

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

西 光 雅 弘^{†1,*1} 秋 田 祐 哉^{†1,†2} 高 梨 克 也^{†2}
尾 嶋 憲 治^{†1} 河 原 達 也^{†1,†2}

講演などの話し言葉音声における節・文境界推定のために、局所的な係り受け情報を用いた段階的チャンキング手法を提案する。節・文境界推定を行ううえで係り受け情報は有効な特徴と考えられるが、話し言葉の文は必ずしも文法的ではなく、またフィラーやいよいよみなどの非流暢現象が含まれている。さらに音声認識器の出力には必然的に認識誤りが含まれており、このようなテキストにおける自動的な係り受け解析は困難である。このため、本研究では隣接する文節間の局所的な係り受けに着目し、サポートベクタマシン(SVM)に基づく段階的テキストチャンキングにおいてこの局所的係り受けを用いた節・文境界の候補の絞り込みを導入する。『日本語話し言葉コーパス』(CSJ)の講演音声における評価において、局所的な係り受けの情報が音声認識結果に対しても頑健に機能し、精度の改善が得られることが示された。

Clause and Sentence Boundary Detection for Spontaneous Speech Using Local Syntactic Dependency Information

MASAHIRO SAIKOU,^{†1,*1} YUYA AKITA,^{†1,†2}
KATSUYA TAKANASHI,^{†2} KENJI OJIMA^{†1}
and TATSUYA KAWAHARA^{†1,†2}

For robust detection of sentence and clause units in spontaneous speech such as lectures, we propose a novel cascaded chunking strategy which incorporates local syntactic information. Application of general syntactic parsing is difficult for spontaneous speech having ill-formed sentences and disfluencies, especially for erroneous transcripts generated by ASR systems. Therefore, we focus on the local syntactic dependency of adjacent words and phrases, and use this information to limit candidates of clause/sentence boundaries, which are detected by classifiers based on SVM (Support Vector Machines). An experimental eval-

uation using spontaneous talks of the CSJ (Corpus of Spontaneous Japanese) demonstrates that the proposed dependency analysis can be robustly performed and is effective for clause/sentence unit detection in ASR outputs.

1. はじめに

現在、会議や講義・講演などの話し言葉音声を対象とした大語彙連続音声認識の研究・開発がさかんに行われている。話し言葉音声認識の先導的なプロジェクトとしては電話会話音声を扱った Hub-5 および EARS が知られており、近年では Rich Transcription¹⁾ や AMI/AMIDA²⁾ などのミーティング音声のプロジェクトが進められている。講演に関しては、学術講演と一般のスピーチを中心とした『日本語話し言葉コーパス』(CSJ)が構築され³⁾、音声認識を含むさまざまな音声言語処理の研究が行われているほか、講義についてもさまざまな大学で音声認識が試みられている⁴⁾⁻⁶⁾。

このような話し言葉音声認識の応用として、講演録・会議録の作成⁷⁾や重要文抽出・要約^{8),9)}、字幕の生成などが考えられる。これらの応用では音声認識器の出力が利用者に提示されるが、話し言葉の音声認識結果を単純に出力しても適切な区切りがなく可読性がよくない。また、重要文抽出や要約を行う際にはあらかじめ文などの単位に分割する必要がある。さらに、これらのアプリケーションでは構文解析などの種々の自然言語処理技術を利用するが、既存の自然言語処理技術は文の単位が明確な新聞などの書き言葉を主な対象として構築されており、単なる単語列である音声認識結果のままでは入力として不適切である。これに対して一定長のポーズの検出による認識結果の分割が一般的に行われているが、特に話し言葉の場合はこのような単位が言語的に均質なまとまりになっていないとは限らず、文や節などと必ずしも一致しない。音声認識の活用のためには、音声認識結果を言語的により適切に分割する技術が求められている。

そこで本研究では、言語的情報やポーズ情報をもとに話し言葉音声を節や文などの言語的な単位に自動分割する節・文境界推定手法を提案する。言語的情報については形態素(単

†1 京都大学大学院情報学研究科
School of Informatics, Kyoto University

†2 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

*1 現在、日本電気株式会社共通基盤ソフトウェア研究所
Presently with Common Platform Software Res. Laboratories, NEC Corporation

語)や品詞がこれまで一般的に用いられてきたが,本研究ではこれに加えて隣接文節間の局所的な係り受け情報を導入する.文節の係り受け関係は推定に有用な情報と考えられるが,音声認識結果では認識誤りの影響を受けて係り受けを正しく抽出することが困難である.これに対して,係り受け関係を隣接する文節に限定することで比較的頑健な抽出が期待でき,これを用いて境界の候補を絞り込むことで音声認識結果における境界推定精度の改善が期待できる.本研究では,サポートベクタマシン(Support Vector Machines, SVM)による段階的なチャンキングにこの絞り込みを組み合わせることで節・文境界推定を実現する.以下では,まず2章において節と文の定義とこれまでの節・文境界推定手法について述べる.次に3章で提案法を詳述し,4章でCSJにおける評価を行う.

2. 話し言葉における節・文境界推定

2.1 節と文の定義

日本語のような自然言語では,まとまった内容を持ち,句点などの記号により終端が明示された言語表現を「文」と定めている.また,文の内部における,述語を中心としたまとまりを「節」としている¹⁰⁾.これに対して,話し言葉では内容のまとまりが明示的に区切られているわけではなく,音声入力では句読点も存在しないことから,このような書き言葉での文の定義を適用することができない.

したがって本研究では,話し言葉における文の単位として,CSJで定められた「節単位」を採用する¹¹⁾.節単位は節間の従属度や自身の完結性をもとに節境界から選択的に定められるものである.節境界と節単位のような階層的な境界の定義は他の言語でも見られ,たとえばLinguistic Data Consortiumにより定められた英語のアノテーション基準では,統語的・意味的な単位(SU)をさらに文レベル(sentence-level SU)および文内レベル(sentence-internal SU)に区別している¹²⁾.このような階層的定義は話し言葉における一般的な基準といえる.

CSJにおける実際のアノテーションでは,まず節境界の位置と種類を自動的に推定し,人手による修正を行ったうえで節単位が認定された.節境界は,その切れ目の大きさによって,絶対境界・強境界・弱境界の3種類に分けられており,各境界の代表的なパターン(節境界ラベル)を表1に,実際の発話での例を図1に示す.絶対境界は通常の文末表現に相当し,強境界は文末表現ではないものの発話の大きな切れ目と考えられ,これら2種類の境界は原則として節単位(すなわち文境界)となる.ただし,絶対境界のうち引用部や連体修飾部に含まれるものや,前後で主題の共有や飛び越し,あるいは節・文の挿入が発生している強

表1 節境界の種類と代表的なパターン
Table 1 Types and typical patterns of clause boundaries.

境界の種類	節境界ラベル
絶対境界	文末・文末候補・と文末など
強境界	並列節「ケド」「ガ」など
弱境界	理由節「カラ」・タリ節・条件節「ナラ」「レバ」など

ただし 接続詞 こういう生活をしてますと《条件節ト》摂取カロリーが大変多くなりまして《テ節》本来だったら 条件節タラ 私の年では一日千五百キロカロリー取れば 条件節レバ 十分なんですが《並列節ガ》大体平均すると 条件節ト 二千五百カロリーぐらいは取ってるんじゃないかなと 引用節 そういう気がいたします[文末]そういうことで 並列節デ 結構年のわりに歩いてるんですが《並列節ガ》

[]: 絶対境界 《 》: 強境界 : 弱境界

図1 節境界の例

Fig. 1 Examples of clause boundaries.

境界などは節単位とはならない(図2A)¹³⁾.一方弱境界は通常は発話の切れ目になることはないと考えられるが,機能的に分割できると判断された箇所のみ文境界となる.たとえば,係り先のない場合や話題導入表現・まとめ表現がある場合などは弱境界でも節単位として認定される(図2B).このほか,人手による修正で「体言止」などの要素が節境界として認定されている.CSJの音声認識テストセット30講演における節境界と節単位との関係を表2に示す.このうち,絶対境界の92%,強境界の74%,弱境界の5.2%,人手による境界の98%が節単位に該当している.

2.2 節・文境界推定

話し言葉のテキストや音声認識結果に対して節や文の認定を自動的に行う節・文境界推定は,話し言葉コーパスの整備やタスクの進展にともなって国内外で広く行われている.欧米では,英語などの放送ニュース音声¹⁴⁾⁻¹⁷⁾および電話会話音声^{16),18),19)}が主な研究対象である.最も一般的な手法は韻律と言語的情報をもとに決定木を用いて認定する手法であり^{15),16)},近年では最大エントロピー法や条件付き確率場(Conditional Random Field, CRF)の枠組みも導入されている¹⁷⁾.推定のための情報としてはポーズもよく利用されている.

一方,日本語を対象とした節・文境界推定としては,対話音声を対象とした統計的言語モ

A. 絶対境界・強境界が節単位とならない例

- 引用節構造
それから 接続詞 このはい 感動詞 そうです [文末] 月曜日ですって 引用節 相手の質問を繰り返したりすることが多々ありますが《並列節ガ》
- 主題の共有
それに対して テ節 従来の手法はやはり逆転現象が見られまして《テ節》ここが有意なくらいに差が逆転しています [文末]
- 挿入節
ホテルの部屋の中も早速夜着いたんですけども《並列節ケドモ》チェックしました [文末]

B. 弱境界が節単位となる例

- 係り先なし
で 接続詞 青山の昔のお話私が住んでた頃の昔のお話こう戻していきたいと 引用節 思うんですが《並列節ガ》現在の青山はあんまりよく知らないものですから 理由節カラ 昔青山小学校も青山だったものですから 理由節カラ あの辺りでも生まれ育ち 連用節 かなり遊んだという トイウ節 ことで 並列節デ 遊んだ場所ということで 並列節デ 青山の中の色々幾つか挙げていきたいと 引用節 思いますが《並列節ガ》
- 話題導入表現
それが分かったのはなぜかと 引用節 いうと 条件節ト コザクラインコってというのは全体的には背中が緑色で 並列節デ 青色がちょっと混ざったような感じなんですけれども《並列節ケレドモ》

図 2 人手により修正された節単位の例

Fig. 2 Examples of sentence boundaries corrected by human annotators.

表 2 節境界と節単位（文境界）の関係

Table 2 Relationships between clause boundaries and sentence boundaries.

	総数	節単体に該当する	該当しない
絶対境界	1,794	1,645 (91.7%)	149 (8.3%)
強境界	1,077	795 (73.8%)	282 (26.2%)
弱境界	4,123	215 (5.2%)	3,908 (94.8%)
人手境界	221	216 (97.7%)	5 (2.3%)
計	7,215	2,871 (39.8%)	4,344 (60.2%)

デルによる手法²⁰⁾、講義音声を対象とした韻律に基づく手法²¹⁾などが行われてきたが、近年は CSJ の講演音声を対象とした研究が多くみられる。日本語では明らかな文末表現が存在することが英語とは異なる特徴の 1 つであり、このため言語的情報を重視した手法が種々提案されている。具体的には、形態素解析における接続コストを利用して文境界候補を求め

る手法²²⁾や統計的言語モデルによる言語尤度とポーズをもとに文境界を定める手法²³⁾、形態素や品詞の情報とポーズを素性として構成した SVM による手法²⁴⁾、同様の素性をういた CRF と係り受け解析を組み合わせた手法²⁵⁾などである。一方、F0 やポーズ・パワーなど、韻律や非言語情報をういた推定についても検討が行われているが²⁶⁾、十分な性能は得られていない。

このように、節・文境界推定に用いる特徴は日本語では形態素や品詞といった言語的情報とポーズが、推定の枠組みとしては SVM や CRF などの機械学習によるモデルが一般的となっている。本研究では、これらの素性をういた SVM による検出手法に対して、新たに隣接文節間の局所的な係り受け関係に基づく境界候補の絞り込みを導入する。文献 25) でも用いられているように、本来係り受け関係は必ずしも隣接した文節に限られない。しかし、現在の自動係り受け解析では係り受けの距離が遠い場合に精度が低下し、文法的でない文が多い話し言葉では深刻な影響を受ける。音声認識結果の場合は認識誤りが含まれるために解析はさらに困難となる。文献 27) では最大エントロピー法を用いて係り受けの判定を行い節境界の検出を行っているが、書き起こしを対象とした評価のみ報告されており、音声認識結果に対する有効性は明らかでない。一方、文献 28) ではパーザにより得られる係り受け関係を英語の SU (すなわち文境界) の検出に利用しているが、音声認識結果における SU の検出では係り受け情報による精度の改善がほとんどない。これに対して、本研究では係り受けを隣接文節間に限定することにより音声認識結果からの頑健な抽出が期待できる。係り受け関係のある隣接文節間には節・文境界が存在する可能性は低いため、これにより境界候補を制限することで節・文境界推定性能が向上すると考えられる。

3. 係り受け情報をういた段階的境界推定

3.1 段階的チャンキング

本研究では、形態素から段階的にチャンキングを行うことにより節・文境界推定を行う。一連のまとめあげ・チャンキングは、SVM に基づくテキストチャンカである YamCha²⁹⁾を用いて行う。提案法の概要を図 3 に示す。

まず、形態素列を CSJ で定義された文節³⁰⁾にまとめあげる。次に 2 段目のチャンキングにおいて、文節内の形態素・品詞情報をもとに直後の文節への係り判定を行い、直後の文節に係る場合にはさらに係り先の文節が述語であるかを規則に基づき判定する。この結果、述語への係り受け以外で直後の文節に係る場合にそれらをまとめあげ、文の主題や述語・格要素にあたるチャンク（構成要素）を生成する。構成要素は節・文とは必ずしも一致しない

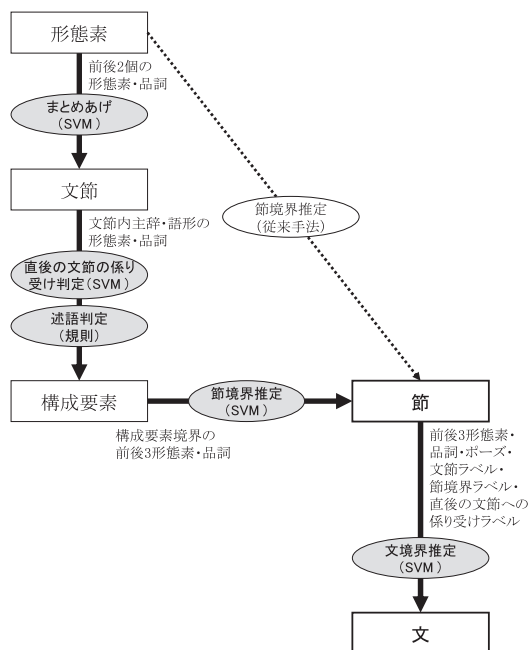


図3 段階的チャンキングのあらまし

Fig. 3 Overview of the cascaded chunking.

まとめりであるが、その境界は節・文境界の候補となりうるため、次段のチャンキング対象をこの構成要素の境界に限定することで節・文境界候補を絞り込む。したがって次の節境界推定では構成要素境界前後の形態素およびその品詞情報を SVM の素性として利用する。最後に文境界推定として、これまでに推定された文節・節境界・直後の文節への係り受けラベルを形態素・品詞・ポーズ情報とあわせて利用しチャンキングを行う。

3.2 係り受け情報を用いた構成要素の生成

係り受け情報は、ある文節が最も依存している他の文節を係り先という形で示したものである。係り先は、2文節に含まれる形態素情報から「格要素と述語」や「連体修飾節の述語と被修飾節」といった2文節間の関係を考慮して決定される。CSJにおける文節間係り受け構造は、「京大コーパス」³¹⁾の基準を原則とし、話し言葉特有の現象に対して新たな基準を設けている³²⁾。本研究では原則としてCSJに付与された係り受け情報に従うが、話し言

葉特有の現象のための新たな基準のうち、係り先が付与されていない文節と倒置による左係りについては、係り受け情報を修正した。具体的には、これらの文節のうち用言（動詞・形容詞・助動詞）・接続詞・フィラー・いいよどもを含まない文節については直後の文節に係るものとしている。これは、現在の係り受け解析技術では正しく「係り先なし」もしくは「左係り」と解析することが困難なためである。また、フィラーのみで構成される文節は基本的に係り受け関係を結ぶ要素とならないため、係り受け情報を分析する際に係り先として考慮しないこととする。

本研究では、文節のチャンキングという観点から、直後の文節への係り受けに着目した。すなわち、直後の文節に係るということ、直後の文節への依存性が強いと仮定し、文節を結合する特徴として用いる。日本語では多くの文節が直後の文節に係り、逆に係り受け解析誤りの多くが直後以外の文節に係る場合であるので、頑健に検出できることが期待される。さらに、日本語文においては格要素となる文節が述語に係り、述語は基本的に節末に存在するという特性を持つ。格要素は日本語の意味を理解するうえで重要な役割を果たしているが、格要素を備えた文節の直後を境界候補として検出することも期待できる。直後の文節への係り受けの有無により認定した単位の例を図4に示す。図4の各行が認定されたそれぞれの単位であり、単位内では文節ごとに空白または斜線で分かち書きされている。これより、直後の文節への係り受けの有無によって、述語に離れて係る主題や副詞などが独立した単位として認定されていることを確認できる。

ここで直後が述語となる格要素も境界候補として検出する必要があるが、このような文節については直後の文節への係り受けの有無だけでは判定できない。そこで、本研究では格要素を備えた文節は述語に係るという特性を利用する。日本語の文節において、その文節に述語を含むか否かを判定することは、体言止めなどの特殊な場合を除いて、文節に含まれる形態素の品詞情報により可能である。したがって本研究では述語判定に基づく境界候補の検出規則を導入する。すなわち、用言（動詞・形容詞・助動詞）を含む文節を述語と定義し、直後の文節に係る場合にその文節が述語であれば、当該文節と述語文節の間を境界候補とする。図4において改行に加えて斜線により区切られているものが、直後の文節への係り受けの有無と述語判定を用いることにより認定される単位（すなわち構成要素）である。図4より、構成要素が文の主題や述語・格要素などにおおむね対応していることが確認できる。

3.3 構成要素の分析

提案手法により生成される構成要素の意味的な妥当性を検証するために、それらの境界と節境界の共起を調査した。また構成要素に関しては、末尾形態素を分析することにより格要

例えば

母さんと お婆ちゃんが

家に // 帰ってくると

もう

門扉のところまで // 来て

はあはあはあはあ // 言いながら // 待ってるんですけども

僕や 父が

夜 // 遅く // 帰ってきて

門のところに // 来ますと

一応

その

彼らは

門まで // 出迎えてくれるんですが

‘//’は構成要素の境界

図 4 直後の文節への係り受けの有無により認定した単位の例

Fig. 4 An example of units produced using local syntactic (bunsetsu) dependencies.

素となる文節の直後が境界となっているかを調べた。

分析データ (CSJ コア講演) に提案手法を適用することにより, 140,534 の構成要素を得た。節境界 (計 52,884 個) のうち 93% (49,385 個) が構成要素の境界と一致しており, 直後の文節への係り受けの有無と述語判定という単純な操作により検出できることが分かった。なお, 一致しない 7% (3,499 個) の節境界はほとんどが弱境界であり, このうちトイウ節が 1,972 個と半数以上を占め, ほかにはテ節・引用節・接続詞などが多くみられた。これらは, たとえば「心から楽しんでるといふ トイウ節 感じがこっちに伝わってきます」のように, 境界の直前・直後に係り受けがあるために構成要素境界とはならない。一方, 節境界に一致しない構成要素 (91,149 個) について, 末尾の形態素を分析した。表 3 はこのうちの上位 5 個を示しているが, 格助詞「が, を, に」, 係助詞「は」などの格要素をなす形態素が上位を占めており, 構成要素全体の 44% (節境界を除くと 68%) については格要素に対応していると考えられる。なお, これ以外の末尾形態素としては, 副詞「ちょっと,

表 3 節単位でない構成要素の末尾形態素 (上位 5 位)

Table 3 Top five words appeared at the end of intermediate units without clause boundaries.

末尾形態素	頻度
格助詞「が」	10,993
係助詞「は」	10,562
格助詞「を」	10,193
格助詞「に」	9,776
接続助詞「て」	8,172

やっぱり」などが見られた。

4. 段階的境界推定の評価

本章では, 提案する段階的境界推定の音声認識結果に対する有効性を評価する。評価実験では CSJ コアデータ (199 講演) のうち音声認識テストセット (計 30 講演) をテストセットとし, これを除いた 169 講演を学習セットとした。テストセットの単語認識精度は 69.8% である。比較のために, 人手による音声の書き起こしに対しても評価を行っている。

4.1 文節へのまとめあげの評価

まず, SVM に基づくテキストチャンカである YamCha を用いて, 音声認識結果の形態素列を文節にまとめあげた。SVM の素性は前後 2 形態素の情報 (表層表現, 読み, 品詞情報) を基本とし, 活用語の活用型・活用形については利用する場合と利用しない場合でそれぞれ評価した。これは, 音声認識においては終止形と連体形の混同が多く, 正しい活用形の情報得られにくいと考えられるためである。以後の実験においても同様の条件で行う。YamCha における多項式カーネルの次数は 3, 解析方向は Right to Left とし, ラベリングスキームには IOE を用いた。

文節へのまとめあげの実験結果を表 4 に示す。書き言葉において SVM に基づく文節へのまとめあげが行われているが³³⁾, 話し言葉においても同様の手法で高精度に文節へのまとめあげが可能であることを確認できた。音声認識結果では書き起こしと比べて 19% の F 値の低下がみられるが, これは単語誤り率 (30.2%) よりはるかに小さいことから, SVM に基づく文節へのまとめあげは認識誤りに比較的頑健に機能しているといえる。なお, 書き起こしと音声認識結果の双方において, 活用型・活用形の利用の有無による精度の違いはみられなかった。

4.2 直後の文節への係り受けの評価

次に, 直後の文節への係り受け判定を SVM (YamCha) に基づく 2 値分類器により行っ

表 4 文節へのまとめあげ精度
Table 4 Accuracy of bunsetsu chunking.

対象	素性	再現率	適合率	F 値
書き起こし	表層表現・読み・品詞	97.9%	98.4%	0.982
	+活用形	98.0%	98.5%	0.982
	+活用型・活用形	98.0%	98.3%	0.982
音声認識結果	表層表現・読み・品詞	80.3%	78.4%	0.793
	+活用形	80.3%	78.4%	0.793
	+活用型・活用形	80.3%	78.3%	0.793

表 5 直後の文節への係り受け解析精度
Table 5 Accuracy of an analysis of local syntactic dependency.

対象	素性	再現率	適合率	F 値
書き起こし	表層表現・読み・品詞	91.6%	88.8%	0.902
	+活用形	92.0%	89.4%	0.907
	+活用型・活用形	91.7%	89.3%	0.905
音声認識結果	表層表現・読み・品詞	75.5%	74.3%	0.749
	+活用形	75.6%	74.4%	0.750
	+活用型・活用形	75.5%	74.4%	0.750

た．SVM の素性には文節内の主辞・語形の形態素と品詞の情報をういた．主辞は助詞・接尾辞を除く文節内の末尾形態素，語形は文節内の末尾形態素である．YamCha のパラメータは，解析方向が Left to Right であること以外は文節へのまとめあげと同一のものをういた．

実験結果を表 5 に示す．CSJ における従来の係り受け解析精度²⁴⁾ はオープン条件で 80.6% である．本研究では，係り受け解析の対象を直後の文節に限定することにより，書き起こしで F 値 0.9 の精度を実現している．また，音声認識結果における F 値の低下は 17.0% で単語誤り率よりも小さく，認識誤りに対しても比較的頑健であるといえる．なお，書き起こしにおいては活用型・活用形の利用により若干の精度の向上がみられるが，音声認識結果においてはその効果が小さくなっている．これは事前の予測のとおり，音声認識結果において終止形と連体形の混同が多いためと考えられる．

4.3 節境界推定の評価

節境界の推定に際しては，節境界の種類によって推定に有用な素性が異なるために，節境界の有無という単純な 2 値分類によるモデル化は適切ではないと考えられる．したがって本研究では節境界推定を表 2 にある 4 種類の節境界と境界以外の計 5 クラスを識別する問題と見なし，SVM (YamCha) の pairwise 法に基づく多値識別により推定している．ただ

し，表 2 の人手境界については頻度が小さいことから以下の評価からは除外している．節境界を推定する手法としては，形態素列の情報をういて推定する手法¹¹⁾ がこれまでに提案され，CSJ においても適用されている³⁴⁾．そこで本実験では，形態素列から節境界を直接推定する手法も YamCha をういて実装し比較を行った．YamCha (SVM) の素性としては，構成要素からの推定では構成要素境界の前後 3 形態素をういた．なお，音声認識結果に対する評価では，認識誤りが含まれるために書き起こしにおける正解と直接比較することができない．このため，書き起こしとのアラインメントに基づき音声認識結果中の適当と思われる箇所に境界ラベルを手作業で付与して音声認識結果用の正解を作成している．

書き起こしと音声認識結果における推定精度を表 6 と表 7 にそれぞれ示す．これらの表で全境界とあるのが節境界としての検出精度であり，この場合は絶対・強・弱の境界の種類の混同を許している．表 6 より，書き起こしでは絶対境界・強境界の推定に関して両手法とも同等の精度を実現している．弱境界に関しては形態素列から推定する手法の方が高い F 値となった．これは直後の文節に係る節境界（連体節など）を構成要素からは推定できないためである．一方表 7 より，音声認識結果では全種類の境界を同等の精度で検出できることが示された．また境界種間の混同を許容した場合の全境界の精度では，構成要素からの推定の方が高い F 値となった．これは直後の文節への係り受けなどの言語情報を考慮した構成要素の境界から節境界を推定することで適合率が向上したためであり，提案手法の有効性を示すものである．なお，音声認識結果において形態素列から絶対境界を推定する場合，活用形を利用することにより推定精度が大きく低下している．これは，CSJ において述語の終止形を絶対境界と定義していることから，音声認識結果における終止形と連体形の混同が大きく影響したと考えられる．

4.4 文境界推定の評価

文境界の推定も，節境界推定と同様に YamCha を利用した．形態素列から文境界を直接推定する手法²⁴⁾ では，形態素や品詞の情報，各講演内で正規化したポーズ長，節境界推定により得られた節境界を SVM の基本的な素性として利用しているが，これに加えて文節境界や局所的係り受けの情報を利用した場合の評価を試みた．文節境界および局所的係り受け情報を利用した文境界推定の結果を表 8 に示す．これらを素性として与えることにより，性能の改善がみられた．これは，特に文境界に該当しない節境界について誤検出が抑えられたことで適合率が向上したためである．図 5 に音声認識結果における文境界推定の結果の例を示す．この認識結果には認識誤りが多く，このような場合は通常の係り受け解析は困難であると考えられる．形態素に基づいた文境界推定では文境界の湧き出し誤りが発生してい

表 6 書き起こしにおける節境界推定精度

Table 6 Accuracy of clause boundary detection on manual transcription.

境界の種類	素性	形態素			構成要素		
		再現率	適合率	F 値	再現率	適合率	F 値
絶対境界	表層表現・読み・品詞	95.3%	97.1%	0.962	96.0%	96.7%	0.963
	+活用形	98.9%	99.5%	0.992	99.2%	99.1%	0.991
	+活用型・活用形	98.9%	99.4%	0.992	99.2%	99.0%	0.991
強境界	表層表現・読み・品詞	96.1%	98.8%	0.974	96.2%	98.8%	0.974
	+活用形	96.9%	98.7%	0.978	97.5%	98.5%	0.980
	+活用型・活用形	97.1%	98.8%	0.979	97.6%	98.6%	0.981
弱境界	表層表現・読み・品詞	96.9%	97.5%	0.972	89.4%	97.4%	0.932
	+活用形	97.7%	97.7%	0.977	90.2%	97.5%	0.937
	+活用型・活用形	97.7%	97.6%	0.976	90.3%	97.6%	0.938
全境界	表層表現・読み・品詞	93.4%	97.5%	0.954	90.6%	98.6%	0.945
	+活用形	95.0%	98.1%	0.965	91.7%	98.8%	0.951
	+活用型・活用形	95.0%	98.1%	0.965	91.8%	98.8%	0.952

表 7 音声認識結果における節境界推定精度

Table 7 Accuracy of clause boundary detection on recognition results.

境界の種類	素性	形態素			構成要素		
		再現率	適合率	F 値	再現率	適合率	F 値
絶対境界	表層表現・読み・品詞	70.1%	80.0%	0.747	69.1%	81.1%	0.746
	+活用形	59.8%	43.9%	0.506	69.1%	80.8%	0.745
	+活用型・活用形	59.6%	43.7%	0.505	69.2%	80.9%	0.746
強境界	表層表現・読み・品詞	68.5%	79.4%	0.736	70.9%	83.8%	0.768
	+活用形	70.9%	79.6%	0.750	70.9%	83.9%	0.769
	+活用型・活用形	70.0%	79.3%	0.743	71.0%	83.8%	0.769
弱境界	表層表現・読み・品詞	63.1%	63.6%	0.634	59.1%	65.5%	0.622
	+活用形	60.9%	62.2%	0.615	59.0%	65.6%	0.621
	+活用型・活用形	60.8%	61.8%	0.613	58.9%	65.6%	0.621
全境界	表層表現・読み・品詞	67.0%	73.0%	0.699	68.8%	80.1%	0.740
	+活用形	64.2%	61.7%	0.629	68.8%	80.1%	0.740
	+活用型・活用形	64.0%	61.4%	0.627	68.7%	80.2%	0.740

表 8 局所的係り受け情報を利用した文境界推定精度

Table 8 Accuracy of sentence boundary detection using local syntactic dependency.

対象	素性	再現率	適合率	F 値
書き 起こし	形態素・ポーズ・節	82.7%	88.0%	0.853
	+文節区切り	83.2%	87.6%	0.854
	+文節・直後への係り	84.0%	92.5%	0.880
音声認識 結果	形態素・ポーズ・節	56.4%	66.0%	0.609
	+文節区切り	56.9%	66.0%	0.612
	+文節・直後への係り	56.0%	68.0%	0.614

- 書き起こし（正解）
「えー普天間基地移設問題についてえー賛成反対にえ分かれてもらう #」
- 音声認識結果・従来法：
「えー不定ま季節問題について # え三正反対にえ分かれてもらう #」
- 音声認識結果・隣接文節の係り受けを用いた推定：
「えー不定ま季節問題についてえ三正反対にえ分かれてもらう #」

は文境界を示す

図 5 音声認識結果における文境界推定結果の例

Fig. 5 Examples of results of sentence boundary detection on automatic transcription.

るが、提案法（構成要素を用いた推定）ではこれを抑制できていることが分かる。

5. おわりに

本研究では、文節の局所的な係り受け情報に着目した、話し言葉のための節・文境界推定を提案した。提案手法は SVM によるテキストチャンカを利用し、形態素から段階的にチャンキングを行って節や文を推定する。チャンキングに際しては、認識誤りを含む音声認識結果からでも頑健に抽出できる特徴として隣接文節間の係り受けに着目し、これに基づく文節のまとめあげを行うことで節・文境界の候補の絞り込みを行う。『日本語話し言葉コーパス』（CSJ）を用いた評価実験において、文節や局所的係り受けの推定が音声認識誤りに対しても頑健であることが示された。これらを用いた段階的な節境界推定では、音声認識結果における評価で形態素列からの直接推定を上回る精度が得られた。さらに文境界推定に関しても、これらの素性の導入により精度が改善された。

謝辞 本研究の一部は、総務省戦略的情報通信研究開発推進制度（SCOPE）「音声認識技術を用いた会議録及び字幕の作成支援システム」により実施された。

参考文献

- 1) Garofolo, J., Laprun, C. and Fiscus, J.: The Rich Transcription 2004 Spring Meeting Recognition Evaluation, *Proc. ICASSP Meeting Recognition Workshop* (2004).
- 2) Renals, S., Hain, T. and Boulard, H.: Recognition and Understanding of Meetings: The AMI and AMIDA Projects, *Proc. ASRU* (2007).
- 3) Furui, S., Maekawa, K. and Isahara, H.: Toward the Realization of Spontaneous

- Speech Recognition – Introduction of a Japanese Priority Program and Preliminary Results, *Proc. Interspeech*, pp.518–521 (2000).
- 4) 富樫慎吾, 山口 優, 北岡教英, 中川聖一: 講義音声の認識・要約・インデックス化の検討, 情報処理学会研究報告, 2006-SLP-62-11 (2006).
 - 5) 根本雄介, 河原達也, 秋田祐哉: スライド情報を用いた言語モデル適応による講義の音声認識と字幕付与, 情報処理学会研究報告, 2007-SLP-66-16 (2007).
 - 6) 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).
 - 7) 南條浩輝, 秋田祐哉, 河原達也: 音声認識を利用した会議録・講演録の作成支援システムの設計と評価, 日本音響学会秋季研究発表会講演論文集, 1-7-13 (2005).
 - 8) 堀 智織, 古井貞照: 講演録作成を目的とした講演自動要約, 日本音響学会秋季研究発表会講演論文集, 2-1-10 (2001).
 - 9) 岩野公司, 広畑 誠, 新中庸介, 古井貞照: 重要文抽出による音声自動要約手法とその客観評価法についての検討, 電子情報通信学会技術研究報告, SP2005-20 (2005).
 - 10) 益岡隆志, 田窪行則: 基礎日本語文法改訂版, くろしお出版 (1992).
 - 11) 丸山岳彦, 柏岡秀紀, 熊野 正, 田中英輝: 日本語節境界プログラム CBAP の開発と評価, 自然言語処理, Vol.11, No.3, pp.39–68 (2004).
 - 12) Linguistic Data Consortium: Simple Metadata Annotation Specification Version 6.2 (2004).
http://www ldc.upenn.edu/Projects/MDE/Guidelines/SimpleMDE_V6.2.pdf
 - 13) 高梨克也, 内元清貴, 丸山岳彦: 『日本語話し言葉コーパス』マニュアル 『日本語話し言葉コーパス』における節単位認定 (2004).
 - 14) Wang, D., Lu, L. and Zhang, H.-J.: Speech Segmentation without Speech Recognition, *Proc. ICASSP*, Vol.1, pp.468–471 (2003).
 - 15) Srivastava, A. and Kubala, F.: Sentence Boundary Detection in Arabic Speech, *Proc. Interspeech*, pp.949–952 (2003).
 - 16) 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).
 - 17) Zimmerman, M., Hakkani-Tür, D., Fung, J., Mirghafori, N., Gottlieb, L., Shriberg, E. and Liu, Y.: The ICSI+ Multilingual Sentence Segmentation System, *Proc. Interspeech*, pp.117–120 (2006).
 - 18) Huang, J. and Zweig, G.: Maximum Entropy Model for Punctuation Annotation from Speech, *Proc. Interspeech*, pp.917–920 (2002).
 - 19) Wang, D. and Narayanan, S.S.: A Multi-pass Linear Fold Algorithm for Sentence Boundary Detection using Prosodic Cues, *Proc. ICASSP*, Vol.1, pp.525–528 (2004).
 - 20) 中嶋秀治, 山本博史: 音声認識過程での発話分割のための統計的言語モデル, 情報処理学会論文誌, Vol.42, No.11, pp.2681–2688 (2001).
 - 21) 野村和弘, 河原達也, 堂下修司: F0 パターンに基づく講義音声の文単位へのセグメンテーション, 電子情報通信学会技術研究報告, SP99-13 (1999).
 - 22) 田島幸恵, 難波英嗣, 奥村 学: 形態素解析器を利用した講演書き起こしの文境界検出について, 情報科学技術フォーラム講演論文集 (2003).
 - 23) 下岡和也, 南條浩輝, 河原達也: 講演の書き起こしに対する統計的手法を用いた文体の整形, 自然言語処理, Vol.11, No.2, pp.67–83 (2004).
 - 24) 下岡和也, 内元清貴, 河原達也, 井佐原均: 日本語話し言葉の係り受け解析と文境界推定の相互作用による高精度化, 自然言語処理, Vol.12, No.3, pp.3–17 (2005).
 - 25) 大庭隆伸, 堀 貴明, 中村 篤: チャンキングと逐次的係り受け解析に基づく話し言葉の文境界検出, 日本音響学会秋季研究発表会講演論文集, 2-2-5 (2006).
 - 26) 小橋修一, 山下洋一: 音声要約のための韻律情報を用いた文境界の自動決定, 日本音響学会秋季研究発表会講演論文集, 3-7-8 (2005).
 - 27) 大野誠寛, 松原茂樹, 柏岡秀紀, 稲垣康善: 節の始端検出に基づく独話文の係り受け解析, 情報処理学会研究報告, 2007-SLP-69-29 (2007).
 - 28) Roark, B., Liu, Y., Harper, M., Stewart, R., Lease, M., Snover, M., Shafran, I., Dorr, B., Hale, J., Krasnyanskaya, A. and Yung, L.: Reranking for Sentence Boundary Detection in Conversational Speech, *Proc. ICASSP*, Vol.1, pp.545–548 (2006).
 - 29) Kudo, T. and Matsumoto, Y.: Chunking with Support Vector Machines, *Proc. NAACL* (2001).
 - 30) 西川賢哉, 小椋秀樹, 相馬さつき, 小磯花絵, 間淵洋子, 土屋菜穂子, 斉藤美紀: 『日本語話し言葉コーパス』マニュアル 文節の仕様について Version 1.0 (2004).
 - 31) 黒橋禎夫, 長尾 真: 京都大学テキストコーパス・プロジェクト, 言語処理学会第 3 回年次大会, pp.115–118 (1997).
 - 32) 内元清貴, 丸山岳彦, 高梨克也, 井佐原均: 『日本語話し言葉コーパス』における係り受け構造付与, 国立国語研究所公開研究発表会予稿集 (2003).
 - 33) Kudo, T. and Matsumoto, Y.: Japanese Dependency Analysis using Cascaded Chunking, *Proc. CoNLL*, pp.63–69 (2002).
 - 34) 高梨克也, 丸山岳彦, 内元清貴, 井佐原均: 話し言葉の文境界—CSJ コーパスにおける文境界の定義と半自動認定—, 言語処理学会第 9 回年次大会, pp.521–524 (2003).

(平成 20 年 6 月 4 日受付)

(平成 20 年 11 月 5 日採録)



西光 雅弘

2004年岡山大学工学部通信ネットワーク工学科卒業。2006年京都大学大学院情報学研究科修士課程修了。現在、日本電気株式会社に所属。音声言語処理の研究に従事。



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。2007年日本音響学会粟屋潔学術奨励賞受賞。電子情報通信学会、日本音響学会、IEEE各会員。



高梨 克也

2000年京都大学大学院人間・環境学研究科博士課程単位取得退学。2000年1月～2005年3月独立行政法人情報通信研究機構専攻研究員を経て、現在京都大学学術情報メディアセンター特定助教。コミュニケーションの組織化を支える認知的・社会的プロセスの解明に従事。2002年人工知能学会研究奨励賞受賞。人工知能学会、言語処理学会、日本認知科学会、社会言語科学会、日本語用論学会各正会員。



尾嶋 憲治

2006年京都大学工学部電気電子工学科卒業。2008年京都大学大学院情報学研究科修士課程修了。現在、トヨタ自動車株式会社に所属。在学中、音声言語処理の研究に従事。



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年から1996年まで米国ベル研究所客員研究員。1998年からATR客員研究員。1999年から2004年まで国立国語研究所非常勤研究員。2001年から2005年まで科学技術振興事業団さきがけ研究21研究者。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。京大博士（工学）。1997年度日本音響学会粟屋潔学術奨励賞受賞。2000年度情報処理学会坂井記念特別賞受賞。情報処理学会連続音声認識コンソーシアム代表、IEEE SPS Speech TC委員、IEEE ASRU 2007 General Chair、言語処理学会理事を歴任。情報処理学会音声言語情報処理研究会主査。日本音響学会、人工知能学会各評議員。電子情報通信学会、言語処理学会、IEEE各会員。