

# 『時事新報』初期の社説 の著者推定

安形輝（亜細亜大学）

上田修一（慶應義塾大学）

背景

實驗  
環境

實驗  
結果

結論



# 福澤全集中の論説記事

- 福澤自ら編んだ明治版全集には無署名論説記事は未収録
- 石河が編んだ大正版より無署名論説記事が全集に入る、昭和版続全集ではさらに追加
- 現行版では全21巻中9巻が論説記事
  - 合計1,553編
- 近年まで全集を基礎資料として福澤研究
  - 収録された論説記事に基づき、福澤のアジア観、例えば『脱亜論』は、特に議論の対象に
  - 福澤の真筆かの論争(井田2001以降 とて)



# 全集中の論説記事選定の疑わしさ

- 石河入社以前記事も石河が選別
- 外的な証拠(書簡、自筆原稿)より
  - 福澤自筆のものが外されている例
  - 福澤以外が執筆したものが含まれている例
- 石河自身の手による論説記事を全集に収録
  - 石河が昭和版『続福澤全集』第5巻附記に自作を全集に含めたことを書き残している

# 従来の著者推定＝井田メソッド

## □ 井田メソッド

- 語彙や文体の特徴から著者を推定
- 井田や平山は客観的な推定手法と主張

## □ 井田メソッドの問題

- 根拠とする語・表現の選定は判定者の判断  
→客観性、再現性に問題

例1) 井田は当初、論説『脱亜論』を石河起草と判定。その後、石河入社以前であることがわかり判定を覆す

例2) 井田、平山、安川の判定結果に違い

# 目標

石河入社  
以前

外的証拠  
から  
福澤以外

従来の  
推定は  
人手

全集論説  
記事は  
真筆？

客観的な  
著者  
推定実験

福澤かそれ以外の  
著作か

『時事新報』初期の  
論説記事

# 研究目的

- **機械的な著者推定手法の適用可能性**
  - 短いなどの不利な条件の論説記事を対象に
- **真筆が明らかな文献が対象の著者推定実験**
  - 福澤の真筆かどうかを判定可能か？

背景

實驗  
環境

實驗  
結果

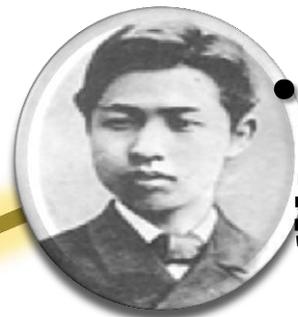
結論



・ 福澤諭吉



・ 中上川彦次郎



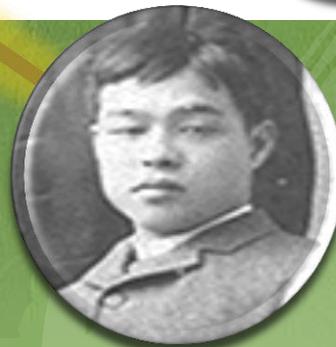
・ 高橋義雄



・ 波多野承五郎



・ 渡辺治



・ 石河幹明 などさらに寄稿者も

著者推定候補者  
(論説記者)

# 論説記者の在籍期間

明治 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34

福澤諭吉



中上川彦次郎



高橋義雄



渡辺治



波多野承五郎



石河幹明



菊池武徳



北川礼弼



堀江帰一



# 実験データの選定

真筆が明らかな

論説記事

雑誌論説記事

その他

高↑優先順位  
↓低

著者	論題(合計16編)	情報源
福澤諭吉	政略	福澤諭吉全集
	医薬分業行はれ難し	福澤諭吉全集
渡辺治	倶楽部ノコト	交詢雑誌
	領事裁判治外法権を論ず	交詢雑誌
	日本国ノ船業	交詢雑誌
高橋義雄	国債論	交詢雑誌
	教育論	交詢雑誌
	人種改良の問・社員某の答	交詢雑誌
中上川彦次郎	遣清特派全権大使	福澤諭吉全集
	朝鮮事変の処分法	福澤諭吉全集
	人民教育ノ説	民間雑誌
	青砥左衛門ノ話	民間雑誌
波多野承五郎	独逸国の武勇	交詢雑誌
	僻地政党者の心得	交詢雑誌
	演説の禁止を解釈す	交詢雑誌
	水月鏡花	交詢雑誌

# データ作成手順



元文献



データ入力

- ・可能ならOCR



データ後処理

- ・新・旧字体統一
- ・カタカナに統一
- ・句読点除去



著者名除去



実験データ

# 元文献の例

- 旧字体
- 印刷不鮮明

→OCR不可能

○演説ノ禁止ヲ解釋ス

波多野生

司法權ハ獨立シテ裁判官ハ獨行セサル可ラサルハ文明世  
 界ノ通論ナリ我國維新ノ初メ司法ノ省ヲ置キ新律改定ノ  
 二律ヲ制定シ近頃マタ新ニ刑法治罪法ヲ編成シテ客年一  
 月一日ヨリ施行セル如キ皆ナ此趣意ニ出テタルモノニ外  
 ナラサルヘシ然レモ趣意ト施行ハ往々ニシテ相背馳シ實  
 務家ト理論家ト毎ニ相反對スルハ誠ニ儘ナラヌ浮世ノ習  
 ヒニシテ亦タ人事ノ據ロナキ次第ナルガ之レヲシテ日ニ  
 一步ニテモ相近キテ而シテ互ニ干格スル所ナカラシムルハ  
 實ニ學者ノ本願ニシテ文明世界ノ急務トスヘキ所ナリ  
 然ルニ我が政府ガ曩ニ發行スル所ノ集會條例第六條第一  
 項ヲ見ルニ

# 用いる著者推定手法

- **圧縮改善係数による著者推定手法**
  - 圧縮プログラムを応用した著者推定手法
  - Benedettoらの手法を改良
- **近代日本文学者8人に対する著者推定で99%を超える推定精度**
  - **ただし、3万文字のデータに対する実験**
    - データを1万文字まで短くした場合でも80%以上の精度と性能低下は少ない
  - 7-zipとの組み合わせ

# 圧縮改善係数による著者推定手法

$$\text{圧縮改善係数} = 2 \cdot \frac{LZ_X + LZ_{A_i}}{L_{X+A_i}} - \frac{LZ_{X+A_i} + LZ_{A_i+X}}{L_{X+A_i}}$$

□ 値が高いほど類似度が高い→類似度順出力

# 著者推定の成功と失敗

データ(著者名+データ番号)

元文献

福澤諭吉01

高橋義雄02

成功

1

福澤諭吉02

福澤諭吉02

失敗

類似度順出力

2

高橋義雄02

福澤諭吉01

3

高橋義雄01

高橋義雄01

4

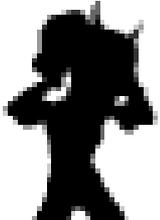
中上川彦次郎01

波多野承五郎02

5

中上川彦次郎03

中上川彦次郎02



# 今回の実験の難度の高さ

- 複数の関係者の存在
  - 起草者、執筆者、添削者、植字工などの存在
- 新聞記事の特性
  - 短いデータ(2,000字程度)
  - 文体がある程度統一
- 福澤以外の真筆文献の少なさ
  - 論説以外の文献
- 技術的な課題
  - 新・旧字体、異体字など
  - 句読点なし



背景

實驗  
環境

實驗  
結果

結論

# 福澤の真筆の判定

- **学術的・思想的に問題になるため**
  - **福澤か福澤以外かの推定**
    - **福澤と福澤以外の著者では立場、年代が異なる**
  - **すべての著者の推定**

著者	論題	成否	出力第1位
福澤諭吉	政略	○	福澤.医薬
	医薬分業行はれ難し	○	福澤.政略
福澤以外 ・渡辺治 ・高橋義雄 ・中上川彦次郎 ・波多野承五郎	倶楽部ノコト	○	中上川.人民
	領事裁判治外法権を論ず	○	渡辺.倶楽
	日本国ノ船業	○	中上川.人民
	国債論	×	福澤.医薬
	教育論	○	波多野.水月
	人種改良の問・社員某の答	○	中上川.人民
	遣清特派全権大使	○	中上川.朝鮮
	朝鮮事変の処分法	○	中上川.遣清
	人民教育ノ説	○	中上川.遣清
	青砥左衛門ノ話	○	中上川.人民
	独逸国の武勇	○	中上川.人民
	僻地政党者の心得	○	波多野.水月
	演説の禁止を解釈す	○	波多野.僻地
水月鏡花	○	波多野.僻地	

著者	論題	成否	出力第1位
福澤諭吉	政略	○	福澤.医薬
	医薬分業行はれ難し	○	福澤.政略
渡辺治	倶楽部ノコト	×	中上川.人民
	領事裁判治外法権を論ず	○	渡辺.倶楽
	日本国ノ船業	×	中上川.人民
高橋義雄	国債論	×	福澤.医薬
	教育論	×	波多野.水月
	人種改良の問・社員某の答	×	中上川.人民
中上川彦次郎	遣清特派全権大使	○	中上川.朝鮮
	朝鮮事変の処分法	○	中上川.遣清
	人民教育ノ説	○	中上川.遣清
	青砥左衛門ノ話	○	中上川.人民
波多野承五郎	独逸国の武勇	×	中上川.人民
	僻地政党者の心得	○	波多野.水月
	演説の禁止を解釈す	○	波多野.僻地
	水月鏡花	○	波多野.僻地

背景

實驗  
環境

實驗  
結果

結論

# 結論

## 機械的な 著者推定 手法

- ・ 再現性、客観性がある
- ・ 既往研究とは異なる点

## 真筆論説 記事実験

- ・ 一定精度での著者推定
- ・ 福澤か否か

# 今後の課題

- **問題となっている論説『脱亜論』に対する著者推定**
  - **データの追加**
    - 論説記者ではない寄稿者の著者候補者への追加
    - 真筆データの追加
  - **推定精度を上げるため**
    - 短いデータの連結といった工夫



**ご清聴ありがとうございました**

安形輝 (亜細亜大学)  
上田修一 (慶應義塾大学)

**『時事新報』初期の社説  
の著者推定**