

## 若者ことばの定着と衰退・消滅： 多項分布型レジームスイッチングによる時期特定の試み

高橋 基規† 原田 健生† 山本 康裕† 柴田 希隆† 石川 泉吹†  
山崎 綾一郎† 山岸 祐己†† 谷口 ジョイ†§  
† 静岡理科大学 ‡ 浜松医科大学 § 国立国語研究所

### 1 はじめに

本研究は、現代日本語における若者ことばについて大規模調査を実施し、その結果を多項分布型レジームスイッチングによって分析することで、過去にどのような変化が生じたのか、またその変化が生成された時期についての特定を試みるものである。本調査で扱う時系列データ（生年）や地理的データ（出身地・居住地）は、情報量が多く、そのまま可視化しても解釈が困難である。よって本研究では、収集データに対してウェイトバック集計による「重み付け」を行った上で、データを単純化し、若者ことばの使用・理解にかかる変化とその時期を特定している。

### 2 調査の概要

#### 2.1 調査方法

現代日本語における若者ことばの使用実態および理解度の把握を目的として以下のような調査を実施した。調査対象は、18 歳から 68 歳の日本語母語話者および高度な日本語運用能力を有する者とし、計 5,334 名から有効回答を得た。調査は、スノーボールサンプリング法によるウェブ調査を採用し、回答者の基本属性（性別、生年、出身地域、現在の居住地）および 50 項目の若者ことばに関する使用・理解について回答を求めた。調査対象となる語は、先行研究に倣い、年鑑の用語辞典 [1] から抽出している。若者ことばの使用・理解に関する回答選択肢としては、「1. 使用する」「2. かつて使用していたが、今は使用しない」「3. 使用しないが、聞いたことがあり、意味もわかる」「4. 聞いたことがあるが、意味はわからない」「5. 聞いたことがない」の 5 段階を設定した。

Establishment and Decline of Japanese Youth Language: An Attempt to Identify Timing by Multinomial Distribution Type Regime Switching Detection

†Motoki TAKAHASHI †Takeo HARADA †Yasuhiro YAMAMOTO  
†Kiryu SHIBATA †Ibuki ISHIKAWA †Ryoichiro YAMAZAKI  
††Yuki YAMAGISHI †§Joy TANIGUCHI  
†Shizuoka Institute of Science and Technology  
‡Hamamatsu University School of Medicine  
§NINJAL

#### 2.2 分析方法

##### 2.2.1 回答のリサンプリング

令和 2 年国勢調査に基づく年齢別人口構成比を参照し、各年齢に対して補正係数 (ウェイト値) を付与することで、年齢ごとの偏りを調整した (人口構成は図 1 を参照)。令和 2 年国勢調査の年齢別人口構成比に基づいて各年齢のウェイト値を算出し、各年齢と各選択肢における回答数をウェイト値の合計 (実数) として扱う。さらに、その年齢ごと選択肢のウェイト合計値の分布と準乱数 [2] に基づいて、再現可能な回答シーケンスを生成しなおした。

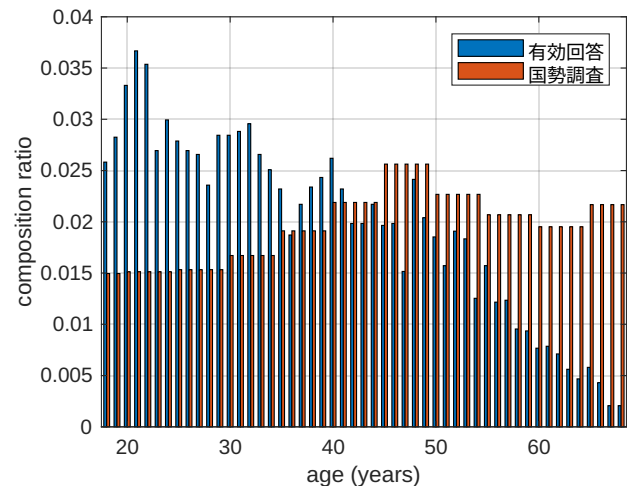


図 1: 年齢別構成比

##### 2.2.2 多項分布型レジームスイッチング検出

多項分布型レジームスイッチング検出は、複雑に変化するデータを単純化し、タイムライン可視化する手法である [3]。本研究では、データが多項分布に従っていると仮定し、多項分布型レジームスイッチング検出をデータの単純化手法として用いることで、若者ことばにおいて、過去に生じた変化とその時期を推定する。また、今回はモデル選択基準として最小記述長原理 (MDL) を採用し、アルゴリズムを自動で終了させる。既存の類似手法 [4] では、各カテゴリの平均値や分散、勾配などの変化を基準としており、事前に変化点数を与える必要があるが、多項分布の最尤推定における最適化問題を貪欲法と局所改善法によって解く本手法においては、全カテゴリの出現分布の変化を基準と

しており、変化点数の決定には情報量基準等を用いることができるため、多項分布の変化が生じた時期や期間に対し、なぜその結果になったのかというある程度の根拠を与えることが可能である。

### 3 結果および考察

事前の残差分析とそのクラスタリングにおいて、他の語とは異なる特徴をもつ語が複数抽出されたため、ここでは「ガクチカ」と「ワクテカ」のレジーム検出結果を示す。「ガクチカ」とは「学生時代に力を入れたこと」の略語であり、大学生の就職活動で用いられる若者ことばである。図 2 から、1990 年代後半生まれまでは、使用・理解の度合いが極めて低いことが見てとれるが、2000 年以降生まれになると「使用する」という回答が顕著に多くなっている。よって「ガクチカ」という語は、20 代半ばまでの若年代代によって使用され、今後、定着していくことが予想される。また、「ワクテカ」という語は、「胸がワクワク、肌や外見がテカテカきらめく状態」を意味する若者ことばであり、インターネット・コミュニティから発生した語である。一般的にこのようなインターネット・スラングは、使用の度合いが低く、理解度は高いことが分かっているが [5]、図 3 においても、同様の傾向が可視化されている。1980 年代後半生まれから 2000 年生まれまでは、「かつて使用した」という回答が多く、それ以降は「聞いたことがない」との回答が増加していることから、この語が既に衰退傾向にあることが示された。

