Bigramの使用による話し言葉用確率文脈自由文法の自動学習

中川聖一[†] 大谷耕嗣[†,☆]

本論文では文のカバー率を改善するための文法規則の自動学習について述べる．この方法は文法規則が登録されていないために解析できない文を解析することを可能にする．システムに入力された文が文法規則が不備なために受理できないとき，システムがこの入力文を使って規則の学習をすることにより文のカバー率を改善する．文法の学習の3つの方法を比較検討する．1つは，文脈自由文法（CFG）の規則の追加学習法，2つめは単語ペアの追加学習法，3つめは非終端記号のペアの追加学習法である．以上の方法について未登録単語を登録した場合としない場合についての生成規則を学習した後で，文のカバー率とパープレキシティについて調べた．さらに単語ペアと文脈自由文法（CFG）の確率化を行うバイグラムと確率文脈自由文法（SCFG）の使用による文法の学習とその評価についても述べ，最後に音声認識に適用した結果を述べる．

Automatic Learning of Stocastic Context-free Grammar for Spontaneous Speech by Integration of Bigram

SEIICHI NAKAGAWA[†] and KOJI OTANI[†,☆]

In this paper, we describe an automatic learning method of stochastic grammar rules for improving coverage of input sentences. It is possible to analyze the sentences that can not be accepted by preregistered production rules. When a sentence which is not accepted by production rules is inputted to the system, the system learns new rules using this input sentence. As a result, the coverage of the sentence can be improved. We propose three methods for learning the grammar. One is to add new production rules to the original context-free grammar (CFG). The second is to add word pairs. The last is to add nonterminal symbol pairs. In registering unknon words or not, we evaluate the coverage of sentences and perplexity using above methods. Next we evaluate the perplexity using bigram and stochastic-CFG instead of word pairs and CFG. Finally we implemented these learning methods on speech recognition and describe the results.

1. はじめに

自然言語理解システムの目的の1つに人間・機械間のコミュニケーションの方法としての人間にやさしいインターフェイスの開発があげられる．我々は，“富士山周辺の観光案内システム”のタスクで自然発話を使った対話システムの開発を行っている[1]．

対話システムはユーザが発話した言葉を受理するための文法を使って入力した文を認識する．しかし，もしユーザがシステムの文法で受理できない文を話したときにはシステムはその文を認識することができない．ところがシステム開発者があらかじめ完全な文法を構築するのは難しい．そこで，この原因による認識間違いを減らすために，ユーザがシステムの文法で受理できない文を発話したときに，その文を使ってシステムの文法に登録されていない単語や規則の学習を行い，これらの文を受理できるようにするシステムの開発を行ってきた[2)~4)]．その結果，新しい入力文に対するカバー率の改善ができるようになった．音声認識用の文法は，入力文のカバー率（受理率）が大きくパープレキシティが小さいのが望ましい．パープレキシティは2のエントロピー乗で定義されるもので，パープレキシティが大きいほど，生成される文の数が多くなる[15)]．直観的にいえば，発声されうる文を受理し発声されえない文を受理しない文法が望ましい．もっと一般的にいえば，発声される頻度・可能性の高い文は高い確率で受理できる確率文法が望ましい[16)]．そのため，よりよい文法構造を見い出すというのが目的でなく，あくまでもカバー率が大きくパープレキシティの小さい

† 豊橋技術科学大学情報工学系
Department of Information and Computer Sciences, Toyohashi University of Technology
☆ 現在，株式会社トヨタテクノサービス
Presently with Toyota Techmo Service Co.

文法を学習するのが目的である．しかし，副次的な結果としてより良い文法構造を獲得できることが期待できる．

文法獲得の過去の研究で，大量のテキストデータからInside-Outsideアルゴリズムで文脈自由文法（CFG）を学習する方法が検討されている[5]が，非終端記号が増えると学習は非常に難しい．効率良い学習のためには何らかの初期文法が必要である．中澤ら[6]はシステムに入力された例文の推測された導出木と現在の文法構造の差から新たな規則が必要となる場所を推定し，生成規則の仮説を，削除・位置交換・挿入・置換を行うことにより，CFGを効率的に学習する方法を調べている．白井ら[7]は，構文構造付きコーパスの内部ノードに非終端記号を与えて確率文脈自由文法（SCFG）を抽出し，その文法を改善することにより適用範囲の広い文法を抽出している．Brill[8]は，句構造の大変ナイーブな状態の知識から始めるアルゴリズムの研究を行っている．これは，トレーニングコーパスで与えられる解析構造を示す括弧と現在の状態の括弧の結果を比較することを繰り返すことによって，システムがエラーを減らすことができる簡単で構造的な変換のセットを学習する．Millerら[9]は文法内のローカルな文脈的な情報を使う方法と，SCFGの導入の効果について調べている．Samuelsson[10]は，解析木のノードに対してのエントロピーをしきい値として使うことによって文法の統合を行う研究を行っている．藤崎ら[11]は確率的有限状態オートマトンの状態遷移図をコーパスから自動的に獲得し，文解析に利用する研究を行っている．

また，音声認識を前提とした研究においては，Wrightら[12]のロバストなパーザのためにCFGとバイグラムの併用を提案している方法がある．しかし，それらは統合されておらず並列に実行される．最近では竹澤ら[13]が，ポーズ情報で区切られた区間を部分木で表現し，部分木出力による音声認識実験に対して，前終端記号バイグラムを利用した再順序付けを行う方法を提案している．これは我々が提案した方法[3]とほぼ類似な考え方である．

これらの研究はWrightらや竹澤らの方法を除いて，いずれも話し言葉の音声認識を目的で開発されたものはなく，言語の制約とロバスト性に欠けると思われる．

本論文では話し言葉用の文法を学習するために3つの方法を提案している．1つめは，規則の不備のためのCFGの生成規則を登録する方法である．2つめは接続可能な単語ペアまたは単語クラスペアを登録する方法である．単語クラスとは意味論的によく似た単語のセット（たとえば，湖名：河口湖，山中湖，精進湖等）のことである．3つめは接続可能な非終端記号のペアを登録する方法である．生成規則の登録は，ボトムアップのパーザを使って新しい適当な単語または非終端記号のペアの登録を行う．

3つの登録方法を比べたところ，単語クラスペアがカバー率とパープレキシティのバランスが一番良いことが分かった．また，CFGと単語（単語クラス）ペアに確率を付与した確率文脈自由文法（SCFG）と単語（単語クラス）バイグラムを学習し，パープレキシティを抑えることを試み，その有効性を示した．

2章でシステムの認識エラーの原因について，3章で生成規則の登録アルゴリズムについて，4章で確率の導入について，5章で本手法のテキスト入力に対する評価結果と音声認識への適用結果について述べる．

2. システムの認識エラーの原因

我々が研究をすすめている“富士山観光案内”のための対話システム[1]での認識エラーは以下の原因によって生じる．

- 文法部のエラー
- 認識部のエラー

前者のエラーはユーザが受理できない未登録単語（語彙外の単語）を含む文を話したときか，生成規則で受理できない文を話した（文法外）ときのエラーである（もちろん，これはシステムの語彙と文法規則がまだ完全ではないともいえる）．たとえば，次のような文が初期文法（システム開発者が人手で作成したもの）では受理できなかった．

× キャンプをしたいんんですけど何千円ぐらいでできるでしょうか．
× 運動したいんですけれども専用のテニスコートもしくはスポーツができる所はあるんでしょうか．
○ コートは何面ぐらいで時間はいくらぐらいですかね．
○ 富士サファリパークの近くに泊まりたいんですけど旅館はどこにありますか．
○ 遊覧船で富士五湖をまわるっていうのも楽しくていいと思います．

このような受理できなかった文を受理できるように文法を人手で修正するのはきわめて難しい．（ここではもちろん，どのような単語列でも受理できるようなトリビアルな文法を意味していない．非文は受理しないという前提がある．）なお，○印のついた例文は学習後受理できるようになった文である．

また後者のエラーは，ユーザが文法で受理できる文

を話したにもかかわらずシステムの音声認識部がユーザの発話を正しく認識できなかった場合のエラーである．

以上の2つの問題を解決することによって認識エラーを減らすことができる．本論文では，前者のエラーを減らすために受理できなかった文を学習に使うことによって，文法の新しい規則を半自動的または自動的に登録する方法について述べる．

前者のエラーには3つの場合がある．それらは以下の登録の欠落である．

- 単語
- 生成規則
- 単語と生成規則

システムは1番目と2番目のエラーについて単語と生成規則の登録を行う．3番目のエラーに対しては現在は対応していない．そこで，3番目のエラーに対しては未登録単語を人間によって完全に登録した後，生成規則の登録の評価を行った．

3. 生成規則登録のアルゴリズム

本章では例文を使った文法規則を登録するアルゴリズムについて述べる．生成規則の登録は文法の学習に3つの方法を使う．1つはボトムアップパーザを使いCFGの新しい適当な生成規則を作り，これらの新しい規則を登録するもので，2つめの方法は部分的にパーズされたストリングの間の単語のペアまたは単語クラスのペアを登録するもので，最後の方法は部分的にパーズされたストリングのそのストリングに対する非終端記号のペアを登録するものである．

3.1 CFGでの規則生成

CFGの生成規則の半自動的な登録方法をここでは簡単に説明しておく[4]．CFGの生成規則登録の要点を以下にまとめた．

(1) システムが入力文の解析に失敗したときに，ボトムアップパーザを使い入力文の部分解析木を作る．

(2) 入力文のすべてをカバーする部分解析木の組合せのうち組合せの数が最小になる組合せを選ぶ．

(3) 先の部分解析木の組合せで，適当な一般性を与えるために**図1**に示す3つの登録方法を使い分ける．

(a) ある部分解析木の非終端記号を適当に変えたとき，入力文が受理できるかどうかを調べる．たとえば，図1(a)で$X \rightarrow P1$を登録したとき，この登録により修正されたシステムの文法を使ってこの入力文が受理可能かどうかを調べる．

(b) 2つの部分解析木の非終端記号をつなげたとき，入力文が受理できるかどうかを調べる．たとえば，図1(b)で$X \rightarrow P1\ P2$を登録したとき，この登録により修正されたシステムの文法を使ってこの入力文が受理可能かどうかを調べる．

(c) 部分解析木の最小の組合せを直接開始記号につなげる．たとえば，図1(c)で開始記号に直接の接続，たとえば$S \rightarrow P_1P_2P_3$を作りそれを登録する．

登録の優先順位は (a) > (b) > (c) とした．なぜなら，この順序が特殊化のしすぎを小さくすると思われるからである．以上の方法によりCFGの生成規則の登録を行う．

(4) 先の登録方法でいくつかの候補があるのなら，ユーザが適当な候補を判断するためにシステムにその候補を使っていくつかの例文を作らせユーザにふさわしいものを選んでもらう．

以上の方法により生成規則の登録を行う．

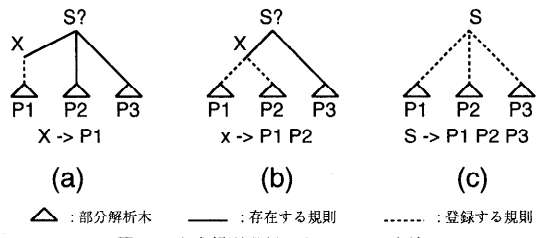


図1 生成規則登録の際の3つの方法
Fig. 1 Three ways for registering production rules.

3.2 単語ペアを使った規則の追加

単語（単語クラス）ペアを使うことによる生成規則の不備により解析できない文のための規則の登録方法について述べる．この登録方法はCFG規則の補助として単語（単語クラス）対制約を登録する方法で，以下の手順で自動的に登録を行う．

(1) システムが入力文の解析に失敗したときに，ボトムアップのパーザを使って入力文の部分解析木を作る．

(2) 入力文をカバーできる部分解析木の組合せで，木の組合せの数が最小となる組合せを見つける．

(3) 部分解析木の組合せの各解析木に隣接する単語（単語クラス）のペアを登録する．つまり，**図2**のように最小の部分解析木の数が2個なら，図中の部分解析木P_1の最後の単語w_1と，P_2の最初の単語w_2の単語（単語クラス）のペア(w_1, w_2)の登録を行う．

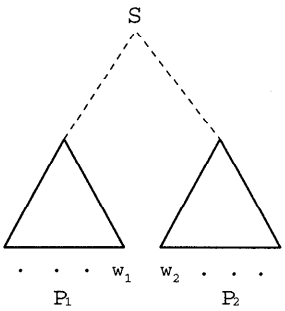


図2 単語対制約の登録
Fig. 2 Registration of word pair.

この方法で登録した単語（単語クラス）ペアは部分解析木を繋ぐ制約として使用する．ただし，実際には，CFGとペアを独立に用いて後続単語の予測を行う．つまり，CFG規則と単語ペアは文脈に関係なくいつも使えるとする．

一方，CFGを使わずに入力文から接続可能な単語のペアを獲得して，その単語ペアだけで文の解析をする方法についても比較検討した．

3.3 非終端記号のペアを使った規則の追加

単語（単語クラス）のペアの代わりに，より接続能力の強い非終端記号のペアの登録方法について述べる．この方法により単語ペアの方法よりカバー率の改善が期待される．登録方法は単語（単語クラス）ペアの場合とほぼ同じである．ただ，3.1節の方法と異なる点は，CFGの規則登録と違って任意の場所で使用可能な非終端記号のペアを登録する点であるが，規則の登録方法と疑似的に等価な関係にある．ペアによって予測される非終端記号（たとえばP_1に対するP_2）より下位の規則の展開に関してはCFGの規則を使う．また登録する部分解析木の非終端記号の組合せについては，一番トップにある記号のペアと一番ボトムにある記号のペアを登録した場合の両方について調べた．ただし，トップの記号より下位の非終端記号のペアは上位のペアに単語予測に関しては完全に包含される．

この方法も単語（単語クラス）ペアを登録したときと同様に非終端記号のペアを部分解析木を繋ぐ制約として使用する．ただし，実際には，CFGとペアを独立に用いて後続単語の予測を行う．つまり，CFG規則と非終端記号ペアは文脈に関係なくいつも使えるとする．

4. 確率の導入

音声認識では文法の評価規準としてパープレキシティ（複雑度）がよく用いられる．パープレキシティは2のエントロピー乗で計算される言語の複雑度を示す規準の1つであり，接続可能な後続単語の予測分岐数に相当し文法の複雑さを示すのによい指標である．

音声認識ではパープレキシティが小さくカバー率が大きい文法が望ましい．そこでパープレキシティを抑えるために確率の導入を単語（単語クラス）ペア，CFG各々に対して行った．

単語（単語クラス）ペアの単語または単語クラスのバイグラムの推定には学習セットとして使用している文を使用して任意の場所の単語のペアまたは単語クラスのペアの出現頻度からバイグラムの確率を推定する．つまり入力文（単語列）に対して，単語W_iの出現頻度$C(W_i)$と単語の2つ組の出現頻度$C(W_{i-1},W_i)$を数え上げる．このときのバイグラム確率は

$$P(W_i|W_{i-1}) = \frac{C(W_{i-1},W_i)}{C(W_i)} \quad (1)$$

で求まる．単語クラスバイグラムについても同様な式を用いて求める．ただ通常の規模では，1000文ぐらいの文を学習に用いた場合，単語ペアの数は十分でなく，データの不足を補うため単語ペアおよび単語クラスペアは学習データのすべての場所を使って求めている．

バイグラムとCFGにより並列で解析する場合のパープレキシティを求めるために，該当単語の予測確率を，CFGで予測される予測単語数の逆数（該当単語が予測されない場合は確率=0）とバイグラムによる予測確率との平均値とした．つまり

$$P(W_i|CFG,bigram) = \lambda P(W_i|CFG) + (1-\lambda)P(W_i|W_{i-1}) \quad (2)$$

（実験では$\lambda = 0.5$）

この逆数をとることによって該当単語の予測単語数（分岐数）とし，これらの相乗平均でパープレキシティを求めた．ただし，バイグラムは任意の場所で適用する．

単語クラスのバイグラムによる単語の出現予測確率は，単語クラスの出現予測確率をその単語クラスに含まれている単語数で割ることによって求まる．また，方法2の場合で，CFGで該当単語が予測されず，バイグラムによる予測確率が0の時には，バイグラムの確率の代わりにユニグラムで求めた確率を使って代用した．

一方，CFGの各規則に対する適用確率は，実験の際に使った評価セットのうち学習前の文法で解析可能な文を使い，これらの文に対する正しい解析木を調べ，使用された生成規則の頻度をカウントし，正規化を行うことによって求める．

5. 評価実験

5.1 評価条件

実験には“富士山周辺の観光案内”の対話システムの文法を使った．システム開発者が対話システムの入力文を想定して，人手で作成したオリジナルな初期文法の詳細は以下のとおりである．なお，パープレキシティは解析できるようになった全評価文を用いて求めている．ただし，学習データ数や言語モデルによって多少文集合が変わるが，評価データ（set 1）のうちオリジナルな元の文法で解析できた39文を使って求めても大差はない[4]．（ ）内の数値はSCFGでのパープレキシティである．

[初期文法]

- 終端記号数（単語数，語彙のサイズ）- 241
- 生成規則数 - 393
- 非終端記号数 - 137
- 前終端記号数（単語クラスのサイズ）- 110
- **パープレキシティ - 76.5(53.4)**

[未登録単語登録後]

- 終端記号数（単語数，語彙のサイズ）- 419
- 生成規則数 - 393
- 非終端記号数 - 137
- 前終端記号数（単語クラスのサイズ）- 198
- **パープレキシティ - 89.8(62.4)**

学習と評価で使われたデータは，53人の発話者にあらかじめ使える単語（名詞と動詞）を教えた条件で集められた計1020文を使用した．これらのうち106文を評価に，残りの914文を学習に使った．**表1**は学習セットの詳細を示している．表中の“受理可能文”は学習前のオリジナルな初期文法で受理された文の数を意味している．“解析失敗の原因”はエラーのタイプを意味している．“単語”と“規則”は単語または生成規則の不備により受理できなかった文の数を示している．“単語&規則”は単語と生成規則がともに不備であり，初期文法からの学習では対象外とした文の数を示している．

表1 学習・評価セットの詳細
Table 1 Detail of learning set.

データセット		受理可能文	解析失敗の原因			合計
			単語	規則	単語 & 規則	
set 1	評価用	39	0	35	32	106
set 2	学習用	51	5	32	18	106
set 3	学習用	58	4	22	16	100
set 4	学習用	31	5	21	43	100
set 5	学習用	41	6	19	34	100
set 6	学習用	50	3	25	22	100
set 7	学習用	43	5	24	28	100
set 8	学習用	57	4	26	13	100
set 9	学習用	38	7	25	30	100
set 10	学習用	50	7	19	32	108
合計		458	46	248	268	1020

学習・評価データで現れる未登録単語をあらかじめ登録した実験も行った．この初期文法の改善により，初期文法では対象外であった“単語&規則”の欄の文についても学習の対象となった（それでもなお，評価セットのうち8文が学習セットに出てこない単語が使われているため評価の対象外となっている）．表中のsetのうちset 1を評価に，その他のsetを学習に使った．受理可能になりうる文数は**初期文法での学習の場合39文＋35文＝74文，未登録単語を登録した場合では98文で，これが本研究の上限となる．**

また評価規準としてカバー率とパープレキシティを用いた．カバー率は評価セットの解析対象文のうち何文が解析できたかを示す．パープレキシティはここではテスト文集合に対するパープレキシティ（テストセットパープレキシティ）を求めている．

なお，規則や単語（単語クラス）の追加によってカバー率は増加するが，確率を使用しない場合は，もとのCFGを含んでいるのでパープレキシティも必ず増加する．

5.2 CFGでの規則登録結果

図3，5に示したCFGの生成規則の登録の実験結果では，システムは生成規則の不備により受理できなかった文の半分が受理できるようになり，受理可能な文のカバー率は39文から58文に増加した．しかし，パープレキシティが学習前に比べて約1.5倍以上になってしまった．また，解析に時間がかかりすぎるのも問題であることが分かった．

5.3 単語/単語クラス/非終端記号のペアを使った規則登録結果

単語（単語クラス）ペアについての学習結果も**図3～6**に示している．

学習セット内の未登録単語を登録しない場合の単語クラスのペアによる学習では，最終的に評価の対象となるセット1の74文中59文まで受理可能になった．また，パープレキシティの増加がそれほどでもなくカバー率の改善もCFGとほとんど変わらない．よって単語クラスペアの登録方法はCFGを使った登録方法よりも良いといえる[4]．

学習セット内の未登録単語を登録した場合の結果は，解析対象文が98文になったこともあるが，カバー数が77文（単語クラスペアの登録時）まで増えた（**図4**）．

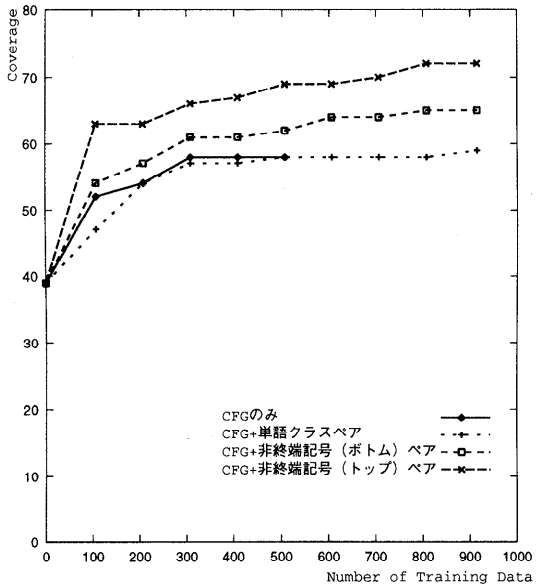


図3 学習によるカバー率の変化（未登録単語登録前：241単語）
Fig. 3 Result for improvement of coverage with 241 words.

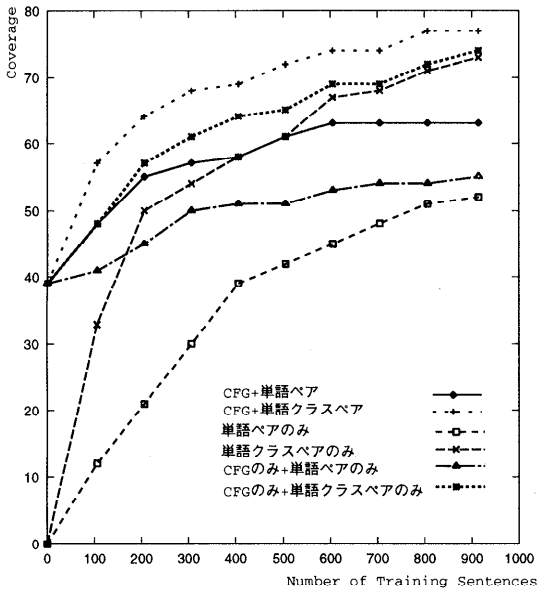


図4 学習によるカバー率の変化（未登録単語登録後：419単語）
Fig. 4 Result for improvement of coverage with 419 words.

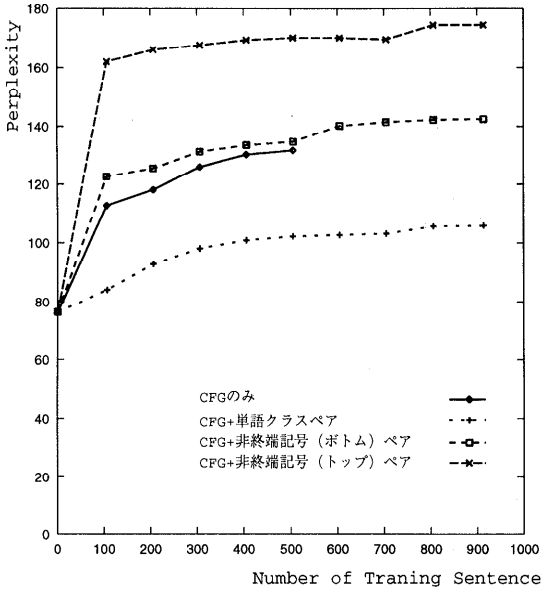


図5 学習によるパープレキシティの変化（未登録単語登録前：241単語）
Fig. 5 Result for improvement of perplexity with 241 words.

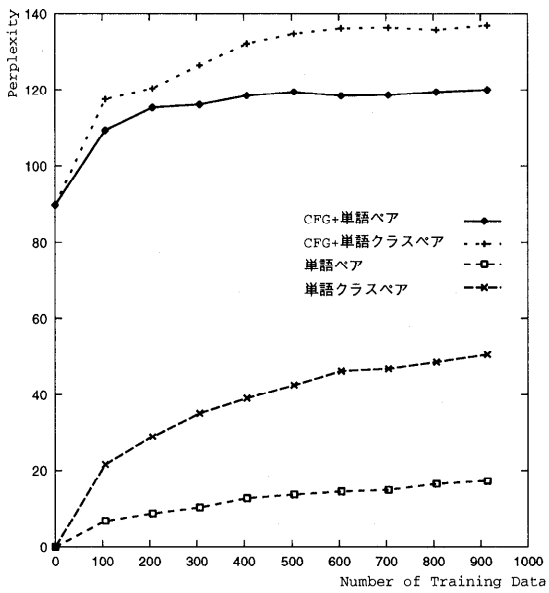


図6 ペアの登録によるパープレキシティの変化（未登録単語登録後：419単語）
Fig. 6 Result for improvement of perplexity with 419 words.

それにともないパープレキシティはかなり増加しているが，これは未登録単語を登録したことにより単語数が約1.7倍になっている点を考慮すれば，パープレキシティの増加は十分に抑えられているといえる．

一方，初期文法のCFGをまったく使わない単語ペアと単語クラスペアだけでの学習法による結果も示している．単語ペアだけの場合は，学習セットがまだ不十分ではあるが，他の方法と比べてまずまずの結果が得られた．また，単語クラスペアだけによる結果はCFGと単語クラスペアを組み合わせた結果と同程度のカバー率が得られ，パープレキシティをかなり小さ

くすることができた．単語クラスによって単語ペアでの学習の場合の学習セットの不足を補うことができることも確かめられた．

また非終端記号のペア（トップおよびボトムのペア）の学習の結果も示しているが，非終端記号のペアの学習によるカバー率の改善は，トップのペアを用いたときにはほぼ完全といってもよい結果となった．しかし，パープレキシティが170以上と非常に増加した．240単語ぐらいの文法でパープレキシティが170では毎回全単語の7割ぐらいを予測していると考えることもでき，音声認識への応用を考えると，この非終端記号のトップのペアを使う方法は良くないという結論になる．ボトムのペアも単語クラスペア＋CFGと比べてカバー数の割にパープレキシティが大きくなりすぎている．よって非終端記号のペアを使う方法は良くないといえる．

以上のことから十分な学習文数がある場合は，カバー率とパープレキシティの総合的な観点から未登録単語をすでに単語クラスに登録してあるのなら単語クラスペアのみによる学習方法が最も良い結果といえる．しかし学習文数が不十分な場合を考えたときには，ペアとCFGを組み合わせる方法のうち単語クラスペアを使う方法が良いといえる．

5.4 単語クラスのバイグラムによる効果

図7 はペアの代わりに確率を使って単語クラスバイグラムを学習した場合のパープレキシティの結果を示している（注：カバーレイジは単語クラスペアと同一である）．「バイグラム＋CFG」はペアをCFGの接続の補助に使った解析とCFGの解析を並列に行った場合で，「バイグラムのみ＋CFGのみ」はペアだけの解析とCFGだけの解析を並列に行った場合の結果である．

図7より単語クラスペアによる確率の導入によって，パープレキシティがかなり改善されることが分かった．また，バイグラムだけで解析を行うようにすればパープレキシティをかなり改善することができるので，単語クラスペアの確率化だけでなくCFG自体の確率化も行うとより効果があると思われる．これについては次節で述べる．

図7より「CFG＋バイグラム」と「CFGのみ＋バイグラムのみ」のパープレキシティは学習セットが少ない所で顕著な違いが出ていることが分かる．これは学習セットが少ないときは，単語クラスペアの接続能力の改善が低いため「CFGのみ＋バイグラムのみ」の場合はカバー率は小さい．一方，「CFG＋バイグラム」の場合は後続単語の予測に単語クラスペアを

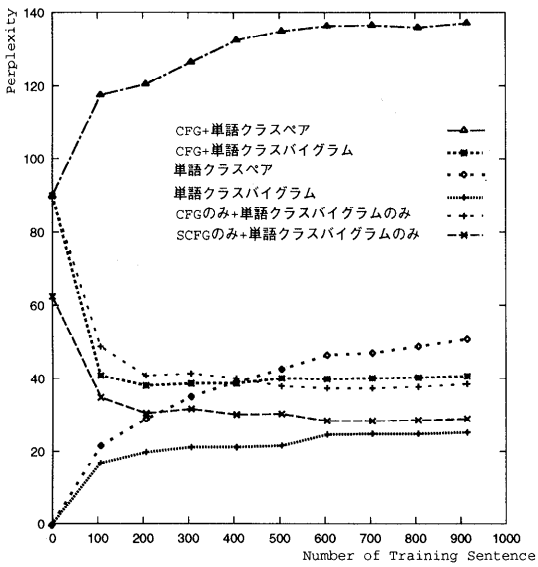


図7 バイグラムによるパープレキシティの変化（未登録単語登録後：419単語）

Fig. 7 Result for improvement of perplexity with bigram.

CFGの補助として使用することによりうまくカバーしているためこのような違いが出ている．一方，学習セットが増えてくれば単語ペアだけの解析でも解析文数がCFGぐらいに増え，適用箇所が増えるのでその差が最終的には逆転している．

式(2)のλの学習データ量ごとの最適な値を求めたのが**図8** である．図から明らかなように学習セットが少ないうちはCFGでの解析に重きを置き，学習が進むにつれ徐々にバイグラムに切り替えていけばいいことが分かる．これは我々の提案している手法がその中間ぐらいの学習データに対応できるように考えており，本手法の正しさが示されているともいえる．また学習効果の高い単語クラスバイグラムで学習した方が単語バイグラムよりもバイグラムへのシフトが速いことも分かる．

5.5 CFGの確率化によるパープレキシティの変化

バイグラムのみの解析とSCFGのみの解析を並列に行った際のパープレキシティを図7に示す．また比較のために同じ手法でCFGに対して行った結果も図7に示してある．

CFGをSCFGにしているのでその分パープレキシティが抑えられている．確率化による影響は学習データが少ない所では効果が大きいが，学習が進むとバイグラムの効果が大きくなりその差が小さくなる．

SCFGと単語（クラス）バイグラムの導入により，中語彙（500単語）程度のタスクなら，1000文ぐら

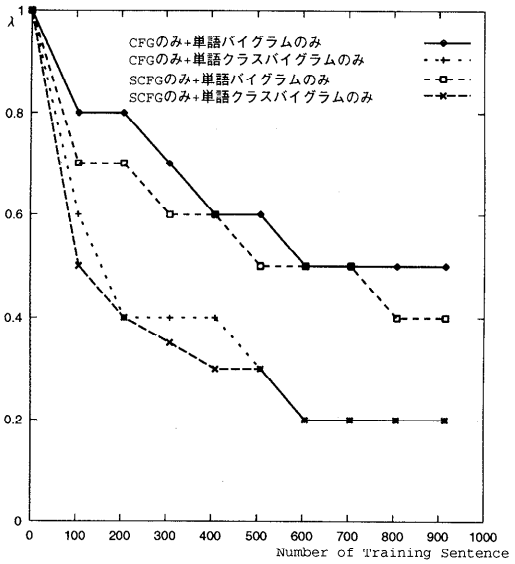


図8 学習データ量ごとのλの最適値（未登録単語登録後：419単語）

Fig. 8 The best weight at every training data set.

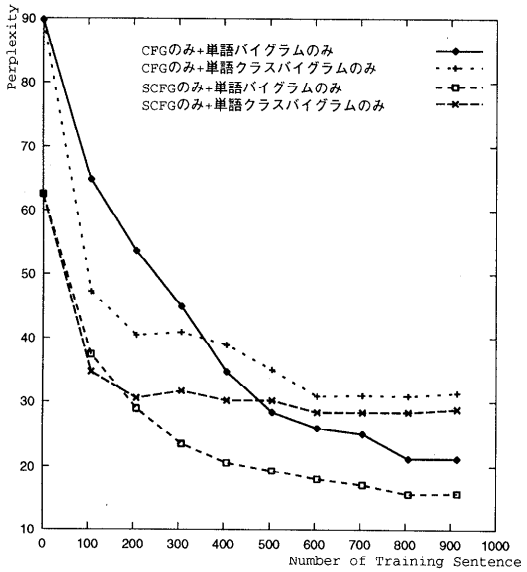


図9 学習データ量ごとの最適なλのときのパープレキシティ（未登録単語登録後：419単語）

Fig. 9 Perplexity with the best weight at every training data set.

いの学習データがあればパープレキシティを20程度にでき，音声認識の精度の向上が期待できることが分かった．なお，1000～2000語彙程度になっても（たとえば，米国ARPAのATISタスク），10000文以上の学習データによるトライグラムを用いれば，やはりパープレキシティを20前後に抑えることができると思われる[14]．しかし，一般のパーソナルなアプリケーションでは，1000文以上の学習データを収録することは大変なので，本論文で提案したような逐次学習法が有用であると考えられる．

5.6 音声認識実験による評価[17]

本節では，今までテキストレベルで評価してきた各登録方法が実際の音声認識においてどれぐらい有効であるかについて述べる．

(a) 実験条件

2人の学生の発声した文を認識対象にして，50文で話者適応化を行った話者適応モデルで認識実験を行った．使用した文法はテキストでの実験で使用したものから間投詞を省いたものである．評価データは，テキストでの実験と同様の106文を使用した．

連続音声認識は5状態4出力分布の連続出力分布型HMM（全共分散行列を使用）に基づく方法を使用した[18]．認識単位となる音節数は113である．音声の分析条件は**表2**のとおりである．音声特徴パラメータとして，10次元のLPCメルケプストラム係数と，動的特徴量としてそれの線形回帰係数を組み合わせて用いている．

表2 音声の分析条件

Table 2 Analysis condition for speech.

サンプリング周波数	12 kHz
ハミング窓幅	21.33 ms（256ポイント）
フレーム周期	8 ms（96ポイント）
LPC分析次数	14次

(b) 規則の確率化を行わない場合

前節までいくつかの文法の学習法をテキストレベルでカバー率などで調べたが，その中で基本となるCFGのみの文法以外にカバー率とパープレキシティのバランスが良かった単語クラスペアを用いた以下の学習法について音声認識実験を行った．

方法1 CFGのみ（文法学習は行わない初期文法のみ）
従来のCFGだけを使って単語予測を行う方法．

方法2 CFG+単語クラスペアをCFGの補間に使用
従来のCFGに加えて，単語クラスペアをCFGの部分解析木の接続に使用して単語予測を並列に行う方法．

方法3 単語クラスペアとCFGで同時に予測を行う
従来のCFGに加えて，単語クラスペアによる予測を並列に使用して単語予測を行う方法．

方法4 単語クラスペアのみ
単語クラスペアのみによって単語予測を行う方法．

結果を**表3**に示す．結果は文正解数(A)と助詞誤りを無視した文正解数(B)についてまとめた．文正

表3 文法を学習した場合の正解音声認識文数
Table 3 Number of correctly recognized utterances by learning of the grammar.

話者	学習文数	方法1		方法2		方法3		方法4	
		A	B	A	B	A	B	A	B
KO	406	21	33	31	50	37	58	36	54
AD	406	34	36	45	52	50	58	44	54
text	406	39		72		64		58	
KO	914	21	33	29	52	41	59	42	58
AD	914	34	36	44	50	53	60	53	60
text	914	39		77		74		73	

A：文正解数（テスト文数：106）
B：助詞誤りを無視した文正解数

解数は認識結果と正解文とが完全に一致した文の数をまとめたもので，助詞誤りを無視した文正解数は助詞については考慮せずに正解と一致した文の数についてまとめたものである（助詞誤りを無視した文正解数を示したのは，助詞の音声認識が難しいのと，ほとんどの助詞は言語処理部で復元できるからである）．textの欄はテキスト入力の場合に解析された文数を表している．

方法2（CFG+単語クラスペアをCFGの補間に用いる方法）を除いて学習文数を増やすと認識率が上昇する傾向にあることが分かる．方法2の認識率が上昇しなかった理由はパープレキシティの大きさによると考えられる．実際認識文数はあまり変わらない結果となっている．方法3（CFG+単語クラスペアで同時に予測を行う）と方法4（単語クラスペアのみの予測）はあまり差がなかった．どの結果も元の文法での結果よりも良くなっているものの，単語クラスペアのみの方法でも良い結果が得られることから，学習データが多い場合には一番良い方法であるといえる．また，学習データが少ない場合には単語クラスペアにCFGでの予測を組み合わせた方が良い結果が得られることから，学習データが少ないうちはCFGを組み合わせた方が良く，学習が進むにつれて単語クラスペアのみの解析にした方が良いというテキストレベルの場合と同じ結論が得られた．

(c) 規則の確率化を行った場合

確率を導入しない場合で認識率があまり上昇しなかった単語クラスペアをCFGの補間に使用する方法（方法2）を除いた3つの方法について，確率を導入した場合の認識率について述べる．前述した以下の3つの学習法についての認識実験を行った．

方法1 SCFGのみ

方法3 単語クラスバイグラムとCFGで同時に予測を行う

表4 確率文法による正解音声認識文数
Table 4 Number of correctly recognized utterances using stochastic grammars.

話者	学習文数	方法1		方法3		方法4	
		A	B	A	B	A	B
KO	914	20	32	50	64	50	64
AD	914	33	35	60	70	60	69
text	914	39		74		73	

A：文正解数（テスト文数：106）
B：助詞誤りを無視した文正解数

方法4 単語クラスバイグラムのみ

結果を**表4**に示す．表3と比べ，方法3と方法4は格段に認識率が向上している．確率を使った言語モデルでもバイグラムのみの方がSCFGのみの結果よりも良く，解析できる文のうち約75%の文が正しく認識できた（助詞誤りを無視すれば約90%）．

以上，テキストで調べてきた文法の学習方法と効果を実際の音声認識でどれぐらいの効果があるかを調べた結果，テキストでの結果と同じ傾向が得られた．

6. むすび

音声認識システムでユーザが受理できない文を入力したときに，新しい規則を登録し，入力文のカバー率を改善するシステムの開発を行い，“富士山観光案内システム”のタスクを使うことにより評価を行った．

CFGの規則の登録法はカバー率ではかなり改善されたがパープレキシティがかなり増加した．非終端記号のペアはトップのペアを使えばカバー率はほぼ完全になるがパープレキシティが大きくなりすぎ，ボトムのペアを使えばカバー率があまり良くならなかった．一方，単語クラスペアを使った方法はカバー率とパープレキシティのバランスが上記の方法と比べて一番良いということが分かった．

確率を導入することにより単語ペアのパープレキシティの増加を抑えることも試みた．CFGを使わずに確率だけで解析する場合には非常に効果があったが，カバー率は十分ではなかった．CFGとバイグラムを組み合わせて使用する方法は，確率を学習するためのデータが不十分な場合にも有効である．バイグラムとSCFGとを並列に使用する場合はカバー率，パープレキシティの両面において非常に効果があった．

大規模なデータベースを利用できる場合はトライグラム等の確率モデルの方がCFGのような文法よりもパープレキシティの面で優れているが，前もって大量のデータが得られないようなアプリケーションではCFGベースの文法の方が利用価値がある．すなわち初期文法はCFGを用い，以後学習データの増加とと

もにN-グラムへの移行（2つの解析に対する重み λ を $\lambda=1$ から $\lambda=0$ に逐次変更することに相当）というのがアプリケーションによっては現実的であると思われる．

音声認識実験による評価においては，テキストレベルとほぼ同等のことがいえた．言語モデルの確率を使わない場合には，学習データが少ない間は単語クラスペアだけでなくCFGを組み合わせた文法を使った方が認識率が良く，学習が進むにつれて単語クラスペアのみで単語予測を行った方が良い認識結果が得られた．

また，確率を用いても確率を用いない場合と同じ傾向であったが，テキストレベルでパープレキシティが下がっていることから認識率の面でさらに良い結果が得られた．テキストレベルから音声認識レベルまで一貫した評価を行った結果によって，学習データが少ない場合の文法の構築において本論文で提案したCFG（SCFG）と単語クラスペア（bigram）を組み合わせた文法が有効であることを示せた．

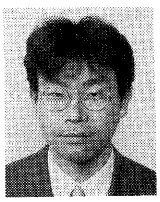
参考文献

1) 山本，伊藤，肥田野，中川：人間の理解手法を用いたロバストな音声対話システム，情報処理学会論文誌，Vol.37, No.4, pp.471–482 (1996).
2) 大谷耕嗣，山本幹雄，中川聖一：例文からの話し言葉用文法の半自動修正法，第50回情報処理学会全国大会論文集，2R-4, Vol.3, pp.59–60 (1995).
3) 大谷耕嗣，中川聖一：単語対制約の追加による話し言葉用文法の自動修正法，第51回情報処理学会全国大会論文集，4H-7，Vol.3, pp.67–68 (1995).
4) 大谷耕嗣，中川聖一：CFGとバイグラムの結合による文法の半自動修正法，情報処理学会音声言語情報処理研究会，95-SLP-9-15, pp.99–104 (1995).
5) 周　旻，中川聖一：日本語及び英語の言語モデルに関する検討，自然言語処理における学習シンポジウム，電子情報通信学会，pp.57–64 (1994).
6) 中澤　聡，濱田　喬：適応パーザによる文脈自由文法の自動修正，第53回情報処理学会全国大会論文集，4M-7, Vol.2 (1996).
7) 白井清昭，徳永健伸，田中穂積：コーパスからの文法の自動抽出，情報処理学会自然言語処理研究会，94-NL-101 Vol.101, pp.81–88 (1994).
8) Brill, E.: Automatic Grammar Induction and Parsing Free Text: A Transformation-Based Approach, *Proc. ACL93*, pp.259–265 (1993).
9) Miller, S. and Fox, H.J.: Automatic Grammar Acquisition, *Proc. Human Language Technology*, pp.268–271 (1994).
10) Samuelsson, C.: Grammar Specialization Through Entropy Thresholds, *Proc. ACL94*, pp.188–195 (1994).
11) 藤崎博也，阿部賢司，横田和章：シミュレーテッド・アニーリング法による状態遷移図の自動獲得，第53回情報処理学会全国大会論文集，7L-6, Vol.2 (1996).
12) Jones, G.J.F., Wright, J.H. and Wrigley, E.N.: The HMM Interface with Hybrid Grammar-Bigram Language Models for Speech Recognition, *Proc. ICSLP-92*, pp.253–256 (1992).
13) 竹澤寿幸，森元　逞：部分木を単位とする構文規則と前終端記号のバイグラムを利用した連続音声認識，情報処理学会音声言語情報処理研究会，95-SLP-9-9, pp.55–62 (1995).
14) *Proceeding of Spoken Language System Technology Workshop*, Morgan Kaufmann (1995).
15) 中川聖一：確率モデルによる音声認識，電子情報通信学会，SP96-101 (1988).
16) 伊田政樹，中川聖一：パープレキシティと音声認識率との関係，日本音響学会論文集，1-3-22 (1996).
17) 中川聖一，大谷耕嗣：CFGとバイグラムの使用による話し言葉用文法の自動学習と音声認識による評価実験，情報処理学会音声言語情報処理研究会，97-SLP-16-3 (1997).
18) 中川聖一，甲斐充彦：文脈自由文法によるOne Pass型HMM連続音声認識法，電子情報通信学会論文誌，Vol.76-D-II, No.7, pp.1337–1345 (1993).

(平成8年12月13日受付)
(平成10年1月16日採録)

中川　聖一（正会員）

昭和51年京都大学大学院博士課程修了．同年同大学情報助手．昭和55年豊橋技術科学大学情報工学系講師．昭和58年同助教授．平成2年同教授．音声処理，自然言語処理，人工知能の研究に従事．昭和52年電子情報通信学会論文賞，1988年度IETE最優秀論文賞．著書：「音声・聴覚と神経回路網モデル」，「確率モデルによる音声認識」，「情報理論の基礎と応用」等．



大谷　耕嗣

平成7年豊橋技術科学大学情報工学課程卒業．平成9年同大学大学院修士課程修了．在学中は自然言語処理の研究に従事．現在，(株)トヨタテクノサービス勤務．