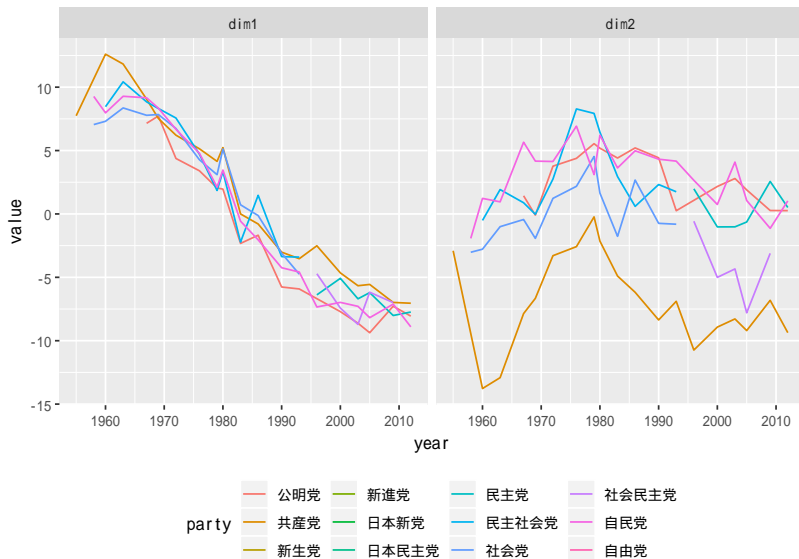
Addendum: Party Embedding in Japan?

- ▶ This addendum replicate Rheault and Cochrane (2019) with Parliamentary Speech in Japan
- ▶ Committees
 - ▶ Budget
 - ▶ Foreign Affairs
 - ▶ Agriculture
 - ▶ Health and Labor
- ▶ Party * (Election Year)

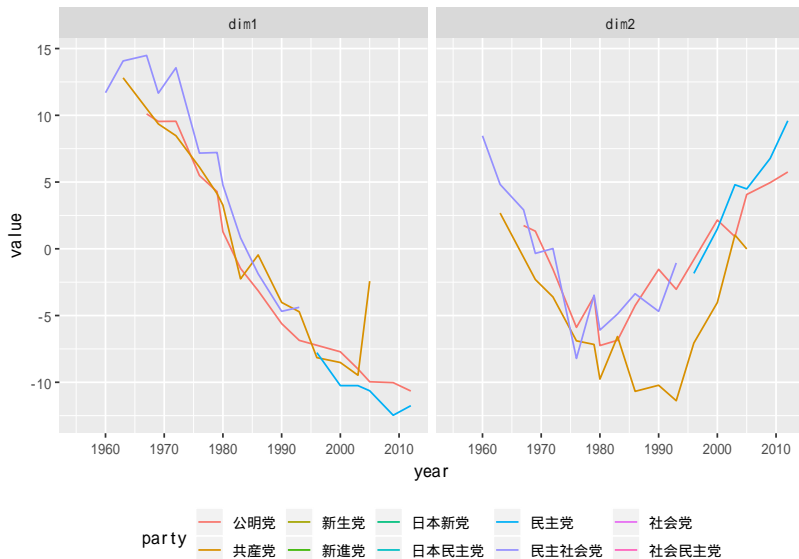
Addendum: Party Embedding in Japan: Budget



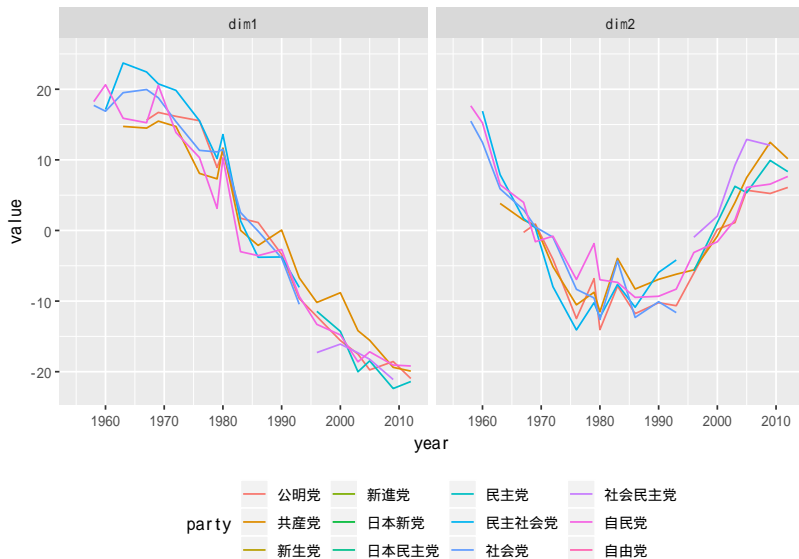
Addendum: Party Embedding in Japan: Foreign Affairs



Addendum: Party Embedding in Japan: Agriculture



Addendum: Party Embedding in Japan: Health and Labor



Addendum: Party Embedding in Japan

- ▶ The first dimension for each of the four committees is always the time trend.
- ▶ The second dimensions of Budget and Foreign Affairs seem to pick up something meaningful.

Party Embedding high score terms (Dim 2) I

Dir	Terms
Pos	社会, 立つ, 選挙, 来る, 述べる, 日本経済, 地方, 示す, 当たり前, 結構, 海外, 検討, 最後, 基地, 世界, 結果, 繰り返す, 悪い, 東京, 行く, 同時に, 大事, 何回, 演説, 中小企業, 必要, 張る, 深刻, 戦争, 研究
Neg	納得, 解決, 協力, 価格, 処理, 米価, 理解, 起こる, 責任, 発生, 払う, 主張, 決定, 改善, 安定, 立場, 思う, 取り上げる, 委員会, 入る, 保証, 発行, 一番, 輸出, 事件, 財政, 非常, 下がる, 大蔵省, 金利

Table: Party Embedding High Score Words: Budget

Party Embedding high score terms (Dim 2) II

Dir	Terms
Pos	全然, 対応, 安保条約, 以上, あり方, 何ら, 防衛, 条約, 全く, 見通し, 絶対, 対処, 事前協議, 現実, 憲法上, 終わる, 対策, 手続, 少し, 批准, 賠償, 法律, 決して, 余り, 仕方, 以前, 当面, 経緯, 最大, 考え方
Neg	自由, 出席, 受ける, 会う, 会談, 大統領, 読む, 意見, 各国, 書く, 重ねる, 訪問, 有力, 存じる, 大使, 思い, 多い, 載る, 人達, 公使, 一人, 北京, 送る, 重要, 上がる, 呼ぶ, 招く, プラス, 今度, 感じ

Table: Party Embedding High Score Words: Foreign Affairs

Party Embedding high score terms (Dim 2) III

Dir	Terms
Pos	働く, 国会, 数字, 法案, 保険, 審議, 趣旨, 現場, 行政, 法律, 言葉, 人達, 読む, 生活, 知る, 国民, 条文, 保健, 法律案, 業務, 自分, 被保険者, 実際, 職場, 資料, 審議会, 性格, 改正案, 健康保険, 中身
Neg	基準, 方式, 要求, 分の, 不足, 確保, 認める, 格差, 方向, 抑制, 補助, 支出, 徴収, 取得, お話, 養成, 差額, 得る, 引き上げる, 改定, 少ない, 助成, 示す, 解消, 引き上げ, 一体, 期待, 要請, 要望, 果たす

Table: Party Embedding High Score Words: Health Labor

Party Embedding high score terms (Dim 2) IV

Dir	Terms
Pos	方法, 報告, 重大, 正確, 起こす, 大きい, 深刻, 場所, キ口, 慎重, 緩和, 心配, やり方, 全然, 深める, 納得, 許す, 失う, 聞く, 実態, 困難, 繰り返す, 全区, 廃止, 意思, 広げる, 臨む, 停止, 切る, 積む
Neg	振興, 共済, 以上, 漁港, 基金, 改良, 農家, 占める, 基本計画, 計画, 年金, 年度, 伺い, 用水, 山間, 工業, 貸付, 需給, 入る, 自給率, 加入, 公団, 園芸, 時間, 交渉, 補助, 漁業, 保険, 農政, 農村

Table: Party Embedding High Score Words: Agriculture

- ▶ The better separation of LDP from other parties in Budget and Foreign Affairs is quite intuitive.
- ▶ The meaning of dimension needs further exploration.

Appendix

High PCA score terms (Dim 1) I

Dir	Terms
Pos	資料, 調査, 見る, 知る, 申し上げる, 最初, 踏まえる, 十分, 早い, 政府, 全国, 聞く, 今後, 重大, 結論, 国民, 調べる, 医者, 思う, 社会保障, 長官, お尋ね, 慎重, 職員, 明確, もう少し, 具体的, 率直, 行政, 国会
Neg	影響, 障害, 措置, 手当, 病院, 職業, 開発, 生活保護, 給与, 労働, 転換, 利用, 対象, 扶養, 規制, 支給, 診る, 政策, 指定, 変わる, 拡大, 不足, 廃止, 機能, 負担, 意思, 与える, 従事, 改正, 移行

Table: PCA High Score Words: Health Labor

High PCA score terms (Dim 1) II

Dir	Terms
Pos	日本, 教科書, 強い, 申し上げる, 段階, 申す, 全国, 先生, 単位, 結論, 答申, 三つ, 団体, 最後, 教育長, 進む, 事項, 教育, 自由, 都道府県, 中国, プルトニウム, 大学院, 心配, 自治, 学科, 明らか, 力国, 会議, 項目
Neg	設ける, 考える, 講じる, 向上, 含める, 確保, 通す, 努める, 充実, 見る, 調べる, 連携, 受ける, 作る, 方法, 話し合う, 講ずる, 通る, 与える, 下げる, 事故, 変える, 設置, 改正, 俣指, 準備, 乗る, 広げる, 図る, 訴える

Table: PCA High Score Words: Education

High PCA score terms (Dim 1) III

Dir	Terms
Pos	関連, 感じる, 保護, 比較, お尋ね, 駆除, お話し, 比べる, 特定, チェック, 適用, 解除, 反映, 一番, 浄化, 排水, 拝見, 委託, 環境庁, 着手, 解剖, 公表, アスベスト, 目的, 水道, 健康, 水質, 乗る, データ, 減少
Neg	分の, 年間, 必要, 議長, 出せる, 段階, 方針, 全国, 中央環境審議会, 至る, 取り組む, 岩手県, 早急, 代表, 丁寧, サミット, 御苦勞様, カ年, ヘクター, この間, 其の様, 環境教育, 予定, 早速, 決意, 九月, 先日, ポイント, 積極的, 現実

Table: PCA High Score Words: Environment

High PCA score terms (Dim 1) IV

Dir	Terms
Pos	求める, 要求, 話し合い, 得る, 進める, 表明, 地元, お話, 反対, 与党, 決議, 委員, 統一, 部分, 大体, 強化, 訴える, 期待, お答え, 削減, 予想, 納得, 皆様, 省庁, 説明, 設ける, 議員, 重要, 全区, 言う
Neg	歴史, 重大, 責任, プラス, 一点, 参る, 悪い, 多い, 事務, 違う, 一つ, 一条, 制度, ページ, 書簡, 安全, 大統領, 安全保障, 常識, 名前, 事件, 言葉, 一言, 国際, 見解, 技術, 事実, 過去, 今日, 性格

Table: PCA High Score Words: Foreign Affairs