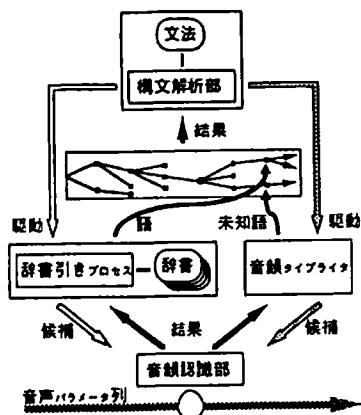


◎ 伊藤克亘 △田中穂積 (東京工業大学)
速水 悟 (電子技術総合研究所)

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ースをめざすシステム niNja (Natural language INterface in JApanese) の試作を行っている [1, 2]。本稿では、このシステムを実際のインタフェースに応用するときに問題となる未知語の処理の手法と、その手法を用いた実験結果についてのべる。ここで提案する手法では、連続音声の中の未知語を検出し、その語を構成する音韻系列の推定もできる。

1 システムの構成

本システムでは、次の図に示すように、構文解析部が音韻タイプライタと辞書引きプロセスを並行して駆動することで、辞書を用いた認識をおこなうと同時に未知語の推定をおこなう。



未知語処理は以下のようにおこなう。

構文解析プロセスが、あらかじめ文法で指定された未知語となりうるカテゴリを予測すると、その辞書と並行して、未知語の音韻系列を推定するための音韻タイプライタを駆動する。このタイプライタは、辞書引きプロセスが処理する候補のスコアをしきい値として駆動されるので、システムが辞書に登録されている語を認識しているときは、ほとんど候補を生成しない。しかし、辞書に登録されてない語を認識しているときには、辞書引きプロセスが生成する候補のスコアは低くなるので、音韻タイプライタのしきい値も小さくなり、生成する候補数も多くなる。音韻タイプライタは未知語をつぎつぎに生成して、駆動するときに設定されたカテゴリの語として構文解析プロセスに送る。未知語は、構文解析以上のレベルでは同じカテゴリの辞書に登録されている語と同じにあつかわれる。

同様に、音韻タイプライタと辞書を併用して未知語処理をする手法が北らによって提案されている [3] が、その手法では未知語の検出については明らかでない。また、その手法と比較すると、本手法では、辞書引きプロセスが生成した候補と音韻タイプライタが生成する候補をひとつにまとめて制御しているので、無駄な処理を

削減でき、動的に未知語の可能性が高い区間で重点的に音韻タイプライタを駆動することができる [2]。

2 実験

2.1 音声資料と言語データ

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) 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。HMM のモデルの数は 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。

実験に用いた音声資料の発声用テキストは 11 文 (文節数は 33) の疑問文などである。このテキストを成人男性の 2 名分を防音室で、8 名分を計算機室で収録した。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。

2.2 認識実験

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|辞典

ここで <> でかこまれた名前は、非終端記号をあらわす。この文法の総単語数は 113、文節の平均分岐数は 8.2 である。なお、文節間には任意の無音区間を許す。

この 11 のテンプレートのうち 10 個には名詞が含まれている。その全てのカテゴリを未知語になりうるとする。この 10 個を 5 個づつふたつのグループに分け、それぞれのグループについてその部分で実際に発話された名詞をそのカテゴリから除いて人工的に未知語が出現するようにする。例えば、上のテンプレートで「ここに本が一冊あります。」という文を認識するときに、カテゴリ <本> の辞書から「本」という語をとりぞいて、未知語が含まれるようにする。こうすると、一度の実験で、話者一人の 11 文のうち 5 文がそれぞれひとつずつ未知語を含むようになる。このような条件で、グループごとに 110 文、のべ 220 文 (うち、未知語を含む文は 100 文) を認識した。

認識実験の結果は、文認識率と文節の挿入・脱落を考慮した文節認識率で評価する。文節認識率は、次のように求める。まず、認識結果の文節系列と正しい文節系列との DP マッチングを行い、文節の置換・挿入・脱落誤りの個数を求める。これらの個数から、次の式を用いて文節認識率を計算する。音韻認識率も同じように求

*Processing Unknown Words in Continuous Speech REcognition System: niNja
By ITOU Katuobu, TANAKA Hozumi (Tokyo Institute of Technology) and HAYAMIZU Satoru (Electrotechnical Laboratory)

める。

$$\text{文節認識率} = \frac{\text{全文節数} - \text{置換} - \text{挿入} - \text{脱落}}{\text{全文節数}}$$

まず、未知語の検出にどのくらい成功したかを調べた。

認識成功		認識失敗		
推定	検出	推定	検出	失敗
13	55	1	6	25

「推定」は未知語の検出に成功し、かつ音韻系列の推定にも成功した文、「検出」は未知語の検出に成功したが、音韻系列の推定には失敗した文、「失敗」は、未知語の検出に失敗した文をあらわす。

未知語を含んだ 100 文のうち、文の認識に完全に成功した(つまり、未知語の音韻系列を正しく推定できた)文は 13 文あった。未知語以外の部分については辞書を用いて正しく認識でき、未知語の部分の推定を誤った文は、55 文あった。つまり、検出のレベルでは、未知語を含む文のうち 68% が正しくあつかえた。

未知語の推定を誤った例を次に示す。

正解	推定結果
hy o- sh i (表紙)	ky o sh i N, o sh i N
n e d a N (値段)	g e d a N, m u d a N
ch o sh a (著者)	ch o sh i o, ky o sh a i
k o t o (こと)	k o t o u, g o u t o u
sh a sh i N (写真)	s a sh i N, s a s u gi N
h o N (本)	h o o N, h o o-

目立った誤りとしては、母音や撥音が実際よりも余分につながってしまうものがあつた。この誤りは、音韻の継続時間長を制御することで防げるだろう。正しく検出された未知語部分の音韻認識率は 50.8% だった。

実際のインタフェイスとして利用する場合のほとんどが、未知語に関しては未知語として検出できれば音韻列の推定が少々不正確でも十分だと考えられる。そこで、未知語を含む文の場合は、未知語以外の部分については辞書を用いて正しく認識でき、未知語の部分では検出さえ正しくできれば、その文節とその文節を含む文の認識は正しいとみなして認識率を計算した。結果を次に示す。

認識率 (%)	未知語なし	未知語あり		
		処理なし	処理あり	
			推定	検出
文	85.5	46.4	41.4	66.4
文節	93.6	68.9	70.3	78.6

「未知語なし」は、辞書から何ものぞかず未知語処理をおこなわずに、認識した場合である。「処理なし」は、辞書から名詞をとりのぞいて未知語がある状態で、未知語処理をおこなわずに認識した場合であり、「処理あり」が本稿で提案した手法を用いた場合である。その中で、「推定」は、音韻系列まで正しいものだけを正解とした場合で、「検出」は上記のように検出に成功したもので正解とした場合である。この結果から、本稿で提案した手法は音韻系列まで完全に推定するのはかなり難しいが、検出するだけでよい場合にはかなり未知語をあつかえるといえる。

また、この実験では、未知語が出現する位置として文頭の場合も文中の場合もあつたが、場所によって検出や推定の精度が変化することはなかった。

3 むすびと今後の課題

今回の実験では、認識率からもあきらかに、未知語処理を行わないときは、正解がえられたのに、未知語処理をおこなうと未知語が含まれない文で正解がえられなくなるものがあった。その中で、未知語を含む結果が、辞書・文法だけを用いて認識された正しい結果よりもよいスコアになる場合があつた。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、ある一定の値以上大きい場合だけ認識結果として採用するという方法が考えられる。[4, 5]。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・棄却するときに、言語的な情報を利用できるシステム構成になっている [1]。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と期待したい。

また、認識を誤った他の場合の認識結果を調べると、実際には、未知語の区間の次の語の部分などを含んだ 10 音韻以上の音韻数の多い未知語を推定している場合が多い。例えば、「本を買いたいと思いますか」という句の「本」が未知語の場合に、「オーカイトエトイモエは薄いです」のように、「買いたいと思います」の部分まで未知語としてしまつて誤つた例がある。このような誤りを防ぐためには、音韻数の多すぎる未知語を抑制するようなヒューリスティクスを導入する必要がある。

さらに、本稿で提案した手法の未知語の検出の精度は、音韻タイプライタの精度に依存するので、音韻タイプライタの精度と、未知語の検出の精度の関係について明らかにする必要がある。

今後は、より複雑なタスクでの評価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していただきました。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。連続音声資料は日本音響学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へ感謝します。

参考文献

- [1] 伊藤他. 拡張 LR 構文解析法を用いた連続音声認識. 信学技報. SP90-74, pp.49-56, (1990-12).
- [2] 伊藤他. 連続音声認識における未知語の扱い. 信学技報. SP91-96, pp.41-47, (1991-12).
- [3] 北他. 連続音声認識における未知語処理. 音響講論, pp.93-94, (1991-3).
- [4] Asadi, et. al. Automatic modeling for adding new words to a large-vocabulary continuous speech recognition system. ICASSP-91, pp.305-308.
- [5] 塚田他. 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. 音響講論, pp.203-204, (1991-3).