

# 敬語意識の経年変化は統計モデルで予測可能か

横山詔一（国立国語研究所）・朝日祥之（国立国語研究所）

## 1. はじめに

### 1. 1. 目的

本研究は、愛知県岡崎市でなされた国立国語研究所による敬語調査のうち「身内敬語についての意見」の変化を経年的に観測したデータが「回答者の生年」と「調査年」の2変数で予測可能であるという仮説を検証する。国立国語研究所は岡崎市で敬語調査を縦断的（経年的）に実施してきた。第1次調査は1953年、第2次調査は1972年におこなわれた。横山詔一・朝日祥之・真田治子（2008）は、この2回分のデータを用いて統計モデルを作り、第3次調査の結果を予測した。以下、本研究は横山・朝日・真田（2008）の予測がどのくらいの的中したのかを2008年秋に実施された第3次調査の実測値を用いて吟味する。

なお、ここでは、敬語意識と、敬語に関する意見を区別しないことがある。

### 1. 2. 岡崎調査デザインの特長

岡崎調査のデザイン（設計）を図1に示す。これは、後で述べる横断法と縦断法を組み合わせた形になっており、ランダムサンプルが横断法に、パネルサンプルが縦断法にそれぞれ対応する。図1と同じデザインは、生涯発達心理学や老年学でも採用されることがあり、「コーホート系列法（cohort sequential method）」と呼ばれている。岡崎調査の具体的なデザインは以下のとおりである。

- 第1次調査：住基台帳を用いてサンプルをランダム抽出（434名）
- 第2次調査：第1次のサンプルを追跡調査（185名）＋新たなサンプルをランダム抽出（400名）→計585名のデータ
- 第3次調査：第1次のサンプルを追跡調査（20名）＋第2次のサンプルも追跡調査（62名）＋さらに新たなサンプルをランダム抽出（306名）→計388名のデータ

図1のようなコーホート系列法は、時代効果、加齢効果、世代効果の3者を分離可能な「もっとも効率的な調査法」として諸学界から高い評価を受けているが、手間がかかる。そのため、大規模な実査データは世界を見渡してもほとんどない。岡崎敬語調査のほかには国内では

山形県鶴岡市で国立国語研究所が行ってきた共通語化調査、海外では米国における「シアトル調査」があるのみのものである。

シアトル調査とは、「知能」の生涯変化を探る目的で実施されている大規模な縦断研究で、老年心理学の分野で世界的に有名である。基本的な調査デザインは岡崎や鶴岡の調査と同様で、米国のシアトル市で1956年から7年ごとに経年的に行われている。パネル追跡調査のほかに、毎回新たなサンプルを500名以上もランダム抽出している。その台帳は社会保険番号の名簿を使っているため、パネルの追跡はほぼ完全にできるようになっている。プロジェクト・リーダーを当初から努めているのは、ペンシルベニア州立大学のK.W.Shaie（シャイア）である（Shaie, 1996; Schaie and Hofer, 2001; Schaie and Willis, 2001）。

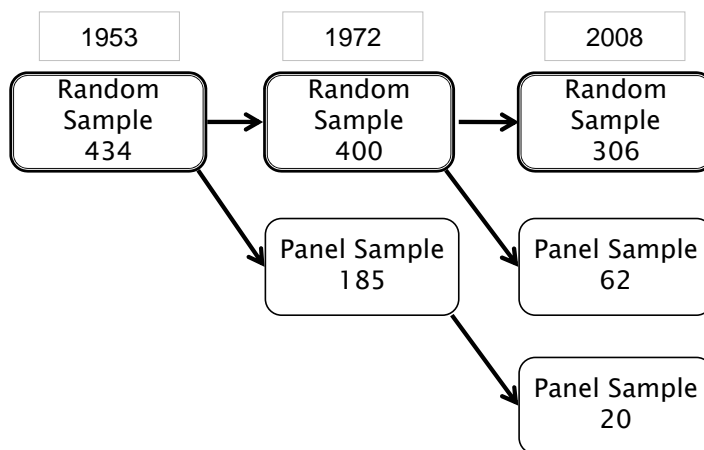


図1 岡崎調査のデザイン

### 1. 3. 言語記憶（言語習得）の統計モデル

「臨界期（言語形成期）に習得した言語運用能力は終生あまり変化せず、それが生涯にわたる言語運用能力を決定づける」という仮定のもと、1回の横断調査から過去の言語使用の姿を経年的に推定しようとする研究を「見かけ上の時間（apparent time）」による言語変化研究という。それに対して、実際の時間経過にそって経年調査をおこなう研究を「実時間（real time）」による言語変化研究という。ちなみに、医学、心理学、老人学などでは、

見かけ上の時間による研究を「横断研究 (cross-sectional study)」, 実時間による研究を「縦断研究 (longitudinal study)」と呼ぶのが普通である。

北米における研究では, 話者の発音が臨界期の後も言語共同体全体の変化と同じ方向に変化するという報告 (Boberg, 2004 ; Sankoff, 2006) のほか, 見かけ上の時間による調査は変化の速度を過小評価する傾向があるという説 (Sankoff, 2006) が出されている。日本では, 井上史雄 (2000) を中心として, 井上史雄・江川清・佐藤亮一・米田正人 (2009), 江川清 (1973) が, 鶴岡市の共通語化調査データを用いて, 成人後も生涯にわたって共通語化が進行している可能性を指摘している。

この点に着目した横山・朝日・真田 (2008) は, 認知科学の知見に基づく多変量解析モデルを作成し, 愛知県岡崎市における敬語意識の変化予測を試みた。本研究は, この横山・朝日・真田 (2008) や横山・真田 (2008) の手法に基づいて多重ロジスティック回帰モデルを作成し, その予測精度を吟味する。

#### ■ 言語変化と臨界期記憶・生涯記憶語について

ここでは臨界期記憶 (言語形成期記憶) と生涯記憶を区分し, 前者が生年に, 後者は調査年に, それぞれ結びついているという前提でデータ解析を試みる。生年と調査年の間に相関関係は存在しないので, 生年と調査年の変数が独立して敬語意識項目のどの選択肢を選ぶかの確率に寄与していると結論づけてよい。統計学的見地から, 生年と調査年を区分して扱うことの妥当性が担保されている。このことを踏まえたうえで, 横山・朝日・真田 (2008) による理論の骨子を以下に示す。

- (1) 生年は, 臨界期の時期を左右する変数である。調査対象者の臨界期は, 生年によってほぼ決まる。臨界期を過ぎた時期が 1950 年代であれば, 1950 年代の社会における敬語使用状況を反映した痕跡が記憶される。そして, その痕跡は変化することなく, その後の生涯を通じて安定して記憶内に保持され, 臨界期記憶という形で脳内に保持される。
- (2) 一方, 調査年は, 臨界期以降の生涯年数に関する変数である。調査年から生年を引き算すれば年齢が求まる。よって, 調査年が新しくなるほど, 同じ生年の調査対象者は臨界期以降の人生を長く経験したことになる。生涯を通じて言語刺激への接触記憶痕跡が累積的に貯蔵されると仮定すれば, 生涯年数に比例した大きさの痕跡強度が形成されるだろう。同じ生年であれば, 調査年が新しいほど生涯記憶の強度は大きいと考えられる。これは先に述べた臨界期記憶とは異なる性質の記憶である。

言語変化が, 調査対象者集団の脳内に貯蔵された臨界

期記憶と生涯記憶という質的に異なる 2 種類の記憶痕跡によってもたらされると仮定すれば, 次の図式を想定することができる。

臨界期記憶+生涯記憶 → 敬語運用の潜在能力  
( 臨界期記憶は生年依存, 生涯記憶は調査年依存 )

#### ■ 記憶メカニズムに基づく多変量解析モデル

生年ならびに調査年と, 臨界期記憶および生涯記憶の関係をまとめると次のようになる。

生年 → 世代効果 → 臨界期記憶  
調査年 → 加齢効果+時代効果 → 生涯記憶

これを敷衍すれば, 説明変数を生年と調査年の 2 変数, 目的変数を敬語意識とする, 以下のモデル式を立てることができる。

臨界期記憶+生涯記憶 → 敬語意識 [1]

先に述べたように, 臨界期記憶は生年と, 生涯記憶は調査年と, それぞれ強い関係がある。よって, 式 [1] は次のように書き換えてよい。

$a_1 \times \text{生年} + a_2 \times \text{調査年} + b \rightarrow \text{敬語意識の変化}$  [2]

式 [2] は, 生年と調査年という 2 つの変数だけで敬語意識の変化過程を説明しようとする言語変化予測式になっている。

さて, 言語変化に関する最近の計量言語学的研究の動向は, 時間だけを説明変数とするのではなく, 場面差などの要因も説明変数に取り込める多変量解析モデルに展開がなされている。この手法はすでに方言研究でのグロットグラムの解析に応用が試みられようとしている (横山・真田, 2007)。この点もふまえると, 岡崎市における第 3 次調査の敬語意識も横山・朝日・真田 (2008) の多変量解析モデルで予測可能であろうと考えられる。具体的には, 生年と調査年を説明変数とする多重ロジスティック回帰モデルの式 [3] を用いて分析を進めることにした。以下, 基本的に横山・朝日・真田 (2008) の記述を引用しながら新たな知見を示す。

$\log [p / (1-p)] = a_1 \times \text{生年} + a_2 \times \text{調査年} + b$  [3]

この式 [3] の右辺は, 先述の式 [2] と等しい。ちなみに, 本研究における統計モデルあるいは多変量解析モデルという用語は, 医学分野で盛んに利用されて

いる「ロジスティック回帰分析 (logistic regression analysis)」と統計学上は同じである。この解析では、臨界期に貯蔵された臨界期記憶だけが生涯を通じて敬語意識の判断を左右し、結果的に生涯記憶の効果がゼロである場合は、生涯記憶に関する変数のパラメータは統計学的に有意にならない。よって2つの記憶の影響関係を計量的に分離することができる。

#### 1. 4. 記憶における生年と調査年の役割

ここまでの議論をまとめると次のようになる。1965年生まれの人は、1970年代の経済高度成長時代に臨界期を過ごしたことになる。その人が10歳前後に過ごした環境が、あるいは、その時代の言語の状況が臨界期記憶に痕跡を残すと考えれば、「生年」は臨界期記憶に相関する変数だと見なしてよい。一方、「調査年」は臨界期以降に過ごした生涯の長さに関係がある。1965年生まれの人が1990年に調査を受けたとすると、その時の年齢は25歳である。調査年が2010年であれば、年齢は45歳になる。臨界期が10歳前後だとすると、調査時の年齢から約10歳を引けば、臨界期以降に過ごした生涯の長さが求まる。したがって、調査年は生涯記憶に相関する変数だと見なしてよい。

この解析手法のメリットは、敬語意識の経年調査データにおいて、臨界期記憶の効果のみが生じているのか、それとも臨界期記憶と生涯記憶の両者の効果があるのか、あるいは生涯記憶の効果だけがあるのか、この3つのいずれが妥当な結論なのかを、主観的・恣意的ではなく、統計学的・客観的にはっきりと判定できる点にある。パラメータが統計的に有意でない場合は、その変数は効果がないのであるから、モデルから除外される。

## 2. 予測の方法と結果

### 2. 1. 経年調査(1953, 1972, 2008)の概要

#### ■ 質問項目

第1次調査報告書(国立国語研究所, 1957)の198ページに掲載されている項目を使用した。すなわち「207家の中でも、年上の人や目上の人には敬語を使わなければならないでしょうか。それとも家の中では使わなくてもいいでしょうか」という質問であった。選択肢は「使うべきだ/時や場所や相手による/使わなくてもいい/不明」であった。第2次調査(第2次報告書117ページ: 国立国語研究所, 1983)。と第3次調査(2008年実施)でも同じ項目を使用した。

#### ■ 調査対象者

愛知県岡崎市の住民基本台帳にもとづいてランダムサンプリングされた住民であった。第1次調査は「不明」

を含めて434名。第2次調査は「無回答」を含めて400名。第3次調査は306名。年齢層は10歳代から50歳代以上まで。

### 2. 2. 統計モデルの作成

第1次・第2次(1953年と1972年)の調査データを用いて第3次調査(2008)の結果を予測するための統計モデルを作成した。具体的には、目的変数を「身内敬語を使うべきだ」という回答の割合、説明変数を生年と調査年の2変数とするロジスティック回帰分析をおこない、最尤推定法でパラメータを推定した。

#### ■ 解析に投入した数値

第1次調査報告書284ページの数表は、「不明」と回答した人数は分母に含めないで算出した百分率を掲出している。それに対して、第2次報告書117ページの数表は、分母に「無回答」を含めて計算した数値を示す。第2次報告書には、さらに、第1次調査の結果について、分母に「不明」と回答した人数を含めて再計算した百分率を掲出している。

本研究は第2次報告書の方針がより望ましいと判断し、それを解析に使用した。よって、第1次調査報告書284ページの数表とは微差がある。

#### ■ 重みの推定

目的変数  $p$  は「身内敬語を使うべきだ」の割合、説明変数は「生年」と「調査年」とした。生年は年代ごとに10年単位でまとめ、その中央の値とした。これは横山・真田(2007)と同じやり方である。

過去2回(1953年と1972年)の結果をロジスティック回帰分析の目的変数にして最尤推定法でパラメータを推定した結果、生年と調査年のいずれもWald検定にて1%水準で有意となり、2変数が敬語意識に影響を与えていることが明らかになった。 $p$ の予測値は式[4]により求まる。

$$p = 1 / [1 + \exp(0.0105 \times \text{生年} + 0.0467 \times \text{調査年} - 111.0888)] \quad [4]$$

## 3. 第3次調査の予測

### 3. 1. 多変量S字カーブによる予測

第1次・第2次のデータから導出した統計モデルの式に、第3次回答者の生年と調査年を代入すれば、「身内敬語を使うべきだ」という回答の割合の予測値が求められる。生年には1951(50歳代)、1961(40歳代)、1971(30歳代)、1981(20歳代)、1991(10歳代)を代入し、調査年には2008を代入して予測値を求めた。

その結果を図2に示す。予測値は○印で、実測値は●印で示した。この図の縦軸は選択確率である。以下、図は確率論にしたがって確率で表示するが、本文の数値などは一般になじみがある百分率(%)で表示する。ここでの百分率は確率を100倍したものである。

第1次・第2次調査は、予測値が実測値によく適合し、解析結果は良好である。予測値が描くカーブに関して、第1次・第2次調査と第3次調査のそれを比べると、調査回数に比例する形でカーブの傾きが小さく平坦になっていることが分かる。つまり、第1次調査から第2次調査までの19年間の変化が、そのままの勢いで現在まで持続しているとはいえ、減少傾向のスピードは鈍化しているため、まだ4%前後の勢力は残っているという予測になっている。ここではS字カーブ理論に基づく予測を行ったので、割合が20%を下回ったあたりから減少傾向のスピードは鈍化するのが普通である。

## 4. 考察

### 4. 1. 予測は的中したか

2008年に実施した第3次調査の実測値を予測値と突き合わせた結果、20歳代と30歳代では高い精度で一致し、予測は的中したと言えよう。一方、40歳代と50歳代は予測よりも多くの人々が「身内敬語を使うべきだ」と回答したことが明らかになった。図2を視察すると、予測の誤差は10%程度しかないものの、本研究の統計的な予測ではとらえきれない要因が潜んでいることが示唆される。このことから、40歳代よりも上の世代では、身内敬語使用の減少スピードは鈍化していると言える。なお、10歳代はサンプル数が10名に満たなかったため、考察の対象とはしない。

第3次調査の結果の結果から、生年(すなわち世代)の違いにより、身内敬語使用の減少スピードに違いが生じていることが明らかになった。中年層以上に身内敬語の保存意識が広まっているようであり、その要因を横山・朝日・真田(2008)の統計モデルに組み込むのが望ましいと考えられる。今後は中年層以上に身内敬語の保存意識が広まっている原因を詳しく究明する必要があるだろう。

**謝辞** 本稿の執筆にあたって岡崎調査のメンバーのみなさまから直接的、間接的に有益なコメントをいただいた。記して感謝申し上げます。

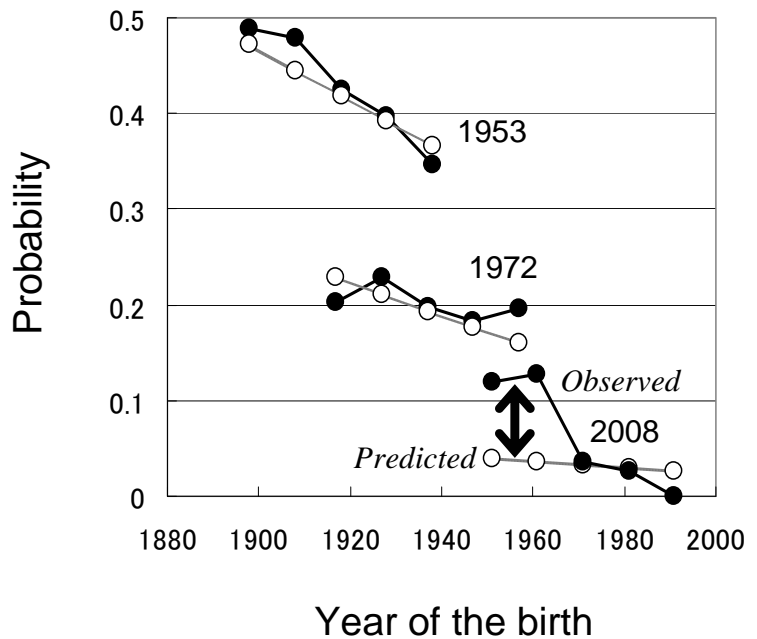


図2 「身内でも敬語を使うべきだ」の予測値と実測値

### 主要な参考文献

- Boberg, C. (2004). Real and apparent time in language change: Late Adoption of Changes in Montreal French. *American Speech*, **79**, 250-269.
- Chambers, J. K. (1998). Social embedding of changes in progress. *Journal of English Linguistics*, **26**, 5-36.
- 井上史雄 (2000). 東北方言の変遷 秋山書店
- 国立国語研究所 (1957). 国立国語研究所報告 **11** 敬語と敬語意識 秀英出版
- 国立国語研究所 (1983). 国立国語研究所報告 **77** 敬語と敬語意識 岡崎における20年前との比較 三省堂
- Labov, W. (1972). *Sociolinguistic patterns*. Philadelphia: University of Pennsylvania Press.
- Sankoff, G. (2006). Age: Apparent time and real time. *Encyclopedia of Language and Linguistics*, 2nd ed, 110-115, Oxford: Elsevier.
- Schaie, K. W., & Willis, S. L. (2001). *Adult development and aging*. 5th ed. New Jersey: Prentice Hall. 【シャイア&ウィルス, 岡林秀樹(訳) (2006). 成人発達とエイジング第5版, ブレーン出版】
- 横山詔一・真田治子 (2007). 多変量S字カーブによる言語変化の解析—仮想方言データのシミュレーション— 計量国語学, **26**, 79-93.
- 横山詔一・朝日祥之・真田治子 (2008). 記憶モデルによる敬語意識の変化予測 社会言語科学, **11**, 64-75.

**連絡先** 横山詔一 (よこやま・しょういち) 〒190-8561 東京都立川市緑町10-2 国立国語研究所 理論・構造研究系 [yokoyama@ninjal.ac.jp]