

音声言語コーパスにおける speaking style の多様性 —転記テキストに着目して—

沈 睿 (早稲田大学人間科学学術院)

菊池 英明 (早稲田大学人間科学学術院)

1 はじめに

郡(郡 2006)によれば, speaking style(口調, 発話様式)は個別言語の特徴記述に必要不可欠である。故に, 一つの言語を習得する際, その言語の speaking style の習得も重要だと思われる。しかし, 現在の外国語教育の分野では, 文法や語彙などの内容が優先されており, speaking style に関する内容が十分に取り込まれていない。一方, コンピュータ技術の進歩に伴い, 大規模コーパスや CALL システムなどの外国語教育への応用に関する研究が盛んに行われているため, 大量の話し言葉を収録した音声言語コーパスを speaking style の習得にも活用できるのではないかと考えられる。その際, まず音声言語コーパスの speaking style の判別ができるようにしなければならない。

近年, speaking style は音声研究の分野で注目されてきたが, speaking style の定義に関しては, 半世紀以来, 明確な定義が提案されていない。当初, Uhlmann は speaking style を「特定利用のための口頭あるいは書かれた表現」として定義した(Uhlmann 1964)。この定義に基づけば, speaking style は話し言葉と書き言葉の両方で重要な指標と言える。Eskenazi はデータに基づいて speaking style を定義することを提案した(Eskanazi 1993)。本稿では主に Eskenazi の提案を用いるため, 後述する。

音声研究において, speaking style は研究目的に応じて定義され, 様々な分野で研究されている。音声の物理特性や音響特徴に焦点を絞った研究が多いが, 発話内容や言語的特徴に注目する研究はまだ少ない。本研究は外国語教育への応用を想定しているため, 話し言葉の音響や韻律の側面から改善をしやすい発話内容や言語的特徴が記述される転記テキストに着目する。なお, 従来の自然言語処理の分野で, 書き言葉に対する文体・ジャンルの判別や著者推定などの研究は多く行われている。品詞率, 語種率と形態素パターンを特徴量とした方法が有効であることが実証されたため, 本研究もそれらの特徴量を speaking style の判別に用いることを想定し, 話し言葉の転記テキストに着目して speaking style の定量化とモデルの構築を試みる。

2 手法

本稿では, Eskenazi が提案した speaking style を表現する 3 尺度を用いる。Eskenazi はデータに基づいて (data-driven) speaking style を定義することを提案した(Eskanazi 1993)。Eskenazi は, 人間のコミュニケーションは, あるチャネルを通じて, 情報(message)が話し手から聞き手へ伝達することであり, speaking style を定義する際, この情報の伝達過程を考慮することが必要であると主張した。Eskenazi によれば, speaking style は「明瞭さ」(Intelligibility-oriented, 以降 I とする), 「親しさ」(Familiarity, 以降 F とする), 「社会階層」(Social strata, 以降 C とする)の 3 尺度で定義できる。Eskenazi によれば, 「明瞭さ」は話し手の発話内容の明瞭さの度合いであり, 情報の読み取りやすさ・伝達内容の理解し

やすさや、読み取りの困難さ・伝達内容の理解の困難さを示す。発話者が意図的に発話の明瞭さをコントロールしている場合も含む。「親しさ」は話し手と聞き手との親しさにより変化する表現様式の度合いであり、家族同士の親しい会話や、お互いの言語や文化を全く知らない外国人同士の親しくない会話などにあらわれる発話様式を示す。「社会階層」は発話者の発話内容の教養の度合いであり、口語的な、碎けた、下流的な表現（社会階層が低い）や、洗練された、上流的な表現（社会階層が高い）を示す。話し手と聞き手の背景や会話の文脈によって変化する場合もある。

本稿の流れについて説明する。まず、speaking style の異なる様々な音声言語コーパスから 6 コーパスをバランスよく選出する。続いて 6 コーパス（カテゴリ）から 10 サンプルずつ音声の転記テキストを選出する。なお、本研究では、サンプルごとの speaking style の集積をサンプルが属するコーパスの speaking style とみなす。続いて抽出したテキストに対し、上述の speaking style の 3 尺度を用いて評定実験を行う。

3 評定実験

本章では、評定実験の詳細について述べる。

3.1 評定者

本評定は、大学生男女 22 名の評定者による。

3.2 刺激

本実験の刺激に音声言語コーパス内の転記テキストを使用する。

3.2.1 音声言語コーパス

なるべく多様な speaking style を含む音声言語コーパスを使用するために、実験で使用する音声の転記テキストを以下の 6 種類の音声コーパス（カテゴリ）から選出した。

(1) 日本語話し言葉コーパス(前川ら 2000)-講演 (CSJ1 と呼ぶ)

日本語話し言葉コーパス (the Corpus of Spontaneous Japanese, CSJ) は、日本語の自発音声を大量に集めて多くの研究用情報を付加した、質・量ともに世界最高水準の話し言葉研究用のデータベースである。本研究では、CSJ に収録された speaking style の中でも、特に学会発表及び模擬講演発表（ひとつのテーマに関してのモノローグ）をまとめて扱う。

(2) 日本語話し言葉コーパス-インタビュー (CSJ2 と呼ぶ)

(1) と同じく CSJ から選出した、インタビュー形式の対話である。講演音声と対話音声の speaking style は大いに違うと思われる所以、今回の実験目的を考慮し、別のカテゴリとした。なお、インタビュアーとインタビュイーとの両方のチャネルの音声を使用した。

(3) 千葉大地図課題対話コーパス(堀内ら 1999) (MAPTASK と呼ぶ)

地図を用いて課題を遂行するための対話コーパスである。

(4) 新入生対話コーパス (FDC と呼ぶ)

大学の研究室に所属して一ヶ月の大学生同士の間の自由対話を収録したコーパスである。本コーパスは初対面の二者の対話音声が、時間経過および二者の親密性の向上とともにどのように変化するのかを調べることを目的としている。

(5)車載環境における質問応答の対話コーパス(宮澤ら 2010) (AUTOと呼ぶ)

本コーパスは、模擬車内環境でドライビングゲームをプレイしたドライバー役被験者と、同乗してナビゲーションを行ったナビゲーター役被験者に対して、走行実験終了後に、実験中の動画を見せながら感想やナビゲーションの的確さをインタビューした際の対話音声である。インタビューはドライバー、ナビゲーターそれぞれに対して実施した。なお、ドライバーとナビゲーターとの両方のチャネルの音声を使用した。

(6)旅行についての対話コーパス(岩野ら 1997) (TRAVELと呼ぶ)

旅行の計画について、面識のある二人の研究室メンバーの間で交わされた自由対話を収録したコーパスである。

3.2.2 転記テキストの加工

上述の 6 種類のカテゴリから 10 個ずつ合計 60 個の音声サンプルを無作為で選出する。Speaking style の最も安定する部分を抽出するため、上記の各音声に付随する転記テキスト中より約 300 字のテキストを切り出す。なるべく発話の内容の影響を避け、発話様式や口調だけで評定してもらうため、テキストの名詞（代名詞は除く）の部分を全部「○○」に自動変換した。

3.3 評定方法

本評定は SD 法を用いる。一つのテキストを読んだ後、3 尺度のそれぞれについて 7 段階で評定してもらう。「明瞭さ」に関して、不明瞭の場合 1、明瞭の場合 7、「親しさ」に関して、親しい場合 1、親しくない場合 7、「社会階層」に関して、低い場合 1、高い場合 7 とする。評定はインターネット上のアンケートサイトを介して行う。評定の前に、尺度についての詳細説明をよく読むように指示した。

4 結果と考察

3 章で述べた評定実験によって得られた結果を本章で述べる。

まず、3 尺度の相関係数を求めた。明瞭さ I と親しさ F の相関係数が .26、明瞭さ I と社会階層 C が .48、親しさ F と社会階層 C が .56 である。今回の評定実験の結果によると、3 尺度が必ずしも独立ではないことが分かった。

さらに、22 名の評定者の評定結果の平均を各サンプルの得点として図 1(X 軸が明瞭さ I、Y 軸が親しさ F、Z 軸は社会階層 C) に示したように 3 尺度空間上にプロットした結果、刺激テキストがコーパスの収録条件や収録環境などの特徴により分かれる仮説とほぼ一致だと分かった。例えば、図 1 の赤い点が旅行についての対話コーパス (TRAVEL) のサンプルを示し、3 尺度空間上に不明瞭・親しい・口語的な位置に集まっている。TRAVEL コーパス内の音声が研究室メンバー同士の間の旅行計画についての自由対話なので、予想される分布と一致すると言える。

5 おわりに

本稿では、従来の書き言葉に対する文体やジャンル判別の手法を話し言葉における speaking style の定量化とモデル化に用いる前に、speaking style を「明瞭さ」 (Intelligibility-oriented), 「親しさ」 (Familiarity), 「社会階層」 (Social strata) の 3 尺度を用いて音声言語コーパスからサンプリングした音声の転記テキストに対する評定実験を行い、音声コーパスの

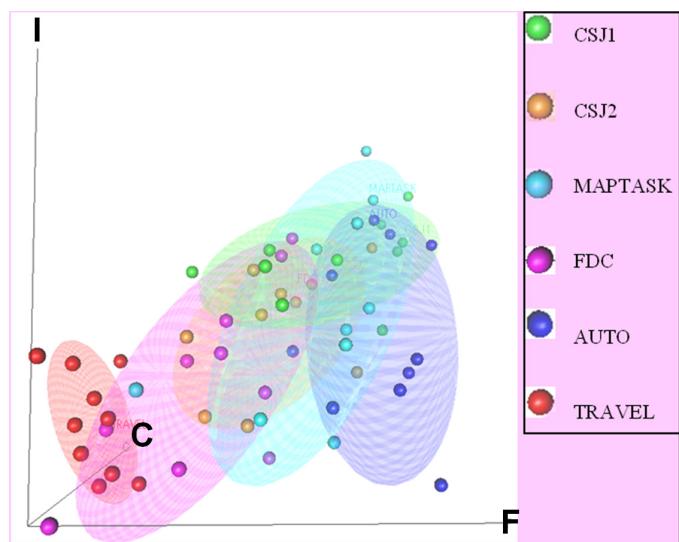


図1 刺激テキストの3尺度空間上の分布

Figure 1 the distribution of text stimuli on
the space of 3-scales of speaking style

speaking style の分布を考察した。その結果、刺激テキストがコーパスの収録条件や収録環境などの特徴により分かれる仮説通り 3 尺度空間上で分かれることが分かった。

今後の方針として、従来の自然言語処理の手法を用い、speaking style の 3 尺度それぞれの判別モデルを構築し、学習者や教師に speaking style の自動判別サービスを提供しようと考えている。

参考文献

- Eskenazi, M. (1993). “Trends in Speaking style Research.” Keynote speech, Proceedings Eurospeech’ 93, Berlin.
- Uhlmann, A. M. (1964). “Meyers Neues Lexikon.” VEB Bibliographisches Institut Leipzig, ausgabe in acht bänden edition.
- 岩野裕利, 杉田洋介, 松永美穂, 白井克彦(1997). “対面および非対面における対話の違い-頭の振りの役割分析.” 音声言語情報処理研究報告, Vol. 15-29, pp. 105-112.
- 郡史郎(2006). “日本語の「口調」にはどんな種類があるか.” 音声研究, Vol. 10-3, pp. 52-68.
- 堀内靖雄, 中野有紀子, 小磯花絵, 石崎雅人, 鈴木浩之, 岡田美智男, 仲真紀子, 土屋俊, 市川熹(1999). “日本語地図課題対話コーパスの設計と特徴.” 人工知能学会誌, Vol. 14-2, pp. 261-272.
- 前川喜久雄, 籠宮隆之, 小磯花絵, 小椋秀樹, 菊池英明(2000). “日本語話し言葉コーパスの設計.” 音声研究, Vol. 4-2, pp. 51-61.
- 宮澤幸希, 影谷卓也, 沈睿, 菊池英明, 小川義人, 端千尋, 太田克己, 保泉秀明, 三田村健(2010). “自動車運転環境下におけるユーザーの受諾行動を促すシステム提案の検討.” 人工知能学会誌, Vol. 25-6, pp. 723-732.