

沈 睿†・菊池 英明††

近年、計算機技術の進歩に伴って大規模言語データの蓄積と処理が容易となり、音声言語コーパスの構築と実用化の研究が盛んに行われている。我々は、speaking style に関心を持つ利用者に音声言語コーパスを探しやすくさせるために、音声言語コーパスの speaking style の自動推定を目指している。本研究では、1993 年に Eskenazi が提唱した speaking style の 3 尺度を導入し、従来の文体・ジャンルの判別や著者推定などの自然言語処理の分野で用いられた言語の形態論的特徴を手がかりとし、音声に付随する書き起こしテキスト(本論文では転記テキストと呼ぶ)に着目した speaking style 推定モデルの構築を試みた。具体的な手続きとしては、はじめに様々な音声言語コーパスから音声に付随する転記テキストを無作為に抽出する。次にこれらの転記テキストを刺激として用い、3 尺度の speaking style の評定実験を行う。そして、評定結果を目的変数、転記テキストの品詞・語種率と形態素パタンを説明変数とし、重回帰分析により 3 尺度それぞれの回帰モデルを求める。交差検定を行った結果、本研究の提案手法によって 3 尺度の内 2 尺度の speaking style 評定値を高い精度で推定できることを確認した。

キーワード:発話スタイル、音声言語コーパス、転記テキスト、推定、回帰モデル

Automatic Estimation of Speaking Style in Speech Corpora Focusing on Speech Transcriptions

RAYMOND SHEN[†] and HIDEAKI KIKUCHI^{††}

Recent developments in computer technology have allowed the construction and wide-spread application of large-scale speech corpora. To enable users of speech corpora to easier data retrieval, we attempt to characterise the speaking style of speakers recorded in the corpora. We first introduce the three scales for measuring speaking style which were proposed by Eskenazi in 1993. We then use morphological features extracted from speech transcriptions that have proven effective in discriminating between styles and identifying authors in the field of natural language processing to construct an estimation model of speaking style. More specifically, we randomly choose transcriptions from various speech corpora as text stimuli with which to conduct a rating experiment on speaking style perception. Then, using the features extracted from these stimuli and rating results, we construct an estimation model of speaking style, using a multi-regression analysis. After cross-validation (leave-1-out), the results show that

[†]早稲田大学人間科学研究科, Graduate School of Human Sciences, Waseda University

^{††} 早稲田大学人間科学学術院, Faculty of Human Sciences, Waseda University

among the three scales of speaking style, the ratings of two scales can be estimated with high accuracy, which proves the effectiveness of our method in the estimation of speaking style.

Key Words: speaking style, speech corpora, transcriptions, estimation, regression model

1 はじめに

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. 海外では、アメリカの Linguistic Data Consortium (LDC) とヨーロッパの European Language Resources Association (ELRA) が言語データの集積と配布 を行う機関として挙げられる.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可視化検索サービスが開発・ 提供されている (Yamakawa, Kikuchi, Matsui, and Itahashi 2009, 菊池, 沈, 山川, 板橋, 松井 2009). 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を検索可能にできることも示されている (Shen and Kikuchi 2011). 検索に用いられるコーパスの属性情報として,収録目的や話者数などがあるが,speaking style も有効な情報と考えられる.郡は speaking style と類似の概念である「発話スタイル」が個別 言語の記述とともに言語研究として重要な課題であると指摘している (郡 2006).Jorden らに よれば、どの言語にもスタイルの多様性があるが、日本語にはスタイルの変化がとりわけ多い (Jorden and Noda 1987). しかしながら、現状では、前述の機関では対話や独話などの種別情報 が一部で提供されているに過ぎない。また、同一のコーパスにおいても話者や収録条件によっ て異なる speaking style が現れている可能性もある.そこで,本研究では 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す.

Speaking style の自動推定を実現するためには、まず speaking style の定義を明確にする必要がある。Joos は発話のカジュアルさで speaking style を分類し (Joos 1968)、Labov は speaking style が話者の発話に払う注意の度合いとともに変わると示唆した (Labov 1972)。Biber は言語的特徴量を用いて因子分析を行い、6 因子にまとめた上で、その6 因子を用いて異なるレジスタのテキストの特徴を評価した (Biber 1988)。Delgado らはアナウンサーの新聞報道や、教師の教室内での発話など、特定の職業による発話を "professional speech" として提案し (Delgado and Freitas 1991)、Cid らは発話内容が書かれたスクリプトの有無を speaking style のひとつの指標

にした (Cid and Corugedo 1991). Abe らは様々な韻律パラメータとフォルマント周波数を制御 することにより小説、広告文と百科事典の段落の3種類のspeaking style を合成した(Abe and Mizuno 1994). Eskenazi は様々な speaking style 研究の考察からメタ的に speaking style の全体 像を網羅した 3 尺度を提案した (Eskenazi 1993). Eskenazi は、人間のコミュニケーションは、 あるチャンネルを通じて、メッセージが話し手から聞き手へ伝達されることであり、speaking style を定義する際,このメッセージの伝達過程を考慮することが必要であると主張した.その 上で、「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 speaking style を定義した. 「明瞭さ」は 話し手の発話内容の明瞭さの度合いであり、 メッセージの読み取りやすさ・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、又これは、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んでいる.「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現れる speaking style 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、砕けた、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)を示している、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.

ここで、本研究が目指すコーパス検索サービスにとって有用な speaking style 尺度の条件を 整理し,Eskenazi の尺度を採用する理由を述べる.まず,一つ目の条件として,幅広い範囲の データを扱える必要がある. 音声言語コーパスは、朗読、雑談、講演などの様々な形態の談話 を含み、それらは話者ごと、話題ごとなどの様々な単位の部分的単位により構成される、限ら れた種類の形態のデータからボトムアップに構築された尺度では、一部の speaking style がカ バーできていない恐れがある.Eskenazi の尺度は,様々な speaking style 研究の考察からメタ的 に構築されたものであり、幅広い範囲のデータを扱える点で本研究の目的に適している。デー タに基づいてボトムアップに構築された他の尺度 (例えば (Biber 1988) など) の方が信頼性の 点では高いと言えるが、現段階では網羅性を重視する、次に、二つ目の条件として、上述した 目的から、コーパスの部分的単位ごとに付与できる必要がある。新聞記事、議事録、講演など のジャンルごとに speaking style のカテゴリを設定する方法では、一つの談話内での speaking style の異なり・変動を積極的に表現することが困難である。一方、Eskenazi の尺度は必ずしも 大きな単位に対象を限定しておらず,様々な単位を対象にした多くの先行研究をカバーするよ うに構築されているため、この条件を満たす。最後に三つ目の条件として、日本語にも有効で あることが求められる. (郡 2006) や (Jorden and Noda 1987) から, speaking style の種類は言 語ごとに異なると言え、特定の言語の資料に基づいてボトムアップに構築された尺度では、他 の言語にそのまま適用できない恐れがある. Eskenazi の尺度はコミュニケーションモデルに基 づいて特定の言語に依存することなく構築されたものであるため、本研究で対象とする日本語

にも他の言語と同様に適用して良いものと考える.

したがって本研究では、Eskenaziの3尺度を用いて音声言語コーパスの部分的単位ごとの speaking style 自動推定を行い、推定結果の集積をコーパスの属性情報として提供することを目指す.これによって、推定された3尺度の値を用いて、例えばコーパス内の部分的単位の speaking style 推定結果を散布図で可視化したり、所定の明瞭さ、親しさ、社会階層のデータを多く含む コーパスを検索するなどの応用を可能にする.

以降、2章では、speaking style 自動推定の提案手法について述べる。Speaking style の推定に用いる学習データを収集するための評定実験については3章で説明する。4章では評定実験結果の分析、speaking style の自動推定をするための回帰モデルの構築および考察を述べる。最後の5章ではまとめおよび今後の方向性と可能性の検討を行う。

2 Speaking style の自動推定手法

Speaking style の自動推定に寄与する要素として、イントネーションや時間構造などの音声的特徴と形態素や統語的構造などの言語的特徴との両方が考えられる。Eskenazi は、speaking style の先行研究のレビューに基づいて音声的特徴と言語的特徴の双方の重要性を述べている (Eskenazi 1993)。郡の調査 (郡 2006) によって示された日本語の口調(speaking style と類似の概念)には、「ですます口調」や「漢文口調」など主に書きことばとしてのスタイルや言語形式によって特徴づけられるものが数多くあげられている。これらから、本研究では speaking style を扱ううえでまず言語的特徴に焦点を当てる。

一方,自然言語処理の分野においては、言語的特徴を手がかりとしたテキストに対する文体・ジャンルの判別や著者推定などの研究が多く行われ、比較的精度の高い成果が得られている.小磯らは現代日本語書き言葉均衡コーパスと日本語話し言葉コーパスにおける7つのサブコーパス(白書,新聞,小説,Yahoo!知恵袋,国会議事録,学会講演,模擬講演)に対して、漢語率、名詞率、接続詞率、副詞率、形容詞率、機能語率を手がかりとする判別分析を行い、約80%の精度でサブコーパスの分類が可能であることを示した(小磯,小木曽,小椋,宮内2009).小山らは形態素出現パタンを手がかりとし、学会における研究発表抄録データの類似性を評価し、いくつかの異なる学会間の類似度をほぼ再現する距離尺度を構成できることを示唆した(小山、竹内2008). Mairesse らは言語的手がかりをパーソナリティの自動認識に用いることを試み、複数の機械学習の手法によって精度の比較と有効な特徴量の検討を行った(Mairesse, Walker, Mehl, and Moore 2007). これらの研究においては、品詞・語種率と形態素パタンなどの形態論的特徴を特徴量とした方法が有効であることがわかった.

上述の理由と先行研究を踏まえた上で、我々はまず音声言語データの形態論的特徴から着手し、従来のテキストの文体・ジャンル判別の手法を用い、音声に付随する書き起こしテキスト

(本論文では転記テキストと呼ぶ) に着目した speaking style 推定モデルを構築することにより, speaking style の自動推定を試みる.

音声的特徴については、上述したように speaking style の推定に有用な特徴であり、今後の導入を検討しているが本稿では扱わない。前述した音声言語コーパスの検索サービスにおいて、形態論的特徴のみに基づく speaking style 推定結果を提供することも、例えば形態論的側面に焦点を当てた日本語教育に用いる資料や、話し言葉における言語情報の話者性変換技術 (水上、Neubig、Sakti、戸田、中村 2013) などの学習データを求めるような需要に応えることが可能と考える。

Speaking style 推定モデルの構築の具体的な流れを図1を用いて説明する.

まず、speaking style の異なる様々な音声言語コーパスから音声の転記テキストを選出し(3.2.1 節詳述)、speaking style の最も安定する部分と思われる最中部の約 300 字程度のテキストを抽出する(3.2.2 節). 続いて抽出したテキストに対し、Eskenazi の speaking style の 3 尺度を用いて speaking style 推定モデルの構築に用いる学習データを収集するための評定を行う(3.3 節). 同時に、UniDic を辞書として用いた MeCab で形態素解析を行い、品詞・語種率、形態素パタンを特徴量として抽出する(4.1 節). 評定実験で得られた評定結果の平均を求め、3 尺度の学習データとする. 最後に重回帰分析により 3 尺度それぞれの回帰モデルを求め、speaking style 推定モデルとする(4.2 節).

構築した speaking style 推定モデルを用いて、任意の転記テキストに対して、 speaking style を自動推定することが可能になる.

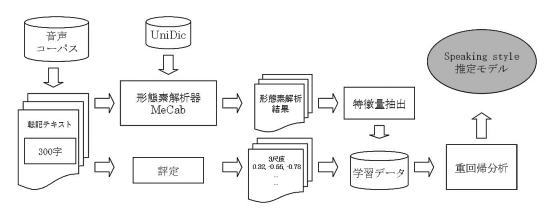


図 1 Speaking style 推定モデルの構築の流れ

3 評定実験

本章では、2章で述べた評定実験の詳細について述べる.

3.1 評定者

本評定は、情報科学を専門とした大学生男女 22 名(その内男性は 15 名)の評定者による。全ての評定者は本研究に関わっていない。

3.2 刺激

刺激には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

Speaking style の多様さと実験のコストの両方を考慮した上で、本評定実験で使用する音声の 転記テキストを、以下の6種類の音声言語コーパス(カテゴリ)から収録条件や話者の役割な どによっての違いを区別せず、それぞれ10サンプルずつ無作為に選出した。

I. 日本語話し言葉コーパス (前川, 籠宮, 小磯, 小椋, 菊池 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 に収録された多様な speaking style の中でも、特に学会講演及び模擬講演を対象とする。

II. 日本語話し言葉コーパス-インタビュー (CSJ2 と呼ぶ)

I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 は著しく異なると思われるので、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の発話を区別しないことにした。このコーパス(カテゴリ)は対面の自由対話である。

III. 新入生対話コーパス (中里, 大城, 菊池 2013) (FDC と呼ぶ)

大学の研究室に所属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。本コーパスは低親密度の二人の対話音声が、時間経過および二人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。このコーパスは大学生同士の間の自由対話である。

IV. 車載環境における質問応答の対話コーパス (宮澤, 影谷, 沈, 菊池, 小川, 端, 太田, 保泉, 三田村 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。質

問がほぼ決まっているため、対話内容が定型文に近く、自由度の低い対話である。なお、インタビュアーとインタビュイーの発話を区別しないことにした。

V. 千葉大地図課題対話コーパス (堀内,中野,小磯,石崎,鈴木,岡田,仲,土屋,市川 1999) (MAPTASK と呼ぶ)

地図課題を遂行するための対話コーパスである。地図課題とは、目標物と経路の描かれた地図を持つ話者(情報提供者)が、目標物のみ描かれた地図を持つ話者(情報追随者)に対し、ルートを教えるという課題である。なお、情報提供者と情報追随者の発話も、話者の面識あり・面識なしなどのパラメータも区別しないことにした。このコーパスは課題による対話である。 VI. 研究室メンバー同士の対話コーパス(岩野、杉田、松永、白井 1997)(TRAVEL と呼ぶ) 旅行の計画を立てるため、二人の研究室メンバーの間で交わされた対話を収録したコーパス

旅行の計画を立てるため、二人の研究室メンバーの間で交わされた対話を収録したコーパスである。このコーパスは高親密度な大学生同士の対話者の間の対面対話である。

3.2.2 転記テキストの加工

上述の 6 種類のコーパス(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に選出する. なお、対話に関しては、話者ごとに一つのサンプルとした. CSJ2 は話者数が少ないため、結果として 4 話者のデータが 2 サンプル以上選択されていた. これ以外について話者の重複はなかった.

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部約300字(約1分の音声の転記量に相当し、speaking style の知覚に十分だと考えるため)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、speaking style だけで評定するよう、テキストの名詞(代名詞は除く)の部分を全て「 $\bigcirc\bigcirc$ 」に自動変換した(図2参照)。名詞の表現の使い分け(例えば「マイク」と「マイクロホン」など)も speaking style の一つと扱えるが、名詞を刺激にそのまま表示すると、本来 speaking style とは独立させるべき話題内容が明確に伝わってしまう恐れがある。したがって、データの話題内容によらない speaking style を推定することを目指すために、話題を強く想起させ得る名詞(代名詞以外)を伏せることにした。なお、転記テキストにある時間情報、フィラー、言いよどみ、笑い、咳などの情報を消し、書式を統一し、図2に例を示したような仮名漢字文字の表記に揃えた上で評定の刺激とした。

こう 魚に 触れる ことができるんですよ それも 凄い 手の 届く 範囲で で触ろうと すると ひゅって 逃げちゃうんですけど



こう ○○に 触れる ○○が できるんですよ それも 凄 い ○○の 届く ○○で で 触ろうと すると ひゅって 逃げちゃうんですけど

図 2 転記テキストからの刺激作成 (例)

評定方法 3.3

本評定には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. 不明瞭の場合を 1. 明瞭の場合を 7. 「親しさ」に関し て、親しくない場合を1、親しい場合を7、「社会階層」に関して、低い場合を1、高い場合を7 とする、評定はウェブブラウザ上のアンケートページを介して行う、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。刺激用の転記テキストはランダマイズして提示する。 テキストを読み終えてから評定するように指示し、評定中何度でも読み直して良いとした、な お、評定実験に先立って、3名の評定者(評定実験の被験者には含まれない)による予備実験を 行い、3 尺度に関しての説明の妥当性および評定の安定性を確認した、その結果、3 尺度 I. F. C の級内相関係数 (Intraclass Correlation Coefficient, ICC) (Koch 1982) の ICC(2,1) (評定者間 の信頼性) はそれぞれ 0.50, 0.79 と 0.82 であり、Landis ら (Landis and Koch 1977) によれば、I は "moderate", F は "substantial", C は "almost perfect" であるため、予備実験における評定 者間の信頼性が高いことを確認できた. 3尺度に関しての説明を若干修正した上で評定実験を 行った. 評定実験で用いた教示の例を図3に示す.

(1/60)(17.60) 〇〇よりは 安いけど 凄い 高いじゃないですか だから 〇〇 買えないから 〇〇から こう 〇〇を 見てるだけで こう 眺めて 楽しんでたんですけど そしたら 〇〇が みんな 入れば いいじゃんて 言って 私は いや やめようよって 言ったんですけど 〇〇に みんな 入ってっちゃって で そしたら 何か 〇〇の 凄い こう 〇〇 〇〇 決めた 〇〇が こう 〇〇を 開け 重い 〇〇を 開けてくれて 凄い いらっしゃいませって 言ってくれたんですけど でも 〇〇 買えないのに で お〇〇の 〇〇に いる 〇〇は 何か お〇〇つぽい 〇〇 ばかりで こう 見るからに 〇〇な〇〇〇〇な 〇〇 〇〇〇〇が 〇〇に 〇〇 〇〇して しかも 〇〇〇〇ぐらいで 〇〇 に 出たんで かなり 嫌がられたと 思うんですけど ちょっと 〇〇 の ちょっと 〇〇の ある〇〇

上のテキストを読み、次の全ての項目について、その印象がどれくらいか、1から7の7段階の評価で選んでください。 各項目について中立の印象は4になります。中立の場合には必ず4にチェックしてください。 テキストは何度読んでも構いません。

1: Intelligibility-oriented(明瞭さ) 不明瞭 01 02 03 04 05 06 07 明瞭 2: Familiarity(親しさ) 親くない 01 02 03 04 05 06 07 親しい 3: Social strata(社会階層) 低い ○1 ○2 ○3 ○4 ○5 ○6 ○7高い

項目に関しての説明:

項目に関しての説明:
1: Intelligibility-oriented(明瞭さ):
発話者の発話内容の明瞭さの度合いです。 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します。 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みます。 明瞭さが低い場合は1, 明瞭さが高い場合は7で、1から7の7段階の評価で選んでください。

発話者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す

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 示します。 親しさが低い場合は1,親しさが高い場合は7で、**1から7の7**段階の評価で選んでください。

3: Social strata(社会階層)

3. Social strate,社工時間 分 発話者の発話内容の教養の度合いです。 口語的な、砕けた、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)を示します。 発話者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ります。 社会階層が低い場合は1,社会階層が高い場合は7で、**1から7の7**段階の評価で選んでください。

図 3 評定実験で用いた教示の例

3.4 評定結果

3尺度は独立した尺度であることを想定しているが、全22名の評定者の評定結果に基づいて尺度間の相関係数を求めたところ、明瞭さIと親しさFの相関係数が-0.27、明瞭さIと社会階層 C が0.48、親しさF と社会階層 C が-0.55 であった。明瞭さI と親しさF の相関は弱いが、社会階層 C と明瞭さI・親しさF との間には中程度の相関が見られた。Eskenazi は 3 尺度の独立性については特に言及していないが、今回の実験で日本語を対象としたため、一部に 3 尺度の独立性が見受けられないことは、日本語での話し方や発話内容の教養度が話者の間の関係に影響されるという傾向によるものだと考えられる。1 章で述べた音声言語コーパス検索サービスでの応用を考えると、属性情報による絞り込み検索や可視化検索などにおいて必ずしも属性間の独立性を保証する必要はないため、尺度の間に相関があっても大きな支障はないと考える。

6 コーパス(カテゴリ)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見るために、全 22 名の評定者の評定結果の平均を用い、I と F (図 F 4 の左上)、F と F (図 F 4 の下部)のそれぞれの F 2 次元散布図と箱ひげ図(図 F 5)

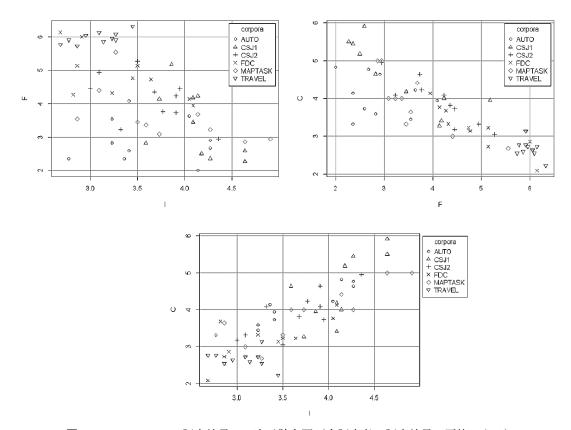


図 4 Speaking style 評定結果の二次元散布図 (全評定者の評定結果の平均による)

を作成した. 図 4 において、CSJ1 は他のコーパス(カテゴリ)に比べて I が高く分布している(有意水準 5 %の t 検定によって、MAPTASK を除いて有意). 図 5 の平均値でも CSJ1 は最も高い. これは、CSJ1 が講演音声であり話者が明瞭な発話を意識しているためと考えられる. CSJ2 は、CSJ1 と比べて F が高く C が低く分布している(t 検定によって有意傾向が見られ、図 5 の平均値についても同様の傾向). これは、CSJ2 が CSJ1 と異なり、講演話者へのインタビューであり、対話の形をとることによると考えられる。FDC は C が低く分布している(有意水準 5 %の t 検定によって、図 5 の平均値では TRAVEL についで 2 番目に有意に低い). これは FDC は大学生同士の間の自由対話であることによると考えられる。AUTO は F が低く分布

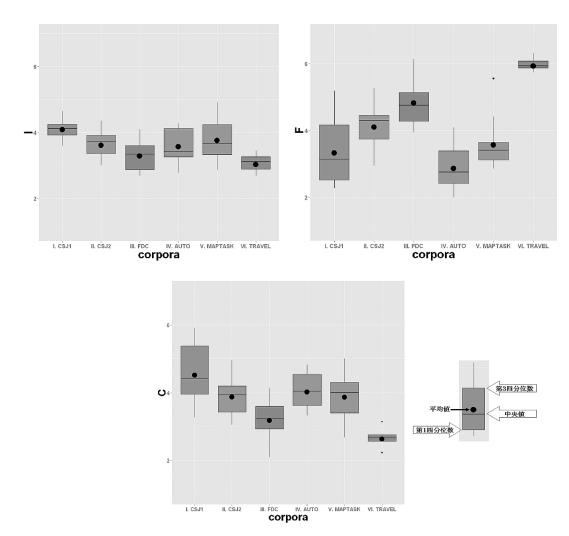


図 5 刺激ごとの評定結果の平均(左上は I. 右上は F. 下は C)

している (有意水準 5%の t 検定によって、CSJ1 を除いて有意、図 5の平均値においても同様の傾向). これは AUTO は実験者と被験者との間の自由度の低い対話であることによると考えられる。MAPTASK は、I が広く分布し、同じ課題対話である AUTO と比較して F や I も 広く分布している。これは、MAPTASK が話者の面識ありとなしの両方を含むことによると考えられる。TRAVEL は他のコーパス(カテゴリ)に比べて I が低く F が高く分布している(有意水準 5%の t 検定によって、FDC (I のみ)を除いて有意、図 5の平均値においても同様の傾向)。これは対話者同士が同じ研究室の大学生であることに加えて、同じ大学生同士の FDC よりも高親密度であることによると考えられる。以上のことから、3.2.1 節で述べた各コーパス(カテゴリ)の speaking style に関する特性が評定値の分布に現れているといえる。一方、3.2.1 節で述べたように、3.2.1 の考察から予想された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。しかし、例えば 1.2.1 と 1.2.1 を選び、上述の考察から予想された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。しかし、例えば 1.2.1 と 1.2.1 を 1.2.1 の 1.2.1 を 1.2.1 を

なお、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 (Koch 1982) を算出した。3 尺度 I, F, C の 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) (平均の信頼性) は 0.72, 0.96 と 0.92 であった。Landis ら (Landis and Koch 1977) によれば、ICCの目安として、0.0–0.20 が "slight"、0.21–0.40 が "fair", 0.41–0.60 が "moderate",0.61–0.80 が "substantial",0.81–1.00 が "almost perfect" とされている。これにしたがえば、ICC(2,1) において F と C は許容範囲内の信頼性といえ、また評定者数が多いことに起因して ICC(2,k) において I, F, C のいずれも評定値の平均の信頼性が高く、モデル構築に用いて良いと考える。なお、本研究が目指すコーパス検索サービスのための尺度としては、評定者によって大きく異ならないものであることが望ましく、上記の結果から I の尺度はこのままでは検索用途に適さない、今後、尺度の説明の見直しなど、検討の必要がある。

4 Speaking style 推定モデルの構築

3章で述べた評定実験によって得られた評定結果と刺激に用いたテキストに現れた特徴量に よってモデル構築を行う.

4.1 特徴量抽出

2章で述べた理由で、本研究では主に品詞・語種率、形態素パタンなどの形態論的特徴を特徴 量とする.

4.1.1 形態素解析

3章で述べた 60 転記テキストに対し, UniDic (伝, 小木曽, 小椋, 山田, 峯松, 内元, 小磯 2007) を辞書として用いた MeCab で自動解析する.

4.1.2 品詞・語種率

各コーパス(カテゴリ)の品詞・語種率(感動詞 int, 助動詞 aux, 動詞 v, 接頭詞 pref, 副詞 adv, 代名詞 pron, 接続詞 conj, 助詞 par, 形容詞 adj, 連体詞 adno, 機能語 func の合計 11 種)を箱ひげ図で示す(図 6 参照)。図 6 の縦軸は品詞・語種率の割合であり、横軸は左から CSJ1, CSJ2, FDC, AUTO, MAPTASK, TRAVEL の順に 6 コーパス(カテゴリ)ごとの結果を示している。なお、個々の品詞・語種率は、転記テキストごとの延べ語数に対する各品詞・語種の延べ語数の割合として求めた。

図 6 に示したように、品詞・語種率の傾向はコーパス(カテゴリ)ごとに異なることがわかった. 例えば、AUTO や MAPTASK のような自由度の低い発話のコーパス(カテゴリ)に比べて、 CSJ2、FDC、TRAVEL のような相対的に自由度の高い発話のコーパス(カテゴリ)の代名詞 (pron) の割合が高い(有意水準 5 %の t 検定によって有意)ことや、AUTO のような定型文に 近く自由度の低い対話のコーパス(カテゴリ)の助動詞 (aux) の割合が高い(有意水準 5 %の t 検定によって有意)ことなど、品詞・語種率が特徴量として speaking style の自動推定に寄与する見込みがあることを示している.

4.1.3 形態素パタン

小山らによると、文章に出現する形態素パタンが文章間の類似性を定量化するための特徴量として有効である (小山、竹内 2008). 沈らは日本語話し言葉コーパスの転記テキストから主に形態論情報や統語論情報などの言語的特徴を手がかりとし、印象形成に寄与すると思われる特徴を 43パタン抽出した (沈、菊池、太田、三田村 2012). 沈らが用いた特徴は全て音声コーパスの転記テキストから抽出したものであり、今回の推定対象となったコーパスの転記テキストの性質と一致している。金水によれば、ある種の日本語の話し方(speaking style と類似の概念)は、その話し手として、特定の人物像を想起させる力を持つ発話スタイルのことを「役割語」と名づけた (金水 2011). したがって、speaking style と特定のキャラクタ像の形成との間に関連はあるものと考え、これらのパタンを本研究に用いた。そのため、我々はこの 43 パタンについて、3 章の評定実験に使用する転記テキストに現れた 23 パタンの出現数を本研究における使用特徴量とした(表 1 参照). なお、刺激ごとに形態素数が異なるため、刺激内のパタン出現数を刺激内の形態素数で正規化した.

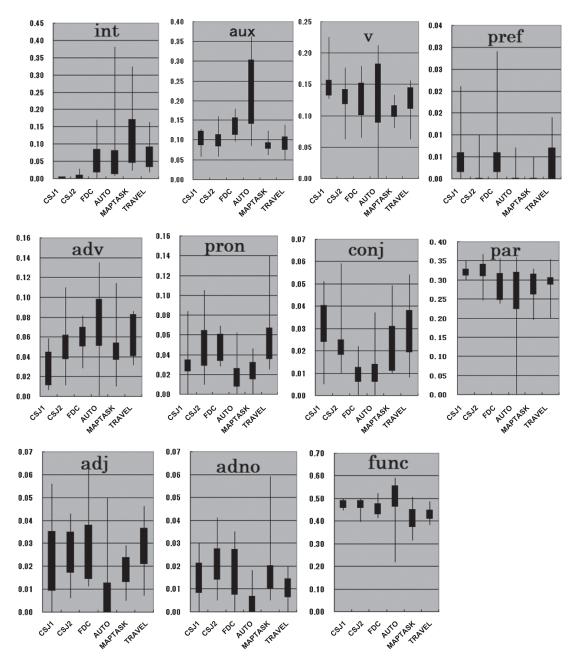


図 6 コーパス (カテゴリ) ごとの品詞・語種率 (転記テキストごとの延べ語数に対する各品詞・語種の 延べ語数の割合を縦軸とした箱ひげ図)

例 形態素パタン No. という と [と](助詞).[言う](動詞) 1 として 2 と [と](助詞) し [為る](動詞) て [て](助詞) よね 3 よ [よ](助詞) ね [ね](助詞) けれども .[けれど](助詞) も [も](助詞) 4 と [と](助詞) か [か](助詞) とか 5 行ってて 6 .[てる](助動詞) て [て](助詞) 7 って [って](助詞).[言う](動詞) っていう 8 なんか [なんか](助詞) なんか 9 ね[ね](助詞) ね よ [よ](助詞) ょ 10 .[全然](副詞) 全然 11 やっぱり 12 .[矢張り](副詞) 凄く 13 .[凄い](形容詞) .[結構](副詞) 結構 14 です .[です](助動詞) |.[ます](助動詞) 15 で [で](接続詞) で 16 こんな .[こんな](連体詞) |.[そんな](連体詞) |.[あんな](連体詞) 17 行っちゃう .[ちゃう](助動詞) 18 .[とく](助動詞) 言っとく 19 こっち [此方](代名詞) | そっち [其方](代名詞) | あっち [彼方](代名詞) こっち 20 じゃ 21 じゃ [で](助詞) 敬語表現 ございます 22 オノマトペ ひゅっと 23

表 1 形態素パタン

(形態素パタンの記法は「出現形 [辞書形](品詞)」の形式で 1 形態素を表し、"." は任意の 1 文字以上の文字列を表し、"A|B" は「A または B」を表す。)

4.2 モデルと考察

Speaking style の自動推定をするため、重回帰分析により推定モデルを構築した。分析には、ステップワイズ変数選択(変数減増法)1の手法で、各コーパス(カテゴリ)の品詞・語種率と形態素パタンを説明変数とし、評定実験の結果の平均を目的変数として最適なモデルを求めた。もと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(Montgomery and Peck 1982)を求めた。その結果を表2に示す。表2に示した結果は、ことわ

¹ 具体的には、統計ソフト R の step 関数を用いる。まず全変数を取り込んで AIC を最も改善させる一つの変数を 削除する。次に AIC が最も改善するように一つの変数を削除あるいは追加する。削除しても追加しても AIC が改 善しないなら止める。

特徴量セット	Intelligibility-oriented (I)	Familiarity (F)	soCial strata (C)
全特徴量	0.54/0.37	0.85/0.81	0.78/0.66
全特徴量(同一コーパスを含めない)	0.36/0.13	0.80/0.74	0.62/0.52
品詞・語種率のみ	0.24/0.13	0.32/0.24	0.36/0.31
形態素パタンのみ	0.36/0.23	0.76/0.67	0.55/0.45

表 2 交差検定 (leave-1-out) の結果 (決定係数/調整済決定係数)

りのない限り全て、テストデータと同一コーパスの他データを学習データに含めて交差検定を 行った結果であるが.交差検定の際に同一コーパスの他データを学習データに含めるかどうか を応用目的に照らして検討する必要がある。同一コーパスの他データを学習データに含めた場 合の交差検定 (leave-1-out) は、同一コーパスの他データの評定値が部分的に既知である場合に 相当する。実際の応用において、推定精度をあげるためにコーパス内の一部の部分的単位に人 手で推定結果を与えてモデル学習に利用することは充分考えられる.そこで.この方法の交差 検定結果を表2の「全特徴量」に示した.一方で.同じコーパスの他データを学習データに含め ない場合の交差検定 (leave-1-out) は、同一コーパスの評定値が全て未知である場合に相当する. 実際の応用においてこうしたケースもあり得るため、この方法の交差検定結果を「全特徴量(同 ーコーパスを含めない)」に示した。3.2.1 節に述べたように、評定実験では speaking style の多 様さと実験のコストの両方を考慮した上で、60 サンプルを評定の刺激としたが、モデル構築の ためのサンプルとしては少ないことが懸念される。サンプル数の少なさの影響を考慮するため に自由度調整済決定係数を求めたところ、同一コーパスが学習データとして存在する場合、Iは 0.37、Fは0.81、Cは0.66であった. 1章で述べたとおり、本研究では、一つのコーパスに対し て複数の部分的単位ごとの推定結果の集積が利用されることを想定しており、全ての部分的単 位に対して正しく推定できていなくても、50%以上の部分的単位を説明できることに相当する 推定精度(決定係数0.50以上)があれば、1章に述べたような応用は実現可能と考える、Fと Cの決定係数は 0.50 を大きく上回っているため応用には十分と考える. 全特徴量(同一コーパ スを含めない)の場合、すなわち、未知のコーパスの場合、自由度調整済決定係数は I が 0.13、 Fが 0.74、Cが 0.52 であり、精度が下がるが、F, C はやはり 0.5 を超えているため応用にとっ て有効と考える. I についてはいずれの場合も推定の精度が高くなく,また C については今後 さらなる改善が必要と考える.

さらに、表2には品詞・語種率のみと形態素パタンのみを使用した場合の決定係数を示す。この結果から、3尺度ともに、品詞・語種率と形態素パタンの両方を使用した場合のモデルの精度は、それぞれを単独で使用した場合よりも高いことがわかり、両方を特徴量として用いることの有効性が示された。

表3に選択された説明変数とその偏回帰係数を示す. Fのモデルにおいて, 形態素パタン「.[で

表3 選択された説明変数と重回帰分析結果における偏回帰係数

	Intelligibility-oriented (I)		Familiarity (F)		soCial strata (C)	
	Explanatory variable	Estimate	Explanatory variable	Estimate	Explanatory variable	Estimate
П	.[けれど](助詞) も[も](助詞)**	35.10	.[けれど](助詞) も [も](助詞),	-25.09	.[けれど](助詞) も [も](助詞)*	34.06
2	よ [よ](助詞) ね [ね](助詞),	19.01	.[です](助動詞) .[ます](助動詞)***	-19.92	よ [よ](助詞)***	-21.21
က	って[って](助詞).[言う](動詞),	-15.16	よ[よ](助詞)***	19.79	.[ちゃう](助動詞),	-13.86
4	と [と](助詞).[言う](動詞)	12.79	.[ちゃう](助動詞)*	17.66	.[です](助動詞) .[ます](助動詞)***	11.34
ಬ	じゃ [で](財詞)**	12.33	.[結構](副詞),	-16.13	連体詞率**	-11.31
9	連体詞率**	-11.30	ね [ね](助詞)***	14.19	と [と](助詞) か [か](助詞)*	-9.44
7	[結構](副詞)	-11.16	接頭辞率,	12.26	.[てる](助動詞) て [て](助詞)*	-7.87
∞	よ[よ](助詞),	-9.21	連体調率*	10.63	助動詞率***	-7.04
6	.[です](助動詞) .[ます](助動詞)***	7.65	と [と](助詞) か [か](助詞)*	90.6	ね [ね](助詞)*	-6.1
10	ね [ね](助調)*	-7.44	助動詞率***	69.9	機能語率***	4.89
11	.[てる](助動詞) て [て](助詞),	-6.02	機能語率***	-5.79	形容詞率**	-2.84
12	助動詞率***	-4.63	と [と](助詞).[言う](動詞)	-5.46	感動 詞率	-1.51
13	機能語率***	3.67	代名詞率,	4.54	$(Intercept)^*$	-1.78
14	副詞率,	3.13	形容詞率*	2.69		
15	動詞率,	2.24	(Intercept)***	2.72		
16	助詞率	1.09				
17	$17 \mid (Intercept)^{***}$	-3.09				

形態素パタンの記法は「出現形 [辞書形](品詞)」の形式で1形態素を表し,""は任意の1文字以上の文字列を表し,"A|B"は「A またはB」 Signif. codes: 0 "***", 0.001 "**", 0.01 "*", 0.05 ",", 0.1 " "

す](助動詞) |.[ます](助動詞)」の偏回帰係数は2番目に絶対値が大きく、です・ます調の出現はFのモデルに負の働きがあるといえる。これは、心的距離が近ければ普通体を用いることが多い (川村 1998) といった先行研究と対応付いている。4番目の「.[ちゃう](助動詞)」(例:行っちゃう)などの音の変化形の出現 (川村 1998) も Fのモデルに大きく貢献することがわかった。他にも、3番目の「よ [よ](助詞)」(例:よ)、6番目の「ね [ね](助詞)」(例:ね)のような終助詞の出現がFのモデルに正に働くことがわかった.「よ」や「ね」などの終助詞は、会話を親しげにしたり、同意を求める雰囲気にしたりする機能を持つという先行研究 (川崎 1989) の主張と対応付いている。Cのモデルにおいては、形態素パタン「.[けれど](助詞) も [も](助詞)」の偏回帰係数が最も大きく、「けれども」などの出現がCのモデルに正の働きがあるとわかった.このことは「けれども」などがよく改まった会話に用いられるという指摘 (川村 1994) と一致している.ほかにも、5番目の「連体詞率」などの形態素パタンが Cのモデルに大きく貢献することがわかった.

5 おわりに

本研究では、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に、音声言語コーパスの speaking style を自動推定して属性情報として付与することを目指す。本研究では、従来のテキストの文体・ジャンル判別の手法を用い、音声に付随する転記テキストの形態論的特徴を手がかりとし、speaking style 推定モデルを構築することにより、speaking style の自動推定を試みた。先行研究より、speaking style を「明瞭さ」(Intelligibility-oriented)、「親しさ」(Familiarity)、「社会階層」(soCial strata)の 3 尺度で定義し、6 コーパス(カテゴリ)から選出した音声の転記テキストを刺激とし、speaking style の評定を行った。評定実験によって得られた評定結果の分布がコーパスの特性と合致していることを確認した上で、重回帰分析を行い、speaking style における 3 尺度のそれぞれの回帰モデルを求めた。モデル構築の際、テキストの文体・ジャンル判別や著者推定などの従来手法において重要性が確認されている品詞・語種率以外に、文書分類や印象形成に有効だと思われる形態素パタンを特徴量として導入した。交差検定を行った結果、特に同ーコーパスのサンプルが学習データとして使用できる場合には、本研究の提案手法によって 3 尺度中 F と C の speaking style 評定値を高い精度で推定できることを確認した.

本稿では、言語的特徴としてまず形態論的特徴に絞って扱ったが、係り受け情報のような統語的特徴も speaking style の推定に役立つ可能性がある。今後、こうした特徴を加えて推定精度の向上を目指す。

本研究で提案した 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。今後の方針

として、speaking style 推定モデルを用いて、一つのコーパスに付随する転記テキストに対して speaking style を自動推定した上で、3尺度の空間上でそのコーパス全体の speaking style を可 視化できるようにすることを目指す.

また、本研究の成果を外国語教育分野において、speaking style の習得支援に生かせるようにする予定である。

参考文献

- Abe, M. and Mizuno, H. (1994). "Speaking Style Conversion by Changing Prosodic Parameters and Formant Frequencies." In *Proceedings of the International Conference on Spoken Language Processing*, pp. 1455–1458.
- Biber, D. (1988). Variation Across Speech and Writing. Cambridge University Press.
- Cid, M. and Fernandez Corugedo, S. G. (1991). "The Construction of a Corpus of Spoken Spanish: Phonetic and Phonological Parameters." In *Proceedings of the European Speech Communication Association Workshop*, 17, pp. 1–5.
- Delgado, M. R. and Freitas, M. J. (1991). "Temporal Structures of Speech: Reading News on TV." In *Proceedings of the European Speech Communication Association Workshop*, **19**, pp. 1–5.
- 伝康晴, 小木曽智信, 小椋秀樹, 山田篤, 峯松信明, 内元清貴, 小磯花絵 (2007). コーパス日本語学のための言語資源: 形態素解析用電子化辞書の開発とその応用. 日本語科学, **22**, pp. 101–122.
- Eskenazi, M. (1993). "Trends in Speaking Styles Research." In *Proceedings of Eurospeech*, pp. 501–509.
- European Language Resources Association (ELRA). http://www.elra.info/
- 言語資源協会 (GSK). http://www.gsk.or.jp/index_e.html
- 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, 14 (2), pp. 261-272.
- 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~. 情報処理学会研究報告, 97 (16), pp. 105-112.
- Joos, M. (1968). "The Isolation of Styles." In Fishman, J. A. (Ed.), Readings in the Sociology of Language, pp. 185–191. The Hague: Mouton.
- Jorden, E. and Noda, M. (1987). *Japanese the Spoken Language*. New Haven & London: Yale University Press.
- 川村よし子 (1994). 上級クラスにおける表現の指導—「改まり度」に応じたことばの使い分け —. 講座日本語教育, **29**, pp. 120–133.

- 川村よし子 (1998). 目上に対して「親しさ」を表す会話のストラテジー. 講座日本語教育, **33**, pp. 1–19.
- 川崎晶子 (1989). 日常会話のきまりことば. 日本語学, 8 (2).
- 菊池英明, 沈睿, 山川仁子, 板橋秀一, 松井知子 (2009). 音声言語コーパスの類似性可視化システムの構築. 日本音響学会秋季研究発表会講演論文集, pp. 441-442.
- 金水敏 (2011). "現代日本語の役割語と発話キャラクタ." 金水敏 (編), 役割語研究の展開, pp. 7–16. くろしお出版.
- Koch, G. G. (1982). "Intraclass Correlation Coefficient." In Kotz S. and Johnson N. L. (Eds.) Encyclopedia of Statistical Sciences, Vol. 4, pp. 213–217. New York: John Wiley & Sons.
- 小磯花絵,小木曽智信,小椋秀樹,宮内佐夜香 (2009). コーパスに基づく多様なジャンルの文体 比較―短単位情報に着目して―. 言語処理学会年次大会発表論文集,pp. 594–597.
- 国立情報学研究所音声資源コンソーシアム (NII-SRC). http://research.nii.ac.jp/src/
- 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, 10 (3), pp. 52-68.
- 小山照夫, 竹内孔一 (2008). 形態素出現パタンに基づく文書集合類似性評価. 情報処理学会研究報告 自然言語処理研究会報告, **2008** (113), pp. 51–56.
- Labov, W. (1972). "The Isolation of Contextual Styles." In Labov, W. (Ed.) Sociolinguistic Patterns, pp. 70–109. Oxford: Basil Blackwell.
- Landis, J. R. and Koch, G. G. (1977). "The Measurement of Observer Agreement for Categorical Data." Biometrics, 33, pp. 159–174.
- Linguistic Data Consortium (LDC). http://www.ldc.upenn.edu/
- 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, **4** (2), pp. 51–61.
- Mairesse, F., Walker, M. A., Mehl, M. R., and Moore, R. K. (2007). "Using Linguistic Cues for the Automatic Recognition of Personality in Conversation and Text." *Journal of Artificial Intelligence Research*, **30**, pp. 457–500.
- 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, **25** (6), pp. 723-732.
- 水上雅博, Neubig, G., Sakti, S., 戸田智基, 中村哲 (2013). 話し言葉の書き起こし文章の話者性の変換. 人工知能学会全国大会論文集, pp. 1–4.
- Montgomery, D. C. and Peck, E. A. (1982). Introduction to Linear Regression Analysis (2nd edition), New York: John Wiley & Sons.
- 中里収,大城裕志,菊池英明 (2013). 音声対話における親密度と話し方の関係. 電子情報通信 学会技術研究報告, **112** (483), pp. 109-114.

- Shen, R. and Kikuchi, H. (2011). "Construction of the Speech Corpus Retrieval System: Corpus Search & Catalog-Search." In *Proceedings of Oriental COCOSDA*, pp. 76–80.
- 沈睿, 菊池英明, 太田克己, 三田村健 (2012). 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, **53** (4), pp. 1269–1276.
- Yamakawa, K., Kikuchi, H., Matsui, T., and Itahashi, S. (2009). "Utilization of Acoustical Feature in Visualization of Multiple Speech Corpora." In *Proceedings of Oriental COCOSDA*, pp. 147–151.

略歴

- 沈 **睿**:2005年中国華東師範大学中国語教育学科卒,2009年早大大学院修士課程了.同年同大学院博士課程入学.2012年より早大人間科学学術院助手.音声言語コーパス,言語教育の研究に従事.言語処理学会会員.
- 菊池 英明:1991年早大・理工・電気卒,1993年同大大学院修士課程了.同年株式会社日立製作所中央研究所入社.早大理工総研助手,国立国語研究所非常勤研究員,早大人間科学部非常勤講師・専任講師・准教授を経て,2012年より早大人間科学学術院教授.博士(情報科学).音声言語,音声対話,ヒューマン・エージェント・インタラクションの研究に従事.人工知能学会,日本音響学会,ヒューマンインタフェース学会,情報処理学会,電子情報通信学会等会員.

(2013 年 8 月 3 日 受付) (2013 年 10 月 12 日 再受付) (2014 年 1 月 6 日 再々受付) (2014 年 2 月 27 日 採録)