

## 若者ことばの使用実態に見られる属性差 - ウェイトバック集計による再評価

原田 健生<sup>1</sup> 山本 康裕<sup>1</sup> 柴田 希隆<sup>1</sup> 石川 泉吹<sup>1</sup> 山崎 綾一郎<sup>1</sup>  
山岸 祐己<sup>1,2</sup> 谷口 ジョイ<sup>1,3</sup>

<sup>1</sup> 静岡理工科大学 <sup>2</sup> 浜松医科大学 <sup>3</sup> 国立国語研究所

### 1 はじめに

「若者ことば」は、移り変わりが激しく、どのような要因によって広がり、どのような語が消失するのか、あるいは消失を免れて、幅広い年代に定着するのか、明らかになっていない。本研究は、若者ことばに関する大規模調査により、その使用・理解の変化について検討するものである。本発表では、調査によって得られたデータについて、ウェイトバック集計を適用し、先行研究と比較した上で、分析結果にどのような違いが生じるかを検証している。既に、無調整のデータを用いた独立性検定および残差分析により、地域差や性差、年代差に関する考察は行ったが、本研究では、回答者分布の偏りを補正したデータを用いて同様の分析を行い、語ごとの定着や衰退・消滅の傾向について考察を行った。

### 2 先行研究

本節では、若者ことばに関する先行研究について、その概要を述べる。本研究における「若者ことば」の定義は、既存研究 [1] に倣い、「主に若年層により、仲間内のコミュニケーションで使用される」「言語規範からの逸脱性を有し、特にメディアを介した伝播などが見られる」といった特徴を有するものとする。これまでの研究では、特定の語の統語的、あるいは語用論的特徴について述べたものが多く [2]、地域差、年代差、性差による使用・理解の違いについて定量的に調査したものは、大学生を対象としたもの [3] のみである。

### 3 研究手法

#### 3.1 調査対象者および調査方法

調査対象は、18 歳以上の日本語母語話者、あるいは高度な日本語運用能力を有している人であり、合計 5,374 名から回答が得られた（性別、年代の内訳は表 1 を参

照）。ウェブ調査により、性別、生年、出身地域、現在の居住地、50 項目の俗語に関する使用・理解について回答を依頼した。調査対象となる語は、昨今の代表的な若者ことば [4] から抽出している。俗語の使用・理解に関しては、「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」という 5 つの選択肢を設けた。

表 1: 調査協力者の内訳（性別・年代）

年代	男性	女性	他	総計
10 代	124	158	7	289
20 代	484	1,023	72	1,579
30 代	307	979	48	1,334
40 代	239	848	35	1,122
50 代	200	552	14	766
60 代	102	150	7	259
70 代	10	15	0	25
総計	1,466	3,725	183	5,374

#### 3.2 調整済み標準化残差

質問項目に対する評価と回答者の属性のクロス集計表における調整済み標準化残差の計算方法について説明する。まず、問題設定として以下を定義する。

- $A = \{a_1, a_2, \dots, a_I\}$ : 質問項目に対する評価（行方向のカテゴリ）
- $B = \{b_1, b_2, \dots, b_J\}$ : 回答者の属性（列方向のカテゴリ）
- $n_{ji}$ : 評価カテゴリ  $a_j \in A$  と属性カテゴリ  $b_i \in B$  の交差セルにおける観測度数
- $n_i = \sum_j n_{ji}$ : 属性カテゴリ  $b_i$  に属する回答者数（列合計）
- $n_j = \sum_i n_{ji}$ : 評価カテゴリ  $a_j$  の合計（行合計）
- $n = \sum_{j,i} n_{ji}$ : 総回答数

このとき、評価カテゴリ  $a_j$  と属性カテゴリ  $b_i$  の交差セルにおける期待度数  $e_{ji}$  は以下の式で計算される。

$$e_{ji} = \frac{n_j \cdot n_i}{n}$$

Attributional Differences in the Use of Japanese Youth Language: A Reassessment Using Weighted Back Tabulation

<sup>1</sup> Takeo HARADA <sup>1</sup> Kosuke YAMAMOTO <sup>1</sup> Ibuki ISHIKAWA <sup>1</sup> Kiryu SHIBATA <sup>1</sup> Ryoichiro YAMAZAKI <sup>1,2</sup> Yuki YAMAGISHI <sup>1,3</sup> Joy TANIGUCHI

<sup>1</sup> Shizuoka Institute of Science and Technology

<sup>2</sup> Hamamatsu University School of Medicine

<sup>3</sup> NINJAL

さらに、観測度数と期待度数の差異を補正した調整済み標準化残差  $r_{ji}$  は以下の式で計算される。

$$r_{ji} = \frac{n_{ji} - e_{ji}}{\sqrt{e_{ji} \left(1 - \frac{n_{j.}}{n}\right) \left(1 - \frac{n_{.i}}{n}\right)}}$$

### 3.3 ウェイトバック集計とクラスタリング

回収サンプルにおいて年齢ごとの分布が日本に住む日本人の人口構成から乖離していたため、分析にあたってはウェイトバック集計を行った。令和2年国勢調査\*に基づいて各年代のウェイト値を算出し、各年代と各選択肢における回答数  $n_{ji}$  をウェイト値の合計(実数)として扱う。クラスタリングでは、調整済み標準化残差行列  $R = [r_{ji}]$  間の L1 ノルム  $\|R_1 - R_2\|_1$  を距離定義とし、群平均法によって階層クラスタリングを行う。今回はクラスタ数を年代のカテゴリ数同等の7としている。

## 4 結果と考察

図1, 2にウェイトバック適応前後の年代カテゴリを用いた階層クラスタリング結果を示す。ウェイトバック適応前の年代カテゴリにおける階層クラスタリングの結果は、クラスタが大規模だったのに対し、適応後では、より多くのクラスタに分散しており、年代差からの詳細な考察が可能となっている。残差分析の結果に基づき、図2のクラスタを検討すると、「り(=了解)」や「びえん(=悲しい、つらい、泣いた)」といった語は、既に衰退傾向を見せている語群であることが示唆された。

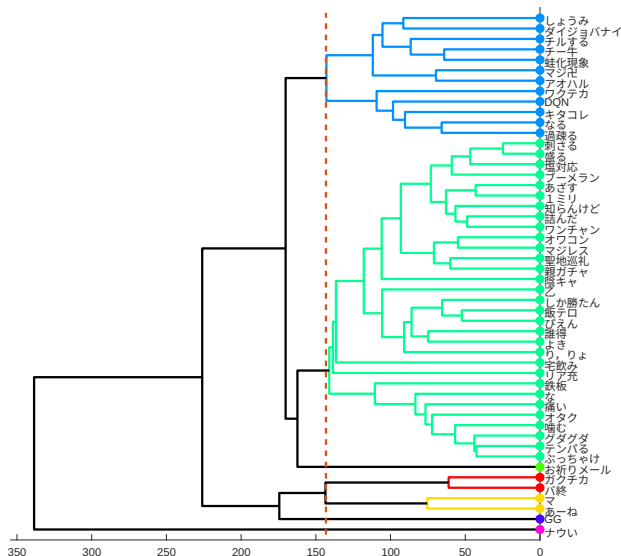


図1: ウェイトバック適応前の年代カテゴリにおける階層クラスタリング (クラスタ数7, 群平均法, L1 距離)

\*[https://www.e-stat.go.jp/stat-search/files?stat\\_infid=000032142410](https://www.e-stat.go.jp/stat-search/files?stat_infid=000032142410)

一方で、「ガクチカ(=学生時代に力を入れたこと)」や「バ終(=バイト終わり)」といった語は、若年世代のみに使用される語群に属する。また、「オタク」、「テンバる」などは、中年世代まで広く浸透し、定着の傾向がある語群に属している。

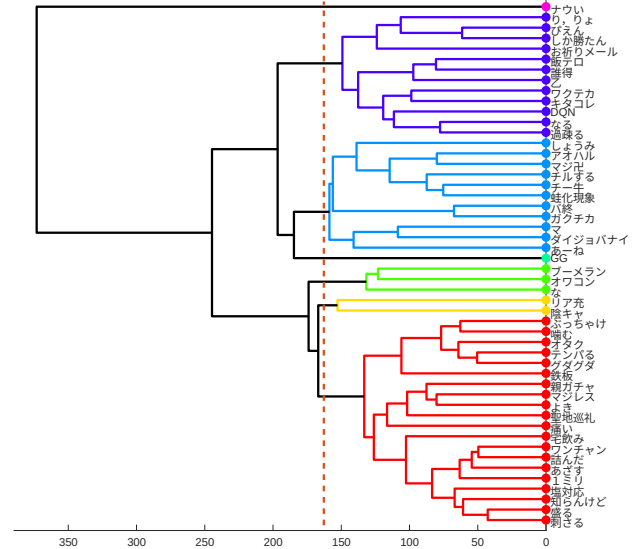


図2: ウェイトバック適応後の年代カテゴリにおける階層クラスタリング (クラスタ数7, 群平均法, L1 距離)

## 5 まとめ

本研究では、若者ことばに関する大規模調査によって得られたデータについて、回答者分布の偏りを補正した上で分析を行い、各語の定着や衰退・消滅の傾向について検証した。その結果、特定の語において地域・年代といった属性別の傾向に違いが見られ、サンプリング補正によって、より精度の高い結果が得られる可能性が示された。これにより、若者ことばの動態をより正確に把握する上で、サンプリング補正をしたうえでの残差分析が有効であることが分かった。

## 参考文献

- [1] 堀尾佳以. 若者言葉の研究: SNS 時代の言語変化. 九州大学出版会, 2022.
- [2] 工藤俊. 若者ことば「マジ卍」の言語的特徴—「やばい」との比較を中心に—. 駒沢女子大学研究紀要, Vol. 26, pp. 63–74, 2019.
- [3] 鍵水兼貴. 「全国若者語調査」結果概観. 専修国文, Vol. 94, pp. 93–117, 2014.
- [4] 現代用語検定協会(編). 現代用語の基礎知識 学習版 2023-2024. 自由国民社, 2023.