

文節間の連鎖制約を取り入れたHMM-LRの音声認識実験

7E-4

保坂 順子 森元 逞
ATR自動翻訳電話研究所

1. はじめに

音声認識の手法として、HMM(Hidden Markov Model)と拡張LRパーザとを組み合わせたHMM-LR方式の研究を進めている[1]。この方式では、CFGで定義された文法から次に発声される音韻を予測し、その音韻をHMMを用いて確認する手法を実現している。現在は文節を単位として音声認識を行っており、高い認識率が得られることが確認されている。しかし、出力された文節候補を組合せ、文をつくった場合、文節間情報が欠けているため、文として成り立たないものが多数出力される。HMM-LRが組み込まれている音声言語翻訳システムSL-TRANS[2]では、多数の文節候補からもっともらしい文候補を取り出すために、文節間の係り受け情報[3]を利用している。

一方この問題のもうひとつの解決方法として、文の文法を利用することが考えられる。ここでは、第1段階として簡単な文の文法をHMM-LRに組み込んで行った認識実験について報告する。

2. 実験

2.1 実験の概要

文法は、従来の文節内文法(G1)に、文節規則の連鎖可能情報のみを与えることにより文の文法(G2)に拡大した。入力文は文節単位に発声された音声で、文に組み立てたものを使った。文節の繋ぎめには、無音区間が入っており、パーザには、その処理が組み込まれている。また、文法にこの処理を起動させる記号を導入した。ここでは、HMM-LRへのG1の適用結果とG2の適用結果との一例を比較することにより、G2の有効性/問題点などについて考察する。

2.2 実験環境

HMM-LRでは、ビーム探索をするが、その際のパラメータ、ビーム幅、1ノードから出る枝数、域値はほぼ同値に設定した。ここでは、音声

認識結果第4位までを扱う。G1、G2共、対象は国際会議に関する問い合わせの会話(37文、83文節)である。G1には、約10種類の文節、計700程度の規則が用意されている。G2には、以下のような連鎖規則が約15種類加えられている。

- (1) 既に、至急、まず、などの副詞、それでは、また、などの接続詞は、出現位置が規定されている。
- (2) ありがとう、などの間投詞は、単独でのみ出現を許す。
- (3) G1では数詞も単独で文節を構成できていたが、G2では特定の環境(料金の記述、地名における番地の記述など)でのみ現れるという制約を付加している。

2.3 実験結果

(実験1)

19: 入力文「どうも失礼いたします」

G1: (39) 159 frames |doumo|

1: go-no(1.880385)

2: go-mo(1.940873)

3: doumo(1.951184)

4: go-o(2.238138)

(40) 386 frames |shitsureiitashimasu|

1: shitsurei-ita-shi-masu(1.877279)

2: shitsurei-shi-masu(2.149714)

3: shitsurei-shi-te-i-masu(2.160686)

4: shitsurei-deki-masu(2.238941)

G1を適用した音声認識では、文節39のように文節候補として単独の数詞がかなり多くあらわれる。また、この結果の尤度のみを考慮して文節を組み合わせると、以下のように、16個の文が候補となる:

- 1: go-no shitsurei-ita-shi-masu
- 2: go-mo shitsurei-ita-shi-masu
- 3: doumo shitsurei-ita-shi-masu
- 4: go-o shitsurei-ita-shi-masu
- 5: go-no shitsurei-shi-masu
- 6: go-no shitsurei-shi-te-i-masu
- 7: go-no shitsurei-deki-masu
- :

これに対し、G2を適用した場合、特に数詞の規制が働き、文候補は2つに減少する。文節40に出力されている動詞句で、ここでは出力されていないものがあるのは、無音処理のためである。

```
G2: VQ-file -> /ifg1/BunParser/Bun/bun-19
*****
1:Q1doumo-Q1shitsurei-ita-shi-masu-Q2-
  (1.873052)
2: Q1doumo-Q1shitsurei-shi-te-i-masu-Q2-
  (1.977608)
*****
```

(注)Q1、Q2は無音区間をしめす。

(実験2)

1: 入力文「もしもし」

```
G1: (1) 205 frames |moshimoshi|
*****
1: moshimoshi(1.848592)
2: ni-zyuu-shi(2.243298)
3: ni-zyuu-shichi(2.300748)
4: ni-zyuu-ichi(2.386748)

G2: VQ-file -> /ifg1/BunParser/Bun/bun-1
*****
1: Q1moshimoshi-Q2- (1.998655)
2: Q1mochii-mashi-ta-Q2- (2.612681)
3: Q1mochii-nai-Q2- (2.624753)
4: Q1ma-chi-mashi-ta-Q2- (2.690399)
```

G1でもG2でも正解が第1位になっている。しかし、G1の場合、一文節からなる文のため、係り受けフィルターの対象外となり、上記の例のように、単独の数詞なども文候補として残される。G2では、構文的には正しい文のみが候補として出力されている。ただし、G2では、文候補2、3、4に見られるように、さらに認識率を高めるには、文間の制約などを利用することが必要である。

(実験3)

17: 入力文「ありがとうございます」

```
G2: VQ-file -> /ifg1/BunParser/Bun/bun-17
*****
1: Q1kata-to-Q1sa-re-masu-Q2- (2.117882)
2: Q1aringatou-gozaime-su-Q2- (2.123770)
3: Q1kaingi-to-Q1sa-re-masu-Q2- (2.129113)
4: Q1imi-to-Q1sa-re-masu-Q2- (2.171893)
```

文節接続規則を導入することにより、候補にあがる文節の種類を制限できる。一方、ある文節数からなる文に対し、それ以外の文節数の文も候補として出力され、文の候補数が必ずしも減少しないことがある。上記の例では、文候補1、3、4で、一文節の入力に対して、二文節の規則が使われている。

また、この例では、まだ十分な文節間の制約が定義されていないため、構文的に正しくない文候補も出力されている。

これらの問題を解決するには、構文情報を制約として強化するとともに、等価な語彙の共起関係や意味素性による選択制約、さらに、前述の文間の制約などを導入することが必要であろう。

以上、実験3にみられるように、今後の改良が必要ではあるが、文レベルの文法の導入が文節レベルの文法に比べて有効であることが、実際のデータにより確認された。

3. まとめ

音声認識への文の文法の適用の是非を検討した。文節内文法に最も単純な連鎖情報を付加することにより、文レベルの文法に拡張し、数詞の出現位置に関する制限を加え、その有効性を実証した。

謝辞

本研究の機会を与えて下さるとともに適切な助言を述べられた樽松明社長に感謝する。また御討論いただいた自動翻訳電話研究所の諸氏に感謝する。

参考文献

- [1] 北・坂野・保坂・川端、SL-TRANSにおける文節音声認識、本予稿集、1989
- [2] 森元・小倉・相沢・樽松、音声言語日英翻訳実験システム(SL-TRANS)の概要、本予稿集、1989
- [3] 柿ヶ原・森元、SL-TRANSにおける文節候補の削減、本予稿集、1989