

令和 2 年 5 月 8 日現在

機関番号：62618

研究種目：挑戦的研究（萌芽）

研究期間：2017～2019

課題番号：17K18501

研究課題名（和文）疫学的統計手法と人工知能学の融合活用による敬語の変化予測研究

研究課題名（英文）Prediction of changes in honorifics by combining epidemiological statistical methods and artificial intelligence

研究代表者

横山 詔一（Yokoyama, Shoichi）

大学共同利用機関法人人間文化研究機構国立国語研究所・言語変化研究領域・教授

研究者番号：60182713

交付決定額（研究期間全体）：（直接経費） 4,800,000円

研究成果の概要（和文）：敬語規範意識の経年変化を疫学的統計手法の代表格であるロジスティック回帰分析モデルで予測する手法を発展させた。研究に用いたデータは、国立国語研究所が愛知県岡崎市で1953年から2008年までの55年間に3回にわたって収集した敬語に関する大規模な社会調査データであった。成果は査読付き学術誌で公刊するとともに、社会言語学や福祉言語学の流れのなかでどのように位置づけられるかを概観し、研究書として出版した。さらに、医療現場で医師が患者とコミュニケーションする際の敬語意識についても経年調査研究を実施した。人工知能学との融合研究については、データがビッグデータではないため、目立つ成果は得られなかった。

研究成果の学術的意義や社会的意義

疫学的統計手法を活用して敬語意識の変化予測に取り組んだ研究は、世界を見渡しても類例が非常に少ない。本研究は、これからの敬語研究に新たな地平を開く第一歩となったと考えられる。また、敬語研究と人工知能学との融合研究を将来的に展開するにあたって必要な条件なども明らかになった。敬語は時代とともに常に変化していくため、日本語の将来像を的確に見通すには敬語の変化を予測する研究が不可欠である。敬語や敬語意識の経年変化に関する科学的研究は、国語施策や国語教育施策のほか日本語教育でも必要とされており、社会的意義は大きい。本研究の方法論や知見は、さまざまな分野に貢献できる可能性がある。

研究成果の概要（英文）：We have developed a method for predicting changes in respectful normative awareness over time using a logistic regression analysis model, which is typical of epidemiological statistical methods. The data used in the study were large-scale social survey data on honorifics collected by the National Institute for Japanese Language and Linguistics in Okazaki, Aichi Prefecture, Japan, three times over a 55-year period from 1953 to 2008. The results have been published in peer-reviewed journals and as a research book with an overview of their place in the sociolinguistic and welfare linguistics streams. In addition, we conducted a longitudinal survey study on physicians' awareness of honorifics when communicating with patients in the medical field. As for the fusion research with artificial intelligence, no notable results were obtained because the data is not big data.

研究分野：社会言語科学

キーワード：敬語意識 言語変化 岡崎敬語調査 鶴岡共通語化調査 パネル調査 トレンド調査 医師の敬語意識  
ロジスティック回帰分析

## 様式 C - 19、F - 19 - 1、Z - 19 (共通)

### 1. 研究開始当初の背景

敬語は、国語施策の立案のほか日本語教育などにおいても重要な課題であり、科学的研究にもとづく知見が求められている。敬語は時代とともに常に変化するため、日本語の将来像を的確に見通すには敬語の変化予測研究が不可欠である。敬語の変化予測は広い学問分野の英知を結集して挑戦すべき最先端の学術的課題であり、社会的必要性も高い。

### 2. 研究の目的

国立国語研究所が愛知県岡崎市で1953年から2008年までの55年間に3回にわたって収集した敬語規範意識と敬語使用に関する大規模データを用いて、次のような課題に取り組んだ。

- (1) 経年的に住民基本台帳等からランダム抽出された話者のデータ(トレンド調査データ)を疫学的統計手法の代表格であるロジスティック回帰分析で解析し、「身内敬語の規範意識」について経年変化の数値予測をおこなった。解析には、国立国語研究所から公開されている岡崎敬語調査データベースを用いた。
- (2) ランダム抽出された人物を長期間追跡したデータ(パネル調査データ)の解析手法について検討した。
- (3) 医療現場での医師と患者のコミュニケーションについて、敬語使用に関する医師の意識を探る調査を実施した。
- (4) 敬語の変化予測に人工知能学の手法を導入できるか検討した。

### 3. 研究の方法

#### (1) ロジスティック回帰分析による身内敬語意識の経年変化に関する分析

母語の運用能力に関して「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、それが生涯にわたって言語運用能力を決定づける」という仮定にもとづいて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。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)」と呼ぶことがある。

欧米においても、実時間による研究が進められている。言語変化のあり方として、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研究のほか、見かけ上の時間による研究は変化の速度を過小評価する傾向があるという説、話者の発音が言語共同体全体の変化とは異なる方向に変化するという研究などが報告されている。日本では、井上(2000)を中心として、井上・江川・佐藤・米田(2009)、江川(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数量モデルを作成し、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(2007, 2010)の手法に基づいて、疫学的統計手法の代表格であるロジスティック回帰モデルを作成し、予測精度を吟味した。解析に用いたデータはトレンド調査データであった。

#### (2) 同一人物を追跡したデータ(パネル調査データ)の解析手法についての検討

岡崎敬語調査データはパネルデータの数が少ないため、パネルデータの解析手法の有用性を評価することが十分にはできない。そこで、国立国語研究所が山形県鶴岡市で1950年から2011年までの61年間に4回にわたって調査した共通語化に関する大規模データ(以下、鶴岡共通語化調査という)の一部を用いてパネルデータの解析手法に関する有用性の評価を試みた。

#### (3) 医療現場での医師の敬語使用に関する意識調査

敬語規範意識の経年変化を予測するための探索的研究の一環として、医療現場における医師と患者のコミュニケーションに関して、敬語使用に関する医師の意識を探るネット調査を実施した。

#### (4) 敬語の変化予測に人工知能学の手法を導入する試み

既存の人工知能システムに岡崎調査データを入力してディープラーニング(深層学習)やサポートベクターマシンなどによる解析をおこなうことを試みた。

### 4. 研究成果

#### (1) ロジスティック回帰分析による身内敬語意識の経年変化に関する分析

第1回調査と第2回調査のデータから導出したロジスティック回帰のモデル式(説明変数は回答者の生年と調査を受けた年)に、第3回調査の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得られる。生年には 1951 (50 歳以上), 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算出した。

その結果を図 1 に示す。予測値は印で, 実測値は印で示した。図 1 の縦軸は「身内敬語を使うべきだ」を選択する確率である。以下, 図の縦軸は確率で表示するが, 以下で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 回調査と第 2 回調査は, 予測値が実測値によく適合し, 解析結果は良好であった。予測値が描く S 字カーブに関して, 第 1 回調査・第 2 回調査と第 3 回調査のカーブを比べると, 調査回数に比例する形で傾きが小さく平坦になっていることが分かる。一般的に, ロジスティック回帰モデルでは, 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。ちなみに, 国立国語研究所 (1983, p.230) は, 第 2 回調査において第 1 回調査よりも家族同士では身内敬語を「使わない」人が有意に増加していると報告している。

結論として, 第 1 回調査と第 2 回調査のデータから導出したモデル式から得られた第 3 回調査の予測値は, 40 歳代と 50 歳以上では実測値との誤差が 8~9% であったが, それら以外の年代では誤差が 4% 未満におさまることが明らかになった。第 3 回調査の 40 歳代と 50 歳以上おける誤差がほかの年代よりも大きくなった理由については, さまざまな側面から検討をおこなったが十分な説明はできなかった。この点の解明は今後の研究に期待したい。

以上の成果は, 査読付き学術誌『計量国語学』に掲載された (横山・朝日, 2018)。また, 社会言語学や福祉言語学におけるロジスティック回帰モデルの位置づけを研究書『社会言語科学の源流を追う』(横山ほか編著, 1998) で解説した。

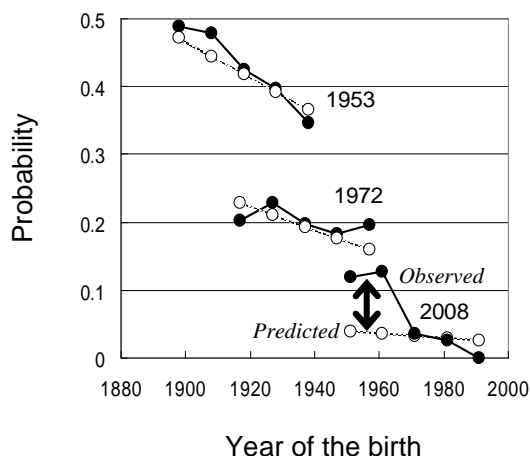


図 1 身内敬語意識の経年変化 (横山・朝日, 2018)

## (2) 同一人物を追跡したデータ (パネル調査データ) の解析手法についての検討

先にも述べたように, 岡崎敬語調査データはパネルデータの数が少ないため, パネルデータの解析手法の有用性を評価することが十分にはできない。そこで, 鶴岡共通語化調査のパネル調査データを用いて分散分析法などの解析手法の有用性を検討した。その成果は研究書『社会言語科学の源流を追う』(横山ほか編著, 1998) の一部として公開された。

パネル調査データを扱える分析手法は種類が限られている。また, トレンド調査データとパネルデータを同時に分析する手法も確立されていない。これらの問題を克服する解析手法の開発がこれからは必要となるであろう。

## (3) 医療現場での医師の敬語使用に関する意識調査

医師 (内科医) 110 名を対象にネット調査を 2018 年 10 月に実施し, Web を介して次の質問について回答してもらった。

質問「成人の患者と話すときの敬語の使い方についてお尋ねします。先生のお考えは, 次の選択肢のどれにいちばん近いですか。一つだけ選んでください。」

- 1 敬語はたくさん使って, できるだけ丁寧に話すべきだ
- 2 ホテルなどサービス業で使われているような接客敬語を使うべきだ
- 3 なるべく簡素な敬語を必要最小限に使うべきだ
- 4 丁寧語「～です・～ます」を使うだけで十分だ
- 5 敬語はなるべく使わず, 気さくに話すべきだ

この質問は, 吉岡泰夫ほか (2008) のネット調査で用いられたものと同じである。今回の結果と吉岡泰夫ほか (2008) の結果を比較した結果, 「敬語はたくさん使って, できるだけ丁寧に」の割合が増加していること等が明らかになった。その理由について, いくつかの説明が考えられるが, 結論を得るにはトレンド調査のような経年調査が不可欠である。

## (4) 敬語の変化予測に人工知能学の手法を導入する試み

岡崎調査で収集されたデータ量がディープラーニング (深層学習) やサポートベクターマシンなどにおいては十分ではないため, 学習がうまくいかないことが分かった。今後はデータ量を格段に増やす工夫 (大規模ネット調査など) が求められる。

5. 主な発表論文等

〔雑誌論文〕 計2件（うち査読付論文 1件/うち国際共著 0件/うちオープンアクセス 1件）

1. 著者名 横山詔一・朝日祥之	4. 巻 31(7)
2. 論文標題 身内敬語意識の55年間の变化に関する数量モデル：岡崎調査データにもとづく検討	5. 発行年 2018年
3. 雑誌名 計量国語学	6. 最初と最後の頁 497-506
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -

1. 著者名 横山詔一	4. 巻 2
2. 論文標題 言語の経年変化をロジスティック曲線で予測する	5. 発行年 2018年
3. 雑誌名 社会言語科学会 シリーズ社会言語科学（社会言語科学の源流を追う）	6. 最初と最後の頁 179-193
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 無
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

〔学会発表〕 計3件（うち招待講演 2件/うち国際学会 0件）

1. 発表者名 横山詔一
2. 発表標題 「病院の言葉」経年研究に向けた予備調査
3. 学会等名 国立国語研究所創立70周年記念シンポジウム 「経年調査の新たな挑戦 日本語の将来を占うために」（招待講演）
4. 発表年 2018年

1. 発表者名 横山詔一
2. 発表標題 言語変化と社会環境
3. 学会等名 日本語学会2017年度秋季大会シンポジウム「ルールを逸脱した表現の産出と許容」（招待講演）
4. 発表年 2017年

1. 発表者名 横山 詔一
2. 発表標題 異体字選好の生態学的モデル
3. 学会等名 機構間連携・文理融合プロジェクト「言語における系統・変異・多様性とその数理」シンポジウム
4. 発表年 2017年

〔図書〕 計1件

1. 著者名 横山 詔一・杉戸清樹・佐藤和之・米田正人・前田忠彦・阿部貴人編	4. 発行年 2018年
2. 出版社 ひつじ書房	5. 総ページ数 296
3. 書名 社会言語科学の源流を追う（社会言語科学会 シリーズ社会言語科学 2）	

〔産業財産権〕

〔その他〕

<p>書籍紹介：シリーズ社会言語科学2 社会言語科学の源流を追う  <a href="https://kotobaken.jp/books/news-181219-02/">https://kotobaken.jp/books/news-181219-02/</a>          創立70周年記念 シンポジウム「経年調査の新たな挑戦 日本語の将来を占うために」  <a href="https://www.ninjal.ac.jp/event/specialists/project-meeting/m-2018/20181222_3-sympo/">https://www.ninjal.ac.jp/event/specialists/project-meeting/m-2018/20181222_3-sympo/</a></p>
--

6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
研究 分 担 者	野崎 浩成  (Nozaki Hironari)  (80275148)	愛知教育大学・教育学部・教授    (13902)	