社団法人 電子情報通信学会
THE INSTITUTE OF ELECTRONICS,
INFORMATION AND COMMUNICATION ENGINEERS

信学技報
TECHNICAL REPORT OF IEICE.
NLC98-31 (1998-10)

# PGLR パーザの異音モデルへの適用について

今井 宏樹　宮崎 一浩　田中 穂積　徳永 健伸
東京工業大学大学院 情報理工学研究科
〒 152-8552 東京都目黒区大岡山 2-12-1
{imai,iri,tanaka,take}@cs.titech.ac.jp

竹澤 寿幸
ATR 音声翻訳通信研究所
〒 619-0237 京都府相楽郡精華町光台 2-2
takezawa@itl.atr.co.jp

あらまし　本論文では，音声認識処理に使用する言語モデルとして PGLR モデルを導入する試みについて述べる．PGLR モデルは，GLR パーザで生成される構文木に対してその生成確率を与える確率モデルであり，文脈依存性を扱うことができる，終端記号の連接情報 (bigram) を同時に扱うことができる，等の利点がある．一方，音声認識においては，精度の良い確率言語モデルを構築することが重要な課題となっている．本研究では，異音を終端記号とする CFG を用いて PGLR モデルを構築し，異音の bigram や trigram のみの確率モデルと比較実験を行ない，それぞれのモデルの性能をテストセットパープレキシティにより評価した．

キーワード　PGLR モデル，異音モデル，音声認識，テストセットパープレキシティ

# Application of a PGLR parser to an allophone model in speech recognition

IMAI Hiroki† MIYAZAKI Kazuhiro† TANAKA Hozumi†
TOKUNAGA Takenobu† TAKEZAWA Toshiyuki‡
†Graduate School of Information Science and Engineering, Tokyo Institute of Technology
2-12-1 Ookayama, Meguro-ku, Tokyo 152-8552, Japan
‡ATR Interpreting Telecommunications Research Laboratories
2-2 Hikaridai, Seika-cho, Soraku-gun, Kyoto 619-0237, Japan
{imai,iri,tanaka,take}@cs.titech.ac.jp takezawa@itl.atr.co.jp

Abstract　We describe an attempt to apply the PGLR model to an allophone model in speech recognition. The PGLR model is a stochastic model that returns the probability of each generated parse tree. The PGLR model has the advantage that it is mildly context sensitive and at the same time readily interfaces with information about the connectability of terminal symbol pairs (so called bigrams). The construction of a good stochastic language model is one important task in speech recognition. We built a PGLR parser using a CFG with allophone symbols and compared our model with both allophone-level bigram and trigram models in terms of test-set perplexity.

key words　PGLR model, allophone model, speech recognition, test-set perplexity

# 1 はじめに

音声認識の分野において，音響モデルとともに精密な言語モデルの作成が重要な課題のひとつとなっている．最近では，単語 n-gram モデルを基本とした言語モデルが主流であり，単語の bigram と trigram を併用して，認識精度を落とさずに計算量を削減する手法が提案されている [11].

しかしながら，音声は言語と深く結び付いており，単純に音声から単語列を取り出すだけでなく，構文処理や意味処理などのより深い言語処理と関連付ける必要があると我々は考えている．このような視点から，我々の研究グループでは，構文解析アルゴリズムのひとつである GLR 法 [9] を基本として，形態素解析や音声認識との統合を目指した研究を進めてきた [12][6][10].

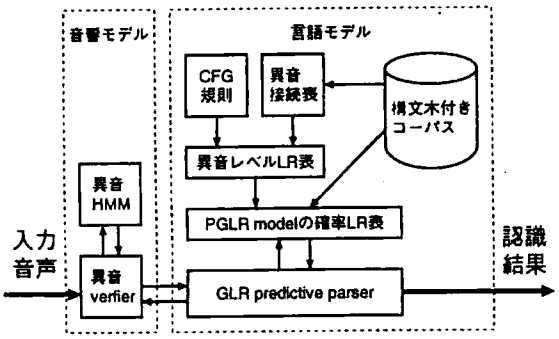


図 1: PGLR モデルを組み込んだ HMM-LR システム

現在は，図 1 に示すような，HMM-LR 音声認識システム [5] を拡張した認識システムの構築を目指している．GLR 法では，LR 表を利用した先読み記号の予測が可能なため，音声認識に応用すれば音素の予測に役立つとされている．このシステムでは，

- n-gram モデルよりも精密な言語モデルの提供
- 音声認識・形態素解析・構文解析の統合的な解析

が期待できる．

本研究では，確率 LR パーザを用いた言語モデルを作成し，n-gram を基本とした統計的言語モデルとの比較を行なうことを目的とする．2 節では，本研究で対象とする PGLR モデルについてその概要を述べ，3 節で異音モデルへの適用について説明する．4 節では，音声対話コーパスを用いた予備実験の結果を報告する．

# 2 PGLR パーザ

PGLR パーザは，PGLR モデル [1][3] と呼ばれる確率モデルを組み込んだ LR 表を用いて構文解析を行なうパーザである．本節では，PGLR モデルの概要および特徴を説明する．

## 2.1 PGLR モデル

PGLR モデルは，GLR パーザ [9] を拡張して，生成された構文解析木に対してその生成確率を付与するための確率モデルである．GLR パーザは，構文的曖昧性を含む一般の CFG も扱えるように LR パーザを拡張したものであり，与えられた CFG からあらかじめ LR 表と呼ばれるプッシュダウンオートマトンを作成し，LR 表に記述された動作にしたがって解析を行なう．GLR パーザでは，初期状態から解析成功までに実行された一連の動作により生成される状態遷移系列が 1 つの解析木に相当する．したがって，状態遷移系列の生起確率が構文木の生成確率と等価になる．

このような考え方で GLR パーザにおける状態遷移系列に確率を与える手法は，Briscoe らにより最初に提案された [1]．しかし，彼らの手法では正規化に問題があり，計算された値が構文木の生成確率とはなっていなかった．Inui らはその問題点を指摘し，構文木の生成確率を正しく与えることができるようなモデルを提案した [3]．我々は，本研究で Inui らのモデルを採用する．ここからは，単に PGLR モデルと称したときは Inui らのモデルを指すこととする．以下に PGLR モデルの構成方法の概略を示す．

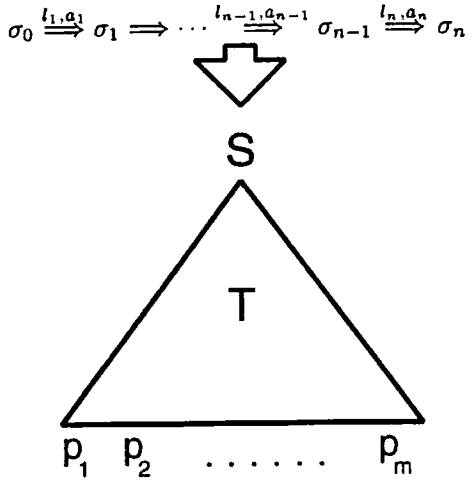


図 2: 解析過程と構文木の関係

図 2 において，$\sigma_i, l_i, a_i$ はそれぞれパーザの状態，先読み記号，実行された動作を表している．初期状態 $\sigma_0$ から受理状態 $\sigma_n$ までのパーザの状態遷移の結果として木 $T$

が生成されるから，その生成確率 $P(T)$ は

$$P(T) = P(\sigma_0, l_1, a_1, \sigma_1, \ldots, l_n, a_n, \sigma_n) \quad (1)$$

と表せる．PGLR モデルでは，各解析ステップの実行される確率は，直前のパーザの状態のみに依存するという仮定を導入し，式 (1) を以下のように近似する．

$$P(T) \approx P(\sigma_0) \cdot \prod_{i=1}^{n} P(l_i, a_i, \sigma_i | \sigma_{i-1}) \quad (2)$$

さらに，LR 表においては，

- $l_i$ と $a_i$ が決まれば $\sigma_i$ は必ず一意に決まる．
- reduce 動作では先読み記号を消費せず，直前の動作と同じ先読み記号が用いられる．よって，reduce 動作では，先読み記号を予測する必要がない．

という 2 つの特徴があるため，式 (2) の各条件付き確率の推定は式 (3),(4) で行なうことができる．

$$P(l_i, a_i, \sigma_i) \approx \begin{cases} P(l_i, a_i | s_{i-1}) & (s_{i-1} \in S_s) \quad (3) \\ P(a_i | s_{i-1}, l_i) & (s_{i-1} \in S_r) \quad (4) \end{cases}$$

ただし，$S_s$ は shift 動作直後に遷移する状態の集合，$S_r$ は reduce 動作直後に遷移する状態の集合をそれぞれ表す．

式 (3),(4) の右辺のそれぞれの条件付き確率は，訓練コーパスを GLR パーザで解析して，式 (3) に対しては状態 $s_{i-1}$ において $l_i$, $a_i$ が発生する回数，式 (4) に対しては状態 $s_{i-1}$，先読み記号 $l_i$ において $a_i$ が発生する回数から，それぞれ学習することができる．そして，LR 表に定義されている各動作に対してこの条件付き確率を割り当てることにより，PGLR モデルが構築される．

### 2.2 PGLR モデルの特徴

PGLR モデルには以下のような特徴がある．

1. 文脈依存性が考慮される
   PGLR モデルでは，その構築過程において LR パーザの状態が考慮される．パーザの状態はその時点までに解析されたスタックの状態に依存するため，左文脈にある程度依存していると考えることができる．
   また，Li によって提案された，bigram LR 表 [6] と呼ばれる終端記号の bigram 確率を LR 表に組み込む手法と比較すると，bigram LR 表では shift 動作後に遷移する状態 $S_s$ のみに対して確率値を割り当てているのに対し，PGLR モデルでは reduce 動作後に遷移する状態 $S_r$ に対しても確率値を割り当てている．
2. 終端記号の連接情報が考慮される
   式 (3) は，スタックの状態 $\sigma_{i-1}$ から次の先読み記号 $l_i$ を予測するモデルとなっている．一方，スタックの状態はそれまで解析済みの入力列 $l_1, \ldots, l_{i-1}$ に関する構文情報を表しているので，終端記号の連接確率 $P(l_i|l_{i-1})$ の前件部分をスタックの状態 $\sigma_{i-1}$ によってさらに細分化していると考えることができる．したがって，PGLR モデルは終端記号の bigram モデルと同様な終端記号の連接情報が反映される．
3. 比較的容易にモデルを学習できる
   モデルの学習はコーパスを解析して使用された各動作の回数を数え上げることにより行なわれる．与えられた各例文に対して正解となる構文構造が与えられていれば，学習のコストは bigram や trigram の学習と比較してもそれほど大きくはならない．

## 3 PGLR パーザの異音モデルへの適用

この節では，まず，音声認識で使用される認識単位と異音モデルの関係について述べ，次に異音モデルを終端記号とする CFG を作成し PGLR モデルを構築する手順を示す．

### 3.1 異音モデル

音声認識の分野においては，音声の認識単位として音素を使用するのが一般的である．しかしながら，同じ音素であっても，前後の音素との連接関係により音響的特徴も変化することが知られている．そこで，認識単位をモデル化する際に，音素の文脈依存性が反映されることが望ましい．

異音モデルは，このような音素文脈依存モデルのひとつである．音素文脈依存モデルでは，全ての文脈パターンを別々に扱うと学習や認識のコストが増大するため，音響的に似ている文脈パターンをクラスタリングするのが一般的である．

異音モデルは，全ての文脈パターンを 1 つの状態で表すところから始めて音響的特徴の差異の大きい文脈から順に状態分割を行なう SSS (Successive State Splitting)[7] と呼ばれる手法を用いて文脈のクラスタリングを行なっている．トップダウンなクラスタリングのため，全ての文脈パターンは必ずいずれかのクラスタに属し，学習用音声データが多くなくても比較的良いクラスタを作成することが可能である，といった利点がある．

我々は，ATR で作成された 1548 個の異音モデルを使用する．この異音モデルの一部を図 3 に示す．

```
元の辞書規則:   <koyuu-meisi> -> ニューヨーク
異音展開後の辞書規則:   <koyuu-meisi> -> <_n> j5 u25 u26 j7 o244 o402 k38 <_u>
異音導出規則:   <_n> -> n1
                <_n> -> n2
                     :
                <_n> -> n25
                <_u> -> u1
                     :
                <_u> -> u48
```

注: 非終端記号は便宜上 "<>" で囲んである．また，"_" で始まる記号は音素を表している．

図 4: 辞書規則の異音列への展開

| 基になる音素体系 (計 26 音素) |
|---|
| - a i k j o zh z u d m g ch |
| ng r sh ts s e b q t w n p h |

| 音素 /b/ の異音 | | |
|---|---|---|
| 音素文脈 | | ラベル |
| 左 | - i j zh u r e q w | b1 |
| 中 | b | |
| 右 | i k j zh z u d m ch ng r sh e b q t n p h | |
| 左 | - i j zh u r e q w | b2 |
| 中 | b | |
| 右 | - a o g ts s w | |
| 左 | a k o z d m g ch ng sh ts s b t n p h | b3 |
| 中 | b | |
| 右 | i k j zh z u d m ch ng r sh e b q t n p h | |
| 左 | a k o z d m g ch ng sh ts s b t n p h | b4 |
| 中 | b | |
| 右 | - a o g ts s w | |

図 3: 音素 /b/ の異音モデル

### 3.2 異音を終端記号とする CFG の作成

異音を終端記号とする CFG を作成するためには，異音列に展開された単語辞書規則を用意する必要がある．具体的には，次のような手順で辞書規則を作成する．

1. 辞書規則の右辺の単語を音素列に展開する．
2. 音素列を前後の文脈を見ながら異音列に変換する．ただし，先頭と末尾の音素は文脈が決定できないのでそのままにしておく．
3. 各音素ごとに，

   音素 → 異音

   なる展開規則を追加する．

図 4 に，「固有名詞 → ニューヨーク」という辞書規則を異音列に展開した例を示す．

### 3.3 PGLR モデルの生成

2.1 節で示した PGLR モデルの定義に基づいて確率 LR 表を生成する手順は，以下の通りである．

1. 学習用例文を異音列に展開し，異音の連接情報を抽出する．
2. CFG と 1. で得られた連接情報から制約伝播アルゴリズム [8] を用いて LR 表を作成する．
3. 生成された LR 表を使用して学習用例文を解析し，使用された shift 動作および reduce 動作の回数を数え上げる．
4. PGLR モデルの定義式 (3),(4) に従い，各動作に対して確率値を割り当てる．

ここで問題となるのが，学習用データの不足によるデータスパースネス問題である．文法の規模が大きくなると LR 表中に定義される動作の数も増えるため，学習用データの解析では使用されない動作が多数現れる．そこで，我々は，あらかじめ全ての動作に対して一定の低い頻度を与えるフロアリングを行なっている．4 節の実験では，フロアリングの値を変化させて，最も良い結果を得た値 0.1 を採用している．

## 4 対話コーパスを用いた評価実験

### 4.1 コーパスと文法

本研究では，音声対話を書き起こした対話コーパスを使用した予備実験を行なった．実験用の対話コーパスとして，ATR で収録された，618 対話，約 21000 文の対話コーパスを使用した．このコーパスには，形態素情報と構文情報がそれぞれ付与されている．文法は，衛藤らがこのコーパスの解析用に作成した日本語句構造 CFG[14] にコーパスから自動的に作成した辞書規則，異音導出規則をそれぞれ結合させたものを用いる．この文法の詳細を表 1 に示す．

今回は，コーパス全体のうち，衛藤の文法体系で解析可能で，かつ半自動的に正解の構文木を付与できた 8244 文を実験に用いた[1]．その中からテストセットとして 400 文をランダム抽出し，残りを学習用データとした．

[1]衛藤文法は，元のコーパスに比べて品詞が細かく分類されている，動詞と後置詞句の係受け関係が文法に記述されている，等の理由により，コーパスの構文木をそのまま使用することができない．

表 1: 実験に使用した文法

| | | | |
|---|---|---|---|
| 総規則数 | 7391 | 異音の種類 | 1548 |
| 辞書規則の数 | 4986 | 細品詞の種類 | 442 |
| 平均規則長 | 5.28 | | |

## 4.2 評価方法

本研究では，言語モデルの評価尺度としてテストセットパープレキシティ[4] を用いた．パープレキシティは，解析のある時点での次に予測されるシンボル (ここでは異音となる) の候補数の平均値を表し，その値が小さいほど良いとされる．音声認識の分野においては，与えられたタスクの難しさの評価や言語モデルの性能比較等において比較的よく用いられる．

PGLR モデルと比較するための言語モデルとして，現在音声認識で一般的に用いられている bigram モデルと trigram モデルを選んだ．また bigram LR 表との比較も行なった．2.2 節で述べたように，PGLR は bigram モデルよりは良い性能を持つことが予想できるが，trigram との性能差については明らかではない．

本研究では，4.1 節のテストセットに対して各手法で文生成確率を求め，そこからテストセットパープレキシティを計算してその値を比較した．ただし，bigram LR 表と PGLR モデルの文生成確率は，確率値の上位 10 位までの構文木の生成確率の和で近似した．曖昧性の多い文に対して全ての構文木の生成確率値を求めるのが困難であること，下位の構文木の生成確率は上位のそれと比較して無視できる程度に小さくなることが予想されること，等がその理由である．

## 4.3 結果と考察

表 2: 各言語モデルのテストセットパープレキシティ

| 言語モデル | test-set perplexity |
|---|---|
| bigram | 3.39 |
| bigram LR 表 | 3.22 |
| trigram | 2.56 |
| PGLR | 2.13 |

表 2 に各言語モデルごとのテストセットパープレキシティの値を示す．4 つのモデルの中で，PGLR モデルがもっとも小さい値を示している．

Imai らは，bigram LR 表はパープレキシティの値において bigram モデルよりは良い結果を示したが，trigram モデルほどは良い結果を得られなかった，と報告している [2]．PGLR モデルは，式 (3),(4) に示されるように，shift 動作直後の状態 $S_s$ と reduce 動作直後の状態 $S_r$ の両方に確率値を割り当てる．一方，bigram LR 表では，式 (4) に相当する，状態 $S_r$ に対しての確率値の割り当てが行なわれない．この $S_r$ への確率値の割り当ての有無が，trigram モデルに対するパープレキシティの相対的な差の要因となっていると考えることができる．

# 5 まとめ

本論文では，PGLR パーザを音声認識の言語モデルに応用する試みについて報告した．ATR の音声対話コーパスを用いた比較実験では，bigram や trigram モデル，および bigram LR 表に比べて低いパープレキシティの値を示した．したがって，音声認識の言語モデルとして有効であると期待される．

今後の課題としては，以下の点が挙げられる．

- 大規模な評価実験
  今回は対話コーパスの一部を使用しての実験だったため，各言語モデルの学習が不十分であった．十分な学習が行なえる程度の規模で実験を行なう必要がある．

- 複数の接続制約の LR 表への組み込み [10] を用いた手法の利用
  今回の実験では，LR 表を作成する際に異音の連接関係のみ LR 表に組み込まれているが，細品詞レベルの連接関係も同時に組み込むことにより，さらなる精度向上が期待できる．現在，同じコーパスを用いてこの手法を用いた実験を行なっており，機を改めてその結果を報告する予定である．

- 音声認識モジュールと統合して認識実験を行なう
  現在使用している PGLR パーザはテキスト解析用であり，音声認識部との統合が行なわれていない．今後，実際の音声データを使用した認識実験を行ない，どの程度の効果があるかを確認したい．現在，結合する音声認識モジュールとして，IPA のディクテーションソフトウェア [13] を利用することを検討中である．


# 謝辞

本研究を行なうにあたり，異音モデルを提供いただいた ATR の林輝昭氏，対話コーパス用 日本語 CFG を提供い

ただいたランゲージウェアの衛藤純司氏にそれぞれ感謝致します.

## 参考文献

[1] T. Briscoe and J. Carroll. Generalized probabilistic LR parsing of natural language (corpora) with unification-based grammars. *Computational Linguistics*, Vol. 19, No. 1, pp. 25–59, 1993.

[2] H. Imai and H. Tanaka. A method of incorporating bigram constraints into an LR table and its effectiveness in natural language processing. In *New Method in Language Processing and Computational Natural Language Learning*, pp. 225–233, 1998.

[3] K. Inui, V. Sornlertlamvanich, H. Tanaka, and T. Tokunaga. A new formalization of probabilistic GLR parsing. In *International Workshop on Parsing Technologies*, 1997.

[4] F. Jelinek. Self-organized language modeling for speech recognition. In A. Waibel and K.F. Lee, editors, *Readings in Speech Recognition*, pp. 450–506. Morgan Kaufmann, 1990.

[5] K. Kita, T. Kawabata, and H. Saito. HMM continuous speech recognition using predictive LR parsing. In *ICASSP89*, pp. 703–706, 1989.

[6] H. Li. Integrating connection constraints into a GLR parser and its applications in a continuous speech recognition system. Technical Report TR96-0003, Department of Computer Science, Tokyo Institute of Technology, 1996.

[7] J. Takami and S. Sagayama. A successive state splitting algorithm for efficient allophone modeling. In *ICASSP92*, pp. I-573–576, 1992.

[8] H. Tanaka, H. Li, and T. Tokunaga. Incorporation of phoneme-context-dependence into LR table through constraint propagation method. *Journal of Japanese Society for Artificial Intelligence*, Vol. 11, No. 2, pp. 246–254, 1996.

[9] M. Tomita. *Efficient Parsing for Natural Language: A Fast Algorithm for Practical Systems.* Kluwer Academic Publishers, 1986.

[10] 綾部寿樹, 徳永健伸, 田中穂積. 複数の接続表の制約の LR 表への組み込み—LR 表工学 (2). 情報処理学会研究報告, NL-117-10, pp. 67–74, 1997.

[11] 李晃伸, 河原達也, 堂下修司. 単語トレリスインデックスを用いた大語彙連続音声認識エンジン JULIUS. 電子情報通信学会技術研究報告, SP98-3, 1998.

[12] 植木正裕, 徳永健伸, 田中穂積. EDR 辞書を用いて形態素解析と統語解析を行なうシステム. EDR 電子化辞書利用シンポジウム論文集, pp. 33–39, 1995.

[13] 河原達也, 李晃伸, 小林哲則, 武田一哉, 峯松信明, 伊藤克亘, 伊藤彰則, 山本幹雄, 山田篤, 宇津呂武仁, 鹿野清宏. 日本語ディクテーションソフトウェア (97 年度版) の性能評価. 情報処理学会研究報告, SLP-21-10, pp. 109–114, 1998.

[14] 田中穂積, 竹澤寿幸, 衛藤純司. MSLR 法を考慮した音声認識用日本語文法 -LR 表工学 (3)-. 情報処理学会研究報告, SLP-15-25, pp. 145–150, 1997.

福山写真製版所