

国会会議録を用いた言語表現の経年変化分析

山口昌也

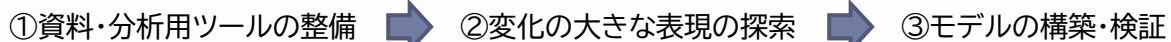
● 本研究の目的

- 国会会議録を対象に経年変化の大きい言語表現を見つけて、その特徴を分析
- 最終的に経年変化のモデルを作成することが目標

● 本研究の特徴

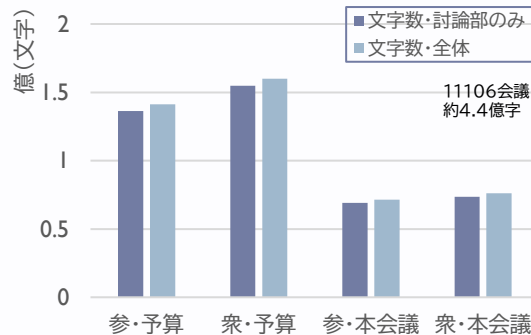
- 国会会議録の特性を利用
(記録期間が長い/議論の手順や発話者の役割が決まっているので、発話条件の均一性が高い)
- 経年変化の大きな表現を探索し、全体的な傾向を分析
(⇒ 特定の言語表現を対象に分析)

● 研究プロジェクトの流れ



● 分析対象の資料の整備

- 収集対象
 - 衆議院・参議院の本会議・予算委員会
 - 1947年～2012年の全議事録を収録
- 研究用のアノテーション
 - 文書構造(ヘッダ, 討議部, フッタ, 附属情報)
 - 会議情報(開催年月日, 回・号)
 - 話者情報(氏名, 生年, 肩書)



● 分析ツールの整備

- 全文検索システム『ひまわり』用のパッケージとして整備
 - 手軽に全文検索, 検索結果の集計が可能
 - XMLファイルなので、『ひまわり』以外でも利用可能
- 経年変化分析用の機能を拡充
 - 検索結果の集計や付与されたアノテーションの集計
 - 文字や単語のn-gramの作成



『ひまわり』



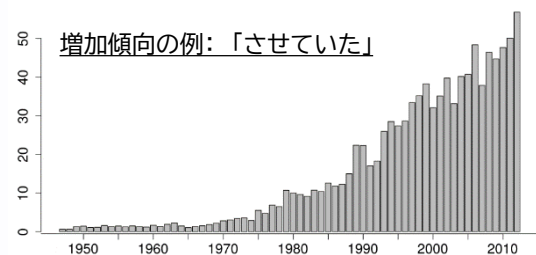
『国会会議録』

『ひまわり』、『国会会議録』パッケージはWeb上で公開

● 経年変化の大きな表現の探索(衆議院・予算)

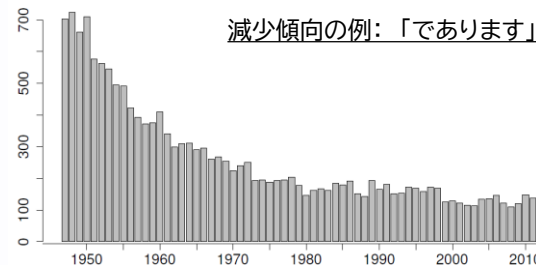
■ 探索方法

- 収録期間の初期(1947-1965)と後期(1999-2012)で、文字5-gramを比較し、調整頻度の変化の大きな文字列を抽出
- 抽出された文字列に対して、全期間での変化を目視で確認し、議事録作成時の表記規則の変更などに起因する候補を除外
(例:「我々」⇒「われわれ」)



■ 探索結果

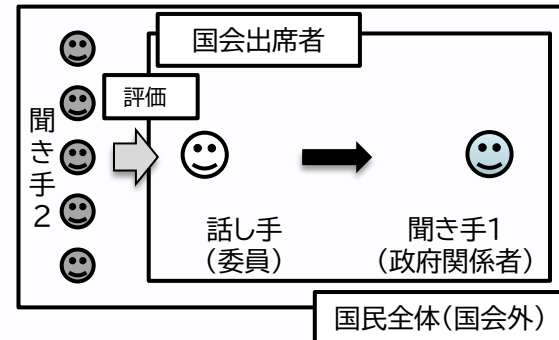
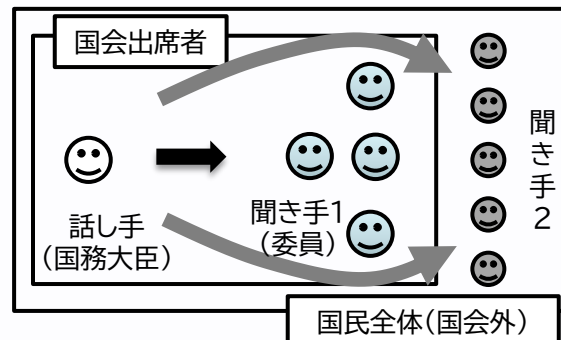
- 減少傾向: 国会特有の表現, 硬い表現
「おきまして」「であります」「ればならぬ」「おるのであ」「いうものは」「しましては」
- 増加傾向: くだけた表現, 丁寧な表現
「んですけど」「ふうに思い」「いるんです」「させていた」「ころでござい」「思っており」



● 経年変化モデルの構築

聞き手の範囲が聞き手1から聞き手2に拡大

聞き手2を考慮した表現に変化



仮説: 丁寧な表現↑, 国会特有の表現↓

仮説: 聞き手2に評価される表現↑

増加傾向: しっかりと、をさせてい、ふうに思い、だきました、っています、いるところ

増加傾向: ふうに思い、んですけれ、思うんです、いるんです、ないんです、させていた、んですね

仮説を否定するような結果はでていない