

## 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

西光 雅弘<sup>†</sup> 高梨 克也<sup>††</sup> 河原 達也<sup>†††</sup>

<sup>†</sup> 京都大学 情報学研究科 知能情報学専攻

<sup>††</sup> 京都大学 学術情報メディアセンター

<sup>†††</sup> 〒 606-8501 京都市左京区吉田二本松町

あらまし 従来の音声言語処理においては、ポーズにより処理単位を設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チャンキングを行う手法を提案する。まず、隣接文節間の係り受けの情報を用いて、構成要素にチャンキングする。この構成要素をポーズまたはフィラーの有無によってチャンキングすることにより、フレーズを生成する。『日本語話し言葉コーパス』(CSJ)で分析を行った結果、構成要素は主題や述語、格要素にあたる意味的なまとまりにおおむね対応しており、節境界の93%を検出できること、また、フレーズの過半数が節境界に対応し、特に絶対境界・強境界の94%を検出できることがわかった。

キーワード 話し言葉, 係り受け, ポーズ, フィラー, チャンキング

## Cascaded Chunking of Spontaneous Japanese using Bunsetsu Dependency and Pause Information

Masahiro SAIKO<sup>†</sup>, Katsuya TAKANASHI<sup>††</sup>, and Tatsuya KAWAHARA<sup>†††</sup>

<sup>†</sup> School of Informatics, Kyoto University

<sup>††</sup> Academic Center for Computing and Media Studies, Kyoto University

Kyoto 606-8501, Japan

**Abstract** This paper proposes a cascaded chunking method of spontaneous Japanese using bunsetsu dependency, pause and filler information. The proposed method generates two new units for spoken language processing: constituent and phrase. A constituent is generated by chunking bunsetsus using the adjacent bunsetsu dependency. A phrase is generated by chunking the constituents using pause and filler information. We made experiments using the Corpus of Spontaneous Japanese (CSJ). Most of constituents correspond to semantic chunks such as subjects, predicates and case frames, and 93% of clause boundaries is detected. More than half of phrases correspond to the clause boundaries, and cover 94% of sentence boundaries.

**Key words** spontaneous Japanese, bunsetsu dependency, pause, filler, chunking

### 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、自発音声(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会議録作成支援や字幕付与などが考えられる。

会議録作成支援や字幕付与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、音声認識結果の効率的な修正や、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、音声認識誤りが不可避であるため、音声認識結果よりそれらを認定することは容易でない。そのため、音声言語処理技術では、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

従来の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。話し言葉音声では文の単位は明確でないが、文のような意味的なまとまりとなる単位を認定する必要がある。

意味的なまとまりとなる言語単位として文や節がある。CSJでは、話し言葉の文に相当するものとして、一つ以上の節から構成される節単位を定義している。節単位の例を図1に示す。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、頑健に節単位を認定することは容易でない。実際、節単位認定の際に用いられている節境界検出プログラムCBAP[2]は、形態素誤りに対して大きく精度が低下することが報告されている[3]。そのため、音声認識誤りが不可避である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2に示す。

本研究では、表層的に抽出できる特徴として係り受けとポーズ・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。このチャンクは、節単位にほぼ対応することが望まれるが、必ずしも一致しないので、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

/例えば/母さんと/お婆ちゃんが/家に/帰ってくと/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/(F その)/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/(F あー)/お前らかみたいな/顔/して/ また/奥に/引き籠ってしまうっていうような/(D 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CSJで定義された文節の境界)

図1 節単位の例

「/例えば/母さん+」  
「と/お婆ちゃんが/家に/帰ってくと/」  
「/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/」  
「/(F その)/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/(F あー)/」  
「/お前らかみたいな/顔/して/」  
「/また/奥に/引き籠ってしまう+」  
「/っていうような/(D 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2 ポーズにより認定した単位の例

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. チャンキングに用いる特徴の検討

### 3.1 分析データ

本研究では、係り受けなどの情報が人手により付与されているCSJのコア199講演を分析データとする。この構成を表1に示す。

CSJにおいては、節境界検出プログラムCBAP-csj[2]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]。本研究ではCSJで定義された節単位を文と定義する。自動検出した節境界には節境界ラベルというタグが付与されており、それらタグ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2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判断される箇所のみが文境界となる。機能的に区切れているかどうかの判

表 1 分析データの構成

総講演数	199	
総形態素数(短単位)	496,617	
総文節数	216,614	
総節数	絶対境界	12,128
	強境界	8,279
	弱境界	32,477
総文数	20,356	

表 2 節境界ラベルの例

レベル区分	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

断は、話し言葉特有の現象である「体言止め」や「倒置」などの箇所の修正と共に、人手により行われる。

CSJにおける文節間係り受け構造は、京大コーパス[4]の基準を原則として、話し言葉特有の現象に対しては新たな基準を設けたものを採用している[5]。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のための左係りについては、係り受け情報を修正する。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。そこで本研究では、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。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。すなわち、係り受け関係にある二文節の間にフィラーのみで構成される文節が存在していたとしても、二文節は隣接しているのみとする。

本研究で用いるポーズは、CSJの転記基本単位の定義[6]に従い、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。

### 3.2 係り受け情報

係り受け情報は、ある文節が最も依存している他の文節に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。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を生成することを考える。

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を用いることとした。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考える。また、日本語では、多くの文節が直後の文節に係ることから、係り受け解析誤りによる影響に対しても頑健であることが期待される。さ

「/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/」  
 「/(F その)/」  
 「/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/」  
 「/(F あー)/」  
 「/お前らかみたいな/顔/して/」  
 「/また/」  
 「/奥に/」  
 「/引き籠ってしまうっていうような/」  
 「/(D な)/」  
 「/本当に/」  
 「/番犬としては/」  
 「/何の/役にも/立たない/犬に/育ててしまいました/」

(/)はCSJで定義された文節の境界

図 3 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

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理解する上で重要な役割を果たしているため、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。直後の文節の係り受けの有無により認定した単位の例を、図3に示す。図3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、単位として認定されていることを確認できる。一方で、直後が述語となる格要素などは、直後の文節への係り受けの有無だけでは認定されていない。そこで、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。

日本語文の文節において、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(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、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4に示す。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。この問題につ

「/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 4 連体修飾する用言の例

「/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/」  
 「/待ってるんですけども/」  
 「/僕や/父が/」  
 「/夜/」  
 「/遅く/」  
 「/帰ってきて/」  
 「/門のところ/」  
 「/来ますと/」  
 「/一応/」  
 「/(F その)/」  
 「/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/」  
 「/(F あー)/」  
 「/お前らかみたくない/顔を/」  
 「/して/」  
 「/また/」  
 「/奥に/」  
 「/引き籠ってしまうっていうような/」  
 「/(D 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 CSJ で定義された文節の境界)

図 5 構成要素の例

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。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の例を図 5 に示す。本稿では、この操作によ

表 3 ポーズの出現箇所分析

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 4 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)

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。

### 3.3 ポーズ情報

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調べた。分析データに出現したポーズの総数は 60,138 であった。分析結果を表 3 に示す。

表 3 より、50%が文境界、節境界に対応することがわかる。また、フィラー・言い直しなどの話し言葉特有の現象である言いよどみにおいても、ポーズが頻出する傾向が見られた。他に、節境界ではないが、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。文境界、節境界に接続詞を加えると、53%のポーズが言語的な区切りの位置に出現していることが確認された。一方で、残り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。そのため、ポーズのみで認定した単位は、均質な言語的まとまりになっているとは限らない。

次に、節境界においてどの程度ポーズが出現しているかを調べた。分析結果を表 4 に示す。表 4 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を確認できる。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とを示している。

### 3.4 フィラー情報

フィラーは言いよどみの一種であり、その頻度は音声認識精度低下の主要因の一つである [7]。一方で、フィラーの出現位置には、意味的なまとまりとなる節境界と相関があることが知られている [8]。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

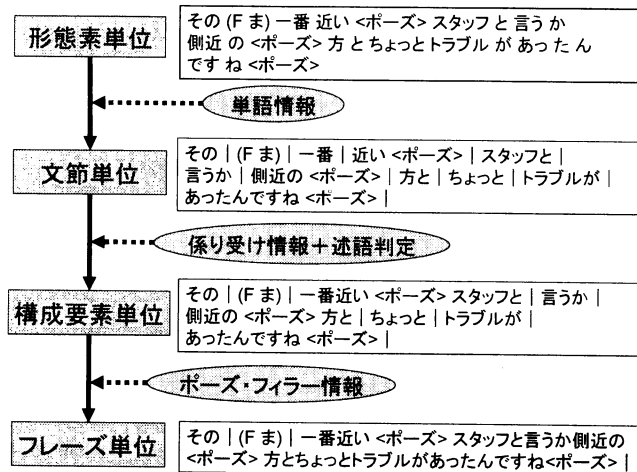


図6 提案する段階的チャンキング

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、フィラーをチャンキングに有用な特徴と考え、用いることとした。

#### 4. 提案手法

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、話し言葉音声を頑健に均質な言語的まとまりにチャンキングする手法を提案する。提案手法の概要を図6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[9]にまとめあげる。本研究では、サポートベクトルマシンに基づくテキストチャンカーであるYamCha[10]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、基本的に述語への係り受けを除いて、直後の文節に係る場合に、それらをまとめあげる。すなわち、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。

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、文境界となりうる境界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。フレーズ

「/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/」  
 「/(F その)/」  
 「/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/」  
 「/(F あー)/」  
 「/お前らかみたくない/顔/して/」  
 「/また/奥に/引き篋てしまうっていうような/」  
 「/(D な)/」  
 「/本当に/番犬としては/何の/役にも/立たない/犬に/育ってしましました/」

(/)はCSJで定義された文節の境界)

図7 フレーズの例

の例を図7に示す。

#### 5. 評価

最初に、文節まとめあげの評価を行った。分析データ199講演のうち、189講演を学習データ、10講演をテストデータとして、文節まとめあげの実験を行ったところ、F値で99の精度を得た。実験に用いた素性は前後2形態素の表記、読み、品詞情報、活用の種類、活用形である。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、ラベリングスキームにはIOEを用いた。なお、学習テキストに含まれるフィラーは削除していない。これは、フィラーを除いた学習テキストに比べ、フィラーを含んだ学習テキストの方が若干精度が高かった(F値で0.3)ためである。これは、フィラーの出現位

表5 構成要素の末尾形態素(上位5位)

末尾形態素	頻度
格助詞「が」	10993
係助詞「は」	10562
格助詞「を」	10193
格助詞「に」	9776
接続助詞「て」	8172

置にも何らかの傾向があることを示しており、フィラーを指標として用いることを支持するものである。

次に、構成要素、フレーズに関する評価を行った。本研究では、3種類の節境界ラベルを意味的な妥当性をもつ指標と考え、提案手法により生成した構成要素、フレーズの境界と節境界の共起を調べることで、意味的な妥当性を検証する。また、構成要素に関しては、フレーズの末尾形態素を分析することにより、格要素となる文節の直後が境界となっているかを検証する。

まず、分析データに提案手法を適用することにより、140,534の構成要素、57,776のフレーズを得た。また、節境界ラベルは52,884であり、そのうち49,385が構成要素の境界と一致していた。すなわち、93%(49385/52884)の節境界が、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、検出できることが示された。次に、節境界に一致しない構成要素91,149の末尾形態素を分析した。末尾形態素の上位5形態素を表5に示す。構成要素の末尾形態素として、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%、節境界を除いた68%については、格要素に対応していると考えられる。残りの末尾形態素の代表例として、接続詞「で、しかし」、副詞「ちょっと、やっぱり」などが見られる。

フレーズは、構成要素をポーズ・フィラーによりまとめあげたものである。そのため、フレーズはポーズ・フィラーの出現箇所に影響を受け、節や格要素などの様々なレベルの言語単位が混在したものとなる。分析の結果、58%(33673/57776)のフレーズが、節境界と一致していた。節境界と一致したフレーズのうち、18,973が文境界となりうる絶対境界・強境界であった。これは、絶対境界・強境界の94%(18973/20228)にあたる。ただし残りの弱境界に関しては、48%(14700/30942)の検出にとどまっている。これは、本稿で提案しているフレーズと弱境界が言語的に異なる性格を持つことを示唆しているが、この点については引き続き調査している。

## 6. おわりに

本研究では、話し言葉を対象として、頑健に検出可能で均質な言語的まとまり(チャンク)の生成手法を提案した。提案手法は、係り受けとポーズ・フィラーの情報を段階的に用いてチャンキングを行う。係り受けの情報によりまとめあげられた構成要素のうち、35%が節境界と一致し、44%が格要素などを備えた境界であることが確認された。フレーズは、構成要素をポーズ・フィラーを用いてまとめあげたものである。ポーズ・フィ

ラーは境界の切れ目が大きいほど出現しやすい傾向があり、生成されたフレーズのうち、58%が節境界と一致した。特に、絶対境界・強境界については、94%検出できた。このように単純な特徴で均質な言語的まとまりを構成しながら、境界の大部分を捕捉できることが示された。

今後は、会議録修正支援や字幕付与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。

## 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] 柏岡秀紀. 独話データのポーズ単位を利用した節境界判定. 情報処理学会研究報告, 2005-SLP-57-15, 2005.
- [4] 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.
- [5] 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.
- [6] 小磯花絵, 前川喜久雄. 『日本語話し言葉コーパス』の設計の概要と書き起こし基準について. 情報処理学会研究報告, 2001-SLP-36-1, 2001.
- [7] T. Kawahara, H. Nanjo, T. Shinozaki, and S. Furui. Benchmark test for speech recognition using the corpus of spontaneous Japanese. In *Proc. SSPR*, pp. 135-138, 2003.
- [8] 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.
- [9] 西川賢哉, 小椋秀樹, 相馬さつき, 小磯花絵, 間瀬洋子, 土屋菜穂子, 斉藤美紀. 文節の様について version 1.0. 『日本語話し言葉コーパス』マニュアル, 2004.
- [10] 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.