

# 鶴岡言語調査データのコーホート分析

中村 隆 データ科学研究系 教授

## ◆ 背景

### ➢ 鶴岡言語調査

- 国立国語研究所が実施
  - 社会調査型言語調査, 定点経年調査 (他に岡崎調査, 北海道調査)
- 鶴岡市における共通言語化に関する調査
  - 第1次調査(1950), 第2次調査(1972), 第3次調査(1991), 第4次調査(2011)
  - 第1次・第2次調査では, 統計数理研究所のメンバーも中心的に活動
  - 第4次調査は, 統計数理研究所調査科学NOE形成事業の一環の連携研修調査として
- ランダム・サンプリング調査(RS調査)とパネル調査(P調査)
  - 各調査次で新しいサンプルを得るとともに, 過去の調査次の協力サンプルを追跡調査する

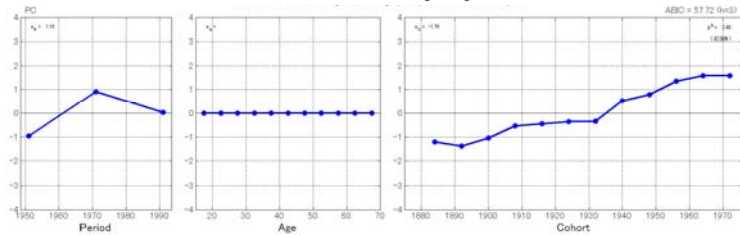
### ➢ 共通語化の機序は?

- コウホート効果(世代差)はあるだろう(世代交代による社会全体の変化)
- 時代効果もあるだろう(年齢や世代を問わない個人の変化)
- 年齢効果はあるのか?(加齢に伴って個人は変わるのか, 変わるとすれば, 共通語を使うようになるのかならないのか)

### ➢ コウホート分析(年齢・時代・世代効果の分離)

- 調査間隔が長すぎる(ほぼ20年間隔)が問題とならないか?
  - コウホート区分を5年幅とすると, 各コウホート区分を支えるデータセル数が少なくなる
  - コウホート区分幅を広げると, 支えるデータセルは多くなるが, 効果が一定とする制約がきつくなる

図2. コウホート分析結果 ([216] スイ「カ」)



## ◆ データ

- 第1次から第3次までの鶴岡RS調査
  - 各調査年次を1951年, 1971年, 1991年として扱った
- 年齢区分は, 15~69歳までの5歳幅
  - 今回の発表では, 男女別には分析していない
- 分析項目
  - 音声31項目, アクセント5項目, 計36項目
  - 各項目の共通語にあたる反応の割合(共通語率)

## ◆ 方法

### ➢ ベイズ型ロジット・コウホートモデル

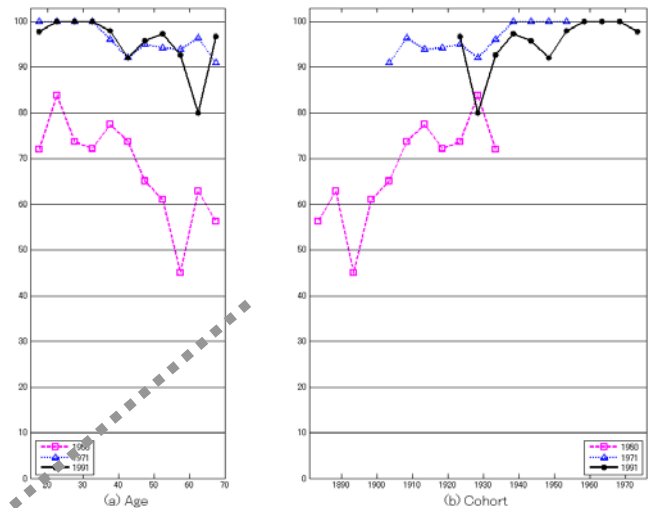
- 各項目の調査時点別年齢階級別の反応度数に2項分布を仮定
 
$$y_{ij} \sim \text{Binomial}(\pi_{ij}, n_{ij})$$
 Log-likelihood:  $\log f(\mathbf{y} | \boldsymbol{\pi}) = \mathbf{y}' \log \boldsymbol{\pi} + (\mathbf{n} - \mathbf{y})' \log(\mathbf{1} - \boldsymbol{\pi})$
- 反応確率のロジットを年齢・時代・世代効果の線形和で表現
 
$$\text{Model: } \boldsymbol{\eta} \equiv \log \boldsymbol{\pi} - \log(\mathbf{1} - \boldsymbol{\pi}) = \boldsymbol{\beta}^G \mathbf{1} + \mathbf{X}_A \boldsymbol{\beta}^A + \mathbf{X}_P \boldsymbol{\beta}^P + \mathbf{X}_C \boldsymbol{\beta}^C$$
- コウホート分析の識別問題を克服するため, パラメータの漸進的変化の条件を事前分布として付加し, ベイズ型モデルを構築
 
$$\frac{1}{\sigma_A^2} \sum_{i=1}^{I-1} (\beta_i^A - \beta_{i+1}^A)^2 + \frac{1}{\sigma_P^2} \sum_{j=1}^{J-1} (\beta_j^P - \beta_{j+1}^P)^2 + \frac{1}{\sigma_C^2} \sum_{k=1}^{K-1} (\beta_k^C - \beta_{k+1}^C)^2 \longrightarrow \text{small}$$
 Prior density:  $\pi(\boldsymbol{\delta} | \boldsymbol{\sigma})$ , Posterior density  $\propto f(\mathbf{y} | \boldsymbol{\delta}(\boldsymbol{\pi})) \pi(\boldsymbol{\delta} | \boldsymbol{\sigma})$
- 赤池情報量規準ABIC最小化法により超パラメータを推定
 
$$\text{ABIC}(\boldsymbol{\sigma}) = -2 \log \int f(\mathbf{y} | \boldsymbol{\delta}) \pi(\boldsymbol{\delta} | \boldsymbol{\sigma}) d\boldsymbol{\delta} + 2h$$

$$[h = \dim(\boldsymbol{\sigma}) + 1 \text{ (number of hyperparameters)}]$$
- パラメータの推定は事後尤度最大化(MAP)による
 
$$\text{MAP estimate: } \hat{\boldsymbol{\delta}} \text{ (or } \hat{\boldsymbol{\beta}}) = \arg \max_{\boldsymbol{\delta}} f(\mathbf{y} | \boldsymbol{\delta}) \pi(\boldsymbol{\delta} | \hat{\boldsymbol{\sigma}})$$

### ➢ 調査間隔問題

- コウホート区分幅を5~12年と変化させ, ABICによりモデル選択

図1. 共通語率【非唇音】 [216] 西瓜(スイ「カ」)



$$\text{Model: } \log[\pi_{ij} / (1 - \pi_{ij})] = \beta^G + \beta_i^A + \beta_j^P + \sum_{k=1}^K x_{i+(j-1)I,k}^C \beta_k^C$$

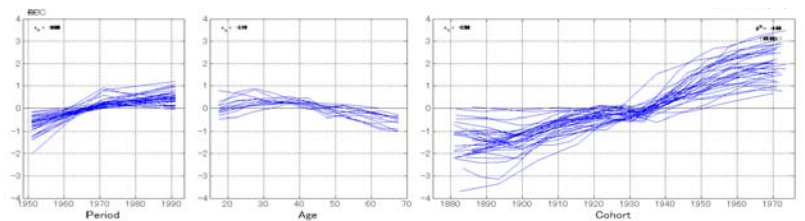
$$\text{Zero-sum constraints: } \sum_{i=1}^I \beta_i^A = \sum_{j=1}^J \beta_j^P = \sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K x_{i+(j-1)I,k}^C \beta_k^C = 0$$

$$\begin{aligned} \beta^G &: \text{grand mean} & \beta^A &= (\beta_1^A, \dots, \beta_I^A)' & \delta^A &= \beta_1^A - \beta_{I+1}^A & \delta^A &= (\delta_1^A, \dots, \delta_{I-1}^A)' & \mathbf{X}_A &= [x_{i+(j-1)I,l}^A] \\ \beta^P &: \text{age effect} & \beta^P &= (\beta_1^P, \dots, \beta_J^P)' & \delta_j^P &= \beta_j^P - \beta_{j+1}^P & \delta^P &= (\delta_1^P, \dots, \delta_{J-1}^P)' & \mathbf{X}_P &= [x_{i+(j-1)I,l}^P] \\ \beta_j^P &: \text{period effect} & \beta^C &= (\beta_1^C, \dots, \beta_K^C)' & \delta_k^C &= \beta_k^C - \beta_{k+1}^C & \delta^C &= (\delta_1^C, \dots, \delta_{K-1}^C)' & \mathbf{X}_C &= [x_{i+(j-1)I,l}^C] \\ \beta_k^C &: \text{cohort effect} & & & & & & & & & \boldsymbol{\delta} &= ((\delta^A)', (\delta^P)', (\delta^C)')' \\ & & & & & & & & & & \boldsymbol{\delta}_l &= ((\delta^A)_l', (\delta^P)_l', (\delta^C)_l')' \end{aligned}$$

## ◆ 結果

### ➢ 分析結果の重ね描きから

- 音声・アクセント36項目の分析結果



### □ コウホート効果

- 変化幅が3効果の中で大きい傾向. 新しい世代ほど共通語率が高い
- 1930年代生まれ以降で上昇が立ち上がる傾向

### □ 時代効果

- 近年ほど共通語率が上昇傾向(70年代から90年代に下降するものもある)
- 50年代から70年代の変化が大きい傾向

### □ 年齢効果

- 存在する. 個人の加齢による変化の他に, 社会移動による割合の変化もあり得る
- 30歳前後をピークに共通語率が下降する傾向

### ➢ 調査間隔問題

- コウホート区分幅のモデル選択が機能する
- コウホート効果が大きいと想定される項目であることも幸い

## ◆ 今後の課題

- 第4次調査の結果を待って再挑戦
  - 男女別データの分析
  - コウホート分析結果からみた項目の布置
- 調査間隔問題への別のアプローチ

## 文献

- 中村 隆 (2005). コウホート分析における交互作用効果モデル再考. 統計数理, 53(1), 103-132.
- 横山 詔一 (2011). 言語変化は経年調査データから予測可能か? 国語研プロジェクトレビュー, 6, 27-37.