音声翻訳において音声認識出力の詳細度は最終結果にどう影響するか？

沢井 康孝 †‡ , 菊井 玄一郎 † , 山本 博史 †
†ATR 音声言語コミュニケーション研究所
‡ 長岡技術科学大学
{yasutaka.sawai , genichiro.kikui , hirofumi.yamamoto}@atr.jp,sawai@nlp.nagaokaut.ac.jp

1 はじめに

音声翻訳システムは音声認識、自動翻訳、音声合成の３つのモジュールから構成される。しかし、これらのモジュールの間での最適な機能分担, すなわち各モジュールの出力する情報内容についてはあまり報告されていない。たとえば、音声翻訳では同音異義語,(同音異表記語) の多義解消が必要であるが、この多義性を音声認識側の言語モデルで解消して漢字（意味）表記レベルの情報を翻訳に渡す場合と、認識側では多義解消を行わず「読み」だけ出力して、翻訳側の訳語多義解消のメカニズムに任せる場合のどちらが全体として最適なのか、については明らかでない。この原因の一つは、自動翻訳等のモジュールが特定の入力形式に適合するように人手で調整された規則や辞書を含むため、入力形式を変更してその効果を検証することが事実上不可能だったためと思われる。

近年、活発な研究が行われている統計的翻訳は、対訳関係にある単語（形態素）列の対の集合のみから翻訳知識を学習するため、学習コーパスの入力（原言語）側の各単語の表現形式を変更する（たとえば、漢字表記の代わりに読みを用いる）ことによって、入力の形態素情報に適合した翻訳モジュールのパラメータを自動的に設定することができる(1)。

本研究では、このことを利用して音声認識から自動翻訳に渡される形態素情報の詳細度と音声翻訳性能の関係を調査する。

以降、2 節では本研究の基本的な考え方について述べ、3 節において実験方法、4 節で今回行った実験結果を示し、5,6 節で実験結果の考察を述べる。

2 基本的な考え方

2.1 前提とする音声翻訳システム

今回使用した ATR の音声翻訳システム は、音声認識部、自動翻訳部及び音声合成部の三つのモジュールをこの順に直列に接続したものである。本研究では、日本語から英語への音声翻訳に限定する。また、最終結果は自動翻訳から出力される英語単語列であるとし、音声合成については対象外とする。音声認識部は音響モデルとして HMM と言語モデルとして単語 N グラムを用いた大語彙連続音声認識処理 [1] であり、翻訳部は統計モデルとして IBM モデル４、探索アルゴリズムとしてビームサーチを用いた統計翻訳処理 [2] である。

音声認識部から自動翻訳部には「単語 ID 列」が渡される。ここで単語 ID 列とは、音声認識用の言語モデルにおける単語（日本語においては形態素）の識別 ID の列である。単語 ID は（字面の）標準形、品詞、読みなどの属性の組み合わせが識別できるように付与されているので、たとえば、同じ字面でも読みや品詞が異なれば別の単語 ID になる。

自動翻訳部はこの音声認識結果を直接入力できるように、学習コーパスの原言語側（入力側）テキストを上記の単語 ID の列に置き換えて統計パラメータの学習を行っている。

2.2 形態素表現の詳細度と機能分担

前節で述べたシステムにおいて、自動翻訳の入力側の単語 ID の定義を粗くして、例えば「読み」以外の違いは無視したとしよう。この場合、同一の読みの単語はマージされて同一の ID が付与される（「橋」も「箸」も同じ語とみなされる）ため、音声認識にとっては同音異表記の多義性がなくなり、見かけ上の認識精度が向上する。一方、自動翻訳部にとっては、上記マージによって増加した橋／箸のような多義性を翻訳多義解消（訳語選択など）の機構でうまく解消する必要がある。

このように、音声認識から翻訳に渡される形態素列の表現形式を変えることにより、前者と後者の（言語的多義解消に関する）機能分担を若干変更することができる。この変更が音声翻訳全体の性能に及ぼす影響を実験的に調査することが本研究の目的である。

(1) 本研究では翻訳の際に統計モデルによるデコーディングに先行する「前処理」は考えない。

3　分析の条件

分析対象コーパス

本研究では旅行会話基本表現集コーパス (BTEC) [2] をランダムに訓練セットと評価セットに分けて使用した。表 1 にこれらのサイズを文数で示す。

表 1: コーパスサイズ

	訓練セット	評価セット
文数	152170	1018

なお、音声翻訳の評価のために、評価セットに対する異なり話者 8 人分の朗読音声データを用いた。

3.1　形態素情報の詳細度

先に述べたように、音声認識部の言語モデル学習用の単語 ID は表層形、読み、正規形、品詞、品詞補助情報などの属性情報の組み合わせに対して付与されている。ここで、正規形とは表記の揺れ（送り仮名、字種など）を正規化して終止形にしたもの、表層形とは正規形を適切に活用させたものである。

表 2: 形態素情報の例

表層形	読み	正規形	品詞	品詞補助
東京	トウキョウ	東京	普通名詞	国名
投げろ	ナゲロ	投げる	本動詞	一段, 命令

これらの属性の中から次の 5 種類の属性の組み合わせを選ぶことによって異なった詳細度の単語 ID 体系を作成した (表 3)。なお、元の単語 ID の詳細度 A である

表 3: 詳細度

詳細度	着目する属性
A	表層形, 読み, 正規形, 品詞, 品詞補助情報
B	表層形, 読み, 正規形, 品詞
C	読み、正規
D	読み、品詞
E	読み

各詳細度に従って、すでに構築されている学習用コーパスの日本語側 ID 列に変更を加えることで新しい学習コーパスを作成し、これを用いて、翻訳処理部で使用する統計パラメータを学習させた。

3.2　学習時の辞書のサイズ

図 1 に日本語コーパスにおける単語 ID 数を示す。形態素属性情報が読みだけ（詳細度 E）のコーパスについて、元のコーパスの約 15 パーセントほどの単語が同一 ID として扱われるようになっている。これは属性情報を削ることによって異なった単語 ID が同一の単語 ID にマージされるため、学習コーパスの原言語部分の辞書サイズ（異なり単語数）は詳細度によって異なる。

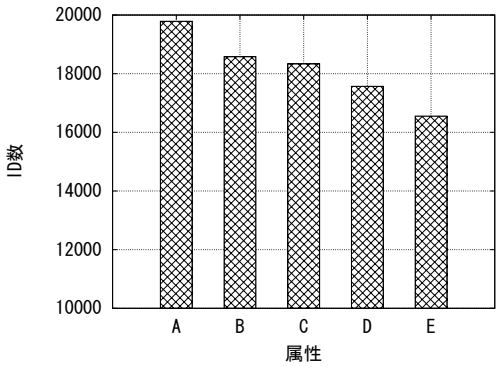

図 1: 日本語コーパス単語 ID 数

単語 ID がマージされたことにより、翻訳側で解消すべき多義語が増加する。このことを平均多義数（文、単語）で評価をしたのが表 4 である。ここで平均多義数（単語) とは変換後の単語が与えられた時の元の単語 ID の確率（相対頻度）を用いて、前者から後者を推定する場合の平均候補数である。なお、平均多義数 (文) とは文全体で考えた平均候補である

表 4: 平均多義数

詳細度	平均多義数 (単語)	平均多義数 (文)
A	1	1
B	1.05	1.56
C	1.18	3.75
D	1.08	1.91
E	1.29	7.74

「読み、品詞」と「読み、正規形」において単語 ID 数は「読み、品詞」の方が圧縮されているにもかかわらず、平均多義数は「読み、品詞」の方が低い。品詞情報が単語を一意に決定するのに大きく作用することがわかる。

3.3　学習された日英辞書

各詳細度において学習された日英翻訳辞書の一部、翻訳先候補の確率値上位 2 つを表 5 に示す。形態素属性情報を削ることで各形態素はあいまいとなるが、翻訳の選択候補については増加、減少の両方が見られた。

3.4　被覆率

二種類のテストセットに含まれる 8 つの認識ファイル及びその正解ファイルの文書が翻訳学習コーパスにどれだけ存在するか、各テストセットのコーパス内被覆率の平均値を表 6 に示す。

[2] 旅行会話で頻繁に使用される表現を書き出して、対訳を付加したコーパス [3]

表 5: 日英候補

詳細度	B		C	
	候補	確率値	候補	確率値
箸	chopstick	0.745	chopstick	0.917
	me	0.126	pair	0.083
	総候補数	4	総候補数	2
橋	bridge	0.826	bridge	0.764
	one	0.039	there	0.091
	総候補数	5	総候補数	5
端	end	0.657	end	0.555
	ball	0.183	dowstairs	0.210
	総候補数	7	総候補数	9
詳細度	D		E	
	候補	確率値	候補	確率値
ハシ	chopstick	0.3715	bridge	0.4808
	bridge	0.3165	chopstick	0.2585
	総候補数	9	総候補数	13

表 6: 文被覆率

詳細度	正解ファイル	認識ファイル
A	0.462	0.451
B	0.467	0.451
C	0.468	0.451
D	0.468	0.453
E	0.473	0.453

4 実験

4.1 評価尺度

各詳細度において翻訳結果の精度を示す指標として、BLUE[6] , WER(Word Error Rate) を使用した。これらは翻訳結果と模範解答を比較することで数値化している。

4.2 実験方法

評価セットの音声データを音声認識部に入力し、元の単語 ID（詳細度 A) による音声認識結果を得た。これを 3.1 節で示した詳細度別の単語 ID 列に変換し、それぞれを（対応する学習コーパスで学習させた）自動翻訳部に与えて翻訳結果を得た。また、音声認識誤りの影響を調べるために、音声を正しく認識した場合の単語 ID 列を各詳細度の単語 ID に変換して同様に自動翻訳結果を得た。なお、処理対象は評価セット全体であるが、パラメータの最適化のためにこれを２つに分け、一方を最適化に用い残りで評価することをそれぞれに対して行い平均した。

4.3 翻訳評価

音声認識精度

図 2 に各詳細度で評価した場合の音声認識精度を示す。詳細度を変化させて認識正解との一致率を調べた値である。形態素属性の詳細情報を減少させることで、属性情報に誤りのある単語が正解に近づくことになる。値は 8 種類ある結果の平均値である。

平均値において見かけ上の認識精度は最大 1 パーセントの上昇が見られる。

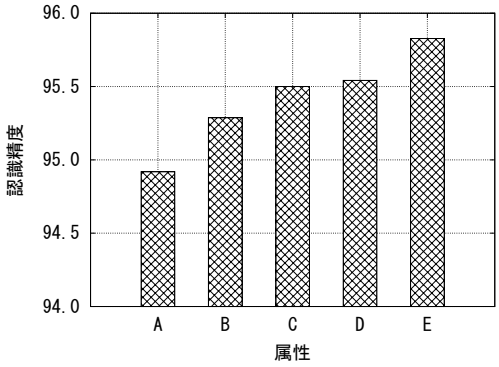

図 2: 音声認識精度

翻訳精度

翻訳単体における評価値及び音声認識、翻訳を合わせた評価値を図 3、4 に示す。

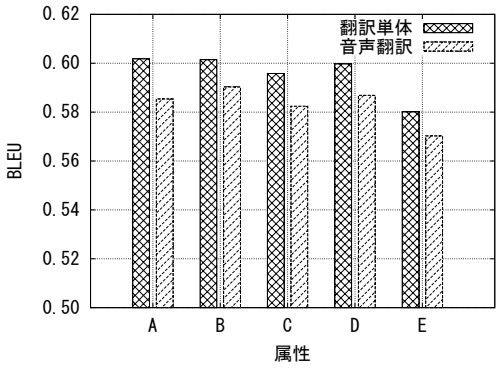

図 3: 翻訳精度:BLEU

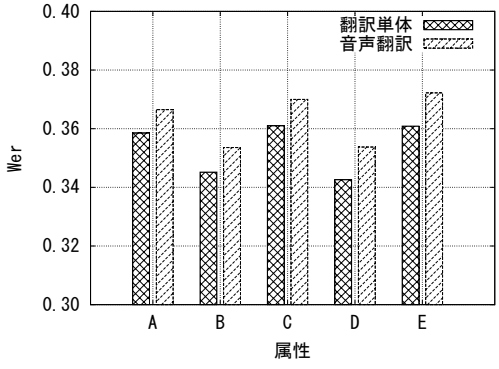

図 4: 翻訳精度:WER

翻訳単体の精度は各詳細度の間で BLEU の値で 0.02、Wer の値で 0.02 の変動幅が存在する。これは BLEU において平均値の 3%の割合である。平均値においてこの結果より、形態素属性の詳細度を粗くしてもあまり性能は変わらないと言える。音声認識と翻訳を合わせた精度は単体の物と比べて、変動幅は同様であり全体的に、低下している。翻訳単体の精度と、音声認識及び翻訳を合わせた精度の差を拡大して図 5 に示す。

音声翻訳精度は翻訳精度と比較して、単語 ID に誤りがあるため、精度は低下している結果となっている。

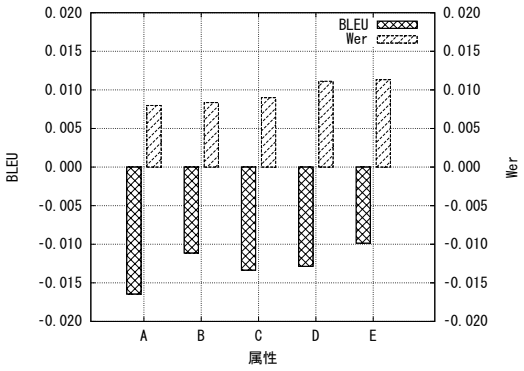

図 5: 翻訳と音声翻訳の差

5 考察

5.1 詳細度と翻訳精度

日本語コーパスの情報を変化させたときの日英翻訳結果は、単語 ID が読みの情報だけの場合でも、全ての情報を付加した場合と同等の翻訳精度が得られた。日本語を英語に翻訳するモデルが、英語の言語モデルと英単語が日本単語に変換される確率モデルによって、翻訳側での多義解消が行えたと推測する。

情報を減らすことにより、翻訳モデルの日本語から英語への対訳辞書については、統合された単語 ID の翻訳先が増える傾向にあった。しかし、増え方については単に統合される前にそれぞれが持っていた翻訳先が単純に足し合わされるだけではなく統合されることで、統合される前には無かった翻訳先が現れていた。また統合された単語以外の単語についても影響が現れていた。

以上より、入力の曖昧性が上昇したことで、通常正解が得られなかった文書が正解を出力できるようになるケースが見られた。しかし曖昧性が上昇したことによって正解であった翻訳が誤りとなってしまうケースも存在する。

僅かではあるが読みだけを利用したモデルが、全ての情報を利用したモデルの精度を上回っている結果もあり、総合的に見て、日本語の情報を減らすことは音声認識の精度を上昇させ、翻訳の精度に付いては多少の変化が見られるが、翻訳精度を大きく低下させる要因になるということは見られなかった。なお、統計的信頼性の議論は今後の課題である。

5.2 各翻訳結果を統合的な利用

上述の通り、形態素属性の詳細度を変えた場合の音声翻訳品質の変化は評価セット全体の平均としてはあまり大きくない。しかし、それぞれの出力を比較すると詳細度ごとに最終的な翻訳誤り箇所が異なることがわかった。このことから、各詳細度の出力から最適なものを選ぶことができれば出力品質が向上する可能性がある。これを検証するために、各詳細度による音声翻訳結果の中から正解に最も近いものを文ごとに選び、翻訳精度を評価した。8 人分の話者に対する値の平均をとると BLEU:0.630、WER:0.283 という結果が得られ、単独の場合よりも特に WER で大幅な改善となった。このことから、5 つの詳細度を適切に使い分けることによって最終性能が改善できる可能性があることがわかった。

6 まとめ

音声認識側で行ってた、多義解消、品詞の付加などの処理を省いて、省いた処理を翻訳側に任せることは、翻訳結果の精度の低下には繋がらないと考える。今後の課題として、複数の詳細度を適切に使い分けることによって最終性能を改善することがあげられる。また、今回は文を単位として選択を行ったが、認識の誤りやすい単語、翻訳の誤りやすい単語など単語単位で属性情報を変化させることも今後の課題である。

謝辞

本研究は情報通信機構の委託研究「大規模コーパスベース音声対話翻訳技術の研究開発」により実施した物である。

参考文献

[1] 伊藤 玄 他 ,“音声認識統合環境 ATRASR の概要と評価報告,” 音響講論,1-P-30, pp.221-222,(2004)

[2] Watanabe,T.et al. , “Exsample-based Decording for Statistical Machine Translation,”MT Summit IX,pp.410-417(2004)

[3] 菊井 玄一郎 他, “対話翻訳のための音声言語コーパスの現状,” 日本音響学会講演, 1-8-25 pp.55-56 (2004)

[4] 安田 圭志 他,“対訳コーパスを用いた翻訳品質自動評価法,” 情報処理学会 47(3):2108-2117

[5] 今村 賢治 他, “機械翻訳自動評価指標の比較,” 言語処理学会 第十回年次大会 pp.452-455(2004)

[6] Kishore Papineni, “BLEU: a Method for Automatic Evaluation of Machine Translation,” ACL pp.311-318 (2002)