

連続音声認識における未知語の扱い Processing Unknown Words in Continuous Speech Recognition

伊藤克亘†

速水 悟‡

田中穂積†

ITOU Katunobu HAYAMIZU Satoru TANAKA Hozumi,

†東京工業大学工学部

Tokyo Institute of Technology

‡電子技術総合研究所

Electrotechnical Laboratory

あらまし 連続音声認識システムを実際に構築する場合には、さまざまな制約から認識する対象の文法や語彙を制限せざるをえない。しかし、システムの柔軟性を増すうえで、未知語が含まれる入力を許容することは不可欠である。連続音声認識システムの未知語が含まれる入力の扱いには、入力中の未知語の有無の検出から、その未知語の意味の推定まで、さまざまなレベルの対応が考えられる。本論文で提案する手法では、未知語が含まれる入力のうち、未知語以外の部分は、辞書や文法を用いて正しい解析をおこなう。一方、未知語の部分については、音韻タイプライターを駆動して実際の音素列に似た音素系列として検出し、その後は辞書引きに成功した語と同じように処理することができる。また、この手法を、連続音声認識システム niNja に導入して、その動作を確認した結果についても報告する。

Abstract This paper introduces a new processing technique of unknown words for a continuous speech recognition system. For a practical system, the user should be allowed to utter the speech with words outside the vocabulary in order to make it more flexible and convenient. There are many problems on processing unknown words in a continuous speech recognition system: the detection whether there exist any unknown words or not, the estimation of its meaning, and that of its transcription etc. A new algorithm to process unknown words is proposed. Parts of unknown words are processed by a phonetic typewriter and other parts are processed with the dictionary and the grammar. Unknown words detected can be processed same as the words in the dictionary in a higher level process. Preliminary results are shown using continuous speech recognition system niNja combined with the proposed method.

1 はじめに

われわれは、音声を含めた日本語による自然言語インタフェイスをめざすシステム niNja (Natural language INterface in JApanese) の試作を行っている [3, 4]。このシステムを実際のインタフェイス (例えば、音声対話システムなど) に応用しようとする、様々な面での限界から語彙を制限しなければならない。

ここで、例えば音声対話システムに応用した場合に、次のような局面ではインタフェイスとして何をすべきであろうか。

- 利用者が、うっかり、システムの辞書にない言葉を発話してしまった場合
- 利用者が言い誤りや言い直しをしてしまった場合

- 利用者がシステムに向かって発話したことがらから知識を獲得したい場合 (「お名前は?」とか「御住所は?」などの質問をシステムがする場合など)

これらの場合、インタフェイスがとるべき対応には以下に示すような様々なレベルが存在するだろう。

1. 未知語の存在を無視して、システムのもつ知識にあわせて強引に認識結果を出す (何も結果を出さない場合も含む)
2. 発話がシステムに扱える範囲にないことを利用者にして、その入力を拒絶する (利用者が発話し直させる場合も含む)
3. 未知語の部分は、とりあえずとばして他の部分はシステムの知識にあわせて正しく認識する

4. 未知語の部分も、とりあえずなんらかの音素系列として認識する
(未知語が固有名詞など、ほとんど名前を表すだけの場合には、この処理だけでも結構実用になると思われる)
5. 未知語の部分、他の部分の解析結果から適当に推論して、発話を適切に理解する

人間は、様々な知識や思い込みなどを用いて誤りをおそれず発話を認識するので、未知語が発話された場合にも、これらの対応が可能である。

本論文では、連続音声認識において未知語の検出を行う手法について提案する。この手法は、構文情報などの言語情報を用いながら、未知語が検出できるので、適当な知識を用いれば上で示した全てのレベルまで拡張できる。

2 従来の研究

未知語や不要語を含んだ発話を扱うのに適した枠組として、スポッティングに基づく手法が提案されている [6, 11]。この手法は、認識結果として、認識後に使う情報を含む語やフレーズ (いわゆるキーワード) だけからなる系列を得ようとする。したがって、1 で示した 3 番目のレベルの処理を行うことが可能である。しかし、スポッティングすべき語やフレーズの数が増えてくると、未知語を含まない発話を入力とした場合でさえ、他の単語の部分との混同などの誤りがおこることが知られている [8]。したがって、未知語や「不要語」を含む発話の場合にも同じような誤りをおこすことが考えられる。また、未知語が本当にシステムで扱う語を部分列として含む場合もあり、そういった場合に適切な扱いをするためには、かなり高度な後処理が要求されるのではないだろうか。また、この手法だけでは、未知語がどんな音素系列か推測することはできない。

スポッティングではなく連続音声を認識する手法としては、辞書や文法といったさまざまな知識を用いて一字一句正確に発話を認識しようとする方法がある。こういった手法で未知語を含んだ連続音声の発話を扱う手法としては、Asadi らの方法が提案されている [1]。日本語に関しては、連続音声の発話を扱う手法はほとんど提案されておらず、いくつか関連した研究があるだけである [12, 10]。

Asadi らの手法 [1] は、未知語用の単語モデルを用意し、システムで扱える文型を記述したテンプレートの固有名詞の部分で、その辞書に登録されている単語のモデルに加えその単語モデルも駆動して認識することで、未知語を検出する。未知語用の単語モデルは 2 つ以上の任意の音素文脈独立音韻モデルがつながるような構造になっており、このモデルを用いれば実質的にはほとんど全ての単語を認識することができる。この手法を用いると、発話が未知語を含むかどうかの検出はできるが、未知語がどのような系列であるかの推定はできない。

北ら [12] は文節ごとに発話した音声を対象として、辞書を用いるシステムと並行して音韻タイプライター [2] を駆動することで未知語を扱う枠組を提案している。しかし、このような手法では、音韻タイプライターによる実際の発話とは異なる誤った認識結果が、辞書などを使ってえられた

正しい認識結果よりもよいスコアになることもあるが、その対処法については述べられていない。また、音韻タイプライターを用いるため、未知語がどのような音素系列であるかを推定することが可能であるが、音韻タイプライターの処理は (特に、正確さを重視する場合には)、非常に計算量が多いので、つねに並行して処理をおこなうというのは、実際のシステムを考慮すると現実的ではなく、何らかの効率的な駆動方法が必要であろう。また、未知語が含まれる場合は、そもそも文節を区切るのが難しいので、文ごとの発話への拡張をおこなう場合には、音韻タイプライターの駆動方法について考慮する必要があると考えられる。

一方、塚田ら [10] は、単語音声に対して半音節のモデルを連結して認識することによって求めた参照用尤度と、辞書を用いて認識をおこなった尤度を比較して、その差が大きいときには、未知語が発話されていると判断する手法を提案している。しかし、参照用の尤度がどの程度の処理量で求められるかは明記されていない。

本論文では、実際の連続音声認識システムで、未知語のない発話に対しては、それほど処理量を多くすることなく未知語の音素系列を推定しながら検出できる手法について述べる。

3 未知語の処理

3.1 未知語を含んだ入力の形態素・構文解析

本研究では、未知語を検出しながら解析をすすめる手法として、堀内らの構文 (形態素) 解析手法 [13] を用いる。この手法は、構文など上位の情報を利用して、バックトラックせずに未知語の検出をおこなえるので、フレーム同期で処理をすすめる miNja のような音声認識システムへの応用が容易である。

この手法は、構文解析の結果を利用して辞書引きをおこなう。構文解析プロセスも辞書引きプロセスも LR 構文解析法を用いて実現するが、辞書引きプロセスは、未知語処理に対応するためレデュースのときは先読みを行わないように拡張した LR 表を用いる (この辞書のことを LR 辞書と呼ぶ)。未知語処理は、未知語となりうる構文カテゴリ (品詞) を指定するとそのカテゴリの辞書引きプロセスに対してだけおこなわれる。こういった処理を可能にするため、辞書は構文カテゴリごとに用意する。

ここでは、「自由が丘に行く」というテキスト入力を、次に示す文法と辞書を用いて解析する場合を例として説明する。

文	→ 名詞句 動詞句	(1)
文	→ 動詞句	(2)
名詞句	→ 名詞 助詞	(3)
動詞句	→ 行く	(4)
名詞	→ 自由 丘	(5)
助詞	→ が に	(6)

名詞は未知語の可能性有

未知語処理は以下のように行う。

1. それまでの構文解析結果にしたがって、次に予測されている構文カテゴリに対応した辞書引きプロセスが駆動される。このとき、未知語となりうる構文カテゴリの辞書引きプロセスでのみ、未知語処理がおこなわれる。例の文法では、名詞だけが未知語となりうるので、解析中に (1) の規則を使うときには、未知語処理をおこなうが、(2) の規則を使うときは、未知語処理をおこなわない。

2. 辞書引きプロセスでは、入力にしたがって LR 表の状態遷移を行う。このとき、すべての未知語を検出するために、LR 表の遷移に成功するしないにかかわらず未知語の処理を行う。未知語の処理は以下のようにする。

(a) まず、それまでの未知語の候補にそのときの入力をつないで未知語の候補の文字列を長くする。

例の場合には、まず、'自' という文字が入力されるので、名詞の「自由」の辞書引きプロセスが自' のところまで成功するが、それとは別に、未知語用に "自" という文字列を構成する。

(b) 辞書引きプロセスのレデュースとシフトにあたる動作を両方行う。レデュースにあたる動作では、そこまでの文字列から構成される未知語をつくる。この未知語は、辞書引きプロセスを駆動したときの構文カテゴリである語として、同じ構文カテゴリの辞書に登録されている語と同じように解析をすすめられる。シフトにあたる動作では、どんどん長い未知語を構成する。

例の場合には、レデュースにあたる動作として「自」という未知語である名詞をつくり、構文解析をつづける。しかし、これが名詞だとすると、規則 (3) より、次に助詞がくるはずだが、実際には、'由' がくるので、文法にあわないため棄却される。シフトにあたる動作としては、次の入力に連結させて長い未知語の候補を作るために "自" という文字列をつくる。

3. 入力なくなるまで 1, 2 をくりかえす。

このアルゴリズムを用いると、例の場合では、

自由 が 丘 に 行く。
自由が丘 (未知語) に 行く。

という 2 つの解析結果がえられる。このように、この解析法では、バックトラックなしに、正しく解析される文字列を部分的に含む未知語がある場合にも未知語を検出しながら構文解析ができる。しかし、この手法には、解析途中で未知語を含んだ解析結果の候補をたくさん生成してしまうという問題点がある。この問題点については、後で述べるような知識を使って枝刈りをおこなうことで、もっともらしい候補を残しながら処理量を減らせば実用上は問題がないと考えられる。

3.2 連続音声認識における未知語を含んだ入力の処理

この手法を連続音声認識へ応用するためには、テキスト入力の場合の「入力の文字」にあたるものを構成しなければならない。

本手法では、「音韻タイプライター」を用いて「入力の文字」にあたる音素系列を構成する。ここで用いる「音韻タイプライター」は [2] と基本的には同じだが、音素が連鎖するかどうかは文法ではなく表を用いている。また、音節 N-gram ではなく音韻 N-gram を用いている。

3.2.1 音韻タイプライター

niNja では、辞書・文法と音韻連鎖統計情報を用いて認識中の仮説の生成・棄却をおこなう [5]。このうち、辞書・文法の部分を日本語としてありうる音韻連鎖の規則に代えると音韻タイプライターとして利用できる。

niNja では、音素文脈依存モデルをあつかうために、先行 2 音素から次に許される音素を一覧にした表を用いている [4]。この表を音韻連鎖の規則として用いる。

本手法では、音韻タイプライターを未知語の有無の検出だけに使うのではなく、未知語がどのような音素の系列かを推定するのにも用いる。したがって、音韻タイプライターの認識精度をなるべく向上させるため必要がある。本手法では、なるべく正しい音素の系列をえるために音韻連鎖の統計情報を用いる。具体的には、音韻連鎖の N-gram モデルとよばれる統計モデルを用いる [5]。

niNja では、それぞれの認識候補について、次の式でスコアを求める。

$$P_{total} = P_{hmm} + w_{gram} P_{gram}$$

ただし、

$$P_{hmm} = \frac{\sum_{t=1}^{N_{frame}} \log P_{hmm_t}}{N_{frame}}$$

$$P_{gram} = \frac{\sum_{p=1}^{N_{phon}} \log P_{gram_p}}{N_{phon}}$$

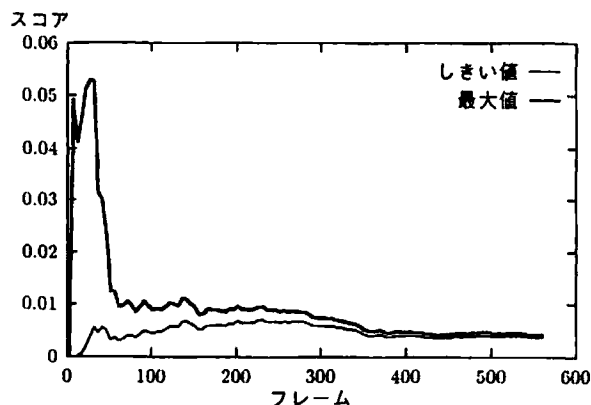
ここで、 N_{frame} はその時点でのフレーム数、 N_{phon} はその認識候補の音韻数、 P_{hmm_t} は各フレームでの音韻モデルと入力を照合してえられるスコア、 P_{gram_p} は各音韻での音韻 N-gram のスコアである。また、 w_{gram} は N-gram モデルの重みである。本稿では、この値を P_{hmm} と P_{gram} のダイナミックレンジを考慮して適当に決める。この式を用いて枝刈りをおこなうことで、音素系列として日本語らしい候補を残すことができるため、処理量をおさえながら認識の精度をあげられる。

3.2.2 音韻タイプライターの制御

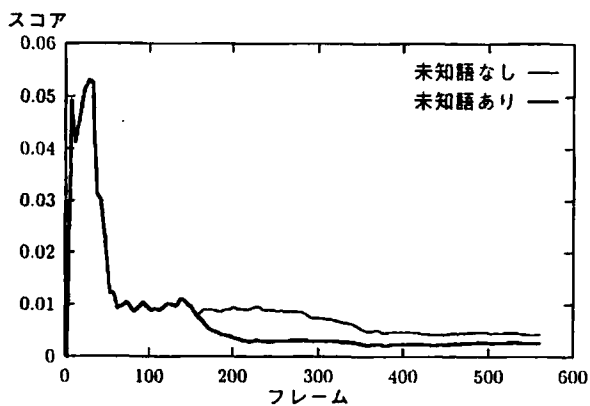
音韻タイプライターも、niNja の枠組みで動作するが、処理途中で候補数が非常に多くなる。実際に連続音声認識システムとして利用することを考えると、未知語処理は辞

書が引けないときに重点的に動作し、辞書が引けるときには余り処理量が多くないことが望ましい。

niNja では、発話がシステムの文法と辞書であつかえる範囲であるとき、それぞれのフレームで最もよいスコアを持つ認識候補(フレームによっていれかわることもある)のスコアをフレーム数で正規化すると、次の図の太い線のように変化する。本システムでは、このスコアから各フレームごとに枝刈りのスコアを決めて、認識候補の枝刈りをおこなう。この枝刈りのスコアは細い線のように、発話がすすむと範囲がせばまるように変化する。



このとき、発話に含まれる単語を辞書から除いて(つまり、システムの文法と辞書ではあつかえない状態にして)無理に認識すると、それぞれのフレームで最もよいスコアを持つ認識候補の正規化されたスコアは、次の図の太い線のように変化する(細い線は、正しく認識した場合)。つまり、除かれた単語(未知語)の部分(図では、200フレームを中心とした100フレームくらいの区間)では、正しく認識できる場合よりも、かなりスコアが悪くなるので傾きが急になる。また、除かれた単語の部分が終わると正しく認識できる場合と同じような傾きになる。

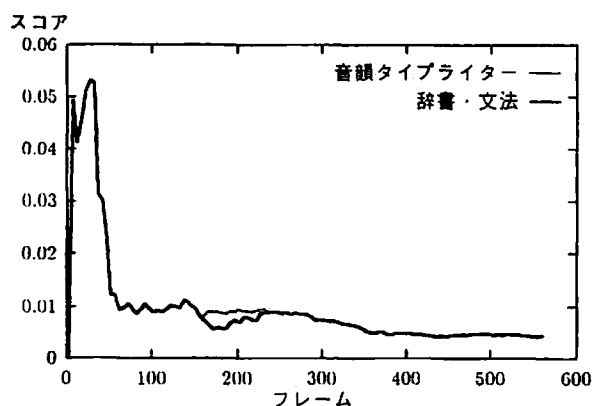


音韻タイプライターを使って認識した場合、それぞれのフレームで最もよいスコアを持つ認識候補のスコアは、(どのような辞書や文法を使っているかには関係なく)正しく認識した場合のスコア以上で、それほど差のないスコアになる。(音韻タイプライターは日本語としてありうる任意の音素系列を生成できるので、正しく認識した候補と同じ系

列になる候補も生成する。)

3.1 に示したアルゴリズムでは少なくとも、未知語のためにひとつの系列が残れば未知語の検出は可能である。したがって、各フレームで辞書引きプロセスが生成している最もよいスコアを持つ認識候補(それまでの部分には未知語が含まれていてもよい)のスコアをしきい値として、それ以上のスコアを持つ認識候補を残すように音韻タイプライターを駆動する。

この手法で、発話に含まれる単語を辞書から除いて認識すると、しきい値は次の図の太い線のように変化する。音韻タイプライターのスコアは細い線のように変化する。(しきい値が、未知語の途中からよくなるのは、タイプライターが未知語を完成させたあとは、また、辞書と文法を用いた解析にもどって、正しい解析をするからである。)



このように制御すると、音韻タイプライターは、発話の未知語の部分では多くの候補を生成するが、未知語でない部分では、それほど多くの候補を生成しない。したがって、無駄な処理を極力さけながら、未知語の部分については、音素系列として正しい解析をおこなうことが期待できる。

3.2.3 アルゴリズム

3.2.2 で述べた制御を用いて、niNja では以下の方法で、辞書・文法を用いた処理と音韻タイプライターをくみあわせる。

1. (音韻タイプライターの初期化) それまでの構文解析結果にしたがって、次に予測されている構文カテゴリに対応した辞書引きプロセスが駆動される。このとき、未知語となりうる構文カテゴリに対しては、通常の辞書引きプロセスに並行して、音韻タイプライターも駆動される。それ以外の構文カテゴリについては、未知語処理をおこなわない場合と同じように解析する。

音韻タイプライターは、それぞれの候補について次のように動作する。

2. それまでの音素系列から駆動できるすべての音韻モデルを 3.2.1 で述べた一覧表で調べ、1 音韻からなる未知語(の部分)用の系列とする。これらの候補は、該当する構文カテゴリに属する辞書に登録されていない語としてあつかわれる。

3. 音韻タイプライターは、ひとつの音韻モデルの照合をおえると、その音韻モデルを含むすべての候補について、それぞれ 3.2.1 でのべたスコアを計算する。そして、そのフレームで、辞書を用いて認識している最も大きなスコアの候補のスコアと比較して、小さければ棄却し、大きければレデュースとシフトの両方の動作をおこなう。レデュースにあたる動作は、そのフレームまでに音韻タイプライターが連結して生成した音素系列をもち、初期化したときにわりあてられた構文カテゴリをもつ未知語を完成させる。以後、この語は、未知語であるということのぞけば、同じ構文カテゴリを持つ辞書に登録されている語と同じようにつかわれる。シフトにあたる動作は、それまでの音素系列から駆動できるすべての音韻モデルをさらに連結して、より長い未知語を生成する。

4. 3 をくりかえす。

ここで用いる音韻タイプライターは、日本語としてあらゆる任意の音韻のくみあわせで候補を生成するので、候補数は非常に多くなる。しかし、おなじ音韻モデルの照合は、まとめて行う [3] ので、照合のための処理量はそれほど増えない。

4 実験

実際に文法・辞書を用いたシステムに本論文の手法を導入して、動作を確認し、処理量などを調べた。

4.1 音声資料

実験に用いた HMM の訓練用音声資料は単語音声と連続音声からなる。単語音声資料の話者は成人男性 5 名で、発声用テキストは音韻バランス単語集合 WD-II (1542 語) [9] である。連続音声資料の話者は成人男性 2 名で、発声用のテキストは ATR 音韻バランス文 150 文である。これらの収録は簡易防音室でおこなった。標本化周波数 15 kHz で A/D 変換を行なった。分析のフレーム周期は 5 ms である。14 次のメルケプストラム係数と、その時間方向の変化量、パワーの時間方向変化量の合計 29 個のパラメータを 1 つのコード帳にベクトル量子化した。コード帳のサイズは 1024 である。

HMM のモデル構造は、すべて 4 状態 3 ループとし、各状態から出る弧の組をタイドアークとした。モデルの数は、43 個 (音素文脈独立) である。

音韻連鎖の統計モデルとしては、トライグラム (N-gram の $N=3$ のもの) を用いた。このモデルは、日本経済新聞 1982 年の新聞記事 19 日分と国際会議に関するキーボード会話・電話会話からなるデータベースから生成した。このデータベースには、合計 11207 文、859311 音韻が含まれている [5]。

実験に用いた音声資料は、防音室または計算機室で収録したもので、これらの話者・テキストは HMM・音韻トライグラムを訓練した資料には含まれておらず、不特定話者・語彙独立な実験条件とした。

4.2 実験結果

実験には、以下に示すような文テンプレートを 11 含む文法を用いた。

<代名詞・場所> <に> <本> <が> <何冊> <あります>

<代名詞・場所> = ここ|そこ|あそこ|どこ

<本> = 本|マンガ|雑誌|アルバム

ここで <> でかこまれた名前は、非終端記号をあらわす。

例に示した文テンプレートを使って解析できる「ここに本が一冊あります。」という文を、上に示した <本> というカテゴリから「本」を除いて、未知語処理をおこなって認識した。<本> というカテゴリは、5 つのテンプレートにあらわれる。

2 人の異なる話者による認識結果を下に示す。(それぞれ 1 位と 2 位のものを示す。)

話者 s0007.00

[1] 5.00e-3 ここに h o N (未知語) が一冊あります。

[2] 4.63e-3 ここに h o N a (未知語) が一冊あります。

話者 s0044.00

[1] 5.80e-3 ここに h o N (未知語) が一冊あります。

[2] 5.45e-3 ここに h o N (未知語) が二冊あります。

次に、もともとの「本」を含む辞書を用いて未知語処理をおこなって上と同じ発話を認識した結果について示す。

話者 s0007.00

[1] 5.00e-3 ここに本が一冊あります。

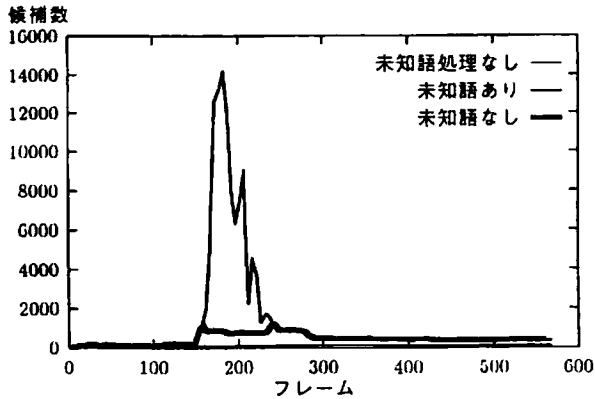
[1] 5.00e-3 ここに h o N (未知語) が一冊あります。

話者 s0044.00

[1] 5.97e-3 ここに本が一冊あります。

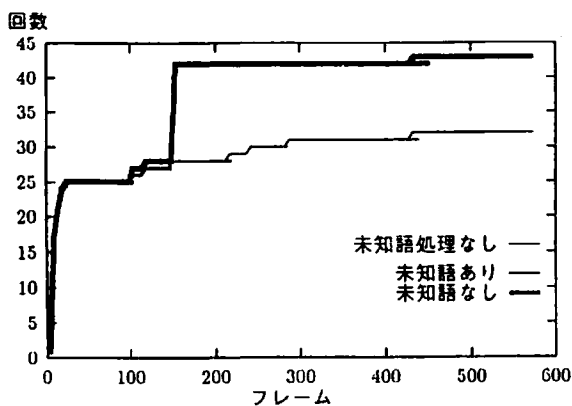
[2] 5.63e-3 ここに本が二冊あります。

処理量がどの程度ふえたかを調べるため、話者 s0007.00 について上の 2 つの場合と未知語処理をおこなわず「本」を含む辞書を用いて認識した場合の途中で生成される候補数の変化を次に示す。図の「未知語処理なし」とは、未知語処理をおこなわず、「本」を含む辞書を用いて認識した場合、「未知語なし」とは、未知語処理をおこない、「本」を含む辞書を用いて認識した場合、「未知語あり」とは、未知語処理をおこない、「本」を含まない辞書を用いて認識した場合を示す。



「本」を除き発話に未知語が含まれるようにして認識すると、未知語の存在する付近で、活発に音韻タイプライターが動作して、候補数が非常に多くなっている。しかし、発話に未知語が含まれないようにして認識すると、しきい値が高いので、未知語が含まれるようにしたときに比べると候補数はかなり少ない。つまり、実際に未知語処理が必要とき以外は、未知語処理による余分な処理はそれほどおこっていない。

次に、音韻モデルの照合回数の変化を示す。



音韻モデルの照合はまとめておこなうので、未知語処理をおこなっても照合回数は候補数ほど多くなりません。

5 検討課題

動作結果にはあらわれていないが、未知語処理をおこなうと、未知語を含まない発話の認識結果として、未知語を含む結果が、辞書・文法を用いて認識された正しい結果よりもよいスコアでえられる可能性がある。この誤りをふせぐ簡単な方法としては、未知語を含む認識結果については、未知語を全く含まない結果と比べて、ある一定のしきい値以上の差がある場合だけ認識結果として採用するという方法がかんがえられる [1, 10]。しかし、人間が発話に含まれる未知語を検出するときには、意味的な知識をはじめとする様々な知識を使っていると考えられる。niNja は、認識中の各フレームで認識候補を生成・棄却するとき、言語的な情報を利用できるシステム構成になっている。したがっ

て、今後は、例えば、どのようなカテゴリの語が未知語になりやすいかなど、未知語の検出に役立つ知識を解明し、実際に、システムに組み込んで、未知語の検出の精度をあげることができるかを調べたい。

この枠組で、言い誤り・言い直しにより生じる文法を逸脱したフレーズ、いわゆる不要語、未知語などの検出と音素系列の推定は可能である。しかし、発話の中での役割は、それぞれ異なるので、発話を理解するためには、これらを区別して扱う必要がある。人間は、これらのフレーズを検出した瞬間にこれらの区別もある程度できていると考えられるので、そういった区別に役立つ知識を明らかにしていく必要があるだろう。

6 むすび

以上、連続音声認識での未知語の処理手法について提案した。われわれは、インタフェイスとしての連続音声認識システムでは、語彙や文法の制限があっても、その制限が利用者に様々な負担をあたえないようにすべきだと考える。そういった負担のひとつを解消するために、未知語を含んだ発話が適切に扱えることを重要視している。

また、語彙や文法の制限もなるべく自然におこなえるよう、実際の音声対話システムを用いたデータをもとに、段階的にシステムを拡張することを目指している [7]。したがって、なるべく自動的に語彙の拡張を行うためにも、システムが発話から自動的に未知語の検出や推定をおこなえることは重要なことであると考えている。

今後は、より複雑なタスクでの実験をおこない、提案した手法の評価・改善をおこなう予定である。

謝辞

本稿で使用したテキストデータベースのうち、日本経済新聞の記事については、NTT 情報通信処理研究所メッセージシステム研究部から、国際会議に関するものについては、ATR 自動翻訳電話研究所から提供いただきました。貴重なデータを使用させていただいたことを感謝いたします。また、連続音声資料は日本音響学会連続音声データベース調査委員会の研究用連続音声データベースの一部であり、関係各位のご尽力に感謝いたします。

また、日頃御討論いただく、東工大田中研の皆様、並びに電総研知能情報部音声研究室の皆様にご感謝いたします。

参考文献

- [1] A. Asadi, R. Schwartz, and J. Makhoul. Automatic modeling for adding new words to a large-vocabulary continuous speech recognition system. In *Proc. ICASSP-91*, pp. 305-308, 1991.
- [2] T. Kawabata, T. Hanazawa, K. Itou, and K. Shikano. Japanese phonetic typewriter using HMM phone recognition and stochastic phone-sequence modeling.

IEICE Transactions, Vol. E 74, No. 7, pp. 1783-1787, 1991.

- [3] 伊藤克亘, 速水悟, 田中穂積. 拡張 LR 構文解析法を用いた連続音声認識. 信学会技術報告, Vol. SP90-74, pp. 49-56, 12 1990.
- [4] 伊藤克亘, 速水悟, 田中穂積. 音素文脈依存 HMM を用いた連続音声認識. 日本音響学会講演論文集, pp. 41-42, 1991.
- [5] 伊藤克亘, 速水悟, 田中穂積. 連続音声認識システム NINJA での音韻連鎖統計情報の利用. 人工知能学会第 5 回全国大会, pp. 599-602, 1991.
- [6] 岡隆一. 文スポッティング型連続音声認識システムによる会話音声認識. 日本音響学会講演論文集, pp. 25-26, 10 1991.
- [7] 速水悟, 伊藤克亘, 田中和世. 音声対話システムの構築とそれを用いた会話音声収集. 信学会技術報告, 12 1991.
- [8] 速水悟, 岡隆一. 連続 DP による連続単語認識実験とその考察. 信学論 (D), Vol. J67-D, No. 6, pp. 677-684, 1984.
- [9] 速水悟, 田中和世, 横山晶一, 太田耕三. 研究用音声データベースのための VCV/CVC バランス単語セットの作成. 電総研彙報, Vol. 49, No. 10, pp. 803-834, 1985.
- [10] 塚田聡, 渡辺隆夫, 吉田和永. 未知語検出・リジェクションのための音声認識の尤度補正. 日本音響学会講演論文集, pp. 203-204, 1991.
- [11] 坪井宏之, 橋本秀樹, 竹林洋一. 連続音声理解のためのキーワードラティスの解析. 日本音響学会講演論文集, pp. 21-22, 10 1991.
- [12] 北研二, 江原暉将, 森元逞. 連続音声認識における未知語処理. 日本音響学会講演論文集, pp. 93-94, 1991.
- [13] 堀内靖雄, 伊藤克亘, 田中穂積. 拡張 LR 構文解析アルゴリズムによる未定義語を含む日本語文の構文解析. 情報処理学会 第 40 回全国大会, pp. 325 - 326, 1990.