

多様な利用シーンに対応する話し言葉翻訳サービスの開発

訪日外国人旅行者数は増加し続け、2020年には4,000万人になると言われており、多くの外国人のお客様が快適にコミュニケーションを図れることが期待されている。また、企業のグローバル化も進み、多言語でのコミュニケーションが行われる機会も増えている。

本稿では、ドコモが開発に取り組んでいる翻訳サービスのうち、多様な利用シーンが想定される、話し言葉を翻訳する会議翻訳、SNS翻訳、接客翻訳について、実現における課題と解決策を解説する。

サービスイノベーション部

たけいち まさと おの たかや
武市 真知 小野 隆哉
ちぢわ ゆうき じゃん いーしん
千々和 祐貴 姜 一欣

1. まえがき

2015年の訪日外国人旅行者数は、過去最高であった2014年をさらに上回り1,974万人に達した[1]。日本政府から発表された2020年のその目標数は4,000万人であり[2]、日本を訪れるすべての外国人旅行者が、ストレスなくコミュニケーションを図れる環境の実現が期待されている[2]。他方、国内全法人の海外現地法人従業員数は、2010年度には499万人であったが、2014年度には575万人まで増え[3]、企業のグローバル化も進んでおり、多言語でのコミュニケーションが行われる機会も増えてきている。

ドコモは、2020年に向けて言語や文化の壁を越えたコミュニケー

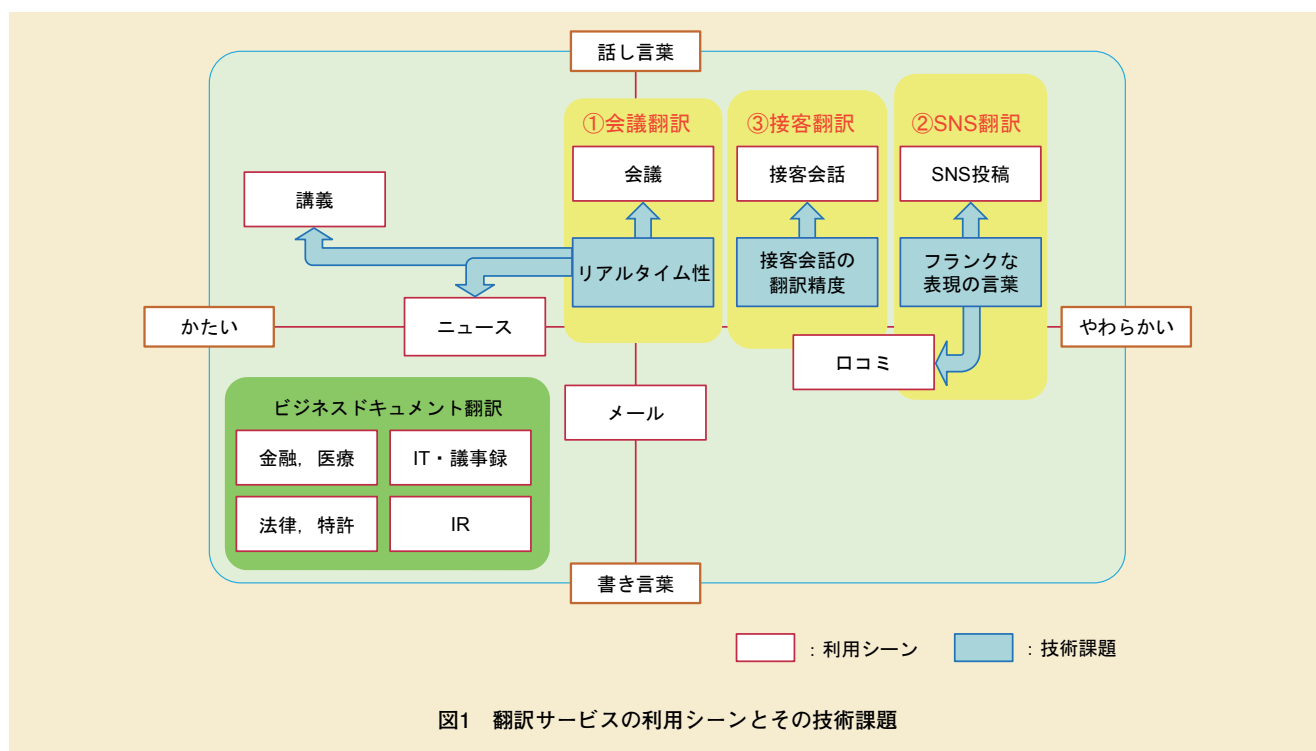
ションを実現するため、音声認識*1技術、機械翻訳*2技術の開発とそれを応用したサービス開発に取り組んでいる。

図1は、翻訳サービスを「話し言葉・書き言葉」と「文章表現のかたい・やわらかい」を表す軸上に分類し、技術課題を明確化したものである。ドコモは、訪日外国人旅行者向けに、図1右上の「話し言葉」かつ「やわらかい」領域を直近のターゲットとし、音声認識と機械翻訳の精度向上をめざしている。また現在、このターゲットに対して以下の3つのサービスを開発している。

①会議翻訳とは、異言語間の会議音声リアルタイムに母国語に翻訳・読み上げ・テキスト表示する翻訳サービスである。

②SNS翻訳とは、SNSで主に使われているフランクな表現の言葉をテキスト翻訳するサービスである。SNSの投稿は一見すると書き言葉に思えるが、「やわらかい」「話し言葉」の表現が多く含まれている。

③接客翻訳とは、店舗などでのお客様と店員との間での接客会話を音声翻訳するサービスである。簡単なコミュニケーションは音声認識と機械翻訳によって気軽に接客を行い、複雑なコミュニケーションは遠隔の通訳者と通話することで、正確で安心なコミュニケーションが行える。2014年試作開発に着手して以降、複数の企業と実証実験を行い、UI*3改善および、音声認



識と機械翻訳の精度向上を繰り返した。そして2016年6月には、はなして翻訳® for Biz^{*4}として商用サービスが提供された。

本稿ではこれら3つのサービスの概要と課題を述べ、それらに対応するドコモの取組みについて解説する。

2. 各翻訳サービスの概要と課題

2.1 翻訳サービスの概要

翻訳サービスは、一般的に音声認識、機械翻訳、音声合成^{*5}の3つの技術から成り立っている。図2にそのシステム構成を示す。

- ①音声認識とは、音声をテキスト化する技術である。まず発話区間検知^{*6}にて発話部分を特定す

る。発話の終了は一定時間以上の無音区間^{*7}によって判定している。その後、雑音除去^{*8}にて背景雑音の除去を行い、音声認識エンジンの音響モデル^{*9}/言語モデル^{*10}を用いて、音声から認識結果テキストに変換する。

- ②機械翻訳とは、テキストを機械的に他言語のテキストに翻訳するための技術である。翻訳前処理として翻訳前の言語を精度向上のために構文解析^{*11}や固有表現クラス化^{*12}といった処理を行い、それから機械翻訳エンジン^{*13}の翻訳モデル^{*14}/言語モデルを用いて指定の言語のテキストに翻訳する。その後、後処理にて辞書の対訳データを用いてクラス化した単語の置換な

どを行い、翻訳結果を作成する技術である。

- ③音声合成とは、翻訳結果テキストから人工的に音声データを作り出し、読み上げる技術である。

コーパスには、音声認識用のコーパスと機械翻訳用の対訳コーパスがある。音声認識用のコーパスとは、発話された音声とテキストをセットにしたデータであり、サービスで蓄積されたログなどから作成する。機械翻訳用の対訳コーパスとは、翻訳前の言語のテキストと翻訳後の言語のテキストを1文ずつ対にした対訳（例：いらっしゃいませ⇔May I help you?）である。

一般的に音声認識、機械翻訳の精度を向上させるには、図1の赤枠で

^{*4} はなして翻訳® for Biz：訪日外国人観光客と日本人店員の接客をサポートするドコモの法人向け翻訳サービス。2016年6月より提供開始。はなして翻訳は(株)NTTドコモの登録商標。

^{*5} 音声合成：テキストから人工的に音声データを作り出し、テキストを読み上げできるようにする技術。

^{*6} 発話区間検知：音声信号から音声が存在する区間とそれ以外の区間を判別する技術。

^{*7} 無音区間：音声が存在しないと判断される区間。

^{*8} 雑音除去：街頭の雑音や他者の発話など、特定の音声を認識するために他の音声を取り除く処理。

^{*9} 音響モデル：認識対象の音素がどのような

周波数特性を持っているかを集計した統計モデル。

^{*10} 言語モデル：形態素の並び方とその並び方の頻度情報を集計した統計モデル。

^{*11} 構文解析：解析対象となる文章の、文節間の係り受けの構造を解析して文法的な関係を明らかにする技術。

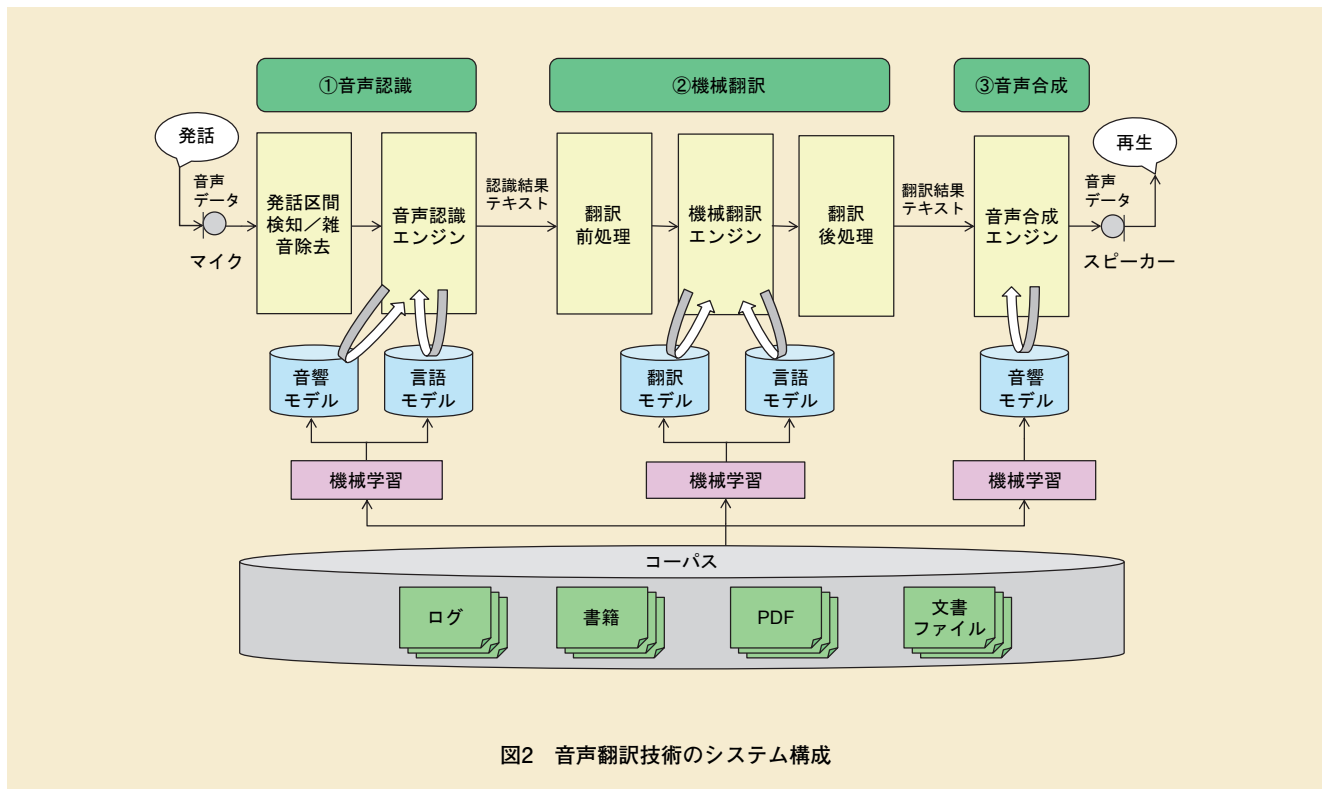


図2 音声翻訳技術のシステム構成

示した利用シーン（会議、SNS投稿、接客会話など）で頻繁に使用される文章を対訳コーパスとして、言語対^{*15}（日英、日中、日韓など）別に大量に収集することが必要である。この大量の対訳コーパスを学習用データとして機械学習^{*16}を行い、利用シーンに特化した翻訳モデルと言語モデルを作成する。同様に音声認識においても、利用シーンに応じたコーパスを学習用データとして、機械学習により音響モデルと言語モデルを作成し、音声合成においても、同様に機械学習による音響モデルを作成する。

2.2 各翻訳サービスの課題

次に各サービスで特有の課題につ

いて述べる。

- ①会議翻訳では、音声認識と機械翻訳の結果表示に「リアルタイム性」が求められるとともに、発話の区切りを正確に検出する必要がある。従前技術では、1人の会議参加者の発話を音声認識する場合に、発話が終了してから音声認識結果のテキストを表示するため、表示までの間は待たされている感覚が強く感じられた。
- ②SNS翻訳では、SNSの投稿において頻出する「フランクな表現の言葉」への対応に課題がある。従前技術では、「おたんじょーびおめでとー」や「これはヤヴァい」といった翻訳前のテキ

ストを分かち書きできない。

- ③接客翻訳では、店員による丁寧な接客と、お客様によるフランクな質問の両方に対応する必要がある。しかし、接客会話に必要な対訳コーパスが不足しているため「接客会話で話される対話文に則した機械翻訳精度の向上」が課題である。

3. 会議翻訳

会議翻訳は、図3のように、異なる言語で話す会議参加者の会話をリアルタイムで音声認識し、他言語に翻訳するサービスである。またキーボード入力によるテキスト翻訳も可能であり、発話ができない状況でもテキストによる会議参加が可能である。

*12 固有表現クラス化：入力文の固有表現をあらかじめ固有名詞を表すラベルに置き換え、機械翻訳後、辞書を用いて置換する技術。

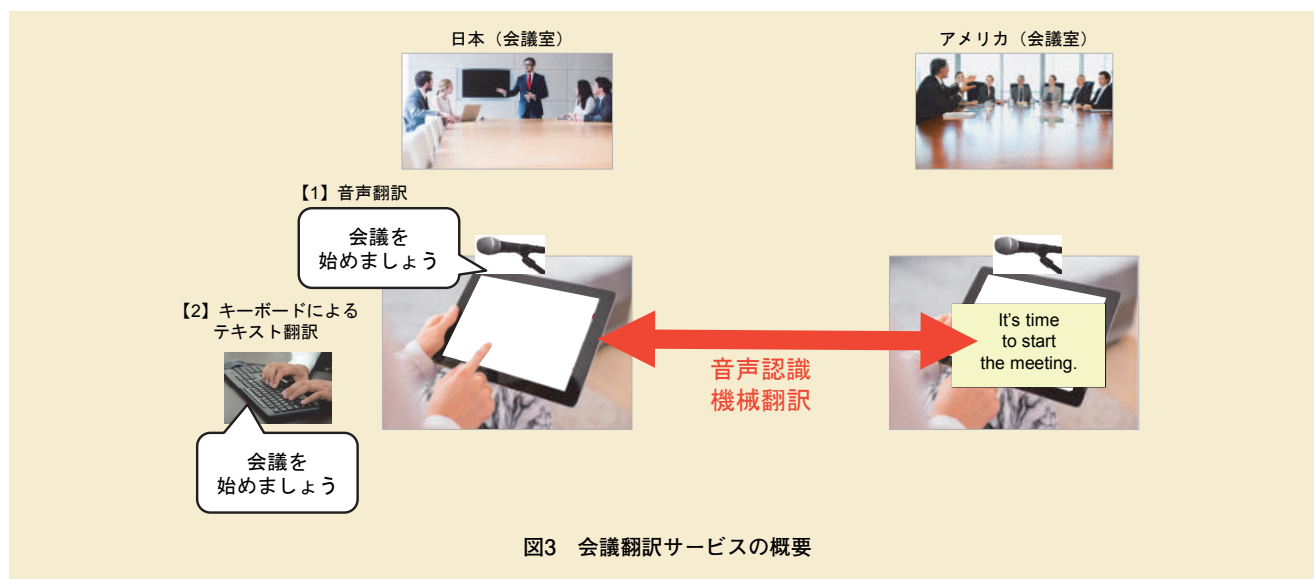
*13 機械翻訳エンジン：機械翻訳向けに学習された言語モデルと翻訳モデルを利用して、統計的にテキストを翻訳するソフトウェア。

*14 翻訳モデル：翻訳前の言語と翻訳後の言語の各文章について、それぞれの単語同士

が意味的に対応するかを計算するために利用される統計モデル。

*15 言語対：翻訳元と翻訳先となる2言語の組合せ（英語・日本語など）のこと。

*16 機械学習：事例をもとにした統計処理により、計算機に入力と出力の関係を学習させる仕組み。



(1)リアルタイム性の向上

テンポよく会議を実施するためには、音声認識も機械翻訳も会議参加者が体感する処理遅延を小さくし、「リアルタイム性」を高くすることが望ましい。

そこで、音声認識の処理において、形態素^{*17}ごとの最尤推定結果を逐次的に表示する処理を行っている。これにより会議参加者の発話と表示の遅延が小さくなり、体感的な「リアルタイム性」を向上させている。

(2)発話区間の自動検知

音声認識を利用するサービスの中には、ユーザが発話区間を明示的に示すために発話・終話をボタンによって制御する場合がある。しかし、会議翻訳の場合は、発話のやりとりが連続的に発生するため、発話・終話のボタンを設けずに発話区間を自動検出できることが望ましい。

また、発話者が次の発話を考える際などに起こる間や、言いよどみに

よって音声の無音区間が続くことで、別の発話として区切られる場合がある。具体的には図4①の場合、「このアプリは音声認識の結果を」「リアルタイムに表示します」のように区切られてしまう。その結果、「機械翻訳」の結果も「This application is speech recognition results.」「Real-time display.」のように発話者の意図と異なる翻訳結果が出力されてしまう。

解決策として、会話中に言いよどみが発生しても連続した発話として検出するために、「無音区間」を判定するパラメータを調整した。具体的には、発話者の各会議会話から言いよどみで発生する途切れの「無音区間」を抽出し、各々の「無音区間」の平均値を算出した。この値を基に発話が区切られる時間として設定するパラメータを最適化し、会話が正しく区切られるか実証した。

これにより図4②の場合は、「こ

のアプリは音声認識の結果をリアルタイムに表示します」といったように、発話の区切れが解消され、「This application is displayed in real-time voice recognition results.」という正確な翻訳結果を出力することができる。

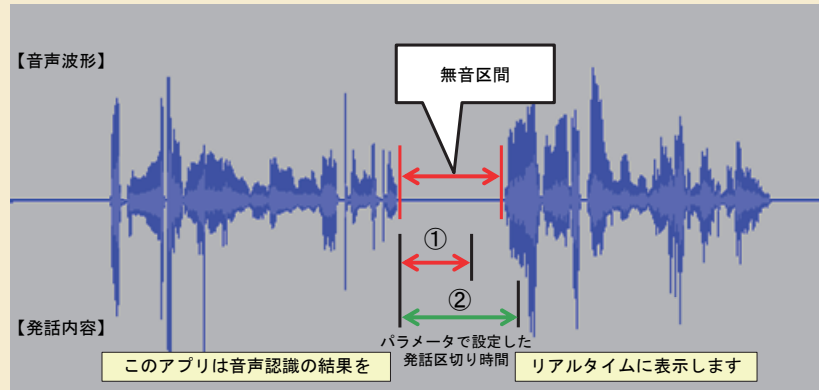
4. SNS翻訳

SNS翻訳とは、SNS投稿でのフランクな言葉を、他言語に翻訳することができるサービスである。このSNS翻訳システムの処理の流れを図5に示す。SNS翻訳システムは、他とは異なり、図2の機械翻訳機能部のみで構成されている。

SNS翻訳では、通常の機械翻訳機能における翻訳前処理に2つの機能を追加した。1つめは文分割機能、2つめは崩れ文字列正規化機能である。

・文分割機能では、一般的な文区切り文字判定に加えて1文中に

*17 形態素：ある言語において、意味を持つ最小の単位。



翻訳結果

【①の場合】
このアプリは音声認識の結果を → This application is speech recognition results.
リアルタイムに表示します → Real-time display.
【②の場合】
このアプリは音声認識の結果をリアルタイムに表示します
→ This application is displayed in real-time voice recognition results.

図4 言いよどみによる「無音区間」

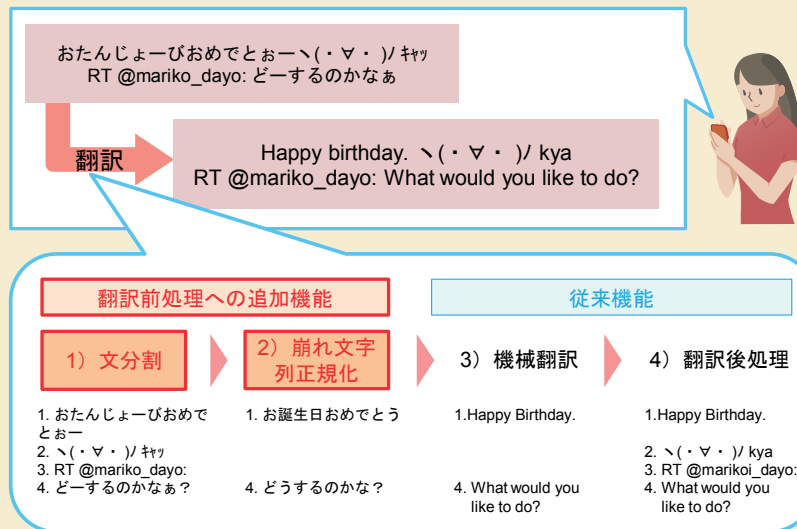


図5 SNS翻訳システムの処理の流れ

複数ある括弧判定，顔文字判定（例：\(.・∀・\)ノ キャ），サービス固有記号判定（例：RT），URL判定，擬音語・擬態語の

判定などにより，文分割を実施している．その結果，翻訳対象とすべきテキストと認めた部分のみを，次の崩れ文字列正規化

機能で処理する．

・崩れ文字列正規化機能では，従前の整った日本語に対する形態素解析を強化し，フランクな表

現にも対応可能とした崩れ形態素解析[4][5]を用いて「表記」に分割する。ここでの表記とは、翻訳前の入力文を日本語の最小単位の単語に分割したものを表す。その後、表記ゆれ辞書と変換候補制御リストを用いて代表表記選択ラティス^{*18}を生成し、機械翻訳エンジンのモデル学習時に利用した膨大なコーパスを利用した言語モデルを用いて、崩れた表記を最適な表記に変換する。さらにネガポジ判定による置換処理により多義性に対応している。

- (1)代表表記選択ラティス生成と崩れた表記の変換の仕組み
代表表記選択ラティス生成と崩れ

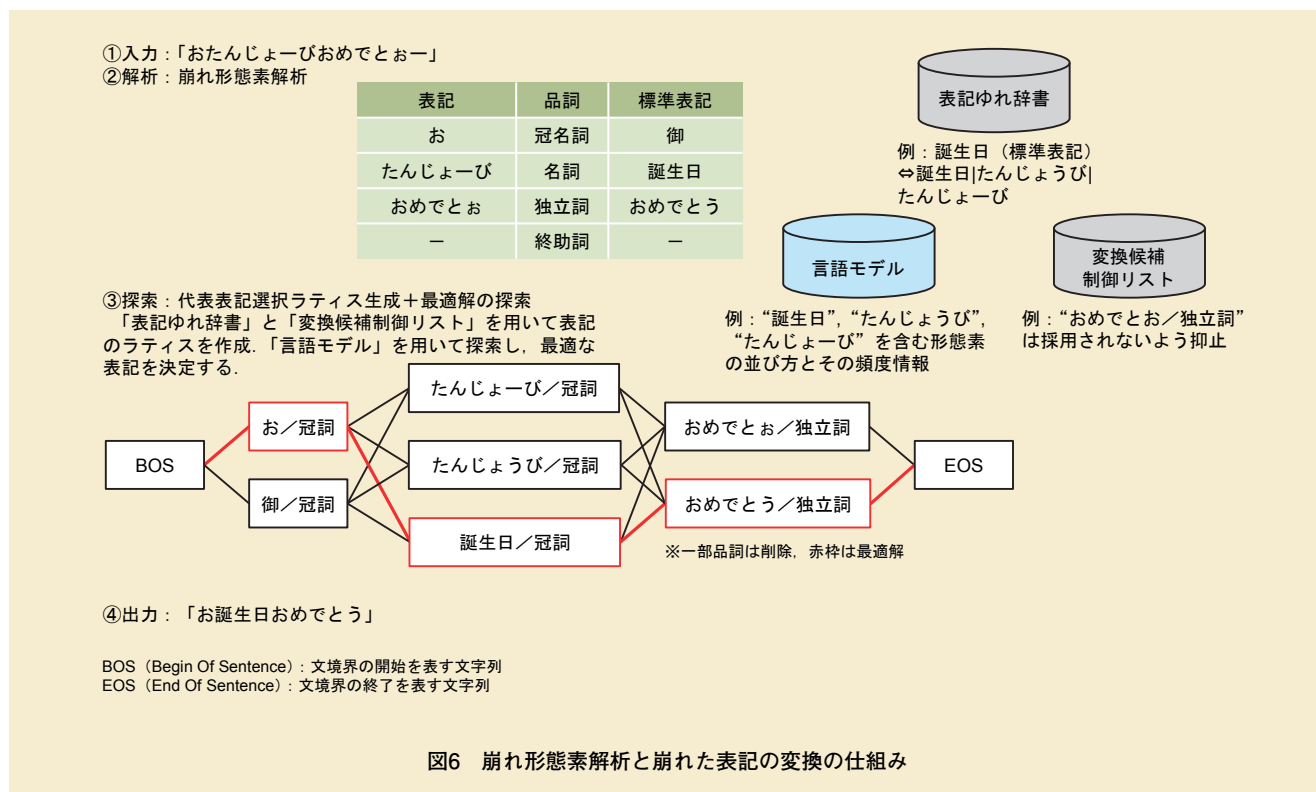
た表記の変換を、図6を用いて解説する。

崩れ形態素解析の結果から入力文の表記と品詞と標準表記^{*19}のリストデータを作成する。入力文「おたんじょーびおめでとー」の場合、表記「お」「たんじょーび」「おめでとー」「ー」に分割される。またそれぞれに標準表記の「御」「誕生日」「おめでとう」「ー」を付与する(図6①②)。この標準表記とは新聞などで用いられる標準的な表記を表す。

次に、標準表記に着目し、標準表記に紐づく入力文に含まれない表記と変換されたくない候補や一意に変換したい候補を含む変換候補制御リストの表記を用いて、選択可能な表記のラティスを作成する。例えば標

準表記、「誕生日」に紐づく表記には「たんじょーび」「たんじょうび」「誕生日」があるため、これをすべて利用して、図6のように表記のラティスを作成する。表記のラティスとは選択可能な表記を列挙したグラフ構造を表す。この表記のラティスに対し、機械翻訳エンジンで利用している言語モデルを用いて表記の系列を探索し、最適な表記を決定する(図6③)。図6の場合には「お」「誕生日」「おめでとう」が最適解となる(図6④)。この代表表記変換によって崩れ文字列を正規化した結果を得る。

なお崩れ形態素解析機能と代表表記変換機能は、NTTメディアインテリジェンス研究所の技術支援・成



^{*18} ラティス: 横方向に形態素の系列を、縦方向に同じ意味であるが表記の異なる形態素を並べた格子状のデータ。

^{*19} 標準表記: ゆらぎのない文法に沿った標準的な書き方に従う字や記号で表された文字列。

果提供を基にして、ドコモで作成した言語モデルや変換候補制御リストを組み込み、翻訳前処理部に実装した。

(2)ネガポジ判定

日本語には多義性があり、ひとことや1文のみでは意図を特定するのが困難な場合がある。例えば、食べ物に関する評判で「これはヤヴァい」といったSNSの投稿を1文ずつ翻訳すると「美味しい」という意味と「不味い」という2通りの解釈がありえる。これを前後文章のネガポジ判定^{*20}を加えることで、ポジティブの場合には「これはすごい」、ネガティブの場合には「これはひどい」に置換する処理を実施する。このような置換処理によって多義性に対応している。

SNS翻訳では、これらの文分割機能と崩れ文字列正規化機能によって正規化された翻訳すべき文章のみを機械翻訳し、翻訳後処理にて入力文の構造に翻訳結果を埋め戻して、ユーザ側に翻訳文として応答している。

また、この翻訳前処理によって、日本語から英語／中国語／韓国語への翻訳における「やわらかい」「話し言葉」の翻訳精度は代表的な他社エンジンに比べて高い性能を確認した。図7にSNS翻訳の日英翻訳における精度比較結果を示す。評価文はSNSのグルメ／コスメ／旅行に関する投稿文を無作為に100文抽出した。それを評価者3名による主観評

価にて次の基準の平均値により評価した。「○評価（3点）：正しく訳されている」と「△評価（2点）：一部誤りが見受けられるが、意味は通じる」の合計が、SNS翻訳エンジンは他社エンジンよりも45点上回った。

5. 接客翻訳

接客翻訳では、日本語を話す店員の丁寧な接客文章を多言語に訳す場合と、お客様の外国語での必ずしも丁寧ではない質問を日本語に訳す場合の両方に対応する高い機械翻訳精度が必要である。接客対応で頻繁に話される文章を対訳コーパスとして集め、翻訳エンジンに学習させることで接客会話専用の翻訳モデルを作成した。また、この対訳コーパスを音声認識エンジンの言語モデルに適

用することで、音声認識エンジンも接客会話に対応させる取組みを行った。

以下に対訳コーパス収集について解説する。

(1)一般的な対訳コーパスの収集

一般的に考えられる対訳コーパス収集の流れを図8に示す。まず、以下の方法で対訳コーパスの元となる対訳データを収集する。

- ・サービス利用規約に則り、商用サービス（はなして翻訳、メール翻訳など）のログを書き起こし、収集する。
- ・販売されている対訳コーパスを利用する。

これらの方法で収集した対訳データに対して、以下の整形を行うことで対訳コーパスを作成する。

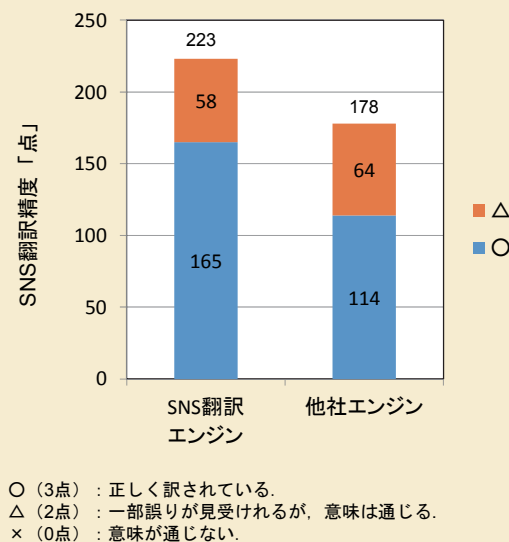
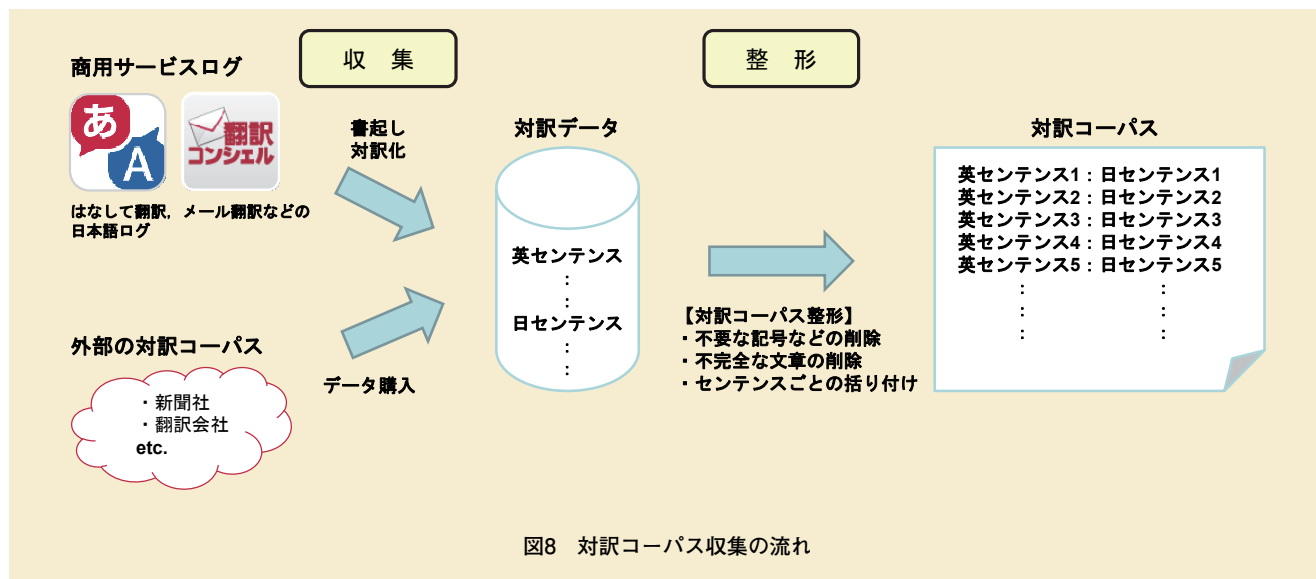


図7 SNS翻訳精度の主観評価結果（日英翻訳）

^{*20} ネガポジ判定：文書による筆者の意図がネガティブ（negative）かポジティブ（positive）かを判定する手法。

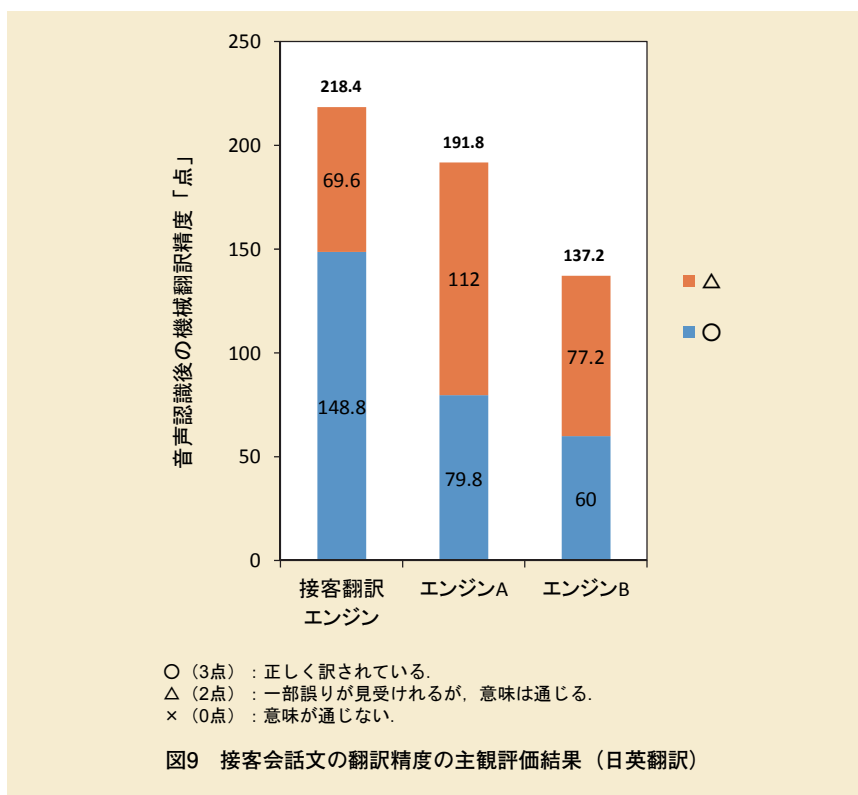


- ・収集したコーパスに含まれている不要な記号などの削除
- ・文章として不完全なコーパスの削除
- ・センテンスごとの括り付け

(2) 人手による対訳コーパスの作成

今回は一般的な収集方法に加え、利用シーンに特化したコーパスを増やす取組みとして、クラウドソーシング*21などを活用し、人手で接客会話を想定した対訳コーパスを作成する作業を行った。この取組みで得られた対訳コーパスを用い、接客用の翻訳モデル、音声認識の言語モデルを作成した。

この結果、各言語ともに音声認識、機械翻訳精度が向上した。図9に日英翻訳での精度比較結果を示す。評価文は書籍やログから無作為に抽出した接客会話200文を作成した。それを評価者5名によって主観評価した。今回、作成した接客会話の音声



認識、機械翻訳の結果の平均値は218.4点となり、代表的な他社エンジンの評価を上回っていることが確認できた。

6. あとがき

本稿では、話し言葉の翻訳とその技術課題を解決するための取組みと

*21 クラウドソーシング：クラウドソーシングは群衆（crowd）と業務委託（sourcing）を組み合わせた造語で、多数の人に分散して業務を委託するという新しい雇用形態。

して、会議翻訳、SNS翻訳、接客翻訳の各種取組みについて解説した。

今後は本技術にさらに磨きをかけて実用性を一層高めるとともに、翻訳文の単語を事前に並べ替えることで翻訳元の言語と翻訳先の言語の語順相違による精度低下を改善する「翻訳事前並べ替え技術」や、話し言葉で頻出する省略語を埋めるための自然対話技術などを取り入れていくことで、さらなる翻訳精度向上を達成できるよう研究開発を進めてい

く予定である。

文 献

- [1] 観光庁：“観光白書（2016年版）。”
- [2] 首相官邸：“明日の日本を支える観光ビジョン,” Mar. 2016.
http://www.kantei.go.jp/jp/singi/kanko_vision/
- [3] 経済産業省 大臣官房調査統計グループ企業統計室 貿易経済協力局貿易振興課：“第45回海外事業活動基本調査結果概要—平成26（2014）年度実績—,” Jul. 2015.
http://www.meti.go.jp/statistics/tyo/kaigaizi/result/result_45/pdf/h2c45kaku1.pdf
- [4] 斉藤 いつみ, 貞光 九月, 浅野 久子, 松尾 義博：“正規—崩れ文字列アライメントと文字種変換を用いた崩れ表記正規化に基づく日本語形態素解析,” 言語処理学会第20回年次大会発表論文集, pp.777-780, Mar. 2014.
- [5] 斉藤 いつみ, 貞光 九月, 浅野 久子, 松尾 義博：“崩れ表記語の生成確率を用いた表記正規化と形態素解析,” 言語処理学会第21回年次大会発表論文集, pp.51-54, Mar. 2015.