

音声認識のための文節構造モデルとその制約について

伊藤 彰則[†] 牧野 正三[‡]

[†] 山形大学工学部 [‡] 東北大学大型計算機センター

音声対話処理など、自由発話を主に扱うタスクに用いるための文節モデルを作成した。このモデルは、日本音響学会連続音声データベースの対話テキストの分析に基づき、そこに出現する表現をカバーするように作成した有限オートマトンである。現在のモデルの規模は、実質語 3386、機能語 615 である。

次に、このモデルを実際に認識に用いることを考慮し、各種の確率モデルとの併用を検討した。有限オートマトンの遷移確率を用いる方法と、N-gram を併用する方法を検討し、bigram との併用で比較的良好な結果を得た。

また、有限オートマトンでは表現しにくい各種の制約について、属性による表現を試みた。属性付きの有限オートマトンをそのまま使うのではなく、属性なしのものに自動変換して使うことを試みた。この手法により、従来有限オートマトンで制御される認識手法をそのまま使うことが可能になる。今回作成した文節モデルにいくつかの属性を導入し、その効果について調査した。

On a Bunsetsu structure model with several constraints for speech recognition

Akinori ITO[†] Shozo MAKINO[‡]

[†] Faculty of Engineering, Yamagata University [‡] Computer Center, Tohoku University

This paper describes a bunsetsu (phrase) model for Japanese spontaneous speech recognition. This model is represented as a finite automaton which covers almost all expressions in dialog transcriptions in the ASJ continuous speech corpus. This model contains 3386 conceptual words and 615 functional words.

Next, stochastic language models are combined with the bunsetsu model. Two types of stochastic models are investigated: a stochastic regular grammar and a N-gram model. When combined with the bunsetsu model, a bigram model gets smaller perplexity.

Finally, several attributes are introduced the bunsetsu model to express constraints between distant words in a phrase. The finite automaton model with attributes is automatically converted to a finite automaton without attributes, which can be easily used for conventional speech recognition schemes.

1 はじめに

音声対話は音声研究の次のターゲットとして盛んに研究が行われている。朗読された音声の認識と異なり、対話音声においては、口語表現・倒置表現・間投詞・言いよどみ／言い直しなどの現象が発生し、また発声もあいまいになりやすい。そのため、対話音声の認識

においては、これらの現象に対処するような認識法を用いることが不可欠である。

筆者らは、このような対話音声を対象とした音声認識を行なうため、特に「文節構造」に着目した文法の整理を行ってきた^[1]。この研究では、人間同士の対話を分析し、その文節構造の有限オートマトンによる

モデル化を試みている。

本稿では、この研究での文節モデルの基本的な考え方、この文節モデルの概略、および文節モデルの精密化について述べる。初めに、このモデルを作成するため、対話音声テキストの分析を行なった。この分析結果を基にして、約4000語の語彙を持つ文節モデルを作成した。次に、このモデルと確率的言語モデルとを融合することによって精度向上を試みる。さらに、有限オートマトンで表現しにくい各種の制約について、属性によってそれらの制約を表現する。

2 本研究での文節のモデル化

「文節」は日本語の文を構成する重要な単位だと考えられている^[2]。もともとの文節の定義には、「続けて発声される」といった音声的な特徴も含まれているが、実際の対話音声においては、言い淀みや言い直しなどの現象によって、この原則は必ずしも成り立たない^[3]。また、統語的な面から見ても、自然な発話においては助詞の脱落が頻繁に発生する^[4]。本研究では、文節に対する「続けて発声される」といった音声的な特徴は特に意識せず、主に文節の構造のみをモデル化する。また、助詞の脱落等の現象を考慮し、文節構造自体にはあまり厳しい制約は設けていない。むしろ、比較的緩い文節構造をもとに、確率などを導入することでモデルの精密化を図る。

本研究における文節構造は、寺村^[5]の文法に基いている。これは従来の学校文法に基づく構造とは異なる。例えば従来の「助動詞」「助詞」といった品詞については、それらが持つ機能によって「アスペクト」「ムード」等に再分類されている。また、このような考え方に基いて、動詞・形容詞などの活用表も独自のものを用いている。「言語の構造を記述する」という意味では、どの文法を採用しても原理的に大きな違いはない。しかし、ここでは、文法の記述性の良さなどから、特にこの文法を採用している。

本研究では、この文節構造を、正規文法(有限オートマトン)で表現している。文節構造を文脈自由文法で記述することも良く行われているが^[6]、多くの場合は正規文法と等価な記法で記述される^[7]。実際の認識を考えた場合でも、文脈自由文法と比較して正規文法の方が制御が容易であるため、ここでも正規文法を採用した。しかし、正規文法では表現しにくい言語現象があるため、属性文法の導入を検討した。

3 文節モデルの概略

3.1 対話テキストの分析とモデルの作成

次に、文節モデルの概略について述べる。この文節モデルは、音響学会連続音声データベース^[8]の対話音声書き起こしテキストの分析に基づいている。この中の

CRL0001～CRL0005

ETL0001～ETL0008

KIT0001～KIT0008

NTU1001～NTU1004

OSA0001～OSA0007

TOH0001～TOH0007

TSU0001～TSU0005

の44対話を分析対象とし、その中に含まれるほとんどの表現をカバーするようにモデルを作成した。

分析にあたっては、テキストデータベースから不要語・あいづち・言い淀みを除いた後、目視により文節に区切った。この結果の3633発話、19019文節が実際の分析対象となる。

分析は、形態素解析のプログラムを用いて半自動的に行われた。まず、従来書き言葉に対して用いていた文節モデル^[9]を用いて対象文節を解析し、誤った解析結果が得られた場合に元のモデルを修正し、全文節が解析できるようになるまでこれを繰り返した。最終的には、128文節(全体の0.67%)が解析不能な文節として残った。これらの文節は、発声時の言い間違い、あるいは書き起こしのミスであろうと思われるが、どちらであるかは判然としなかった。

解析できた文節から、実質語3740、機能語978が抽出された。このうち、表記の違い(漢字書きとカナ書き、長音化等)と思われるものを一つにまとめると、実質語3386、機能語615となった。これらのおおまかな内訳は、表1の通りである。

作成された文節モデルの概略を図1に示す。モデル内の品詞は、大きく実質語(名詞、動詞、形容詞、形容動詞、副詞、連体詞)と機能語とに分かれる。

3.2 実質語

名詞は「普通の名詞」「サ変名詞」「副詞的な名詞」に下位分類された。このうち「副詞的な名詞」は、単独で副詞的な働きをする以外に、名詞に後続する助詞などを取って副詞句を構成するもので、「多少」「一番」「今度」「いくら」「少し」などが含まれる。実際には、これをさらに「数量詞的に用いられるもの」「そ

表 1: モデルに含まれる品詞の内訳

品詞		モデル内 の出現数	テキスト 内出現数	表記の重複 を除いた数
名詞	普通の名詞	2378	2320	2249
	サ変名詞	189	186	181
	副詞的名詞	91	88	71
	計	2658	2594	2501
動詞		437	428	370
形容詞		103	99	87
形容 動詞	普通	85	83	80
	名詞的	20	19	18
	計	105	101	98
副詞		207	199	175
連体詞		46	45	33
その他の実質語		184	176	122
計		3740	3642	3386
活用語尾		182	128	105
後置詞		160	118	112
COPULA		11	11	11
助動詞		85	75	60
形式名詞		52	51	33
接尾辞		34	26	21
その他の機能語		454	363	273
計		978	772	615

れ以外」に下位分類している。動詞と活用語尾については、これまで我々が用いてきた活用体系^[9]を採用している。しかし、基本的な活用(カ行・サ行変格活用を含めて)以外に、特殊な活用をする動詞がいくつかあったため、これらの活用を追加した。具体的には、「行く(語幹末子音がkで促音便)」、「ゆく(タ系語尾がない)」、「いらっしゃる(命令形語尾が「い」)」等である。また、関西方言のCopula「や」の活用と、同じく関西方言でタ系語尾しかとらない助動詞(「見してもーた」の「もー」等)の活用を追加した。また、後述する助動詞と活用語尾との融合を扱うため、「基本音便形」という形を設けた。形容動詞は、一般の形容動詞(「きれいだ」等)の他に、名詞的な形容動詞という下位分類を設けた。これは「いろいろ」「ぎりぎり」「苦手」等で、「～な」という普通の形容動詞の語尾以外に、「～の」という名詞に後続する後置詞にも接続する。副詞には、

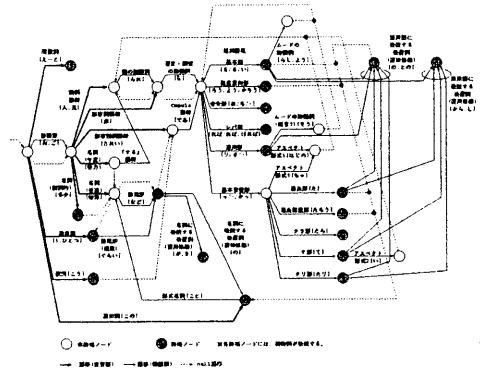


図 1: 文節モデルの概略

「今だに」「このようにして」等の慣用表現も含めた。表1中の「その他の実質語」には、間投詞(「あー」「えーと」等)、挨拶表現(「ありがとうございます」「どうも」等)が含まれる。

これらの実質語とは別に、数量詞を設定した。通常の「数字+単位」だけでなく、「いくつ」「いくら」のようなものも数量詞と同じ扱いにしている。

3.3 機能語

後置詞については、ある一定の規則による接続を許した。この接続表の一部分を表2に示す。この接続表から、後置詞をA～Hの8つのグループに分け、それらのグループ間の接続可能性を実際の規則としている。表中の+は、COPULAを表す。また接続表中の○は「接続が可能」であることを、*は「接続が可能だが、それに相当する表現を別項目として立ててある」ことを、また?は「判断に揺れがある」ことを示している。これらの助動詞とは別に、「ね」「ですね」等の終助詞を登録し、これらの要素は任意の文節の最後に接続するとした。グループ間の接続規則は表3の通りである。

COPULAには、従来の「だ」「です」「である」とその否定形だけでなく、関西方言に見られる「～や」とその否定形「～やない」も登録した。

助動詞は、「～している」「～しておく」「～しはじめる」等のアスペクトを表すもの、「～してもらう」等の授受を表すもの、「～していただく」「～してくださる」等の敬語表現に分けた。助動詞のうち、「～ている」「～てもらう」の形の接続をするものに

表 2: 後置詞の接続表 (部分)

	+	A	B	E	H	C	G	F	D
後置詞	だ い う	も ら か	も り り は も は	で に ど の か で を も	ど の か か	に ま ど か で ま で ら の で と や へ も の か			
先									
H	○	○	○	○	○	○	○	○	○
か	○	○	○	○	○	○	○	○	○
で	○	○	○	○	○	○	○	○	○
から	○	○	○	○	○	○	○	○	○
C	○	○	○	○	○	○	○	○	○
に	○	○	○	○	○	○	○	○	○
まで	○	○	○	○	○	○	○	○	○
E	○	○	○	○	○	○	○	○	○
に	○	○	○	○	○	○	○	○	○
F	○	○	○	○	○	○	○	○	○
へ	○	○	○	○	○	○	○	○	○
も	○	○	○	○	○	○	○	○	○
を	○	○	○	○	○	○	○	○	○
B	○	○	○	○	○	○	○	○	○
が	○	○	○	○	○	○	○	○	○
で	○	○	○	○	○	○	○	○	○
より	○	○	○	○	○	○	○	○	○
より	○	○	○	○	○	○	○	○	○
でも	○	○	○	○	○	○	○	○	○
は	○	○	○	○	○	○	○	○	○
A	○	○	○	○	○	○	○	○	○
しか	○	○	○	○	○	○	○	○	○
なら	○	○	○	○	○	○	○	○	○
も	○	○	○	○	○	○	○	○	○
G	○	○	○	○	○	○	○	○	○
と	○	○	○	○	○	○	○	○	○
の	○	○	○	○	○	○	○	○	○
D	○	○	○	○	○	○	○	○	○
も	○	○	○	○	○	○	○	○	○
が	○	○	○	○	○	○	○	○	○

については、「～てしまう」→「～ちゃう／～じゃう」、「～ておく」→「～どく／～どく」のように、語幹によって形の変わる要素がある。接尾辞については、概数表現(「～ぐらい」、「～ほど」等)と、それ以外(「～など」、「～自身」、「～方面」等)とを分けた。概数表現は、名詞だけでなく、数量詞や動詞にも後続する可能性がある。

表 3: 後置詞グループ間の接続規則

先行グループ	後続グループ
A	+
B	+
C	+ A B E H
D	なし
E	+ A
F	+ A
G	なし
H	+ A B C E F G H

4 確率モデルとの併用

この文節モデルは、分析したテキストに表れる表現のほとんどを網羅するように作成された。そのため、このモデルをそのまま認識システムに用いても、あまり強い制約にならない可能性が高い。しかし、これ以上文法的制約を強くすれば、対話音声のような「崩れた」音声を扱うことができなくなる。そこで、文法的制約は比較的緩いままにして、そこでの制約の弱さを確率的言語モデルによって補うという方法が考えられる。

今回は、有限オートマトンに遷移確率を使うという方法と、bigramと組みあわせる方法の2通りを試してみた。

4.1 遷移確率を用いる方法

有限オートマトンの各状態遷移について、遷移確率を付与することによりモデルの精度を上げようという方法である。この方法は、<カテゴリー-単語>または<単語-カテゴリー>のbigramによる確率モデルに近い(厳密には、それより能力が高い)。

状態遷移確率は、モデルの作成に使ったテキストデータベースを使って推定した。まず形態素解析システムを使って対象文を解析し、各要素の出現頻度を調査した。ただし、形態素解析にあいまいさが残った場合、それを目視で一意化することせず、あいまいさの度合で頻度を重みづけしている。例えば、「しないです」なという表現の場合、

- s [verb] - ina [neg] - i [indicative] - des [copula] - u [indicative] - ne [pp]

- s [verb] - ina [neg] - i [indicative] - desune [pp]

の2通りの分析がなされる。このうち、あいまいさがない“s-ina-i”の部分は1回、あいまいさがある“des-u-ne”と“desune”は0.5回と数えている。このようにして分析した結果から、文節モデルの各アークの遷移確率を求め、perplexityを計算した。

4.2 bigram との併用

有限オートマトンを遷移する候補について、単語bigramで得点付けすることによって候補を絞ることができる。これは、先行する単語によって状態を分割した新たな確率つき有限オートマトンを用いることと等価である(図2)。

ここで用いた単語 bigram は、次のようにして作成

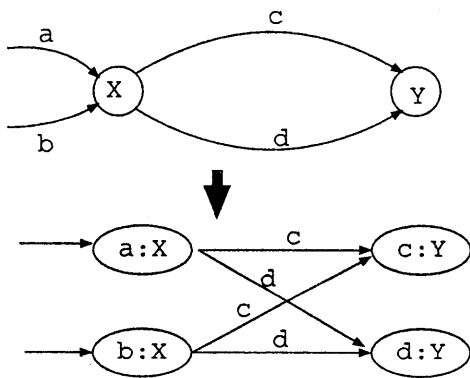


図 2: 有限オートマトンと bigram の併用

表 4: 各言語モデルの perplexity

	perplexity	ngram 総数
(A) 有限オートマトン	403.37	-
(B) 確率オートマトン	217.19	-
(C) auto+bigram	93.71	-
(D) auto+bigram+B/O	117.20	-
(E) bigram	99.365	13167
(F) trigram	72.302	23850
(G) 4-gram	65.698	34626

した。まず、前節の形態素解析結果を使ってテキストを単語に分け、曖昧性のある文は目視によって一意化した。このようにして単語分けされたテキストから単語 bigram を求めた。この bigram と有限オートマトンを組み合わせたときの単語 perplexity を求めた。また、bigram を求めたテキストに出現しない組み合わせが出現することを考慮して、バックオフ平滑化(文献^[10]による方法)を用いた場合についても検討した。

さらに、比較のため、bigram, trigram, 4-gram のみによる perplexity も求めた。

4.3 各手法の perplexity

以上の言語モデルの perplexity は表 4 のようになった。表中の (A) は各遷移に一律な確率を与えた有限オートマトンモデル、(B) は上記の遷移確率を与えた有限オートマトンモデルである。(C) は bigram と併用したもの、(D) はバックオフ平滑化を使った

bigram と併用したものである。(E)～(G) は各確率モデルを単独で使った場合である。

5 属性による制約の表現

5.1 正規文法で表現しにくい制約

次に、この文節モデルへの属性を用いた制約の導入を試みる。今回作成した文節モデルは正規文法(有限オートマトン)であるために、主に隣接した単語間での制約しか表現されていない。しかし、実際の文節には、離れた要素間の相互作用が存在する。例えば、次のような場合に離れた要素間の関係が問題になる。

「有る」という動詞には助動詞「～(て)いる」が接続しないが、本モデルではこの(実際には許容されない)「有っている」という文節を「有/って/い/る」と分解するため、この関係は「有る」の語幹「有」と助動詞語幹「い」との相互作用ということになる。このような離れた要素間の作用をうまく表現する方法の 1 つは、正規文法よりも上のクラスの文法、例えば文脈自由文法を用いてモデルを記述することである。しかし、文脈自由文法を用いたモデルでは、実際の処理が複雑になるので、あまり望ましい方法とは言えない。

そこで、今回は次のような手法でこの制約を表現した。まず、言語的な制約を属性付きの有限オートマトンで記述し^[11]、これを自動的に属性なしの有限オートマトンに変換する。

5.2 属性付き有限オートマトン

本稿で言う「属性付き有限オートマトン」とは、属性文法^[12]の文脈自由文法のかわりに、(与えられた有限オートマトンと等価な)正規文法を用いたものである。属性文法の属性には合成属性と継承属性があるが、ここでは継承属性のみが用いられる。

属性付き有限オートマトンの例を図 3 に示す。この例では、W という属性が用いられている。便宜上、開始状態 (A) においては、すべての属性が 0 であるとする。この例では、「走」という遷移において属性 W に 0 が代入され、「送」という遷移では W に 1 が代入されている。さらに、状態 C から D への遷移「ある」は、属性 W が 1 の場合のみ許容されている。このような属性付けにより、このモデルから「送ってある」という文節は生成されるが、「走ってある」という文節は生成されない。

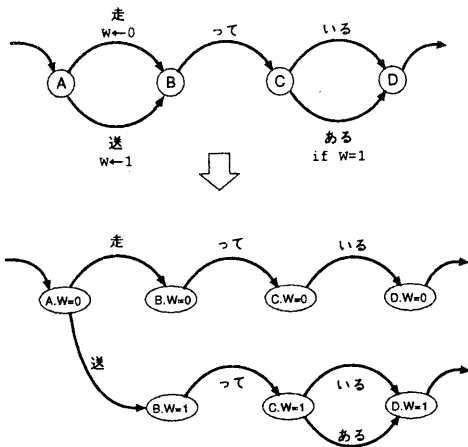


図 3: 属性付きオートマトンの変換

このような属性付きオートマトンをそのまま認識時の処理に用いることもできるが、有限個の属性がそれぞれ有限個の属性値しか取らない場合は、属性付きオートマトンの記述能力は単なる有限オートマトンと等価である。ここでは、取り扱いの容易さを考えて、属性付きの有限オートマトンを属性なしの有限オートマトンに変換し、最終的には属性なしのオートマトンとして扱うことにする(図 3)。

属性付きのオートマトンを属性なしに変換するには、基本的に次のように考えればよい。属性の種類を $A = \{a_1, a_2, \dots, a_n\}$ とし、 a_i の取る属性値を $V_i = \{v_1^{(i)}, v_2^{(i)}, \dots, v_{m_i}^{(i)}\}$ とする。各属性とそれが取りうる値の組の集合 AV を

$$AV = \bigcup_{i=1}^n \bigcup_{j=1}^{m_i} \{(a_i, v_j^{(i)})\}$$

とすると、元の有限オートマトンの状態 $F = \{s_1, s_2, \dots, s_k\}$ に対して、 $F' = F \times AV$ を変換後の有限オートマトンの状態とする。これにより、状態だけで属性とその値を表現することができる。

これを単純に行うと、状態数が(属性の数×属性値の数)倍に増加する。例えば、2 値の属性を 3 つ導入すると、状態数は 6 倍になってしまう。しかし、各状態遷移について「ありえない属性値の組み合わせ」があることを考えると、この方法で生成される状態には非常に無駄が多くなる。そこで本稿では、実際に出現する属性値の組み合わせだけを考慮した変換法を用い

表 5: 補助動詞に関する動詞の属性

属性名	意味	属性値
V	アスペクト的性質	0(瞬間動詞), C(継続動詞), S(状態動詞)
W	自制性	0,1

表 6: 補助動詞の接続条件と属性

補助動詞		V	W
いる	条件	S 以外	任意
	結果	S	0
ある	条件	任意	1
	結果	S	0
みる	条件	任意	1
	結果	不変	不変
はじめる	条件	C	任意
	結果	0	0
かかる	条件	S 以外	任意
	結果	0	0

た。この方法では、最初に各属性に default の属性値“”を与えておき、これ以外の属性の代入があった部分だけ別な状態を派生させるというものである。この手法により、ありえない属性値の組み合わせに対応する状態を生成せずにすむ。

6 文節オートマトンにおける制約

次に、実際に文節モデルに対して導入する制約について述べる。一般的な文節のモデルについての制約はいくつかあるが^[9]、ここでは次のような制約について検討を行った。

1. 補助動詞についての制約:

補助動詞には、「～している」などのアスペクトを表すもの、「～してみる」などの「もくろみ」の助動詞、「～してもらう」などの授受の助動詞が含まれる。ここでは、補助動詞が接続する元の動詞に属性を付与し、その属性値に基づいて接続制限を行う。ここで用いた属性を表 5 に示す。また、これらの属性に対して、いくつかの補助動詞の接続条件と接続後の属性値を表 6 に示す。

表 7: 属性による制約の導入の効果

	制約なし	制約導入後	
		条件付	条件無視
状態数	220	1713	1922
状態遷移数 (null 遷移を含む)	4603	12298	14177
perplexity (一様確率)	403.4	359.3	363.0
perplexity (bigram+B/O)	117.2	113.0	120.6

2. 活用語尾

活用語尾に関しては、推量・非過去形（「食べよう」など）と命令形（「食べろ」など）については $W=1$ の動詞のみという制約を設けた。

3. 接頭辞「お」

動詞の連用形に接頭辞「お」が先行した形式は、「お考え」のように名詞として用いられることがある。これまでのモデルでは、この形式を元の動詞とは別な項目として登録していたが、今回これを属性による表現に変更した。接頭辞「お」を遷移した候補に属性 $O=1$ を与え、 $O=1$ のときのみ連用形を名詞として扱うことにした。

7 制約導入の評価

次に、実際に文節モデルに制約を導入した場合の効果について調べた。まず、制約を導入した文節オートマトンを属性なしの有限オートマトンに変換した際の、状態数と(条件を無視)という項目は、元のモデルとの比較のために、属性付きオートマトンを変換する際に、状態遷移の条件を無視して変換したものである。次に、それぞれのモデルを用いて、状態遷移確率一定とした場合の perplexity を求めた。この結果を表 7 に示す。この結果から、制約の導入により、状態数が約 8.5 倍、状態遷移が約 3 倍になっていることがわかる。単純な展開では状態数・遷移数とも 12 倍になるので、それと比較するとかなり少なく抑えられていると言える。また perplexity は、オリジナルの 403.4 から 363.0 に減少した。さらに、bigram とバックオフ平滑化を併用した場合についても perplexity を求めたところ、オリジナルの 117.2 から 113.0 に減少した。

8 まとめ

対話音声の認識のための文節構造モデルについて述べた。このモデルは音響学会連続音声データベースの書き起こしテキストの分析に基づいて作成されたものであり、有限オートマトンで表現されている。この文節モデルに確率モデルを導入することで、モデルの精密化を図った。遷移確率の導入と bigram モデル併用の 2 つの場合について検討し、bigram との併用によってある程度の効果が得られることを確かめた。また、有限オートマトンでは表現しにくい「離れた単語間の制約」について、属性を使った表現を試みた。さらに、属性付き有限オートマトンを属性のない有限オートマトンに展開して用いる手法を提案した。

今後は、このモデルを用いた認識システムの作成、確率モデルを含む文節間制約の表現、より小さいタスクに対するモデルの適応などの検討を行っていきたい。

参考文献

- [1] 伊藤, 岡本, 牧野: 「対話音声認識のための文節構造モデル」, 電子情報通信学会技術報告 SP94-27 (1994-6)
- [2] 橋本信吉: 「国文法体系論」, 岩波書店 (1959)
- [3] 村上仁一, 嵯峨山茂樹: 「自由発話音声認識における音響的および言語的な問題点の検討」, 信学技報 SP91-100 (1991-12)
- [4] 山本, 小林, 中川: 「音声対話文における助詞落ち・倒置の分析と解析手法」, 情報処 vol.33, No.11, pp.1322-1330 (1992-11)
- [5] 寺村: 「日本語のシンタクスと意味」 vol. II, くろしお出版 (1985)
- [6] 北, 川端, 斎藤: 「HMM 音韻認識と拡張 LR 構文解析法を用いた連続音声認識」, 情報処理学会論文誌, Vol.31, No.3 (1990-3)
- [7] 首藤, 榎原, 吉田: 「日本語の機械処理のための文節構造モデル」, 信学論, Vol. J62-D, No. 12, pp. 872-879 (1979-12)
- [8] 小林, 板橋, 速水, 竹沢: 「日本音響学会研究用連続音声データベース」, 音響学会論文誌 vol.48, No.12 (1992)
- [9] 伊藤, 小川, 岡田, 牧野, 城戸: 「日本語 Dictation System における文節構造モデル」, 信学技報 SP87-104 (1987-12)

- [10] G. Bordel, I. Torres and E. Vidal: *Back-off Smoothing in a Syntactic Approach to Language Modelling*, Proc. ICSLP-94, Vol. II, pp.851-854 (1994)
- [11] 伊藤, 牧野: 「属性文法による文節内文法の制約の表現の検討」, 情処講論 4G-2 (1991-10)
- [12] 佐々政孝: 「プログラミング言語処理系」, 岩波書店 (1989)