

話しことばにおけるポーズ節の考察

2Q-6

保坂順子

衛藤純司

ATR 音声翻訳通信研究所

日本アイアール株式会社

1 はじめに

自然な発話を音声認識、言語解析し、その意味を理解するための構文規則に関する研究を進めている。我々の音声翻訳システムでは、文節ごとに区切った発話を処理対象としてきた。そのため、音声認識部では、文節に基づく言語情報を用いてきた。一方、言語解析部では、音声認識された文を入力とし、文を単位とした言語情報を用いてきた。しかし、自然な対話では文節ごとに区切った発話をせず、意味的なまとまりをポーズで示しながら話を進めていると思われる。そこで、我々は、より自然な発話の処理を目指して、ポーズ節を構文規則の基本単位とすることを考える。本稿では、まず、我々が音声認識と言語解析で使っている構造の違いを示す。次に、自由発話の音声データを基にポーズの位置、頻度などを調べる。さらに、この結果を基に、ポーズ節に基づいた構文規則を構築し、音声認識に応用する。

2 音声認識と言語解析の構文構造

我々の音声翻訳システムでは[1]、文節ごとの切れ目にポーズを入れて文を発話し、それを音声認識の入力として扱っていた。そのため、音声認識部では、文節に基づく構文情報を制約として使っていた[2]。一方、音声認識に続く言語解析部では、音響的に、また構文的に尤もらしいと認識された文を文字列として入力し、文単位の処理を行っていた。図1に示すように、音声認識では、文節の中を最初にまとめあげ、その後それぞれの文節の関係を考慮しながら文を組み上げていく。言語解析では、中心となる述語から出発し、その補語との関係を見ながら

文の解析を進めていく：

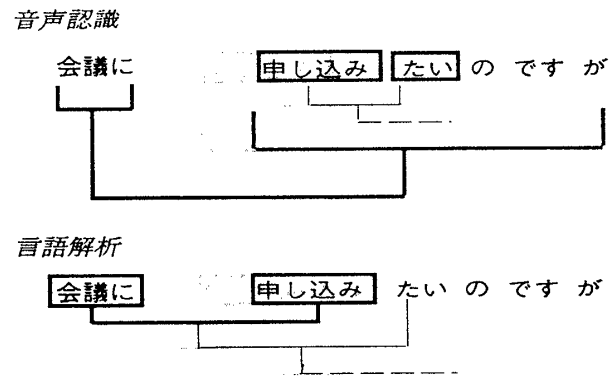


図1: 構造の違い

図1 から明らかなように音声認識と言語解析では、処理単位と構造が異なるため、これまで、同一の構文規則を適用することは不可能だった。

自由発話を同一の構文規則で扱うには、ポーズ節を基本単位とするのが最適であろう。自然な対話では、意味の理解が重要であるため、意味を表すのに適したところにポーズをおくからである。また、ポーズは文法的な区切りと一致することも分かっている[3]。

3 自由発話のポーズの位置と頻度

ポーズと構文の関係に関する研究は、音声合成の立場からも行なわれている[4]。しかし、対象データが読み上げたものであるなど、自然さに欠ける。

本稿では、電話による、会議の申し込み、講演時間の変更などに関する自然な対話を基に、ポーズが挿入される位置、頻度などを調べる。この自由発話データは、40 対話からなっている。対象としたのは、3 人の問い合わせ者の発話で、489 ターン 544 文である。ポーズの検出は人手で行なった。

表1に、どのようなものの後にポーズが出現しているのかを、出現していない頻度と比較して示す：

表1: ポーズの位置と頻度

ポーズの前	ポーズあり	ポーズなし
感動詞	253	35
接続助詞	122	22
接続詞	57	49
後置詞句	49	9
名詞 / 代名詞	34	21
ですね	33	2
副詞	17	15

表1 から、「はい」「ええ」などの感動詞のあと、及び、「けれども」「ので」「が」のような接続助詞のあとではポーズがおかれる場合の方が圧倒的に多いことが分かる。後置詞句や「実はですね」「それではですね」のように「実は」「それでは」などを強めるために使う「ですね」の後でポーズがおかれる頻度はそれほど高くないが、ポーズがおかれない場合に対する割合は高い。

4 ポーズ節を使った音声認識

3 節の調査に基づき、ポーズ節を基本単位とする構文規則セットを 3 種類、試作した: 文法 1: 言語解析のための文法を変更したもの [1]、文法 2: 文法 1 の中から実際に使われる規則、及び単語だけを取り出したもの、文法 3: 音声認識のための文法を変更したもの [2]。表 2 にそれぞれの構文規則セットの大きさとパーブレキシティを示す:

表2: 構文規則の規模と複雑度

	文法 1	文法 2	文法 3
規則数	2655	841	2326
単語数	1466	383	751
音韻パーブレキシティ	5.99	3.23	3.96

表 2 に示した構文規則を音声認識に適用した。実験の対象としたのは、国際会議に関する問い合わせ 137 文である。このテストセットは、文単位に発話されており、ポーズが自然に入っている。認識には HMM-LR パーザを使い [5]、ビーム幅 100 で実験

した。表 3 に、認識実験結果を示す¹。

表3: 音声認識率

ランク	累積認識率 (%)		
	文法 1	文法 2	文法 3
1	25.0	47.4	50.0
2	29.5	48.9	54.2
3	31.1	49.6	55.1
4	31.1	51.1	55.9
5	31.1	51.1	55.9

表 3 から分かるように、認識率はそれほど高くない。最小の文法 2 でも、1 位で 47% の認識率にとどまっている。文法 3 は、文節単位の発話では 70% の文認識率を得ている。しかし、自由発話を入力とした場合、認識率が 20% 低下している。

5 おわりに

本稿では、自由発話の音声認識と言語解析のために、ポーズ節を基本単位とすることを提案した。そのため、544 文からなる自由発話データを使い、ポーズの挿入傾向を調べ、その結果を構文規則にとり込んだ。さらに、従来の音声認識にこの構文規則を適用し、自然な発話の認識実験を行なった。

自動検出では動詞句の中にもポーズが検出されることがある。今後、さらにポーズの使用傾向を調査し、文法的な区切りと合わないポーズの扱いを考えるとともに、構文規則を改良していく予定である。

参考文献

- [1] 竹沢, 森元, 谷戸, 鈴木, 嵯峨山, 樽松 (1993): "ATR 音声言語翻訳実験システム ASURA", 第 46 回情報処理学会全国大会
- [2] 保坂, 竹沢, 江原 (1991): "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
- [3] 杉藤 (1988): "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 2, 明治書院
- [4] 海木, 匂坂 (1991): "局所的句構造に基づくポーズ長の分析", 信学技法 SP91-130
- [5] 北 (1992): A Study on Language Modeling for Speech Recognition, ATR テクニカルレポート TR-L-0273

¹文法 1 と文法 3 を適用した場合、最後まで処理できたものはそれぞれ 132 文、118 文だった。