社団法人 情報処理学会 研究報告
IPSJ SIG Technical Report

2006－SLP－61 (4)
2006／5／11

隣接文節間の係り受け情報に着目した話し言葉のチャンキングの評価

西光 雅弘　河原 達也　高梨 克也

京都大学 情報学研究科 知能情報学専攻
〒606-8501 京都市左京区吉田本町

あらまし

会議録作成支援や字幕付与などの音声言語処理を指向して、話し言葉を“適当な”単位に区分化することを考える。従来、話し言葉音声では、ポーズに基づいて発話単位を設定することが多いが、ポーズが文や節の境界と対応しない場合が多く、均質な言語的まとまりにならない。一方、話し言葉の節や文の境界を機械学習に基づいて検出する方法も研究されているが、音声認識結果に対してはF値が70%台であり、誤検出箇所に関して意味づけを見いだすのが難しい。これに対して本研究では、話し言葉の非定型性や音声認識誤りに頑健であると考えられる局所的な特徴、具体的には隣接文節間の係り受けに着目して、チャンキングを行う。述語判定や係り受けタイプ判定を組み合わせることにより、文の主題や述語・格要素におおむね対応する「構成要素」を抽出する。『日本語話し言葉コーパス』(CSJ)で分析・評価を行った結果、隣接文節間に絞ることで係り受け解析が高い精度でできること、構成要素に基づいて音声認識結果に対してもより頑健に節境界を検出できることが示された。

A Evaluation of a Cascaded Chunking Method of Spontaneous Japanese using Local Bunsetsu Dependency

Masahiro Saikou　Tatsuya Kawahara　Katsuya Takanashi
School of Informatics, Kyoto University, Kyoto 606-8501, Japan

Abstract The paper addresses chunking of spontaneous Japanese oriented for speech summarization and closed caption generation. Conventionally, inter-pause unit (IPU) has been a basic unit in annotation and processing of spontaneous speech, but pauses are not necessarily related with sentence or clause boundaries. On the other hand, automatic detection of sentence or clause boundaries based on machine learning techniques is not so reliable for speech recognition results. For robust chunking against ill-formedness in spontaneous speech and speech recognition errors, we focus on local bunsetsu dependencies. Combined with detection of predicates and classification of dependencies, we define a unit named “constituents”, which apparently corresponds to subjects, predicates and case frames. With analysis and evaluation using the Corpus of Spontaneous Japanese (CSJ), it is shown that we can perform more reliable dependency structure analysis by focusing on adjacent bunsetsus and more robust detection of clause boundaries by extracting constituents.

1 はじめに

従来の自然言語処理技術においては、文が処理の基本単位として用いられてきた。しかし、話し言葉においては、文の単位が自明でない。『日本語話し言葉コーパス』(CSJ)では、話し言葉の「文」に相当する単位として節単位を定義しているが、その認定には、前後の形態素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、頑健に節単位を認定することは容易でない。そのため、従来の音声言語処理においては、ポーズを用いて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節や文と必ずしも一致せず、流暢な箇所では全く区切れない反面、言いよどみの多い箇所では数多く途切れる結果となる。

そこで我々は、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の情報に着目し、これをポーズ・フィラーの情報と組み合わせて、段階的にチャンキングを行う方式を提案している[2]。この処理の概要を図1に示す。まず、形態素列を文節にまとめあげる。次に、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、文の主題や述語・格要素にあたる構成要素にチャンキングする。最後に、構成要素間に存在するポーズ・フィラーを用いて、さらに大きなチャンクであるフレーズを生成する。

本稿では、特に構成要素を生成する過程で用いる局所的な言語処理について詳しく述べるとともに、CSJのコアを用いて評価した結果を報告する。最後に、提案手法により生成される字幕の例を示す。

2 局所的な係り受け情報の利用

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、二文節に含まれる単語情報から、「格要素と述語」や「連体修飾節の述語と被修飾語」といった二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」[3]の基準に準拠しながら、話し言葉特有の現象に対して新たな基準を設けている[4]。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・言いよどみを含まない文節については、直後の文節に係るものとした。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、係り受け関係にある二文節の間にフィラーが存在しても、二文節は隣接しているとみなす。

話し言葉においては、言いよどみなどの現象や、発話途中でのプランの変更により、(主語と述語間のような)長い距離の係り受けの関係が書き言葉ほど厳密に保持されないことが多い。このような話し言葉に対しても頑健な言語的特徴として、隣接文節間の係り受けの有無に着目する。すなわち、直後の文節に係るということを、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、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。意味解析においても有用な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。

一方で、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されないので、述語判定の処理を導入する。本研究では、用言(動詞・形容詞・助動詞)を含む文節を述語と定義し、直後の述語に係る場合は、当該文節と述語文節の間も境界候補とする。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、以下の例のような意味的に不自然なチャンクが生成される。

/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)

今回は、連体修飾する用言を含む述語に関しては、格要素を備えた文節の直後を境界としないことで対処する。

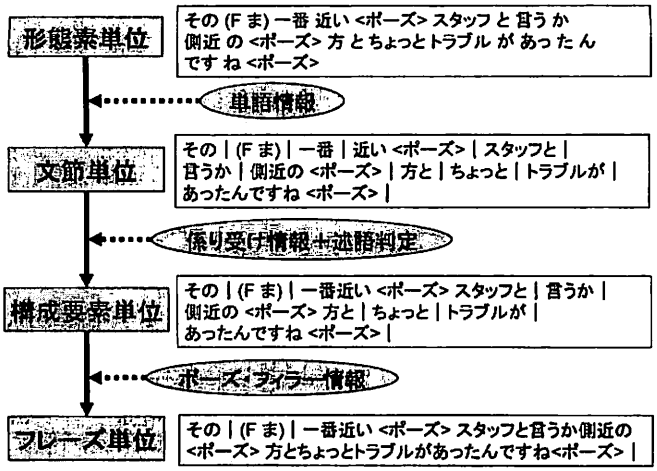

図 1: 局所的な係り受け情報に着目した話し言葉のチャンキング

1 章で述べたように、以上の処理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。

3 実装と評価

3.1 文節・構成要素のチャンキング精度

以上で述べた処理を実装し、評価を行った。

本稿では、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。

まず、文節へのまとめあげには、サポートベクトルマシン (SVM) に基づくテキストチャンカである YamCha[5] を用いた。SVM の学習の素性には、前後 2 形態素の情報（表記・読み・品詞）を用いた。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。

なお予備実験において、学習テキストに含まれるフィラーを削除しない方が若干精度が高いことを確認したので、フィラーは（F タグ付きで）残したまま解析している。これは、フィラーの出現位置にも有意な傾向があり、フィラーがチャンキングの指標として有用であることを示している。

実験結果を表 1 に示す。従来、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。

表 1: 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 2: 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

表 3: 構成要素の生成精度

対象	再現率	適合率	F 値
音声認識結果	87.4%	69.9%	0.777

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 値が約 19%低下しているが、これは単語誤り率よりかなり小さい。

次に、直後の文節に係るか係らないかの判定を、別の SVM に基づく二値分類器によって実現した。ここでは、文節へのまとめあげと同様に、YamCha を用いた。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。そのため、特に音声認識結果に

表 4: 節境界ラベルの例

区分	節境界ラベル
絶対境界	文末、文末候補、と文末
強境界	並列節「ガ」「ケド」「シ」など
弱境界	タリ節、条件節「ナラバ」「レバ」など

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 2 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 [6] によると、open テストで 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 90%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 値の低下は約 17%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、述語判定と連体修飾の例外処理を実装・統合して、構成要素を生成した。音声認識結果に対する精度は表 3 に示す通り、F 値で 78%程度である。

3.2 節境界の推定精度

提案手法により生成される構成要素は、意味的なまとまりであるが、CSJ で定義されている節などとは必ずしも一致しない。そこで、構成要素を基に節境界の推定を行った。

CSJ においては、節境界検出プログラム CBAP-csj[7] を用いて節境界を自動検出した後、人手により修正を施すことで、話し言葉の文に相当する節単位を認定している [1]。自動検出した節境界には、節境界ラベルが付与されており、そのラベルは直後の切れ目の大きさによって、絶対境界・強境界・弱境界という 3 レベルに区分されている。節境界ラベルの例を表 4 に示す。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人手で判断される箇所のみが文境界となる。

本実験では、CBAP-csj で自動認定された 3 レベルの節境界と、それ以外に人手による修正の際に認定された境界（体言止めなど）を YamCha を用いて推定した。この場合、計 5 クラス（4 種の境界と境界以外）を識別する問題となるが、ここでは各クラス対の組合せを識別する N*(N-1)/2 種類の識別器を作成し、最終的にそれらの多数決で決定する pairwise 法を用いた。

表 5: 節境界推定における活用型・活用形の影響

推定元	素性	再現率	適合率	F 値
形態素列	表層・読み・品詞	95.3%	97.1%	0.962
	[ASR]	70.1%	80.0%	0.747
	+活用型・活用形	98.9%	99.4%	0.992
	[ASR]	59.6%	43.7%	0.505
構成要素	表層・読み・品詞	96.0%	96.7%	0.963
	[ASR]	69.1%	81.1%	0.746
	+活用型・活用形	99.2%	99.0%	0.991
	[ASR]	69.2%	80.9%	0.746

表 6: 節境界推定精度

推定元	境界の種類	再現率	適合率	F 値
形態素列 (7046)	絶対境界 (1784)	95.3%	97.1%	0.962
	[ASR]	70.1%	80.0%	0.747
	強境界 (1066)	96.1%	98.8%	0.974
	[ASR]	68.5%	79.4%	0.736
	弱境界 (3975)	96.9%	97.5%	0.972
	[ASR]	63.1%	63.6%	0.634
	全境界 (7046)	93.4%	97.5%	0.954
	[ASR]	67.0%	73.0%	0.699
構成要素 (6699)	絶対境界 (1776)	96.0%	96.7%	0.963
	[ASR]	69.1%	81.1%	0.746
	強境界 (1064)	96.2%	98.8%	0.974
	[ASR]	70.9%	83.8%	0.768
	弱境界 (3640)	89.4%	97.4%	0.932
	[ASR]	59.1%	65.5%	0.622
	全境界 (6699)	90.6%	98.6%	0.945
	[ASR]	68.8%	80.1%	0.740

（上段が書き起こし、下段 [ASR] が音声認識結果）

また、節や文境界を推定する手法としては、形態素列の局所的な情報に基づいて、CBAP-csj のように規則ベースで行ったり [1]、あるいは SVM を用いて機械学習を行うもの [6] が、これまでに CSJ に対して適用されている。

そこで本実験では、形態素列から節境界を推定する手法も YamCha を用いて実装し、比較した。SVM の学習の素性としては、構成要素からの推定では構成要素境界の前後 3 形態素、形態素列からの推定では前後 3 形態素の情報を用いた。

まず、SVM の学習に与える単語の素性として、活用型・活用形を利用することの影響を調べた。絶対境界に対する結果を表 5 に示す。正しい形態素情報が付与された書き起こしでは、活用形の情報を追加することで、終止形と連体形の曖昧性がなくなり、精度の改善がみられている。しかしながら、音声認識結果においては、形態素列から絶対境界を推定する場合、活用型・活用形を含めることにより、推定精度が大きく低下している。音声認識の際の終止形

と連体形の混同により、精度が低下したためである。他の境界についてはそれほど有意な差はなかったが、上記の違いは大きいので、活用型・活用形を含めない素性を用いて全体の評価を行った。

その結果を表6に示す。書き起こしからの絶対境界・強境界の推定に関しては、形態素列からでも構成要素からでも同等の精度となっている。ただし、弱境界に関しては、形態素列から推定する方が高いF値となった。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り、この点を考慮すると、両手法はほぼ同等の精度といえる。表中の括弧内の数字は、検出可能な節境界の総数を示しており、境界候補を絞りこんだ形である構成要素から検出可能な節境界は少ないものの、そこから高い精度で節境界を抽出できていることがわかる。音声認識結果[ASR]に対しては、4種各々の境界の精度については大きな差はないが、境界種間の混同を許容した場合の全境界の精度では、構成要素から推定する方が高いF値となった。これは、主に前述の通り、構成要素を介した方が適合率が高くなるためであり、提案手法の有効性を示すものである。

4 字幕生成の例

本研究で提案した構成要素・フレーズの応用例として、講演などの字幕の作成が考えられる。そこで、字幕放送の経験則[8]を参考にして、1行15文字以内の字幕を生成した。具体的には、構成要素単位で15文字になるまで文字列を結合し、異なるフレーズに含まれる構成要素は結合しないことで字幕を生成する。これにより生成される字幕を図2に示す。また比較のため、形態素、文節、ポーズ・フィラーにより認定した単位それぞれについても、同様に15文字になるまで結合し字幕とした例を図3、4、5に示す。ただし、ポーズ・フィラーにより認定した単位においては、一つのまとまりが15文字以上のものが多く存在するので、そのようなまとまりについては15文字で分割している。

形態素情報のみに基づく字幕（図3）においては、「門/まで」のように自立語と助詞などの付属語が分割されている。文節の単位を用いた字幕（図4）においては、「僕らの/顔」のように修飾語と被修飾語のまとまりが考慮されない。ポーズ・フィラーにより認定した単位を用いた字幕（図5）では、意味的なまとまりを多く生成できているが、しばしば非常に単位が長くなり、均質な字幕となっていない。

これに対して、構成要素・フレーズを用いた字幕は、均質的でかつ意味的なまとまりとなる字幕を生成できている。

実際に付与される字幕は、音声の書き起こしをそのまま提示するのではなく、人手によって語の言い換えや不要な語句の削除を行うことが一般的である。これらのより高度な言語処理については今後の課題としたい。

5 おわりに

話し言葉においても頑健に抽出できる特徴として、直後の文節への係り受けの情報に着目し、これを用いて段階的にチャンキングを行う方法を提案・実装した。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チャンキングが行えることを確認した。今後は、字幕付与や会議録作成支援などに提案手法を適用していく予定である。

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, 2003.

[2] 西光雅弘, 高梨克也, 河原達也. 係り受けとポーズ・フィラーの情報を用いた話し言葉の段階的チャンキング. 電子情報通信学会技術研究報告, SP2005-137, NLC2005-104 (SLP-59-48), 2005.

[3] 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.

[4] 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. 平成15年度国立国語研究所公開研究発表会予稿集, 2003.

[5] 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.

[6] 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.

[7] 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.

[8] 福島孝博, 江原暉将, 白井克彦. 文単純化のための文字数圧縮規則. 言語処理学会第5回年次大会, pp. 221–224, 1999.

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
待ってるんですけれども
僕や父が夜遅く帰ってきて
門のところに来ますと一応
彼らは門まで
出迎えてくれるんですが
僕らの顔を見ると
お前らかみたいな顔して
また奥に
引き篭ってしまうっていうような
本当に番犬としては
何の役にも立たない犬に
育ってしまいました

図 2: 構成要素・フレーズを用いて生成した字幕の例

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れども僕
や父が夜遅く帰ってきて門の
ところに来ますと一応彼らは門
まで出迎えてくれるんですが僕ら
の顔を見るとお前らかみたいな顔
してまた奥に引き篭ってしまう
っていうような本当に番犬として
は何の役にも立たない犬に育って
しまいました

図 3: 形態素を用いて生成した字幕の例

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
待ってるんですけれども僕や父が
夜遅く帰ってきて門のところに
来ますと一応彼らは門まで
出迎えてくれるんですが僕らの
顔を見るとお前らかみたいな
顔してまた奥に
引き篭ってしまうっていうような
本当に番犬としては何の役にも
立たない犬に育ってしまいました

図 4: 文節を用いて生成した字幕の例

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
待ってるんですけれども
僕や父が夜遅く帰ってきて
門のところに来ますと一応
彼らは門まで出迎えてくれるんで+
すが
僕らの顔を見ると
お前らかみたいな顔して
また奥に引き篭ってしまう
っていうような
本当に番犬としては何の役にも立+
たない
犬に育ってしまいました

('+' は一つのまとまりが 15 文字以上のため分割した箇所)

図 5: ポーズ・フィラーを用いて生成した字幕の例