

音声認識のための連鎖制約としての文法

4 E - 7

保坂 順子 小暮 潔 小倉 健太郎

ATR自動翻訳電話研究所

1. はじめに

音声認識には、単語連鎖の統計的性質を利用したもの、文法を利用したもの、文節間の係受けを利用したもの、意味情報を利用したものなど、様々な角度から言語情報を取り入れる試みがなされている^[1]。本稿では、そのうち、音声認識のための文法構築について考える。

2. 連鎖制約としての文法の構築

ここでは、文法を単語間の連鎖制約として捉え、必要に応じ制約を局所的に強める、という方法をとる。

従来、文法構築は、構造を明らかにする、という主旨で、文は、名詞句と動詞句からなるという規則(S → NP VP)が示すように、トップダウン式に行われてきた。また、近年、この言語学の研究成果を自然言語処理に取り入れようとする動きもある^[2]。HPSG/JPSGを機械翻訳に積極的に取り入れているのも、その一例である^[3]。しかし、音声認識に同量の情報を利用することは、処理の爆発につながる。

従って、ボトムアップ式に、比較的緩い制約から出発して問題となる部分の制約を局所的に強めていく方が有効だと思われる。以下、単語間の連鎖制約をなくしたときの音声認識結果を考察し、制約を強める箇所を明示し、そこに強い制約を加え、音声認識を試みる。

音声認識実験には、CFGの枠組みで定義した規則により次の音素を予測する、という手法を実現したHMM-LR^[4]を利用する^{注1}。この方式では、現在、文節単位の高度な認識が確認されている。

2.1. 連鎖制約無しの音声認識結果の分析

(1)、(2)は、連鎖制約をなくした時の音声認識実験結果である。

```
(1) 229 frames [sochirawa]
*****
Parsing time: CPU-time = 190875 msec,
Elapsed-time = 196 sec.
> 1: sochira-wa
   2: sochira-wa-hu
   3: sochira-hu-wa
   4: sochira-hu-wa-hu
   5: sochira-wa-hu-hu
*****
```

(>)は正しく音声認識されたものを示す。)

```
(2) 376 frames [arigatougozaimasu]
*****
Parsing time: CPU-time = 363752 msec,
Elapsed-time = 679 sec.
   1: ari-nga-to-wa-eN-hu-su-su-su
   2: ari-nga-to-wa-eN-hu-su-su
   3: ari-nga-to-wa-eN-hu-su-su-u
   4: ari-nga-to-wa-eN-su-su
   5: ari-nga-to-wa-eN-hu-su-su-su-u
*****
```

(ng は、鼻音化したガ行音、N は撥音を表す)

HMM-LRでは、左から右への探索を行っているので、文節のはじめが誤って認識されると、文節の正しい認識は望めない。しかし、(1)からわかるように、文節のはじめが正しく認識されても、文節末の認識は、不安定である。

(2)は、'tou'が、'to'と誤って認識され、それ以降正しい音素列が見つけられなかった例である。

(1)(2)共に、認識を誤った時点から、音素数の少ない単語^{注2}を認識候補に選択している。これを回避するためには、音素数の少ない単語の連鎖制約をより厳密にすることが考えられる。音素数の少ない単語としては、助詞、代名詞、数詞、副詞の一部などがあるが、ここでは、特に助詞を中心に話を進める。

2.2. 助詞の連鎖制約

助詞は、音素数の少ない単語であり、また、文節末にくるので、HMM-LRでは特に注意を要するところである。助詞に関しては、必須格としての、「が、を、に」及び、主題、対照の「は」^[5]などの研究、または、個々の助詞に関する用法の考察^[6]は行われているが、その連鎖に言及しているものは「ダケ、ノミ、バカリ、クライ」と格助詞の語順^[7]を扱ったものなどに留まり、総体的に扱ったものは、少ない。通則とされている助詞相互の結合規則^[8]では、例えば、副助詞+格助詞、及び、副助詞+係助詞により、
*食べることながら頭にある
*食べることしかは頭にある
が、許されてしまう。

そこで、データベース検索支援システム^[9]を利用し助詞の連鎖の詳細化をはかった。ここで

は、特に問題となる、音素数の少ない疑問代名詞とそれに後続する助詞との連鎖を調べる。

ATRでデータベース化が進んでいる、電話会話、キーボード会話(約8万語)で実際に使用されている疑問代名詞は、いくら、どこ、どちら、どなた、いずれ、いつ、何、どれ、誰、であり、最高2つの助詞が後続している。このデータベース検索に基づき、

全ての疑問代名詞に、

が、を、に、と、も、でも、の
か+{が、を、に、と、も、は、で、の}
の後続を許し、一部のものには、
まで・から+{に、で}
くらい・ばかり・だけ+
{なんて、なら、こそ、しか、でも}
の後続も許す。また、助詞の後続可能性を図1に示す。

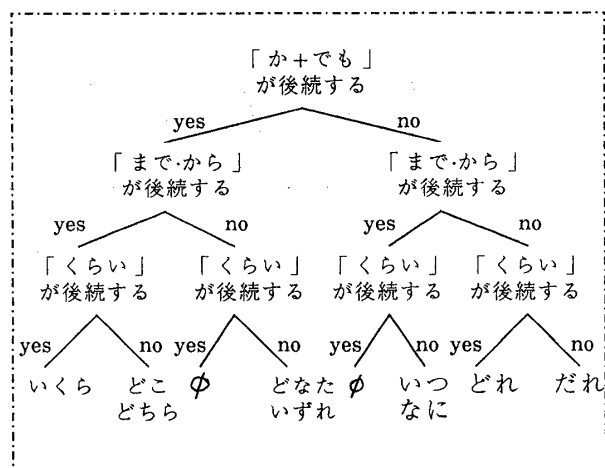


図1 助詞の後続可能性

2.3 連鎖制約の音声認識への応用

約1400の規則で、緩い制約を加えたものと、2.2.で述べた疑問代名詞と助詞との連鎖制約を加え、約1500の規則にしたものを使い音声認識した結果を以下に示す。緩い制約を加えた結果の(3)の2、3、4が、強い制約を加えた結果の(3')では消え、3位に正解が入っている。

国際会議に関する電話での問い合わせ対話、37文、83文節について、実験の結果、第3位までの認識率が、79.5%から85.5%に上昇した。

(3)229 frames |soudesu|

Parsing time: CPU-time = 34815 msec,

Elapsed-time = 35 sec.

- 1: sou-desu-shi
- 2: itsu-o-desu-shi
- 3: itsu-to-desu-shi
- 4: itsu-to-o-desu-shi
- 5: i-soudesu-shi

(3') 229 frames |soudesu|

Parsing time: CPU-time = 32232 msec,

Elapsed-time = 77 sec.

- 1: sou-desu-shi
- 2: i-soudesu-shi
- > 3: sou-desu
- 4: i-soudesu
- 5: sou-desu-ka

3. おわりに

音声認識に特に有効に働くと思われる文法の構築に、局所的に連鎖制約を強めることを提案した。ここでは、音声認識において問題となる音素数の少ない単語の中でも、特に、疑問代名詞と助詞の連鎖制約を扱った。

日常会話では、言い間違い、言いよどみ、などが頻繁に表れる。疑問代名詞と助詞との連鎖制約を強めることにより生成されなくなった「いつをですし」、「いつとですし」、「いつとをですし」なども、実際には発声される可能性がある。しかし、音声認識結果を、文法的に正しい文だけを受け付ける言語処理システム(例えば、機械翻訳システム)への入力とする場合には、これらを認識できるようにすることよりも、文法的に正しい文に対する認識率を向上させることの方が重要である。このように、連鎖制約を強めていくことは、音声言語の統合的処理の一部としての音声認識システムとしては、適切な方法である。

今後は、この文法構築アプローチにそって、文法を音声認識に、より適合させる予定である。

謝辞 本研究の機会を与えて下さると共に適切な助言を頂いたATR自動翻訳電話研究所 樽松明社長、同データ処理研究室 森元暹室長に感謝します。

注釈

- (注1) ビーム幅16、1本の枝からの枝分かれ10の設定で実験した。
- (注2) 動詞、形容詞など、活用のあるものは、活用したものを1単語としている

参考文献

- [1] 好田正紀 1988:「音声認識における言語処理」,人工知能学会誌 vol.3 No.4
- [2] 郡司隆男 1987:『自然言語の文法理論』,産業図書
- [3] 小暮・堂坂・加藤 1989:「SL-TRANSにおける言語解析」第39回情報処理学会全国大会
- [4] 北・坂野・保坂・川端 1989:「SL-TRANSにおける文節音声認識」,第39回情報処理学会全国大会
- [5] 久野晴 1973:『日本文法研究』,大修館
- [6] 国立国語研究所 1951:『現代語の助詞・助動詞』,秀英出版
- [7] 久野晴 1983:『新日本文法研究』,大修館
- [8] 水谷・石綿・荻野・賀来・草薙 1983:『文法と意味I』,朝倉日本語新講座 3,朝倉書店
- [9] 橋本・小倉・江原・森元 1990:「対話データベースを用いた各種言語現象の検索」,本予稿集