音声認識・理解における
統計的言語処理の研究

１９９９年 １月

政瀧 浩和

目次

第１章 序論

1.1 研究の背景と目的

計算機の性能向上・価格の低下・周辺技術の進歩等に伴い，企業内のOA化の普及，インターネットの流行等，一般の人間が計算機を使用する機会は，近年急速に増加している．このため，計算機の操作性を向上させるマン　マシンインターフェースの重要性はますます高まっている．人間同士でコミュニケーションを取るとき，言葉，特に話し言葉は最もよく用いられる手段であり，手を使わなくても良いことや，多少の距離があっても伝達可能であること等，コミュニケーションのための非常に便利な手段である．従って，マン　マシンインターフェースの向上のためには，人間の発話を計算機が認識・理解し，発話内容に応じて動作を行う「音声認識」,「言語理解」が有効であるという考えは，ごく自然であると考えられる．このため，音声認識および言語理解の技術はかなり古くから盛んに研究されてきた．

近年の音声認識技術の進歩には目覚ましいものがある．これは，米国のARPA(Advanced Research Projects Agency)が主催する，Wall Street Journalおよび北米の種々の経済新聞(NAB: North American Business News)記事の読み上げ音声を認識するプロジェクトが強力に進められ，さまざまな研究機関が性能向上に向けて技術開発を競い合ったことが大きく貢献している．最近では，連続音声認識の性能が65,000語彙で単語認識率93%程度とかなり高い性能が報告されている[1][2][3]．また，CPUの高速化，ディスク・メモリの大容量化等，計算機性能の著しい向上により，PC（パーソナルコンピュータ）でも実時間に近い処理が可能となり，ここ数年間で音声認識技術は，かなり実用レベルに近づいた感がある．最近のARPAプロジェクトでは，放送ニュースの認識や自然発話の認識等，さらに困難な課題に研究の対象が移行する一方，

ATIS(Air Travel Information System)システム[4]と呼ばれる，音声による航空路線の案内システムのような，音声理解システムも盛んに研究されている．

一方，日本においても，ディクテーションシステムを前提とした日本語大語彙音声認識は高い性能が報告されており[5][6]，自然発話音声認識の研究[7]や，音声認識結果を他国語に翻訳して異国語間の会話を可能とする音声翻訳システム[8]のような音声理解システムの研究も盛んに行われている．

音声認識技術において中心的役割を果たすのは，波形の特徴量から音素の識別を行うための音響モデルである．音響モデルには，DP マッチングを用いる手法が盛んに研究されていたが，1980 年代に隠れマルコフモデルを用いた手法が提案されて以来，音声の認識精度は大幅に向上した．しかし，現在の技術レベルでは，連続音声認識における音響モデルの性能は，音素認識率で 60〜70 パーセント程度であり，現状では音響モデルだけで実用的な認識率を得ることはできない．このため，認識の単位を単語とし，単語間の接続を制約する「言語モデル」とを組み合わせることにより，連続音声認識の性能を飛躍的に向上させている[9]．

従来，連続音声認識の言語モデルとしては，文脈自由文法等の文法的ルールを用いる手法が研究されていたが[10]，上述の ARPA プロジェクトで有効性が認められた言語モデルは，N-gram と呼ばれるモデルである．N-gram 言語モデルは，直前の(N-1)単語から次の単語を確率(遷移確率)で予測するモデルであり，学習用に与えられたテキストデータから遷移確率を推定することにより，自動的に構築される「統計的モデル」である．N-gram は極めて単純な言語モデルでありながら，構築の容易さ，統計的音響モデルとの相性の良さ，認識率向上や計算時間短縮の効果が大きい等の理由で，連続音声認識には非常に有効であることが確認されている．最近では殆んどの連続音声認識システムにおいて，統計的言語モデル N-gram が使用されていると言っても過言ではない．

このように，連続音声認識の言語制約として，その有効性が広く確認されている N-gram 言語モデルであるが，実用的な連続音声認識システムを構築する際には，次のような問題を解決しなければならない．

問題 1) 日本語の N-gram 学習用データを収集するのは困難

英語等の文章は，単語がスペースで区切られているため，テキストデータから単語を単位とした N-gram を直接構築することが可能であるが，日本語の文章は単語の区切りがなく，テキストデータから直接 N-gram を構築することはできない．通常，日本語の連続音声認識では，文章を品詞の単位で分割した「形態素」を単位とした形態素 N-gram が使用される．しかし，形態素 N-gram を構築するためには，テキストデータをあらかじめ形態素解析しておく必要がある．精度の高い形態素 N-gram を構築するためには形態素が付与された大量のテキストデータが必要であるが，正確な形態素解析には手作業が必要なため，日本語の N-gram 学習用のデータを大量に収集するのは容易ではなく，連続音声認識システム構築の際のネックとなっている．

問題 2) 少量の学習データから精度の良いモデルを得るのは困難

N-gram はパラメータ数が多いため，学習には通常，新聞記事等の大量のテキストデータが用いられる．しかし，より自然なインターフェースを実現するためには，話し言葉のような自然発話の認識が必要である．しかし，自然発話のデータを大量に収集することは容易ではない．特に，日本語の場合は，問題 1)のように，テキストに対して形態素解析を行う必要があり，さらにデータの収集が困難になる．学習データ量が少ない場合，N-gram の精度は低下し，システム構築上の大きな問題となる．

また，音声認識では統計的言語モデル N-gram が使用される一方，音声認識のアプリケーションとして研究されている音声理解システムにおいては，言語理解の手法として，文法的ルールを用いた構文解析等の手法が盛んに用いられている．このため，音声理解システムにおいては，次の問題が発生する．

問題 3) 音声認識と言語理解との相性が良くない

音声認識と言語理解とを組み合わせた音声理解技術において，連続音声認識では，言語モデルとして統計的モデル N-gram が盛んに使用されているが，N-gram は局所的な制約しか持たないため，認識誤りが生じると非文法的な認識結果が得られる場合がある．この場合，構文解析等による従来の言語理解手法では解析不能に陥り，言語理解ができなくなることが予想される．

本論文は，これらの問題を解決し，連続音声認識および音声理解システムの構築を容易にし，かつ性能を向上させるための研究を行った成果をまとめたものである．

1.2 本論文の構成

本論文の構成は以下の通りである．

第2章では，本論文の理解に必要な基本技術の概説を行う．まず，確率統計的手法に基づく連続音声認識・音声理解の概念を示し，言語現象をモデル化するために用いられるN-gramモデル，および隠れマルコフモデルについてその動作原理，理論等について解説する．

第3章では，まず，日本語連続音声認識および理解における単位について考察し，形態素が有効であることを示す．形態素を単位としたN-gramを構築する場合，学習データとして，形態素が付与されたテキストデータが必要となる．このため，文章を自動的に形態素に分割する，形態素解析の技術が重要である．本論文では，形態素解析のための言語モデルとして，『品詞と可変長形態素列の複合N-gram』を提案する．品詞と可変長形態素列の複合N-gramは，基本的には品詞を単位としたN-gramであるが，特定の形態素は品詞から分離させ，さらに特定の形態素列を結合させて新たな単位とするN-gramモデルである．品詞と可変長形態素列の複合N-gramはパラメータ数が少なく，少量のデータからでも比較的精度の高い言語モデルを得ることができる．

第4章では，音声認識の精度を向上させるため，第3章の形態素解析でも使用した『品詞と可変長形態素列の複合N-gram』による連続音声認識手法を提案する．複合N-gramは，言語モデルとしての精度が高いことに加え，頻繁に出現する形態素列を結合させて長い単位として扱うことにより，音声認識の探索空間の削減が可能であるため，連続音声認識の精度を向上させることができる．

第5章では，目的のタスクのデータがごく少量しか集まらない場合でも精度の高い言語モデルを得る手法として，タスク適応の手法を検討する．タスク適応は，他のタスクのデータで構築したN-gramを目的のタスクの少量のデータに適応させることにより，目的のタスクにおける言語モデルの精度を向上させる手法である．本論文では，従来のタスク適応手法である，独立N-gramの線形結合に代わり，理論的に整備されている確率推定方法である，最大事後確率(MAP)推定を用いた手法を提案し，その有効性を示す．

第6章では，音声認識と言語理解との相性向上を目的として，統計的モデルである隠れマルコフモデルを用いた言語理解手法を提案する．言語理解に統計的手法を用いることにより，音声認識の誤りによる非文法的な文の理解が可能になるほか，学習により理解モデルが自動的に構築できるため，話し言葉の理解システムも容易に構築できるという長所がある．

最後の第7章は結論であり，本研究の成果を要約し，今後の展望・課題について述べる．

第２章 統計的言語モデル概説

2.1 まえがき

近年の連続音声認識技術は，音響モデルとして隠れマルコフモデル，言語モデルとしてN-gramモデルという確率・統計的モデルが盛んに用いられ，大きな成果を上げている．本論文ではこれら２種類の統計的モデルを対象として，精度向上や言語処理への応用に関する研究を行っている．このため本章では，本論文の理解の補助のため，これらのモデルの動作原理やその問題点等について解説する．

まず2.2節では，確率統計的処理による連続音声認識・理解の基本原理を示す．続く2.3節では，連続音声認識の言語モデルとして近年盛んに使用され，また本論文の主な研究対象ともなっているN-gram言語モデルについて概説する．また，N-gram言語モデルの問題点を示し，それらを解決するために提案されている従来手法について解説する．最後の2.4節では，隠れマルコフモデルについて概説を行う．隠れマルコフモデルは，連続音声認識の音響モデルとして盛んに用いられているモデルであるが，本論文第６章ではこれを言語理解のためのモデルとして使用している．

2.2 統計的処理による連続音声認識・理解の原理

連続音声認識技術は，音声信号 $X(x_1, x_2,\ldots,x_T)$ から実際の発話内容である単語列 $W(w_1,w_2,\ldots,w_N)$ を識別する技術である．確率統計的手法による連続音声認識は，入力波形 X が与えられた時に確率的に最も高い単語列 W を識別する手法であるが，これは次式のように表される．

$$\hat{W} = \arg\max_{W} P(W|X) \qquad (2.1)$$

ベイズの定理を用いると，$P(W/X)$ は

$$P(W|X) = \frac{P(X|W)P(W)}{P(X)} \qquad (2.2)$$

と変形でき，式(2.1)は次のように書き換えることができる．

$$\hat{W} = \arg\max_{W} \frac{P(X|W)P(W)}{P(X)} \qquad (2.3)$$

右辺分母 $P(X)$ は音声波形の特徴量 X が生起する確率を意味するが，この確率は右辺の最大化には無関係な量であるから，上式は結局次式と等価になる．

$$W = \arg\max_{W} P(X|W)P(W) \qquad (2.4)$$

本式が確率統計的処理による連続音声認識の基本となる式である．本式は『 $P(X/W)$ 』，『 $P(W)$ 』，『 $arg\,max_W$ 』の三要素から成り立っているが，それぞれの意味を以下に示す．

・$P(X/W)$：音響モデルによる評価

単語列 W から入力波形の特徴パラメータ X を出力する確率で，この確率を与えるモデルは音響モデルと呼ばれている．近年では音響モデルとして，隠れマルコフモデルが盛んに用いられている．

・$P(W)$：言語モデルによる評価

単語列 W が出現する確率で，この確率を与えるモデルは言語モデルと呼ばれている．近年最も盛んに用いられる言語モデルはN-gram言語モデルである．

・$arg\,max_W$：サーチ（探索過程）

考えられる全ての単語の組み合わせから，音響モデルおよび言語モデルで与えられる確率が最大になる単語列を探索する．代表的な手法にはビームサーチ，単語グラフサーチ[7]，A*サーチ[11]等がある．

実際の連続音声認識では，式(2.4)の確率の対数を用い，

$$W = \arg\max_{W} \{\log P(X|W) + \log P(W)\} \qquad (2.5)$$

として評価を行う．これは，確率の乗算によるアンダーフローを防ぐためである．また，通常，音響モデルと言語モデルの確率のダイナミックレンジの差を調整するために，言語モデルの確率 $P(W)$に言語重みなる定数 λを乗じて確率計算を行う．

$$W = \arg\max_{W} \{\log P(X|W) + \lambda \log P(W)\} \qquad (2.6)$$

次に，統計的処理による言語理解手法について述べる．本研究では，言語理解は，文章 $W(w_1,w_2,\ldots,w_N)$ から，その文章の概念を表現するシンボル $S(s_1,s_2,\ldots,s_T)$ を獲得する過程であると定式化する．確率統計的手法では Wから最も確率の高いシンボル列 S を求めることにより実現されるが，これは以下のように表すことができる．

$$S = \arg\max_{W} P(S|W) \qquad (2.7)$$

ベイズの定理を用い，先述の連続音声認識の導出と同様の変形を行えば，上式は，

$$S = \arg\max_{W} P(W|S)\,P(S) \qquad (2.8)$$

と変形できる．本論文では，第 6 章において，$P(W/S)$ をモデル化するために隠れマルコフモデルを，$P(S)$をモデル化するために N-gram(Bigram)を使用する方法を提案している．

2.3 N-gram モデル概説

2.3.1 N-gram モデルの動作原理

統計的言語モデルは，単語列 W の生成確率を与えるモデルである．単語列 W は L 単語からなるとすると，生成確率は次式のように表される．

$$P(W) = P(w_1,w_2,\ldots,w_L) \qquad (2.9)$$

この確率をそのまま求めるのは非常に困難であるため，本式を次のように変形してみる．

$$P(W) = \prod_{t=1}^{L} P(w_t|w_1^{t-1}) \qquad (2.10)$$

$$\left(= P(w_1)P(w_2|w_1)P(w_3|w_1w_2)\cdots P(w_L|w_1,w_2,\ldots,w_L) \right)$$

但し，　w_x^y は単語列 W の x 番目から y 番目の単語列を意味する．

$P(w_1)$ は単語 w_1 が文の最初に出現する確率である．また，　$P(w_2/w_1)$ は，単語 w_1 の次に単語 w_2 が出現する確率である．これらの確率は適度な量のテキストデータから比較的容易に求めることができると考えられる．しかし，$P(w_L/w_1,w_2,\cdots,w_{L-1})$ は，L-1 単語列の次に単語 w_L が出現する確率で，L が大きいと L-1 単語の組み合わせは膨大な数になり，この確率を求めるのは不可能に近い．

そこで，t 番目の単語 w_t は直前の N-1 単語列のみに依存すると考えると，単語列 W の生成確率は次のように近似できる．

$$P(W) \approx \prod_{t=1}^{L} P(w_t|w_{t-N+1}^{t-1}) \qquad (2.11)$$

N が比較的小さければ，N-1 の単語列の組み合わせはそれほど大きくなく，大量のテキストデータがあれば，求めるのは不可能ではない．本式で表されるモデル，すなわち直前の N-1 単語から次の単語への遷移を確率として（遷移確率）与えるモデルが N-gram 言語モデルである．

通常，直前の 1 単語から次の単語への遷移確率を与える Bigram(2-gram)，あるいは直前の 2 単語から次の単語への遷移確率を与える Trigram(3-gram)がよく用いられる．Bigram および Trigram は次のように表される．

・Bigram:　　$P(W)=\prod_{t=1}^{L}P(w_t|w_{t-1})$　　(2.12)

・Trigram:　　$P(W)\cong\prod_{t=1}^{L}P(w_t|w_{t-2},w_{t-1})$　　(2.13)

但し，w_i,w_{ij} は単語 w_i と w_j とが連続して出現する状態を表す．

これらの遷移確率は大量の言語データから推定される．通常，遷移確率の推定には最尤推定が用いられる．以下に，最尤推定による単語 Bigram と Trigram の遷移確率の計算方法を示す．

・Bigram:　　$P(w_t|w_{t-1})=\frac{n(w_{t-1},w_t)}{n(w_{t-1})}$　　(2.14)

・Trigram:　　$P(w_t|w_{t-2},w_{t-1})=\frac{n(w_{t-2},w_{t-1},w_t)}{n(w_{t-2},w_{w-1})}$　　(2.15)

但し，$n(\#)$は単語または単語列 '#' のデータ中の出現頻度を表す．

このように，単語 N-gram の遷移確率は，単語および単語列の出現頻度から極めて容易に計算できる．

なお，N-gram 言語モデルの精度の優劣は通常パープレキシティという値により評価される．パープレキシティの算出方法等に関しては、巻末の付録に記載している．

2.3.2 N-gram の問題点

単語 N-gram のパラメータ数，すなわち単語遷移の組み合わせは V^N（V は語数）であり，N を大きくするとパラメータ数が爆発的に増加するため，それぞれの値の推定が困難になるという大きな問題が存在する．例えば，語彙が 10,000 語の時，Trigram のパラメータ数は $10{,}000^3=10^{12}$ (=1 兆) となり，それぞれのパラメータを推定するためには，数兆語からなる膨大な量のテキストデータが必要となるが，これほどの大規模のデータを収集することは事実上不可能に近い．

このため，補間（平滑化）と呼ばれる，学習テキスト上に出現しない単語遷移に対しても 0 でない確率を与える手法が提案されている．代表的な例は削除補間[12]と back-off smoothing[13]であるが，いずれも N-gram の遷移確率を低次の N-gram で補う方法である．以下，それぞれの方法を簡単に示す．

・削除補間法（線形補間法）[12]：

最尤推定で求められた 1-gram,2-gram,...,N-gram 遷移確率の線形和を N-gram の遷移確率とする方法．

$$\overline{P}(w_t|w_{t-N+1}^{t-1})=\sum_{i=1}^{N}\lambda_i P(w|_{t-i+1}^{t-1})(=\lambda_1 P(w_t)+\lambda_2 P(w_t|w_{t-1})+\lambda_3 P(w_t|w_{t-2},w_{t-1})\ldots)$$

(2.16)

$$(\text{但し，}\ \sum_{i=1}^{N}\lambda_i=1)$$

重み λ_i は，学習データの一部を削除しながら残りのデータで N-gram 確率の推定を行い，削除する部分を変化させながら，学習データ全体にわたって推定を行うことによって適切な値が求められる．

・back-off smoothing[13]：

求める遷移確率の N 単語列がデータ上に存在する場合，最尤推定により求まる確率を減じ(Discount)，存在しない場合は(N-1)-gram の遷移確率に係数を乗じた値を用いる．

$$\overline{P}=\begin{cases}\alpha P(w_t|w_{t-N+1}^{t-1}) & (n(w_{t-N+1}^{t})>0\text{の場合})\\ \beta P(w_t|w_{t-N+1}^{t-1}) & (n(w_{t-N+1}^{t})=0\text{の場合})\end{cases}\qquad(2.17)$$

但し，α は Discount 係数と呼ばれる定数で Turing 推定[14]という方法により求められる．また β は正規化のための係数である． back-off smoothing は，近年盛んに使用されている平滑化の手法法であり，さらに精度を高めるためにさまざまな手法が提案されている[15][16][17].

また，N-gram のパラメータ数を削減し，少量のデータからのパラメータ推定の信頼性を増すため，次の 2 種類のモデルが提案されている．

・クラス N-gram[18][19][20]：

複数の単語をまとめてクラスとして扱い，クラス間の遷移を考えることによりパラメータ数を削減し，推定量の信頼性を高めるものである．一般に下式で表される．

$$P(w_t|w_{t-N+1}^{t-1}) \cong P(w_t|c_t)P(c_t|c_{t-N+1}^{t-1}) \qquad (2.18)$$

（c_t は単語 w_t の属するクラスを意味する）

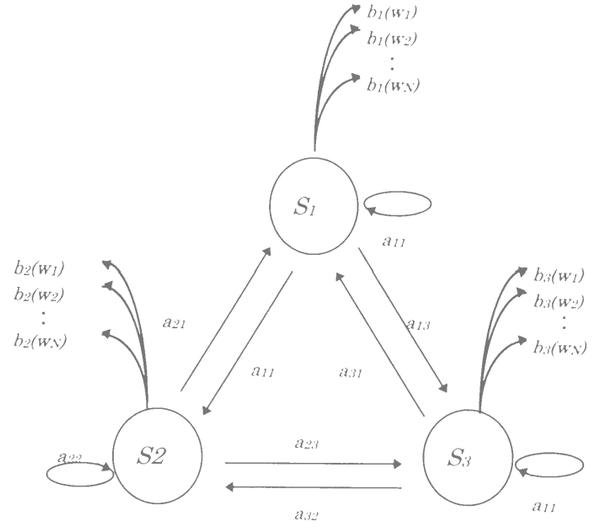
$P(c_t|c_{t-N+1}^{t-1})$ は直前の N-1 単語列に対応するクラス列から次の単語の属するクラスへの遷移確率を表し，$P(w_t|c_t)$ は，次クラスから次単語が出現する確率を意味する．クラス数が 100 の時，Trigram の全てのクラス間の遷移の組み合わせは $100^3=10^6$ (=100 万)であるから，単語 N-gram に比べてパラメータ数は極めて少なく，比較的信頼できる遷移確率を求めることができる．

・可変長単語列 N-gram[21][22][23]：

単語列を結合したものを N-gram の単位として扱うもので，固定長の N-gram と比較して，局所的に N を大きくさせる効果があり，パラメータ数の増大を抑えながら，より長い単語間の関係を表現するものである．

2.4 隠れマルコフモデル概説

N-gram モデルは出力シンボルを単語とした場合のマルコフモデルである．マルコフモデルは，出力シンボルの系列によって一意に状態遷移が決まる決定性確率有限状態オートマトンに対応している．これに対して，隠れマルコフモデル（HMM: Hidden Markov Model）は，出力シンボルの系列にからは一意に状態遷移が決まらない非決定性確率有限状態オートマトンに対応している．図(2.1)に隠れマルコフモデル(3 状態)の概念図を示す．



図(2.1) 隠れマルコフモデルの概念図

なお，音声認識では，出力シンボルが波形の特徴量という連続量に対応するため，離散的シンボルではなく正規分布等の連続出力分布とする場合が多い．しかし，HMM を言語処理に用いる場合は，単語を出力シンボルとした離散分布による出力確率を用いる．

隠れマルコフモデルを言語モデルとして使用する場合の動作は以下のように表現できる．

1. 初期状態ではいずれかの状態にある．状態 s_i にある確率を π_i とする．
2. 単語が入力されると，状態 s_i から状態 s_j へ確率 a_{ij} で遷移する．
3. 遷移先の状態 s_j から単語 w_k を確率 $b_j(w_k)$ で出力する．

隠れマルコフモデルからシンボル列が出力される確率を求めるアルゴリズムは，前向きアルゴリズム(forward algorithm)と呼ばれている．前向きアルゴリズムでは，考えうる全てのパス上の確率の和を計算するが，これに対し最大の確率を与えるパスのみの確率値を求める方法としてビタビアルゴリズム

(Viterbi algorithm)も使用される．音声認識に隠れマルコフモデルを用いる場合，前向きアルゴリズム，ビタビアルゴリズムのどちらを用いた場合でも認識性能に大差はないとされており，計算量の少ないビタビアルゴリズムがよく用いられる．また，シンボルを出力する毎に乗算を繰り返すと確率が極めて小さな値になるため，通常，確率の対数を取り，対数確率の和として計算する．

マルコフモデル(N-gram)のパラメータ推定には通常最尤推定が用いられるが，隠れマルコフモデルの場合，状態遷移系列が非観測であるため，直接最尤推定することができない．このため，隠れマルコフモデルでは EM アルゴリズム（隠れマルコフモデルでは，特に Baum-Welch アルゴリズム[24]と呼ばれる）を用いた繰り返しアルゴリズムによりパラメータを推定する．

HMM は，ある状態に留まっている間は，シンボルの出力確率は同一であり，状態が遷移するとシンボルの出力確率が変化するという特徴を持つ．音声認識の基本単位となる音素は，音素の先頭部分の特徴と，音素の終了部分の特徴は異なり非定常的な信号であるが，短い区間では定常的な信号である．HMM は，上記のような特徴により，音素のように，全体としては非定常的な信号ではあるが，局所的には定常的な信号という特徴を持つ信号列をモデル化するのに適したモデルである．

言語は，同じ意味を表す表現でも，倒置等の現象により，文全体を考えれば，語順が大きく入れ替わったりするが，局所的には比較的似た構造をしていると考えられ，隠れマルコフモデルにより効果的なモデル化が可能であると考えられる．

第３章 連続音声認識のための高精度な N-gram による形態素解析

3.1 まえがき

本章では，日本語の連続音声認識・理解の統計モデルの学習に必要な言語コーパスを，効率的に整備することを目的として，統計的言語モデル『品詞と可変長形態素列の複合 N-gram』による日本語形態素解析を提案する．

N-gram が最初に連続音声認識の言語モデルとして適用されたのは英語であるが，英語の文章は単語がスペースで区切られているため，テキストデータがあれば単語を単位とした N-gram が容易に構築でき，連続音声認識の構築も容易である．しかし，日本語の文章は文字が連続しており，単語の境界が明らかではない．このため，日本語の N-gram を構築する場合は，英語とは異なり，テキスト文のデータを収集しても，単語の切り離しを行わなければ，単語 N-gram を構築することはできない．しかし，数万，数十万文程度のまとまった量の日本語文章に対し，単語の切り出しを全て人手で行うのは，かなりの作業量が必要であり，連続音声認識用言語モデルの学習に必要な言語コーパスを収集するのは用意ではない．

この問題を解決するため，本章では，まず 3.2 節に連続音声認識に適した単位について考察を行い，形態素が有効であることを述べる．3.3 節では，統計的モデル N-gram による形態素解析手法を示し，品詞 N-gram，および形態素 N-gram よりも精度を向上させる『品詞と可変長形態素列の複合 N-gram』(以化，複合 N-gram と略）を提案する．複合 N-gram は，基本的には品詞を単位とした N-gram であるが，言語モデルとしての精度を高めるため，特定の形態素は品詞クラスから分離させ独立して扱い，さらに特定の形態素列を結合さ

せて新たな単位として扱うモデルである．このため，品詞という単位では表現できない形態素独自の特徴を表現でき，かつ長い範囲の形態素間の連接関係を効率的に表現することができるモデルである．また，品詞から未知語が出現する確率を考えることにより，未知語の形態素解析を行う手法も提案する．3.4節では，実験により，形態素解析における複合N-gramの有効性を示すと共に，言語コーパス整備方法に関して考察を行う．

3.2 連続音声認識・理解のための単位

英語等の言語では，単語を単位とした連続音声認識システムが構成されている．しかし，日本語の文章は連続した文字列から構成されているため，単語の定義が明らかでなく，音声認識のための適切な単位が明確ではない．連続音声認識のための適切な単位として，以下の条件を満たすが重要であると考えられる．

・認識単位から読みへの対応が明確であること

日本語の場合，漢字の読みや，助詞の「は」「へ」等，その文字だけでは読みが決定できない場合が多い．できる限りバリエーションの少ない読みが決定できる認識単位を選択することが，連続音声認識の探索空間の削減に有効である．

・ポーズ（無音区間）の挿入位置が限定できること

連続音声といっても，発声文中にはポーズが挿入される．連続音声認識では，認識単位間にはポーズの挿入が可能であるとして認識を行うため，連続音声認識の探索空間の削減および認識率の向上のためには，認識単位をできるだけ長くしてポーズが挿入される個所を少なくし，かつ認識単位中にはポーズは挿入されないように決定するのが好ましい．

・言語理解システムとの整合性が良いこと

音声認識のアプリケーションとして，音声認識と機械翻訳等の言語理解とを結合した音声理解システムが考えられる．言語理解には形態素解析・構文解析

等の言語解析が必要である．このため，認識誤りにより言語解析が不能に陥ることが少ない単位が望まれる．

・単位の長さが適切であること

N-gramは前の単位列から次の単位を出力する確率を与えるモデルであるから，単位の長さが長い方が，より長い範囲の接続関係を表現することができ，言語モデルとしての性能は向上すると考えられる．しかし，単位を長くし過ぎると，認識単位の種類が増加し，パラメータ数が大幅に増加するため，限られた量のデータからでは正確なパラメータ推定が困難になる．このため，単位が長すぎると，逆に言語モデルの精度は低下する可能性がある．このため，適度な長さがあり，種類数が少ない認識単位が最も好ましい．

日本語の音声認識の単位としては，音素や音節を単位とする手法[25]が提案されているが，単位の長さが短いため言語制約としての性能は低く，連続音声認識の単位しては不向きであると考えられる．また，文字あるいは文字列を単位とした手法が提案されているが[23][26][27]，助詞の「は」「へ」や，漢字の音読み・訓読み等の問題より，単独の文字・文字列から，それらの読みを正確に推測するのは困難である．さらに，文字列を単位とした場合は，ポーズが単位の中に挿入される可能性もある等，連続音声認識の単位としては好ましくない．

先述の条件をよく満たす単位として，形態素が挙げられる．形態素を認識の単位とした場合，事前に形態素に読みを与えておく必要はあるが，各文字毎に読みを与える方法よりも正確な読みの付与が可能である．また，通常ポーズは形態素の前後に挿入されると考えられ，各音素・音節や文字の前後にポーズが挿入される可能性があるとする方法よりも，音声認識の探索空間を狭めることができ，認識率も向上すると考えられる．さらに，音声理解システムを考えた場合，音声認識の単位を形態素とすることにより，認識結果に対し言語理解部で改めて形態素解析を行う必要がない．以上より，連続音声認識では，形態素を認識単位とする方が望ましいと考えられる．

しかし，形態素を単位としてN-gramを構築する場合，テキストデータから形態素を切り出す必要がある．形態素解析の作業をすべて人手で行うのは相当な作業量が必要となり，言語コーパスの整備の上で大きな問題となる．このた

め，形態素切り出し作業を省力化するため，自動で形態素解析を行うことが重要であると考えられる．

3.3 統計的言語モデルによる形態素解析

日本語の形態素解析は，文の文字列 L から，それに対応する形態素列 $\hat{W}$ を獲得することである．統計的手法では，L に対して最も高い確率を与える形態素列 $\hat{W}$ を探索することにより形態素解析を実現する．これは，以下の式で与えられる．

$$\hat{W} = \arg\max_{W} P(W|L) \qquad (3.1)$$

ベイズ則により，本式は下式のように変形される．

$$\hat{W} = \arg\max_{W} \frac{P(L|W)P(W)}{P(L)} \qquad (3.2)$$

本式において，$P(L)$ は右辺の最大値を与えるためには無関係な量である．従って，式(3.2)は下式と等価となる．

$$\hat{W} = \arg\max_{W} P(L|W)P(W) \qquad (3.3)$$

右辺の確率 $P(L|W)$ は形態素から文字列を与える確率であるが，これは，形態素列 W の表記と文字列 L が一致する場合は必ず 1 であり，一致しない場合は 0 である．また，確率 $P(W)$ は，形態素列 W の生成確率である．従って，統計的手法による形態素解析は，与えられた文字列と一致する全ての形態素列の中から，生成確率が最も高くなる形態素列を探索することによって実現できる．形態素列 W を $w_1, w_2 \ldots w_m$ とすれば，形態素 N-gram を用いると，その生成確率 P(W) は次のように表される．

$$P(W) = \prod_{t=1}^{m} P(w_t|w_{t-N+1}^{t-1}) \qquad (3.4)$$

w_x^y は x 番目から y 番目までの形態素列を表す．

形態素解析は，文章に品詞を付与することが目的であるから，形態素解析には品詞を単位とした N-gram がよく用いられる[20]．これは，2.3.2 節で述べたクラス N-gram の一種で，品詞をクラスとして用いる手法である．品詞 N-gram は以下のように，品詞間の遷移確率と，品詞から形態素が出現する確率とを用いて，形態素 N-gram を近似するモデルである．

$$P(w_t|w_{t-N+1}^{t-1}) \cong P(w_t|c_t)P(c_t|c_{t-N+1}^{t-1}) \qquad (3.5)$$

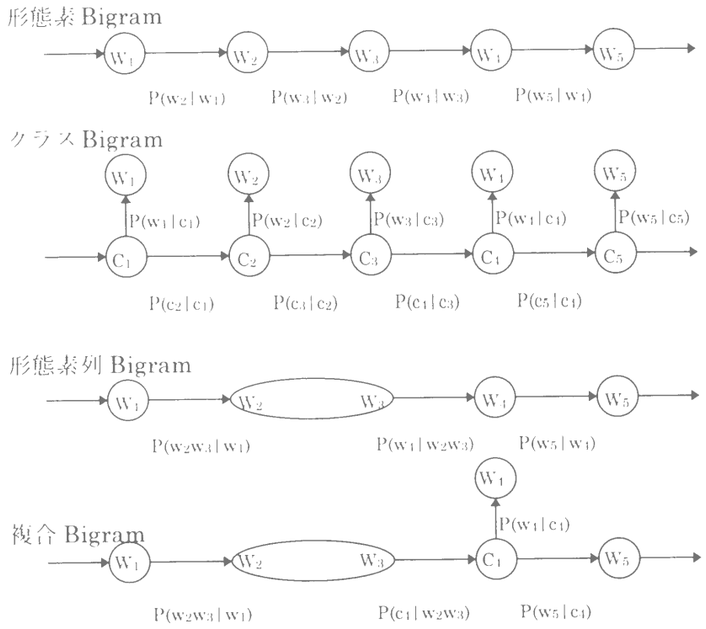
上式で，$P(c_t|c_{t-N+1}^{t-1})$ は品詞間の遷移確率，$P(w_t|c_t)$ は品詞クラス c_t から形態素 w_t が出現する確率である．

3.4 品詞と可変長形態素列の複合 N-gram

3.4.1 品詞と可変長形態素列の複合 N-gram 概説

2.3 節で述べたように，形態素 N-gram はパラメータ数が極めて多いため，少量のデータからは正確なパラメータ推定が困難である．この解決手法として，2.3 節のような遷移確率の平滑化（補間）の方法が提案されている[12][13][15]-[17]．しかし，これらの手法を用いても，少量のデータから精度の高い言語モデルを得ることは困難であると考えられる．また，パラメータ数を削減するために単語（形態素）を品詞等のクラスに分類し，クラス間の遷移確率を考える方法も提案されているが[18]-[20]，モデルの自由度が小さいため，言語モデルの精度は低下すると考えられる．

これらの問題を解決するため，「品詞と可変長形態素列の複合 N-gram モデル」を提案する[29]．本モデルは，出現頻度の低い形態素はまとめて品詞クラスとして扱い，出現頻度の高い形態素は品詞クラスから分離させ独立して扱い，さらに出現頻度の高い形態素列を結合させたものである．従って，品詞クラス N-gram と可変長形態素列 N-gram とのそれぞれの長所を生かしながら，それぞれの短所を補い合うため，出現頻度が低い形態素への信頼性を高めると共に，出現頻度が高い形態素に関する予測精度を高めることができる．



図(3.1) 各種 Bigram の確率計算方法の比較

Bigram を例にして，形態素 N-gram，クラス N-gram，可変長形態素列 N-gram，複合 N-gram との比較を図(3.1)に示す．

複合 N-gram は，品詞クラス，形態素，形態素列を同時に扱うため複雑なモデルとなるが，表現を簡単にするため，複合 N-gram を次の 3 種類のクラス間の N-gram として表現する．

A) 品詞クラス

B) 独立した 1 形態素のみで構成されるクラス

C) 1 連接形態素列のみで構成されるクラス

このクラス分類を用いると，複合 N-gram による文の生成確率は，下式のクラス N-gram の形で与えることができる．

$$P(w_1^L) = \prod_{t=1}^{K} P(ws_t|c_t)\,P(c_t|c_{t-N+1}^{t-1}) \qquad (3.6)$$

但し，ws_t は文章を上記のクラス分類を用いた場合の，t 番目の形態素列（単独の形態素も含める）を意味する．また，K は文章の形態素列の個数を表し，($K \leq L$)である．例として，次の文章（7 形態素）を考える．

「わたくし - 橋本 - と - 言 - い - ま - す」

「橋本」は出現頻度が高くないため，固有名詞クラスとして扱う方が適切であると考えられる．「わたくし」および「と」は日本語の文章で頻繁に出現する形態素であるため，品詞クラスより分離して単独で扱う．また，「言 - い - ま - す」は日本語で頻繁に用いられるフレーズであるため，結合させて一単位として扱う方が効果的であると考えられる．以上のクラス分類を用いた場合，例文の生成確率は，次の式で与えられる．

$$\begin{aligned} &P(w_1^L) = P(\text{わたくし}\,|\,\{\text{わたくし}\})P(\{\text{わたくし}\}) \\ &\cdot P(\text{橋本}\,|\,<\text{固有名詞}>)P(<\text{固有名詞}>|\,\{\text{わたくし}\}) \\ &\cdot P(\text{と}\,|\,\{\text{と}\})P(\{\text{と}\}\,|\,<\text{固有名詞}>) \\ &\cdot P(\text{言います}\,|\,[\text{言います}])P([\text{言います}]\,|\,\{\text{と}\}) \end{aligned} \qquad (3.7)$$

但し，<>，{}，[] はそれぞれ，クラス A) B) C) に属していることを表す．B)および C)のクラスは，形態素（列）とクラスの出現頻度は等しいため($P(w_t)=P(c_t)$)，上式は次のように変形することができ，複合 N-gram と等価であることがわかる．

$$\begin{aligned} &P(W_1^L) = P(\text{わたくし}) \\ &\cdot P(\text{橋本}\,|\,<\text{固有名詞}>)P(<\text{固有名詞}>|\,\text{わたくし}) \\ &\cdot P(\text{と}\,|\,<\text{固有名詞}>) \\ &\cdot P(\text{言います}\,|\,\text{と}) \end{aligned} \qquad (3.8)$$

3.4.2 エントロピー最小化基準による複合 N-gram の自動生成

より少ないパラメータで，次形態素予測精度の高い効率的な複合 N-gram を得るためには，初期クラスから独立させる形態素，および結合させる形態素列を適切に選択する必要がある．このため，品詞クラスを初期クラスとし，初期クラスから形態素独立によるクラス分離，および形態素列結合によるクラス分離という，2 種類のクラス分離を逐次的に行うことによって，複合 N-gram のためのクラス分類を決定する方法を提案する．形態素独立，および形態素列結合候補の決定は，式(3.9)により求められるエントロピーを最小にさせる候補を 1 つのみ選択する．

$$H(c_i) = \sum_i P(c_i) \sum_k P(ws_k|c_j) P(c_j|c_i) \log_2 \{P(ws_k|c_j) P(c_j|c_i)\} \qquad (3.9)$$

但し $ws_k \in c_j$

エントロピーはあいまいさを表す尺度であり，また，エントロピーを H としたときパープレキシティは 2^H で与えられる．すなわち，エントロピーが小さいことはあいまいさが小さく，また，次形態素予測の分岐も少なく，言語モデルの精度が高いことを意味する．従って，クラス分離を行う際に，常にエントロピーを最小にする候補を選択する本手法は，より少ないパラメータで精度の高い複合 N-gram を生成するのに適した手法であると考えられる．なお，本手法において，エントロピーの減少は常に正になることが保証されており，クラス分離によって，学習データに関してエントロピーは単調に減少する．

次に，複合 N-gram の生成アルゴリズムの詳細を示す（図(3.2)，図(3.3)参照）．

1) 初期設定： 全形態素を品詞クラスに分類する

全ての形態素 w_x に対して $w_x \in c_\xi$

2) クラス分離：（a）～(c）の手順でクラス分離を行う

a) 分離クラス候補のリストアップ： i. ii. の 2 種類のクラス分離を考える

i. 形態素の品詞クラスからの分離

品詞クラスに属している任意の形態素に対して，その形態素の独立したクラスと，残りの形態素からなる品詞クラスとに分離する．

$c_\xi \rightarrow w_x \oplus c_\xi \setminus w_x$ （c_ξは形態素 w_x の属するクラス）

（ $a \setminus b$ は a の要素から b を除くこと意味する ）

ii. 連接形態素の結合によるクラス分離

既に初期クラスより分離されている形態素のクラス，および形態素列クラス間の任意の 2 クラスに対して，結合して生じる新たな形態素列クラスと，その形態素列に連接しない形態素クラスとに分離する．

$w_x \oplus w_y \rightarrow \{w_x w_y\} \oplus \{w_x \tilde{w}_y\} \oplus \{\tilde{w}_x w_y\}$

$\{w_x w_y\}$ は連接形態素 $w_x w_y$ のクラス

$\{w_x \tilde{w}_y\}$ は形態素 w_y が後続しない形態素 w_x のクラス

$\{\tilde{w}_x w_y\}$ は形態素 w_x より後続しない形態素 w_y のクラス

b) エントロピー減少量の計算： a) の各分離候補に対して，エントロピー減少量を計算する．

i. 形態素の品詞クラスからの分離

$$\Delta H = H(c_i) - H(c_i \setminus c_\xi + \{w_x\} + c_\xi \setminus w_x)$$

ii. 連接形態素の結合によるクラス分離

$$\Delta H = H(c_i) - H(c_i \setminus (\{w_x\} \oplus \{w_y\}) + \{w_x w_y\} + \{w_x \tilde{w}_y\} + \{\tilde{w}_x w_y\})$$

c) 分離クラスの決定：

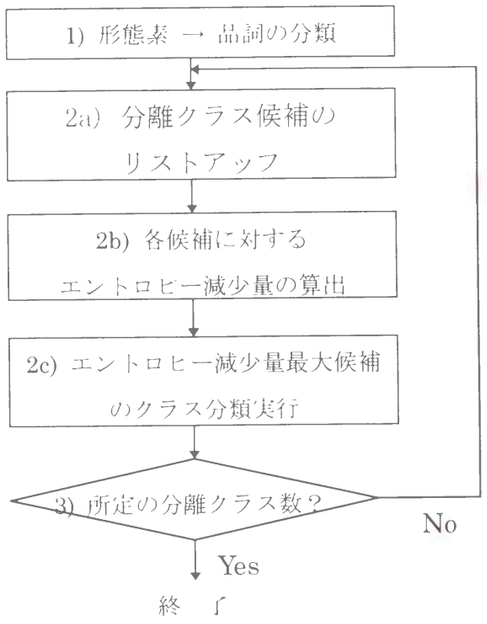
a) の候補内で b) のエントロピー減少量を最大のものを 1 候補のみ選択し，クラス分離を行う．

3) 生成終了

分離クラス数が所定の個数に達したら生成終了．

そうでない場合は，(2) のクラス分離を繰り返す．

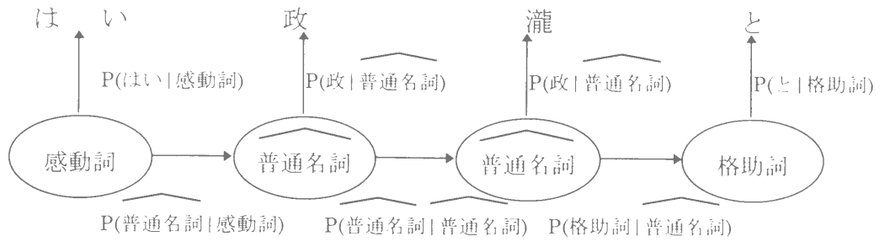
図(3.2) 複合 N-gram の自動生成アルゴリズム



図(3.3) 複合 N-gram の自動生成フローチャート

3.4.3 未知語を含む文の形態素解析

未知語の形態素解析を行うために，品詞クラス c_ξ に対して，同一品詞の未知語のためのクラス $\hat{c}_\xi$ を導入する．クラス $\hat{c}_\xi$ は，任意の文字を 1 文字を出力するクラスであり，同一未知語クラス $\hat{c}_\xi$ が連続した場合は，それらをまとめて一つの未知語とみなす．図(3.4)に，「政瀧」という未知語を含んだ文の品詞 Bigram を使用した形態素解析の処理例を示す．以下に，$\hat{c}_\xi$ に関する確率の導出を行う．



図(3.4) 未知語の形態素解析処理

Turing 推定[14]によると，データ上に r 回出現する形態素は，次式の r^* 回と推定される．

$$r^* = (r+1)\frac{n_{r+1}}{n_r} \qquad (3.10)$$

但し，n_r はデータ上に r 回出現した形態素の種類数を表す．従って，r 回出現する形態素 w の品詞からの出現確率 $P(w|c_\xi)$ は，

$$P(w|c_\xi) = \frac{r^*}{N(c_\xi)} \qquad (3.11)$$

と表される．これを，クラス c_ξ に属する全ての形態素について計算し，1 から引いた残りが品詞 c_ξ から未知語が出現する確率 $P(\hat{c}_\xi)$ である．

$$P(\hat{c}_\xi) = 1 - \sum_{w \in c_\xi} P(w|c_\xi) \qquad (3.12)$$

品詞 c_ξ の未知語の文字 l の出現する確率 $P(l|\hat{c}_\xi)$ は，全ての文字が等しい確率で出現すると仮定し，未知語出現確率 $P(\hat{c}_\xi)$ から均等に割り当てる．

$$P(l|\hat{c}_\xi) = \frac{P(\hat{c}_\xi)}{V} \qquad (3.13)$$

但し，V は文字の種類数とする．また，$P(\hat{c}_\xi|\hat{c}_\xi)$ は，未知語が連続する確率であるが，未知語の長さが二項分布に従うと仮定すると，その品詞に属する語 w の文字列長 $len(w)$ より下のように求められる．

$$P(\hat{c}_\xi \mid \hat{c}_\xi) = \sum_{w \in c_\xi} \frac{len(w)-1}{len(w)} \qquad (3.14)$$

3.5 評価実験および考察

3.5.1 N-gram モデルの形態素解析精度評価実験

自然発話旅行会話データベース[30]を用いて形態素解析の評価実験を行った．本データベースには，間投詞や感動詞のほか，ら抜き表現，助詞落ち等の自然発話特有の言語現象が頻出する．データベースは，1,334 対話，44,091 文，559,711 形態素から成り，語彙は 7,724 語である．このうち，約 4 分の 1(334 対話, 11,321 文，137,691 形態素)を評価用データとし，残り(1, 000 対話，32,770 文，402,020 形態素)を言語モデル学習に使用した．

形態素解析精度の比較対象として，複合 N-gram と形態素 N-gram，および品詞 N-gram を構築した．複合 N-gram は，活用形，および活用型を含めた 234 品詞を初期状態とし，最大 2,000 クラスまで分離を行い，500 分離おきにデータを採取した．また，形態素 N-gram，品詞 N-gram，複合 N-gram ともに，形態素および品詞クラスの遷移確率を back-off Smoothing[13]により学習データに出現しない形態素および品詞クラス遷移に対して 0 でない確率を与えた．また，本節の実験では，辞書には学習データ，評価データに出現する全ての形態素が登録されており，未知語は存在しない．但し，学習データに出現しない形態素に対する遷移確率は，全てのモデルにおいて 1/(100*語数) という確率を与えた．これは，他に候補が無い場合はこの形態素を割り当てるために，0 でない小さい値を与えることを目的としている．

形態素の正解率の評価には，音声認識で広く用いられている単語正解率(Accuracy)にならい形態素 Accuracy を用いた．形態素 Accuracy(%)は下式で表される．

$$100 \times \frac{W - S - D - I}{W} \qquad (3.15)$$

但し，W:正解の形態素数，S:置換誤り形態素数，D:削除誤り形態素数，
I:挿入誤り形態素数を表す．

表(3.1) 各種言語モデルの形態素解析性能比較(品詞のみの評価)

	形態素 N-gram	品詞 N-gram	複合 N-gram(クラス数)			
			500	1000	1500	2000
Bigram	98.90	98.56	99.13	99.07	99.02	99.01
Trigram	98.95	98.94	99.17	99.08	99.01	99.03

表(3.2) 各種言語モデルの形態素解析性能比較(品詞と読みを含めた評価)

	形態素 N-gram	品詞 N-gram	複合 N-gram(クラス数)			
			500	1000	1500	2000
Bigram	98.54	96.78	98.62	98.64	98.62	98.63
Trigram	98.64	97.12	98.68	98.68	98.61	98.66

通常，形態素解析では，形態素の分割が正しく，かつ付与された品詞が正しければ，正解とみなされる．この場合の形態素 Accuracy(%)を表(3.1)に示す．また，形態素解析結果を音声認識に用いることを考えると，同一品詞の形態素でも読みが異なるものは，別単位として扱うことが好ましい．形態素に読みまで考慮した場合の形態素 Accuracy(%)を表(3.2)に示す．但し，読みの推定は，表記が同一の形態素でも読みが異なるものは別の形態素として扱い，異なる単位として N-gram を構築し形態素解析を行うことにより実現している．

表中で下線を付した値が，その次数の複合 N-gram の最高の形態素正解率を示す．表(3.1)および(3.2)より，形態素付与のみの評価および読みまで含めた評価のいずれの場合も，複合 N-gram の最も高い形態素正解率は，同次数の形態素 N-gram および品詞 N-gram よりも高い正解率を得ることができ，複合 N-gram の他のモデルに対する優位性が実験的に示された．形態素正解率最高の値を与える分離クラス数は，品詞のみの評価の場合は分離クラス数 500，読みも含めた評価の場合は分離クラス 1000 であり，それ以上増やしても逆に形態素正解率は低下する傾向にある．これは，クラス数が増加すると共に，パラメータ数も増加するため，各パラメータの確率推定が正しく行われないことに起因すると考えられる．このため，適切なクラス数を決定する必要があるが，これは，ニューラルネットワークの学習回数の決定等で用いられる Cross

Validation の手法を用いることにより，適切なクラス数を実験的に求めることができる．以下に手順を示す．

1. 学習データの一部を仮想的なテストデータとする
2. クラス数を徐々に増加させながら N-gram を学習する
3. 仮想的なテストデータに対し形態素解析を行い，
   形態素解析の性能が頭打ちになる所をクラス数とする

3 種類のモデルを比較すると，品詞 N-gram は読みを含めた評価の場合に他のモデルと比較して形態素正解率が著しく低下している．これは，ある形態素の読みはその前後の形態素の読みに影響されると考えられるが，品詞という枠組みでは，前後の読みの関係が表現できないためと考えられる．形態素 N-gram と複合 N-gram では，読みまで含めた形態素を単位として扱うことができるため，このような大きな低下は見られない．また，複合 N-gram と形態素 N-gram との正解率の差は大きくはないが，3.3.5 節で示した未知語処理の容易さとを考えると，複合 N-gram が有利である．

次に，品詞 Bigram，形態素 Bigram，複合 Bigram の形態素解析結果および遷移確率の比較によりそれぞれの有効性を比較した．

例 1） わたくし 一 人 です ので

品詞 Bigram 「わたくし いちり です ので」

形態素 Bigram 「わたくし ひとり です ので」

複合 Bigram 「わたくし ひとり です ので」

これは，品詞 Bigram において最も誤りの多かった，数詞に対する読みの付与の誤りの例である．「一人」の発音は「ひとり」であるが，品詞 Bigram ではこういった形態素同士の接続ではなく，「数詞」→「普通名詞」という品詞という間での接続しか考慮することができず，出現頻度の最も高い「いち」「り」が選択されている．形態素 Bigram では，形態素間の接続を直接表現することができるため，「ひと」→「り」の遷移確率が「いち」→「り」，「いち」→「にん」等の遷移確率よりも高くなり，「ひとり」という正確な読みの付与が可能である．また，複合 Bigram では，「ひと」「り」がいずれも初期品詞クラスより分離されて，単独の形態素として扱われているため，形態素 Bigram と同様に正しい読みを付与することが可能である．

例 2） はい それ で 結構 です

品詞 Bigram: はい(感動詞) それで(接続詞) 結構(形容名詞) です(判定詞)

形態素 Bigram: はい(感動詞) それ(代名詞) で(格助詞) 結構(形容名詞) です(判定詞)

複合 Bigram: はい(感動詞) それ(代名詞) で(格助詞) 結構(形容名詞) です(判定詞)

この例では，「それで結構」という慣用的なフレーズの表現ができることが重要でなると思われる．品詞 Bigram では，こういった形態素間の特殊な接続関係が表現できず「接続詞」→「形容名詞」の確率と，「代名詞」→「格助詞」→「形式名詞」との確率を比較した結果前者の方が確率が大きくなったため誤りが起きたと考えられる．形態素 Bigram，複合 Bigram では例 1）と同様，こういった特定の形態素間の関係が表現できるため，正しい形態素解析が可能となる．

例 3）（到着は何時ごろに）おなりでしょうか

品詞 Bigram お(接頭辞) なり(本動詞) でしょうか(準体助動詞)

形態素 Bigram お(接頭辞) なり(副助詞) でしょうか(準体助動詞)

複合 Bigram お(接頭辞) なり(本動詞) でしょうか(準体助動詞)

形態素 Bigram では，パラメータ数が(語数)^N という膨大な量になり，求めようとする遷移確率の N 形態素列が学習データに出現しない場合は平滑化によってのみ確率が与えられるため，遷移確率はかなり小さい値となる．このため，「お(接頭辞)」→「なり(副助詞)」という遷移確率が「お(接頭辞)」→「なり(本動詞)」の遷移確率よりも高くなり，このような誤った形態素解析結果となった例である．品詞 Bigram では，接頭辞→本動詞という確率が接頭辞→副助詞という遷移確率よりも高いため，正しい結果が得られている．複合 Bigram は，「なり」という本動詞が初期品詞クラスから分離されなかったため，品詞 Bigram と同様に「お(接頭辞)」→本動詞という遷移確率が高くなり，正しい結果が得られている．

例 4) 同泊 の 方 の お 名前 （を頂けますでしょうか）

品詞 Bigram　：どうはく の ほう の お なまえ

形態素 Bigram：どうはく の ほう の お なまえ

複合 Bigram　：どうはく の かた の お なまえ

方(かた，ほう)，および後(あと，のち)等の読みの付与は非常に困難である[31]．本例において「方」に（かた）という読みの付与を行うためには，言語モデルが「同泊の方」という表現ができることが求められる．複合 Bigram の場合，「の 方(かた)」という 2 形態素が結合され 1 単位としてされたため，Bigram でも「同泊」→「の，方(かた)」という 3 形態素間の表現が可能であった．しかし，形態素 Bigram では，「同泊」→「の」，「の」→「方(かた)」という 2 形態素間ずつの表現できないため，正確な読みの付与ができなかった．

例 5) （お支払いは）カード で さ れ (ますか)

品詞 Bigram　　カード(普通名詞) で(格助詞) さ(本動詞) れ（助動詞）

形態素 Bigram　カード(普通名詞) で(格助詞) さ(補助動詞) れ(助動詞)

複合 Bigram　　カード(普通名詞) で(格助詞) さ(補助動詞) れ(助動詞)

複合 Bigram では「さ(補助動詞)」「さ(本動詞)」共に，初期品詞クラスより分離され，「で(格助詞) さ(本動詞)」という形態素列が学習データに出現しなかったため，形態素 Bigram と同様に形態素解析結果に誤りが生じた．

以上，形態素解析結果の誤り傾向として代表的な 5 例を挙げたが，複合 N-gram は出現回数の低い形態素は品詞として扱うことにより学習のスパース性を解決でき，また出現回数の高い形態素は形態素を単位として扱うことによりモデル化能力が向上できるため，品詞 N-gram および形態素 N-gram よりも高い精度が得られることが分かった．さらに，複合 N-gram は複数の形態素列を単位として扱うこともできるため，局所的に形態素 N-gram よりもモデル化能力に優れるため形態素 N-gram よりも形態素解析の向上が可能であることが分かった．しかし，例 5) のように，形態素が品詞から分離されたために正確な確率推定ができず，誤りを生じる例もあったため，さらに精度を向上させるには，形態素 N-gram の確率を品詞 N-gram で補間する等により，さらに言語モデルのロバスト性を向上させることが重要であると考えられる．

また，形態素解析結果を分析した結果，形態素 N-gram および複合 N-gram の誤りの大半は，「三階(さんがい)」「三階(さんかい)」等複数の読みが可能な場合，および「で(格助詞)」「で(判定詞)」のように形態素解析が非常に困難な場合もしくは正解の揺れ，による誤りが大部分であった．これらの場合，実際には誤りではないため，実際の形態素正解率は表(3.1)および(3.2)の数値よりも，もっと高いと考えられる．

3.5.2 学習データ量と形態素解析精度との関係

前節の実験で，約 40 万語のデータより構築した複合 N-gram モデルは，読みまで考慮した形態素正解率が 98%以上の，高い解析率が得られることが分かった．しかし，40 万語の形態素データを集めることは容易ではなく，連続音声認識に使用する N-gram を学習するための，大量の形態素データを容易に集めるという本研究の目的と矛盾する．従って，データ量が少ない時にどの程度の形態素解析率が得られるかは，本論文の趣旨において重要なことである．

これを調査するため，前節の実験で用いたデータを量を 1/2,1/4 から最小 1/64 とした時の形態素正解率を調べた．言語モデルには，複合 Bigram の分離クラス数 500 と 1000 を用い，形態素正解率は読みも含めた場合の形態素 Accuracy(%)で評価した．実験結果を表(3.3)に示す．

表(3.3)より，データ量が減少するのに比例して，形態素正解率は低下することが分かる．しかし，データ量が全体の 1/64 の場合は，形態素数がわずか 6,306 であるが，このような非常に少ない量の学習データから構築したモデルでも，95%程度の比較的高い正解率が得られる．95%の形態素正解率は自動で形態素解析を行うには高い精度とは言えないが，自動形態素解析の結果を見て，人手で誤り個所を修正するような，半自動の形態素解析としては，使用に耐える性能であると考えられる．

全学習データを使用した場合は，複合 Bigram の分離クラス数 1000 の場合が分離クラス数 500 の場合よりも正解率が高いが，データ量が減少するにつれて，正解率は逆転している．これは，パラメータ数の多い分離クラス 1000 のモデルでは，データ量が少ない場合では，正確なパラメータ値を推定することが困難になることが原因であると考えられる．

表(3.3) 複合 N-gram の学習データ量と形態素解析性能の関係

	データ量（全学習データに対する割合）						
	1/64	1/32	1/16	1/8	1/4	1/2	1/1
形態素数	6,306	14,293	25,931	50,794	101,227	200,105	402,020
クラス数500	94.15	95.87	96.97	97.53	98.02	98.45	98.62
クラス数1000	94.27	95.66	96.85	97.50	97.98	98.41	98.64

以上より，大量の形態素データを得るためには，まず，数千形態素程度のデータを人手で作成し，クラス数の少ない複合 N-gram を構築して半自動の形態素解析を行い，数十万形態素程度のデータが集まった段階で，クラス数の大きい複合 N-gram を構築し，その後は自動で形態素解析を行う，というのが効果的な手段であると考えられる．以下に，この作業に必要なコストについて検討した．

まず，最初の N-gram 学習用の形態素データを作成する必要があるが，これは，1 形態素のデータ作成に 1 分あれば十分であるとして，6,000 形態素のデータ作成にかかる時間は，1 分×6,000＝6000 分=100 時間である．これを基にして作成した複合 N-gram で 95%程度の正解率が得られるため，形態素解析したデータの修正には，1 形態素あたりでは形態素データ作成時の 1/20 の 3 秒程度で可能であると考えられる．40 万形態素のデータを作成するためには，40 万×3 秒＝120 万秒＝2 万分＝約 333 時間となる．一日 8 時間労働としても，2 ヶ月程度で正解率 98%以上の形態素解析システムの構築が可能であることになる．また，修正を行うだけなら比較的単純な作業であり，多数の人間で平行して行うことができるため，さらにシステム構築の期間を短縮することが可能である．

3.5.3 ルールベースの形態素解析との比較実験

形態素解析システム JUMAN[32](Version 3.5)との比較により，従来のルールベースの形態素解析に対する有効性を示す．但し，我々の形態素解析と JUMAN とでは，用いている形態素の体系や辞書に登録されている形態素の語数等が異なるため，できるだけ公平になるよう次のような方法で比較を行った．

表(3.4) ルールベースの形態素解析との比較

	形態素数	品詞付与正解率 (%)	読み付与正解率 (%)
JUMAN	2,158	95.83	94.21
複合 N-gram	2,129	99.91	99.91

・辞書サイズの均等化

辞書サイズが，本論文の実験では約 7 千語であるのに対し JUMAN では約 58 万語あり，さらに，本論文の実験では評価データの全ての形態素が登録されている等，条件は JUMAN が圧倒的に不利である．このため，名詞，動詞，形容詞等の自立語の語彙を我々のシステムと同一にした．但し，「えー」「あのー」等の語は我々のシステムでは間投詞としているが，JUMAN には間投詞という品詞は存在しないため感動詞とした．

・評価方法

我々のシステムと JUMAN とでは形態素の体系が異なり，評価データに対して JUMAN の形態素体系の正解は存在しない．このため，提案手法および JUMAN 共に，形態素解析結果を目視して正誤の判定を行った．但し，形態素の切り分けや品詞の判断は専門家でも困難な部分もあるため，明らかに誤りであると判断できる個所のみを誤りと判断している．また，評価データ約 1 万文を，目視により全て検査するのは時間を要するため，最初の 200 文のみを評価の対象とした．

3.5.1 節の実験で，最も正解率の高かった複合 Trigram（クラス数 1000）と JUMAN に関し，形態素数と形態素解析の品詞付与および読み付与正解率とで比較した結果を表(3.4)に示す．

表(3.4)より，形態素数はほぼ同じであり，両システムの形態素体系は同程度の長さであることがわかる．形態素解析の精度に関しては，品詞付与で約 4%，読み付与で約 5%と本論文の提案手法の方が優れている．JUMAN の誤り個所

を調べると，大部分は感動詞と，数字の読みに関する誤りである．以下に代表例を示す(下線の個所は誤りを意味する)．

例 1)「たぶんえー大丈夫だと思います」
→「たぶん(副詞) え(動詞) ー(記号) 大丈夫だ(形容詞) と(助詞) 思い(動詞) ます(接尾辞)」

例 2)「九月十一日ご一泊」
→「きゅう つき じゅう いち にち ご いち はく」

これらの誤りの大部分は，接続ルールや重みを変更することで対応できると考えられる．しかし，そのためには，相当数のルールの追加・変更が必要になると考えられる．このような，修正を行うためには，試験的に形態素解析を行って形態素解析の誤り個所を見つけ，誤りの個所が修正でき，かつ正解個所の解析結果は変化しないように接続ルールや重みを変更する必要があると予想される．この作業を行うためにはルールの作成において相応の経験・知識を持つ人が，相当な時間をかける必要があると考えられる．

これに対して N-gram では，前節の実験でデータ量が増えるに比例して形態素解析率は向上していることから，形態素解析の誤り部分を修正するだけで形態素解析精度が向上でき，日本語において多少の文法的知識を持つ人なら容易に作業が可能であり，ルールベースの方法より精度の改善が容易であると考えられる．

3.5.4 未知語を含む文の形態素解析実験

未知語を含む文の形態素解析実験を行った．学習・評価には，3.5.1 節の実験と同一データを使用した．但し，辞書には学習データに出現した形態素しか登録しておらず，評価データのみにしか出現しない形態素が未知語となる．このような未知語は 632 語存在し，評価データ中の 137,691 形態素中の 859 形態素(約 0.6%)を占める．但し，形態素 N-gram は，この処理は行えないため，品詞 N-gram と複合 N-gram のみで比較実験を行った．但し，処理時間の都合上，両モデル共に Bigram のみを用いた．

表(3.5) 未知語を含む文の形態素解析性能

品詞 Bigram	複合 Bigram（分離クラス数）			
	500	1000	1500	2000
97.66	98.31	98.26	98.24	98.26

形態素解析の評価は，品詞付与の形態素 Accuracy(%)のみで評価した．これは，現在の我々の形態素解析システムでは，未知語に対し読みを付与する機能がないためである．未知語に読みを付与するためには漢字毎の読みの情報があることが最低条件となるが，現在そのようなデータを持ちあわせていないことがその理由である．また，未知語，特に固有名詞の読みは人間でも間違う場合が多く，これを自動で行うのは技術的にも困難であると考えられる．表(3.5)に結果を示す．

未知語処理を行った場合でも，複合 Bigram が品詞 Bigram よりも高い正解率を得た．辞書に全語彙が登録されている 3.5.1 節の実験では正解率が 99.13%であったから，0.8%程度低下はしているものの，98%以上の比較的高い正解率が得られた．

未知語の形態素解析誤りを分析したところ，「防音」が「防」と「音」のように，1 形態素が複数の形態素に分割された例が多数見られた．これは，「音」という形態素が辞書に登録されているため「防」という文字のみが未知語として解析された結果生じた現象である．「防」も「音」も両方普通名詞であるから，これらの語を結合させることにより，誤りを低減することが可能であると考えられる．

3.6 あとがき

本章では，連続音声認識用の N-gram 言語モデルの学習に必要な言語コーパスの整備コストの削減を目的として，品詞と可変長形態素列の複合 N-gram を用いた形態素解析手法提案した．形態素解析実験の結果，最大 99.17%の精度であり，読みまで考慮した結果でも最大 98.68%の精度を得ることができた．これは，従来の品詞 N-gram および形態素 N-gram よりも高い精度であり，

提案手法の有効性が示された．また，実験により，一万語程度の少ない学習データから学習したモデルでも 94%程度の比較的高い精度が得られることを確認した．さらに，品詞から未知語の出現確率を考えることにより未知語を含む文の形態素解析が行えるよう改良を行い，実験の結果，未知語が登録されている場合と比較して形態素解析精度の低下は 0.8%程度であることを確認した．

今後の課題としては，品詞から分離された形態素に関する遷移確率を品詞 N-gram で補間する等の方法で確率推定の信頼性をさらに向上させることが形態素解析の精度向上の上で重要であると考える．また，未知語に関しては，同一品詞の未知語と既知語とを結合させ，新たな未知語と考えること等により形態素解析率を向上させることが必要であると考えられる．さらに，音声認識に有効に活用するためには，未知語に対して読みを自動的に付与する手法の開発も行いたい．

第４章 品詞と可変長形態素列の複合 N-gram による連続音声認識

4.1 まえがき

N-gram 言語モデルは，大語彙連続音声認識の言語制約として盛んに用いられているが，2.3.2 節で述べたように，パラメータ数が非常に多いため，精度の高いモデルを構築するには，通常，新聞記事等の数千万～数億語からなる膨大な量のテキストデータが用いられる．しかし，音声認識のアプリケーションとして，実用化を前提に行われている研究においては，各種情報案内システムや，旅行会話の音声翻訳等，使用するタスクを限定したシステムが盛んに研究されている．このようなシステムでは，目的のタスクに合わせてデータを収集する必要がある．しかし，目的のタスク毎に，数千万～数億語といった莫大な量のテキストデータを収集するのは非常に困難である．このため，通常の形態素 N-gram に代わり，少量のテキストデータからでも高い認識精度を得ることを目的として，第 3 章の形態素解析で用いた『品詞と可変長形態素列の複合 N-gram』を連続音声認識の言語モデルに使用することを提案する．

4.2 節では，連続音声認識の単位として形態素列を単位とすることを検討し，さらに品詞および可変長形態素列を単位とする複合 N-gram を提案する．4.3 節では評価実験を行い，パープレキシティおよび連続音声認識の基準によりその有効性を示す．

4.2 品詞と可変長形態素列の複合 N-gram

前章 3.2 節では，連続音声認識・理解における単位について考察し，形態素を認識単位とすることの有効性について述べた．しかし，形態素を認識の単位とした場合でも，次のような問題が残る．

問題 1) 日本語には短い形態素が多く認識性能が低下する

日本語の形態素には，助詞・語幹・語尾 等，1 音節程度のかなり短い形態素が多く存在する．一般的な傾向として，短い形態素ほど出現頻度が大きいため，補間(平滑化)により，先行形態素に関わらず，常に高い遷移確率が与えられる．このため，探索空間が広がり計算コストが大きくなり，また湧き出し誤りが発生し易くなる．

問題 2) 長い範囲の形態素接続関係の表現ができない

N-gram はパラメータ数の問題により，通常 N は 2,3 程度が用いられる．すなわち，せいぜい 2 および 3 形態素間の接続関係しか表現できない．しかし，実際の発話を考えると，「～をお願いします」「～だと思うんですが」等，複数の形態素が連なったフレーズ的な言い回しが頻繁に出現するが，このような言い回しを効果的に表現することができない．

これらの問題点を解決するため，特定の形態素列を結合させた形態素列を単位として用いる方が，連続音声認識の性能向上に有効であると考えられる．しかし，このような形態素列を単位とした言語モデルを考えた場合，次の問題が生じる．

問題 A) パラメータ数増大の問題

形態素に加え，形態素列を認識単位に追加する場合，追加された形態素列の分だけ，認識単位の種類数が増大する．N-gram のパラメータ数は，V^N (V は語数，N は N-gram の次数) であるから，認識単位の種類が増大するとパラメータ数が大幅に増大し，パラメータ推定が困難になるため，言語モデルの精度が低下する恐れがある．

問題 B) ポーズ位置の問題

連続音声認識において，通常，ポーズは認識単位の前後の任意の位置で挿入可能であるとして探索処理を行う．しかし，形態素列を認識の単位とした場合，ポーズが形態素列の中に挿入される可能性もある．例えば，先例の「～をお願いします．」は「～を (ポーズ) お願いします．」と発声される可能性もあり，通常の処理では，認識不能となってしまう．

問題 A)に対して，第 3 章の形態素解析で用いた『品詞と可変長形態素列複合 N-gram』を用いる手法を提案する．通常の形態素列 N-gram では，形態素 N-gram よりもパラメータ数が大きくなるが，複合 N-gram では出現頻度が小さい形態素に関しては，品詞を単位として扱うことにより，パラメータ数を少なくすることができ，パラメータ推定の信頼性を高めることができる．また，3.3.4 節で示した，エントロピー最小化基準による複合 N-gram の生成方法を用いれば，与えられたパラメータ数で最適に近い精度の言語モデルが得られ，形態素 N-gram よりも少ないパラメータで精度を向上させることも可能である．

問題 B)に対して，我々は次のような方法で解決した．まず，形態素の読み(発音)を次のような形式で与える．

を： o{|-}

お願い： o n e g a i{|-}

“{|-}”はポーズが挿入されても，挿入されなくても良いことを意味する．全ての形態素の後にこのようなポーズ音素を挿入することにより，任意の形態素間でポーズが挿入されても認識可能となる．形態素を結合して形態素列を生成する場合，ポーズ音素も含めて読みを結合させる．この処理により「をお願い」は，

をお願い： o{|-}o n e g a i{|-}

となり，助詞の「を」の直後にポーズが挿入された場合でも認識が可能になる．

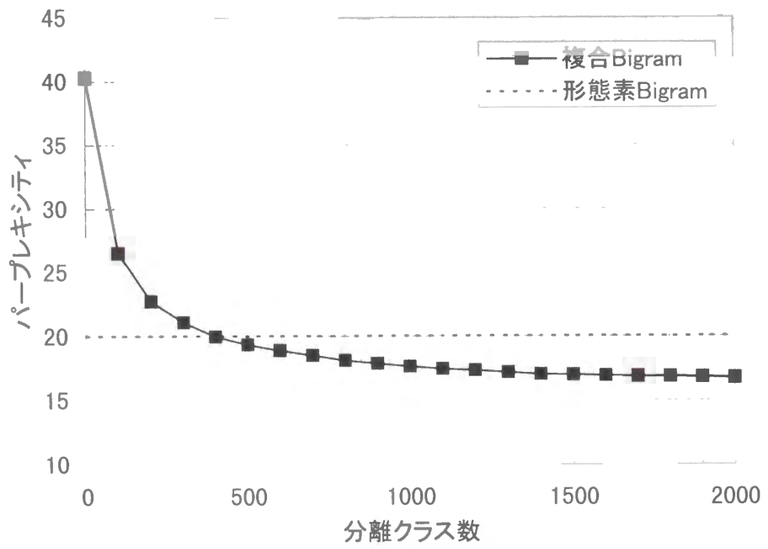
4.3 評価実験および考察

4.3.1 言語モデルの性能評価実験

提案モデルの性能を確認するため，パープレキシティ，およびパラメータ数について従来の形態素 N-gram との比較実験を行った．実験に用いたデータは ATR の自然発話旅行会話データベース[30]の，27,054 文，475,009 形態素（6,396 異なり形態素）から構成される．このうち，約 4 分の 1（6,763 文，118,732 語）をランダムに選択して評価用テストセットデータとし，残り（20,291 文，356,277 語）を言語モデルの学習用として使用した．

学習データ量は新聞記事[5]等と比較すると極めて少量なため，パラメータ数の少ない Bigram で比較実験を行った．複合 Bigram は，活用形，および活用型を含めた 156 品詞を初期状態とし，最大 2000 クラスまで分離を行い，100 分離おきにデータを採取した．また，複合 Bigram・形態素 Bigram ともに，クラス，および形態素の遷移確率は，削除補間法[12]を用いて，学習データに出現しない形態素遷移に対して確率を与えた．

クラスの分離に伴う複合 Bigram のテストセットパープレキシティの値の変化の様子を形態素 Bigram の値と共に図(4.1) に示す．



図(4.1) 分離クラスに対する複合 Bigram のパープレキシティの変化

表(4.1) 形態素 Bigram と複合 Bigram との性能比較

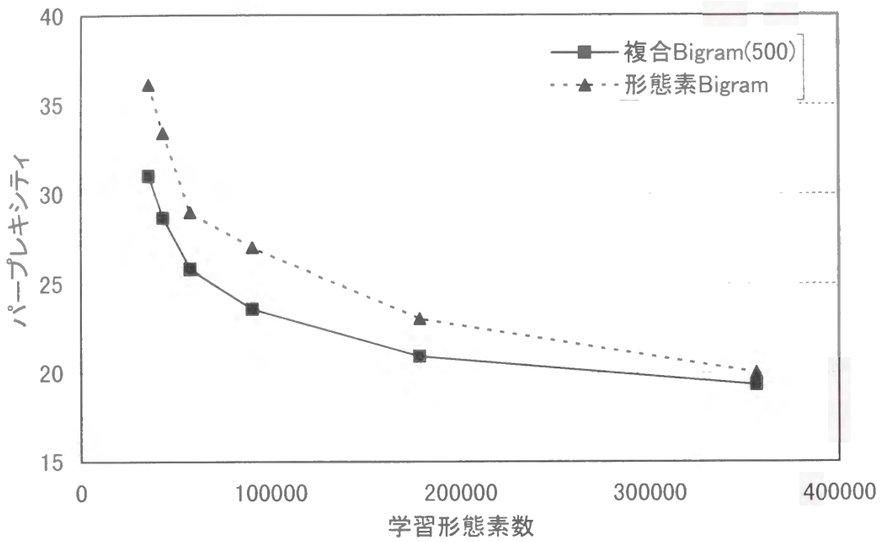
	形態素 Bigram	複合 Bigram(分離クラス数)				
		0	500	1000	1500	2000
パープレキシティ	20.02	40.27	19.34	17.64	16.98	16.75
パラメータ数	51,018	8,109	28,046	42,700	52,586	60,748

図(4.1)より，複合 Bigram のテストセットパープレキシティは，分離クラス数が増加するに従って単調に減少している．分離クラス数が 400 の時形態素 Bigram と同程度の値となり，分離クラスがそれ以上の場合は従来の形態素 Bigram よりもパープレキシティが低くなる．従って，分離クラスが 400 以上で，複合 Bigram は形態素 Bigram よりも高い精度をとなる．但し，複合 Bigram は，分離クラス数が 1,500 を超えると，パープレキシティの減少が頭打ち状態となるため，この辺りが本モデルの性能の限界であると考えられる．

表(4.1) には，パープレキシティ，およびパラメータ数について，形態素 Bigram と分離クラス数 500 おきに採取した複合 Bigram との比較を示す．但し，パラメータ数は．学習テキスト上で実際に観測した形態素遷移（複合 Bigram の場合は，品詞・形態素列の遷移も含む）の数を表す．

表(4.1)より，複合 Bigram(500) （括弧内の数字は分離クラス数を表す．以下同様）は形態素 Bigram よりもパープレキシティ値が 3%低いが，パラメータ数は形態素 Bigram に比べて 45%も少ない．また，複合 Bigram(1000)は形態素 Bigram よりもパープレキシティは 12%低いが，パラメータ数は 14%少ない．従って，複合 Bigram は，与えられたパラメータで極めて効果的な表現が可能なモデルであることが確認できた．

また，各モデルの頑健性を比較するため，学習データ量を変化させた時のテストセットパープレキシティ値の比較を図(4.2)に示す．但し，図(4.2)において，横軸の値は学習に用いた形態素数を表す．



図(4.2) 学習データ量とパープレキシティとの関係

図(4.2)より，全ての学習データを用いた時は，形態素 Bigram と複合 Bigram(500)とのパープレキシティの差は僅か 0.7 である．しかし，学習形態素数を減少させた時，複合 Bigram(500)，は形態素 Bigram よりもパープレキシティの増加は比較的小さい．この結果が示すように，複合 Bigram はパラメータ数が少ないため，頑健性が高く，少ない学習データでも，従来の形態素 Bigram より精度の高い言語モデルが構築できる．

表(4.2) 形態素 Trigram と複合 Trigram との性能比較

	形態素 Trigram	複合 Trigram(分離クラス数)				
		0	500	1000	1500	2000
パープレキシティ	15.13	38.76	16.65	15.16	14.61	14.48
パラメータ数	148,711	19,086	19,086	104,043	159,781	173,129

次に，N を増加した場合の様子を知るため，Trigram での比較を表(4.2)に示す．複合 Trigram の作成にあたっては，単純に複合 Bigram と同じクラス分類を用いた．表(4.2)より，複合 Trigram は，クラス数 1000 で形態素 Trigram と同程度のパープレキシティを示し，それ以上のクラス数では形態素 Trigram よりも低くなっている．従って，Trigram においても，複合 N-gram の有効性を示すことができた．複合 N-gram の生成時にエントロピー減少量を複合 Bigram によって計算しているため，これを基準としたクラス分類は Bigram に対して適切なクラス分類となるが Trigram に対して適切なクラス分類とはなっていないと考えられる．クラス分離の際に，複合 Trigram でのエントロピー計算を行うことにより，複合 Trigram の性能はさらに向上すると予想される．

4.3.2 他手法との比較実験

複合 N-gram の生成方法の有効性を示すため，他手法との比較実験を行った．従来提案されている手法で，複合 N-gram に類似した手法として文献[33][23]の 2 手法が挙げられる．文献[33]では，自立語のみを品詞クラスとして扱い，付属語は独立した形態素として扱う手法が提案されている．また，文献[23]では，文字列の頻度により，結合する文字を決定する方法が提案されている．これら 2 手法を組み合わせて，

1. 初期状態を品詞クラスとし
2. 全ての付属語を品詞クラスより分離し
3. 付属語対を頻度の高い順に結合させる

という手法により，複合 N-gram を生成する手法を検討した．なお，付属語として以下の品詞を用いた．これらの品詞に属する形態素は 583 語である．

- 助詞類（格助詞・引用助詞等 9 種類）
- 助動詞
- 補助動詞
- 語尾
- 接尾辞・接頭辞

表(4.3) 従来法による複合 N-gram のパープレキシティ

	538	1000	1500	2000
複合 Bigram	29.81	27.60	28.00	28.36
複合 Trigram	27.15	25.04	25.38	25.72

付属語を分離した時点(分離クラス数 583)，および付属語対の結合による分離クラスとの合計が 1000,1500,2000 の場合における，複合 Bigram，および複合 Trigram のテストセットパープレキシティの値を表(4.3)に示す．

表(4.3)より，従来法の組み合わせにより生成した複合 N-gram は，Bigram および Trigram 共に分離クラス数 1000 まではパープレキシティは減少するが，それ以上のクラス分離を行っても逆に性能が劣化している．すなわち従来法の組み合わせでは効果的な複合 N-gram が生成できない．

また表(4.1)と表(4.2)および表(4.3)とを比較すると，本論文で提案した手法により生成された複合 N-gram のパープレキシティは，Bigram および Trigram 共に従来法の組み合わせにより生成された複合 N-gram の値よりもはるかに低く，提案手法は与えられたパラメータで効果的な複合 N-gram が生成できることが確認できた．これは，従来手法では品詞および出現頻度という直感的な尺度でクラス分離を行っているのに対し，提案手法ではエントロピーという情報理論の基準に基づきクラス分離を行っているためである．

4.3.3 連続音声認識の評価実験

連続音声認識に適用し，複合 N-gram の有効性の評価実験を行った．表(4.4)に示す条件に基づいて音響モデルを作成し，単語グラフによる連続音声認識法[7]により認識候補を探索し，形態素 Bigram，複合 Bigram（分離クラス 0,500,1000）との性能比較を行った．認識対象文は，データベース中のホテル予約タスクより選択した 16 対話であり，これらの会話については，言語モデルの学習の対象外である．また，認識対象語は次の 2 通りで実験した．

・辞書 Fullset:　データベースに出現する全ての形態素（6,396 語）

・辞書 Subset:　ホテル予約タスクに出現する形態素（1,321 語）

表(4.5) 音響分析条件

分析条件	
サンプリング周波数	12KHz
ハミング窓	20ms
フレーム周期	10ms
使用パラメータ	16 次 LPC ケプストラム+Δケプストラム + Log パワー + Δ log パワー
音響モデル	
Hmnet[34] (ML-SSS[35]) 男女別不特定話者モデル [36] 800 状態 5Mixture	

表(4.6)形態素認識率の比較

	形態素 Bigram	複合 Bigram		
		0	500	1000
辞書 Subset	54.62	48.46	59.74	58.13
辞書 Fullset	-	47.28	58.23	60.41

各言語モデルで尤度 1 位の文認識候補の形態素 Accuracy を表(4.6)に示す．なお，メモリ容量と計算時間の都合上で，Fullset の辞書では，形態素 Bigram の認識実験を行うことができなかった．

表(4.6)より，辞書 Subset の場合の比較では，複合 Bigram の分離クラス数 500，および 1000 の場合で，形態素 Bigram よりも認識率がおよそ 4〜5%程度向上しており，連続音声認識結果においても，複合 Bigram が優れていることが確認できた．複合 Bigram(500)と形態素 Bigram とでは，パープレキシティの差は小さいが，これは，複合 Bigram では，語尾・格助詞等，比較的出現頻度の高い短い形態素が結合されており，それらの形態素の湧きだし誤りの発生が抑えられることが認識率の向上に寄与しているためである．

4.4 あとがき

本章ではN-gram言語モデルの次形態素予測精度，およびパラメータ推定の信頼性向上の両立を図るため，品詞および可変長形態素列を単位とする複合N-gram連続音声認識に適用し，その効果を検討した．言語モデルの評価実験の結果，品詞および可変長形態素列を単位とする複合N-gramは，パープレキシティ値の比較により，従来の形態素N-gramよりも次形態素予測精度が高いことが分かり，また頑健性が高く，少ないデータ量でも比較的精度の高いモデルが構築できることを確認した．また，連続音声認識に適用した結果，通常の形態素Bigramよりも高い認識性能を得ることが確認できた．さらに，インプリメントの観点からも，パラメータ数の少ない複合Bigramは形態素Bigramよりも有利であり，大容量のメモリを必要とする大語彙の連続音声認識システムの構築も容易である．

第５章 最大事後確率推定による N-gram 言語モデルのタスク適応

5.1 まえがき

第3章および第4章では，『品詞と可変長形態素列の複合N-gram』による形態素解析，および連続音声認識を提案した．『品詞と可変長形態素列の複合N-gram』はパラメータが少なく，かつ高い精度が得られるため少量の言語コーパスからでも比較的精度の高い形態素解析および連続音声認識が可能である．しかし，本モデルを用いても，数万形態素程度のコーパスがなければ満足できる精度を得ることはできない．

現在の連続音声認識では，性能を向上させるために，タスクを限定したシステムを構築する場合が多く，タスク毎に言語コーパスを整備する必要がある．しかし，タスク毎に数万単語（形態素）程度の言語データを整備するには大きなコストを必要とする．特に，日本語の場合は，英語等の言語のように文章において単語の区切りが明確ではない．このため，通常は形態素を単位としたN-gramを使用するが，この場合文章データ収集の作業量の他に，形態素解析の作業量も必要となる．第3章で述べたように，形態素解析を自動的に行うにしても質の高い形態素データを整備するためにはある程度の作業量が伴う．

本章では，目的のタスクに関して言語コーパスのデータ量が極めて少量しかない場合でも精度の高いN-gram言語モデルを構築する手段として，N-gram言語モデルのタスク適応について考える．タスク適応は，目的のタスクのデータは少量でも，他のタスクのデータは大量に存在する場合，他のタスクの大量のデータで学習したN-gramを，目的のタスクの少量のデータに適応させる方法である．タスク適応により，データ量不足の問題を解決し，かつ適応によっ

て目的のタスクの言語的特徴を効果的に表現することができるため，目的のタスクに関して N-gram 言語モデルの精度を向上させることができる．

タスク適応の手法として，複数のタスクの N-gram を線形結合する方法が提案されているが[25][37][38]，我々は，最大事後確率推定（以後 MAP 推定と略： Maximum A-posteriori Probability Estimation）を用いた手法を提案する．MAP 推定は，理論的に整備されたパラメータ推定法であり，実際音響モデルの話者適応の手法として用いられその有効性が確認されている方法である．また，従来の適応方法と比べて，形態素組み毎に適応の度合いを決定できいるため，より精密な適応が可能であると考えられる．

5.2 節では，MAP 推定による N-gram 遷移確率の導出を行い，5.3 節では MAP 推定のタスク適応への適用方法を示す．5.4 節では実験により，MAP 推定によるタスク適応の効果，従来の線形結合による適応との比較，連続音声認識における効果によりその有効性を示す．

5.2 最大事後確率推定による N-gram 遷移確率

5.2.1 最大事後確率推定の概念

通常，N-gram の遷移確率は，最尤推定(以後 ML 推定と略（Maximum Likelihood Estimation）)を用いて推定される．ML 推定では，観測したサンプル値 x に対して，尤度関数 $f(x|p)$ を最大にする値として確率値 p_{ML} が定められる．

$$p_{ML} = \arg\max_p f(x|p) \qquad (5.1)$$

N-gram の場合，推定対象の確率 p は，直前の $N\text{-}1$ 形態素列 h から次の形態素 w への遷移確率 $p(w|h)$ である．観測サンプル，すなわちテキストデータにおいて形態素列 h が $c(h)$ 回観測され，その内形態素 w が後続する場合（確率 p）が $c(h,w)$ 回，w 以外の形態素が後続する場合（確率 $1\text{-}p$）が $c(h)\text{-}c(w)$ 回であるから，尤度関数 $f(x|p)$ は，

$$f(x|p) = p^{c(h,w)}(1-p)^{c(h)-c(h,w)} \qquad (5.2)$$

となる．$f(x|p)$ の最大化条件 $d/dp\ log\ f(x|p) = 0$ を解くことにより，N-gram の遷移確率は次のように計算される．

$$p_{ML}(w|h) = c(h,w) / c(h) \qquad (5.3)$$

これが 2.2 節で示した，最尤推定による N-gram の遷移確率を求めるための式である．本式より，最尤推定では，もし形態素列 h,w が観測データ上で出現しない場合，$c(h,w)=0$ であるから，遷移確率は 0 と推定される．

これに対して，MAP 推定では，観測したサンプル値 x に対して，事後関数 $l(p|x)$ を最大化する値として，確率を推定する．

$$p_{MAP} = \arg\max_p l(p|x) \qquad (5.4)$$

ベイズ則を用いると，本式は次のように変形できる．

$$p_{MAP} = \arg\max_p f(x|p)g(p) \qquad (5.5)$$

ここで，$g(p)$ は，確率 p に関して何らかの情報より先見的に与えられる事前分布である．従って，MAP 推定を用いると，N-gram の遷移確率はある事前知識より得られる分布 $g(p)$ に従う変数とし，この事前分布と実際に観測されたサンプル値とを用いて，実際の遷移確率が推定される．このため，観測データで出現しない形態素遷移に対しても，事前知識により 0 でない遷移確率を与えることができる可能性がある．

5.2.2 最大事後確率推定による N-gram 遷移確率の導出

本節では，MAP 推定による N-gram の遷移確率を求める方法を示す．但し，変数の定義は前節と同一である．

まず，遷移確率 p の事前分布 $g(p)$ としてベータ分布（$ap^{\alpha-1}(1-p)^{B-1}$，a は正規化のための定数） を用いる．ベータ分布を用いる理由は次の 2 点である．

・ベータ分布は 2 項分布の共役分布であり，MAP 推定によるパラメータの解が直接計算可能である．

・パラメータ α，β を変化させることにより，様々な形状の分布を表すことができる．

式(5.5)の MAP 推定の定義に従うと，遷移確率 p_{MAP} は，尤度関数 $f(x|p)$ と事前分布 $g(p)$ とを用いて次の式を満たす値として求められる．

$$p_{MAP}(w|h) = p^{c(h,w)}(1-p)^{c(h)-c(h,w)} ap^{\alpha-1}(1-p)^{\beta-1} \equiv L(p) \qquad (5.6)$$

$L(p)$ が最大となるための条件 $d/dp\ log\ L(p) = 0$ を p について解くと，形態素 Bigram の遷移確率は次のように求められる．

$$p_{MAP(w|h)} = \frac{c(h,w)+\alpha-1}{c(h)+\alpha+\beta-2} \qquad (5.7)$$

αおよびβは，事前分布であるベータ分布のパラメータであるが，これらは次のように求めることができる．

ベータ分布の平均μおよび分散σ^2は以下の式となることが知られている[17][19].

$$\mu = \frac{\alpha}{\alpha+\beta} \qquad (5.8)$$

$$\sigma^2 = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} \qquad (5.9)$$

これらの式を α，$\alpha+\beta$ について解くと，

$$\alpha = \frac{\mu^2}{1-\mu}\sigma^2 - \mu \qquad (5.10)$$

$$\alpha+\beta = \frac{\mu(1-\mu)}{\sigma^2} - 1 \qquad (5.11)$$

が得られる．

以上より，事前分布の平均μ・分散σ^2を式(5.10)および(5.11)に代入することによりαおよび$\alpha+\beta$が得られ，これらの値と，テキストデータから形態素列 h および h,w の出現頻度を求めることにより得られる $c(h)$, $c(h,w)$ とを式(5.7)に代入することにより，MAP 推定による形態素 N-gram の遷移確率 $p_{MAP}(w|h)$ を求めることができる．

5.3 最大事後確率推定のタスク適応への応用

5.3.1 タスク適応における事前・事後知識

MAP 推定を行うためには，事前分布および観測サンプルを定義する必要がある．本研究では，MAP 推定をタスク適応に応用するため，事前分布をタスク毎の遷移確率の分布とし，また，観測サンプルを目的のタスクのテキストデータと定める．この定義の下で，MAP 推定によるタスク適応 N-gram の遷移確率を求める手順を以下に示す．

複数のタスク毎の遷移確率の分布を MAP 推定に用いる事前分布とした場合，これをベータ分布に従うと仮定した場合，式(5.10)および(5.11)より，この分布の平均および分散からαおよびβを求めることができる．但し，出現頻度を考慮するため，平均および分散は加重平均および加重分散を用いる．これらの値は，次式により計算される．

$$\mu = \frac{\sum_i c_i(h)p_i(w|h)}{\sum_i c_i(h)} \qquad (5.12)$$

$$\sigma^2 = \frac{\sum_i c_i(h)p_i(w|h)^2}{\sum_i c_i(h)} - \mu^2 \qquad (5.13)$$

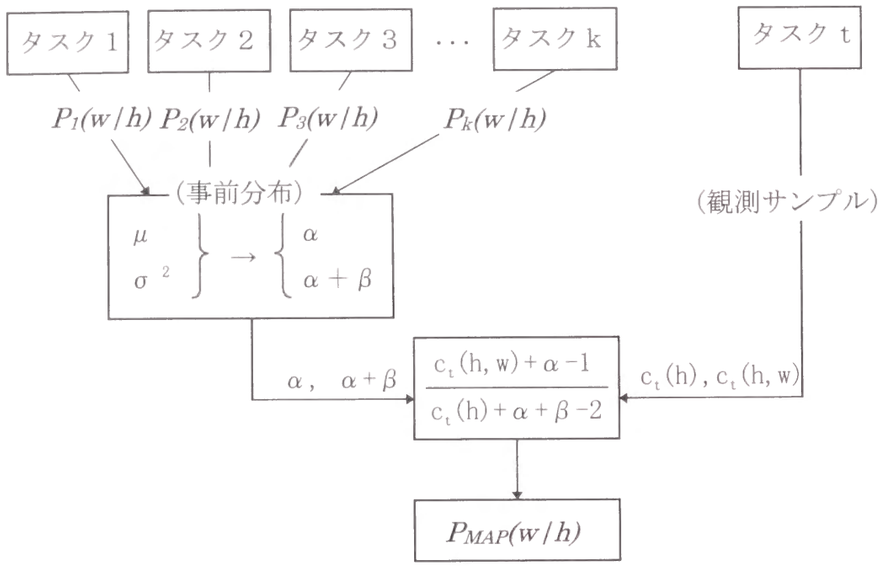
本式において，$c_i(h)$ はタスク i のテキスト内の $(N-1)$ 形態素列 h の出現頻度，$p_i(w/h)$ はタスク i における形態素列 h から w への遷移確率である．但し，各タスク毎の遷移確率 $p_i(w/h)$ は最尤推定によって求める．

観測サンプルを目的のタスクのテキストデータとすると，前節の $c(h)$ および $c(h,w)$ は次のように表される，

・$c(h)$：目的のタスクのデータ中の形態素列 h の出現頻度

・$c(h,w)$：目的のタスクのデータ中の形態素列 h,w の出現頻度

これらμ，σ^2, $c(h)$, $c(h,w)$ を前節の式(5.7) (5.10) および(5.11)に代入することにより，N-gram の遷移確率 $p_{MAP}(w_n/w_1^{n-1})$ が得られる．この遷移確率の計算を，全ての N 形態素列の h,w について行うことにより MAP 推定によるタスク適応 N-gram を生成することができる．図(5.1)に，これらの一連の手順を図示した．



図(5.1) MAP 推定によるタスク適応 N-gram の遷移確率算出

5.3.2 back-off smoothing による遷移確率の平滑化

前節で，MAP 推定によるタスク適応の基本原理を述べたが，実際に言語モデルとして使用するには，2 つの問題がある．1 つは，平滑化の問題である．不特定タスクの大量のデータを用いても，出現しない形態素列が存在し，MAP 推定の事前分布の平均・分散を求めることができない．従って，平滑化によりテキストに出現しない形態素組みに対して，遷移確率を与える必要がある．もう 1 つの問題は，本研究で提案するタスク適応手法は，全ての遷移確率を独立に求める手法であるため，遷移確率の和が 1 になるとは限らない．連続音声認識等に適用する際は特に支障はないが，パープレキシティ等の計算により他の言語モデルとの比較を行う場合は，正しい評価ができない．本節では，近年盛んに用いられている back-off 平滑化[13]の手法を応用して，これらの問題を解決する方法を示す．

遷移確率を求めようとする N 形態素列 w_1^n $(=h,w)$ が不特定タスクデータに含まれる場合は，既に示した MAP 推定によるタスク適応手法により遷移確率 $p_{MAP}(w_n/w_1^{n-1})$ を求め，Turing 推定により，確率 $p_{MAP}(w_n/w_1^{n-1})$ $(=p_{MAP}(w/h))$ を小さく（Discounting）し，遷移確率 $P_s(w_n/w_1^{n-1})$ とする．但し，Discounting の係数は全タスクのデータにおける形態素列 w_1^n の出現頻度 $c_I(w_1^n)$ を用いて計算する．Discounting により生じた確率の余剰分を w_1^n が不特定タスクデータに含まれない形態素連鎖に対して，(n-1)-gram の遷移確率 $P_s(w_n/w_2^{n-1})$ に比例して配分する．こうして，より低次の N-gram の遷移確率を再帰的に割り当てることにより高次 N-gram の遷移確率を求める．

以上をまとめると，タスク適応 N-gram の平滑化後の遷移確率 $P_s(w_n|w_1^{n-1})$ は次式で表される．

$$P_s(w_n|w_1^{n-1}) = \begin{cases} c_I(w_1^{n-1})>0\text{の場合} & \tilde{P}(w_n|w_1^{n-1}) \\ c_I(w_1^{n-1})=0, c_I(w_2^{n-1})>0\text{の場合} & \alpha(w_1^{n-1})P_s(w_n|w_2^{n-1}) \\ c_I(w_1^{n-1})=0, c_I(w_2^{n-1})=0\text{の場合} & P_s(w_2^{n-1}) \end{cases} \quad (5.14)$$

上式で，$\tilde{P}(w_n|w_1^{n-1})$ はタスク適応により得られる確率に Discount 係数をかけたものである．

$$\tilde{P}(w_n|w_1^{n-1}) = \frac{c_I(w_1^n)+1}{c_I(w_1^n)} \frac{n_{c_I(w_1^n)+1}}{n_{c_I(w_1^n)}} p_{MAP}(w_n|w_1^{n-1}) \quad (5.15)$$

但し，n_k は不特定タスクテキスト中に k 回出現する形態素列の種類数である．また同式で，$\alpha(w_1^{n-1})$ は正規化のための係数であり，次のように求められる．

$$\alpha(w_1^{n-1}) = \frac{1-\sum_{w_n:c_I(w_1^n)>0} \tilde{P}(w_n|w_1^{n-1})}{1-\sum_{w_n:c_I(w_1^n)>0} \tilde{P}(w_n|w_2^{n-1})} \quad (5.16)$$

以上の Back-off 平滑化を応用した手法を用いることにより，求める N-gram 遷移確率の N 形態素遷移がデータ中に出現しない場合は，(N-1)-gram 以下の低次の遷移確率によって確率値を与えることができる．また，式(5.16)

において α を求める際に正規化を行うため，遷移確率の和は自動的に 1 に正規化される．

5.4 評価実験および考察

5.4.1 言語モデルのタスク適応効果の評価実験

提案したタスク適応の有効性を確認するため，評価実験を行った．実験に用いたデータは，ATR 自然発話データベース[30]で，1,098 会話，449,070 形態素(のべ)，6,797 (異なり)形態素からなる．また，このデータベースは表(5.1)に示すように，15 タスクから構成されている．これらのデータの内，約 4 分の 1 の会話をランダムに選んでテストセットとして，残りの会話を学習セットとして使用した．但し，各タスクから最低でも 1 会話はテストセットとして選択している．

表(5.1) タスク内容

タスク番号	会話数	内容
1)	491	ホテルのサービス
2)	351	ホテルの部屋の予約
3)	50	旅行パックの問い合わせ
4)	36	ホテルの会議室の相談・予約
5)	28	交通手段の問い合わせ
6)	24	ホテルの部屋の相談
7)	22	飛行機のフライトの予約
8)	22	バス・列車の切符の問い合わせ
9)	20	レンタカーの問い合わせ
10)	14	コンサートのチケットの予約
11)	12	レストランの予約
12)	8	レストランの予約
13)	8	料理の注文
14)	8	道案内
15)	4	ショッピング

言語モデルとしては，次の 3 種類のモデルを考える．

・不特定タスクモデル：全タスクのテキストで作成した N-gram

・特定タスクモデル：　各タスクのテキストのみで作成した N-gram

・タスク適応モデル：　MAP 推定により各タスクに適応させた N-gram

これらのモデルをタスク毎に，形態素 Bigram，Trigram で作成した．モデル・タスク毎のテストセットパープレキシティ値を表(5.2)に示す．

タスク適応モデルのパープレキシティは，不特定タスクモデルと比較して，タスク全体の平均で約 29% (Bigram)，21% (Trigram) 低くなっている．特定タスクモデルと比較しても，約 29% (Bigram)，31% (Trigram)低く，提案したタスク適応手法により言語モデルの精度が向上することが確認できた．また，タスク適応モデルのパープレキシティは，全てのタスクで Bigram，Trigram の両方において，不特定タスクモデル・特定モデルのいずれよりも低く，安定してタスク適応の効果が得られることが分かった．

表(5.2) 各モデルのタスク別パープレキシティ

番号	形態素数		不特定タスクモデル		特定タスクモデル		タスク適応モデル	
	Train	Test	Bigram	Trigram	Bigram	Trigram	Bigram	Trigram
1)	136,175	42,698	23.177	17.954	22.922	18.261	22.159	17.391
2)	118,124	38,697	14.844	10.080	13.843	9.942	13.404	9.553
3)	19,471	6,610	26.539	17.398	23.934	17.201	20.042	14.364
4)	15,302	5,075	31.285	24.706	38.158	32.851	27.994	23.041
5)	10,791	2,983	24.191	16.563	21.772	16.577	17.471	13.187
6)	8,802	2,999	17.136	11.199	14.666	11.402	11.867	8.712
7)	8,617	2,722	21.114	14.186	18.391	14.649	14.644	11.060
8)	8,537	2,193	21.148	14.296	14.225	11.307	12.769	10.208
9)	8,567	2,528	25.171	18.167	26.007	20.822	19.141	14.922
10)	5,036	1,608	16.592	10.832	14.060	10.929	10.915	7.815
11)	5,326	1,439	12.982	8.887	12.179	9.631	9.350	6.908
12)	3,578	1,165	32.918	19.395	25.369	18.372	18.034	12.756
13)	2,378	1,075	30.288	22.405	34.246	32.202	18.185	16.735
14)	2,572	908	35.545	27.109	46.611	42.088	25.396	21.947
15)	1,750	509	44.156	34.232	47.543	44.545	25.948	23.186
平均	−	−	25.139	17.827	24.928	20.719	17.821	14.119

不特定タスクモデルと特定タスクモデルのパープレキシティを比較すると，Bigram では，特定モデルのパープレキシティの方が不特定モデルよりも低い値を示す場合が多いが，Trigram では，不特定タスクモデルの方が特定タスクモデルよりも低い場合が多い．これは，形態素 Bigram では，学習データのスパース性が低いため，特定のタスクの言語特徴を効果的に表現することのできる特定タスクモデルの方が有利であるが，Trigram では，学習データがよりスパースであるため，特定タスクの少ない量のデータでは，信頼できるパラメータ推定が行われていないことが原因と考えられる．従って，タスク適応を行うと，大量のデータを用いたことにより，学習データのスパース性が解決でき，さらに，適応を行うことにより，そのタスクの言語特徴を表現できたと考えられる．

データ量が少ない 12),15) 等のタスクでは，タスク適応によるパープレキシティの減少の割合が大きくなる傾向にある．特にタスク 15)では，不特定タスクモデルと比較して 41%(Bigram)および 32%(Trigram)，特定タスクモデルと比較して 45%(Bigram)および 48%(Trigram)と，パープレキシティの減少が非常に大きい．すなわち，目的のタスクのデータが少量しか集まらない場合に，タスク適応を使用する効果が大きいと言える．

上記実験では，MAP 推定の事前分布を推定する際に，適応先タスクのデータを含めて推定を行った．これは，全タスクのデータを集めても約 45 万形態素と，N-gram を構築するにはかなり少量であり，正確な事前分布を求めることは困難であると考えたためである．しかし，音声認識の結果から逐次的に適応を行う場合等，事前分布に適応先タスクのデータを常に含めることができるとは限らない．このため，事前分布の推定に適応先タスクのデータを含めない場合について実験を行った．使用したデータ，およびその他の条件は，上記実験と同一である．パープレキシティを表(5.3)に示す．

表(5.3) 事前分布に適応先タスクのデータを含めない場合のタスク適応 N-gram のパープレキシティ

タスク番号	Bigram	Trigram
1)	24.096	19.608
2)	14.273	10.414
3)	23.147	19.201
4)	29.945	26.039
5)	19.514	16.801
6)	12.932	10.409
7)	16.611	13.959
8)	14.763	12.674
9)	21.364	18.491
10)	12.181	9.420
11)	10.094	8.134
12)	22.693	19.247
13)	19.718	18.131
14)	27.609	25.528
15)	31.752	29.239
(平均)	20.046	17.153

表(5.2)および表(5.3)より，事前分布の推定に適応先タスクのデータを含めない場合においても，ほとんどのタスクにおいて MAP 推定による N-gram はタスク特定・不特定モデルよりもパープレキシティが小さく，全タスクの平均では，タスク不特定モデルと比較して約 20%(Bigram)および 4%(Trigram)，特定タスクモデルと比較して約 20%(Bigram)および 17%(Trigram)パープレキシティが減少している．従って，事前分布の推定に適応先タスクのデータを含めない場合においても MAP 推定によるタスク適応の効果が得られることが確認できた．しかし，事前分布の推定に適応先タスクのデータを含めた場合よりパープレキシティの減少効果は小さく，特に適応データが量が多い 1),2)等のタスクにおいては，逆に不特定および特定タスクモデルよりもパープレキシティが大きい場合も存在する．

従って，適応データがある程度多く事前に与えられる場合は，適応データも含めて事前分布の推定を行うことにより効果的なモデルが得られることが明

らかになった．また，事前に多くのデータが得られない場合は，適応データを事前分布に含めなくてもタスク適応の効果が認められ，適応データの量が多くなった時に，何らかのタイミングでそれまでの適応データを事前分布に取り入れるのが本手法による効果的な適応方法であると考えられる．

次に，Bigram を例に，不特定タスク，特定タスクモデル，タスク適応モデルの特性を比較を，タスク 15)における形態素間の遷移確率値の実例により示す．

例 1)	じゃあ	その	漢方薬	下さ	い
・不特定タスク Bigram：	3.83×10^{-2}	9.68×10^{-4}	1.93×10^{-1}	5.78×10^{-1}	
・特定タスク Bigram：	1.39×10^{-1}	3.97×10^{-2}	1.85×10^{-1}	5.55×10^{-1}	
・タスク適応 Bigram：	6.65×10^{-2}	2.17×10^{-2}	1.93×10^{-1}	7.28×10^{-1}	

全体的に，タスク適応 Bigram は，不特定タスク Bigram と特定タスク Bigram との中間的な値を示している．「その」→「漢方薬」という遷移では，不特定タスク Bigram の遷移確率がかなり小さい値になっている．これは，「その」という形態素は，日本語のあらゆる文章で頻繁に出現する形態素であるが，「その」の次に「漢方薬」という形態素が出現する確率は非常に低いと考えられる．しかし，本タスクでは，薬局における買い物というタスクであるため，「漢方薬」という形態素は通常の会話に比べ非常に高い．このため，特定タスク Bigram では，3.97×10^{-2} という比較的大きな値となっている．タスク適応 Bigram では，2.17×10^{-2} という値で，特定タスク Bigram よりは小さな値であるが，不特定タスク Bigram と比較すると 20 倍以上大きな値であり，そのタスクの言語特徴をより効果的に表現するというタスク適応の目的を果たしている．

例 2)	はい	お願いします	(文末)
・不特定タスク Bigram：	6.99×10^{-4}	9.24×10^{-1}	
・特定タスク Bigram：	1.00×10^{-6}	5.03×10^{-1}	
・タスク適応 Bigram：	3.50×10^{-4}	9.23×10^{-1}	

「はい」→「お願いします」という遷移確率が，特定タスク Bigram において非常に小さな値となっている．これは，タスク 15)の学習データが少ないためこの表現が一度も出現しなかったため，遷移確率の平滑化によって与えられている値である．「はいお願いします」という表現は，日本語では普通の表現であるため，全学習データ中には含まれる表現であるため，不特定タスクでは 6.99×10^{-4} という値を得ることができる．タスク適応 Bigram では 3.50×10^{-4} という不特定 Bigram の半分程度の値を示している．これは，タスク適応により，そのタスクの学習データに含まれない表現でも，他のタスクのデータを用いることにより比較的大きな値を得ることができた例である．

以上 2 例により，そのタスクの言語特徴が効果的に表現でき，および目的のタスクの学習データ量が少なくても他のタスクのデータを用いることにより補うことができるというタスク適応の目的を実現していることが分かった．

5.4.2 線形結合によるタスク適応との比較実験

従来法との比較のため，[25][37][38]等で提案されている，複数のタスクの線形結合によるタスク適応手法の実験を行った．本手法は，下式のように各タスク i 毎の遷移確率に重み λ を乗じ，全タスクについて和をとるものである．

$$P(w|h)=\sum_i \lambda_i P_i(w|h) \qquad (5.17)$$

但し，重み係数の和は 1($\Sigma_i \lambda_i = 1$)であり，それぞれの重み係数の値 λ_i は削除補間法 [12] によって求められる．また，それぞれのタスクの N-gram は Back-off Smoothing [13] によりあらかじめ平滑化してある．線形結合により適応を行った結果を表(5.4)に示す．

線形結合によるタスク適応を行った N-gram のパープレキシティは，不特定タスクモデルと比較して，タスク全体の平均で約 24%(Bigram)および 10%(Trigram) 低く，また特定タスクモデルと比較して，約 23% (Bigram)，22%(Trigram)低く，適応によるパープレキシティの減少の効果が見られる．しかし，減少の割合は事前分布の推定に適応先タスクのデータを含めた場合の MAP 推定によるタスク適応モデルの方が大きく，本論文で提案した手法の方がより有効であることが実験結果から明らかになった．

表(5.4) 線形結合によるタスク適応 N-gram のパープレキシティ

タスク番号	Bigram	Trigram
1)	23.860	20.248
2)	11.305	10.721
3)	21.571	16.088
4)	28.179	24.154
5)	18.752	15.176
6)	13.132	10.550
7)	16.161	13.529
8)	13.697	11.537
9)	20.380	16.719
10)	12.151	9.635
11)	10.143	8.148
12)	20.158	15.358
13)	19.801	17.827
14)	19.801	23.531
15)	26.921	28.328
平均	19.190	16.103

また，線形結合によるタスク適応では，不特定タスク・特定タスク N-gram よりもパープレキシティが大きくなっているタスクも存在するが，MAP 推定による適応モデルは，全てのタスクにおいて不特定タスク，および特定タスク N-gram よりもパープレキシティが低く，本論文で示した MAP 推定によるタスク適応手法により，タスクに依存せず安定して適応の効果が得られることが分かった．

5.4.3 連続音声認識におけるタスク適応効果の評価実験

連続音声認識におけるタスク適応の効果を調べた．認識の対象は，前節のタスク 8)，バス・列車の切符の問い合わせタスクとした．音響パラメータ・音響モデルは第 4 章で行った実験と同一条件とした．また，単語グラフサーチ[7]を用いて認識結果を探索した．

表(5.5) 各モデルの連続音声認識精度(形態素 Accuracy %)の比較

不特定タスク Bigram	特定タスク Bigram	タスク適応 Bigram
66.65	73.62	71.82

特定タスクモデルはそのタスクに出現しない形態素への遷移確率が 0 になり，音声認識においては，不特定タスクモデル・タスク適応モデルと比較して有利になると考えられるため，認識対象語彙を 8)タスクに出現する 713 語に限定した．

言語モデルの学習には，表(5.2)と同じく 8,537 形態素を用い，評価データの 2,193 形態素について認識を行った．不特定タスクモデル，特定タスクモデル，タスク適応モデルに関して，形態素 Bigram モデルを連続音声認識に適応した場合の形態素 Accuracy(%)を表(5.5)に示す．

表(5.5)より，タスク適応 Bigram は不特定タスク Bigram よりも 8.17%高く(改善率 24.9%)，また，特定タスク Bigram よりも 1.2%高い(改善率 4.5%)認識率が得られた．タスク適応 Bigram は，不特定タスク Bigram より，認識率が大幅に向上したことより，タスク適応の効果は示せた．しかし，特定タスク Bigram と比較すると，パープレキシティが約 30%程度低下しているのに比べると，認識率の向上は予想外に小さかった．この原因としては，次の 2 点が考えられる．

1. 言語データの偏り

   特定タスクモデルは学習データがわずか 8,537 形態素しかないのにかかわらず，比較的高い認識率が得られている．これは，データベースを収録する際，チェックシートを用いて会話の制御を行ったため，固有名詞や，日付，電話番号等で同一形態素がよく用いられ，評価セットは学習セットに極めて近い内容であったことが原因と考えられる．

2. パープレキシティと認識率との相関の低さ

   一般的にパープレキシティと音声認識率との間には相関があるとされているが，最近ではこれらの間には相関が低いという報告もされており[39][40]，パープレキシティが低くても，認識率が必ずしも向上するとは

限らない．このため，認識率との相関性が高い尺度を検討し，再度タスク適応の評価を行うべきであると考えられる．

5.5 あとがき

本章では，最大事後確率（MAP）推定による N-gram のパラメータ推定の方法を示し，タスク適応の適応方法を提案した．実験の結果，タスク適応によるパープレキシティの減少効果が確認され，数千語程度の少量のテキストを用いるだけで，適応前のモデルよりも大幅に精度の良い N-gram が構築できることがわかった．また，連続音声認識に適用した結果，形態素認識率 1.2%の向上(改善率 4.5%)が得られ，連続音声認識においても有効であることが示せた．

しかし，パープレキシティによる減少効果に比べて音声認識率改善効果が予想以上に小さかったため，この原因を明らかにし音声認識に有効な適応手法等を考えることが今後の課題である．

第６章 隠れマルコフモデルによる頑健な音声言語理解

6.1 まえがき

近年，隠れマルコフモデルによる音響モデル，および N-gram による言語モデルを用いた連続音声認識が盛んに研究されており，日本語でも，数万語彙の連続音声認識における単語認識率が 90%程度と報告されており，実用レベルに近づきつつある[5][6]．音声認識技術を用いたアプリケーションとしては，読み上げ文をそのまま出力するディクテーションシステムの研究が盛んであるが，旅客機案内システム[41]，電話番号案内システム[42]，音声翻訳システム[8]等，音声認識結果を理解し，ユーザーに情報を提供するいわゆる音声理解システムも盛んに研究されている．

現在，音声理解システムのための言語理解の手法は，システムが扱うことのできる文型を限定した手法[41][42]，キーワードを用いた手法[43]，文法ルールを用いて構文解析を行う手法[44]等が提案されている．発話内容の文型を限定する手法は，理解のための処理が容易であり理解率の向上が期待できるが，限定された文型以外の入力に対処できず，柔軟なシステム構成をとりにくい．一方，キーワードを用いた手法は，より自由な発話を扱える利点はあるが，キーワードのみでは正確な理解を得るのは困難であると考えられ，文献[43]ではユーザーインターフェースによりキーワード間の単語の補完を必要としている．また，文法ルールを用いた構文解析による手法は，文型を限定する手法よりは，発話の自由度が高いが，文法的に正しい文章でないと理解できず，自然発話や認識誤りを含む文等の非文法的な文の理解は困難であると考えられる．

本章では，より自然な発話を扱うことができ，多少の認識誤りを含む文に対しても頑健な理解も可能であり，かつ正確な理解が可能な音声理解システムの構築を目標として，統計的処理による音声理解手法を検討した．提案する手法は，中間表現の各要素から文の生成確率を与える隠れマルコフモデル，および

中間表現の各要素間の共起確率を用いた方法である．統計処理による手法は，文章を解析して言語理解に必要な情報を自動的に学習できるため，専門家による文法ルールの作成を必要としない上に，文法として規定しにくい自由発話に対する理解も期待できる．また，隠れマルコフモデルは，文の構造を状態の遷移として表現することができるためキーワードによる理解系よりも正確な理解ができる可能性があり，また，文の局所的な構造をモデル化するため，局所的な音声認識誤りに対しても頑健な理解が行える可能性がある．

また，連続音声認識において，隠れマルコフモデル，N-gram といった統計的モデルが盛んに用いられており，その有効性が広く認識されているため，言語理解においても効果が期待される．

6.2 節では，本研究の対象とした音声理解システムについて説明する．6.3 節では統計的モデルによる言語理解の手法を提案し，6.4 節で実験によりその有効性を示す．

6.2 音声理解システム概要

6.2.1 システム概要

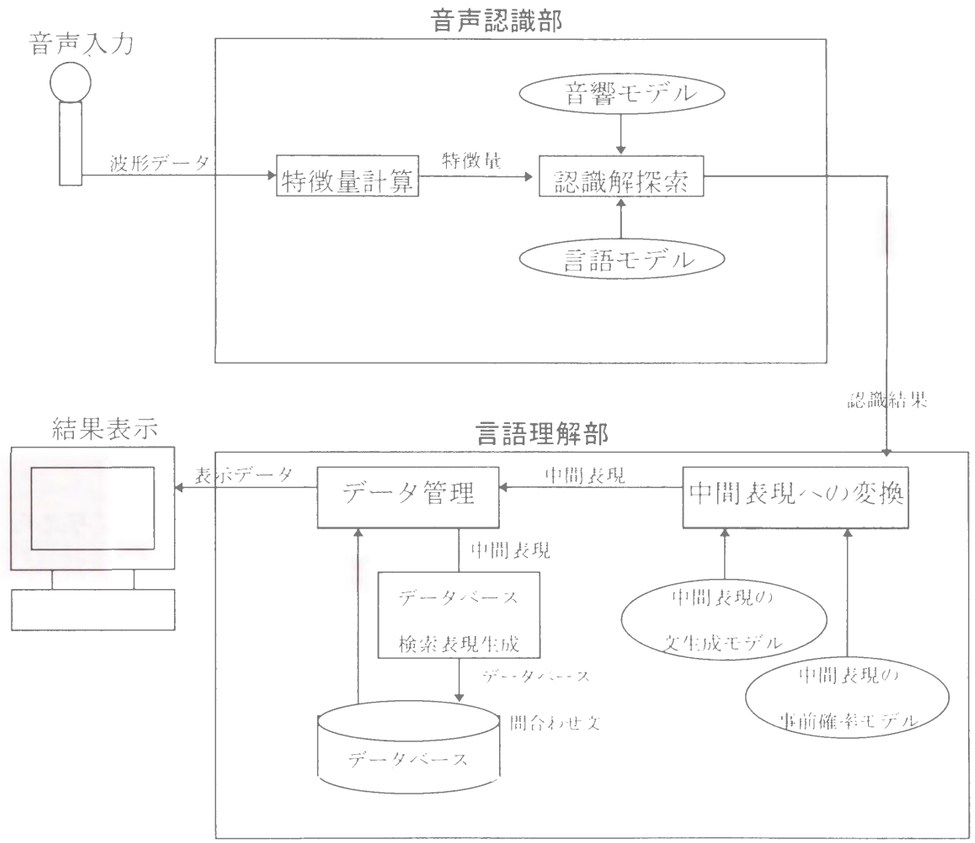
今回開発した音声理解フロトタイフシステムは，スキー場の案内を対象としている．これは，入力された音声を理解し，スキー場のデータが入力されたデータベースへアクセスし，ユーザーの要求する情報を表示するシステムで，次の 3 種類の動作を行うことができる．

- ユーザーが要求する条件を満たすスキー場の検索 (SHOWLIST)
- 各スキー場のデータ（県・標高差・リフト数等 12 項目）の表示 (SHOWVALUE)
- スキー場の地図の表示 (SHOWIMAGE)

システム全体の構成を図(6.1)に示す．本システムは，主に「音声認識部」と「言語理解部」からなる．

音声認識部では，入力された波形データに対し特徴量計算を行った後，隠れマルコフ網による音響モデル[34][35]，および複合 N-gram による言語モデル

[29]を用いて，単語グラフサーチ[7]により認識解の探索を行い，認識結果を出力する．言語理解部では，音声認識部で得られた認識結果の単語（形態素）列を中間表現に変換する．さらにデータベース検索用表現（SQL 問い合わせ文）を生成することによりデータベースにアクセスし，検索されたデータをユーザーの要求に応じて表示する．



図(6.1) 音声理解システム概要

6.2.2 言語理解部概要

本システムにおける言語理解部の目的は，認識結果をデータベース検索用表現に変換することである．しかし，データベース検索用表現は，文脈に直接関係のないデータベース言語特有のキーワード等が含まれるため，認識結果文から検索用表現に直接変換するのは効率的でない．このため，本システムでは，データベース言語特有のキーワードを除いた中間表現を用い，「入力文から中間表現への変換」，「中間表現からデータベース検索用表現への変換」と二段階の変換を行うことを考える．中間表現は，データベース検索用表現へ正確かつ容易に変換できるように設計されている．また，中間表現には，データベース検索の条件と共に，ユーザーの要求動作のタイプも含まれている．

本システムで用いた中間表現は次の要素から構成される．

- R_(コマンド名): 要求動作の指定 (Request)
- O_(対象物名): 動作の対象 (Object)
- D_(ドメイン名): データベースの検索項目 (Domain)
- C_(比較方法): データベース検索の比較条件 (Comparison)
- V_(値): データベース検索の値 (Value)

中間表現は，これらの要素の列として表現され，次のフォーマットで与えられる．

R_(コマンド名)　O_(対象物名)　D_(ドメイン名)　C_(比較方法)　V_(値)

以下に，自然言語から中間表現への変換例を挙げる．

例 1)
入力文：
「標高差が 1000m 以上のスキー場を教えて下さい.」
中間表現：
R_SHOWLIST　O_スキー場名　D_標高差　C_>=　V_1000

例 2)
入力文：
「八方尾根スキー場の標高差は何メートルですか.」
中間表現：
R_SHOWVALUE　O_標高差　D_スキー場名　C_=　V_八方尾根

例 3)
入力文：
「八方尾根のゲレンデマップを見せて下さい.」
中間表現：
R_SHOWIMAGE　O_ゲレンデ地図　D_スキー場名　C_=V_八方尾根

データベース検索用表現は SQL 言語のサブセットを用いている．中間表現からデータベース検索用表現への変換例を以下に示す．

中間表現：
R_SHOWLIST　O_ スキー場名　D_ 標高差　C_ >=　V_ 1000
データベース検索表現：
SELECT　スキー場名　FROM　スキー場データ
WHERE 標高差　>=　1000

中間表現の O_， D_， C_， V_ の各要素をデータベース表現の下線の部分に挿入するだけで機械的に変換が可能である．

言語理解部の一連の動作例を図(6.2) に示す．言語理解部は，認識結果が入力されると，次の順序で処理を行う．

1. 認識結果から中間表現への変換
2. 中間表現からデータベース検索用表現を生成
3. 条件に適合するデータをデータベースから検索し，対象の情報を獲得
4. 対象物名に対して中間表現のコマンド名で規定された動作を実行

入力文：「八方尾根スキー場の標高差を教えてください」

1. 中間表現への変換
   “R_SHOWVALUE　O_標高差 D_スキー場名　C_=　V_八方尾根”

2. データベース検索表現への変換
   “SELECT 標高差 FROM　スキー場データ
   WHERE　スキー場名　=　八方尾根”

スキー場データ

スキー場名	県	標高差	リフト数	
志賀高原	長野	500	27	
野沢温泉	長野	1100	29	
妙高赤倉	新潟	800	26	
八方尾根	長野	1000	34	
栂池高原	長野	700	25	

3. “スキー場名=八方尾根”を検索
4. 標高差を出力

図(6.2) 言語理解部の動作例

6.3 統計的処理による言語理解

前節で述べたように，本システムでの言語理解は，「入力文から中間表現への変換」，「中間表現からデータベース検索用表現への変換」と二段階の変換を行う．中間表現からデータベース検索用表現への変換は機械的に可能であるため，入力文から中間表現への変換が，本システムの言語理解における重要な役割を果たす．本章の研究の主眼点は，この「入力文から中間表現への変換」を統計的手法により実現する点にある．

入力文から中間表現への統計的手法による変換の原則は，形態素系列 W が与えられたとき，確率的に最も高い中間表現列 $\hat{S}$ を得ることである．これは，次式で表される．

$$\hat{S} = \arg\max_{S} P(S|W) \qquad (6.1)$$

これにベイズ則を用いて変形すると次式と等しくなる（2.2 節参照）．

$$\hat{S} = \arg\max_{S} P(W|S)P(S) \qquad (6.2)$$

本式において，$P(W|S)$ は，ある中間表現 S から入力文の形態素列 W を生成する確率である．この確率を求めるために，隠れマルコフモデル(Hidden Markov Model)を用いる．一方，$P(S)$ はある中間表現 S が出現する確率で，入力文とは独立に求められる事前確率である．この確率を求めるために，中間表現の各要素間の共起確率を用いる．それぞれのモデルに関しては，続く 6.3.1 および 6.3.2 で説明する．また，これらのモデルを用いて，入力文から中間表現を得る方法を 6.3.3 節に述べる．

6.3.1 隠れマルコフモデルによる文生成モデル

確率 $P(W|S)$，すなわち中間表現から入力形態素列の生成確率を与えるモデルとして，隠れマルコフモデル(Hidden Markov Model: 以下 HMM と略)を用いる．HMM は，複数の状態から構成され，形態素が入力される毎に，状態 i から状態 j へ確率 a_{ij} で遷移し，遷移後の状態 j から確率 $b_j(w_k)$ で形態素 w_k を出力するモデルである（2.4 節参照）．HMM を用いることにより，出力確率 $b_j(w_k)$ により，中間表現の各要素と形態素との共起関係を表すことができ，また，状態遷移確率 a_{ij} により，形態素の並びとして表される文の構造をも統計的に反映することができるため，単なるキーワードに基づく理解系に比べて，統計的により精度の高い理解が可能であると考えられる．

隠れマルコフモデルは，中間表現の各要素毎に作成し，文が入力されると，全てのモデル毎に独立に動作し，全てのモデルそれぞれが入力文の形態素全体

に対して生成確率を計算する．中間表現 Sから入力文 Wの生成確率は，次式のように6.2.2節で示した5要素それぞれの文生成確率の積として近似する．

$$P(W|S) = \prod_{t=1}^{5} P(W|s_t) \qquad (6.3)$$

但し，s_t は中間表現 S の t 番目の要素を表す

音声認識では通常 left-to-right 型，すなわち一方通行型のモデルが盛んに使用されている．しかし，言語理解のための隠れマルコフモデルとして，6.2.4節 図(2.1)のように，全ての状態間の遷移が可能なエルゴディック隠れマルコフモデルモデルを用いた．これは，自然発話の多彩な言い回しに対応可能な理解モデルを構築するため，構造をあらかじめ決定することは避け，理解に必要な言語構造の特徴を，モデルのパラメータ推定により自動的に獲得することを狙ったためである．

隠れマルコフモデルのパラメータの推定には，通常，最尤推定(Maximum Likelihood Estimation, 以下ML推定と略)が用いられるが，中間表現間の各要素の特徴をより明確に表現し，識別の精度を向上させるため，本研究では識別誤り最小化学習(Minimum Classification Error Training, 以下MCE学習と略)[45][46]を用いた．MCE学習は，正解と誤りとの距離を表す識別誤り関数を最小化するように行われる学習である．MCE学習を本モデルに適用する際，中間表現の要素のグループ(R_*,D_*,…)毎に行われ，正解の中間表現列に含まれる要素を正解，同一グループに属するの他の要素を誤りとして学習を行った．但し，MCE学習を行う際のHMMの初期パラメータは，ML推定による値を用いた．

6.3.2 要素の共起確率による中間表現の事前確率

中間表現は，各要素がランダムに出現するわけではなく，例えば，ユーザーの要求がスキー場のリストを表示する(R_SHOWLIST)場合，その対象は必ずスキー場名(O_スキー場名)になる等，中間表現の要素の共起関係が存在する．この共起関係を確率的に表現するため，中間表現の各要素間のBigramとして表される共起確率を用いた．この時，中間表現の事前確率 $P(S)$ を，次式によって求められる．

$$P(S) = P(s_1)\prod_{t=2}^{5} P(s_t|s_{t-1}) \qquad (6.4)$$

それぞれの確率は最尤推定により次式で容易に求められる．

$$\begin{aligned} P(s_1) &= N(s_1)/L \\ P(s_t|s_{t-1}) &= N(s_t, s_{t-1})/N(s_{t-1}) \end{aligned} \qquad (6.5)$$

但し，$N(\#)$は中間表現データ中の要素 '#' の出現回数を表し，L は学習データの文数を表す．

6.3.3 入力文から中間表現への変換アルゴリズム

入力文から中間表現への変換を行うためには，文生成確率 $P(W|S)$，および中間表現の事前確率 $P(S)$ の積 $P(W|S)P(S)$ の最大値を与える中間表現列を求めればよい．

図(6.3)に中間表現列を求める手法の概念図を示す．まず，文が入力されると，中間表現の各要素に対応する全てのHMMにおいて，$P(W|s_t)$ を計算する．次に，確率 $P(W|S)P(S)$ を計算する．これは，式(6.3)，および(6.4)より，次式のように求められる．

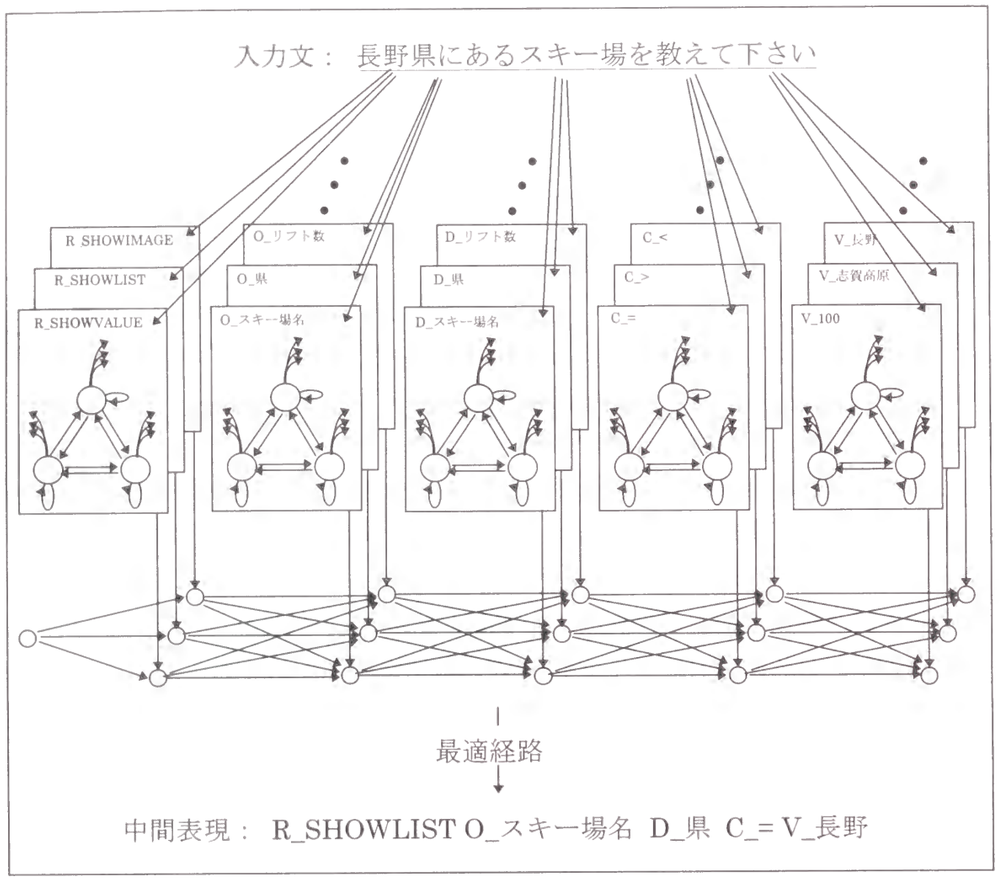
$$P(W|S)P(S) = \{\prod_{t=1}^{5} P(W|s_t)\}\{P(s_1)\prod_{t=2}^{5} P(s_t|s_{t-1})\} \qquad (6.6)$$

右辺を変形すると，次のようになる．

$$P(W|s_1)P(s_1)\prod_{t=2}^{5} P(W|s_t)P(s_t|s_{t-1}) \qquad (6.7)$$

本式を用いると，中間表現の各要素の順番(R_* , O_* , D_* , C_* , V_*)にHMMと共起確率との確率との積を計算し，それらの最大値を与える中間表現の要素列 s_t $(1 \leq t \leq 5)$ を求めることにより，入力文から中間表現への変換結果を得ることができる．本システムにおいて，中間表現の要素の組み合わせは，

3·14·14·6·108 = 381,024 通りあり，この中から最適解を求める必要があるが，本論文では，中間表現の要素の順序を固定し，要素間の共起確率をBigramの形式で表現したことにより，ビタビアルゴリズムで $P(W|S)P(S)$ の最適解を高速に求めることができる．



図(6.3) 入力文から中間表現への変換概念図

6.4 評価実験および考察

6.4.1 言語理解部の評価実験

提案手法による言語理解の性能評価実験を行った．言語理解部を単独に評価するため，まず，正解文からの言語理解率を評価した．実験に用いたデータは，スキー場案内システムのために収集している会話で，現在，2,700 文（34,098 形態素）あり，全ての文章に，それに対応する中間表現を人手で付与している．

この内，2,618 文（33,017 形態素）を言語理解モデルの学習に使用し，残りの 82 文（1,081 形態素）を評価用のデータとした．なお，中間表現の要素は 145 種類ある．

言語理解のためのモデルは，最尤推定(ML)による隠れマルコフモデル，および，識別誤り最小化(MCE)学習を行った 2 種類のモデルを用意した．隠れマルコフモデルの状態数は，全てのモデルで同一数とし，1 から 10 まで変化させて実験を行った．評価には言語理解率を用いた．但し，言語理解率は，入力文章から中間表現へ正確に変換できた割合であり，中間表現の全ての要素が正しく変換できた場合のみ正解とする．表(6.1)にこれらの条件での言語理解率(%)を示す．

隠れマルコフモデルの学習に ML 推定を用いた状態数が 1 の時で理解率は 82.9%であり，状態数を大きくしても理解率は向上せず，逆に低下する傾向にある．これに対して，MCE 学習を用いた場合は，状態数の増加に従って理解率が向上しており，状態数 6 の時に言語理解率は最大 91.5%まで向上している．これは，本来状態数が多くなるほど，モデルの自由度が増し，より精度の高い表現が可能であるが，MCE 学習を行うことにより，ML 推定では困難であった中間表現の要素の識別が効果的に行われた結果と考えられる．

状態数 1 の隠れマルコフモデルは語順を全く考慮することができず，状態数を複数にすることにより語順を考慮できるようになる．従って，状態数の増加に従って理解率が向上するという実験結果より，正確な理解のためには，出現形態素の種類だけでなく，語順すなわち文構造の把握が重要であり，提案モデルは隠れマルコフモデルの状態遷移という形で理解に必要な文構造の表現が自動獲得できたと考えられる．

表(6.1) 各種隠れマルコフモデルに対する言語理解率

状態数	1	2	3	4	5
ML	82.9	80.5	80.5	78.0	81.7
ML+MCE	82.9	87.8	87.8	90.2	90.2

6	7	8	9	10
79.2	79.2	79.2	78.0	80.5
91.5	90.2	89.0	89.0	90.2

6.4.2 音声理解システムの評価実験

言語理解部を音声認識部と接続し，音声理解システムとしての性能を評価した．

連続音声認識システム部の条件を表(6.2) に示す．なお，音声認識部の言語モデルの学習には，前節の言語理解モデルの学習に用いたデータと同一データを使用した．言語理解部へは，音声認識結果の尤度最大候補のみを用いて処理を行った．また，言語理解部の隠れマルコフモデルは，前節の実験で文理解率の最も高かった，6 状態の MCE 学習を行ったモデルを使用した．音声認識率，および音声理解率を表(6.3)に示す．

表(6.2) 音声認識実験条件

音響分析条件	
標本周波数	12kHz
窓関数	20ms ハミング窓
フレーム周期	10ms
パラメータ	16 次元 LPC ケプストラム＋Log パワー ＋16 次元ΔLPC ケプストラム+ΔLog パワー
音響モデル	
モデル	隠れマルコフモデル網[34][35]
状態数	803 状態
確率分布	5 混合 正規分布
言語モデル	
モデル	品詞と形態素列の複合 N-gram[29]
クラス数	54 品詞クラス＋300 分離クラス

表(6.3) 音声理解率

形態素 Accuracy	文認識率	音声理解率
91.4	59.8	73.2

表(6.3)より，音声認識部の文認識率は約 60%であるのに対して，音声理解率は約 73%と文認識よりも約 13%程度高い値を示している．すなわち，認識誤りを含む文から正しく理解された例が全体の約 13%あり，多少の認識誤りが発生しても正しい理解が得られることが確認できた．

誤認識文のうち，正しく理解できた文の代表例を以下に示す．

例 1)

正解文：

「妙高杉の原の上級コースの割合を教えて下さい.」

認識結果：

「妙高杉の原の上級コースが割合を教えて下さい.」

例 2)

正解文：

「ハチ北で最長コースは何メートルぐらいですか.」

認識結果：

「ハチ北てさ以上コースは何メートルぐらいですか.」

例１は，認識結果に助詞の置換誤りを生じており，文の構造解析は可能であるが，意味的をなさない文である．例２は，認識誤りの結果文法的に解析不可能な文となっている．これらの場合は，文法ルールを用いた構造解析では理解が困難であると考えられる．しかし，提案手法では，確率的に最も高い中間表現列を理解結果として選択するため，多少の誤りが生じ，意味的および文法的に理解不能な文が入力されても，正しく理解できる場合があることが分かった．提案手法において，例２のような誤りが中間表現に正しく変換できた理由として，以下の２つが考えられる．

- 中間表現の各要素が文全体から判断するため，「コース」「何メートル」等の表現から，「最長」という形態素が認識されなくても「最長コース」を意味する隠れマルコフモデルの確率は高くなる．
- 中間表現の要素間の共起確率を用いたことにより，その他の要素との共起関係で高い確率が得られた．

表(6.4) 音声認識誤りの種類と理解率の関係

誤りの種類	認識誤り数	理解誤り数	理解正解数
助詞誤り	10	1	9
重要キーワード誤り	16	16	0
その他の誤り	7	4	3

次に，認識誤りを生じた 33 文の理解状況について分析を行った．誤りの種類としては，例 1 のような助詞の置換，脱落に関する誤り，スキー場名，数値等の重要キーワードの認識誤り，それ以外の誤りに分類し，それぞれの誤り分類に関して，理解正解数，誤り数を調べた．結果を表(6.4) に示す．

表(6.4)より，助詞誤りについては，認識誤りを生じた 10 文の内 9 文が正しく理解できており，助詞誤りに関しては，非常にロバストな理解が可能であることが分かった．その他の誤りは，例 2 の「で最長」→「てさ以上」や，「割合」→「ありあり」等，重要語ではない名詞等に誤りを生じた例であるが，この種の誤りの場合でも 7 文中 3 文と半数近くが理解に成功しており，本論文で提案した手法の理解誤り文に対する頑健性が実証できた．

音声認識誤りに対するロバストな理解として，文献[47]ではヒューリスティックスによる助詞落ち・誤りの訂正，キーワード抽出による意味解析等を用いる手法が提案されており，30.3%の文認識率に対して 54.5%の文が正しく理解されたとされている．これを，認識誤りに文における理解正解率で本論文と比較する．本論文では表 4 より認識誤り 33 文の内 12 文が正しくされていることより認識誤りに文における理解正解率は 36.4%である．文献[47]では全体の 69.7%(100%-30.3%)が認識誤り文で認識誤り文から正しく理解できた割合が全体の 25.9%であるため認識誤りに文における理解正解率は 37.2%となり，両手法はほぼ同じ程度の値となる．また，文献[47]では，助詞落ち・誤りに関しては，90%程度の訂正が可能とされており，本論文と同程度のレベルである．また，語彙サイズは本論文が 288 語に対し，文献[47]では 241 語であり同程度の規模のシステムであると考えられる．従って，認識誤りに対する理解のロバスト性は，両者ほぼ同等の能力であると考えられる．しかし，文献[47]のヒューリスティックスやキーワード抽出方法では，タスクが大規模になるとルールの作成やキーワードの選択が困難になると考えられる．これに対して，本論文の手法は統計量による方法であるためデータから自動学習できるため，システム構築の際には本論文の手法の方が有利であると考えられる．また，文献[47]とは異なり，本論文では誤り訂正のための特別な手法を用いていないにも関わらず同程度のロバスト性があるため，誤り訂正の手法とを組み合わせることにより，音声理解の精度をさらに向上させることもできると考えられる．

表(6.4)より，重要キーワードの認識誤りは全誤り 33 文中 16 文存在と約半数を占め，当然ながらいずれの文も正しく理解されなかった．このため，音声理解率向上のためには，認識結果の複数の候補から理解を行う等の手法が必要であると考えられる．

6.5 あとがき

自然発話文および音声認識の誤り文の正確な理解を目的として，隠れマルコフモデルおよび共起確率という統計処理により自然言語から中間表現への変換を行うことを特徴とする音声言語理解手法を示した．実験の結果，隠れマルコフモデルに識別誤り最小化学習を行うことにより，最高 91.5%の言語理解率が得られ，また，音声認識システムと結合した際の音声理解率は 73.2%であり，統計量のみによる手法としては比較的に高い言語理解率が得られた．実験結果の考察・解析により，正確な理解のための文構造を自動獲得することができると考えられ，また，確率的に理解を行うことにより，認識誤りにより意味的および文法的に解析不能な文に対しても，正しい理解が得られる場合があることを確認した．

提案した言語理解手法は統計的手法を用いているため，より自然な発話の理解も可能であると考えられる．また，今回開発した程度の規模のシステムでは，文法ルールの作成やキーワードの選択等の作業を必要とせず，数千文程度のデータを自動学習して比較的良好な結果が得られるため，比較的短時間でのシステム構築が可能である．

今回提案した手法は，比較的簡単な条件での言語理解を対象とした方法であったが，今後はさらに複雑な条件を扱えるようアルゴリズムの改良を行い，音声翻訳システム等，さまざまな音声理解システムへ応用してゆきたい．

第７章 結論

本論文は，音声認識・理解における統計的言語処理に関する研究成果をまとめた．研究の内容は，主に言語モデル N-gram の精度向上のための研究，および統計的モデルによる音声理解手法の 2 点である．

精度の高い N-gram 言語を構築するために重要な点は次の 2 点である．

1. 大量で質の高い言語コーパスをいかに整備するか
2. 少量の言語コーパスから，いかに精度の高い言語モデルを構築するか

1.の問題を解決するため，第 3 章で少量のデータから精度の効果的な言語表現を得ることのできる品詞と可変長形態素列の複合 N-gram を提案し，少量のデータで学習した複合 N-gram を用いて形態素解析を行うことにより，大量で質の高い言語コーパスを比較的容易に整備する方法を示した．

2.の問題を解決するため，第 4 章では品詞と可変長形態素列の複合 N-gram を音声認識に適用し，比較的少量のデータから構築した言語モデルでも比較的高い認識率が得られ，また従来の形態素 N-gram よりも高い認識率が得られることを確認した．また，第 5 章では，目的のタスクの言語データ量が極めて少ない場合，他のタスクのデータから得られる N-gram を最大事後確率推定により目的のタスクの言語特徴に適応させることにより，言語モデルの精度を向上させるタスク適応手法を提案した．実験の不特定タスク N-gram や特定タスク N-gram よりも認識の精度を向上させることができた．

第 6 章では，自然発話の理解，および音声認識の誤りにロバストな言語理解を目的として，隠れマルコフモデルという統計的モデルを用いた言語理解手法を提案した．実験の結果，言語理解率 90.2%という比較的高い値を得ることができた．また，連続音声認識と結合させた音声理解率では，文認識率 59.8%に対し音声理解率が 73.2%と，音声理解率が音声認識率よりも約 13%向上し，認識誤りに対する理解のロバスト性を示すことができた．

以上，本論文では統計的手法による音声言語理解の研究成果を示した．これらの方法を用いれば音声認識・理解の精度を従来のものより向上させることができる．しかし，これで完全な言語モデルができたわけではない．最後に，今後の課題・展望を述べる．

（N-gram に関して）

N-gram はパラメータ数が膨大であるため，限られた量の言語データでは正確な確率が求まらないという大きな問題がある．大量のデータを用いて，補間（平滑化）等の処理を行っても，4-gram 程度が性能的に限界であるとされている．従って，たかだか 4 単語(形態素)間の局所的な関係しか表現できない．これに対して，従来の生成文法等では，文全体の関係を表現できる．しかし，言語現象は非常に例外の多い現象であるから，言語の全てを表現するためには，莫大な数のルール，しかも矛盾のないルールを作成する必要があり，これは極めて困難な作業であると考えざるを得ない．今後は，統計モデルと文法ルールとの両方の長所を生かし，大局的かつ精密な表現ができるような手法の開発が重要であると考えている．

（音声理解に関して）

本論文では，隠れマルコフモデルを用いた言語理解手法を提案したが，非常に小さいタスク内での有効性を確認したにとどまっており，さらには従来の構文解析等の手法と直接比較を行い有効性を確認したわけではない．音声理解は，音声認識ほど盛んに研究されているわけではなく，音声認識のHMM＋N-gram のように確立された手法は存在していない．逆に言えば，まだ精度改善の余地は多いにあると考えられる．現在の主流である構文解析による手法や本論文のような単純な統計的モデルを用いた手法で，人間と同程度の理解が得

られるとは考えにくい．今後は，さまざまな手法を試行錯誤しながら，精度の改善を試みる必要があると思う．筆者の個人的な考えだが，真に人にやさしいマンマシンインターフェースのためには，音声の理解は必須であると考える．もっと，音声理解に対する研究が盛んになることが望まれる．

謝辞

本論文をまとめるに当たり，懇切に御指導，御教授を賜りました京都大学大学院情報学研究科知能情報学専攻堂下修司教授に心から感謝の意を表します．

本論文の作成に際し，懇切にご指導を賜りました京都大学大学院情報学研究科知能情報学専攻河原達也助教授に心から感謝の意を表します．

本研究の主要な部分はATR音声翻訳通信研究所にて行ったもので，在職中に直接御指導，御教示を賜りました第一研究室匂坂芳典室長に深謝致します．研究の機会を与えていただいたATR音声翻訳通信研究所山崎泰弘前社長（現在KDD研究所），および山本誠一社長には深く感謝致します．また，日頃有益な御意見を頂きました，松永昭一主任研究員（現在NTTヒューマンインターフェース研究所），中村篤主任研究員ならびに研究室の皆様に感謝します．さらに，共同研究を通じて御討論頂いた京都大学工学部情報工学科の久木和也氏（現在日立製作所システム開発研究所）に感謝します．

本論文を作成するにあたり，さまざまな面でサポートしていただいた住友金属工業株式会社システム研究開発部の新井三鉉グループ長，および中村雅巳博士に感謝致します．

最後に，研究活動全般を支え，協力していただいた多くの方々にこの場を借りてお礼申し上げます．

参考文献

[1] Proc. ARPA speech recognition workshop, Morgan Kaufmann Publishers, 1996.

[2] P.C.Woodland et.al, "The 1994 HTK large vocabulary speech recognition system", Proc. ICASSP95', Vol.1, pp.73-76, 1995.

[3] L.R.Bahl et.al: "Performance of the IBM large vocabulary continuous speech recognition system on the ARPA Wall Street Journal task", Proc. ICASSP95', Vol.1, pp.41-44, May 1995.

[4] R.Pieraccini and E.Levin: "Stochastic representation of semantic structure for speech understanding", Proc. EUROSPEECH-91, Vol.2, pp.383-386, 1991.

[5] 松岡 達雄, 大附 克年, 森 岳至, 古井 貞熙, 白井 克彦: "新聞記事データベースを用いた大語い連続音声認識", 信学論 Vol.J79-D-II No.12, pp.2125-2131, December 1996.

[6] 西村 雅史, 伊東 伸泰: "単語を単位とした日本語ディクテーションシステム", 信学論 Vol.J81-D-II No.1, pp.10-17, January 1998.

[7] 清水 徹, 由本 博史, 政瀧 浩和, 松永 昭一, 匂坂 芳典: "大語い連続音声認識のための単語仮説数削減", 信学論 Vol.J79-D-II, No.12, pp.2117-2124, December 1996.

[8] 竹沢 寿幸, 森元 逞, 匂坂 芳典, ニック キャンベル, 飯田 仁: "日英音声翻訳システム ATR MATRIX", 第56回情処全大, 6Q-07, March 1998.

[9] L.R.Bahl, F.Jelinek, and R.L.Mercer: "A maximum likelihood approach to continuous speech recognition", IEEE Trans. on Pattern Analysis and Machine Intelligence, pp.179-190, 1983.

[10] 永井 明人, 鷹見 淳一, 峯崎山 茂樹: "逐次状態分割法(SSS)と音素コンテキスト依存LRパーザを統合したSSS-LR連続音声認識システム", 信学技報, SP92-33, 1992.

[11] 河原 達也, 北岡 教英, 堂下 修司: "A*探索に基づいたフレーズスポッティングによる頑健な音声理解", 信学論, Vol. J79-DII, No.7, pp.1187-1194, 1996.

[12] F.Jelinek and R.L.Mercer: "Interpolated estimation of Markov source parameters from sparse data", Proc.Workshop Pattern Recognition in Practice, pp.381-397, 1980.

[13] S.M.Katz: "Estimation of probabilities from sparse data for the language model component of a speech recognizer", IEEE Trans. on Acoustics, Speech, and Signal Processing, pp.400-401, 1987.

[14] I.J.Good: "The population frequencies of species and the estimation of population parameters", Biometrika, Vol.40, no.3and4, pp.237-264, 1953.

[15] P.Placeway, R.Schwartz, P.Fung and L.Nguyen: "The estimation of powerful language models from small and large corpora", Proc. ICASSP-93, Vol.II, pp.33-36, 1993.

[16] R.Kneser and H.Ney: "Improved backing-off for m-gram language modeling", Proc. ICASSP-95, Vol.1, pp.181-184, May 1995.

[17] 川端 豪, 田本 真詞: "二項事後分布に基づく N-gram 言語モデルの Back-off 平滑化", 信学技報 SP95-93, pp.1-6, December 1995.

[18] P.F.Brown et al.: "Class-Based n-gram models of natural language", Computational Linguistics, Vol.18, No.4, pp.467-479, 1992.

[19] 田本 真詞, 川端 豪: "連接共起に注目した単語のクラスタリング", 信学技報 SP93-125, pp.55-62, January 1994.

[20] M.Nagata: "A stochastic Japanese morphological analyzer using a forward-DP backward A* N-best search algorithm", COLING-94, pp.201-207, 1994.

[21] E.P.Giachin: "Phrase bigrams for continuous speech recognition", Proc. ICASSP-95, Vol.1, pp.225-227, May 1995.

[22] S.Deligne and F.Bimbot: "Language modeling by variable length sequences. theoretical formulation and evaluation of multigrams", Proc. ICASSP-95, Vol.1, pp.169-172, May 1995.

[23] 伊藤 彰則, 好田 正紀: "かな・漢字文字列の連鎖統計による言語モデル",信学論 Vol.J79-D-II No.12, pp.2062-2069, December 1996.

[24] L.E.Baum and J.A.Eagon: "An inequality with applications to statistical prediction for functions of Markov processes and to a model frecology", Bull. Am. Math. Soc., 73, 1967.

[25] S.Matsunaga, T.Yamada and K.Shikano: "Task adaptation in stochastic language models for continuous speech recognition", Proc. ICASSP'92, Vol.1, pp.165-168, 1992.

[26] 山田 智一, 川端 豪, 松永 昭一, 鹿野 清宏: "かな・漢字の文字列連鎖情報を利用した統計的言語モデル", 信学技報 SP91-26, pp.65-72, June, 1991.

[27] 山田 智一, 川端 豪, 松永 昭一, 鹿野 清宏: "音声認識におけるカナ・漢字連鎖確率に基づく統計的言語モデル", 信学論 (A), Vol.J77-A, no.2, pp.198-205, 1994.

[28] 高木 幸一, 古井 貞熙: "形態素の読みの確率を考慮したニュース音声のディクテーション", 音響講論 1-6-5, pp.9-10, March 1998.

[29] 政瀧 浩和, 松永 昭一, 匂坂 芳典: "連続音声認識のための可変長連鎖統計言語モデル", 信学技報, SP95-73, pp.1-6, 平成 7 年

[30] T.Morimoto et al.: "A speech and language database for speech translation research", ICSLP, pp.1791-1794, September 1994.

[31] 渕 武志, 松岡 浩司, 高木 伸一郎: "保守性を考慮した日本語形態素解析システム", (http://lambda.cipl.cae.ntt.co.jp/jtag/).

[32] 黒橋 禎夫, 長尾 真: "日本語形態素解析システム JUMAN version 3.5", (http://www-lab25.kuee.kyoto-u.ac.jp/nl-resource/juman.html)

[33].Wakita, J.Kawai and H.Iida: "An evaluation of statistical language modeling for speech recognition using a mixed category of both words and parts-of-speech", ICSLP, pp.530-533, September 1996.

[34] 鷹見 淳一, 嵯峨山 茂樹: "逐次状態分割法による隠れマルコフモデル網の自動生成", 信学論 Vol.J76-D-II, No.10, pp.2155-2164, October 1993.

[35] M.Ostendorf and H. Singer: "HMM Topology design using maximum likelihood successive state splitting", Computer Speech and Language, Vol.11, pp.17-41, 1997.

[36] 小坂 哲夫, 鷹見 淳一, 嵯峨山 茂樹: "話者混合 SSS による不特定話者音声認識", 日本音響学会講演論文集 2-5-9, pp.135-136, October 1992.

[37] R.Kneser, and V.Steinbiss: "On the Dynamic Adaptation of Stochastic Language Models", Proc. ICASSP'93, Vol.2, pp.585-588, 1993.

[38] P.Clarkson and A.Robinson: "Language Model Adaptation using Mixtures and an Exponentially Decaying Cache", Proc. ICASSP'97, Vol.2, pp.799-802, 1997.

[39] 中川 聖一, 伊田 正樹: "連続音声認識のタスクの複雑さを表す新しい尺度", 信学論 Vol.J81-D-II, No.7, pp.1491-1500, July 1996.

[40] 伊藤 彰則, 好田 正紀: "N-gram タスク適応の認識実験による評価", 音響講論 1-6-20, pp.43-44, March 1998.

[41] 坂井 信輔, 畑崎 香一郎, 水野 正典, 渡辺 隆夫: "音声入力を用いたパソコンネットワーク旅客機空席案内システムの試作", 信学技報, SP94-89, pp.29-36, January 1995.

[42] 内藤 正樹, 黒岩 眞吾, 武田 一哉, 山本 誠一, 谷戸 文廣: "大規模内線電話受付システムの試作", 信学技報, SP94-90, pp.37-42, January 1995.

[43] 遠藤 充, 伊藤 達朗, 星見 昌克, 二矢田 勝行: "音声による文例検索方法の検討", 音響講論 2-Q-12, pp.163-164, March 1997.

[44] S.Seneff: "TINA: A Natural Language System for Spoken Language Applications", Computational Linguistics, Vol.18, No.1, March 1992.

[45] W.Shou, C.H.Lee and B.H.Juang: "Minimum error rate training based on N-best string models", ICASSP93, pp.652-655, 1993.

[46] 村上 哲範, 武田 一哉, 河井 恒, 山本 誠一: "行列によるトレリス計算を用いたHMM の文レベルでの識別学習", 信学技報, SP95-25, pp.1-6, July 1995.

[47] 山本 幹雄, 伊藤 敏彦, 肥田野 勝, 中川 聖一: "人間の理解手法を用いたロバストな音声対話システム", 情処論, Vol.37, N0.4, pp.471-482, April 1996.

関連発表論文

〔主論文〕

1. 政瀧 浩和，松永 昭一，匂坂 芳典："品詞と可変長単語列の複合 N-gram の自動生成"，電子情報通信学会論文誌，Vol.J81-D-II，No.9，pp.1929-1936（1998 年 9 月）．

2. 政瀧 浩和，匂坂 芳典，久木 和也，河原 達也："最大事後確率推定による N-gram 言語モデルのタスク適応"，電子情報通信学会論文誌，Vol.J81-D-II，No.11，pp.2519-2525（1998 年 11 月）．

3. 政瀧 浩和，匂坂 芳典："品詞と可変長形態素列の複合 N-gram を用いた日本語形態素解析"，自然言語処理，Vol.6，pp.41-57（1999 年 1 月）．

4. 政瀧 浩和，谷垣 宏一，匂坂 芳典："統計的モデルによる音声言語理解"，電子情報通信学会論文誌，D-II（1999 年 2 月 掲載予定）．

〔国際学会報告〕

1. H.Masataki and Y.Sagisaka: "Variable-Order N-gram Generation by Word-Class Splitting and Consecutive Word Grouping", Proc. of ICASSP 96, pp.188-191 (May, 1996)

2. H.Masataki, Y.Sagisaka, K.Hisaki and T.Kawahara: "Task Adaptation Using MAP Estimation in N-gram Language Modeling", Proc. of ICASSP 97, pp.783-786 (April, 1997)

3. T.Shimizu, H.Yamamoto, H.Masataki, S.Matsunaga and Y.Sagisaka: "Spontaneous Dialogue Speech Recognition Using Cross-Word Context Constrained Word Graphs", Proc. of ICASSP 96, pp. 145-148 (May, 1996)

〔研究会資料〕

1. 政瀧 浩和，松永 昭一，匂坂 芳典："連続音声のための可変長連鎖統計言語モデル"，電子情報通信学会技術報告，SP95-73，pp.1-6（1995 年 11 月）．

2. 政瀧 浩和，匂坂 芳典，久木 和也，河原 達也："MAP 推定を用いた N-gram 言語モデルのタスク適応"，電子情報通信学会技術報告，SP96-103，pp.59-64（1997 年 1 月）．

3. 政瀧 浩和，谷垣 宏一，匂坂 芳典："統計的処理による音声・言語理解モデル"，電子情報通信学会技術報告，SP97-98，pp.25-32（1998 年 1 月）．

〔講演報告〕

1. 政瀧 浩和，松永 昭一，匂坂 芳典："連続音声認識のための品詞・単語可変長 N-gram"，日本音響学会平成８年度春季研究発表会講演論文集，1-P-17，pp.195-196（1996 年 3 月）．

2. 政瀧 浩和，匂坂 芳典，久木 和也，河原 達也："MAP 推定による N-gram 言語モデルの適応"，日本音響学会平成９年度春季研究発表会講演論文集，1-6-3，pp.5-6（1997 年 3 月）．

3. 政瀧 浩和，谷垣 宏一，匂坂 芳典: “統計的手法による認識結果から中間表現への変換を用いた音声理解システム”，日本音響学会平成１０年度春季研究発表会講演論文集，1-6-7，pp. 13-14（1998 年 3 月）.

4. 政瀧 浩和: “MAP 推定に基づく N-gram 言語モデルの自動分類されたコーパスへの適応”, 日本音響学会平成１０年度春季研究発表会講演論文集，1-6-19，pp. 41-42（1998 年 3 月）.

5. 清水 徹，山本 博史，政瀧 浩和，松永 昭一，匂坂 芳典: “単語グラフと可変長 N-gram を用いた大語彙自然発話音声認識”，日本音響学会平成８年度春季研究発表会講演論文集，1-P-18，pp. 197-198（1996 年 3 月）.

6. 谷垣 宏一，政瀧 浩和，匂坂 芳典: “決定木を用いた発話の意味タグ推定”，日本音響学会平成１０年度春季研究発表会講演論文集，1-6-2，pp. 3-4（1998 年 3 月）.

7. 内藤 正樹，政瀧 浩和，Harald Singer，塚田 元，匂坂 芳典: “日英音声翻訳システム ATR-MATRIX における音声認識用音響・言語モデル”，日本音響学会平成１０年度春季研究発表会講演論文集，2-Q-20，pp. 159-160（1998 年 3 月）.

付録）パープレキシティの算出方法

パープレキシティは，言語モデル性能の評価基準として用いられる値である．以下にパープレキシティの算出方法を示す．

言語は，形態素(単語) 列を生成する情報源であると考えられる．言語 L において，言語モデル M の形態素列 $W\ (=w_1 \cdots w_n)$ に対する生成確率を $P(w_1 \cdots w_n)$ とすれば，言語モデル M の言語 L におけるエントロピー $H_0(L)$ は次式により計算される(以下の式において，対数の底は 2 である).

$$H_0(L) = -\sum_{w_1 \cdots w_n} P(w_1 \cdots w_n) \log P(w_1 \cdots w_n) \qquad (8.1)$$

1 形態素当たりのエントロピーは次のようになる．

$$H(L) = \frac{1}{n} \sum_{w_1 \cdots w_n} P(w_1 \cdots w_n) \log P(w_1 \cdots w_n) \qquad (8.2)$$

この値は，十分長いテキスト W を用いて，

$$H(L) = -\frac{1}{n} \log P(w_1 \cdots w_n) \qquad (8.3)$$

として計算される．$H(L)$ は言語 L から生成される形態素を特定するために必要な情報量(ビット）であり，ある形態素の後には平均して $2^{H(L)}$ 個の形態素が後続可能であることを示している．すなわち，

$$PP = 2^{H(L)} \qquad (8.4)$$

は情報理論的な意味での形態素の平均分岐数を表しており，PP はパープレキシティ(perplexity)と呼ばれる．言語のパープレキシティが小さいほど，形態素の分岐数が少なく，形態素を特定するのが容易であり，言語モデルの性能が優れていることを表す.

自然言語は複雑であり，大量のテキストデータを用いて言語モデルを学習しても，言語の全ての特徴をモデル化できるわけではない．このため，言語モデルの評価には，テストセットパープレキシティと呼ばれる評価値が一般に用いられる．テストセットパープレキシティは，言語モデルの学習に使用したデータとは別に評価用のデータを用意し，評価データに対して上記の(8.3)および(8.4)式により計算されたパープレキシティである．一般に，パープレキシティ

により言語モデルの評価を行う場合は，テストセットパープレキシティを用いる．

実際にハーフレキシティを求める際には，次の a, b の２点を考慮する必要がある．

a. 未知語の扱い

テストセットパープレキシティを求める場合，評価データには学習データに出現しない形態素，いわゆる「未知語」が出現する場合がある．未知語は，辞書に登録されていない「未登録語」と，辞書には登録されているが学習データには出現しない「未学習語」とに分類できる．本論文では，学習データおよび評価データに出現する形態素全てをあらかじめ辞書に登録し，評価データに出現する未知語は全て「未学習語」として扱っている．しかし，未学習語への遷移確率は 0 となり，エントロピーおよびパープレキシティは無限大となり，言語モデルの評価ができない．未知語は，それ自体が研究テーマになるほどの困難な問題であるが，本論文では，以下に示す単純な手法を用いて未知語への遷移確率を計算し，パープレキシティを求めている．

辞書に登録されている語彙のサイズを Vとしたとき，未知語への遷移確率 p_0 を以下の式で与える．

$$p_0 = \frac{1}{KV} \qquad (8.5)$$

Kは定数で，本論文ではいずれの実験においても $K=100$ としている．しかし，未知語への遷移確率を与えることにより，遷移確率の和が 1 よりも大きくなる．このため，既知語に関しては，元の遷移確率を p とした時，

$$p' = p \cdot (1 - \frac{1}{K}) + \frac{1}{KV} \qquad (8.6)$$

と補正し，遷移確率を 1 に正規化している．なお，言語モデルを連続音声認識に適用する際も，(8.5)および(8.6)式による遷移確率を用いている．

b. 形態素列 N-gram の扱い

本論文の第 3 章および第 4 章で使用している「品詞と可変長形態素列の複合 N-gram」では，特定の形態素列を結合し，言語モデルのための新たな単位として扱うが，この場合のパープレキシティの算出方法を示す．

例として 4 形態素からなる文 $W(=w1,w2,w3,w4)$ を考える．形態素 Bigram の場合，文 W の遷移確率は，

$$P(W) = P(w_1) \cdot P(w_2 \mid w_1) \cdot P(w_3 \mid w_2) \cdot P(w_4 \mid w_3) \qquad (8.7)$$

のように，４形態素間の遷移確率の積として表される．一方，複合 Bigram において，例えば w_2 と w_3 を結合させた場合，文 W の生成確率は，

$$P(W) = P(w_1) \cdot P(w_2, w_3 \mid w_1) \cdot P(w_4 \mid w_2, w_3) \qquad (8.8)$$

のように，3 形態素列間の遷移確率の積として求められる．しかし，この形態素列の単位はあくまでも言語モデルでの単位であり，エントロピー $H(L)$ は，1 形態素当たりのエントロピーであるから，元の形態素数 4 を用い，

$$H(L) = -\frac{1}{4}\{\log P(w_1) + \log P(w_2, w_3 \mid w_1) + \log P(w_4 \mid w_2, w_3)\} \qquad (8.9)$$

として計算する．