

帰納的学習と静的な翻訳規則を融合した音声翻訳手法の提案

笹岡久行*

荒木健治†

桃内佳雄‡

柄内香次†

旭川工業高等専門学校*

北海道大学大学院工学研究科†

北海学園大学工学部‡

1 はじめに

大規模なコーパスの整備、音声認識技術の進歩並びに計算機の処理能力向上などの要因により、パーソナルコンピュータの上でも実時間の範囲で動作可能な連続音声認識手法の提案がなされ、高い認識精度を実現している [1, 2, 3]. しかし、実際の問題として、音声認識システムが 100% の認識精度を達成するのは大変困難なことである。多くの機械翻訳手法の研究では、原言語の入力文は正しいものとして翻訳処理を進めるため、入力文に音声認識誤りが含まれることは音声翻訳システムには非常に大きな問題となる。

一方、機械翻訳手法に関する研究においても従来から行われている書き言葉を翻訳対象としたものだけではなく、話し言葉を対象とした翻訳手法の提案がなされている [4]. 古瀬らの研究 [4] では、「旅行会話」での使用を想定し、模擬会話を書き下した実験データにおいて、実時間の範囲で高い翻訳成功率を実現したことを報告している。しかし、この実験データは話し言葉を正確に書き下したものであり、上述した音声認識結果に誤りが含まれる問題を解決してはいない。

音声翻訳システムは入力音が音声であり、その入力の容易さから高い需要はあるが、実用的なシステムは多くはない。音声翻訳における主な問題点としては、(1) 音声認識処理の認識誤りの問題、(2) 話し言葉の処理の問題、(3) 処理速度の問題などがあげられる [5]. 本研究では、音声翻訳システムの実現を目指し、主に (1) および (2) の問題について考える。本稿では、本研究における音声翻訳における基本的な考え方を述べ、提案手法を基に作成した実験システムの概要について説明する。次に、上述した問題 (1) の解決を目指した音声認識誤りの訂正手法について述べ、この訂正手法に対しての評価実験について報告する。今回の実験は、「買い物」の場面での会話文を実験データとし、音声認識システムの 82.5% という高い音声認識精度から、提案した訂正手法により 0.7 ポイントの精度向上があり、音声認識精度を向上できる可能性があることを示した。最後に、帰納的学習により用例から自動的に獲得された翻訳規則と人手により作成された静的な翻訳規則を融合して用いる翻訳手法について述べる。

2 音声翻訳に対する基本的な考え方

近年の機械翻訳処理手法に関する研究では大規模なコーパスを利用する処理手法に関する研究が盛んに行われて

いる [6, 7]. それ以前には人手により作成された規則だけを用いた手法 (規則に基づく手法) の研究が盛んであった。しかし、これらの翻訳手法を用いる場合、人手により作成される翻訳規則の網羅性や規則作成のためのコストの大きさなどが問題となる。これに対して、コーパスに基づく手法では学習データを用意することで、人手により用意された規則よりも網羅性の高い規則を自動的に獲得することができる。しかし、これらの手法ではコーパスにそれほど多く出現しない言語現象への対処が問題となる [8]. さらに、処理する対象分野毎に大量の学習データが必要となる。音声翻訳は対象が話し言葉であり、低頻度の言語表現が出現する。また、翻訳対象分野毎に数万文もの学習データを用意するのも大変なコストがかかる。そのために、コーパスに基づく翻訳手法のみで精度の高い音声翻訳を行うのは困難である。

一方、我々も遺伝的アルゴリズムを適用した帰納的学習による機械翻訳手法 (GA-ILMT) を提案し、旅行者用の会話文においてその有効性が確認している [9, 10]. この手法では、実例から翻訳規則を学習することにより、比較的少量の学習データでも翻訳対象分野に適応でき、文脈に合った高品質な翻訳が可能であることを報告している。GA-ILMT では、入力文とその翻訳文に遺伝的アルゴリズムの交叉や突然変異を施し、多様な翻訳規則を獲得することに成功している。しかし、それらの多様な翻訳規則を獲得するための学習処理およびそれらの規則を組み合わせる翻訳処理のために、GA-ILMT は遺伝的アルゴリズムの各処理を行わない翻訳手法と比較して処理時間は増大する。さらに、GA-ILMT において獲得された翻訳規則の中には、翻訳処理にとって有効な翻訳規則に加えて、有効に働かないものも含まれており、それらも処理時間の増大および翻訳率の低下を招く一因となっている。この有効ではない翻訳規則の増加および処理時間の増大は、実時間の範囲での翻訳処理が必要となる音声翻訳において問題となる。

そこで、提案手法では「静的な翻訳規則」と「用例から獲得した規則」とを融合して処理を進める。「静的な翻訳規則」とは、実際の言語現象に対する内省などに基づいて人手により作成される規則のことである。また、「用例から獲得した規則」とは実際の翻訳対象とその訳文の間から帰納的学習を用いてシステムが自動的に獲得する翻訳規則のことである。「静的な翻訳規則」は、人手により作成されるために大量の学習データを必要とせず、ま

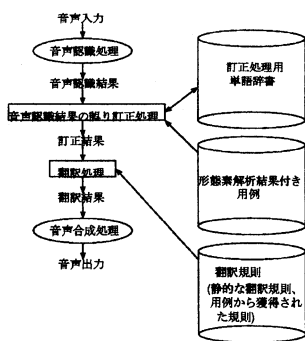


Fig. 1 実験システムにおける翻訳処理の流れ

た、比較的低頻度の言語現象へも対処しうる適切な規則となりうる。一方、「用例から獲得した規則」により翻訳規則の網羅性を高め、翻訳対象の文脈に適合した翻訳が可能になる。

本研究では2種類の翻訳規則（「用例から獲得した規則」と「静的な規則」）を融合して利用するが、このようにした理由は、実際にその翻訳する対象分野において利用され、正誤判定結果が出されるまで、「静的な規則」と「用例から獲得した規則」のどちらを優先的に適用するのかを決定できないと考えたからである。翻訳では、1つの単語あるいは単語列に複数の訳語が存在する場合には、訳語選択の問題があるので、利用する規則の優先順位を決定するのは困難である。さらに、提案手法では新たな翻訳規則を獲得する際にも「静的な規則」を利用する。これにより、用例だけを用いた規則獲得では獲得が困難であった翻訳規則も獲得可能になると考えられる。

3 システムの概要

我々が実現を目指す音声翻訳システムでは、翻訳対象をユーザが音声で入力し、その入力を音声認識して文字列に変換する。この文字列には、誤りが含まれている可能性があるために、これを確信度および用例からの辞書を用いてシステムが自動的に誤りを訂正する。この処理については第4章において説明する。そして、その結果を翻訳し、その翻訳結果を音声合成し、ユーザに音声として出力する。なお、本研究における実験システムでは音声認識および音声合成については既存のシステム[1, 2, 3]を利用した。この理由は、上述したような近年の音声認識技術の向上、および、音声合成については人間らしさを除くと実用上は問題無いからである。Fig.1に実験システムにおける翻訳処理の流れを示す。

4 確信度と用例を用いた自動訂正手法

4.1 処理概要

我々は、単音節の認識結果から誤りを復元する手法を提案し、その有効性を確認している[11]。しかし、連続音声の認識結果が比較的高い精度で得られている現状を踏まえると、連続音声認識結果の誤りを訂正する手法が望まれる。また、統計的な情報を利用した、日本語文の誤りの検出・訂正手法の提案と漢字仮名混じり文における有効性の確認がなされている[12]。しかし、この手法では統計的な情報の学習に非常に大量な学習データを必要とすること、低い頻度でしか出現しない文字の並びを処理し難いこと、さらには文字数が多い場合の訂正処理において訂正候補を得るためには膨大な計算が必要となるなどの問題がある。

提案する訂正手法は、以下の通りである。

1. 音声認識における確信度を認識結果の各単語に付与
2. 予め決定されている確信度に対する閾値を基に誤り位置を推定
3. 推定された誤り位置における他の認識候補を取得
4. 訂正処理用単語辞書を参照し、認識候補の中から訂正単語を決定
5. 誤り位置にある単語と決定された単語と置換し、訂正結果として出力

手順1における確信度は、音声認識システム[1, 2, 3]により得ることができる数値である。文献[1, 13]において述べられているように、この数値は、人間が行う単語単位分割を調査した上で決定された統計モデルによって計算されており、音声認識において、他の候補となる単語に比べてどれくらい確らしいかを示す相対的な値である。この値がとりうる範囲は、-100~100となっている。手順2における確信度に対する閾値は、予備実験において、音声認識結果の正認識と誤認識の割合を確信度に応じて調査し、決定している。この閾値以下の確信度を付与された単語を誤りと推定している。もし、この段階で閾値を下回る確信度を持つ単語がなければ、そこで処理を終了する。手順3における誤り単語における他の認識候補とは、上述したシステム[1, 2, 3]から得られる単語の集合であり、音声認識結果となりえた単語の集合である。もし、誤り位置における他の認識候補が存在しない誤り位置については、そこで処理を終了する。手順4における訂正処理用単語辞書には、処理開始前に人手により、対象分野において必要と思われる単語を予め登録しておく。さらに、処理対象に新たに出現した単語は、順次、システムがこの辞書に追加する。これにより、辞書の網羅性を高めている。また、辞書には単語の字面に加

え、適切な訂正処理に使用された回数、訂正処理に使用された回数、不適切な訂正処理に使用された回数も併せて登録されている。それらの数値は、処理毎に更新され、候補の単語の集合において確らしさを判断する際に利用する。具体的には、システムが各候補の各数値を順次比較し、処理に用いる単語を決定している。

4.2 処理例

以下に処理例を示す。図中における () の中の数字は各単語に付与された確信度である。

ユーザからの入力文
MD プレーヤーはありますか
音声認識結果
NT(-17) プレーヤー (5) は (4) あります (10) か (14)
「NT」以外の候補となる単語
MD, 求人, MT, NGO, 友人, 他
訂正結果
MD プレーヤーはありますか

この例では、「MD プレーヤーはありますか」という文をユーザが音声入力したところ、「NT プレーヤーはありますか」という文が認識結果として出力された。しかし、音声認識処理では、単語「NT」に対して、非常に低い確信度を与えている。そこで、この単語をこの確信度から誤りが含まれている可能性がある単語と推定する。この例では、訂正処理用単語辞書において、「MD」の方がより確らしい単語として登録されており、これにより正しい音声認識結果へと訂正される。

4.3 品詞情報の利用

字面情報のみに基づいて音声認識誤り訂正処理を行うと、訂正規則適用の際に問題が生じることが明らかになった [14]。本訂正手法では、規則が適切な訂正処理に使用された度数、処理に使用された度数および不適切な訂正処理に使用された度数を順次参照し、適応する規則の優先順位を決定している。出現頻度が高い助詞などの付属語は、出現度数が低い地名などの固有名詞と比較して、訂正処理に利用される回数が多くなる。そのために、出現度数が高い単語は訂正処理において利用されやすくなり、不適切な訂正処理においても利用されていた。これを防ぐために、形態素解析の結果から得られる品詞情報を参照し、不適切な訂正処理を判断するようにした。具体的には、訂正処理を行った結果の品詞並びと用例中の文の品詞の並びを比較する。そして、訂正処理後の文の品詞並びが用例中に存在しない場合は、不適切な訂正処理を行ったとシステムが判断し、その訂正処理を却下する。

4.4 音声認識誤り訂正処理に関する実験

音声認識結果に対する音声認識誤り訂正処理に関する実験を行った。本実験システムでは、音声認識には日本アイ・ビー・エム(株)の「ViaVoice」[1, 2, 3]を、形態素

解析には「茶筌」[15]を利用した。実験データは文献[16]の会話文の中から103文(474語)を利用した。また、予備実験の結果から、訂正処理を行う単語の確信度の閾値は0とした。以下の表に、実験結果を示す。

Table 1 音声認識の精度

	正認識	誤認識	合計
認識結果	391 (82.5%)	83 (17.5%)	474

Table 2 誤り位置特定の精度

閾値以下で正認識であった単語	121 (61.7%)
閾値以下で誤認識であった単語	75 (38.3%)
閾値以下であった単語	196 (100.0%)

Table 3 訂正結果の精度

	正認識	誤認識	合計
認識結果	394 (83.2%)	80 (16.8%)	474

Table 1 は、音声認識システムの出力結果の精度を表し、今回の音声認識システムの精度が良いことが確認できる。Table 2 は、予め定めた閾値以下の単語、つまり、訂正処理を施す対象となる単語についての認識精度を表している。この表から閾値以下にも数多くの正翻訳が存在していることが確認でき、この閾値についても今後研究を進める予定である。そして、Table 3 は訂正処理を施した後の全体の認識精度を示しており、正認識率が0.7ポイント向上している。

5 翻訳処理

上述したように本研究における翻訳手法では、「静的な規則」と「用例から獲得した規則」の双方を融合して利用する。

「静的な規則」は、実際の会話例の言語現象に対する内省等により人手により作成し、実験システムが有する翻訳規則辞書に登録した。また、「用例から獲得した規則」は既に提案している「帰納的学習を用いた翻訳手法」[17]と基本的には同様の手法で規則を獲得する。翻訳対象と形態素解析結果が付与された訳文の間から、学習データにおける共通部分と差異部分を抽出し、それらの原言語と目的言語の組みを翻訳規則として、実験システムが有する翻訳規則辞書にシステムが自動的に登録する。ただし、本研究では翻訳文には形態素解析結果が付与されており、さらに、人手で作成された「静的な翻訳規則」との間からも規則が獲得可能な場合は獲得を行う。翻訳処理の際には、その蓄えられた翻訳規則を利用して処理を進める。

上述したように、我々が既に提案している GA-ILMT[9, 10]では、入力文と翻訳文の対に交叉処理や突然変異処理を施し、多様な翻訳規則を獲得する。しかし、有効な翻

訳規則に加え、有効には働かない訳規則をも獲得してしまい、処理時間の増大および正訳率の低下を招いている。単純に静的な訳規則を GA-ILMT に獲得された訳規則と融合しただけでは、交叉や突然変異による誤った原言語と目的言語の文の対の生成を防ぐことはできず、その結果、有効ではない訳規則の生成を抑制することはできない。さらに、処理時間増大の問題も残る。そのために、GA-ILMT において獲得される訳規則と静的な訳規則を融合することで、音声翻訳が抱える問題を解決することはできないと考えられる。一方、我々は GA-ILMT[9, 10] における訳例生成の際に解析的知識を導入し、正訳例の生成において有効性上がることを確認している [18]。しかし、文献 [18] において考察されているように、正訳率を向上させるには、解析的知識を訳例生成処理にだけ導入するのではなく、訳処理やフィードバック処理の各処理にも導入しなくてはならず、さらなる処理時間の増加が予想される。このような理由から、本研究では帰納的学習を用いた機械翻訳手法 [17] に対して静的な訳規則を融合している。

6 おわりに

本稿では、音声翻訳システムが抱える問題点をあげ、その問題の一つである音声認識結果の精度向上のために、音声認識誤りの自動訂正手法を提案した。音声認識精度が高い音声認識結果に対しても本手法を用いることにより精度が向上することの可能性が示された。しかし、提案した音声認識誤り訂正手法では、音声認識誤りのなかで認識の選択候補中に正解となる単語が存在する場合にしか対応できない。つまり、脱落誤りや付加誤りへの対応を検討する必要がある。今後、実験システムにおける訳処理に関して評価実験を行い、その結果を報告する予定である。

謝辞

本研究の一部は、北海学園大学ハイテク・リサーチ・センターの研究費による補助のもとに行われた。

参考文献

- [1] 西村雅史, 伊藤伸泰, “単語を認識単位とした日本語ディクテーションシステム”, 信学論, Vol.81-D-II, No.1, pp. 10 - 17, 1998.
- [2] 西村雅史, 伊藤伸泰, 山崎一考, “単語を認識単位とした日本語の大語彙連続音声認識”, 情処学論, vol.40, No.4, pp.1395 - 1403, 1999.
- [3] 西村雅史, “日本語ディクテーションシステムの現状と今後の課題”, 信学技術報告 NLC99-26, Vol.99, No.523, pp.7 - 12, 1999.
- [4] 古瀬蔵, 山本和英, 山田節夫, “構成素境界解析を用いた多言語話し言葉翻訳”, 言語処理学会論文誌, vol. 6, No. 5, pp.63-91, 1999.
- [5] 長尾真 (編), “自然言語処理”, 岩波書店, 1996.
- [6] 佐藤理史, “アナロジーによる機械翻訳”, 認知科学モノグラフ 4, 共立出版, 1997.
- [7] 北研二, “確率的言語モデル”, 言語と計算 4, 東京大学出版会, 1999.
- [8] 松本裕治, 徳永健伸, “コーパスに基づく言語処理の限界と展望”, 情処学会誌, Vol.41, No.7, pp.793 - 796, 2000.
- [9] Kenji Araki and Koji Tochinal, “Performance Evaluation for Adaptable Machine Translation Method”, Proceedings of the Conference of PA-CLING '97, pp. 1 - 6, 1997.
- [10] 越前谷博, 荒木健治, 桃内佳雄, 枅内香次, “旅行者用英会話文における GA-ILMT の有効性について”, 情処技術報告 NL-128, Vol.98, No.99, pp. 119 - 126, 1998.
- [11] 荒木健治, 宮永喜一, 枅内香次, “多段階分割復元法による誤りの多い文字列からの原文の復元”, 情処学論, vol.30, No.2, pp.169-178, 1989.
- [12] 荒木哲郎, 池原悟, 塚原伸幸, 小松康則, 田川崇史, 橋本憲久, “m 重マルコフ連鎖モデルを用いた日本語文の誤字・脱落・誤挿入誤り文字列の検出と訂正法”, 信学論, Vol.J83-D-II, No.6, pp.1516-1528, 2000.
- [13] 日本アイ・ビー・エム株式会社 ナショナル・ランゲージ・サポート, “SMAPI 解説書”, 日本アイ・ビー・エム株式会社, 1999.
- [14] 笹岡久行, 荒木健治, 桃内佳雄, 枅内香次, “音声翻訳における音声認識誤りの確信度と用例を用いた自動訂正手法の提案”, 平成 12 年度電気関係学会北海道支部連合大会, pp.250, 2000.
- [15] 松本裕治, 北村啓, 山下達雄, 平野善隆, 松田寛, 高岡一馬, 浅原正幸, “形態素解析システム『茶釜』 version 2.2.1 使用説明書”, 奈良先端科学技術大学院大学 松本研究室, 2000.
- [16] 原島一男, “店員さんの英会話”, 文化堂印刷, 1999.
- [17] 荒木健治, 枅内香次, “多段階共通パターン抽出法を用いた訳例からの帰納的学習による翻訳”, 情報処理北海道シンポジウム'91, pp.47 - 49, 1991.
- [18] 工藤晃一, 荒木健治, 桃内佳雄, 枅内香次, “学習型機械翻訳手法に適用された遺伝的アルゴリズムにおける知識による制約の有効性について”, 信学論, Vol.82-D-II, No.11, pp. 2035 - 2047, 1999.