

多言語音声翻訳のためのマルチ・エンジン翻訳技術

Multi-engine Translation Technology for Multi-lingual Speech-to-Speech Translation

隅田 英一郎、パウル ミシャエル、今村 賢治、大熊 英男

ATR 音声言語コミュニケーション研究所

{ Eiichiro.sumita, Michael.paul, Kenji.imamura, Hideo.okuma } @ atr.jp

1. はじめに

ATR は、多言語音声翻訳の基盤技術を研究している [1]。そのため、話し言葉を高精度に翻訳する技術の実現を目指して、対訳コーパスを用いた機械学習に基づく翻訳手法を研究・開発した。

本稿では、提案手法とその翻訳性能について報告する。提案手法では、単一の入力文を異なる特性を持つ複数の翻訳エンジンで処理し、その出力集合から最良の訳文を統計的なモデルに基づいて選択する。ATR が構築した大規模な多言語対訳コーパスを使って評価し、提案手法は対象言語によらず高品質の翻訳を実現できることを確認した。

2. 提案技術の位置づけ

多言語音声翻訳技術はグローバリゼーションの深刻な課題である言語バリアを克服するために不可欠である。

ところが、翻訳システムの開発は膨大な時間と費用を要する。特に、多言語を処理する場合のコストは膨大である。単一方向のシステムを構築するコストに比べて、N 言語なら $N*(N-1)$ 倍のコストがかかる。そのため、大幅なコストの削減が必須となる。

また、よく知られているように、話し言葉と書き言葉は様々な面で異なっており、既存の書き言葉用の翻訳システムを音声翻訳に移植するコストも小さくない。

これらの背景のもとに、ATR では話し言葉の多言語対訳コーパスを構築し、機械学習に基づく翻訳手法を開発してきた。

3. 話し言葉の多言語対訳コーパス

ATR は旅行対話のドメインで 2 種類の対訳コーパスを開発し[2]、これらに基づいて、翻訳システムの構築と評価を行った。

(1) BTEC : 幅広い話題の基本文を多言語¹で集めた大規模な²コーパス

(2) MAD : 翻訳システムを介した 2 者の対話を集めた、話し言葉の特徴を反映する小規模な³対訳コーパス

4. 翻訳アルゴリズム

我々は、翻訳の研究開発プロジェクトを Corpus-Centered Computation (C-cube)と呼んでいる[3]。C-cube ではコーパスをその中心にしている。すなわち、コーパスから翻訳知識を抽出し、翻訳品質はコーパスを参照して測定し、自動測定した品質によってシステムを最適化し、また、一方で、コーパスの精練・正規化・換言を行う。本稿では、翻訳アルゴリズム概要について説明する⁴。

対訳コーパスを使った翻訳技術には、Example-Based Machine Translation (EBMT) [4]と Statistical Machine Translation (SMT) [5]の 2 つの主要な枠組がある。ここで、翻訳技術の現状⁵を見てみると、次々に新しいアルゴリズムが提案され、翻訳品質も改善している。しかし、まだ、翻訳誤りを起こさないアルゴリズムはないし、今後も生まれまいだろう。

¹ 日本語、英語、中国語、韓国語を対象としている。

² 規模は、約 70 万文である。

³ 規模は、約 1 万文である。

⁴ 換言については[6] [7]、文分割については[8]を参照。

⁵ 現在の世界の翻訳技術の到達点については[9] [10] [11]を参照。

そこで、単一のアルゴリズムに頼るのではなく、複数の翻訳エンジンを用い、その結果を利用するという発想が生まれる。そのため、C-cube では、以下で説明するように、EBMT と SMT を平行して研究し、さらに両者を融合する新たな手法を研究した。

表 1 に C-cube の各翻訳エンジンの特徴をまとめてある。順に簡単に説明する。

表 1 翻訳エンジンの比較

	EM	D-cube	HPAT	HPATR	SAT	HPATR2
枠組	EBMT	EBMT	EBMT	(EB+S)MT	(EB+S)MT	SMT
翻訳単位	文	文	句	句	文	句
カバー率	(特に)狭い	狭い	広い	広い	広い	広い
翻訳品質	(特に)優	優	良	良	優	優
処理速度	高速	高速	高速	中速	中速	中速
コーパス以外のリソース		シソーラス 対訳辞書	パーザ シソーラス	パーザ		パーザ

4.1.1. EM

入力文とソース側が完全に一致する対訳を検索し、あれば、そのターゲット側を返す。シンプルであるが、高速かつ高品質な訳文を実現する優れた方法であり、他の手法を組合せることで、高精度の話し言葉翻訳を実現できる。

4.1.2. D-cube

文の編集距離を使った類似文検索による EBMT である[13]。編集距離の計算の置換操作に関して、シソーラスを使った単語の意味的な類似性を加味しているのが特徴である。用例翻訳は通常、句を単位に処理するが、D-cube では文を単位にしており、このことで高い翻訳品質を達成した。

また、大量の対訳用例と入力文の編集距離の計算を高速化するために、用例のグループ化、グラフ表現、A*探索を用いた手法を開発した[14]。

4.1.3. HPAT

構文に基づく EBMT である。変換パターンで、ソースのパターンとターゲットのパターンの対応を定義する。入力文はまず、変換パターンのソース側を文法として、構文解析される。次に、ソース側の構文木が、ターゲット側の構文木に、変換パターンに従って写像される。ターゲット側の構文木のターミナルノードは対訳

4.1. 複数の翻訳エンジン

C-cube では、3種の EBMT (EM、D-cube、HPAT)、2種の EBMT と SMT の融合(HPATR、SAT)と1種の SMT (HPATR2) を提案した。

この他に、PMHMTM と MSEP がある[12] が、本稿では、最終評価に用いた手法に説明を限定する。

辞書で変換される [15]。

変換パターンは自動的に獲得されるので、不適切なパターンも混入しうる。品質自動評価手法を用いて、これを削除する手法を提案し、品質改善の効果を確認した[16]。

4.1.4. HPATR

用例に基づく HPAT に、統計的な語彙モデルと言語モデルを導入し、拡張したものである。変換時には訳語を確定せず、可能な候補を列挙するに止める。生成時に語彙モデルと言語モデルに基づく確率が最大になる訳文を選択する[17]。

これは、用例翻訳の計算機構で訳文候補を出し、統計翻訳のモデルで最適な翻訳を求めるための統合方式の一つであり、用例翻訳単独の HPAT に比べ、訳質を改善できた。

4.1.5. SAT

SAT は入力文と訳文の候補からなる「種」から始めて、統計翻訳の翻訳モデル・言語モデルに基づいて訳質が改善するように訳文を編集する。改善が得られなくなるまで greedy に探索し最適な訳文を求める SMT である[18]。

SAT では対訳コーパスから検索された類似用例を「種」とする。類似尺度として、①情報検索で用いられる tf/idf と②編集距離の組み合わせたものを用いる。①で内容の類似性を、

②で構文の類似性を考慮し、両者をバランスさせた類似検索を実現した。

4.1.6. HPATR2

構文に基づく SMT である[19]。語彙モデルと句モデルに文脈自由文法による構文情報を併せて訳文を生成し、語彙モデルと句モデルだけからなる一般的な SMT を上回る翻訳精度を達成できる点が特徴である。

4.2. 最良訳の選択

翻訳品質はシステム毎、入力文毎に異なる。例として、入力文「お支払いは現金ですかクレジットカードですか」に対する D-cube、HPAT、SAT の翻訳を表 2 に示す。

各入力文毎に複数の翻訳エンジンの出力から最良のものを選択できれば、翻訳精度を改善できる。我々は、複数の翻訳候補を相互に比較するために、統計翻訳のモデルに従って、入力と各訳文候補の対の妥当性を測る手法を提案した[20]。本方式は言語モデルに翻訳モデルを加えた点と検定を行う点が従来方式と異なり、そのことによって、より良い性能を達成できる。

表 2 訳文のヴァリエーション例

- | |
|---|
| 1. Would you like to pay in cash or with a credit card? |
| 2. Is the payment cash? Or is it the credit card? |
| 3. Could you cash or credit card? |

5. 評価

ATR のコーパス BTEC と MAD を用いて、日英・日中の翻訳実験を行って評価した。

- コーパスサイズを、2 万、16 万、50 万、70 万に変化させた。
- 機械翻訳の自動評価尺度の mWER を用いた。mWER とは multiple Word Error Rate の略であり、機械翻訳結果と複数の参照訳との編集距離を入力長で正規化したものの最小値。値が小さい方が良い。また、主観評価との相関が高いことが分かっている。

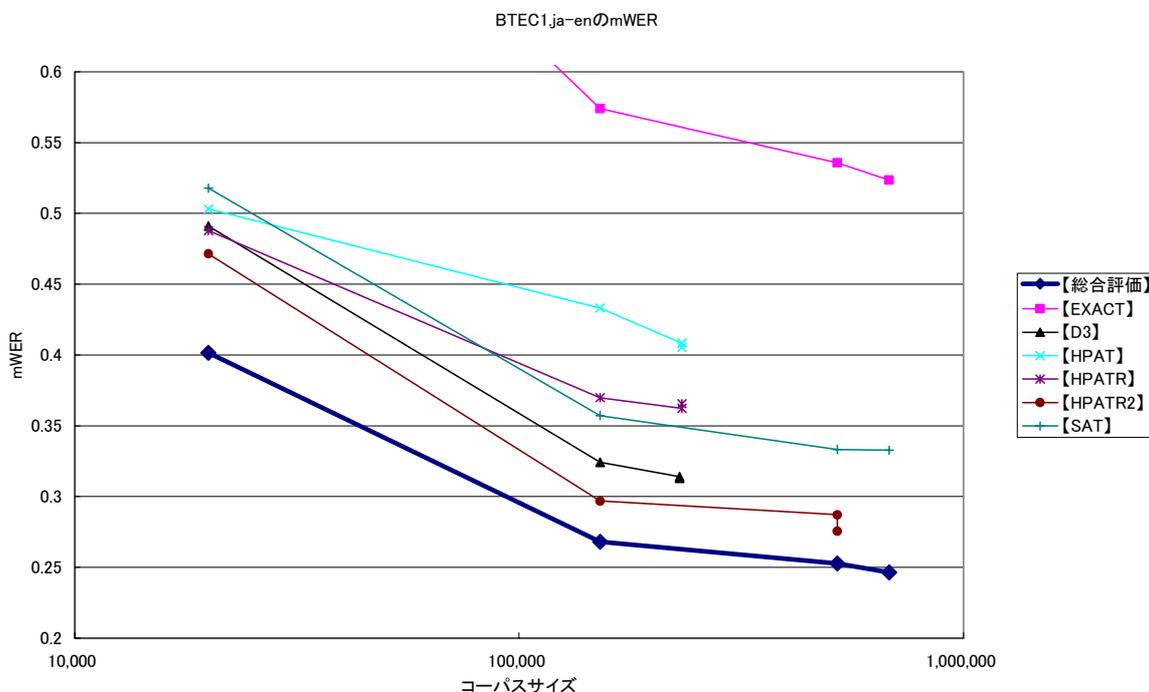


図 1 サイズと訳質 (BTEC)

5.1. 結果

図 1 に主要な評価結果を示す。

- 全ての翻訳エンジンにおいてコーパスサイズ増加に伴い、mWER 値が下がる。
- 4.2 節の選択により、mWER 値はどの要素

エンジンよりも低くなる。
どのテストセットでも、どの言語方向でも同じ結果が得られた。

5.2. 翻訳性能の意味

達成された mWER はどれだけ役に立つ翻訳品質なのだろうか？

ATR では、英語能力の検定試験として有名な TOEIC スコアに着目し、同スコアが既知である人間による翻訳文と機械による翻訳文を比べて、機械のスコアを推定する方法[21]を開発し活用している。

5.1 節の性能は、海外部門のビジネスマンの平均スコアに匹敵し充分役に立つ性能に達した。

6. まとめ

ATR は多言語コーパスと翻訳のための機械学習アルゴリズムを開発し、高精度な翻訳部の構築手法として有用であることを実証した。

謝辞

本研究は情報通信研究機構の研究委託「大規模コーパスベース音声対話翻訳技術の研究開発」により実施したものです。類語新辞典体系の研究利用を許可して下さった(株)角川書店に深謝いたします。また、過去・現在の C-cube のメンバーに感謝します。

参考文献

1. Satoshi, Nakamura, et. al., "The ATR Multilingual Speech-To-Speech translation System," the IEEE Transactions on Speech and Audio Processing, 2005 (to appear).
2. Genichiro Kikui, Toshiyuki Takezawa, Seiichi Yamamoto. Multilingual Corpora for Speech-to-Speech Translation Research, ICSLP 2004(8th International Conference on Spoken Language Processing), Vol.2, pp. 357-360, October, 2004.
3. Sumita, E., "Corpus-Centered Computation," Proc. of S2S workshop of 40th ACL, pp. 1—8, (2002).
4. Nagao, M., 1984. A Framework of a Mechanical Translation between Japanese and English by Analogy Principle, in A. Elithorn and R. Banerji (eds), Artificial and Human Intelligence, Amsterdam: North-Holland, pp. 173--180.
5. Ney, H., 2001. Stochastic Modeling: From pattern classification to language translation, in Proc. of the ACL 2001 Workshop on DDMT, pp. 33--37.
6. 下畑 光夫, 渡辺 太郎, 隅田 英一郎, 松本 裕治. パラレルコーパスからの機械翻訳向け同義

- 表現抽出. (2003)情報処理学会論文誌, Vol44, No11, pp2854-2863.
7. Andrew, Finch, Taro Watanabe, Yasuhiro Akiba and Eiichiro Sumita. Paraphrasing as Machine Translation, JNLP, 11 (5), pp.87—111, October 2004.
 8. Takao DOI and Eiichiro SUMITA, "Splitting Input for Machine Translation Using N-gram Language Model Together with Similarity," IEICE TRANS. INF. & SYST., Vol. E88-D, No. 6, pp. 1256-1264, 2005.
 9. 隅田 英一郎, 佐々木 裕, 山本 誠一. 機械翻訳システム評価法の最前線. 情報処理, Vol.46, No.5, 通巻 483 号, pp.552-557. (2005,5)
 10. <http://www.slt.atr.jp/IWSLT2004/archives/000619.html>
 11. <http://penance.is.cs.cmu.edu/iwslt2005/proceedings.html>
 12. Michael Paul, Takao Doi, Youngsook Hwang, Kenji Imamura, Hideo Okuma, Eiichiro Sumita. Nobody is Perfect: ATR's Hybrid Approach to Spoken Language Translation, Proc. of IWSLT 2005, pp.55-62 (2005,10)
 13. Eiichiro Sumita. An Example-based Machine Translation Using DP-matching between Word Sequences.(2001), ACL-2001 (39th Annual Meeting of the Association for Computational Linguistics) Workshop on Data-driven MT, pp. 1-8.
 14. 土居 蒼生, 隅田 英一郎, 山本 博史. 編集距離を使った用例翻訳の高速検索方式と翻訳性能評価. (2004,4) 情報処理学会論文誌, Vol.45, No.6, pp.1681-1695.
 15. 今村 賢治. 構文解析と融合した階層的句アライメント. (2002) 自然言語処理, Vol.9, No.5, pp.23-42.
 16. 今村 賢治, 隅田 英一郎, 松本 裕治. 自動評価を用いた機械翻訳規則のフィードバッククリーニング. (2004,8)情報処理学会論文誌, Vol.45, No.8, pp.2068-2077.
 17. Kenji Imamura, Hideo Okuma, Taro Watanabe, and Eiichiro Sumita. Example-based Machine Translation Based on Syntactic Transfer with Statistical Models. The 20th International Conference on Computational Linguistics (COLING 2004), pp. 99-105, 2004.
 18. Watanabe, T., and Sumita, E., "Example-based decoding for Statistical Machine Translation," MTS-IX, pp. 410-417, (2003)
 19. Kenji Imamura, Hideo Okuma, Eiichiro Sumita. Practical Approach to Syntax-based Statistical Machine Translation. (2005,9), Proc. of MTS-X, pp. 267-274.
 20. Akiba, Y., Watanabe, T., and Sumita, E., "Using Language and Translation Models to Select the Best among Outputs from Multiple MT systems, Proc. of Coling, pp. 8—14, (2002).
 21. 菅谷他, 音声翻訳システムと人間との音声翻訳能力評価手法の提案と比較実験, 信学論, J84-D-II, 11, pp. 2362-2370, 2001.