係り受けとポーズ・フィラーの情報を用いた話し言葉の解析単位の段階的チャンキング

西光 雅弘[†]　　高梨 克也[‡]　　河原 達也[†‡]

[†] 京都大学 情報学研究科 知能情報学専攻

[‡] 京都大学 学術情報メディアセンター

e-mail: saikou@ar.media.kyoto-u.ac.jp

1　はじめに

近年,『日本語話し言葉コーパス』(Corpus of Spontaneous Japanese: 以下 CSJ と記す) に代表される，話し言葉を収集したコーパスの作成にともない，話し言葉を対象とした音声言語処理の研究が進展している．その応用として，音声認識技術を用いた字幕付与や会議録作成支援などが考えられる．

字幕付与や会議録作成支援といった音声言語処理においては，音声認識技術に加えて，高度な言語処理が必要不可欠である．例えば，字幕付与では，可読性の高い字幕を実現するための文分割，文圧縮といった処理が必要となる．従来，このような言語処理技術においては，文がその処理単位として用いられてきた．しかし，話し言葉においては，文の単位が自明ではない．CSJ では，話し言葉の文に相当する単位として,「節単位」とよばれる統語的・意味的な妥当性を備えた単位を定義しているが，その認定には高度な言語情報が必要である [1]．さらに，音声認識を用いた場合は，認識誤りが不可避であるため，それらの認定は一層困難である．そのため従来は，ポーズに代表される物理的な音響情報を用いることにより，その処理単位を認定することが大半である．しかし話し言葉では，ポーズにより認定した単位は文や節と必ずしも一致せず，均質な言語的まとまりになっているとは限らない．

本研究では，話し言葉音声を対象として，頑健に抽出できる特徴として直後の文節への係り受けとポーズ・フィラーの情報に着目し，それらを段階的に用いることにより，音声認識結果を文節，文の主題や述語・格要素にあたる構成要素，節にあたる意味的なまとまり (フレーズ) にチャンキングを行う手法を提案する．

2　音声言語処理における処理単位

CSJ で定義されている節単位は，統語的・意味的な妥当性を備えており，音声言語処理に有用な単位と考えられる．しかし，その認定には前後の形態素情報による節境界認定と，人手による修正が施されており [1]，音声認識誤りを含む表層的な言語情報を用いて，節単位を認定することは容易でない．そのため，音声言語処理においては，ポーズを用いて処理単位を認定することが大半である．しかし，ポーズにより認定した単位は文や節と必ずしも一致せず，一般に均質な言語的まとまりとはならない．ポーズにより認定した単位を図 1 に示す．

「/で/」
「/よく/(F あのー)/」
「/毎年/夏が/近付いてくると/」
「/沖縄の/」
「/(F あの)/」
「/キャッチコピー+」
「と/写真+」
「を/よく/雑誌の/裏とかに/出てますよね/」

('/' は CSJ で定義された文節の境界，'+' はポーズにより文節が分割された箇所)

図 1: ポーズにより認定した単位の例

本研究では，頑健に抽出できる特徴として，直後の文節への係り受けとポーズ・フィラーの情報に着目し，それらを段階的に用いてチャンキングを行うことにより，ボトムアップに処理単位を生成する．まず，係り受けの情報に着目し，文の主題や述語，格要素にあたるチャンクを生成する．本稿では，これを構成要素とよぶ．さらに，ポーズ・フィラーの情報によって，その構成要素をチャンキングし，より大きなチャンクを生成する．本稿ではこれをフレーズとよぶ．

生成されるこれらの処理単位は，高度な言語処理を要しない表層的な情報を用いて認定されており，音声認識誤りなどに対しても頑健であることが期待される．

3　係り受けとポーズ・フィラー情報の分析

3.1　分析データ

本研究では，CSJ のコア 199 講演を分析データとして，着目する係り受けとポーズ・フィラーの情報について分析する．

CSJ に付与された文節間係り受け構造は，京大コーパスの基準を原則とし，話し言葉特有の現象に対して新たな基準を設けたものを採用している．また，本研究で用いるポーズは，CSJ の転記基本単位の定義に従

「/視覚刺激として/」
「/入ってきた/画像情報から/」
「/音韻情報を/」
「/検出している/可能性が/」
「/あります/」
('/' は CSJ で定義された文節の境界，下線部が連体修飾する用言)

図 2: 連体修飾する用言の例

い，フィラーは，形態素短単位に F タグが付与された感動詞，D タグが付与された言いよどみとする．

CSJ コアにおいては，着目する情報以外に，話し言葉の文に相当する節単位，節単位認定の際の節境界ラベルなども付与されている．本研究では，節単位を文と定義する．節境界ラベルは，節境界検出プログラム CBAP-csj[2] を用いて自動検出されたものであり，それらは直後の切れ目の大きさによって，絶対・強・弱という 3 レベルに区分されている．自動検出された節境界のうち，絶対境界・強境界は基本的に文境界となり，弱境界は機能的に区切れていると判断される箇所のみが文境界となる．

3.2 係り受け情報

本研究では，文節のチャンキングという観点から，係り受け情報を用いて，構成要素となる単位の生成を考える．

音声認識誤りや係り受け解析誤りに対する頑健性を考慮すると，局所的な情報により，単位を認定することが望ましい．そこで，直後の文節への係り受けの有無に着目する．すなわち，直後の文節への係りを，直後の文節への依存性が強いと考え，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ることから，係り受け解析誤りの多くがそれ以外の場合であるので，頑健に抽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もつ．格要素は日本語の意味を解析する上で重要な役割を果たしているので，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．ただし，直後が述語となる格要素は，直後の文節への係り受けの有無だけでは認定できない．そこで，このような文節の直後も境界候補として検出するため，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ことを考える．

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，文節に含まれる語の品詞情報により可能である．そこで本研究では，用言を含む文節を述語と定義し，直後の文節に係る場合，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こととした．ただし，連体修飾する用言の場合，この規則を適用することによって，意味的に不自然なチャンクが生成される．具体的には，連体修飾する述語とその格要素は分離されるが，連体修飾する述語と被修飾語は結合されるといったことが起こる．そのような例を図 2 に示す．今回は，連体修飾する用言を含む述語に関しては，その直後を境界としない．この問題については，本来，連体修飾節と被修飾語の関係を分析して対処すべきであるが，その対処手法については今後の課題としたい．

本稿では，この操作により得られる単位を構成要素とよぶ．構成要素は，おおむね文に含まれる主題や述語，格要素などに対応している．

表 1: ポーズの出現箇所の分析

ポーズの出現箇所		頻度
節境界 (文境界を含む)	絶対境界	11362 (19%)
	強境界	6186 (10%)
	弱境界	12491 (21%)
言いよどみ		6906 (12%)
接続詞/接続表現		2008 (3%)
その他		21185 (35%)

表 2: 節境界におけるポーズの出現傾向

レベル区分	ポーズの出現率
絶対境界	95% (11362/11959)
強境界	75% (6186/8279)
弱境界	40% (12491/30942)

3.3 ポーズ・フィラー情報

ポーズは，音声認識誤りに対して頑健に検出可能な物理的な音響情報であることから，多くの音声言語処理において，その処理単位の認定に用いられてきた．そこで，ポーズが出現する箇所について分析し，意味的な妥当性を備えた文や節の境界とどの程度対応付けられるかを分析した．分析結果を表 1 に示す．

表 1 より，50%のポーズが節・文境界に対応していることがわかる．また，フィラーなどの言いよどみ，談話構造の把握に重要な役割をもつ接続詞においても，ポーズが多く出現することが確認された．一方で，35%のポーズに関しては，その出現位置に言語的な役割を確認することはできなかった．その他のポーズが生起する箇所は，格要素を備えた文節や強調部分など多岐にわたっていた．これは，ポーズの生起要因が多様であることを示している．

次に，3 レベルの節境界（絶対・強・弱）におけるポーズの出現傾向を分析した．分析結果を表 2 に示す．表 2 より，境界の切れ目が大きいほどポーズが出現しやすい傾向が確認された．これは，言語的な区切りとなる位置に出現するポーズを選択することにより，文境界となりうる絶対境界・強境界を高い精度で検出できるこ

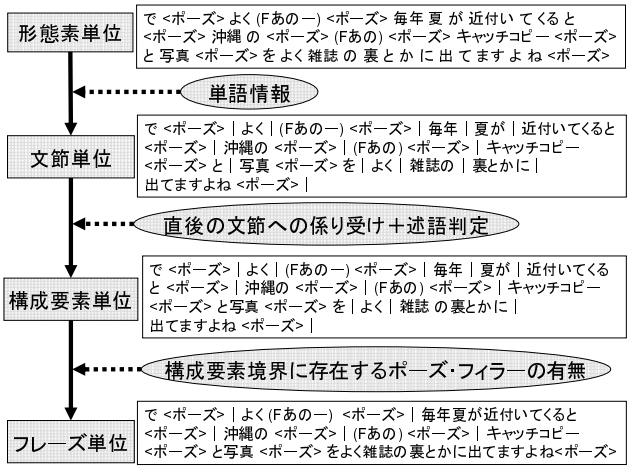

図 3: 提案する段階的チャンキング

とを示している．また，フィラーにおいても同様の傾向が確認されている [3]．その理由としては，人間は発話内容を節や句のような単位で生成しており，フィラーはその生成処理が何らかの理由で滞った場合に，出現するためと考えられている．またフィラー自体がポーズの一種とも考えられる．そこで本研究では，フィラーについてもチャンキングに有用な特徴と考え，用いることとした．

4 段階的チャンキングによる単位の生成

本研究では，3 章の分析に基づいて，係り受けとポーズ・フィラーの情報を段階的に用いることにより，ボトムアップに処理単位を生成する手法を提案する．提案手法の概要を図 3 に示す．

提案手法は，3 段階のチャンキングにより構成される．1 段目のチャンキングでは，語の最小単位である形態素を，CSJ で定義された文節にまとめあげる．本研究では，サポートベクトルマシン (SVM) に基づくテキストチャンカーである YamCha[4] を用いて，文節にまとめあげる．

次に，2 段目のチャンキングにおいて，直後の文節に係るか係らないかの判定と，直後の文節に係る場合，その文節が述語であるか否かの判定を行うことにより，構成要素を生成する．この操作では，直後の文節に係れば，直後の文節への依存性が強いと考えてチャンキングする．ただし，格要素は日本語の意味を理解する上で重要な役割を果たしていることから，格要素を備えた文節は述語に係るという特性を利用して，格要素を備えた文節と述語を分離する．直後の文節に係るか係らないかの判定は，SVM に基づく二値分類器によって実現する．本研究では，文節へのまとめあげと同様に，YamCha を用いる．

最後に，3 段目のチャンキングでは，2 段目で生成された構成要素を，ポーズ・フィラーの情報を用いてまとめあげる．具体的には，隣接する構成要素間にポーズもしくはフィラーがなければまとめあげ，それらがあれば境界とする．ポーズは，境界の切れ目が大きいほど，出現頻度が高くなる傾向があることから，文境界などを高い精度で検出することが期待される．この操作により生成された単位をフレーズとよぶ．

表 3: 文節へのまとめあげ精度

対象	再現率	適合率	F 値
書き起こし	97.9%	98.4%	0.982
音声認識結果	80.3%	78.4%	0.793

表 4: 直後の文節への係り受け解析精度

対象	再現率	適合率	F 値
書き起こし	91.6%	88.8%	0.902
音声認識結果	75.5%	74.3%	0.749

5 実験と評価

本実験では，CSJ 公開版の音声認識テストセット（計 30 講演）を評価に用い，これを除いたコアデータ（168 講演）を学習セットとした．テストセットの音声認識精度は 69.8%である．また，音声認識によるフィラーの検出（認識）率は，再現率 65.8%，適合率 55.8%，F 値 0.604 である．

まず，文節へのまとめあげの評価を行った．SVM の学習の素性には，前後 2 形態素の情報（表層表現，読み，品詞情報）を用いた．一般には，活用形なども素性として用いるが，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が得られにくいため用いていない．予備実験より，活用形を素性として用いない場合のほうが，用いる場合と同等もしくはそれ以上の精度であることを確認している．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．実験結果を表 3 に示す．従来より，書き言葉においては，SVM に基づく文節へのまとめあげが行われている．同様の手法で，話し言葉においても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．また，音声認識結果では書き起こしに比べて F 値が 19.2%低下しているが，これは単語誤り率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．

次に，直後の文節への係り受け判定の精度を評価した．SVM の学習の素性には，文節内の主辞・語形の単語情報を用いた．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 に与えるパラメータは，解析方向が Left to Right であることを除いて，文節へのまとめあげと同一である．係り受け解析の評価は文節単位で行う必要がある．その

表 5: 音声認識結果からの構成要素・フレーズの生成精度

対象	再現率	適合率	F 値
構成要素	87.4%	69.9%	0.777
フレーズ	77.8%	76.0%	0.769

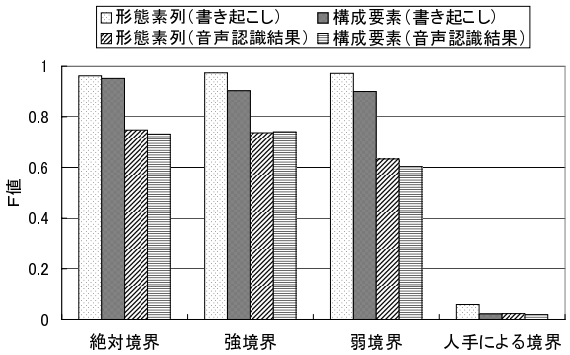

図 4: 節境界推定精度

ため，特に音声認識結果においては文節へのまとめあげ精度が問題となる．本実験では，文節単位で再度 DP マッチングを行った上で精度を推定した．実験結果を表 4 に示す．従来の話し言葉の係り受け解析精度は，下岡らの報告 [5] によると，open テストで 80.6%とある．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 程度の精度が実現できている．また，音声認識結果における F 値の低下は 17.0%であるので，認識誤りに対しても比較的頑健であるといえる．

最後に，提案手法に基づいて，音声認識結果より構成要素・フレーズを生成し，評価を行った．書き起こしに対する精度は表 5 に示す通り，F 値で 0.77 程度であり，認識誤りに比較的頑健であるといえる．

6 節・文境界推定

本稿で提案する構成要素・フレーズは，意味的なまとまりであるが，節や文とは必ずしも一致しない．そこで，構成要素・フレーズから節・文境界を推定することを試みた．

本研究では，CSJ で定義されている 3 レベルの節境界と，体言止めなどの人手により認定された境界を節境界と定義し，SVM に基づく多値分類器によりそれらを推定した．評価実験では，形態素列より推定する手法も実装し，比較した．学習の素性には，構成要素からの推定に構成要素の末尾 3 形態素，形態素列からの推定に前後 3 形態素の情報を用いた．実験結果を図 4 に示す．書き起こしでは，形態素列からの推定の方が若干高い精度となった．これは構成要素から直後の文節に係る節境界（連体節など）を推定できないためである．一方，音声認識結果においては，両手法が同等の精度であることを確認できた．また，体言止めなどの人手により認定された境界は，機械学習に基づく手法での認定が困難であることがわかった．

表 6: 文境界推定精度

対象	推定元	再現率	適合率	F 値
書き起こし	形態素列	83.0%	87.9%	0.854
	フレーズ	84.4%	87.5%	0.859
音声認識結果	形態素列	73.9%	81.7%	0.776
	フレーズ	64.3%	71.0%	0.675

次にフレーズより文境界を推定した．文境界推定は節境界推定と同様に，SVM に基づく手法により実現する．評価実験では，形態素列より文境界を推定する手法 [5] も実装し，比較した．学習の素性には，下岡らの手法 [5] を参考にして，フレーズの末尾 3 形態素の情報，節境界推定により得られた節境界，1 講演で正規化したポーズ長を用いた．実験結果を表 6 に示す．書き起こしでは，フレーズからの推定が高い F 値を得た．一方で，音声認識結果では，形態素列からの推定の方が高い F 値となった．これは，フィラーの検出精度の影響により，フレーズにおける文境界の再現率が低いことが原因と考えられる．

7 おわりに

本研究では，音声認識結果から頑健に抽出できる特徴として，直後の文節への係り受けとポーズ・フィラーの情報に着目し，それらを用いて段階的にチャンキングを行うことにより，文の主題や述語・格要素にあたる構成要素，節にあたる意味的なまとまり（フレーズ）を生成する手法を提案した．CSJ を用いた評価実験により，音声認識誤りに対しても頑健に，それらの処理単位が生成できることを確認した．また，処理単位より節・文境界が一定の精度で推定できることを確認した．今後は，字幕付与や会議録修正支援などの音声言語処理に提案手法を適用し，評価を行う予定である．

参考文献

[1] 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, 1997.

[2] 丸山岳彦, 柏岡秀紀, 熊野正, 田中英輝. 日本語節境界検出プログラム CBAP の開発と評価. 自然言語処理, Vol. 11, No. 3, pp. 39–68, 2004.

[3] 渡辺美知子, 伝康晴, 広瀬啓吉, 峯松信明. フィラー出現確率予測における節の種類と接続長. 日本音響学会秋季研究発表会講演論文集, 3-1-11, 2005.

[4] T. Kudo and Y. Matsumoto. Chunking with support vector machines. In *Proc. of the 2nd Meeting North American Chapter of the Association for Computational Linguistics*, 2001.

[5] 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.