様式Ｃ－１９

科学研究費助成事業（科学研究費補助金）研究成果報告書

平成24年5月31日現在

機関番号：14301
研究種目：若手研究（B）
研究期間：2009～2011
課題番号：21700193
研究課題名（和文）話し言葉の統計的モデル化に基づく自動整形

研究課題名（英文）Study of automatic style transformation of spoken transcripts based on statistical modeling of spontaneous speech

研究代表者
秋田　祐哉（AKITA YUYA）
京都大学・学術情報メディアセンター・助教
研究者番号：90402742

研究成果の概要（和文）：
話し言葉音声認識の応用を広げるための基盤技術研究の一環として，話し言葉の自動整形手法の研究を行った．話し言葉には口語表現や冗長表現が含まれ，また句読点も存在しないため音声認識結果は可読性が低く，このままでは活用が容易ではない．これに対して，話し言葉と書き言葉の関係のモデル化に基づく表現の修正や，話し言葉における句読点のモデル化に基づく挿入からなる自動整形について検討を行った．

研究成果の概要（英文）：
To enhance the availability of spontaneous speech recognition, we conducted a study of automatic transformation of spoken transcripts to the document style. Since spoken transcripts contain a variety of colloquial and redundant expressions, and no punctuation marks are inserted, readability of such transcripts is quite low. In this study, we specifically focused on transformation of text styles by modeling the correspondence between spoken and documented texts, and automatic punctuation for spoken transcripts.

交付決定額

（金額単位：円）

	直接経費	間接経費	合　計
2009年度	1,100,000	330,000	1,430,000
2010年度	1,300,000	390,000	1,690,000
2011年度	900,000	270,000	1,170,000
年度			
年度			
総　計	3,300,000	990,000	4,290,000

研究分野：音声情報処理
科研費の分科・細目：情報学　知覚情報処理・知能ロボティクス
キーワード：話し言葉，音声認識，自動整形，モデル化

１．研究開始当初の背景

近年，講演や講義，あるいは議会などの話し言葉を対象とした音声認識の研究が進展している．話し言葉音声認識の技術は実用的な水準となりつつあることから，研究・開発の対象は認識技術単体から活用方法へと拡大してきている．具体的には，講演録や会議録の作成，字幕の生成，音声翻訳などのシステムが挙げられる．この際，話し言葉の単純な書き起こしでは可読性が低い点が音声認識を活用する上での共通の問題となっている．音声認識器は認識した単語列を単に逐次的に出力するだけであるため，話し言葉特有の口

語表現や間投詞などの冗長表現がそのまま出力され，また単語列が句読点で適切に区切られることもない．この結果，認識結果は音声認識のタスクによらず著しく読みにくいものとなっている．さらに，認識結果を後続の自動処理（たとえば機械翻訳などの言語処理）と組み合わせる際にも，冗長で区切りのない単語列は処理を難しくする原因となる．話し言葉音声認識の応用性を高めるするためには，認識にともなう基本・基盤技術として，このような話し言葉を可読性の高い整形文（書き言葉）に自動的に変換する手法の確立が不可欠といえる．

２．研究の目的

(1) 話し言葉から書き言葉への変換
- a. 話し言葉と書き言葉の言語表現の差異について，テキストデータベースを利用して定性的な特徴を抽出するとともに，統計的・量的な傾向を確認する．
- b. 音声認識結果に対してこれらの定性的・統計的分析に基づき変換を行うモデル・枠組みを検討する．
- c. 会議や講義などの複数のタスクにおける音声認識結果に本手法を適用し，有効性を検証する．

(2) 句読点の挿入
- d. 話し言葉データの言語表現や間（ポーズ）と，このデータに人手で挿入した句読点との関係を調査し，人間による句読点挿入の傾向を明らかにする．また，上記のスタイル変換と句読点挿入の関係についても調査する．
- e. この分析に基づいて，句読点挿入に有効な言語・音響的特徴を選定しモデルを検討する．
- f. 会議や講義などの複数のタスクにおける音声認識結果に本手法を適用し，有効性を検証する．

３．研究の方法

(1) 話し言葉と書き言葉の関係の統計的分析

話し言葉と書き言葉（整形文）が対応づけられた音声・テキストデータベースを利用して，修正の対象となる話し言葉表現とその修正結果，また句読点の挿入位置を抽出し，統計的な分析をもとに傾向を明らかにする．分析には会議や講義などの複数のタスクのデータベースを利用する．

(2) モデルの枠組みの検討

上記の統計的分析に基づき音声認識結果に対して表現の修正と句読点挿入を行う枠組みを設計する．検討に当たっては，音声認識に必然的に認識誤りが含まれることを考慮し，このような場合でもできるだけ頑健に整形が機能することも重視する．

(3) 整形処理の実験的評価

モデルの設計と実装・構築が終了次第，評価を実施する．これにより設計上・実装上の問題を明らかにし，随時モデルにフィードバックするとともに，拡張にも役立てる．

４．研究成果

(1) 国会会議録における整形の分析

会議録の作成時には，単語レベルにおける発言の整形から，節や句の単位での入れ替え，発言そのものの削除などが行われる．このうち単語レベルの整形は，フィラーや文末表現の削除，助詞などの挿入，口語表現の置換の3種類に大別できる．表1に，衆議院予算委員会の書き起こし（計77K単語）から会議録を作成する場合に，単語レベル以外も含むすべての作業により削除・挿入・置換された単語の数と，総単語数に対するそれぞれの割合を示す．表1より，合計で11.7%の単語が整形の対象となっていることがわかる．このうち4分の3が削除処理であり，その42%がフィラー，18%が文末表現であった．挿入処理は43%が「い」または「いる」を補うもので，このほか38%は脱落した助詞の挿入である．置換処理には種々の修正が含まれるが，40%は口語表現をより書き言葉に近い表現に改めるものであった．ここではこれら5種類の表現に特に着目し分析を行う．これらの表現の総単語数は5,352で，整形対象の59%に相当する．なお，削除や置換の処理においては，節・句・発言レベルの修正も相当数含まれているが，ここではこれらは対象としない．

表 1: 代表的な整形処理と頻度

	単語数	割合	例
削除	6,760	8.78%	
– フィラー	2,862	3.72%	えー，この，やはり
– 文末表現	1,199	1.56%	ですね，～と，～が
挿入	926	1.20%	
–“いる”	399	0.52%	てる→ている
– 助詞	354	0.46%	は，を，が，に
置換	1,345	1.75%	
– 口語表現	538	0.70%	けど→けれど， じゃあ→では
合計	9,031	11.73%	
総単語数	77,007	100%	

これら5 種類の整形箇所について，実際にどのように認識されたのかを分析した．表2に分析の結果を示す．まず削除箇所について，フィラーは2,225箇所（78%）が正しく認識された．脱落誤りとなって，結果として会議録（整形文）と一致したものが316箇所（11%），別の単語への認識誤りは321箇所（11%）であった．文末表現は1,073箇所（90%）が正しく認識され，38箇所（3%）が脱落し，88箇所（7%）で置換誤りが発生した．「いる」の挿入箇所については，326箇所（82%）が発声通り認識され，8箇所（2%）では挿入された後の単語として認識された．誤認識は65箇所（16%）である．助詞の挿入箇所では，169カ箇所（48%）が発声通り（助詞が入らず）認識された．62箇所（18%）で挿入すべき助詞が音声認識の段階で含まれており，123箇所（35%）で誤認識が発生した．口語表現の置換については，375箇所（70%）で認識に成功した．脱落誤りはほとんどなく，置換誤りとあわせた合計の誤認識は111箇所（21%）であった．また52箇所（10%）は整形後の表現として認識されたものである．これらをまとめると，5種類の整形箇所（5,352箇所）の77.9%において正しく認識ができていることが確認された．音声認識としては誤りであるが，整形結果と一致するため問題のない箇所（476箇所）を含めると86.8%となり，単語正解率とほぼ同等となる．これより，整形の対象箇所は十分に認識できており，自動整形技術の適用可能性が裏付けられたといえる．

表 2: 各整形箇所における音声認識の結果

種類	書起しと一致	整形と一致	認識誤り
フィラー削除	2,225	316	321
文末表現削除	1,073	38	88
“いる” 挿入	326	8	65
助詞挿入	169	62	123
口語表現置換	375	52	111
合計	4,168	476	708

(2) 講演における句読点の自動挿入

本研究では，『日本語話し言葉コーパス』（CSJ）の講演音声の書き起こしに対して人手により句読点の挿入を行い，その傾向を分析した．対象としたのはCSJ で「コア」と呼ばれる講演のうち177 講演（学会講演70・模擬講演107）で，これらの書き起こしに対して専門の速記者3名によりそれぞれ独立に句読点の付与を行った．177 講演の合計の単語数は365,285である．

表3に3名の作業者（A・B・C）ごとの読点と句点の総数を示す．表1から，読点の数が作業者によって顕著に異なることがわかる．図1は各作業者により付与された読点の重複の度合いを示しているが，3名とも一致した読点は15,027箇所あり，これはA・B・Cの各作業者が付与した読点のそれぞれ51%・64%，76%である．一方，それぞれ20%・8%・7%の読点が，単一の作業者のみにより付与されている．これらより，多くの読点が作業者により異なる場所に付与されていることがわかる．

表3: 作業者ごとの句点・読点の総数

作業者	読点	句点
A	29,393	16,958
B	23,371	16,972
C	19,854	16,969

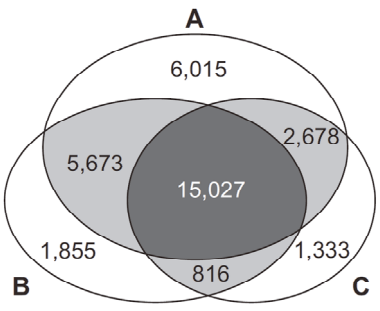

図 1: 作業者間の読点の重複の度合い

これらの句読点の自動挿入のために，本研究ではCRFに基づく識別器を構成す

る．CRFの実装にはCRF++を利用し，識別に利用する特徴は単語（出現形），品詞（大分類），文節境界，直後の文節への係り受け情報およびポーズで，これらの特徴はそれぞれ前後3単語分まで識別器に入力される．形態素解析はChaSen+IPADIC，文節境界の推定と直後の文節への係り受け推定は解析器Cabochaによって自動的に行う．

本研究では，3名のアノテータにより挿入された句読点をもとに，アノテータに共通する句読点のラベルとアノテータ個別の句読点のラベルを計6種類用意した．まず，「アノテータ共通」の句読点ラベルとして，3名のアノテータ間の共通性に基づき“3”・“2+”・“1+”の3 種類のラベルを定めた．これに対して，それぞれのアノテータにより付与された句読点をそのまま「アノテータ個別」のラベル“A”・“B”・“C” として用いる．

まず共通の読点について自動挿入を検討する．ここでは，“3”・“2+”・“1+” のアノテータ共通句読点ラベル3種類でそれぞれCRF のモデルを直接的に学習し，それぞれの挿入を行う．さらに，アノテータ個別句読点ラベル“A”・“B”・“C”を用いて対応する個別モデルを学習し，これらの3つのモデルの挿入結果を基に投票を行う手法も導入する．投票の方法としては，どれか1つ以上のモデルが投票した場合に句読点を挿入する“Any”，2つ以上のモデルの投票による“Majority”，全てのモデルの投票による“Consensus”の3種類を行う．これらは“1+”・“2+”・“3”のラベルから直接学習したモデルとそれぞれ比較可能である．

表4にモデルの直接学習と個別モデルの投票による読点挿入の結果を示す．評価データにはCSJコアの177講演を使用し，10-foldの交差検定を行っている．これらの結果から，異なるラベルを用いて独立に学習された複数のモデルの組み合わせは，ある基準に基づいて選択的に付与されている読点の挿入には有効であり，直接的な学習は任意に置かれうる読点をよくモデル化しているといえる．

表4: アノテータ共通の読点の挿入結果（F 値）

評価ラベル		1+	2+	3
直接学習	学習ラベル	1+	2+	3
	F 値	0.822	0.758	0.620
モデル投票	学習ラベル	A,B,C	A,B,C	A,B,C
	投票の種類	Any	Majority	Consensus
	F 値	0.810	0.756	0.642

次にアノテータ個別の読点の自動挿入について検討する．個別の句読点ラベルで学習された個別モデルに加えて，表4で任意に置かれうる読点に対して高い再現率・適合率を実現した1+モデル，さらに個別モデルと1+モデルの補間手法の結果を表5に示す．評価は表4と同様の交差検定によるものである．これらのモデルの中で，補間手法が最も高い性能を実現した．個別モデルを強化する上で，他のアノテータの情報を組み合わせることが有用であることがわかった．

表5: アノテータ個別の読点の挿入結果（F 値）

評価ラベル	A	B	C
個別モデルのみ	0.785	0.743	0.661
1+モデルのみ	0.793	0.737	0.655
重み付き補間	0.795	0.758	0.689

５．主な発表論文等
（研究代表者、研究分担者及び連携研究者には下線）

〔雑誌論文〕（計2件）

(1) 秋田祐哉，三村正人，河原達也．会議録作成支援のための国会審議の音声認識システム．電子情報通信学会論文誌，Vol.J93-D，No.9，pp.1736-1744，2010.（査読あり） http://search.ieice.org/bin/summary.php?id=j93-d_9_1736&category=D&lang=J&year=2010

(2) Yuya Akita and Tatsuya Kawahara. Statistical transformation of language and pronunciation models for spontaneous speech recognition. IEEE Trans. Audio, Speech & Language Process., Vol.18, No.6, pp.1539-1549, 2010.（査読あり） DOI: 10.1109/TASL.2009.2037400

様式Ｃ－１９

〔学会発表〕（計10件）

(1) 渡邉真人，秋田祐哉，河原達也．予稿の話し言葉変換に基づく言語モデルによる講演音声認識．日本音響学会春季研究発表会，2012年3月15日，神奈川大学（横浜市）．
(2) Y uya Akita and Tatsuya Kawahara. Automatic comma insertion of lecture transcripts based on multiple annotations. Interspeech 2011, 2011 年 8 月 31 日, Florence, Italy.
(3) 秋田祐哉，河原達也．講演に対する読点の複数アノテーションに基づく自動挿入．情報処理学会音声言語情報処理研究会，2011年7月21日，定山渓グランドホテル瑞苑（札幌市）．
(4) 秋田祐哉，河原達也，政瀧浩和．衆議院会議録作成における音声認識システム －言語モデル－．日本音響学会春季研究発表会，2011年3月11日，早稲田大学（東京都新宿区）．
(5) 秋田祐哉，三村正人，Graham Neubig，河原達也．国会音声認識システムの音響・言語モデルの半自動更新．情報処理学会音声言語情報処理研究会，2010年12月20日，国立オリンピック記念青少年総合センター（東京都渋谷区）．
(6) Yuya Akita, Masato Mimura, Graham Neubig and Tatsuya Kawahara. Semi-automated update of automatic transcription system for the Japanese national congress. Interspeech 2010, 2010 年 9 月 27 日, 幕張メッセ（千葉市）.
(7) 秋田祐哉，河原達也．講演における読点の個人的傾向のモデル化と自動挿入．日本音響学会秋季研究発表会，2010年9月14日，関西大学（大阪府吹田市）．
(8) 秋田祐哉，河原達也．国会音声における認識文と整形過程の分析．日本音響学会春季研究発表会，2010年3月9日，電気通信大学（東京都調布市）．
(9) 秋田祐哉，河原達也．講演の書き起こしに対する読点の自動挿入．日本音響学会秋季研究発表会，2009年9月16日，日本大学（福島県郡山市）．
(10) Y uya Akita, Masato Mimura and Tatsuya Kawahara. Automatic transcription system for meetings of the Japanese National Congress. Interspeech 2009, 2009 年 9 月 7 日, Brighton, UK.

６．研究組織

(1)研究代表者
秋田　祐哉（AKITA YUYA）
京都大学・学術情報メディアセンター・助教
研究者番号：90402742