Internal Use Only (非公開)

TR-I-0133

NETgramを用いたHMM英単語音声認識の改善

Improvement of HMM-based English Word Recognition using NETgram

丸山 活輝　中村 雅己　川端 豪†　鹿野 清宏‡

Katsuteru MARUYAMA, Masami NAKAMURA,
Takeshi KAWABATA and Kiyohiro SHIKANO

1990.2

概要

自動翻訳電話の要素技術である英語音声認識を実現するために、HMMを用いて英単語の認識実験を行ってきたが、音響的特徴だけでは認識が困難な場合があり、認識率を向上させるためには言語情報の利用が必要であると考えられる。文章中の単語の音声認識結果の誤りを訂正するために、ニューラルネットワークモデル(NETgram)による品詞情報を用いて単語列を予測を行い、HMM英単語認識結果の改善を行った。また、従来のTrigram確率モデルでの実験を行い、学習効率などの性能を比較した。単語予測モデルは、Bigram, Trigramの2種類の単語組数のモデルを用いた。その結果、学習データが十分ある場合には、確率モデルを用いた方が効果があるが、学習データが少ない場合には、NETgramの方が有効であることがわかった。

ATR Interpreting Telephony Research Laboratories
ATR 自動翻訳電話研究所
©ATR Interpreting Telephony Research Laboratories
©ATR 自動翻訳電話研究所
†現在、NTT基礎研究所勤務
‡現在、NTTヒューマンインターフェース研究所勤務

目次

1. はじめに 1

2. HMM音韻連結学習によるHMM英単語認識 2

3. NETgramによる単語予測 5

3.1 NETgram 5

3.2 N-gram への拡張 7

3.3 NETgramの学習 9

4. NETgramを用いたHMM英単語音声認識の改善 12

5. 実験結果 15

6. まとめ 19

参考文献 20

付録-1 HMM英単語認識に用いたデータの例 21

付録-2 Brown Corpus Text Database 22

付録-3 Brown Corpus における品詞の分類 23

1.はじめに

自動翻訳電話の要素技術である英語音声認識を実現するために、Hidden Markov Modelを用いて英単語の認識実験を行ってきたが[1]、音響的特徴だけでは認識が困難な場合があり、認識率を向上させるためには言語情報の利用が必要であると考えられる。

文章中の単語の音声認識結果の誤りを訂正する方法として、次に続く単語の出現確率を用いて単語列を予測する方法が知られている[2][3]。しかし確率情報を用いる従来の方法(Trigram確率モデル)では、有効な統計量を獲得するために膨大なデータを必要とし、また、未知の単語列に対しては予測が困難であるという問題があった。そこで、これらの問題を解決するためにニューラルネットワークモデル(NETgram)が提案された[4][5]。

本報告では、NETgramによる品詞情報を用いて単語列を予測することによりHMM英単語認識結果の改善を行ったので報告する。また、従来のTrigram確率モデルでの実験を行い、学習効率などの性能を比較した。単語予測モデルは、Bigram, Trigramの2種類の単語組数のモデルを用いた。

2.HMM音韻連結学習によるHMM英単語認識

音響処理としては、HMM音韻連結学習[1]による英単語認識手法を用いる。英語の音韻のラベル付けは日本語に比べてはるかに困難であるため、ラベリングされていない音声データを用いて実験を行っている。従って、単語内の時間情報がなく、音韻を正確に切り出すことはできないため、音韻単位での学習はできない。そこで学習時に学習単語に対応する音韻モデルを結合、分解しながらforward-backwardアルゴリズム[6]によりラベルなしで学習を行う方法を採用する。

図1にHMM音韻連結学習アルゴリズムを示す。学習の対象となる音韻モデルは、無音モデル1個を含む57個である。まず、単語の発音記号を記載した単語-音韻変換テーブル用いて、学習用単語の音韻列に従い単語モデルをつくる。この際、音韻モデルを連結したモデルの両端に無音モデルをつけ音声始終端も自動的に決定できるようにする。次に学習用単語モデルに対し、forward-backwardアルゴリズムによりHMMパラメータの学習を行う。単語モデルで学習した後、もと通り音韻モデルに分解する。

図2に認識アルゴリズムを示す。認識実験は、学習された音韻モデルを連結して辞書単語(542語)の単語モデルを作り、評価用単語に対する生起確率を求める。本実験では、辞書単語中で確率の大きいものから上位10位をHMM認識結果として用いた。

HMMによる認識実験では、セパレートベクトル量子化手法を用いた。セパレートベクトル量子化は、スペクトル、パワー等の特性の異なる幾つかの音声の特徴パラメータに対し複数個のコードブックを作成し、それらのコードラベルの組合せで入力ベクトルを表現する方法である。この手法はKai-Fu Lee[7]らにより提案され認識能力の向上が報告されている。また、ATRでも日本語の音韻認識において有効性が確認されている[8]。本実験では以下の3種類の特徴パラメータを用いている。

・スペクトル(WLR)
・スペクトルの時間変化(DCEP)[9]
・パワー(POW)

音声データとしては、特別の発声訓練を受けていない米国人男性1人により発声された英単語2834語(542種類)を使用する。発声内容は、国際会議の問合せや、ホテルの予約等の電話英会話文377文を単語に区切って発声したものである[10]。これらのデータにはラベリングは施されていない。これらの音声データ中の1487語(190文)でHMMの学習を行い、1347語(187文)を評価データとして単語認識を行った。付録1に実験データの会話文の1例を示す。

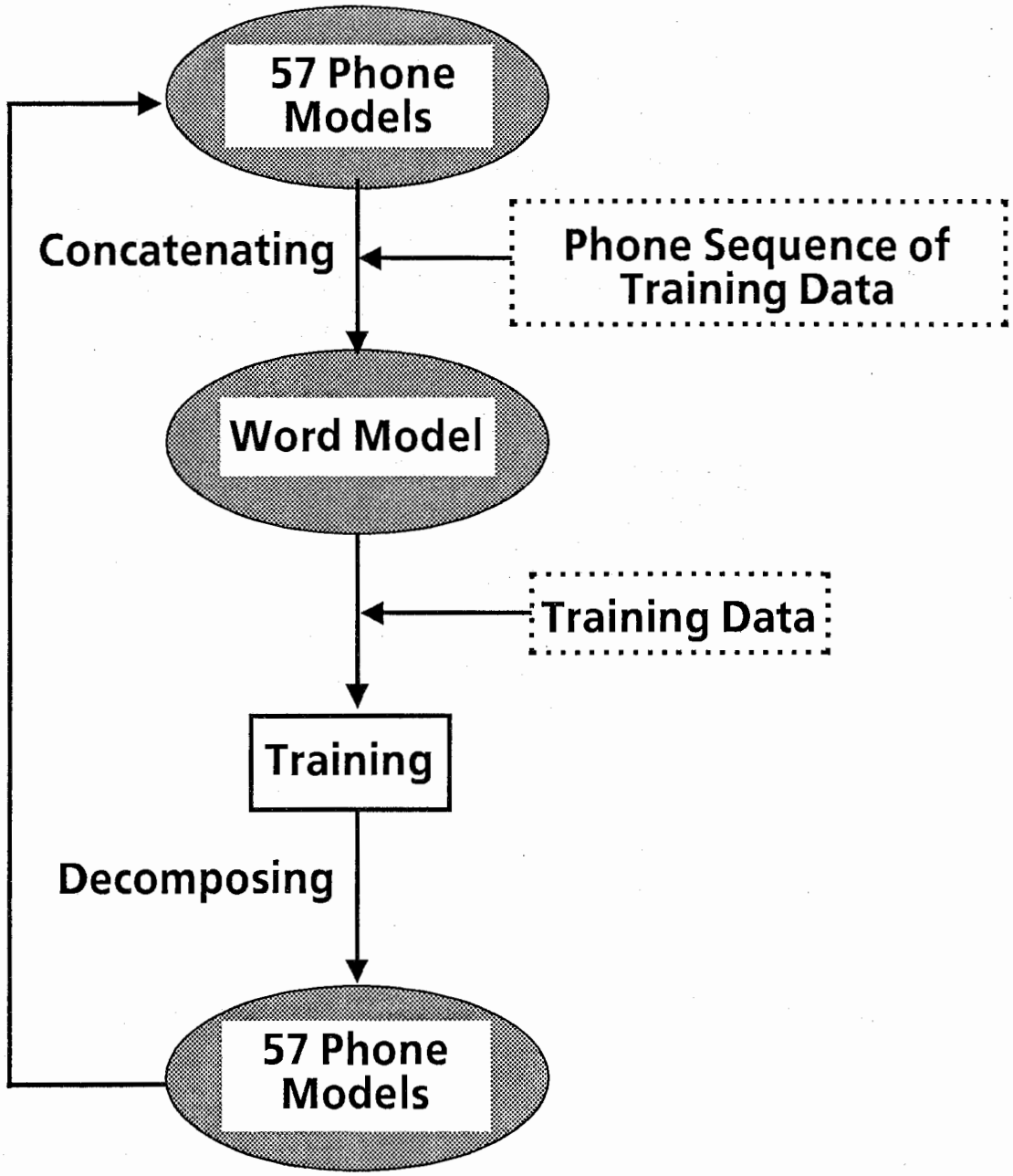

図1　HMM音韻連結学習

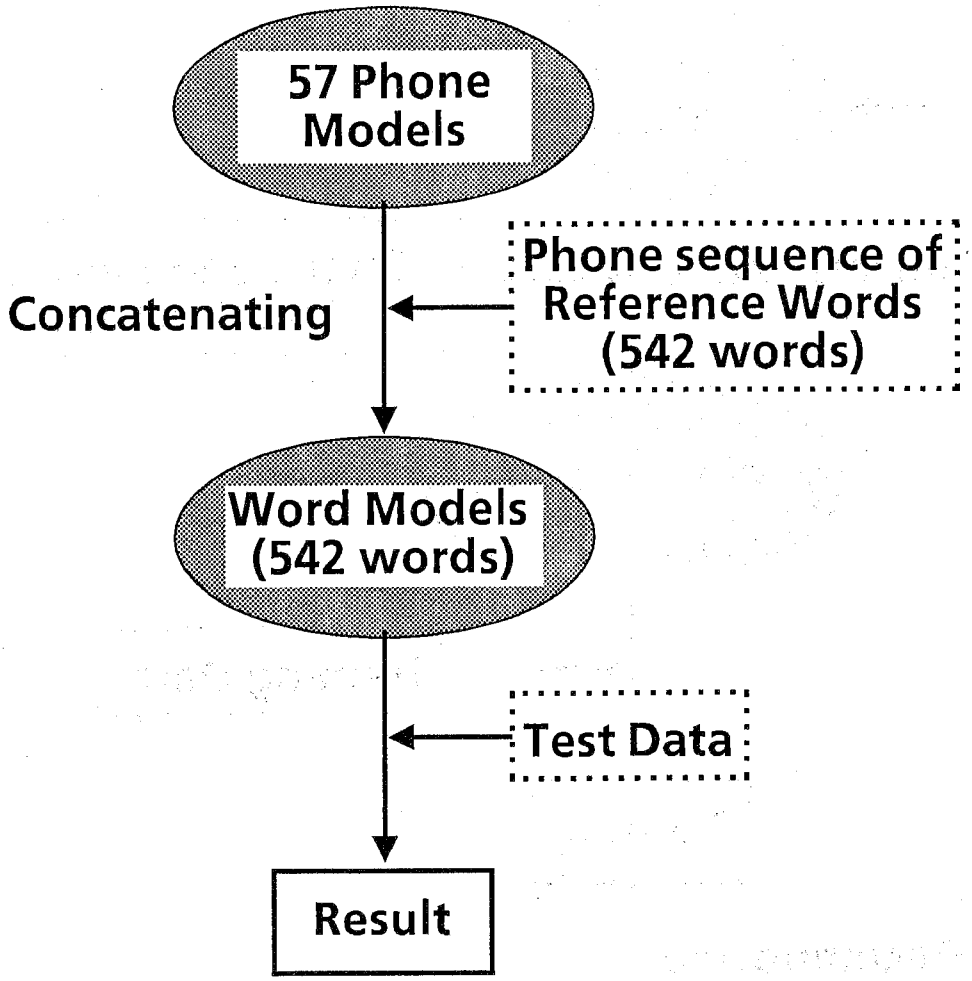

図2　HMM音韻連結学習による英単語認識

3.NETgramによる単語予測

3.1 NETgram

文章中の単語の音声認識結果の誤りを訂正する方法として、次に続く単語の出現確率を用いて単語列を予測する方法が知られている[2][3]。このような単語列予測の場合、過去いくつまでの単語情報を利用するかが、予測能力を向上させるためのキーポイントとなる。しかし、確率情報を用いる従来の方法(以下、確率モデルと記す)では、有効な統計量を獲得するために膨大なデータを必要とし、単語組数(gram数)を多くすれば計算機の記憶容量が増加すると言う問題があった。そこで、これらの問題を解決するためにニューラルネットワークモデル(NETgram)が提案された[4][5]。NETgramは、フリーパラメータの数を指数的に増加させることなくBigram, Trigram, $\cdots$, N-gramと容易に拡張可能なモデルである。(図3)

NETgramの基本となるBigramネットワークは、図4に示すような2つのHidden Layer(以下、HLと略す)をもつ、4層フィードフォワードネットワーク構造である。それぞれのHLがInput Layer、Output Layerの特徴のアナログ符号化を学習することを期待するためである。入力に対し出力は次の単語のカテゴリーを学習させるため、HLは単に単語カテゴリーの符号化ではなく、文章内で隣接する単語カテゴリー間の言語情報を持った符号化がなされることが期待できる。

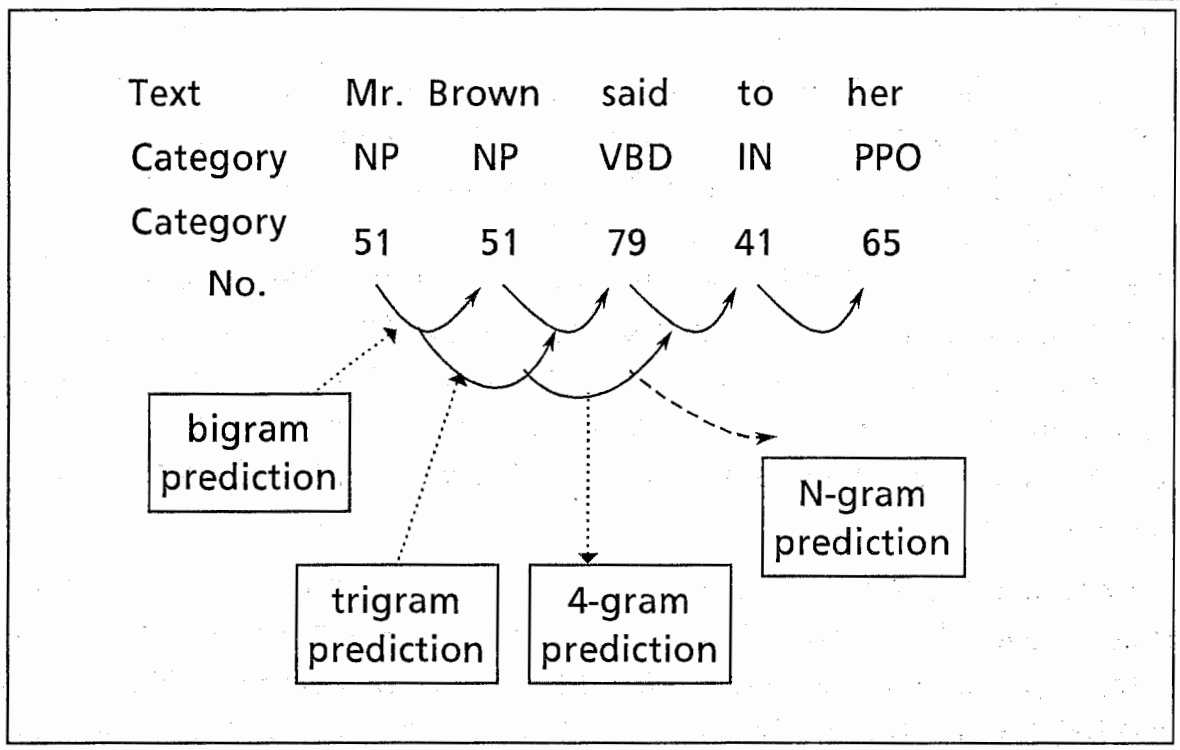

図3　Brown Corpus を用いた単語予測モデル

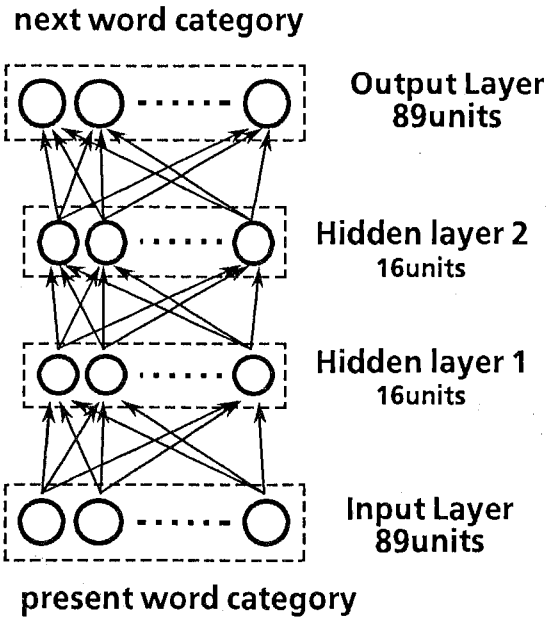

図4　Bigramネットワーク

3.2 N-gram への拡張

図5に示すようにNETgram では、Bigram ネットワークからTrigram ネットワークへの拡張は、新たなInput Layer からBigram ネットワークのHidden Layer 1 (図中HL1) にリンクウェイトセットw1' を結合させることによって行う。リンクウェイトセットw1' には、初期値に0 を与える。こうすることにより、Bigram ネットワークモデルで学習した出力と同じ値から学習を始めることができる。Trigram ネットワークから4-gram ネットワークへの拡張は、HL1 を追加し、N-2, N-3 のInput Layer から新たなHL1 へリンクウェイトセットw1, w1' を結合させる。

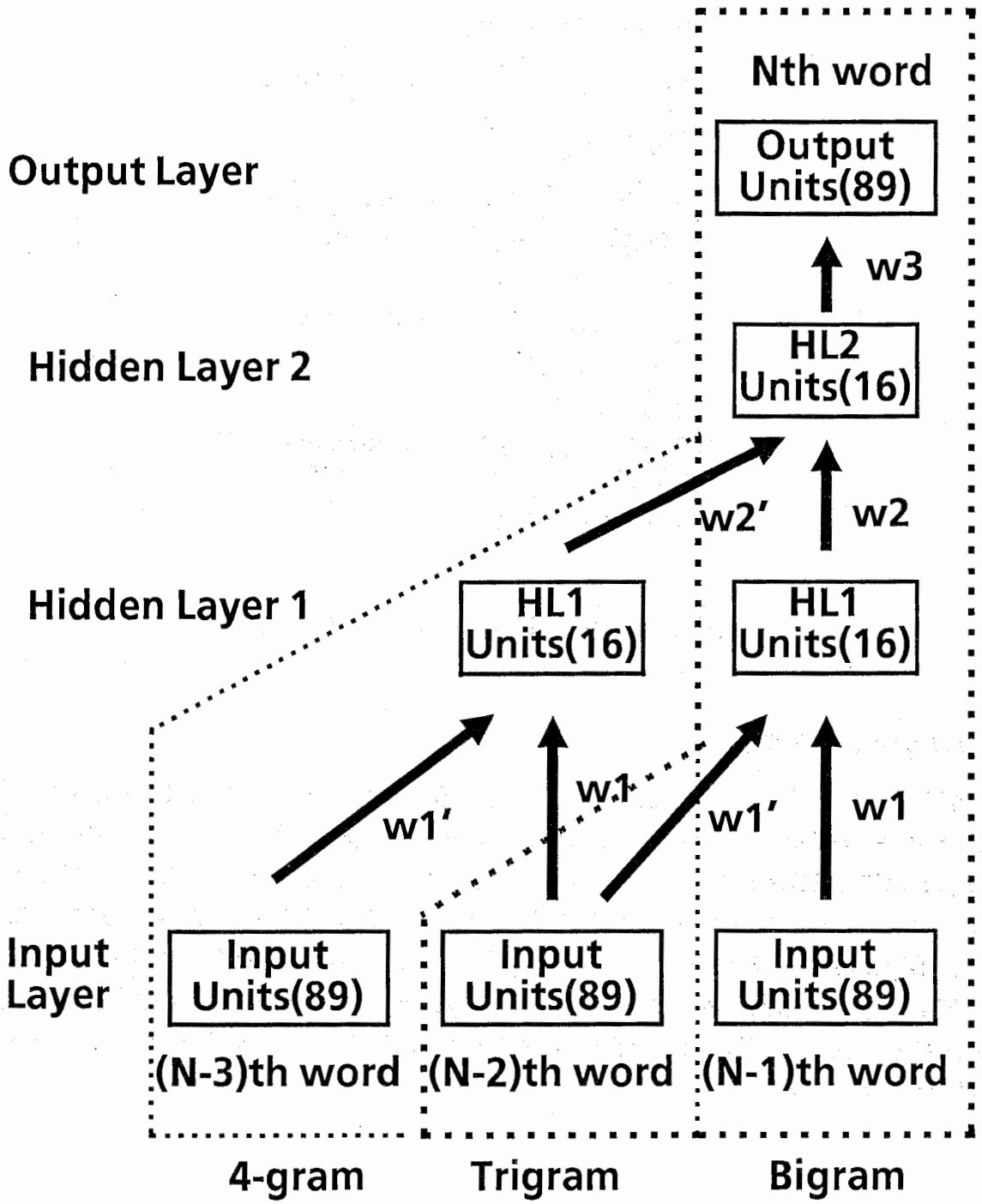

図5　NETgramの単語組数(gram数)の拡張

3.3 NETgramの学習

NETgram の学習、ここでは特に Trigram の学習について述べる。(図6)
入力には、Brown Corpus テキストデータベース[11] の文章の先頭の単語から句読点の1つ前までの単語カテゴリー(ここでは単語の品詞)を順に与えていく。入力は、そのカテゴリー番号のユニットのみ1を与え、他のユニットには0を与える。出力にも、そのカテゴリー番号のユニットのみ1を教え、他のユニットには0を教え、バックプロパゲーション法[12] により学習する。まず、Bigram ネットワークの学習を行い、次にその学習されたリンクウェイトの値を初期値として Trigram の学習を行う。

学習に用いるBrown Corpus テキストデータベース中の単語は、89種類のカテゴリー(品詞)に分類され、それぞれにカテゴリー番号がつけられている。付録2にテキストの内容を、付録3に分類されたカテゴリーを示す。本実験では、単語予測モデルのデータを10^{-6}で flooring しており、10^{-6}より大きいデータを有効データとして、それぞれのモデルでの有効データ数と全データ数(Bigram は 89^2、Trigram は 89^3)の対する割合を表1に示す。

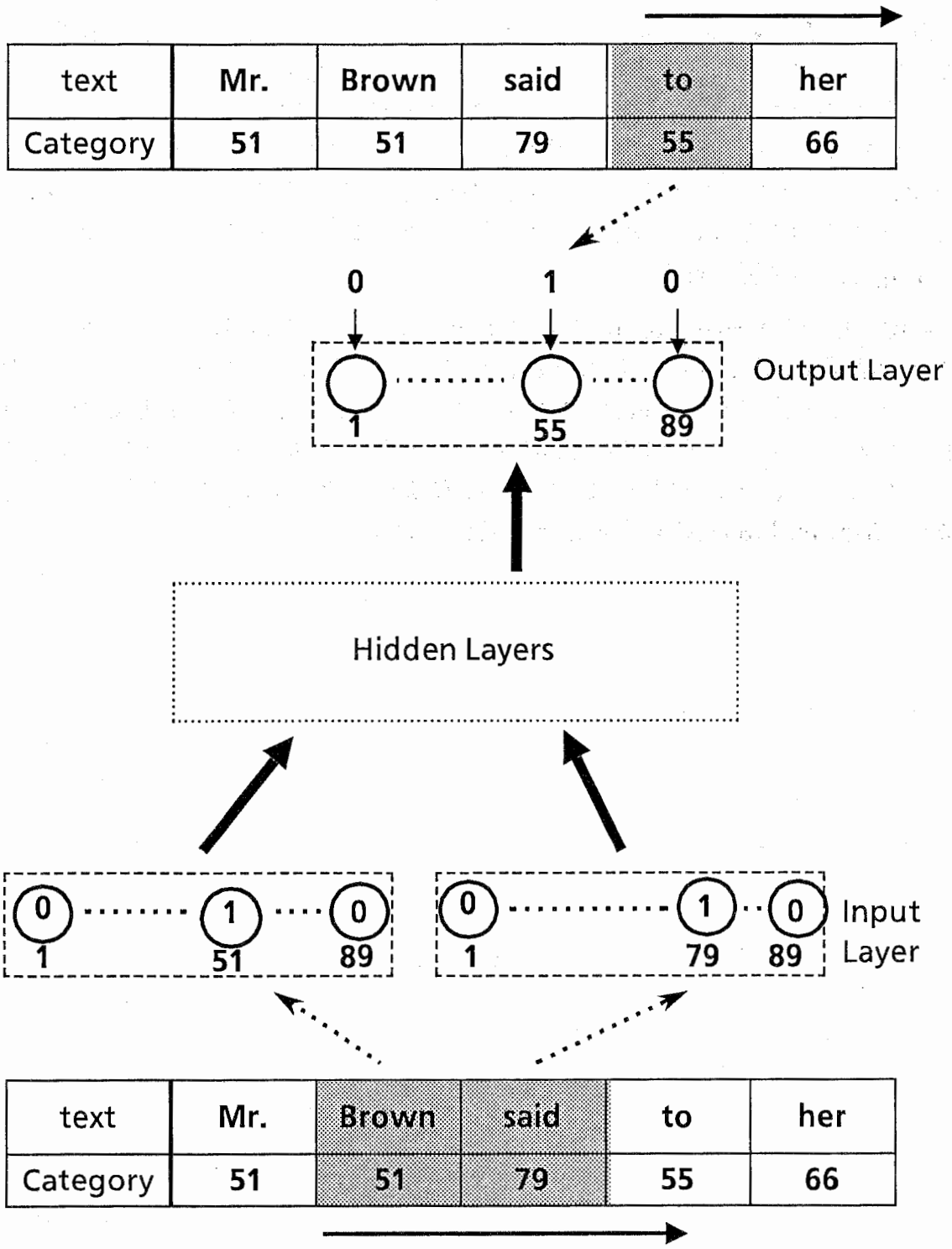

図6　NETgarmの学習（Trigram モデル）

表1　単語予測モデルに含まれる
有効データ数

Model	Training Sentences	NETgram	Statistical Model
Trigram	512	441,257 (62.6)	3,673 (0.5)
	1,024	442,028 (62.7)	5,610 (0.8)
	2,048	——	8,727 (1.2)
	4,096	456,831 (64.8)	12,944 (1.8)
	6,144	——	16,210 (2.3)
	8,192	——	18,013 (2.6)
	30,000	456,509 (64.8)	21,892 (3.1)
Bigram	512	4,915 (62.1)	973 (12.2)
	1,024	4,973 (62.8)	1,273 (16.6)
	2,048	——	1,697 (21.4)
	4,096	5,074 (64.1)	2,135 (27.0)
	6,144	——	2,430 (30.7)
	8,192	——	2,568 (32.4)
	30,000	5,217 (65.9)	3,347 (43.2)

( )内は、有効データ数の全データに対する割合(%)

4. NETgramを用いたHMM英単語音声認識の改善

HMM を用いた単語認識結果と NETgram(あるいは確率モデル) による品詞予測値から最適単語列を決定する。(図7)

文中 i 番目の単語を w_i、カテゴリー(本実験では品詞番号)を C_i とし、Trigram の単語予測値、すなわち w_{i-2}, w_{i-1} に続く単語 w_i の出現確率 $P(w_i/w_{i-2}\,w_{i-1})$ を以下のように近似する。

$$
\begin{aligned}
&P(w_i/w_{i-2}\,w_{i-1})\\
&\quad\simeq P(w_i/C_{i-2}\,C_{i-1})\\
&\quad= P(C_i/C_{i-2}\,C_{i-1})\\
&\qquad\times\{P(w_i/C_{i-2}\,C_{i-1})/P(C_i/C_{i-2}\,C_{i-1})\}\\
&\quad\simeq P(C_i/C_{i-2}\,C_{i-1})\{P(w_i)/P(C_i)\}
\end{aligned}
$$

まず、w_{i-2}, w_{i-1} から予測される w_i の出現確率は、カテゴリー C_{i-2}, C_{i-1} からの予測として近似される。

$$P(w_i/w_{i-2}\,w_{i-1}) \simeq P(w_i/C_{i-2}\,C_{i-1})$$

また、w_i の出現確率と、C_i の出現確率から以下の近似をする。

$$P(w_i/C_{i-2}\,C_{i-1})/P(C_i/C_{i-2}\,C_{i-1}) \simeq P(w_i)/P(C_i)$$

したがって、$P(C_i/C_{i-2}\,C_{i-1})$ と、$P(w_i)/P(C_i)$ を使って単語 w_i の出現確率を求める。$P(C_i/C_{i-2}\,C_{i-1})$ は、NETgramの出力値、あるいは確率モデルの予測確率値であり、カテゴリー C_i である単語における w_i の頻度である $P(w_i)/P(C_i)$ は、単語のカテゴリーを記述した辞書から得られる。

ここでは、$P(w_i/w_{i-2}\,w_{i-1})$ の対数をとり ST_i を定義する。

$$ST_i = \log P(C_i/C_{i-2}\,C_{i-1}) + \log(\,P(w_i)/P(C_i)\,)$$

以上は Trigram モデルの場合であるが、n-gram モデルのについても $P(C_i/C_{i-2}\,C_{i-1})$ を $P(C_i/C_{i-(n-1)}\,C_{i-(n-2)}\cdots C_{i-1})$ 置き換えることにより同様に計算できる。

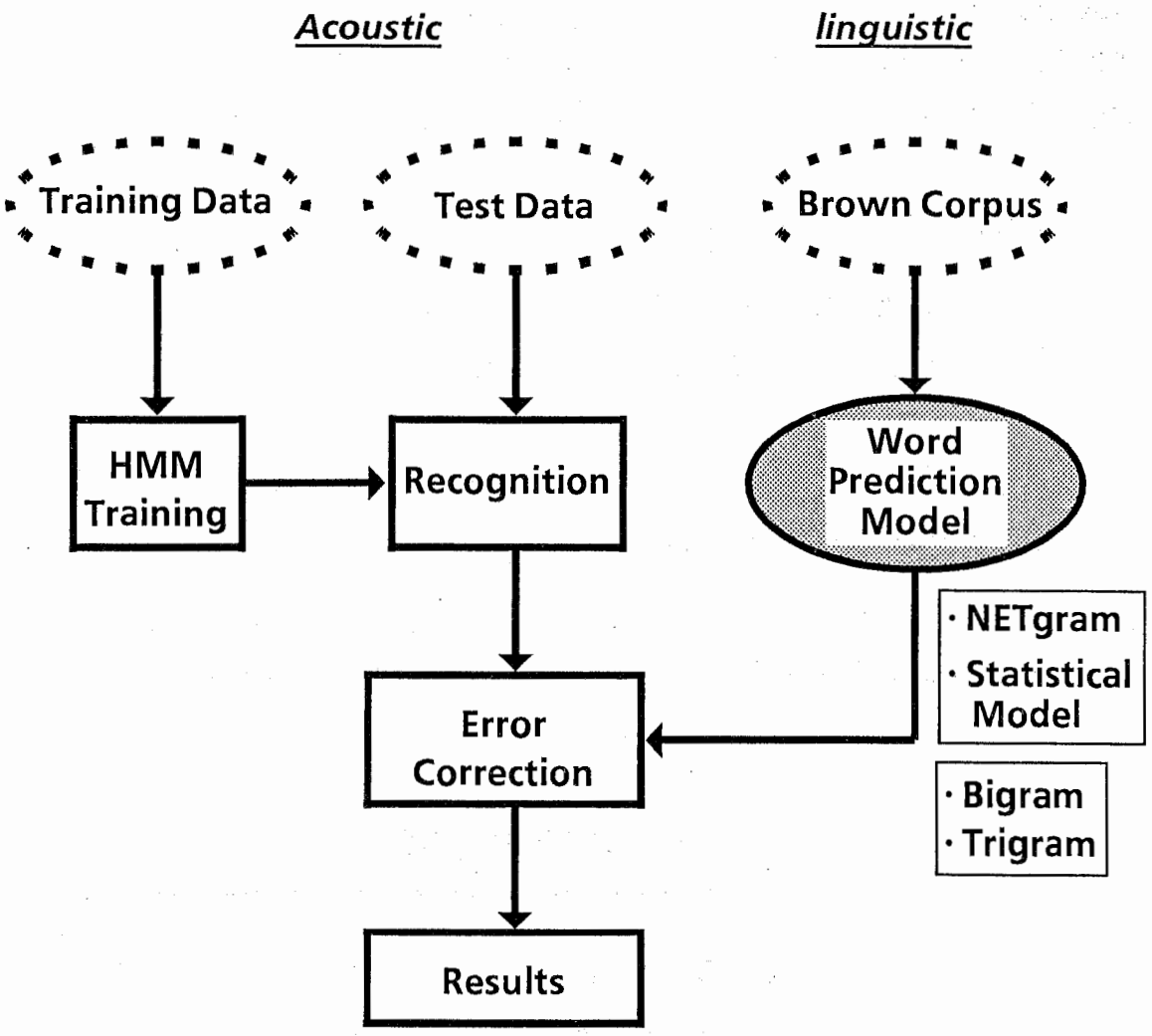

図7　単語予測モデルを用いた
HMM英単語音声認識結果の改善

次に、文中 i 番目の単語の j 番目の単語候補を $w_{i,j}$ とする。(図8)　単語候補として、HMM認識結果から上位10位までの単語を用いているが、複数のカテゴリーを持つ単語はそれぞれカテゴリー毎に異なる単語として扱うので、単語候補は10個以上($j \geq 10$)となる。単語 $w_{i,j}$ の類似度 $SW_{i,j}$ を、HMMによる音響的類似度 $SH_{i,j}$ とTrigramモデルによる言語上の類似度 $ST_{i,j}$ から以下のように決定する。

$$SW_{i,j} = (1-\omega) \cdot SH_{i,j} + \omega \cdot ST_{i,j}$$

ここで、 ω は重みを決める定数で $0<\omega<1$ とする。

そして、文頭から i 番目 までの $SW_{i,j}$ の和 $G_{i,j}$ を

$$G_{0,j} = SW_{0,j} \qquad (i = 0)$$
$$G_{i,j} = max(SW_{i,j} + G_{i-1,j}) \qquad (i \neq 0)$$

として計算する。文末まで計算後、 $G_{N-1,j}$ (N は文に含まれる単語数) が最大である単語 $w_{N-1,j}$ までの単語列をバックトレースして求め、認識結果とする。

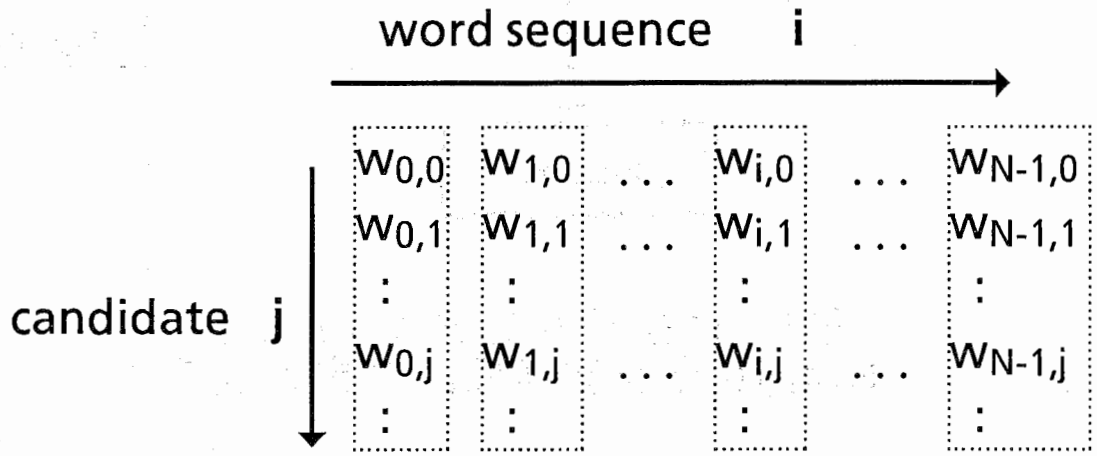

図8　単語列と単語候補

5. 実験結果

単語予測モデルとしてNETgramと確率モデルのそれぞれについてBigram、Trigramのモデルを用い、学習データ数を変えて実験した結果を表2に示す。学習データの1文は平均24単語である。なお、HMMのみによる単語認識率は81.0%である。

TrigramモデルでNETgramと確率モデルを比較すると、NETgramの場合、学習データを増加させても学習データが1,024文からは認識率の向上はあまり見られないのに対し、確率モデルでは学習データの増加に従って認識率が向上しており、学習データ30,000文では確率モデルの方が高い認識率となっている。したがって、学習データが十分多くある場合よりも、むしろ正確な確率モデルを得るには不十分な少ない学習データの時にNETgramはその効果を発揮することがわかる。

確率モデルの場合は、学習データに含まれない未知の単語列は学習データの増加に伴って減少する。すなわち有効な予測値を持つデータが増加するが、NETgramでは少ない学習データでも未知の単語列に対する予測値を持っており、そのデータ数は学習データを増やしても余り変わらないことが、NETgramで学習データを増やしても認識率が向上しない理由の一つであると考えられる。

Bigramモデルの場合、確率モデルはTrigramモデルと同様に学習データの増加に従って認識率が向上しているが、NETgramは学習データ4,096文で認識率が飽和している。これは、NETgramのフリーパラメータの許容の限界であろうと考えられる。

Trigramモデルについて実験結果を分析した結果、例えばwillは、9個がwellに誤認識されていたが、NETgramによりそのうち7個が正しく認識されるようになった。(表3) 改善効果の見られなかった例としては、doの場合、HMMの認識結果で18個あった誤認識が1個しか改善されていない。これは、評価データ中でdoの多くは"Do you ～？"のような疑問文で使われているが、単語予測モデルの学習に用いたBrown Corpusは会話文が含まれていないため疑問文が少なく、会話特有の言語構造の学習が不十分であるためと考えられる。

表2　各種単語予測モデルによる
認識率の改善結果(%)

Model	Training Sentences	NETgram	Statistical Model
Trigram	512	86.3 (0.005)	85.5 (0.002)
	1,024	86.9 (0.008)	85.4 (0.002)
	2,048	——	86.2 (0.003)
	4,096	86.9 (0.011)	86.6 (0.006)
	6,144	——	86.9 (0.006)
	8,192	——	87.5 (0.006)
	30,000	87.2 (0.009)	87.7 (0.006)
Bigram	512	85.7 (0.004)	86.0 (0.003)
	1,024	86.5 (0.001)	86.3 (0.005)
	2,048	——	86.6 (0.007)
	4,096	86.9 (0.006)	87.3 (0.014)
	6,144	——	87.5 (0.011)
	8,192	——	87.7 (0.011)
	30,000	86.9 (0.008)	87.7 (0.011)

(　)内は、ωの大きさを示す

表3　認識結果の改善された単語の例
(誤認識単語の数)

Prediction Model	Training Sentences	and	I	on	will
———	———	3	12	6	9
NETgram	512	0	7	5	2
	1,024	1	7	5	4
	4,096	1	5	5	3
	30,000	1	5	5	3
Statistical Trigram	512	2	10	4	2
	1,024	2	10	4	6
	4,096	1	9	3	2
	30,000	2	6	4	3

HMM認識実験で用いたデータ187文中の約30%が疑問文であるのに対し、単語予測モデルの学習に用いたデータは約3%しか疑問文が含まれていない。そこで、単語予測モデルの学習データとして約30%が疑問文となるようにBrown Corpusがら文を選びだして作った学習データで確率モデルを作り実験した結果を表4に示す。

疑問文を増やすことによって認識率の向上が見られる。特にBigramについては大きく向上している。同様に、この実験で用いた学習データでNETgramの学習を行うことにより認識率を向上できることが期待できる。

表4　認識率の改善結果(%)
(確率モデル)

Model	Training Sentences	疑問文 約3%	疑問文 約30%
Trigram	512	85.5	——
	1,024	85.4	86.3
	2,048	86.2	86.9
	4,096	86.6	——
	6,144	86.9	——
	8,192	87.5	——
	30,000	87.7	——
Bigram	128	——	86.2
	512	86.0	87.6
	1,024	86.3	88.0
	2,048	86.6	87.7
	4,096	87.3	87.9
	6,144	87.5	——
	8,192	87.7	——
	30,000	87.7	——

6. まとめ

HMM英単語音声認識結果を改善するためにNETgramと確率モデルのTrigram, Bigramモデルを用い、学習データ数を増加させて認識実験を行い、比較した。

学習データが十分ある場合には、確率モデルを用いた方が効果があるが、学習データが少ない場合には、NETgramの方が有効である。

しかし、単語候補としているHMM単語認識結果の10位までに正解が含まれていない単語があることや、HMM単語認識で使用している実験データが話し言葉であるのに対して単語予測モデルの学習に用いたBrown Corpusはほとんどが書き言葉であるという問題を解決することによりさらに認識率の改善を行うことができると考えられる。

参考文献

[1] 丸山、花沢、川端、鹿野:「HMM音韻連結学習を用いた英単語音声の認識」,信学技報 SP88-119(1989.1)

[2] F.Jelinek, "Continuous Speech Recognition by Statistical Methods", Proceedings of the IEEE, Vol.64, No.4 (1976.4)

[3] 鹿野:「Trigram Model による単語音声認識結果の改善」, 信学技報 SP87-23(1987, 6)

[4] 中村、鹿野:「英文テキストデータからのニューラルネットによる単語列予測モデルの検討」,信学技報 SP88-26(1988, 6)

[5] M.Nakamura, K.Shikano : "A Study of English Word Category Prediction Based on Neural Networks",ICASSP 89, S13.10(1989.5)

[6] Levinson,S.E, Rabiner,L.R. and Sondhi,M.M. : "An Introduction to the Application of the Theory of Probabilistic Functions of a Markov Process to Automatic Speech Recognition" , The Bell System Technical Journal, vol.62, No.4 (April,1983)

[7] Lee,K.F., Hon,H.W. : "Large Vocabulary Speaker-Independent Continuous Speech Recognition Using HMM" , ICASSP88(Aprl,1988)

[8] 花沢、川端、鹿野:「HMM音韻認識におけるセパレートベクトル量子化の検討」,音講論集2-P-8(1988.10)

[9] S.Furui, : "Speaker Independent Isolated Word Recognition Using Dynamic Feature of Speech Spectrum" , IEEE Trans.ASSP-34,1(1986)

[10] 丸山、鹿野:「英語音声データベースの作成」、ATR Technical Report TR-I-0048(1988.10)

[11] Brown University, "Brown Corpus", Tech. Report, Brown University (1967)

[12] D.E.Rumelhart et al., "Parallel Distributed Processing", M.I.T. Press (1986)

付録-1　HMM英単語認識に用いたデータの例

話者A,Bによる電話会話文

A Hello.

B Hello.

A I would like to attend the conference.

B Would you like us to send you the details?

A Yes, please. Could you tell me how I make an application?

B We will send you an application form.

A Thank you. I don't intend to give a paper, so is it all right if I just attend as a listener?

B Yes, it is. You can register and pay at the conference in August.

A How much is the registration fee?

B $100 U.S.

A Is it possible to pay in yen?

B I'm sorry, but we would like you to pay in dollars.

A Is it $100 for three days?

B That's right, hotel and food expenses are extra.

A I see. If I can only attend on the first day, is it possible to make a cancellation?

B Do you mean be entitled to a refund?

A Yes, is it possible to get 2/3 of the fee back?

B I'm sorry, but it is not possible.

A I see. What language will the conference be in?

B Both English and Japanese.

A I do not have an English capability. Will simultaneous interpretation be provided?

B Yes.

A I would like to converse with some of the participants. Will I be able to have an interpreter at my disposal?

B Yes.

A I see. Thank you very much.

B Do you have any other queries?

A No, thank you.

B O.K. Goodbye.

○ Informative Prose

- Press : Reportage
- Press : Editorial
- Press : Reviews (theatre, books, music, dance)
- Religion
- Skills and Hobbies
- Popular Lore
- Bells Letters, Biography, Memoirs, etc.
- Miscellaneous
- Learned

○ Imaginative Prose

- General Fiction
- Mystery and Detective Fiction
- Science Fiction
- Adventure and Western Fiction
- Romance and Love Story
- Humor

付録-3　Brown Corpus における品詞の分類

品詞番号	品詞略号	品詞の種類
1	.	文末(; ? . !)
2	(	左丸括弧
3	)	右丸括弧
4	*	not
5	--	ダッシュ
6	,	コンマ
7	:	コロン
8	ABL	pre-定限詞(quite rather)
9	ABN	pre-定限詞(half all)
10	ABX	pre-定限詞(both)
11	AP	post-決定詞(many several next)
12	AP$	所有格post-決定詞
13	AT	冠詞
14	BE	be
15	BED	were
16	BEDZ	was
17	BEG	being
18	BEM	am
19	BEN	been
20	BER	are
21	BEZ	is
22	CC	等位接続詞(and or)
23	CD	基数詞(one two 2)
24	CD$	所有格基数詞
25	CS	従位接続詞(if although)

品詞番号	品詞略号	品詞の種類
26	DO	do
27	DOD	did
28	DOZ	does
29	DT	単数決定詞(this that)
30	DT$	所有格単数決定詞
31	DTI	単数決定詞/複数決定詞(some any)
32	DTS	複数決定詞(these those)
33	DTX	決定詞/相関等位接続詞(either neither)
34	EX	存在のthere
35	FW	外来語
36	HV	have
37	HVD	had(過去)
38	HVG	having
39	HVN	had(過去分詞)
40	HVZ	haz
41	IN	前置詞
42	JJ	形容詞
43	JJR	比較形容詞
44	JJS	意味的最上級形容詞(chief top)
45	JJT	形態素的最上級形容詞(biggest)
46	MD	法助動詞(can should will)
47	NN	単数、質量名詞
48	NN$	所有格単数、質量名詞
49	NNS	複数名詞
50	NNS$	所有格複数名詞

品詞番号	品詞略号	品詞の種類
51	NP	固有名詞
52	NP$	所有格固有名詞
53	NPS	複数固有名詞
54	NPS$	所有格複数固有名詞
55	NR	副詞的名詞(home today west)
56	NR$	所有格副詞的名詞
57	NRS	複数副詞的名詞
58	OD	序数詞(first 2nd)
59	PN	名詞的代名詞(any- every- no- some-)
60	PN$	所有格名詞的代名詞
61	PP$	所有格人称代名詞(my our)
62	PP$$	第2所有格人称代名詞(mine ours)
63	PPL	単数再帰/強意人称代名詞(myself)
64	PPLS	複数再帰/強意人称代名詞(ourselves)
65	PPO	目的格人称代名詞(me him it them)
66	PPS	3人称単数主格代名詞(he she it one)
67	PPSS	主格人称代名詞(3人称以外) (I we they)
68	QL	定限詞(very fairly)
69	QLP	post-定限詞(enough indeed)
70	RB	副詞
71	RB$	所有格副詞
72	RBR	比較級的副詞
73	RBT	最上級的副詞
74	RN	名詞的副詞(here then indoors)
75	RP	副詞/不変化詞(about off up)

品詞番号	品詞略号	品詞の種類
76	TO	不定詞のto
77	UH	感嘆詞
78	VB	動詞
79	VBD	動詞(過去)
80	VBG	動詞(現在分詞、動名詞)
81	VBN	動詞(過去分詞)
82	VBZ	動詞(3人称単数現在)
83	WDT	WH決定詞(what which)
84	WP$	所有格WH代名詞(whose)
85	WPO	目的格WH代名詞(whom which that)
86	WPS	主格WH代名詞(who which that)
87	WQL	WH定限詞(how)
88	WRB	WH副詞(how where when)
89		文頭のブランク