大語彙連続音声認識における連鎖語の追加による語彙拡大の効果

和田陽介[†] 小林紀彦[†]
中野裕一郎[†] 小林哲則[†]

大語彙連続音声認識において，形態素連鎖を単語として追加登録することによりどの程度性能が向上するか，および，追加する形態素連鎖を選ぶ価値基準の違いが性能に与える影響について検討した．連鎖語の選定基準としては，高頻度の形態素連鎖を選定する方法，エントロピーの減少に貢献する形態素連鎖を選定する方法，延べ単語数の減少に貢献する形態素連鎖を選定する方法，の3つを試みた．毎日新聞のテキストコーパスとJNAS音声データベースを用いて，テストセットパープレキシティおよび連続音声認識実験における単語誤り率を評価した．どの手法を用いても追加連鎖語数を増加させるにつれて性能は向上したが，選択手法による顕著な差は見られず，性能向上のためには選択手法よりは選択語数が主要因となることが分かった．バイグラム言語モデルを用いるとき，連鎖語の語彙追加により，テストセットパープレキシティは最高で33%減少し，連続音声認識実験における単語誤り率は21%減少した．また，トライグラム言語モデルを用いるとき，テストセットパープレキシティ，単語誤り率ともに最高で19%減少した．

Effect of Vocabulary Extension using Word Sequence Concatenation for Large Vocabulary Continuous Speech Recognition

YOSUKE WADA,[†] NORIHIKO KOBAYASHI,[†] YUICHIRO NAKANO[†]
and TETSUNORI KOBAYASHI[†]

Vocabulary extension is utilized to improve the language model for large vocabulary continuous speech recognition (LVCSR). In this method, we make new words by concatenating some morpheme sequences and add them to the vocabulary. We tested three methods of extension: frequency-based extension, entropy-based extension and total-word-number-based extension. We tested their effects in terms of perplexity and recognition accuracy using Mainichi newspaper articles and JNAS speech corpus. As the results, all three methods contributed to improve the performance. There were no significant difference among these three methods. The size of the extended vocabulary was the primal factor for the performance. In case of bigram-based experiments, the best results gave 33% reduction of perplexity and 21% reduction of the word error rate. As for the trigram-based experiments, the best results gave 19% reduction of perplexity and also of word error rate.

1. まえがき

本論文では，日本語大語彙連続音声認識における，認識の基本単位の設定について検討を行う．

近年，大語彙連続音声認識システムは，単語Nグラムによる統計的言語モデルと，HMMを用いた音響モデルの組合せによって構成されるものがほとんどとなってきており，統計的言語モデルの研究はその重要さを増している．統計的言語モデルに関しては，様々な観点から研究が進められているが[1)~3)]，特に日本語の場合，認識の基本単位となる「単語」に明確な定義がなく，単語として何を選ぶべきかという問題からして大きな問題となる（本論文において「単語」とは，認識システムの辞書上に登録された，認識のための基本単位を表す語として用いる）．

従来，多くの日本語連続音声認識システムにおいて，形態素が単語として採用されてきた[4),5)]．しかしながら，日本語形態素は比較的短い音素列で構成されるものが多く，連続音声認識時に誤りを生じやすい．単語間の音響的な距離比較を容易にし，脱落・挿入を減じる意味からは，単語は長めの形態素列によって構成することが望まれる．また，拡大情報源の1記号あたりの情報量はその情報源のエントロピーに近づくというよく知られた事実からは，言語モデルの基本単位としても，より長い形態素列が有利であることが予想され

† 早稲田大学理工学部
School of Science and Engineering, Waseda University

る．特に，近年主流となっているバイグラムやトライグラムなどの短い履歴しか用いない単純な言語モデルを用いる場合，形態素列を基本単位とすれば，言語モデルの影響の及ぶ範囲が広がり，モデルの精度が向上することが期待できる．

このような観点から，本研究では，大語彙連続音声認識において，形態素連鎖を単語として採用し，これを基本語彙に追加することを試みる．

形態素や文字の連鎖に注目して，パープレキシティあるいは音声認識率を向上させようとした研究は，すでにいくつか試みられている．Masataki ら[6]は，クラスNグラムを言語モデルとして採用する認識システムのために，エントロピーを基準としながら，クラスの分割，単語の連結を行い，認識の基本単位を設定する方法を提案している．しかし，この方法は，クラスNグラムとの併用を前提とした手法であり，実験の結果は，本研究で目指す，単語Nグラムにおける形態素連鎖語登録の効果を知るうえで，直接の参考にはならない．伊藤ら[7]は，文字単位の出現予測において，文字列の出現頻度に注目し，高頻度文字列を新たに予測単位とすることの効果を報告している．また，森ら[8]は，クロスエントロピーを基準として，予測単位として効果的な文字列，形態素列を選定する方法を提案している．しかし，これらの研究は，エントロピーを減少させることが興味の中心であり，大語彙連続音声認識での評価は行われていない．エントロピーは，単語の予測しやすさに関する一面的な評価にすぎず，認識システムの性能を直接的に評価するものではない．他の手法による効果と比較するためにも，音声認識実験による評価は不可欠である．認識実験まで行った例としては，Giachin[9]の研究があるが，小語彙での実験にとどまっている．

本研究においては，大語彙連続音声認識実験を通じて，形態素連鎖の語彙追加の効果を評価することを目的とする．このような単純な処理の効果が言語モデルの複雑化などによる効果に比べ，どの程度に位置づけられるかを調べる．また，形態素連鎖語の選定方法に関しても，エントロピーを最小化する基準にとどまらず，形態素連鎖の頻度基準，あるいは延べ単語数を最小化基準するなどについて検討する．これらの語彙セットの選定においては，部分形態素列の重なりが複数の形態素列間に存在するため，ある形態素連鎖語の採用が他の連鎖語の採用の価値に影響を与える．このため，厳密な意味で基準を最適化する語彙セットを求めるのには，膨大な計算が必要となる場合が多い．ここでは，少量の計算で準最適な語彙セットを構成する手法についても提案する．

2. 連鎖語の選出のための基本アルゴリズム

本研究では，5000 種類の形態素を基本語彙として，その組合せによってできる形態素連鎖から，ある評価基準を最適化するする形態素連鎖の集合を，50〜1000 の規模で選択し，これを語彙に加えるものとする．ここでは，語彙として追加される形態素連鎖を形態素連鎖語あるいは単に連鎖語と呼ぶ．

考慮すべき形態素連鎖の種類は莫大な数であり，また，1つの形態素連鎖を単語として採用するかどうかが，他の形態素連鎖の採用の価値に影響を与える性質の問題であることから，最適な解を求めることは，実質的には不可能である．このため，同種の問題においては，様々な近似手法がとられている．一般的には，一度計算した評価関数値をもとに，上位候補の採用にともなう評価関数値の影響を無視（あるいは近似的に補正）して採用候補を求める方法や，順次，候補の採用にともなって評価関数値を再計算しながら，そのつど最適な評価関数値をあたえる候補を採用していく方法[6),9)]などがとられる．前者の方法の精度が悪いのはもちろんであるが，後者の方法にしたところで，あくまで近似手法であり，また，語彙および選択候補数が多くなるにつれて，計算量は膨大になるという問題を含む．

このため，ここでは，以下の選定アルゴリズムを採用した．

連鎖語選定アルゴリズム

1. 連鎖語の候補を適当数リストアップする．
2. 以下を，収束するまで繰り返す：
   - 2-1. 各候補 w_i について，それを採用したときの，評価関数の上限推定値 $F_{sup}(w_i)$ と 下限推定値 $F_{inf}(w_i)$ を求め，それぞれの値で，候補をソートする．上限推定値でソートして得られる順位表を上限表，下限推定値でソートして得られる順位表を下限表とする．
   - 2-2. 以下を収束するまで繰り返す：
     - 2-2-1. 上限表で選択予定数 N に順位づけられた候補 w_{SN} の上限推定値 $F_{sup}(w_{SN})$ を求め，下限表で $F_{sup}(w_{SN})$ 以上の評価関数値を持つ候補の採用を決定する．
     - 2-2-2. いくつかの単語の採用が決定したことによる，各候補の上限推定値に対する影響を計算し，これを補正したうえで，上限表を作成しなおす．

2-3. 下限表で選択予定数 N に順位づけられた候補 w_{IN} の下限推定値 $F_{inf}(w_{IN})$ を求め，上限表で $F_{inf}(w_{IN})$ 以下の評価関数値しか持たない候補を棄却し，候補を削減する．

3. 上限表の上位 N 位に含まれる，採用未決定の候補を採用する．

このアルゴリズムにおいては，一般に 2. までの段階ですべての採用を確定できないため，3. の段階で近似が入る．しかしながら，他候補の採用にともなう評価関数値の影響を無視する方法に比べれば信頼性は高い．また，2-2-2. における上限値の修正は簡素化が可能であるから，評価関数の再計算は 2-1. に集約できる，一括して複数の候補を採用できることから，評価関数の再計算を進めながら逐次的に採用を決める方法に比べれば計算量を抑えることができる．

3. 連鎖語の選出のための評価基準

3.1 高頻度の形態素連鎖語を採用する基準

最初の基準としては，形態素連鎖の出現頻度を選んだ．

高頻度形態素連鎖語の選定にあたっては，前章の方法における上限推定値と下限推定値を決める必要がある．連鎖語 w_i の頻度基準における上限推定値 $C_{sup}(w_i)$ としては，学習テキストを形態素解析して得た形態素列データにおける w_i を構成する形態素連鎖の頻度を用いた．各形態素連鎖の計数は，他の形態素連鎖の計数と独立して行われるため，重複した計数が生じる．このため，値は真値より高くなり，w_i の上限推定値としての性質を持つ．下限推定値 $C_{inf}(w_i)$ としては，学習用形態素列データに対し全連鎖語候補との最長一致基準を適用して得た単語列データにおける w_i の頻度を用いた．ここでは，複数の形態素連鎖語に対応する可能性のある形態素は，原則的には長い方の形態素連鎖語に含められることになるため，形態素連鎖の重複した計数はない．また，最終的に採用されるかどうか分からない連鎖語も含めて解析を行うことから，各形態素列の頻度は真値より少なめに見積もられることになり，推定下限値としての性質を持つ．

上限値の補正には，採用された連鎖語の部分列となっている形態素連鎖語について，その上限値から採用された連鎖語の下限推定値を差し引く方法をとった．採用語を v とすれば，

$$\hat{C}_{sup}(w_i) = C_{sup}(w_i) - C_{inf}(v), \qquad (1)$$

と表現できる．ここで，$\hat{}$ は補正後の値を表すものとする．初期形態素連鎖語候補としては，5 形態素連鎖までを考慮したうえで，上限推定値の上位を用いた．

3.2 エントロピーを減少させる基準

第 2 の基準は，1 形態素あたりのエントロピーの減少である．従来，連鎖語を扱う研究の多くは，これに準じる基準を設けている．

テストセットに対し，1 形態素あたりのエントロピーは，以下の式で与えられる．

$$H = \frac{1}{\sum |w_i|} \sum -\log P_M(w_i). \qquad (2)$$

ここで，w_i は学習コーパス中の単語を表し，$P_M(w_i)$ は採用する言語モデルが与える w_i の出現確率を表す．$|w_i|$ は，単語 w_i を構成する形態素の数であり，この総和で情報量の総和を除すことは，単語セットの違いを補正して，1 形態素（連鎖語採用前の基本単位）あたりのエントロピーを求める意味を持つ．

このエントロピーを，膨大な形態素連鎖語候補に対して，その採用候補の組合せを変えるたびに算出しなおすのが理想であるが，これは非常にコストが高い．そこでここでは，各形態素連鎖語の採用にかかわるエントロピーの評価基準として，全体のエントロピーを計算する代わりに，問題となる形態素連鎖を含む形態素列のみを対象とした部分エントロピーを考える．

バイグラム言語モデルを採用するものとして，形態素連鎖 $a{\cdot}b$ を考えるとき，これを含む全形態素列 $x_i{\cdot}a{\cdot}b{\cdot}y_j$ について，$a{\cdot}b$ を連鎖語として採用する前のエントロピー $H_0(a{\cdot}b)$ は，次式で表される．

$$\begin{aligned} H_0(a{\cdot}b) &= -\sum_{i,j} P(x_i,a,b,y_j)\{\log P(a|x_i) \\ &\quad + \log P(b|a) + \log P(y_j|b)\} \\ &= -\Big\{\sum_i P(x_i,a,b)\log P(a|x_i) \\ &\quad + P(a,b)\log P(b|a) \\ &\quad + \sum_j P(a,b,y_j)\log P(y_j|b)\Big\}. \end{aligned} \qquad (3)$$

一方，形態素連鎖語 $a{\cdot}b$ を採用後のエントロピー $H'(a{\cdot}b)$ は，次式で表される．

$$\begin{aligned} H'(a{\cdot}b) &= -\sum_{i,j} P(x_i,a,b,y_j) \\ &\quad \times \{\log P(a,b|x_i) + \log P(y_j|a,b)\} \\ &= -\Big\{\sum_i P(x_i,a,b)\log P(a,b|x_i) \\ &\quad + \sum_j P(a,b,y_j)\log P(y_j|a,b)\Big\}. \end{aligned} \qquad (4)$$

ここで，連鎖語 $a{\cdot}b$ を採用した時のエントロピー変化

の期待値 $H_0 - H'$ を，評価関数 $E(a{\cdot}b)$ とすることにすれば，

$$\begin{aligned} E(a{\cdot}b) &= H_0(a{\cdot}b) - H'(a{\cdot}b) \\ &= -\Big\{ \sum_i P(x_i, a, b) \\ &\quad \times \{\log P(x_i|a) - \log P(x_i|a,b)\} \\ &\quad + \sum_j P(a, b, y_j) \\ &\quad \times \{\log P(y_i|b) - \log P(y_j|a,b)\} \Big\}, \end{aligned} \tag{5}$$

が得られる．同様に，3 形態素の連鎖語 $a{\cdot}b{\cdot}c$ についての評価関数 $E(a{\cdot}b{\cdot}c)$ は，

$$\begin{aligned} E(a{\cdot}b{\cdot}c) &= H_0(a{\cdot}b{\cdot}c) - H'(a{\cdot}b{\cdot}c) \\ &= -\Big\{ \sum_i P(x_i, a, b, c) \\ &\quad \times \{\log P(x_i|a) - \log P(x_i|a,b,c)\} \\ &\quad + \sum_j P(a, b, c, y_j) \\ &\quad \times \{\log P(y_j|c) - \log P(y_j|a,b,c)\} \\ &\quad + P(a,b,c)\{\log P(c|b) - \log P(c|a,b)\} \Big\}, \end{aligned} \tag{6}$$

となる．

選択にあたって，評価関数の上限推定値 $E_{sup}(w_i)$ としては，各形態素連鎖の頻度を独立に計数して上記評価関数を計算したときの値を用いた．

$$E_{sup}(w_i) = E(w_i) = H_0(w_i) - H'(w_i). \tag{7}$$

下限推定値の算出にあたっては，まず 3.1 節に述べたと同じ方法によって，頻度の上限推定値 $C_{sup}(w_i)$ と，下限推定値 $C_{inf}(w_i)$ を求め，これらの値でエントロピー基準の上限推定値を按分することで得た．

$$E_{inf}(w_i) = \frac{C_{inf}(w_i)}{C_{sup}(w_i)} E_{sup}(w_i). \tag{8}$$

本来形態素解析しなおすことで，バイグラムなどの確率そのものが変化するため，この方法による下限推定値は，下限値としての条件を満たさない．しかし，こうすることで，単語組の計数が不要になり，劇的に計算量を減じることができるため，これを採用した．同様に，採用する連鎖語の決定にともなう上限推定値の補正についても，確定した形態素連鎖語に応じて，各形態素連鎖語の頻度の上限推定値を補正し，修正前の頻度の上限値と修正後の頻度の上限値で按分する方法を用いた．

$$\hat{E}_{sup}(w_i) = \frac{\hat{C}_{sup}(w_i)}{C_{sup}(w_i)} E_{sup}(w_i). \tag{9}$$

3.3 延べ単語数を減少させる基準

第 3 の基準は，コーパスに現れる延べ単語数を減少させる基準である．これは，1 単語あたりの平均形態素数を最大化するのと同じことである．音声認識時における脱落・挿入を抑えるため，できるだけ長い単語を作ろうとすることに相当する．

一般に K 連鎖の形態素連鎖を採用すれば，その形態素列の頻度の $K-1$ 倍だけ延べ単語を減じることができる．これを考慮し，連鎖語 w_i の延べ単語減少基準における評価関数の上限推定値 N_{sup} として，次式を用いた．

$$N_{sup}(w_i) = W(w_i) \cdot C_{sup}(w_i). \tag{10}$$

ここで，C_{sup} は頻度の上限推定値であり，$W(w_i)$ は重みで，その初期値は w_i を構成する形態素数から 1 減じた値である．下限値 $N_{inf}(w_i)$ としては，3.1 節と同じく頻度の下限値 $C_{inf}(w_i)$ とした．

$$N_{inf}(w_i) = C_{inf}(w_i). \tag{11}$$

下限推定値の算出にあたって，$K-1$ の重みを掛けないのは，K 連鎖の形態素連鎖を採用したとしても，その部分列となっている $K-1$ 連鎖の形態素連鎖語が同時に採用されるならば，K 連鎖の語を採用による延べ単語の減少効果はたかだか頻度分しかないためである．

採用する連鎖語の決定にともなう評価関数の上限推定値の補正は，次の方法によって行った．採用した語 v を部分列として含む連鎖語 w_i については，評価関数における重み計数に以下の変更を与えた．

$$\hat{W}(w_i) = \min(W(w_i), |w_i| - |v|). \tag{12}$$

採用した語 v に部分列として含まれる連鎖語 w_i については，評価関数における頻度の上限推定値を，3.1 節と同様に補正した．

$$\hat{C}_{sup}(w_i) = C_{sup}(w_i) - C_{inf}(v). \tag{13}$$

$\hat{N}_{sup}(w_i)$ は，$\hat{W}(w_i)$ と $\hat{C}(w_i)$ の積で表される．

4. 実　　験

上記 3 基準を用いて，連鎖語をそれぞれ 50，100，250，500，1000 語選出し，テストセットパープレキシティ，1 単語あたりの平均形態素連鎖数，連続音声認識における単語認識率の観点から評価した．また，連鎖語選出のための学習データには，1991 年 1 月から 1994 年 9 月までの毎日新聞 45 カ月分の記事を用いた．なお，2 章に述べた選出アルゴリズムにおけるステップ 2. の段階で選出できた割合（高い信頼度で選出した割合）は，頻度，エントロピー，延べ単語数

の各基準について，それぞれ平均62%，55%，40%であった．

4.1 テストセットパープレキシティによる評価

式(2)に示した，1形態素あたりのエントロピー H を用いて，1形態素あたりのテストセットパープレキシティを，

$$P = 2^H, \tag{14}$$

として求めた．

図1，**図2**は，各方法で求めた語彙セットによる1形態素あたりのテストセットパープレキシティを表す．横軸は，基本語彙に追加した形態素連鎖語の数である．言語モデルの学習には，1991年1月から1994年9月までの毎日新聞45カ月分の記事を用いた．図1は，毎日新聞の1994年10月から12月までの3カ月分のうち，出現単語が認識実験で用いる基本語彙5千語に閉じている約26000文を選んでテストセットとして用いた場合の結果であり，図2は，ASJ JNASにおける5Kセット（基本語彙5千語に閉じている文の集合）から選んだ200文をテストセットとして用いた場合の結果である．JNASの200文は，後で認識実験に用いるデータセットと同じものである．

図1から，新聞記事の大規模コーパスによって評価するとき，バイグラム言語モデルによるテストセットパープレキシティは，語彙を追加するとともに低下していることが分かる．バイグラムのテストセットパープレキシティにおける最良の結果は，エントロピー基準で1000語選んだ場合で，ベースライン（語彙追加のない場合）の80.0から61.3まで低下した．これは，これは23%の減少にあたる．トライグラム言語モデルによるテストセットパープレキシティは，追加語彙500程度で，43程度に収束している．最良の結果は，延べ単語数を最小化する基準で1000語選んだ場合で，テストセットパープレキシティは，ベースラインの47.0から43.0まで低下した．これは，9%の減少にあたる．手法間の差異はごくわずかであるとともに，バイグラムのテストセットパープレキシティの最適値を与える語彙セットが必ずしもトライグラムのテストセットパープレキシティの最適値を与えないことが分かる．

図2のJNAS200文をテストデータセットとして用いた評価実験の結果においては，図1に比べ，語彙追加の効果が若干高いことと，選択手法間の差異が若干大きいことを除いて，おおむね図1と同様の結果を得ている．手法間の差異の多くは，データ数が少ないことが原因と考えられる．バイグラムのテストセットパープレキシティにおける最良の結果は，延べ単語数

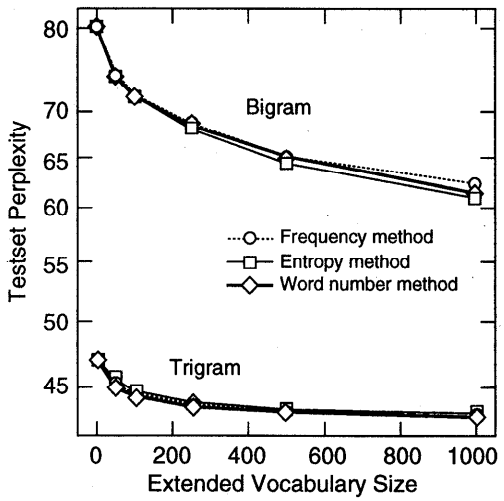


図1 追加連鎖語数とテストセットパープレキシティの関係．評価データは毎日新聞3カ月

Fig. 1 Test-set perplexity as a function of extended vocabulary size. Test-set consists of 26000 sentences from Mainichi newspaper articles for 3 months.

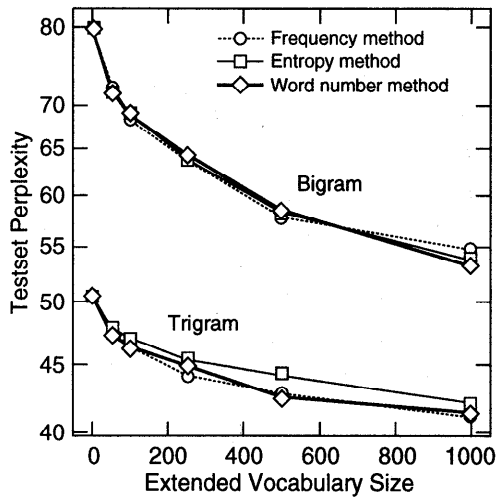


図2 追加連鎖語数とテストセットパープレキシティの関係．評価データは認識実験に用いるJNAS 200文

Fig. 2 Test-set perplexity as a function of extended vocabulary size. Test-set consists of 200 sentences from JNAS speech corpus.

を最小化する基準で1000語選んだ場合で，ベースラインの79.8から53.1まで低下した．これは33%の減少にあたる．トライグラム言語モデルによるテストセットパープレキシティにおける最良の結果は，頻度基準で1000語選んだ場合で，ベースラインの50.6から41.1まで低下した．これは19%の減少にあたる．

語彙追加の効果が，新聞記事データよりはJNAS200文の方が大きいことから，ここで選んだ認識用のテス

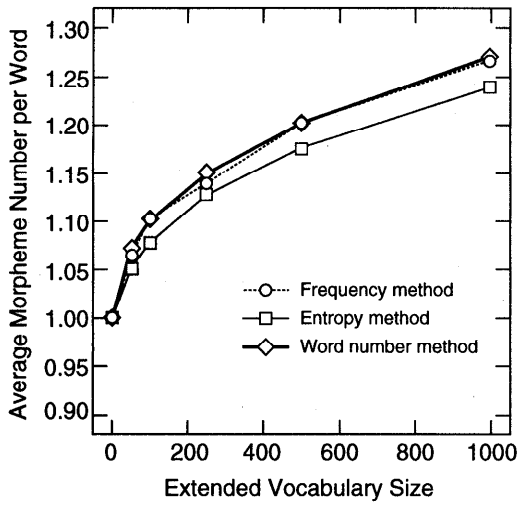


図 3 追加連鎖語数と平均形態素連鎖数の関係．評価データは認識実験に用いる JNAS 200 文

Fig. 3 Average morpheme number per word as a function of extended vocabulary size. Test-set consists of 200 sentences from JNAS speech corpus.

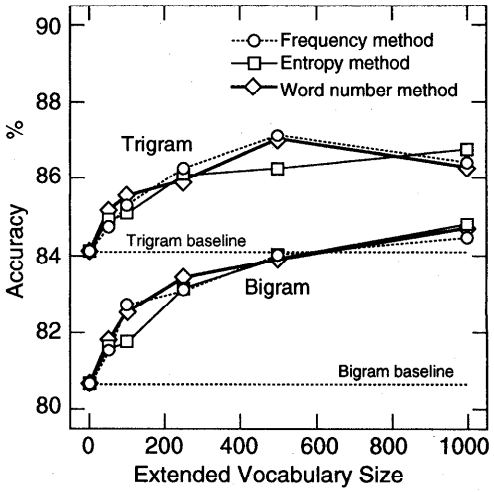


図 4 基本語彙 5 千における連続音声認識実験結果．評価データは JNAS 男性 20 名 200 文

Fig. 4 Performance of 5000 word continuous speech recognition as a function of extended vocabulary size. Test-set consists of 200 sentences uttered by 20 male speakers which is from JNAS speech corpus.

トセットは，平均的な新聞記事よりは連鎖語追加の効果が出やすいセットということができる．

4.2 平均形態素連鎖数による評価

図 3 は，1 単語あたりの平均形態素数を表す．評価データは，JNAS の 200 文である．

効果は，延べ単語数を評価基準とした方法において最も高いが，頻度を基準とする方法との差はごくわずかである．延べ単語数を基準とした方法を用いるとき，1 単語あたりの平均形態素数は，1.27 となった．これは，延べ単語が 約 21%減少したことにあたる．

4.3 大語彙連続音声認識率による評価

連続音声認識実験を次の条件で行った．

テストデータ： 日本音響学会 JNAS 連続音声データベース[15] における 5K セットから選んだ，男性 20 人の発声による 200 文．

特徴パラメタ： 12 次 MFCC，パワー，およびその差分からなる，26 次元ベクトル．

音響モデル： 子音は後続音に依存したバイフォン．母音は，前後の音に依存したトライフォン．5 状態 3 ループの HMM で，各状態は，4 混合正規分布で表現．

言語モデル： 基本 5 千語の語彙セットによる，バイグラム，トライグラム，および，3 章に述べた 3 基準のそれぞれで，50，100，250，500，1000 語の連鎖語を登録したときの，バイグラム，トライグラム．

デコーダ： 木構造辞書を用いたフレーム同期のワンパスビームサーチ[11] に，束ね[12),13)] 処理を導入してサーチを簡素化するとともに，トライグラム言語モデルまで扱えるよう拡張したもの[14]．単語内のみ音素文脈依存性を考慮し，単語間は音素文脈不依存の音響モデルを用いて接続した．

評価尺度： 単語正解精度および単語誤り率．それぞれの値は，正解単語数を C，挿入単語数を I，置換単語数を S，脱落単語数を D とするとき，以下のように表される．

$$\text{単語正解精度} = \frac{C-I}{C+S+D}$$

$$\text{単語誤り率} = \frac{S+D+I}{C+S+D}$$

なお，正解単語数の計数にあたっては，「および」と「及び」など，異なる表記でも同じ意味の単語は，正解として数えた．

図 4 に，連続音声認識実験の結果を示す．

この図よりバイグラム言語モデルを用いた場合には，どの手法を用いても，追加する連鎖語の数を増加するにともなって，認識率が向上していることが分かる．最も改善効果が高かったのは，エントロピー基準で 1000 の連鎖語を追加登録した場合で，単語正解精度は追加前の 80.6%から 84.8%まで改善できている．これは，約 21%の誤りを改善したことにあたる．また，500 語連鎖語を追加した場合の，バイグラム言語モデルの性能は，トライグラムのベースラインとほぼ同等であり，1000 語追加した場合には，トライグラ

表 1 認識実験における誤りの内訳
Table 1 Itemized count of errors.

選択手法（選択数）	置換数（改善率）	脱落数（改善率）	挿入数（改善率）
ベースライン	225 (–)	71 (–)	42 (–)
頻度（500）	188（16%）	46（35%）	39（7%）
エントロピー（1000）	195（13%）	46（35%）	40（5%）
延べ単語数（500）	192（15%）	47（34%）	36（14%）

ムのベースラインを超える性能となっている．

トライグラム言語モデルを用いた場合には，性能は500語追加あたりでピークをとっている．最も改善効果が高かったのは，頻度基準で500語追加した場合で，単語正解精度は，84.1%から87.1%に改善している．これは，約19%の誤りを改善したことにあたる．

語彙拡大による誤り傾向の変化を見るために，各選択基準での最良の結果を与えた選択語数における，誤りの内訳を**表1**に示す．この表によれば，連鎖語の追加による語彙拡大の効果は，特に脱落の減少に表れている．

5. 検　　討

拡大情報源の1記号あたりの情報量はその情報源のエントロピーに近づくというよく知られた事実からは，連鎖語の考慮によって，連鎖語の1形態素あたりの情報量は情報源のエントロピーに近づくことが予想される．このため，バイグラム，トライグラムのような考慮する履歴の短い単純な言語モデルを用いた場合でも，モデルエントロピーが下がることが期待されたが，実験によって，これが確かめられた．

音声認識においては，この言語モデルの精度の向上に加え，認識単位が長くなることによる脱落・挿入の可能性の減少，読みの精度向上（複数の読みが可能な2つの単語が並ぶ場合，それらを連鎖語として登録しない場合にはその組合せすべての読みが許されることになるが，連鎖語として登録する場合には，1通りの読みが確定して与えられる），音響モデルの音素文脈依存処理の容易性（音響モデルの連結にあたり，連鎖語内での形態素の境は単語内として処理されるため，単語内音素文脈依存処理で扱われる部分が増える）などのメリットが生じる反面，語彙が増加することにともなうサーチスペースの増大というデメリットが生じる．実験の結果から見れば，多くの条件において，形態素連鎖語の語彙追加のメリットは，デメリットを上回っていたということができる．

連鎖語の追加が性能向上に結び付かなかった数少ない例として，トライグラム言語モデル利用時において，頻度基準と延べ単語最小化基準で1000語連鎖語を追加した結果が500語追加の結果を下回った例があげられる．この例においては，追加語彙数500で，テストセットパープレキシティがすでに飽和していたため，それ以上の語彙追加は，単語予測精度を向上させる観点からはメリットがない一方で，語彙の追加によって探索空間が広がり，相対的にビーム幅が減少するというデメリットが増えたため，追加語数500程度で性能がピークを持ったものと考える．

6. む　す　び

大語彙連続音声認識において，形態素連鎖を新たな単語として語彙登録することの効果を調べた．

形態素連鎖語の選出方法としては，頻度の高い連鎖語を選出する方法，エントロピーの減少に貢献する連鎖語を選出する方法，延べ単語数の減少に貢献する連鎖語を選出する方法，の3つの方法を試した．テストセットパープレキシティおよび単語正解精度の観点からは，おおよそエントロピー基準において，連鎖語の追加語数が少ないとき，他の基準に比べ効果が劣る一方，追加語数が多くなると効果が高くなる傾向が見受けられた．しかしその差はわずかであり，選択手法よりは，選択語数の影響が大きかった．

どの手法を用いても，バイグラム言語モデルを用いた場合には，連鎖語の追加語数の増加にともなって，パープレキシティも単語正解精度も改善した．トライグラム言語モデルを用いた場合には，連鎖語の追加語数500程度でテストセットパープレキシティは収束し，単語正解精度もピークとなった．連鎖語の語彙追加によって，バイグラム言語モデルのテストセットパープレキシティは，最高33%減少し，大語彙連続音声認識実験における単語誤り率は，最高21%減少した．トライグラム言語モデルの場合は，テストセットパープレキシティ，単語誤り率ともに，最高で19%減少した．

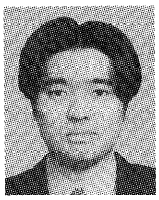
連鎖語追加を1000語追加したときのバイグラム言語モデルは，語彙追加のないトライグラム言語モデルを超える性能を実現した．連鎖語を追加すれば，バイグラム言語モデルで十分高い性能を実現できることは，トライグラムの実装にともなうデコーダの複雑化の問題を避けて通ることができるため，認識システムの簡素化の観点からも価値が高い．

謝辞 本研究には，CD-ROM版毎日新聞記事データ，RWC形態素解析ツールキット，ASJ新聞記事読み上げ音声コーパスを使用した．記して，関係諸氏に感謝申し上げる．

参考文献

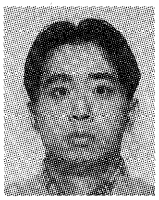
1) Deligne, S. and Bimbot, F.: Language modeling by variable length sequences: theoretical formulation and evaluation of multigrams, *IEEE Proc. ICASSP95*, pp.169–172 (May 1995).
2) Masataki, H., Sagisaka, Y., Hisaki, K. and Kawahara, T.: Task adaptation using MAP estimation in N-gram language modeling, *IEEE Proc. ICASSP97*, Vol.2, pp.783–786 (Apr. 1997).
3) Simons, M., Ney, H. and Martin, S.: Distant bigram language modelling using maximum entropy, *IEEE Proc. ICASSP97*, Vol.2, pp.787–790 (Apr. 1997).
4) Matsuoka, T., Ohtsuki, K. Mori, T. and Furui, S.: Japanese large-vocabulary continuous speech recognition using a business-newspaper corpus, *Proc. ICSLP 96*, pp.22–25 (Sep. 1996).
5) Lee, A., Kawahara, T. and Dohshita, S.: An efficient two-pass algorithm using word trellis index, *Proc. ICSLP98*, pp.1831–1834 (Dec. 1998).
6) Masataki, H. and Sagisaka, Y.: Variable-Order N-gram Generation by Word-Class Splitting and Consecutive Word Grouping, *IEEE Proc. ICASSP96*, Vol.1, pp.188–191 (May 1996).
7) 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（D-II），Vol.J79-D-II, No.12, pp.2062–2069 (1996).
8) 森 信介，山地 治，長尾 眞：予測単位の変更による n-gram モデルの改善，信学技報，NLC97-48, SP97-81, pp.35–42 (Dec. 1997).
9) Giachin, E.: Phrase bigrams for continuus speech recognition, *IEEE Proc. ICASSP95*, pp.225–228 (Apr. 1995).
10) Lee, C.H. and Rabiner, L.R.: A Frame-Synchronous Network Search Algorithm for Connected Word Recognition, *IEEE Trans. Acoust., Speech, and Signal Proc.*, Vol.37, No.11, pp.1649–1658 (1989).
11) Ney, H., Haeb-Umbach, R., Tran, B.-H. and Oerder, M.: Improvements in beam search for 10000-word continuous speech recognition, *Proc. ICASSP92*, Vol.I, pp.9–12 (Mar. 1992).
12) 渡辺隆夫，吉田和永，畑崎香一郎：バンドルサーチ法を用いた連続音声認識の高速化，信学論，J75-D-II, No.11, pp.1761–1769 (1992).
13) 伊藤克亘，速水 悟，田中穂積：音素文脈依存モデルと高速な探索手法を用いた連続音声認識，信学論，J75-D-II, No.6, pp.1023-1030 (1992).
14) 中野裕一郎，小林哲則：ビームサーチとバンドルサーチを併用したフレーム同期型連続音声認識における高次言語モデルと音素環境依存型音響モデルの簡易実装とその効果，音講論 (Mar. 1999).
15) Itou, K., Yamamoto, M., Takeda, K., Takezawa, T., Matsuoka, T., Kobayashi, T., Shikano, K. and Itahashi, S.: The design of the newspaper-based Japanese large vocabulary continuous speech recognition corpus, *Proc. ICSLP98* (Dec. 1998).
16) 小林紀彦，中野裕一郎，肥田木康明，小林哲則：統計的言語モデルにおける付属語の扱いに関する一考察，音講論，2-1-6, pp.59–60 (Sep. 1997).

(平成 10 年 10 月 6 日受付)
(平成 11 年 2 月 8 日採録)



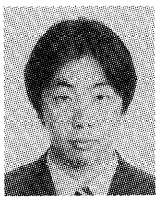
和田 陽介

1975 年生．1998 年早稲田大学理工学部電気電子情報工学科卒業．現在同大大学院修士課程在学中．大語彙連続音声認識における言語モデルの研究に従事．日本音響学会会員．



小林 紀彦

1973 年生．1997 年早稲田大学理工学部電気電子情報工学科卒業．現在同大大学院修士課程在学中．大語彙連続音声認識における言語モデルの研究に従事．日本音響学会会員．



中野裕一郎

1974 年生．1997 年早稲田大学理工学部電気電子情報工学科卒業．現在同大大学院修士課程在学中．大語彙連続音声認識におけるデコーダの研究に従事．日本音響学会会員．

小林 哲則（正会員）

1957 年生．1980 年早稲田大学理工学部電気工学科卒業．1985 年同大大学院博士課程修了．工学博士．同年法政大学講師．同助教授，早稲田大学助教授を経て，現在早稲田大学理工学部電気電子情報工学科教授．音声言語処理，動画像処理，知能ロボット等の研究に従事．電子情報通信学会，日本ロボット学会，人工知能学会，日本音響学会，IEEE，ACM 等会員．