

国立国語研究所学術情報リポジトリ

『日本語話し言葉コーパス』に捉えられた言語変異現象

メタデータ	言語: jpn 出版者: 公開日: 2020-06-29 キーワード (Ja): キーワード (En): 作成者: 前川, 喜久雄, 小磯, 花絵, 菊池, 英明, 間淵, 洋子, 斎藤, 美紀 メールアドレス: 所属:
URL	https://doi.org/10.15084/00002949

『日本語話し言葉コーパス』に捉えられた言語変異現象

前川 喜久雄 小磯 花絵 菊池 英明 間淵 洋子 斎藤 美紀
(独)国立国語研究所研究開発部門

1 目的

言語学的には同一とみなされる音韻や語が様々に異なる形で発音される言語変異現象は自然言語の基本的な特徴のひとつである。言語変異を説明する要因には、言語的要因と社会的要因が共に関与することが知られている[1]。

『日本語話し言葉コーパス』(以下 CSJ)には、一定の方針に基づいて発話スタイルが異なる音声組織的に収録されており、さらに、それらが聴き手に与える印象が印象評定値という形で提供されているから、言語変異研究の有益な研究資料となると思われる[2]。

以下では、これまでに実施した予備的な解析結果を報告するが、紙幅の関係でグラフや表の大部分はポスターに譲ることとする。また CSJ に関する用語は説明なしに利用しているため、不明な用語については本研究会の前川の講演を参照していただきたい。さらに、以下の分析は CSJ の構築過程の折々に実施してきたものなので、利用したデータの量や分析手法は現象ごとに異なっていることもお断りしておきたい。

2 データ

今回の分析でとりあげた言語変異現象は以下の4種類である。1) 母音の無声化、2) 長母音の短呼、3) 「デウ」の「ジャ」への縮約、4) 助詞「ノ」の撥音化、ただし、3)と4)は品詞の相違によってそれぞれ二種類に下位区分できるから、実際には6種類の現象を検討した。1)、2)は音声・音韻、3)、4)は形態レベルの変異現象である。

3 全体的評価

表1に、6種の変異×6種の社会的要因=36種の組み合わせのそれぞれに対して、一元配置の分散分析を個別に適用した結果を示す[3]。とりあげた社会的要因は、1) 講演の種類別(学会講演か模擬講演か)、2) 発話速度、3) 発話スタイルの印象評定値、4) 自発性の印象評定値、5) (笑)タグ(笑いながらの発話)の有無、6) 話者の性別である。評定値はいずれも1~5の範囲に分布する。

すべての変異について最低でも4種の要因が統計的に有意な効果を示しており、一方、発話速度以外の要因は、最低で4種の変異について有意である。これによって、CSJ に記録された言語変異の平均値が社会的要因によって有意に変動していることが確認された。以下では、その変動の具体的ありさまを、言語的要因の分析も含めて個別に検討してゆくことにする。

4 母音の無声化

母音の無声化は典型的には狭母音が前後を無声子音に挟まれた環境で生起するから、音声生理学的な必然性をもつ現象である。要因2)の有意性はこの点で納得のゆく結果である。一方、無声化は語の意味に影響することがない音声レベルの変異である点で他の現象と著しく異なっている。そのため、発話スタイルによる影響は弱いものと予想されたが、表1では要因3)、4)も高い有意性を示した。この事実は、無声化にもやはり話者によって(無意識のうちに)制御されている側面があり、それが自発音声においては、顕著に観察されている可能性を示唆している[3,4]。

5 長母音の短呼

母音の長短は意味の対立に関わる日本語の音韻特徴である。この現象には、言語的要因と社会的要因が共に強い効果を有することが判明している[5]。興味深いことに、言語的要因(語種、長母音の語中位置、長母音自体のアクセント核の有無など)は、それぞれが統計的に有意な効果を示し、さらに複雑な交互作用を示すのに対し、表1にとりあげた社会的要因は、いずれも単独では有意な主効果を示さず、交互作用項(講演種別*(笑)と性別*(笑))だけが統計的な有意性を示す。(笑)は、講演時に話者がどれだけリラックスしているかの指標とみなすことができるから、要因3)、4)と強く関連しているものと考えられる。

6 語の融合

「デウ」の融合形「ジャ」の生起頻度は、「デ」の品詞(格助詞か助動詞か)によって大幅に異なる。格助詞が平均 1.7%であるのに対し、助動詞は 42.6%に達する。しかし、このような頻度差にも関わらず、いずれの品詞にも要因3)、4)との相関が明瞭に観察されることから、形態論レベルの変異は、話者にははっきりと意識されていることがわかる [2]。

融合現象は、表1の要因以外に話者の年齢(生年代)も影響する。回帰樹(regression tree)という統計手法を利用して、融合形の生起率を予測してみた。

助動詞にだけ注目すると、生起率が最低となるのは、1950年代以降に生まれた話者による、相対的に発話スタイルが高く(評定値 4 以上)、自発性も低い(評定値 2 以下)学会講演で、平均 1.7%である。一方、融合率が最も高いのは、自発性が高く(3 以上)、発話スタイルが低い(2 以下)模擬講演で、平均値は 77.1%に達する(次節末も参照)。

7 助詞の撥音化

助詞の「ノ」には格助詞と準体助詞があり、これらはいずれも「ン」と発音される。撥音化の生起頻度は融合現象以上に品詞によって変動しており、格助詞の平均が 0.5%であるのに対し、準体助詞は 51%に達する。また、準体助詞に関しては、後続語(短単位)の品詞による影響も顕著であり、助動詞が後続する場合の撥音化率が 78.5%であるのに対し、助動詞以外の品詞の場合は 0.5%にとどまっている。ここで助動詞というのは断定の助動詞(ダないしデス)であり、助詞「ノ」の撥音化はほとんどがこの環境で生起している[6]。

撥音化についても回帰樹による予測を試みた。助動詞はまず講演タイプによって下位区分され、その後後続語の品詞によって下位区分される。一方、格助詞は、まず発話速度によって3グループに下位区分され、発話速度の遅い順に、性別、発話スタイル、講演種別によって下位区分される。この分類によって平均撥音化率は、準体助詞では最低 0.3%から最高 83%まで、また格助詞では同じく 0.1%から 2.2%まで変動する。回帰樹による分類結果は、語の融合や助詞の撥音化についても、或る程度まで定量的に正確な予測が可能であることを示唆している。

8 アクセント句末音調

ここまで、表 1 にとりあげた変異現象を個別に検討した。最後にCSJのコアに付与された韻律ラベルを利用して、アクセント句末音調を検討する。

「これから | 東京方言の | 複合語アクセントについて | 研究発表を | いたします」という発話を例にとれば、縦線で区切られた区間が韻律の上でのまとまりをなすことができる。こ

表1: 外的要因の変異現象の平均値への影響(一元配置分散分析結果)

変異現象	N	要因1 講演種別	要因2 発話速度	要因3 スタイル	要因4 自発性	要因5 (笑)の有無	要因6 性別
母音無声化	300,018	****	****	****	****	N.S.	****
長母音短呼	47,886	****	N.S.	****	****	****	****
/zya/1(格助詞)	1,730	****	N.S.	****	***	***	N.S.
/zya/2(助動詞)	1,707	****	N.S.	****	****	****	N.S.
/no/1(格助詞)	32,317	**	N.S.	****	***	N.S.	****
/no/2(準体助詞)	16,900	****	****	****	****	****	****

有意水準 **** < .0001; *** < .001; ** < .01; N.S. >= .01

れがアクセント句であり、その句末音節には多彩な句末音調(イントネーション)が観察される。句末音調は若者言葉や職域方言の特徴としてしばしば言及されることがある。

ここでは、生起頻度が高い上昇音調(X-JToBIのラベルではL%H%)と上昇下降音調(L%HL%)の二種類に注目する。これら二種類の句末音調の生起頻度(すなわちアクセント句の総数に占める各音調の生起数)は種々の社会的要因との相関のあり方に関して全く逆のふるまいを示すことが明らかになった[3]。

すなわち、図1が示すように、上昇音調の生起率が発話スタイルの上昇とともに増加し、自発性の上昇とともに減少するのに対して、上昇下降音調の生起率は発話スタイルの上昇とともに減少し、自発性の上昇とともに増加する。

講演種別との関係では、上昇音調が学会講演に多く模擬講演に少ないのに対して、上昇下降音調は模擬講演に多く学会講演に少ない。さらに、話者の性別に関しては、上昇音調は男性、上昇下降音調は女性話者によってより多く用いられる傾向がある。

9 まとめと展望

上に紹介した予備的解析の結果から分かるように、CSJには種々の変異現象が記録されている。総じてCSJは言語変異研究の資源として非常に有効であると判断できる。今回は報告に到らなかったが、いわゆる可能形のゆれやアクセントのゆれなど、一般の耳目を集めやすい現象に関する研究も可能である。

今後は、これらの変異現象も含めてCSJ全体を解析し、現代日本語における言語変異の全体像を把握し、さらに個々の現象間に潜む相関関係の解明を進めたい。その研究成果は、社会言語学の知見を深めるにとどまらず、アクセントないし発音辞典類や音声認識用発音辞書の改良に貢献するだろうと考えている。

参考文献

- [1] 前川: “音声分野における統計的研究法—音声の変異をめぐって—”, 日本語学, 20, 5, pp.144-156 (2001).
- [2] 前川: “『日本語話し言葉コーパス』を用いた言語変異研究”, 音声研究, 6, 3, pp.48-59 (2002).
- [3] Maekawa, K. et al.: “Use of a large-scale spontaneous speech corpus in the study of linguistic variation”, *Proc. 15th International Congress of Phonetic Sciences (ICPhS 2003)*, Barcelona, pp.643-646 (2003).
- [4] Maekawa, K. & H. Kikuchi: “Corpus-based analysis of vowel devoicing in spontaneous Japanese—An interim report—”, Unpublished manuscript.
- [5] 前川: “話し言葉における長母音の短呼”, 国語学会 2002 年度春季大会要旨集, pp.43-50 (2002).

- [6] 小磯 他: “話し言葉における助詞の撥音化現象の実態”, 第 10 回社会言語科学会研究大会予稿集, pp.215-220 (2002).

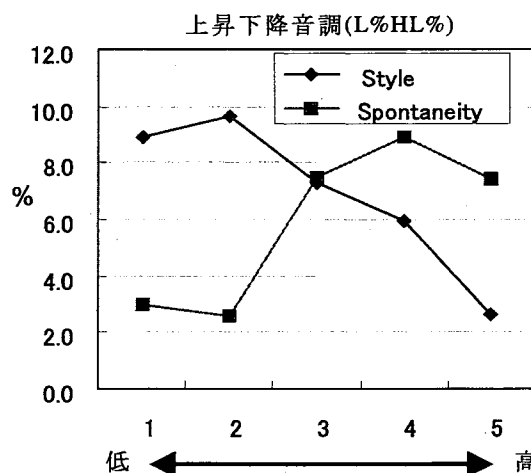
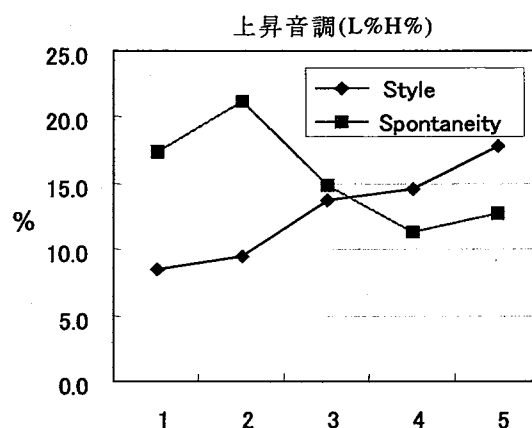


図1: アクセント句末音調の生起率と発話スタイル及び自発性に関する印象評定値との関係