TR-IT-0092

混合 N-gram を用いた認識システムの評価

伊田 政樹　　河井 淳　　脇田 由実

1995.2

Abstract

音声認識処理において、認識性能を向上させるため言語情報を利用する手法が研究されている。音声認識処理と翻訳処理を統合することを考えた場合、この双方で用いる言語モデルを統一せねばならない。しかしながら本研究室で用いているシステムでは、音声処理においては、認識性能の観点から単語 N-gram を、翻訳処理においては open な単語も扱えるように自立語は品詞、付属語は単語の N-gram (以下混合 N-gram と記す) を用いている。今後、音声認識部と翻訳処理部の言語モデルを統一化するために、翻訳処理部で用いている混合 N-gram を音声認識処理にも使用することで、N-gram を作成するにあたって用いたコーパスに対して単語 open な文の認識も可能にした場合の認識システムの性能を評価した。さらに、認識結果を翻訳した場合の翻訳正解率を確認し、TDMT でリカバー可能な誤認識の分析を行なった。

ATR Interpreting Telecommunications Research Laboratories

目次

混合 N-gram を用いた認識システムの評価

伊田政樹　河井淳　脇田由実

平成 7 年 2 月 23 日

1　はじめに

音声認識処理において、認識性能を向上させるため言語情報を利用する手法が研究されている。音声認識処理と翻訳処理を統合することを考えた場合、この双方で用いる言語モデルを統一せねばならない。しかしながら本研究室で用いているシステムでは、音声処理においては、認識性能の観点から単語 N-gram を、翻訳処理においては open な単語も扱えるように自立語は品詞、付属語は単語の N-gram (以下混合 N-gram と記す) を用いている。今後、音声認識部と翻訳処理部の言語モデルを統一化するために、翻訳処理部で用いている混合 N-gram を音声認識処理にも使用することで、N-gram を作成するにあたって用いたコーパスに対して単語 open な文の認識も可能にした場合の認識システムの性能を評価した。さらに、認識結果を翻訳した場合の翻訳正解率を確認し、TDMT でリカバー可能な誤認識の分析を行なった。

2　音声処理部および翻訳処理部の概要

本実験を行なうにあたり、使用した音声認識システムおよび翻訳処理システムについて述べる。

2.1　単語 N-gram を用いた音声認識

文音声認識を行なうにあたって、音響情報のみで高い認識性能を得ることは困難である。そこで、音声認識処理に統計的な言語情報を利用することで perplexity を低下させ、認識性能を向上させている。従来の音声認識では単語 trigram を用いて前の単語列から現在の単語に遷移する確率を求め、音声認識に利用している。単語の trigram を用いた音声認識システムのアルゴリズムは以下のようになる。[1]

言語情報として単語の trigram を用いた場合、求める解は、文候補 $l(w_1, w_2, \ldots, w_N)$ を以下のように定式化して、これを最大化する文 $w_1, w_2, \ldots, w_N$ を選び出すことである。

$$\sum \log(P_a(w_i)) + \alpha \times \sum \log(p(w_{i+2}|w_i, w_{i+1}))$$

ここで $P_a(w_i)$ は単語 w_i の音響尤度、$p(w_{i+2}|w_i, w_{i+1})$ は単語 w_i の次に w_{i+1} が現れた時に w_{i+2} に遷移する確率、α は音響尤度と言語との連鎖確率を結びつける結合定数である。認識単位を単語とした場合のアルゴリズムを表 1 に、システムの構成図を図 1 に示す。

本実験では単語 trigram を用いるところを混合 bigram・混合 trigram・品詞 bigram・品詞 trigram(混合 N-gram・品詞 N-gram に関しては後述) に置き換えて実験を行なった。

2.2　変換主導型翻訳 (TDMT)

本節では、経験的知識を活用して翻訳処理を行なう変換主導型翻訳 (Transfer-Driven Machine Translation , 以下 TDMT と記す) システムについて述べる。[2]

TDMT では、経験的知識として蓄積された用例の中から入力表現にベストマッチする用例を意味距離計算により求め、ベストマッチした用例の対訳情報を使って翻訳結果を作る。変換の経験的知識は言語表現の表現形式によりストリング・パタン・文法の 3 つのレベルに分類され、入力文の性質に応じて各レベルの変換知識を使い分けて翻訳処理を行なう。ストリングレベルの変換知識は表層語句のみで対応関係が表されたもの (例: ありがとうございました→ Thank you.)、パタンレベルの変換知識は文法属性を表現しない X のような記号と表層語句で表したもの (例: X は Y です→ X' be Y')、文法レベルの変換知識は品詞などの文法カテゴリーによって言語表現を記述したもの (例: 名詞 1 名詞 2 →名詞 2 of 名詞 1) である。

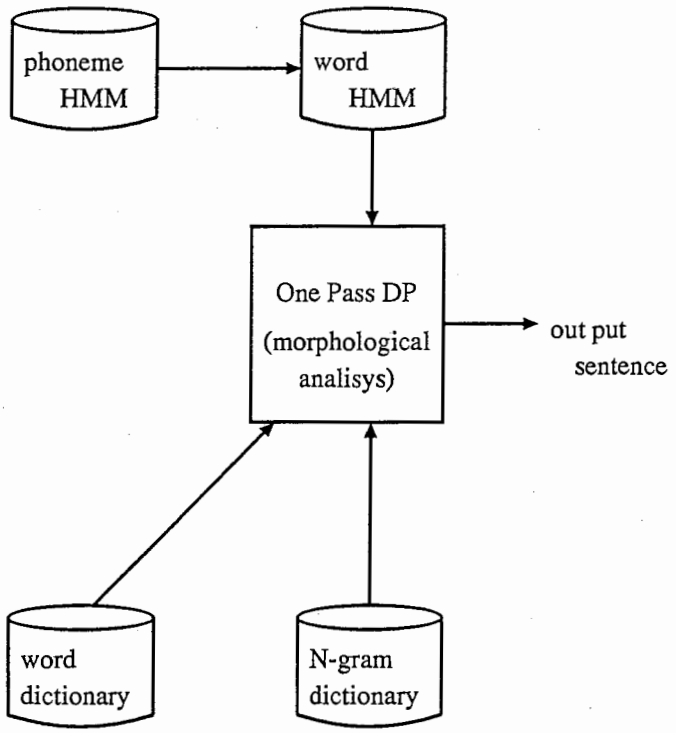

図 1 単語 N-gram を用いた音声認識システムの構成図

TDMT の基本構成は図 2 に示したように変換モジュールを中心に各モジュールが協調した処理を行ない翻訳結果をつくり出す。その処理の流れは次の通りである。

1) 入力文に対し、変換知識の原言語側を組み合わせた原言語構造を作る。
2) 原言語構造の部分構造ごとに意味距離計算に基づき最尤の部分構造へ変換し、目的言語構造を作る。
3) 2) で得られた目的言語構造の中から意味距離の総和に基づき最尤の構造を決定する。

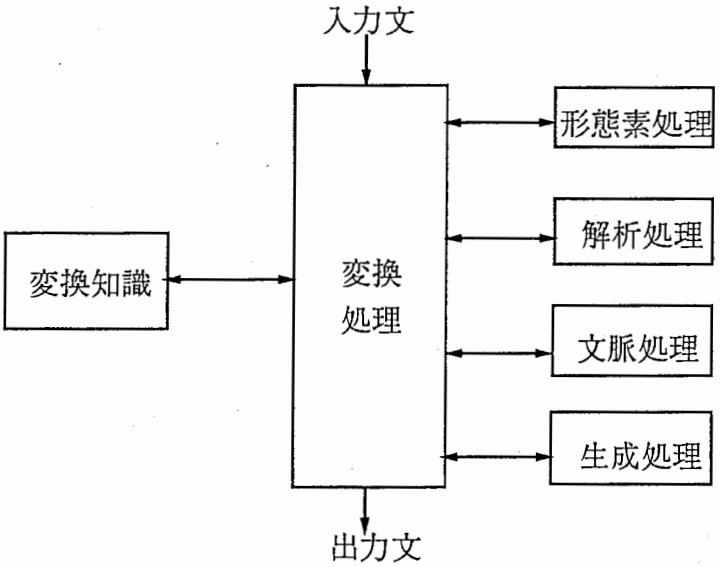

図 2 TDMT の基本構成

3　混合 N-gram・品詞 N-gram を用いることによる単語 open な文の認識実験

前述の音声認識システムにおいては N-gram が単語連鎖として記述されているために N-gram の作成に用いたコーパスに対して単語 open な文の認識は不可能である。しかしながら、N-gram を品詞連鎖として記述することで、単語 open な文の認識が可能となる。

ここでは、N-gram の記述に品詞を用いることが単語 open な文の認識に対してどの程度有効であるかを検討するため、単語 N-gram と混合 N-gram・品詞 N-gram との比較を行なった。また、認識結果を翻訳した場合の翻訳率も確認した。

3.1　混合 N-gram・品詞 N-gram

混合 N-gram とは、N-gram の記述に品詞を用いる際に自立語は品詞で分類してまとめたカテゴリとして扱い付属語は各単語独立したカテゴリとして扱った N-gram のことである。これに対し品詞 N-gram は、すべての単語を品詞で分類したカテゴリで N-gram を記述したものである。本実験で用いた混合 N-gram・品詞 N-gram のカテゴリ数を表 1 に示す。

表 1　N-gram のカテゴリ数

N-gram	カテゴリ数
混合 N-gram (open)	209
混合 N-gram (closed)	211
品詞 N-gram (open)	43
品詞 N-gram (closed)	43

3.2　実験条件

実験は単語 open な文の認識実験と closed な文の認識実験の比較を単語 bigram・単語 trigram・混合 bigram・混合 trigram・品詞 bigram・品詞 trigram を用いてそれぞれ行なった。さらに、各認識結果の文字列 (漢字かなまじり文) を TDMT に与えて日英翻訳を行なった。

N-gram の連鎖確率値の計算には、ＡＴＲの対話データベースの中にある国際会議の予約に関するデータ 1734 文から、単語 open な文の認識実験にはテストデータ 94 文を除いた 1640 文、closed な文の認識実験にはテストデータを含む 1734 文のテキストデータを用いた。評価文はこのデータベースの中の国際会議の問い合わせに関する文 (通称モデル会話)261 文から、モデル会話に含まれない 1473 文中に 1 度も現れない単語を 1 つ以上持つ (単語 open) 文、94 文をピックアップした。これらの重複の関係を図 3 に示す。

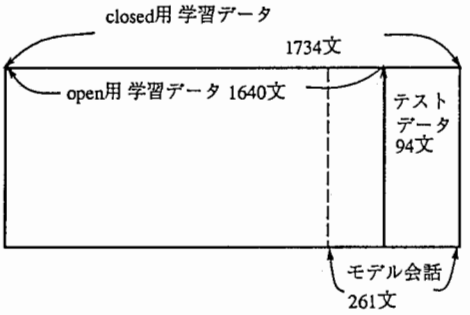

図 3　学習データとテストデータの関係

また、文認識アルゴリズムの単語予測部が単語 N-gram を使う仕様であるため、今回作成した混合 N-gram および品詞 N-gram の出現頻度を元の単語 N-gram の対応する単語連鎖の出現頻度に置き換えることで単語 N-gram として扱える疑似混合 N-gram・疑似品詞 N-gram に変換して使用した。したがって、厳密には混合 N-gram・品詞 N-gram ではないがそれぞれ混合 N-gram・品詞 N-gram として扱って差し支えない。その他の実験条件を表 2 に示す。なお、スムージング処理は行なっていない。

表 2　文音声認識の実験条件

基本アルゴリズム	Continuous mixture HMM + Beam Search + 混合 N-gram または品詞 N-gram
Mixture 数	最大 14 (各音素によって異なる)
1 音素あたりの状態数	3-state 4-loop left-right model
使用パラメータ	LPC ケプストラム 16 次 + パワー + Δ パワー + Δ ケプストラム 16 次
ウインド幅	20ms
フレーム周期	5ms
HMM の学習音声	テストデータと同一話者の 2670 単語発声
音素カテゴリ数	52 音素
認識単語数	1567
ビーム幅	1024
duration control	なし
未知語処理	なし
認識単位	文
発声様式	朗読発声
発声内容	国際会議の申し込み (通称モデル会話)
言語情報	混合 N-gram または品詞 N-gram
N-gram の連鎖確率の推定に使用した	1640 文章　15732 単語　(open data)
テキストデータ (学習データ)	1734 文章　16622 単語　(closed data)
テストデータ	94 文章　890 単語

3.3　実験結果・考察

実験結果を表 3 および図 4 に示す。比較のために同じ学習データ (closed) を用いて作成した単語 N-gram による認識結果も併せて示す。

表 3　認識実験の結果 認識率 [%]

	単語 bigram	混合 bigram		品詞 bigram		単語 trigram	混合 trigram		品詞 trigram	
N-best	closed	open	closed	open	closed	closed	open	closed	open	closed
1	73.4	24.5	42.6	20.2	31.9	90.4	21.3	87.2	37.2	85.1
2	85.1	27.7	50.0	24.5	37.2	95.7	21.3	93.6	42.6	93.6
3	86.2	27.7	50.0	26.6	38.3	97.9	22.3	95.7	42.6	93.6
4	86.2	27.7	50.0	28.7	41.5	97.9	22.3	95.7	42.6	93.6
5	87.2	28.7	51.1	28.7	41.5	97.9	22.3	96.8	42.6	93.6
6	87.2	28.7	51.1	28.7	41.5	97.9	22.3	96.8	43.6	94.7
7	87.2	28.7	51.1	28.7	41.5	97.9	22.3	96.8	43.6	94.7
8	87.2	28.7	51.1	28.7	41.5	97.9	22.3	96.8	43.6	94.7
認識時間 (sec/ 文)	82.0	70.4	70.5	70.3	69.7	90.8	88.1	86.6	104.8	93.4

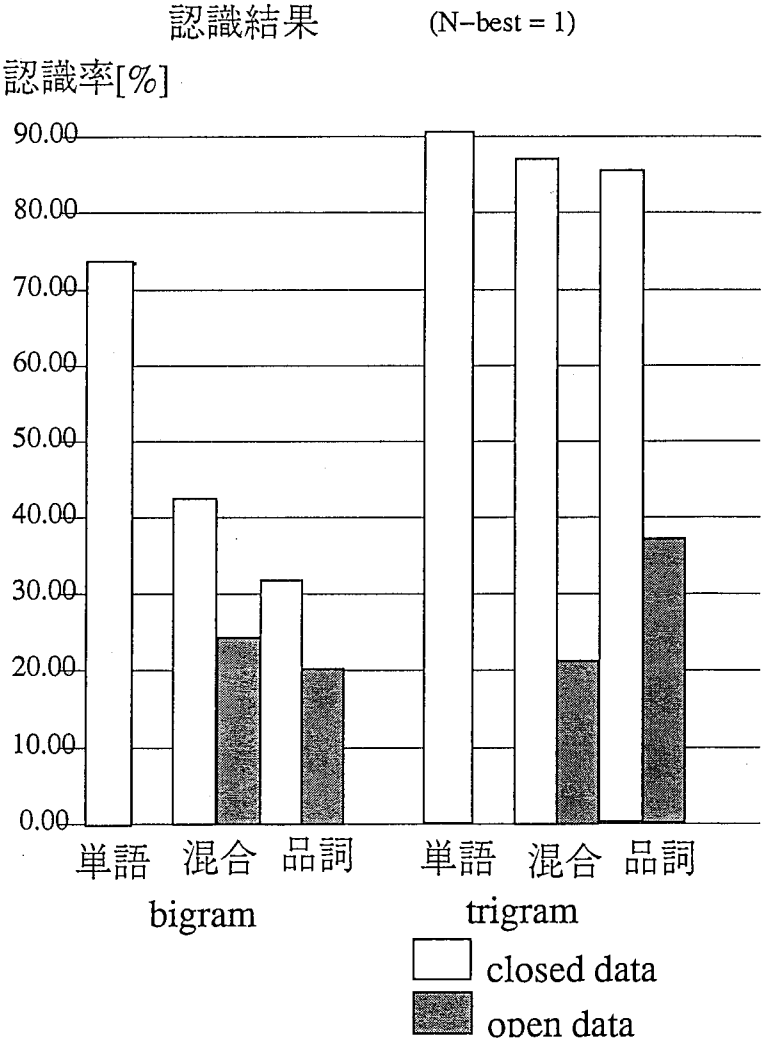

図4 認識率

認識結果より、

- closed data に対する認識では、単語 N-gram を用いた場合と比べて、N-best が 1 の場合、混合 bigram で 20.8 % 品詞 bigram で 41.5 % の認識率の低下が見られたのに対し、混合 trigram・品詞 trigram ではそれぞれ 3.2 % ・ 5.3 % しか低下しなかった。
- 混合 N-gram を用いた場合、品詞 N-gram に比べて bigram の時は open data ・ closed data ともに認識精度が向上しているのに対し、trigram の時は closed data に対しては向上しているものの open data に対しては認識率が低下している。
- closed data に対しては trigram を用いた場合と bigram を用いた場合との認識率の差が大きく現れているが、open data に対しては trigram を用いても認識率はあまり上昇しない。

ということがわかる。この原因を考察するために今回作成した N-gram の規模とテストデータとの関係について調べた。以下の評価は認識第 1 位のみに着目する。表 4 および表 5 に示す。

表4 各 N-gram の規模

	混合 bigram	品詞 bigram	混合 trigram	品詞 trigram
	open → closeed	open → closeed	open → closeed	open → closeed
N-gram 規則数	997 → 1035	388 → 400	2423 → 2583	1259 → 1359

open : 1640 文で生成された N-gram 数
closed : 1734 文で生成された N-gram 数

表5　テストデータのカバーと認識結果の関係

	N-gram(注1)	認識結果	混合 bigram	品詞 bigram	混合 trigram	品詞 trigram
a.	○	○	23文	19文	20文	35文
b.	○	×	30文	46文	1文	8文
c.	×	×	41文	29文	73文	51文
d.	×	○	0文	0文	0文	1文
N-gramの存在する文			53文	65文	21文	43文
N-gram存在時の認識率(注2)			43.4	29.2	95.2	81.0

※注1　文中の全ての単語連鎖においてN-gramが存在する場合○
※注2　N-gramが存在する文に対する認識率 a / (a + b)

これらの表より、

- trigramを用いた場合、テストデータを学習データに加えることによって増加するN-gramの増加傾向 ($\frac{2583-2423}{94}$ = 1.71規則／文) は、学習データを与えた時 ($\frac{2423}{1640}$ = 1.47規則／文) と同程度(表4)であり、今回学習に用いたコーパスではテストデータの単語連鎖をカバーするに足りなかった。このことは、表5においてN-gramが存在しない文 (c+d) が多いことからもわかる。(表5においてN-gramが存在しないにもかかわらず認識できている文は、「国際会議場」を「国際」「会議」「場」の3単語に分割することでN-gramが存在するために認識可能となっていると予測される。)

- bigramによる認識においては、テストデータ94文のうち混合bigramで56.4 %、品詞bigramでは69.1 %の文がそれぞれカバーされており、bigramが存在しないこととこと以外に誤認識の原因があると考えられる。(表5)

- 品詞N-gramを用いた方が混合N-gramを用いた場合より多くの文をカバーできる。しかしながら、N-gramがあるにもかかわらず認識に失敗する例が多くなる。N-gramの存在する文に限定するならば混合N-gramを用いた方が優秀な結果となる。(表5)

- trigramの場合にはN-gramの存在を保証すればopen dataであってもclosed dataと同等の認識率を得ることができる。(表3・表5)

がわかる。以上より考察できることを以下に示す。

- 高精度の認識処理を行なうためにはどの種類のN-gramを用いたとしてもtrigram以上の厳しい言語規則を用いるべきである。

- 今回用いた程度の規模のコーパスでtrigramを作成した場合テストデータ内に未登録の規則が大量に存在する結果となる。認識精度の向上には、テストデータに対して適当な質・量のデータを用いてtrigramを作成することが必要であると考えられる。

認識結果をTDMTによって翻訳処理を行なった結果を表6および図5に示す。翻訳処理によってリカバーできた誤認識はわずかであった。この点について次章で述べる。

表6　翻訳実験の結果 翻訳正解率 [%]

	単語 bigram	混合 bigram		品詞 bigram		単語 trigram	混合 trigram		品詞 trigram	
N-best	closed	open	closed	open	closed	closed	open	closed	open	closed
1	78.7	26.6	43.6	27.7	36.2	93.6	30.9	89.4	46.8	86.2
2	88.3	30.9	52.1	30.9	39.4	97.9	30.9	94.7	51.1	93.6
3	89.4	31.9	52.1	34.0	39.4	100.	31.9	96.8	51.1	93.6
4	90.4	31.9	52.1	36.2	42.6	100.	31.9	96.8	51.1	93.6
5	91.5	34.0	53.2	36.2	42.6	100.	31.9	96.8	51.1	93.6
6	91.5	34.0	53.2	36.2	42.6	100.	31.9	97.9	52.1	94.7
7	91.5	35.1	54.3	36.2	42.6	100.	31.9	97.9	52.1	94.7
8	91.5	35.1	54.3	36.2	42.6	100.	31.9	97.9	52.1	94.7

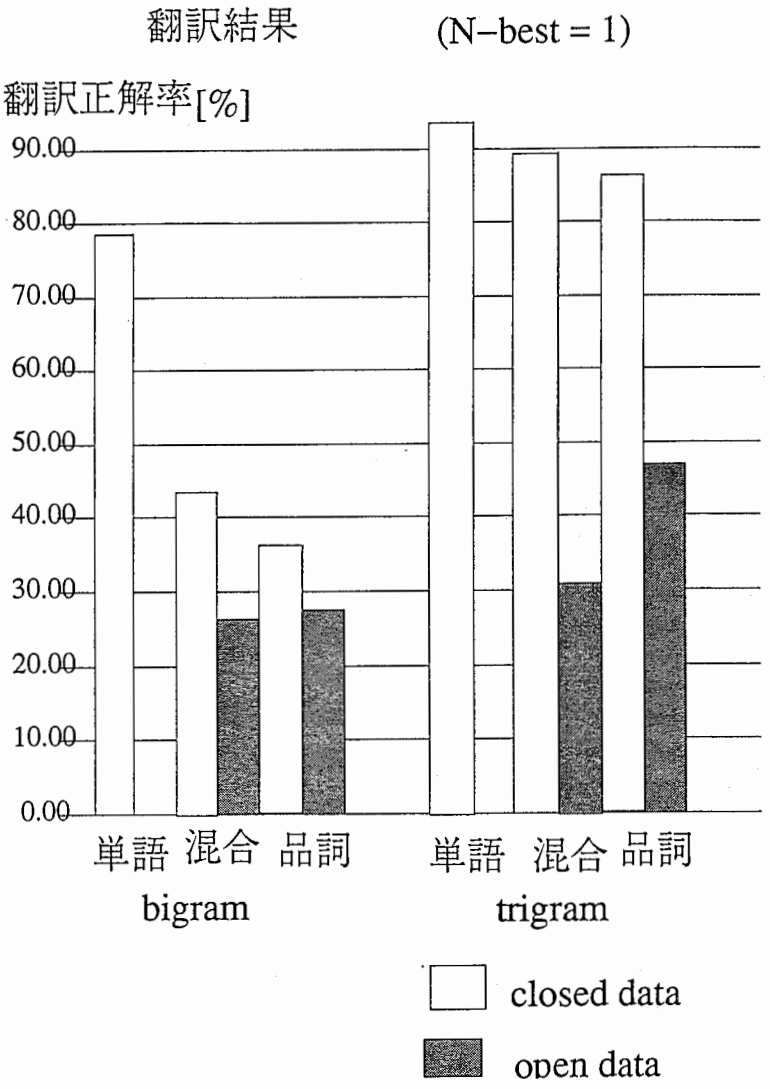
翻訳結果 (N−best = 1)
翻訳正解率[%]
90.00
80.00
70.00
60.00
50.00
40.00
30.00
20.00
10.00
0.00
単語
混合
品詞
単語
混合
品詞
bigram
trigram
closed data
open data

図5 翻訳正解率

4 翻訳処理部による誤認識のリカバー特性に関する実験

音声認識処理部による出力が完全に正解でなかったとしても翻訳処理部でその誤認識をリカバーすることができることもある。ここでは、音声認識処理部による誤認識の傾向と、TDMT でリカバー可能な誤認識について実験・考察する。

前章の実験結果のうち混合 N-gram を用いた closed な文に対する認識・翻訳実験の結果について検討する。認識率および翻訳正解率を表 7 に示す。

表 7 認識・翻訳実験の結果

	bigram		trigram	
N-best	認識率 [%]	翻訳正解率 [%]	認識率 [%]	翻訳正解率 [%]
1	42.6	43.6	87.2	89.4
2	50.0	52.1	93.6	94.7
3	50.0	52.1	95.7	96.8
4	50.0	52.1	95.7	96.8
5	51.1	53.2	96.8	96.8
6	51.1	53.2	96.8	97.9
7	51.1	54.3	96.8	97.9
8	51.1	54.3	96.8	97.9

結果を分析し、誤認識例とともに以下に示す。例は (正解文 → 1位出力) の順で示す。

- 認識誤りの中には単語の長さ (継続時間) が不適当なものが含まれており、音韻レベルまたは単語レベルでの継続時間制御の必要性が考えられる。
  - 例
    - ～一人部屋をお取りしました。→～一人でお取りしました。
- TDMT でリカバーできるものは誤り音節数の少ない場合に限られるが、誤り音節数が少ない場合でもリカバーできない場合もある。
  - リカバー可能な例
    - ～質問したいんですが。 → ～質問したいのですが。
  - リカバー不可能な例
    - ～割引を行なっておりません。 → ～割引を行なっておりました。
- 文の意味・文脈からのリカバーの可能性がある。
  - 例
    - ご住所は～よろしいですね。→ ご住所は～よろしいです。
- 誤認識であると判定できない文章が出力されることもある。
  - 例
    - ～していただかなくてはなりませんが。→～していただかなくてはなりませんか。
    - ～でございますね。→～でございます。

5 まとめ

本実験のまとめとして、以下のことがいえる。

- closed data の文認識においては、 trigram による認識では混合 N-gram・品詞 N-gram ともに音声認識の際の言語情報として有効に機能していることがわかる。
- open data の文認識にこの手法を適用した場合、今回の実験では、コーパスが小さいために認識率は低下した。しかし、 N-gram が存在するならば closed data と同等の認識性能が得られていることから、コーパスが大きくすれば trigram を用いると open data であっても closed data に近い認識性能が期待できる。
- trigram を用いた場合は混合 N-gram・品詞 N-gram を用いた場合においても言語情報による制約が厳しいために認識率の低下は僅かであるが、 bigram を用いた場合には、制約がさらに緩やかになるため認識率の大幅な低下を避けることができず、 bigram 程度の制約では認識処理を行なうのは難しいと思われる。
- 翻訳処理部でのリカバーには限界があるので、誤認識と判定できない文章などに対しては認識性能を向上させることが望まれる。

今後の課題として、

- 音響処理部に継続時間制御を導入することの検討
- タスクと N-gram 作成のためのコーパスの質・量を再検討し、 open data の認識に耐え得る N-gram の作成を行なうこと
- 疑似 N-gram ではなく、純粋な混合 N-gram・品詞 N-gram を用いて単語予測を行なった場合の評価を行なうこと

の三点が挙げられる。

謝辞

本研究の機会を与えて頂いた音声翻訳研究所 山崎泰弘社長、第三研究室 飯田仁室長、古瀬蔵主任研究員に深く感謝致します。

参考文献

[1] 村上仁一，“単語の trigram を利用した文音声認識と自由発話認識への拡張”，ATR Technical Report，TR-IT-0031，(1993)．

[2] 古瀬蔵 隅田英一郎 飯田仁，“経験的知識を活用する変換主導型機械翻訳”，情報処理学会論文誌，Vol.35，No.3，pp.414-425 (1994)．