

博士学位論文審査報告書

大学名 早稲田大学
研究科名 人間科学研究科
申請者氏名 沈 睿
学位の種類 博士 (人間科学)
論文題目 音声言語コーパスにおけるspeaking styleの自動推定
Automatic Estimation of Speaking Style in Speech Corpora
論文審査員 主査 早稲田大学教授 菊池英明 博士 (情報科学) (早稲田大学) (知覚情報処理)
副査 早稲田大学教授 森本豊富 Ph. D. (UCLA) (文化人類学)
副査 早稲田大学教授 松居辰則 博士 (理学) (早稲田大学) (知能情報学、感性情報学)
副査 筑波大学名誉教授 板橋秀一 工学博士 (東北大学) (知覚情報処理)

計算機技術の進歩に伴い大規模言語データの蓄積と処理が容易になり、音声言語コーパスの構築と活用が盛んになされている。音声科学の研究分野ではデータ重視の傾向が増し、信頼性や再現可能性を示すために音声言語コーパスが活用されている。音声工学・言語工学の分野では、統計的なモデルによる手法が主流となっており、コーパスは欠かせない手段となっている。多くの研究機関で構築される音声言語コーパスの流通性を高めるために、海外では、アメリカの Linguistic Data Consortium (LDC) やヨーロッパの European Language Resources Association (ELRA) など、言語データの集積と配布を行う機関が設けられている。これらの機関では、様々な研究分野からの利用者に所望のコーパスを探しやすくさせるために検索サービスが提供されている。日本国内においても、国立情報学研究所音声資源コンソーシアム (NII-SRC) や言語資源協会 (GSK) などの音声言語コーパスの整備・配布を行う機関が組織され、コーパスの属性情報に基づいた検索サービスが開発・提供されている。申請者はコーパスの属性検索と可視化検索を同時に提供することで、コーパスに関する知識の多少に関わらず所望のコーパスが検索可能になることを示した。コーパス数の増加に伴ってこうした検索機能の充実への期待が高まっている。

利用者がコーパスを選定する際、目的に合致しているかどうかを見極めるために speaking style は最も重要な情報の一つといえる。Speaking style は、「親しい友人との日常会話」「議会でのあらたまったスピーチ」など、発話時の様式の違いを表す概念であり、近年様々な研究分野において注目が集まっている。日本語は speaking style の変化がとりわけ多いと言われ、日本語の音声言語コーパス利用者は、目的に合致した speaking style のコーパスを注意深く選定する必要がある。しかしながら、speaking style の標準的な分類方法は定まっておらず、また大量の音声言語データを含むコーパスに対してどのように表現すべきかも明らかでない。そのため、現状では、前述の機関では対話や独話などの種別情報が一部

提供されているに過ぎない。そこで、speaking style に関心を持つ利用者に所望の音声言語コーパスを探しやすくさせるため、音声言語コーパスにおける部分的単位ごとの speaking style の自動推定を可能にし、コーパスの属性情報としてより詳細な speaking style の集積を提供することを目指して本研究は行われた。

本論文は全 7 章から構成される。

第 1 章では、上述の音声言語コーパスおよびその検索機能に関する先行研究を概観するとともに、著者が構築した日本語音声言語コーパス検索サービスの概説が加えられている。

第 2 章では、音声言語コーパス検索における speaking style の必要性を述べ、とりわけ言語教育における音声言語コーパス利用の状況と、そこでの speaking style 情報の有用性を整理している。

第 3 章では、speaking style に関する先行研究を概観したうえで、Eskenazi の 3 尺度(明瞭性、親密性、社会階層)を解説している。

第 4 章では、speaking style 自動推定システムを構築する手法の概要が説明されている。具体的には、まず音声言語コーパスの speaking style 評定を行い(第 5 章で詳述)、主に形態論的な特徴に注目して説明変数とし、speaking style 評定値を目的変数とした重回帰モデルを構築している(第 6 章に詳述)。

本章ではさらに、音声言語コーパス検索にとって必要な speaking style 定義の条件を明確にしたうえで、条件を満たす尺度として 3 章で紹介された Eskenazi の 3 尺度を採用することを述べている。Eskenazi の 3 尺度とは、数多くの speaking style 関連の研究論文を概観したうえで、言語の違いやレジスタ(言語使用域)に依存しない普遍的な尺度をメタ的に構築したものである。他の多くの尺度が特定の言語やレジスタに対して提案された範囲限定的なものであり、カテゴリカルであるのに対して、この尺度は網羅性が高く連続的に数値で表現可能である点で本研究の趣旨に合致している。

第 5 章では、speaking style 自動推定の実現に向けた第一段階として行われた音声言語コーパスの speaking style 評定について述べている。具体的には、22 名の被験者に音声言語コーパスの転記テキストの抜粋を刺激として提示して、Eskenazi の 3 尺度それぞれを 7 段階で評定させる実験が行われた。その際、扱うデータの網羅性を考慮して、種類の異なる 6 つの音声言語コーパスが選定され、それぞれ 10 データが無作為に抽出された。データから刺激を作成する際には、Speaking style の安定する部分として、転記テキスト中部約 300 字が抽出され、発話の内容の影響を避けて speaking style だけで評定されるようテキストの名詞(代名詞は除く)の部分が全て「○○」に自動変換された。その他、コーパスやデータによって異なる表記方法はできるだけ統一された。評定実験を終えた後、6 コーパスの speaking style 評定値が実際にどのように分布してコーパスの特性と合致しているかを検証している。その結果、speaking style が多様になることを意図して選定された 6 コーパスについて、予想されたようにコーパスの特性を表すように評定値が分布しており多様なデータを確保できたと言える。さらに評定値の安定性を確認するために、3 尺度の級内相関係数を算出している。ICC(2, 1)(評定者間の信頼性)はそれぞれ 0.11、0.53 と 0.35 であり、ICC(2, k)(評定平均の信頼性)は 0.72、0.96 と 0.92 であった。ICC(2, 1)において親密性と社会階層は許容範囲内の信頼性といえ、また ICC(2, k)によれば 3 尺度のいずれも評定値の平均の信頼性は高く、モデル構築に用いて良いと言える。

第6章では、第5章の実験によって得られた speaking style 評定値と刺激に用いたテキストに現れた言語特徴量によってモデルを構築して評価する過程が述べられている。推定に用いる特徴量として、申請者らが別途行った研究で導かれた形態論的特徴が導入されている。この研究では、任意のテキストに対して形態論レベルの変換を行うことによって特定のキャラクター像を想起させるテキストが生成できることを示している。その過程で得た43種類の形態論的特徴と、他の先行研究に基づいて得た品詞・語種率を説明変数、speaking style 評定値を目的変数として、重回帰分析により推定モデルが構築された。その際、ステップワイズ変数選択(変数減増法)の手法で最適なモデルを求めた。求めたモデルの有意性を検定するために、「全ての偏回帰係数がゼロである」という帰無仮説を立てたF検定を行った。その結果、3尺度共に有意水準1%でモデルの有意性が証明された。さらに、モデルの信頼性を確認するため、交差検定(leave-1-out)によりモデルの決定係数を求めた。サンプル数の少なさを考慮して自由度調整済決定係数を求めたところ、同一コーパスの他データが学習データとして存在する場合、明瞭性は0.37、親密性は0.81、社会階層は0.66と、明瞭性を除いて高い精度で推定可能であることが示された。重回帰モデルにおける偏回帰係数の絶対値が大きい特徴は、いずれも従来の先行研究で speaking style に関係が深いとされていたものであり、モデルの学習結果が妥当であることが確認されている。

本研究で提案された speaking style の自動推定手法によって、コーパスの部分的単位ごとの speaking style を推定した結果はコーパス全体の speaking style の判断材料となる。音声科学や音声工学などの多分野で需要が増している音声言語コーパス検索において、コーパス選定に有用な speaking style の情報を効率よく高精度で推定して提供できる本技術の意義はきわめて高い。

なお、本論文(一部を含む)が掲載された主な学術論文は以下のとおりである。

- [1] 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: 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, 人工知能学会論文誌, Vol.25, No.6, pp.723-732 (2010)
- [2] 沈 睿, 菊池 英明, 太田 克己, 三田村 健: 音声生成を前提としたテキストレベルでのキャラクタ付与, 情報処理学会論文誌, Vol.53, No.4, pp.1269-1276 (2012)
- [3] R. SHEN, K. SUNAOKA: Construction and Application of a Text Corpus of Newspaper Articles about Disasters in Chinese Education, The Journal of Modernization of Chinese Language Education, Vol.1, No.2, pp.43-52 (2012)
- [4] 沈 睿, 菊池 英明: 音声言語コーパスにおける speaking style の自動推定—転記テキストに着目して—. 言語処理学会論文誌, Vol.21, No.3 (印刷中)

以上のことに鑑みて、本審査委員会は本論文が博士(人間科学)の学位を授与するに十分値するものと認める。

以 上