

## 自然発話の言語現象と音声認識用日本語文法

竹沢 寿幸      田代 敏久      森元 暉

ATR 音声翻訳通信研究所

自然で自発的な発話のための音声認識用日本語文法について検討を行なった。本稿では、ポーズ情報で区切られた区間を対象とする、部分木に基づく文法を提案する。まず、ATR 音声言語データベース中の、自然で自発的な発話に出現する言語現象の典型例を紹介する。次に、朗読音声および自然発話に挿入されるポーズについて調査する。さらに、部分木を出力する音声認識系の予備実験を行ない、有効性を確認する。また、文法の被覆率とその複雑度の関係についても述べる。最後に、自然発話音声認識のための日本語文法の設計開発方針および現状を述べ、今後の研究の進め方を展望する。

## Linguistic Phenomena of Spontaneous Speech and Japanese Grammar for Speech Recognition

Toshiyuki Takezawa, Toshihisa Tashiro, Tsuyoshi Morimoto

ATR Interpreting Telecommunications Research Laboratories

We have been studying how to design and develop a Japanese grammar for spontaneous speech recognition. In this paper, we propose a grammar based on partial trees separated by pause-like information. First, we introduce an overview of linguistic phenomena in ATR Spoken Language Database. Second, we examine how to insert pauses in read and spontaneous speech. We have conducted preliminary experiments of speech recognition based on partial parse trees. We also discuss the relationship between the coverage of the grammar and its perplexity. Finally, we mention the current status of our Japanese grammar for spontaneous speech recognition and the future direction of our research.

### 1 まえがき

音声翻訳ないし音声対話システムの構築を目指して、自然で自発的な発話を対象とする連続音声認識の研究を進めている。その一環として、音声認識用日本語文法の検討を行なっている。音声認識における言語モデルの1つとしての日本語文法には、次の利点がある。

- 単語や品詞の統計的な接続情報 (典型的にはバイグラムやトライグラム) のみでは、発話としてあり得ないような系列が生成されることがある。統語的な制約を利用すれば、統語的にあり得ない系列を除去することができる。
- 音声翻訳ないし音声対話システムを構成しようとする場合、音声認識系からの出力としては、音素系列よりも単語系列、単語系列よりも部分木系列

が望ましい。音声認識用日本語文法を利用すれば、部分木仮説を生成することができる。

本稿では、ポーズ情報で区切られた区間を対象とする、部分木に基づく文法を提案する。まず、音声認識用日本語文法を設計するために、自然で自発的な発話の調査をする必要がある。その基礎資料として、我々が現在収集を進めている音声言語データベースを利用し、従来の研究では「不適切」とされていた言語現象を分類・整理する。次に、朗読音声および自然発話に挿入されるポーズ位置について調査する。さらに、部分木を出力する音声認識系の予備実験を行ない、有効性を確認する。また、文法の被覆率とその複雑度の関係についても述べる。最後に、自然発話音声認識のための日本語文法の設計開発方針および現状を述べ、今後の研究の進め方を展望する。

表 1: 自然で自発的な発話の言語現象の典型例

言い淀み	[えー] わたくし一人だけです。 近代美術館のすぐ隣りに【あの】なりますので。
気づき	【あっ】 チケットの受け取りの方法はどうなってるんです。 次の列車までは. . . 【あ】 四分しかないですね。
応答	【ええ】 フォードアの小型車だったら何でもいいんですけど。
言い直し	それと、そのホテルの電話番号を教えて(いたた)いただけますか。
言い誤り	はい、お見えになります。
格助詞の欠落	なにかお好みの条件 <が> 有りますか。 お名前とお電話番号のほう <を>、頂けますでしょうか。
助動詞の欠落	年齢は四十歳 <です>。
述語の欠落 1	【あの】 きょうは何時まで開いていますか。 残念ですが、きょうはお休みなんです。 水曜日ですから。 【あ】 水曜日がお休みなんですか。 ええ、あしたは十一時から九時まで <開いています>。
述語の欠落 2	それで、ご予算の方は <いくらですか>。
言い換え	市街地、交通の便利なところがいいですね。
繰り返し	(十時) 朝の十時からですんで一、夜の七時、これで九時間ですか。
連続文	そうですか、鴨川屋ですか。 分かりました、それはすごく良さそうですね。
箇条発話	九月十日、土曜日、二名様、「のぞみ九号」博多行き、東京から京都まで、禁煙席ですね。
必須格の欠落	チケット持ってるんですけども、<便を> 変更したいんですけども。 五月三日は <チケットが> ありますでしょうか。
終助詞の欠落	で、お席の方は、普通席、グリーン席、どちらになさいます <か>。
融合文	八時間コースは今の説明いたしました二つのコースと、 昼食のほうが新都ホテルでご用意いたしております。
語用論レベル	【そうですねー】 新幹線があれだから. . . 夕方六時ごろです。

## 2 自然で自発的な発話の言語現象

### 2.1 ATR 音声言語データベース

調査対象は、ATRで収集を進めている音声言語データベース [1, 2] である。特に音声認識と言語翻訳で共通に利用するための対話を調査対象に選んだ。特徴の1つは、日本語話者と英語話者の、通訳者を介した対話を収集している点にある。通訳の質を高めるために、日英方向と英日方向の2名の通訳者を介している。「近未来の音声翻訳システム」研究の基礎資料を目指しているため、通訳者は1回の発話(発声単位)毎に逐次的に通訳を行なう。通訳者が正確に伝えられるために、1回の発話は10秒以内としている。また、相手の話している間に割り込むことは禁止した。話題は複数存在するが、すべて「旅行に関する対話」である。

### 2.2 言語現象の分類・特徴・頻度

音声翻訳ないし音声対話システムを構築するという立場から言語現象の分類調査を行ない、文献 [3] に結果を報告した。そのうち典型的な例を表1に、頻度の多いものを表2に示す。

表2: 不適格な言語現象と頻度(上位のもの)

265: 言い淀み(文頭)	32: 中止文(接続助詞終止)
131: 気づき(文頭)	27: 提題助詞の欠落
84: 助動詞の欠落	25: 連体助詞の欠落
71: 言い淀み(文中)	23: だ型文
53: 格助詞の欠落(を)	22: つづり
47: 特殊構造(箇条発話)	20: 直示
46: 応答(文頭)	19: サ変名詞述語
37: 述語の欠落	18: 後置詞句の名詞修飾
36: 連続文	18: 融合文

## 3 ポーズの調査

ポーズはもともと息継ぎのために置かれるもので、発話の過程にとってはあくまでも外的なものである。しかし、いったんその過程に入り込むと、統語的・意味的構造の制約を受けないわけにはいかない。

一方、聞き手は、ポーズで区切られたまとまりを手がかりにして発話を理解している。ポーズをすべて取り除くと、聞き手はまったく発話を理解することができなくなることを立証した実験もある [4]。

そこで、ポーズ情報で区切られた区間を部分木で表現するような文法を設計しようと考えた。文法設計のために、文法開発者が構文木の形に注意してポーズ位置の観察を行なった。文法開発者の主観によりポーズの前後に何らかの構文木が構成できると判定できる箇所が存在するポーズを、文法開発者の主観で短ポーズと長ポーズに分けてマークアップした。比較のために、朗読音声と、

自然で自発的な発話音声に対して作業を行なった。音声波形切り出しおよび音素ラベル付与担当者が認定したポーズを物理現象とみなし、その間にどのような関係があるか調べる。

### 3.1 朗読音声の調査

まず、朗読音声に対して、音声波形切り出しおよび音素ラベル付与担当者が認定したポーズの分布を表3に示す。対象は、男性話者1名による「国際会議問い合わせ」に関する261文である。

表3: 朗読音声に対するポーズ長の分布

ポーズ長 [ms]	頻度	[%]
10- 100	44	27.0
100- 200	33	20.2
200- 300	18	11.0
300- 400	15	9.2
400- 500	13	8.0
500- 600	9	5.5
600- 700	14	8.6
700- 800	8	4.9
800- 900	2	1.2
900-1000	2	1.2
1000-1100	1	0.6
1100-1200	1	0.6
1200-1300	2	1.2
1300-1400	1	0.6
合計	163	99.8

同じ朗読音声に対する、文法開発者によるポーズの調査結果を表4に示す。

表4において、文法開発者が短ポーズと認識したにも関わらず、音声波形切り出しおよび音素ラベル付与担当者がポーズと認定しなかった箇所を、次に△マークで示す。ここで▼は長ポーズを示す。

登録用紙は△至急▼送らせていただきます。  
 もう△登録用紙はお持ちでしょうか。  
 それでは△よろしくお願ひします。  
 こちらは△会議事務局です。  
 代理人が参加する場合は、▼あらかじめ△こちらまでお知らせください。  
 それでは早急に△その案内書を送ってください。  
 どちらのホテルが△会議場に近いですか。  
 京都ホテルと△京都プリンスホテルは▼予約できます。

実際に、音声聞いてみると確かに短ポーズがあると判断してもおかしくないように聞こえる。波形を観察してもポーズが存在するように見えるが、パワーがわずかに残っているように見える箇所もある。

表 4: 朗読音声に対する文法開発者によるポーズの調査結果

ポーズ長 [ms]	短ポーズ		長ポーズ		非検出	
	頻度	[%]	頻度	[%]	頻度	[%]
0-	8	4.7				
10- 100	31	18.1	1	0.6	12	7.0
100- 200	19	11.1	13	7.6	1	0.6
200- 300	3	1.8	15	8.8		
300- 400	1	0.6	14	8.2		
400- 500			13	7.6		
500- 600			9	5.3		
600- 700			14	8.2		
700- 800			8	4.7		
800- 900			2	1.2		
900- 1000			2	1.2		
1000- 1100			1	0.6		
1100- 1200			1	0.6		
1200- 1300			2	1.2		
1300- 1400			1	0.6		
小計	62	36.3	96	56.4	13	7.6
	頻度			[%]		
総計	171			100.3		

一方、音声波形切り出しおよび音素ラベル付与担当者がポーズと認定したにも関わらず、文法開発者がマークアップしなかったものが13ある。これらはすべて文節区切りの位置にあったので、意図的なものではない。見落としの理由は主に2つ考えられる。

- 15ms程度の短ポーズが頻度として多い。そのような極めて短いポーズは、主観により「ある」とも「ない」ともいえるものである。つまり、個人差の範囲と考えられるものである。
- 直前の文節区切り位置に長ポーズが存在する場合、直後の文節位置にある短ポーズを見落とす傾向が見られる。

短ポーズと長ポーズには、重なりもあるものの、分布に明らかな差が見られる。

### 3.2 自然で自発的な音声の調査

次に、自然で自発的な発話に対して、音声波形切り出しおよび音素ラベル付与担当者が認定したポーズの分布を表5に示す。対象は、音声言語データベースから選んだ12対話である。

その12対話のうち1対話に対して、文法開発者が短ポーズと長ポーズに分けてマークアップした結果がある<sup>1</sup>。それと音声波形切り出しおよび音素ラベル付与担当者が認定したポーズとの比較結果を表6に示す。

<sup>1</sup>12対話すべてに対してポーズの調査を行なった。短ポーズと長ポーズに分けたものはそのうち1対話のみであった。

表 5: 自然で自発的な発話に対するポーズ長の分布

ポーズ長 [ms]	頻度	[%]
10- 100	63	22.3
100- 200	50	17.7
200- 300	36	12.8
300- 400	35	12.4
400- 500	26	9.2
500- 600	32	11.3
600- 700	16	5.7
700- 800	7	2.5
800- 900	7	2.5
900-1000	1	0.4
1000-1100	4	1.4
1100-1200	1	0.4
1200-1300	1	0.4
1300-1400	2	0.7
1400-1500		
1500-1600		
1600-1700		
1700-1800		
1800-1900	1	0.4
合計	282	100.1

表 6: 自然で自発的な音声に対する文法開発者によるポーズの調査結果

ポーズ長 [ms]	頻度		
	短ポーズ	長ポーズ	非検出
10- 100	5	1	8
100- 200	2	2	4
200- 300	3	4	2
300- 400	2	1	
400- 500	1	1	
500- 600		1	2
600- 700		1	
700- 800			
800- 900		3	
900- 1000			
1000- 1100			
1100- 1200		1	
小計	13	15	16
総計	44		

表6で、500ms～600msに2つマークアップされなかったものがある。いずれも英日逐次通訳者によって発話された中にあり、次に△マークで示す。ここで▼は長ポーズを示す。

現在、▼ニューヨークシティホテル△(570ms)に△(75ms)お泊まりですね。  
電話番号は△(115ms)二零三、▼四四三、▼一七零零△(515ms)で△(15ms)よろしいでしょうか。

これらは、ポーズの前後に適切な構文木が構成できないと判断されたものである。それ以外の非検出データに関しては、朗読音声とほぼ同じ傾向が見られ、同様の解釈が可能である。

ところで、日本語話者の音声と、英日逐次通訳者の音声には、ポーズ位置に関して、有意な差が感じられた。例えば、この例のような発話の仕方は、日本語話者にはほとんど見られず、英日逐次通訳者にしばしば見られた。今回の調査では、日本語話者の音声に重点を置いた。

表3と表5、表4と表6の日本語話者の音声と比べると、似た傾向が見られる。つまり、英日逐次通訳者の発話を除けば、自然で自発的な発話に対して、ポーズを考慮して、部分木を単位とする日本語文法が設計開発できそうな感触が得られた。

## 4 予備実験

### 4.1 部分木を出力する音声認識系の検討

自由で自発的な発話に対して、文全体をカバーする文法を書くことは難しい。一方、音声言語解析(格解析)部への入力として部分木を採用することは魅力的である。そこで、部分木を出力する音声認識系が有効ではないかと考え、朗読音声を対象に、予備実験を行なった。

音声データは、男性話者1名による「国際会議問合わせ」に関する94文である。音素モデルとしては、600状態のHMnet特定話者モデル[5]を使った。実験に使用したマシンはHP9000/755である。実験結果を表7に示す。

SSS-LR[6]が従来手法、SSS-PLRが部分木を出力できるように改良したものである。文法は、両者とも、文献[7]と同じ文単位の文法を採用した。音素タイプライタ[8]で、音素 accuracy の計算式の分母が小さいのは、長母音を考慮したためである。LRパーザを組み合わせた音声認識系からは長母音が出力できるが、音素タイプライタ型からは長母音の出力はできない。

この予備実験によると、部分木を出力することによって、第1位の認識率は若干落ちるものの、第5位までの累積では同程度であることがわかる。音素 accuracy で比較すると、部分木出力のほうがよい。理由は主にビーム探索をしているためである。SSS-LR法では文になり得る仮説が認識過程で伸ばせなくなってしまい、長い文の最後の部分の音素列が適切に作れないことがある。しかし、SSS-PLR法では全体を部分木の集合として取り扱えるので、長い文でも何らかの音素列を最後まで作ることができる。

朗読音声を対象とする予備実験結果は、部分木を出力する音声認識系が有効であることを示唆している。

### 4.2 複雑度(perplexity)の推定

次に、文法的设计にあたって、複雑度(perplexity)の推定に関する予備実験を行なった。自然で自発的な発話に含まれる言語現象の典型例として、言い淀み(間投詞)を取り上げ、既存の「適格な文法」(文献[7]で利用した文節内文法と文節間文法を組み合わせたもの)に言い淀み(間投詞)を組み入れた場合に、どの程度、音素複雑度が変わるか比較した。その結果を表8に示す。言い淀み(間投詞)としては、次の9種類の典型的なものを扱った。

「あの」「あのー」「えと」「えーと」「えー」  
「あ」「じゃ」「うーん」「うーんと」

もし言い淀みがすべての文節区切りに出現できるとすると、複雑度は最大となる。しかし、もし言い淀みの出現を文の先頭のみ許容すると、複雑度の値は小さくなる。この結果は、もし文法を音声認識過程で直接的に利用しようとする、慎重に文法およびメカニズムを設計しなければ、急激に音素複雑度が悪化することを示唆している。

## 5 試作した音声認識用日本語文法の概要

### 5.1 文法の構成と解析例

我々の従来の音声認識用日本語文法は、文節内文法と文節間文法の2段階に分かれていた[9]。部分木を出力する機能を用意することによって、文節という概念が文法体系の中で特別な位置を占めなくなった。そこで、新しい文法は1段階になっている。その結果、音声認識部と言語解析部[10]のインタフェースを取ることが容易になった。

具体的に▼これは、どういう内容の▼ものなんですか。

では、予約の方▼確認させていただきます。

前の文では、「これは」(後置詞句)と「どういう内容の」(連体修飾句)という2つの句が1つのポーズ節(ポーズの間をポーズ節と呼ぶ)に含まれている一方、「どういう内容のもの」(名詞句)という1つの句がポーズによって分割されている。

後の文では、「では」(副詞句)と「予約の方」(名詞句)という2つの句が1つのポーズ節に含まれており、また、「予約の方」に接続すべき格助詞「を」が省略されている。

改良されたメカニズムおよび日本語文法では、これらの文は図1、図2のように解析されることになる。

表 7: 部分木を出力する音声認識系の検討

条件	ビーム幅	認識率		音素認識率				CPU 時間 [s]
		1 位まで	5 位まで	accuracy	ins	del	sub	
SSS-LR	100	54/94 (57%)	62/94 (66%)	1613/2642 (61.1%)	283	209	537	66.1
	250	63/94 (67%)	71/94 (76%)	1766/2642 (66.8%)	241	172	463	119.0
	500	66/94 (70%)	76/94 (81%)	2066/2642 (78.2%)	140	143	293	203.0
SSS-PLR	100	51/94 (54%)	62/94 (66%)	2257/2642 (85.1%)	138	57	200	67.5
	250	59/94 (63%)	71/94 (76%)	2354/2642 (88.8%)	111	32	155	103.4
	500	62/94 (66%)	76/94 (81%)	2405/2642 (90.7%)	97	23	127	175.4
音素タイプライタ	—	10/94 (11%)		1952/2522 (77.5%)	117	99	354	53.3

表 8: 音素複雑度 (perplexity) の比較

文法	テストセット	音素複雑度
適格な文法	文節区切りの文	2.49
適格な文法	文	3.23
言い淀み (間投詞) つき文法	文節の位置に言い淀み出現を許容しない文	3.39
言い淀み (間投詞) つき文法	文節の位置に言い淀み出現を許容する文	4.06

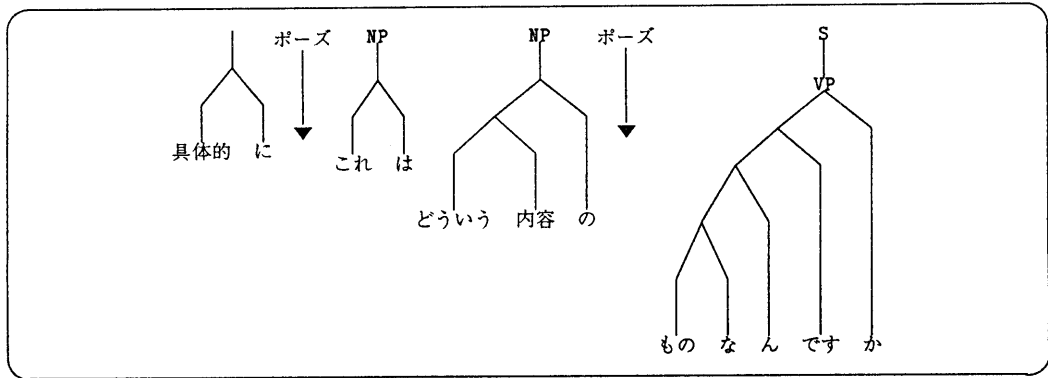


図 1: 部分木による表現例 1

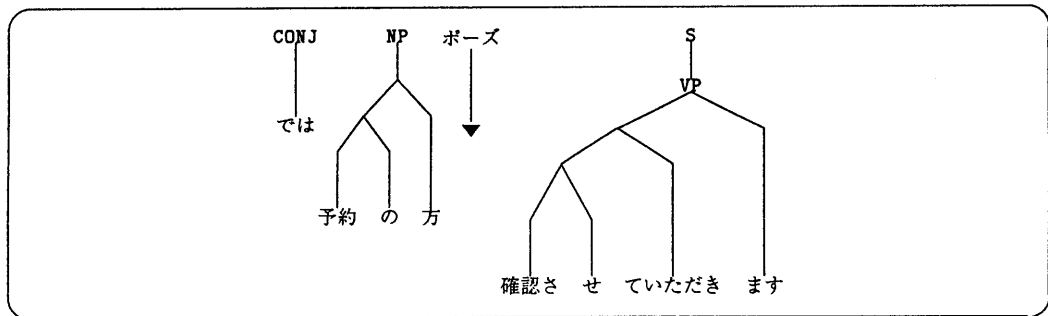


図 2: 部分木による表現例 2

## 5.2 不適格な言語現象の処理方針

自然で自発的な発話に含まれる不適格な現象には、文法で扱えるものと扱えないものがある。例えば、表1の言い直しと言い誤りを文法で扱うのは困難なので、文法では扱わない。

極端な言い直しならば、発話者が気付いて再入力してくれるかもしれない。言い直しの前後にポーズが検出できれば、部分木出力で対処できるかもしれない。

「はい、お使いになります。」という言い誤りが、文法を使って「はい、お使いになれます。」と認識されてもかまわない。音声認識としては誤りであっても、あとの言語解析部にとってはうまく解釈可能となるだろう。

言い直しと言い誤り以外の表1の現象は基本的に文法で扱える。出現頻度の多いものから優先的に対処する。構文規則に取り込む場合は、ポーズ節内に関しては、音声認識に対する制約(特にトップダウン予測による接続制約)が働くことが期待できる方向への記述を行なう。ポーズ節の間にわたる関係については、無理な規則化は避ける。

## 5.3 不適格な言語現象の処理事例

### 5.3.1 言い淀み

言い淀んだ時に発せられる「あー」とか「えー」とかいう発声は、語彙項目として登録するにはなじまないものである。将来は、音素モデルを分けることも考えられる。言い淀みをどこにでも現れうるように記述すると、表8に示すように音声認識へ悪影響を与える。言語現象の調査結果(表2あるいは[3])によると、大半は文頭に出現し、文中のものもポーズの直後に出現することが多いため、文頭(ポーズの直後を含む)に現れるもののみを規則で扱うことにした。

例文1 [えー] そちらの口座の方に請求させていただきます。

例文2 [あー] アールシーエービルディングの六十五階ですか。

### 5.3.2 呼びかけ

「旅行に関する会話」では、ホテルの担当者が顧客に直接呼びかける次のような表現がしばしば出現する。

例文3 鈴木様、お待ちしております。

例文4 お客様、お名前をお願いいたします。

この種の表現は、文頭、文中、文末のどこにも出現しうるが、言い淀みの対処と同様に考え、文頭(ポーズの直後を含む)に出現するものだけを認めた。

### 5.3.3 助詞の省略

日本語では、名詞に後続する助詞が省略されることがある。

例文5 ルームサービスのメニュー以外に何か食べ物くを>作っていただくこと<が>できますか。

例文6 こちら<は>エアバシフィックニューヨーク支店でございます。

これをそのまま構文規則に反映させ、助詞を伴わないすべての名詞句が述語に係ることができるというような規則を与えると、音声認識の際に曖昧性が非常に増える<sup>2</sup>。そこで、助詞が省略された名詞句は部分木として出力し、述語との係り受け関係は、音声言語解析(格解析)部に委ねることにした。

ただし、例文6のように、話者自身を指す「こちら」や「わたくし」だけは、特別なカテゴリを与えて、述語に直接係るようにした<sup>3</sup>。

### 5.3.4 述語の省略

動詞や助動詞が省略されていて、述語に相当する表現が欠けていたり不完全であったりする文である。

例文7 田中健史様。

例文7は、助動詞「ですね」が省略されている。単独の名詞句として不完全なまま出力される。

### 5.3.5 特殊構文(箇条発話)

ホテルや列車の予約の際に、確認のために予約項目を読み上げるものである。

例文8 鈴木和子様、八月の十日から十二日まで、シングルルームシャワー付き二泊ですね。

例文9 はい、ワシントンディエシーへ、メトロライナー十二時二十分発、ユニオン駅でよろしいですね。

これらの発話は、外部の状況がそのまま言語に反映されているといったもので、あらかじめどのような順序で出現するかを予測するのは難しい。しかも、統語的なカテゴリのみで規則化することは困難である。さらに、ポーズで区切って発話される傾向がある。そこで、無理な規則化を避け、部分木として出力する。

<sup>2</sup>規則化する手間をかけても、音声認識に有利に働かそうにないなら、手間を減らそうと判断した。この判断は直感的なものである。部分木で対処するよりも、規則化するほうが、単語複雑度(perplexity)で有利ならば、規則化すべきなので、今後調査する予定である。

<sup>3</sup>従来の文法で既にそうだったので、体系を踏襲した。

### 5.3.6 融合文

2つの文が1つの文に無理に押し込められたといった感じのする文である。

例文 10 連絡先は、今、ウエストサイドホテルの三〇三号室にいるんですよ。

この文は、「連絡先はウエストサイドホテルの三〇三号室です」という文と「わたしは、今ウエストサイドホテルの三〇三号室にいるんですよ」という文が融合したものである。ただ、このような解釈は、意味と文脈を考慮してはじめて可能なので、統語的な情報だけでは不可能である。したがって、この文は、「連絡先は」が「いる」に係るような文として解析される。

## 6 議論および今後の展望

### 6.1 $N$ -gram モデルと構文規則ベース型

$N$ -gram モデルは統計的な言語モデルの典型である。コーパスさえあれば、自動学習 (automatic learning) が可能である。しかし、訓練 (training) のために大量のテキストが必要である。新聞記事のような書き言葉ではなく、話し言葉を対象とすると、大量に書き起こしテキストを集めるには莫大な手間およびコストがかかる。しかも、領域 (ドメイン) や課題 (タスク) に依存してしまう。

一方、構文規則ベース型における構文規則は、領域や課題にあまり強く依存せず開発可能と期待できる。したがって、語彙項目 (lexical entry) さえ変更すれば、複数の課題/領域にポータブル可能である。ヒューリスティックあるいは統計に基づくスコアを付与することは容易である。

したがって、音声翻訳ないし音声対話システムのフロントエンドとしては、構文規則に基づき部分木系列をスコア付きの仮説として出力する、音声パーザ (speech parser) が有用であろう。

ただし、語彙サイズが大規模となったたり、発話様式が激しく変化する状況に対応しようとする、音声認識過程で直接的に構文規則を言語モデルとして利用する手法が最適かどうかはわからない。探索手法と言語モデルの組み合わせ方に関しては、対象とする応用分野 (語彙サイズや発話様式を規定する条件) に応じて、さまざまな手法を精力的に研究する必要がある。

### 6.2 今後の展望

まず、設計および試作の済んだ音声認識用日本語文法を用いて、連続音声認識の評価実験を進める。1万語程度までの大語彙を意識した検討を行ないたい。音声認識用日本語文法のみにごծならず、確率・統計・クラスタに基づく言語モデルとの組合せも試みる。一方で、音声認識過程での探索手法についても精力的に研究を進め

る必要がある。さらに、音声対話システムを考慮した音声認識手法の研究を行なう予定である。

## 7 むすび

ATR 音声言語データベースに含まれる自然で自発的な発話の言語現象を概観し、ポーズの調査や部分木を出力する音声認識系の検討、文法の複雑度の評価を交えながら、試作した音声認識用日本語文法の概要を報告した。今後は、設計および試作の済んだ音声認識用日本語文法を用いた連続音声認識の評価実験および効率的な探索手法に関する研究、さらに音声対話システムを考慮した音声認識手法の研究を進めていく。

## 謝辞

本研究を進めるにあたり、適切な助言をいただいた ATR 音声翻訳通信研究所第四研究室 石崎雅人 主任研究員に感謝します。さらに、調査や実験を進めるうえで支援をいただいた衛藤純司氏、林輝昭氏に感謝いたします。

## 参考文献

- [1] Morimoto, T., Uratani, N., Takezawa, T., Furuse, O., Sobashima, Y., Iida, H., Nakamura, A., Sagisaka, Y., Higuchi, N., Yamazaki, Y.: "A Speech and Language Database for Speech Translation Research", *Proceedings of 1994 International Conference on Spoken Language Processing (ICSLP '94)*, pp. 1791-1794 (1994-09).
- [2] 竹澤寿幸, 古瀬誠, 中村篤: "音声言語データベース — 話し言葉を収集し、音声的・音語的特徴を探る —", *ATR ジャーナル*, No. 17, pp. 4-5 (1994 秋).
- [3] 竹澤寿幸, 田代敏久, 森元暹: "音声言語データベースを用いた自然発話の言語現象の調査", *人工知能学会 言語・音声理解と対話処理研究会 (第 10 回)*, SIG-SLUD-9403-3, pp. 13-20 (1995-02).
- [4] 杉藤美代子: "談話におけるポーズとイントネーション", *日本語と日本語教育 第 2 巻 — 日本語の音声・音韻*, 明治書院 (1993).
- [5] 廣見淳一, 嵯峨山茂樹: "逐次状態分割法による隠れマルコフ網の自動生成", *信学論*, Vol. J76-D-II, No. 10, pp. 2155-2164 (1993-10).
- [6] 永井明人, 廣見淳一, 嵯峨山茂樹: "逐次状態分割法 (SSS) と音素コンテキスト依存 LR パーザを統合した SSS-LR 連続音声認識システム", *信学技報*, SP92-33 (1992-06).
- [7] Takezawa, T., Morimoto, T.: "An Efficient Predictive LR Parser Using Pause Information for Continuously Spoken Sentence Recognition", *Proceedings of 1994 International Conference on Spoken Language Processing (ICSLP '94)*, pp. 1-4 (1994-09).
- [8] 大脇浩, Singer, H., 廣見淳一, 樽松明: "音素配列構造の制約を用いた音素タイプライタ", *信学技報*, SP93-113 (1993-12).
- [9] 保坂順子, 竹澤寿幸, 江原暉輝: "対話データベースを利用した音声認識のための構文規則", *情報処理学会 自然言語処理研究会*, 83-13 (1991).
- [10] 田代敏久, 森元暹: "音声言語処理のための構文解析ツールキット", *情報処理学会 自然言語処理研究会*, 106-12, pp. 67-72 (1995-03).