情報処理学会論文誌 Vol.54 No.2 463–470 (Feb. 2013)

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

秋田 祐哉[1,a)] 河原 達也[1]

受付日 2012年5月31日, 採録日 2012年11月2日

概要：音声認識による講演などの書き起こしの可読性と有用性を高めるためには，句読点を自動的に挿入することが不可欠である．本論文では，単語・係り受け・ポーズの情報を素性とする条件付き確率場（Conditional Random Fields, CRF）に基づく読点の自動挿入について述べる．読点の挿入箇所は人により大きく異なるため，我々は複数のアノテータによる句読点ラベルを利用して，アノテータ個別および共通の挿入傾向をモデル化した．そして，これらを投票と補間の枠組みにより組み合わせる．『日本語話し言葉コーパス』（CSJ）の講演を用いた評価実験では，モデルの組合せにより，それぞれのアノテータの読点と，すべてのアノテータに共通する読点について高い挿入精度が得られることが示された．

キーワード：講演音声，音声認識，読点挿入，複数アノテーション，条件付き確率場

Automatic Comma Insertion in Lecture Transcripts Based on Multiple Annotations

Yuya Akita[1,a)] Tatsuya Kawahara[1]

Received: May 31, 2012, Accepted: November 2, 2012

Abstract: To enhance readability and usability of speech recognition results, automatic punctuation is an essential process. In this paper, we address automatic comma prediction based on conditional random fields (CRF) using lexical, syntactic and pause information. Since there is large disagreement in comma insertion between humans, we model individual and common tendencies of punctuation using annotations given by multiple annotators, and combine these models by voting and interpolation frameworks. Experimental evaluations using lectures of the CSJ demonstrated that the combination of these punctuation models achieves higher prediction accuracy for commas agreed by all annotators and those given by individual annotators.

Keywords: lecture speech, speech recognition, comma insertion, multiple annotations, conditional random fields

1. はじめに

音声認識の研究対象は，講義 [1] や講演・演説 [2]，議会 [3] など，さまざまな話し言葉音声に拡大してきている．このような話し言葉音声認識は音声翻訳や字幕付与，また音声の文書化への貢献が期待されているが，音声認識システムの出力には一般に句読点が含まれないことが 1 つの問題となっている．長時間に及ぶ話し言葉音声に対して，読みやすい字幕や文書を効率的に実現するためには，音声認識結果を適切な単位に自動的に区切って句読点を付与する必要がある．また，音声認識の後段に行われる機械翻訳などの自然言語処理においても，句読点の付与されたテキストを入力として想定しているため，これらの単位は不可欠である．このように句読点の自動挿入は，人間の利用・機械の利用のいずれの場合でも重要な課題となっている．

音声の書き起こしに対する句読点の自動挿入については，主に句点（すなわち文境界推定）を対象として，放送ニュースや電話会話などのタスクで多くの研究が行われてきた．最大エントロピー法やサポートベクタマシン（SVM），条件付き確率場（CRF）などの機械学習の枠組みに基づいて，韻律やポーズ・言語的情報を用いて挿入を行う手法が

[1] 京都大学学術情報メディアセンター
Academic Center for Computing and Media Studies, Kyoto University, Kyoto 606–8501, Japan
[a)] yuya@media.kyoto-u.ac.jp

一般的である [4]．日本語では，たとえば『日本語話し言葉コーパス』（CSJ）[5] の講演を対象に，ポーズと言語的情報を素性とする SVM により推定を行う手法が提案されている [6], [7]．また，句読点とは異なるものの，講演の字幕に区切り（改行）を入れるという観点からの研究も行われている [8]．このほか，大学講義を対象として，音声認識の複数仮説に基づく文整形とともに句点を挿入する手法が報告されている [9]．これに対して，読点やコンマの推定に関するこれまでの研究は限られている [10], [11], [12]．文献 [10] はポルトガル語の放送ニュース音声を対象として最大エントロピー法により推定を行うもので，素性としては単語や品詞のほかに話者や音声セグメントの境界情報などが利用されている．文献 [11] は英語の放送ニュース音声が対象で，CRF などによる推定に構文情報を導入することで読点の推定精度の改善が得られている．日本語の読点については，書き言葉の新聞記事を対象として，文献 [12] により読点の用法の分類と分析，および最大エントロピー法に基づく自動推定が提案されており，評価実験で 0.76 の F 値を得ている．

読点の挿入は句点よりも高頻度でかつ主観的であるため，たとえば音声認識結果に句読点を挿入して字幕として提示するような場合は，多くの人が一致して挿入するような箇所に読点を置くことが妥当といえる．これとは反対に，字幕字数の制限がある状況では，読点が挿入可能な箇所をできるだけ多く求めて改行の候補とするといったことも考えられる．一方，音声認識結果を編集して講演録などを作成する場合は，作業者は多くの場合は 1 人であるから，作業者の主観に沿って一貫した読点が挿入されていると効率的である．しかし，これまでの研究では正解として利用されている句読点ラベルは単一のもので，このような共通性や個人性を考慮したうえで挿入を行っているわけではない．

そこで本研究では，講演音声を対象に複数のアノテータにより付与された異なるラベルを利用して，アノテータに共通する一般的な読点と個別の読点について挿入をモデル化することを目指す．まず，講演の書き起こし中の各アノテータによる句読点について，アノテータ間の相違について分析を行うとともに，句読点の周辺における言語的な情報やポーズ情報との関係についても分析を行う．そして，これらの分析に基づいて素性を定め，CRF の枠組みに基づく自動挿入を検討する．複数のラベルにより複数のモデルが構築可能であるが，異なるラベル間で一致する読点があることから，モデルの組合せにより読点挿入の性能が向上することが期待できる．このため本研究では，共通の読点と個別の読点のそれぞれの場合に，どのようなモデルや組合せ手法が有効であるかを比較する．これら共通の読点と個別の読点の挿入について，本研究では CSJ の講演音声において評価を行う．これにより，検討した素性の有効性について明らかにするとともに，複数の読点ラベルを用いる効果についても示す．

2. コーパスとアノテーション

句読点の分析やモデル化を行うためのデータとしては，新聞や Web のニュースなどのテキストコーパスを使用することが考えられる．しかしこれらのテキストコーパスは書き言葉またはそれに近いスタイルであるため，講演のような自発的な音声とは句読点の頻度や位置が異なり，適切なデータとはいえない．また，重要な手がかりであるポーズの情報が含まれないという問題もある．

本研究では，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 177 講演（学会講演 70・模擬講演 107，総単語数 365,305）である．CSJ には音声・書き起こしに加えてポーズや非流暢現象などのアノテーションが含まれているが，句読点は付与されていない．そのため，プロフェッショナルの速記者 3 名をアノテータとして，それぞれ独立に句読点の付与を行った．手順としては，まず CSJ の書き起こしに対して整形作業を行い，この整形テキストに対して 3 名が独立に句読点を付与した．書き起こしに対して整形と句読点付与を行ったテキストの例を**図 1** に示す．

整形作業は通常講演録を作成する際に行うもので，フィラーや口語表現を対象とする一次整形と，文末表現の修正からなる．いずれも単語レベルで削除・挿入・置換を行う作業であり，句や節を入れ替えるなどの処理は行っていない．文体の変換は行わないが，整形により文末表現が挿入される場合はいわゆる「ですます」調（敬体）としている．作業にあたり，本研究では整形の対象と内容について基準を定め，句読点挿入を行うアノテータとは別の作業者により整形を実施した．基準の主な内容は次のとおりである．

(1) フィラー・非言語音：CSJ によりあらかじめタグの付与されているフィラーや非言語音はすべて削除する．このほか，感動詞やフィラー的な挿入節も削除する．
(2) 冗長表現：冗長な表現や重複表現は，係先を持たない場合には削除する．付属語や機能語「ですね」「なんですね」などは削除する．代名詞による言い換えも削除する．
(3) 口語表現：「じゃ」（では），「さして」（させて）など，口語（発音の怠け・変化など）は修正する．文末・文節末の終助詞・間投助詞は，疑問型以外については削除する．
(4) 助詞：格助詞が欠落している場合は挿入する．格助詞が間違っている場合は適切な助詞に置換する．
(5) 接続詞：文頭の接続詞「で」は削除するほか，冗長な接続詞は削除する．口語的な接続詞は標準的な日本語に置き換える．

文末表現については，体言止めなど，文末が省略されて

(1) CSJ の書き起こし
そんな風に思いながら (F んー) ドラマの一シーンみたいな感じで
(F えー) 少しですね自分で文章を作りながら中に入ってっちゃったような感じがします
(F あのー) なかなか経験できないことだと思います
(F んー)(F まー) できるならば遭難でなくてですね無人島に行って (F ま) そういう経験したいなと (F ま) 文章作りながら思いました

(2) 整形処理後のテキスト
そんな風に思いながらドラマの一シーンみたいな感じで
少し自分で文章を作りながら中に入っていってしまったような感じがします
なかなか経験できないことだと思います
できるならば遭難でなくて無人島に行ってそういう経験をしたいと文章を作りながら思いました

(3) 句読点付与後のテキスト
アノテータ A:
そんな風に思いながら、ドラマの一シーンみたいな感じで、少し自分で文章を作りながら、中に入っていってしまったような感じがします。なかなか経験できないことだと思います。できるならば遭難でなくて、無人島に行って、そういう経験をしたいと、文章を作りながら思いました。

アノテータ B:
そんな風に思いながら、ドラマの一シーンみたいな感じで。少し自分で文章を作りながら中に入っていってしまったような感じがします。なかなか経験できないことだと思います。できるならば、遭難でなくて無人島に行って、そういう経験をしたいと文章を作りながら思いました。

アノテータ C:
そんな風に思いながら、ドラマの一シーンみたいな感じで。少し自分で文章を作りながら中に入っていってしまったような感じがします。なかなか経験できないことだと思います。できるならば遭難でなくて、無人島に行ってそういう経験をしたいと、文章を作りながら思いました。

図 1 整形処理と句読点付与の例
Fig. 1 An example of annotation of punctuation marks.

いる場合には補う．また，文末が次の文と接続されていて文として切れていない場合は，CSJ で定義されている節単位境界のみを対象として修正を行う．この修正は原則として文末の接続表現を削除し，直後に接続詞を補う．修正の対象表現と内容のパターンは定義されているが，接続詞の追加の有無は作業者によって文脈から判断されている．

句読点の付与を行った 3 名のアノテータはプロフェッショナルの速記者であるので，本研究では句読点付与に際して具体的なガイドラインを提示していない．したがって句読点は速記者各自の基準と内省により付与されている．アノテータは作業の際に音声を聴取せず，整形済みのテキスト（図 1 (2)）のみを参照しているが，図 1 (2) にあるように整形済みテキストは原則として CSJ の節単位ごとに改行されているため，これがアノテータの判断に影響した可能性はある．

表 1 各アノテータによる句読点の数
Table 1 Numbers of punctuation marks by annotators.

アノテータ	読点	句点
A	29,393	16,958
B	23,371	16,972
C	19,854	16,969

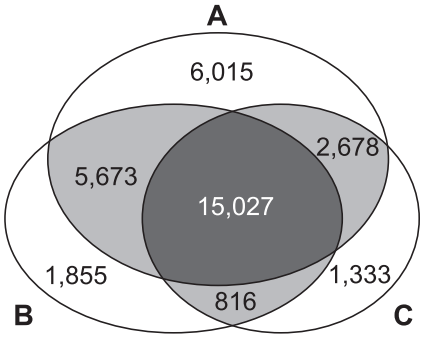

図 2 3 名のアノテータによる読点の重複の度合い
Fig. 2 Overlap of positions of commas given by the three annotators.

3. 句読点の挿入傾向の分析

本章では複数のアノテータにより付与された句読点の差異について調査する．また，言語的情報やポーズがどの程度読点と関連しているのかについても調べる．なお，これらの分析に先立って，句読点の付与されたテキストを構文解析器 Cabocha[*1] により自動解析して単語・文節への分割を行った．以降の分析における単語・文節の単位は Cabocha によるもので，CSJ における短単位・長単位ではない．

3.1 句読点の頻度と重複度合い

まず，書き起こしに付与された句点と読点の数，およびアノテータ間の重複について比較する．**表 1** に 3 名のアノテータ（A・B・C）ごとの読点と句点の総数を示す．句点の総数は 3 名のアノテータともほぼ同等であるのに対して，読点の数はアノテータによって顕著に異なる．最も少ないアノテータ C は，最も多いアノテータ A の 3 分の 2 程度である．**図 2** は各アノテータ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 カ所であり，これは A・B・C の各アノテータが付与した読点のそれぞれ 51%・64%・76%である．一方，A の読点の 20%（6,015 個），B の読点の 8%（1,855 個），C の読点の 7%（1,333 個）がそれぞれ単一のアノテータのみにより付与されている．多くの読点がアノテータにより異なる場所に付与されているが，これはたとえプロフェッショナルの速記者であっても，読点の数や位置は主観的影響を受けることを示している．なお，句点では 16,462 カ所（各

[*1] http://code.google.com/p/cabocha/

表 2 読点の直前に高頻度にみられる文節
Table 2 Frequently observed bunsetsu units followed by commas.

アノテータ A		アノテータ B		アノテータ C	
のは	172	そして	94	まず	47
して	80	ように	59	それから	41
のが	72	結局	45	また	35
時に	58	つまり	45	これは	24
これは	42	これは	38	そして	22
今	38	言うと	29	のが	19
しかし	37	また	28	例えば	19
ことで	35	私は	27	のは	17
為に	34	それで	25	ような	16
時は	33	即ち	24	特に	13
中で	33	例えば	23	あるいは	12
勿論	32	のは	21	更に	11
私は	32	それから	16	逆に	10
実際	31	言って	15	ことが	10
ことを	30	本当に	12	して	10

数値はそれぞれの出現回数である．

アノテータが付与した句点の 97%）で 3 名の一致がみられており，句点にはアノテータによる揺れが少ないことが確認された．

3.2 代表的な言語表現

日本語の文における典型的な読点の用法には，(1) 節の終端を明示，(2) “A，B，C” のような複数の要素の列挙，(3) 文内の係り受け構造（どの文節がどの文節に係っているか）の明確化，(4) 単語列が読みやすくなるよう分割，の 4 つが考えられる．このうち (1)〜(3) は英語のような他の言語でも共通する用法であるのに対して (4) は日本語に特有の用法である．また，特に (3) と (4) の用法は主観的でありさまざまな読点の入り方がみられる．そこで，読点の挿入における個人的な傾向を分析するために，読点とともに出現する言語的表現について調べる．

具体的にどのような箇所でアノテータ間に違いが生じるか調べるため，アノテータ 1 名のみにより付与された読点（すなわち図 2 における白地の部分 3 カ所）について，前後の単語・文節の比較を行った．**表 2** に，アノテータごとの，読点の直前に高頻度に観測された文節を示す．この結果，特徴的な傾向としてアノテータ A では格助詞（「は」「が」など）の直後，たとえば「〜いうことは」といった文節の直後への挿入が多くみられた．アノテータ B・C は接続詞の後に読点を挿入する傾向があったが，具体的な単語は異なり，たとえばアノテータ B では「そして・つまり・それで・すなわち」など，アノテータ C では「あるいは，それから」や副詞「まず」などの後において多数の挿入が観測された．

表 3 読点と係り受けの相関
Table 3 Correlation between commas and dependency structure.

アノテータ	隣接文節に係る	係り先がない・隣接文節以外	その他
A	2,779	26,573	41
B	2,000	21,343	28
C	2,018	17,801	35

3.3 係り受け構造

句読点は句・節や文の切れ目に置かれることから，これらを規定する文節の係り受け情報は有用な情報といえる．実際，係り受け解析と文境界推定を組み合わせることも提案されている [13]．ただし，音声認識の出力に対して句読点を付与する場合，認識誤りのために係り受け解析が大きく損なわれる．このため，本研究では直後の文節への係り受け情報 [14] に着目する．これは係り受け解析結果のうち隣接する文節に係るもののみを抽出したもので，長い係り受け構造と比較して，認識誤りに対して頑健な推定が期待できる．また，このような係り受けがある場合は文節の結び付きが強く，読点が置かれにくいと考えられる．

読点について，隣接する文節に係る箇所に置かれた場合と，係り先がない箇所または隣接する文節以外に係っている箇所に分けて調べた結果を**表 3** に示す．なお，読点は文節内に含まれるものがあり，表 3 ではその他として分類している．表 3 から，隣接する文節に係る箇所に置かれた句読点は全体の 1 割程度であり，読点が置かれにくいことが確認できた．

3.4 ポーズとの相関

音声入力に対して句読点挿入（文境界推定）を行う際に，ポーズの情報は有用であると考えられる．本研究のアノテータは句読点の付与の際に音声を聴取していないためポーズを手がかりとしては使用していないが，参考のためにポーズと句読点との対応を調査した．

図 3 はポーズ長を 0.1 秒ごとに区切って作成したポーズ長のヒストグラムで，ポーズ箇所のうち句点または読点が付与された内訳をあわせて示している．本研究では CSJ で人手により付与された転記基本単位の時刻情報をもとにポーズを抽出して利用したが，0.2 秒未満のポーズは CSJ ではアノテーションされていないため図 3 には含まれていない．なお，どのアノテータの句読点でも同様の分布を示したため，ここではアノテータ A の場合の統計を示している．

図 3 から，ポーズが長くなるほど句点に対応する割合が大きくなるのに対して，読点の割合は変動が小さいことが分かる．ポーズの存在する箇所のうち，読点が挿入された割合は平均で 29.6%である．ポーズの存在を前提としない場合の読点挿入の割合は 8.0%（単語総数 365,305 のうち

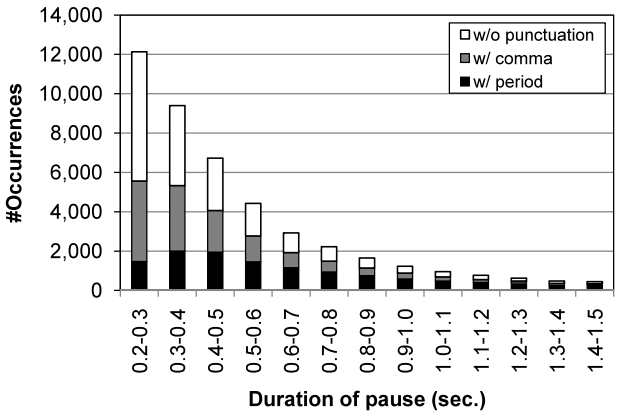

図 3 ポーズと句読点の相関

Fig. 3 Correlation between duration of pauses and punctuation marks.

29,393 個）であるから，ポーズの検出により読点の挿入される可能性が大きくなることは明らかである．これらの分析から，ポーズの出現は読点の予測の手がかりとなりうるが，長さの情報はそうではないといえる．

4. 読点の自動挿入

4.1 CRF によるモデル化

前章の分析をふまえて，句読点を自動で挿入するための，CRF に基づく識別器を構成した．CRF の実装には CRF++ [*2] を利用する．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ある．なお，形態素解析は ChaSen+IPADIC [*3]，文節境界の推定と直後の文節への係り受け推定は Cabocha によって自動的に行われている．ポーズの素性としては 0.2 秒以上のポーズの有無を抽出して利用する．図 3 で示したようにポーズ長と読点との間の相関は強くないため，ポーズ長は素性として使用しない．これらの素性はそれぞれ前後 3 単語分まで識別器に入力される．

この挿入手法について，2 章で述べた CSJ の 177 講演の書き起こしにおける 10-fold の交差検定により評価を行った．評価の指標として，本論文では次式で定義される再現率・適合率および F 値を用いる．

$$\text{再現率} = \frac{\text{正しく付与された読点（または句点）ラベルの数}}{\text{読点（または句点）ラベルの総数}} \quad (1)$$

$$\text{適合率} = \frac{\text{正しく付与された読点（または句点）ラベルの数}}{\text{出力された読点（または句点）ラベルの数}} \quad (2)$$

$$\text{F 値} = \frac{2 \times \text{再現率} \times \text{適合率}}{\text{再現率} + \text{適合率}} \quad (3)$$

再現率・適合率が改善されれば，より適切な位置や間隔で句読点が挿入されるといえるため，句読点挿入後の文の読みやすさや，後段の機械処理の性能の改善も期待できる．以降で示す各指標の値は，特に記述のない限り，この交差検定における平均値である．

4.2 素性の比較

まず，それぞれの素性の効果を測るために，素性のさまざまな組合せについてモデルを学習し評価を行った．ここでの正解には，一般的な選択手法であると考えられる多数決，すなわちアノテータ 3 名中 2 名以上が句点または読点と判断した箇所を用いた．**表 4** に，CRF の学習に用いた素性の組合せと，それぞれの場合の句点・読点の再現率・適合率・F 値を示す．

表 4 より，読点の予測ではどの素性も決定的ではなく，それぞれの素性が読点挿入における異なる要因を表現しているため，すべての素性により相乗的に性能が改善されていくことが分かる．すべての素性を利用した場合の読点挿入の F 値は 0.758 となった．一方，参考として句点についても評価を行ったところ，すべての素性を利用した場合の F 値は 0.980 となったが，単語のみを素性としてもこれに近い精度が得られた．これは，人手により編集された書き起こしで評価を行っているため，典型的な文末表現が容易に検出できることが理由と考えられる．句読点全体の F 値は 0.851 である．なお，表 4 にあるように単語素性の貢献が大きいことから，この句読点挿入モデルは講演ドメインに依存していることが分かる．

4.3 アノテータ共通の読点の挿入

次に，CRF に基づく読点挿入手法を異なる種類の句読点ラベルに対して適用し，評価を行った．なお，以降の実験では読点のみを評価の対象とし，前節におけるすべての素性を使用する．

本研究では読点ラベルとして 6 種類を用意した．まず，「アノテータ共通」の読点ラベルとして，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．“3” ラベルは，3 名のアノテータが一致して付与した読点のみをラベルとして用いるものである．同様に，“2+” ラベルは少なくとも 2 名により付与されたもの，“1+” ラベルは任意の 1 名以上により付与されたものである．これらのラベルは複数の人間の判断により選択されたものであるので，一般的であると考えられる．これに対して，それぞれのアノテータにより付与された読点ラベルを「アノテータ個別」のラベル “A” ・ “B” ・ “C” として用いる．

共通の読点の挿入に際しては，“3” ・ “2+” ・ “1+” のアノテータ共通読点ラベル 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検討する．投票の方法としては，ど

[*2] http://crfpp.sourceforge.net/
[*3] http://chasen-legacy.sourceforge.jp/

表 4 句読点の自動挿入における素性の組合せの比較
Table 4 Comparison of results by various combination of features.

使用した特徴	読点		
	再現率	適合率	F 値
単語	0.611	0.729	0.665
単語+文節境界	0.647	0.764	0.700
単語+文節境界+直後係り受け	0.698	0.768	0.731
単語+品詞	0.624	0.764	0.687
単語+品詞+文節境界	0.679	0.768	0.721
単語+品詞+文節境界+直後係り受け	0.713	0.774	0.742
単語+品詞+文節境界+直後係り受け+ポーズ	0.734	0.784	0.758

使用した特徴	句点		
	再現率	適合率	F 値
単語	0.972	0.969	0.971
単語+文節境界	0.975	0.974	0.975
単語+文節境界+直後係り受け	0.978	0.983	0.981
単語+品詞	0.974	0.973	0.974
単語+品詞+文節境界	0.976	0.973	0.975
単語+品詞+文節境界+直後係り受け	0.979	0.983	0.981
単語+品詞+文節境界+直後係り受け+ポーズ	0.975	0.984	0.980

表 5 アノテータ共通の読点ラベルに対する挿入結果
Table 5 Insertion results for commas common to annotators.

（1）モデル直接学習

評価ラベル	1+	2+	3
学習ラベル	1+	2+	3
再現率	0.814	0.734	0.559
適合率	0.830	0.784	0.695
F 値	0.822	0.758	0.620

（2）A，B，C モデルによる投票

評価ラベル	1+	2+	3
学習ラベル	A,B,C	A,B,C	A,B,C
投票の種類	Any	Majority	Consensus
再現率	0.774	0.729	0.633
適合率	0.849	0.786	0.652
F 値	0.810	0.756	0.642

1+：1 名以上のアノテータにより付与された読点
2+：2 名以上のアノテータにより付与された読点
3：すべてのアノテータにより付与された読点
A/B/C：それぞれのアノテータにより付与された読点

れか 1 つ以上のモデルが投票した場合に読点を挿入する “Any”，2 つ以上のモデルの投票による “Majority”，すべてのモデルの投票による “Consensus” の 3 種類を考えた．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．換言すれば，直接モデルを学習した場合は投票がラベル作成の時点で行われているのに対して，個別モデルによる投票は挿入の段階で行われることになる．

表 5 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．評価ラベルが “3” の場合，すなわちすべてのアノテータに共通の読点の場合，“3” モデルによる F 値は 0.620 であった．一方，評価ラベルが “1+” の場合，すなわち読点の置かれうるすべての点を予測すべき場合は，“1+” モデルによる F 値は 0.822 であった．これらの結果は，読点の置かれうる点の予測が，共通の（必ず置かなければならない）読点の予測に対して比較的容易であることを示している．これらの結果と投票の結果を比較すると，“Consensus” 投票では “3” モデルよりも高い F 値が得られたのに対して，“Majority” 投票は “2+” モデルの場合とほぼ同等であり，“Any” では “1+” モデルに比べて高い性能が得られなかった．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合せは，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．

4.4 アノテータ個別の読点の挿入

次にアノテータ個別の読点のモデル化について検討する．個別の読点ラベル “A” ・“B” ・“C” に対して，それぞれで学習された個別モデルにより挿入を行い評価を行った．これに加えて，表 5 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+” モデルについても評価を行った．さらに個別モデルと “1+” モデルの補間手法も導入する．CRF の枠組みではすべての出力候補に対して確率が計算され，この確率に基づいて識別が行われる（最大の確率を得た候補が結果として選択・出力される）．ここで，素性ベクトル X が入力された場合に，個別モデルと “1+” モデルが出力候補 O に対して与える確率をそれぞれ $P_{\mathrm{personal}}(O|X)$ および $P_{1+}(O|X)$ とすると，次式のように

表 6 アノテータ個別の読点ラベルに対する挿入結果
Table 6 Insertion results for personal commas.

評価ラベル			A	B	C
個別モデル	A	再現率	0.772	0.811	0.796
		適合率	0.799	0.668	0.557
		F 値	0.785	0.732	0.655
	B	再現率	0.617	0.712	0.665
		適合率	0.845	0.776	0.616
		F 値	0.713	0.743	0.640
	C	再現率	0.501	0.554	0.617
		適合率	0.855	0.752	0.711
		F 値	0.632	0.638	0.661
1+モデルのみ		再現率	0.832	0.877	0.859
		適合率	0.758	0.635	0.529
		F 値	0.793	0.737	0.655
重み付き補間（A/B/C & 1+）		再現率	0.803	0.793	0.741
		適合率	0.786	0.725	0.644
		F 値	0.795	0.758	0.689

A・B・C および 1+については表 5 を参照のこと．

これらの確率を補間して最終的な識別に利用する．

$$P(O|X) = \lambda P_{\text{personal}}(O|X) + (1-\lambda)P_{1+}(O|X). \quad (4)$$

補間重み λ について，本実験では 0.1 刻みですべての場合を評価し，最も高い性能を得た 0.6 に事後的に設定した．

表 6 に，A・B・C の 3 名の個別の読点ラベルについて，個別モデル，“1+” モデル，および補間手法を用いて挿入した結果を示す．個別モデルについては，評価ラベルに対応する本人のモデルを用いた場合が最も大きな F 値となることが分かる．これら本人のモデルに対して “1+” モデルを補間することで最も高い性能を実現した．個別モデルを強化するうえで，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といえる．

4.5 音声認識結果における評価

最後にこれらの挿入モデルを講演音声の認識結果に適用し評価した．この実験では，CSJ の音声認識テストセットから 5 講演を使用する．テストセットの総単語数は 11,619 で，単語誤り率は 15.0%であった．2 章で述べた書き起こしに対する整形作業に相当する処理として，ここでは自動整形手法 [15] を音声認識結果に適用している．

表 7 に音声認識結果における挿入結果と，対応する人手の書き起こしにおける挿入結果を示す．使用したラベルは，共通ラベル “1+”・“2+”・“3” と個別ラベル “A”・“B”・“C” の 6 種類で，ここではモデルの補間や投票は行っていない．書き起こしにおける結果と比較して，音声認識結果では F 値に低下がみられる．ただし低下の割合は単語誤り率と同程度であるから，性能の低下はおおむね音声認識誤りの箇所に限られていると考えられる．したがってこれらのモデルは音声認識結果に対しても有効であるといえる．

表 7 音声認識結果における読点挿入の結果
Table 7 Results of comma insertion on automatic transcripts.

評価ラベル		1+	2+	3	A	B	C
人手による書き起こし	再現率	0.819	0.740	0.525	0.779	0.711	0.663
	適合率	0.822	0.795	0.725	0.771	0.788	0.796
	F 値	0.821	0.767	0.609	0.775	0.748	0.723
音声認識結果	再現率	0.696	0.619	0.423	0.658	0.588	0.533
	適合率	0.686	0.663	0.589	0.644	0.655	0.657
	F 値	0.691	0.640	0.492	0.651	0.619	0.588

1+・2+・3 については表 5 を参照のこと．

5. おわりに

本論文では，単語・ポーズ・係り受け情報を素性とする CRF に基づく，講演音声の書き起こしへの読点の自動挿入について述べた．まず，プロフェッショナルの速記者であっても読点の挿入傾向が人により異なることを確認し，句読点と言語表現や係り受け構造，ポーズ情報との関係について分析を行った．そして，これらを素性とする CRF に基づく句読点の自動挿入器を構成した．この際，複数のアノテータにより付与された異なる読点ラベルを用いることで，アノテータに共通する読点のモデルと，アノテータ個別の読点のモデルが学習される．これらのモデルを組み合わせることで，アノテータに共通の基準，およびアノテータ個別の基準に基づく読点に対して挿入性能の改善を得ることができた．

謝辞 本研究は JST CREST および科学研究費補助金によって行われた．

参考文献

[1] Glass, J., Hazen, T., Cyphers, S., Malioutov, I., Huynh, D. and Barzilay, R.: Recent Progress in the MIT Spoken Lecture Processing Project, *Proc. Interspeech*, pp.2553–2556 (2007).
[2] Alberti, C., Bacchiani, M., Bezman, A., Chelba, C., Drofa, A., Liao, H., Moreno, P., Power, T., Sahuguet, A., Shugrina, M. and Siohan, O.: An Audio Indexing System for Election Video Material, *Proc. ICASSP*, pp.4873–4876 (2009).
[3] Akita, Y., Mimura, M., Neubig, G. and Kawahara, T.: Semi-automated Update of Automatic Transcription System for the Japanese National Congress, *Proc. Interspeech*, pp.338–341 (2010).
[4] Liu, Y., Shriberg, E., Stolcke, A., Peskin, B., Ang, J., Hillard, D., Ostendorf, M., Tomalin, M., Woodland, P. and Harper, M.: Structural Metadata Research in the EARS Program, *Proc. ICASSP*, Vol.5, pp.957–960 (2005).
[5] Furui, S., Maekawa, K. and Isahara, H.: Toward the Realization of Spontaneous Speech Recognition –Introduction of a Japanese Priority Program and Preliminary Results, *Proc. ICSLP*, pp.518–521 (2000).
[6] Akita, Y., Saikou, M., Nanjo, H. and Kawahara, T.: Sentence Boundary Detection of Spontaneous Japanese Using Statistical Language Model and Support Vector Ma-

chines, *Proc. Interspeech*, pp.1033–1036 (2006).
[7] 清水 徹，中村 哲，河原達也：音声翻訳単位の推定における句読点情報の効果，情報処理学会研究報告，2008-SLP-74-22 (2008).
[8] 村田匡輝，大野誠寛，松原茂樹：読みやすい字幕生成のための講演テキストへの改行挿入，電子情報学会論文誌，Vol.J92-D, No.9, pp.1621–1631 (2009).
[9] Fujii, Y., Yamamoto, K. and Nakagawa, S.: Improving the Readability of ASR Results for Lectures using Multiple Hypotheses and Setence-level Knowledge, *IEICE Trans. Inf. & Syst.*, Vol.E95-D, No.4, pp.1101–1111 (2012).
[10] Batista, F., Caseiro, D., Mamede, N. and Trancoso, I.: Recovering Punctuation Marks for Automatic Speech Recognition, *Proc. Interspeech*, pp.2153–2156 (2007).
[11] Favre, B., Hakkani-Tur, D. and Shriberg, E.: Syntactically-informed Models for Comma Prediction, *Proc. ICASSP*, pp.4697–4700 (2009).
[12] 村田匡輝，大野誠寛，松原茂樹：読点の用法的分類に基づく自動読点挿入，情報処理学会研究報告，2010-SLP-81-8 (2010).
[13] Oba, T., Hori, T. and Nakamura, A.: Improved Sequential Dependency Analysis Integrating Labeling-based Sentence Boundary Detection, *IEICE Trans. Inf. & Syst.*, Vol.E93-D, No.5, pp.1272–1281 (2010).
[14] 西光雅弘，秋田祐哉，高梨克也，尾嶋憲治，河原達也：局所的な係り受けの情報を用いた話し言葉の節・文境界の推定，情報処理学会論文誌，Vol.50, No.2, pp.544–552 (2009).
[15] Neubig, G., Akita, Y., Mori, S. and Kawahara, T.: Improved Statistical Models for SMT-based Speaking Style Transformation, *Proc. ICASSP*, pp.5206–5209 (2010).

秋田 祐哉 （正会員）

2000 年京都大学工学部情報学科卒業．2002 年同大学院情報学研究科修士課程修了，2005 年同博士後期課程修了．京都大学博士（情報学）．2005 年より京都大学学術情報メディアセンター助手（現，助教）．音声言語処理の研究に従事．2006 年度日本音響学会粟屋潔学術奨励賞，2009 年度情報処理学会山下記念研究賞，2011 年度情報処理学会喜安記念業績賞，2012 年度科学技術分野の文部科学大臣表彰科学技術賞を受賞．電子情報通信学会，日本音響学会，IEEE 各会員．

河原 達也 （正会員）

1987 年京都大学工学部情報工学科卒業．1989 年同大学院修士課程修了．1990 年同博士後期課程退学．同年京都大学工学部助手．1995 年同助教授．1998 年同大学大学院情報学研究科助教授．2003 年同大学学術情報メディアセンター教授．現在に至る．この間，1995～1996 年米国・ベル研究所客員研究員．1998～2006 年 ATR 客員研究員．1999～2004 年国立国語研究所非常勤研究員．2006 年～情報通信研究機構短時間研究員・招へい専門員．音声言語処理，特に音声認識および対話システムに関する研究に従事．京都大学博士（工学）．1997 年度日本音響学会粟屋潔学術奨励賞，2000 年度情報処理学会坂井記念特別賞，2011 年度情報処理学会喜安記念業績賞，2012 年度科学技術分野の文部科学大臣表彰科学技術賞を受賞．IEEE SPS Speech TC 委員，IEEE ASRU 2007 General Chair，INTERSPEECH 2010 Tutorial Chair，IEEE ICASSP 2012 Local Arrangement Chair，言語処理学会理事，情報処理学会音声言語情報処理研究会主査を歴任．日本音響学会，情報処理学会各代議員．電子情報通信学会，人工知能学会，言語処理学会，IEEE 各会員．