

3.5.2 ユニバーサルコミュニケーション研究所 多言語翻訳研究室

室長 隅田英一郎 ほか9名

多言語翻訳システムの構築に必要な対訳データと翻訳アルゴリズムの研究開発

【概要】

本研究室は、人と人との言葉の壁を克服するため、日本語と英語のような異なる言語間の翻訳の研究を実施している。対訳データ(原文と訳文の対を集積したもの)に基づいて翻訳する手法を採用し、自動化やコミュニティとの協業など新たな手法によって対訳データの構築を効率化し、同手法の基盤になる大規模な対訳データを構築してきている。

平成 25 年度は、従来までの翻訳アルゴリズムを改善し、特に、長文(話し言葉(10 語以上)、書き言葉(20 語以上))に関して高精度翻訳を実現した。この成果は、多数の技術移転、受賞、第一級国際会議、論文誌採録、という形で、産学官から広く評価されている。

【平成 25 年度の成果】

(A) 【話し言葉(10 語以上)の高度化】前年度に収録した同時通訳者による模擬通訳の対訳コーパスを活用して、同時通訳の要となる「入力途中から出力する」五月雨(漸次)翻訳(図1「五月雨翻訳」からも分かるように、従来技術で実現できる逐次通訳にはなく同時通訳にはある効用は応答時間の短縮であり、コミュニケーション促進に寄与すると期待される)の入力の分割点を分析し、さらにコーパスから学習できる句対応モデルと構文に基づく自動分割法の検討と翻訳アルゴリズムの変更を開始した。

(B) 【書き言葉(20 語以上)の高度化】長文の翻訳に不可欠な解析技術を改良(精度の向上と記憶容量の削減を同時に実現)し、特に、40 語以上の長文翻訳の翻訳率を大幅に改善した(表1)。改良内容は以下のとおり。

- (i) 構文データを翻訳に適した構造に変換
- (ii) 変換後のデータの単語ラベルを、別途 TEXT から学習したクラスで置換
- (iii) 潜在的な文法記号を使った確率的文脈自由文法を学習

(C) 既学習のモデルを事前確率とし、追加対訳を学習する「階層的なモデルによる逐次学習」(図2)を提案し、従来法にあった性能劣化を克服した。

(D) KDDI 株式会社による音声翻訳の採用(おはなしアシスタント、図3)など翻訳技術を8件実用化した。

(E) 音声翻訳を次の3領域で展開。①国際展開(U-STAR): 音声翻訳のサーバ・クライアントのプログラムをオープンソースとして公開し、参加機関数を増大させた。②医療分野への展開: 医療分野において、日本語の模擬対話とその英語訳からなる対訳コーパス(1 万文)を作成した。翻訳精度と英語音声認識性能が大幅に向上した。③聴障者支援アプリの改良: 高齢者対応及び利用者の声を踏まえた GUI の改良を行い、ハイブリッド方式を実現した。

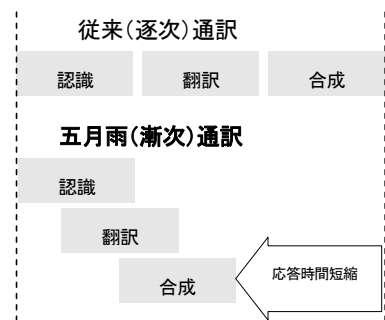


図1 「五月雨翻訳」の効用

表1 40 語以上の長文の翻訳性能の大幅な改善(特許英日)

	翻訳(通じる)率 (%)		
	全体	40 語未満	40 語以上
改良前	67	71	54
改良後	84	87	75

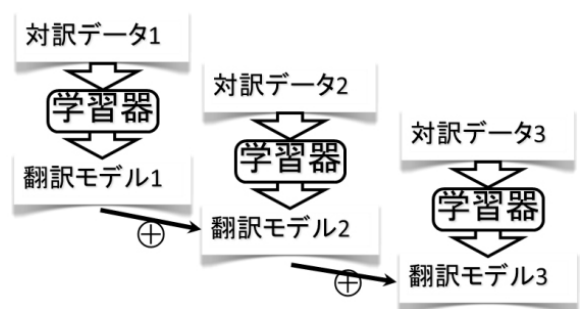


図2 逐次学習

(F) ① 国際会議 Workshop on Future Directions in Translation Research (WFDTR、10月)と、②コンペ型国際会議 NTCIR/PatentMT (6月)及び③ International Workshop on Spoken Language Translation (IWSLT、12月)の開催によって多言語翻訳分野の発展へ貢献した。

(G) さらに、年度計画を上回る成果は以下のとおりである。

- 【機械翻訳のための係り受け木における教師なし品詞推定、Akihiro Tamura, Taro Watanabe, Eiichiro Sumita, Hiroya Takamura, Manabu Okumura. Part-of-Speech Induction in Dependency Trees for Statistical Machine Translation. The 51st Annual Meeting of the Association for Computational Linguistics (ACL 2013).】原言語の単語に加えて目的言語の単語から、隠れ状態として品詞タグ Z を推定する(図4)ことにより、単言語の既存知識に基づく品詞タグにある翻訳に向いていない点を自動的に修正してシステムの精度改善を実現した。

- 【ラベル系列の識別に基づく語順推定モデル、Isao Goto, Masao Utiyama, Eiichiro Sumita, Akihiro Tamura, Sadao Kurohashi. Distortion Model Considering Rich Context for Statistical Machine Translation. The 51st Annual Meeting of the Association for Computational Linguistics (ACL 2013).】相対的な語順を考慮でき距離の影響を学習できる手法を提案し、翻訳システムの大幅な精度改善を実現した。

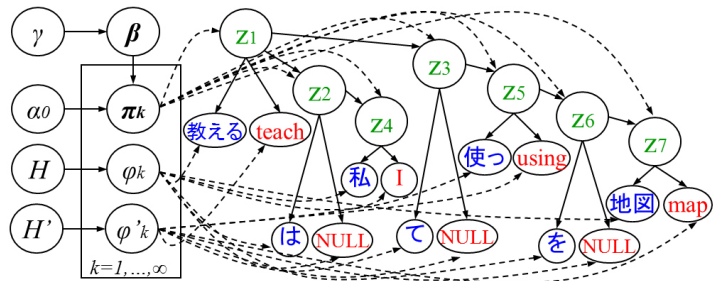


図4 機械翻訳のための品詞推定

- 【加法型ニューラルネットワークによる翻訳、Lemao Liu, Taro Watanabe, Eiichiro Sumita, Tiejun Zhao. Additive Neural Networks for Statistical Machine Translation. The 51st Annual Meeting of the Association for Computational Linguistics (ACL 2013).】局所的な非線形モデルと非局所的な線形モデルとを結合し、文単位に学習(ニューラルネットワークにより、素性表現を自動的に学習する)。

(H) 主要国際論文誌6件(5件 ACL、1件 EMNLP)・最難関国際会議6件(1件 AI Magazine、1件 Machine Translation、1件 ACM Transactions on Interactive Intelligent Systems (TiS)、3件 ACM Transactions on Asian Language Information Processing (TALIP))で採録された。

(I) 「ネットワーク音声翻訳技術の実用化」で、第11回産学官連携功労者表彰総務大臣賞(平成25年8月29日、NICT多言語翻訳研究室室長・隅田英一郎、株式会社フィート代表取締役社長・小林照二、成田国際空港株式会社代表取締役社長・夏目誠)など3件受賞した(図5)。

(J) 開発したプログラムをオープンソースとして(cicada、http://www2.nict.go.jp/univ-com/multi_trans/cicada/、expgram、http://www2.nict.go.jp/univ-com/multi_trans/expgram/)公開した。また、講習会を開催(平成26年3月4~7日)し、自動翻訳技術の教科書を出版(コロナ社、平成26年1月22日)した。



図3 KDDI採用



図5 総務大臣賞受賞