言語間の並べ替えを考慮した同時音声翻訳のための訳出タイミング決定法*

Graham Neubig, 藤田朋希, Sakriani Sakti, 戸田智基, ○中村哲 (奈良先端大)

1 はじめに

音声翻訳システムは長年の研究・開発で精度が大幅に向上している一方，同時性の高い翻訳である同時通訳は未だに実現されていない．その理由は，音声翻訳システムの 3 つの処理部である音声認識 (ASR)，機械翻訳 (MT)，音声合成 (TTS) の関わり合いにある．通常，認識が終了した時点で翻訳が開始され，翻訳が終了した時点で合成が開始されるため，発話開始から合成終了まで大きな遅延が生じる．

従来の音声翻訳システムは「文」を処理の単位とする [1]．このため，翻訳は文が終了するまで開始されず，長い文であれば翻訳処理にも多くの時間を要する．これに比べて，人間の同時通訳者は文をより細かい単位に分割することで遅延を最低限に抑える [2]．

本研究では，遅延の問題を解決するために，文が終了する前に翻訳を開始する手法を提案する．具体的には，統計的フレーズベース翻訳 [3] で用いられるフレーズテーブルに着目し，その中の情報を用いて入力を文より短い単位に分割する手法を提案する．翻訳の単位を短くすることでスピードの大幅な向上を実現する際に，単位が短すぎると，正確な訳出を行うための文脈情報が失われる．つまり，スピードと精度は反比例の関係にある．本提案では，翻訳に最も適切な単位を選択するために，翻訳対象となる言語対の並べ替えやすさを考慮したパラメータを導入し，スピードと精度のトレードオフの調整を可能とする．

実験的評価において，日英，英日，仏英などの言語対における実験で提案法の有効性を検証する．その結果，提案法により音声翻訳の遅延が大幅に改善され，導入したパラメータによりスピードと精度のバランスを調整できることが分かった．また，言語的情報を利用せず，ポーズのみで文を分割した手法との比較を行い，高い同時性が求められる状況において，提案法が精度を維持したまま 20%のスピード向上を実現できることも明らかになった．

2 関連研究

音声翻訳の同時性に着目した先行研究は多くないが，関連する報告はいくつかある．まず，漸次的係り受け解析を行い，係り受け木上のルールを用いて音声翻訳の同時性を向上させる手法は Ryu らにより提案されている [4]．この手法は言語学者の知識を取り入れており，日本語の倒置という現象を利用することで英日翻訳において遅延の問題をうまく取り扱っている．その一方，対象言語に精通した言語学者と精度の高い漸次的係り受け解析器が必要となり，新たな言語対への適応は容易ではない．また，音声認識の無音区間に着目した手法も Bangalore らにより提案されている [5]．この手法は言語対に関わらず容易に適応可能であるが，韻律情報しか用いておらず，翻訳対象となっている言語対の言語的特徴を利用しない．

本研究では，この 2 つの手法の間をとり，翻訳対象の言語対の言語的特徴を考慮しながらも，対訳データのみから学習可能な訳出タイミング決定法を提案する．こうすることにより，手軽なシステム構築を保ちながら，より正確に適切な訳出タイミングを決定できると考えられる．

表 1　フレーズテーブルと右確率

原言語	目的言語	右確率
私	I	0.8
私 は	I	0.9
男	man	0.2
男 です	am a man	0.6
何	what	0.9
何 時	what time	0.7
何 時 から	from what time	0.5
プレー	play	0.2
でき	can	0.7
でき ますか	?	0.95

3 提案手法

提案手法は，広く用いられる統計的フレーズベース機械翻訳の枠組みに基づく [3]．フレーズベース翻訳では，複数の単語からなるフレーズを翻訳の基本的な単位とし，原言語フレーズと目的言語フレーズの対応を記述するフレーズテーブルを用いて翻訳を行う．その例を表 1 に示す．このテーブルは対訳データに対して単語アライメントとフレーズ抽出を行うことで自動的に構築可能である．3 列目の右確率について，3.2 節で詳しく述べる．以降の節では，既存のフレーズベース翻訳システムが与えられた際，そのシステムのフレーズテーブルを用いて訳出のタイミングを決定する方法について述べる．

3.1 フレーズテーブルを用いた翻訳単位決定

フレーズテーブルは対訳コーパスのみから学習可能であるため，対訳データさえあればどの言語対に

*A method for deciding translation timing in speech translation considering reordering between languages.
by Graham NEUBIG, Tomoki FUJITA, Sakriani SAKTI, Tomoki TODA, and Satoshi NAKAMURA
(Nara Institute of Science and Technology)

表 2　単位決定と翻訳の結果

単位	結果
私 は	I
男 です	am a man

対しても構築可能である．このため，本研究では多言語に対応し，かつ言語対の言語的特徴を捉える情報源としてフレーズテーブルに着目する．

まず，原言語文 $F = f_1 \dots f_J$，入力済みでまだ翻訳されていない単語のキャッシュ $G = g_1 \dots g_K$ を定義する．音声認識器の入力を漸次的に処理する場合を想定し，F が 1 単語ずつ入力されるとする．各単語 f_j が入力された時点で，f_j とその前の未処理の単語を全て翻訳するか，次の単語が入力されるのを待つかを決定する．この決定を行うために，まず f_j を G に追加する．その次，G に含まれている単語をフレーズテーブルに含まれるフレーズの原言語側と照合する．G と同等の単語列からなる原言語側を持つフレーズがフレーズテーブルに含まれている場合，G を翻訳せずに，次の入力を待つ．G がフレーズテーブルに一致しなければ，G の最後の単語以外 $(g_1 \dots g_{K-1})$ に対して翻訳処理を行い，G に最後の単語のみに置き換える $(G \leftarrow g_K)$．

この手続きを行うことで，並べ替えを許さないフレーズベース翻訳を行った時と類似した結果が得られる．フレーズが一致する限り G に単語を追加していくため，未翻訳の単語列から始まる最長のフレーズが翻訳の単位として選ばれる．

具体例として，原言語文を「私 は 男 です」が与えられ，表 1 のフレーズテーブルに基づいて翻訳の単位を決定する場合を考える．まず，f_1 が G に追加され，$G = \{$“私”$\}$ となる．「私」がフレーズテーブルに存在するため，すぐに翻訳せずに，次の入力を待つ．その次，f_2 が入力され $G = \{$“私”, “は”$\}$ となり，これもフレーズテーブルに存在するためすぐに翻訳を開始しない．また，f_3 を追加し，$G = \{$“私”, “は”, “男”$\}$ が得られるが，今回はフレーズテーブルに存在しないため，翻訳を行う．$g_1 \dots g_{K-1}$ に当たる $\{$“私”, “は”$\}$ を翻訳エンジンへ送り，G を g_K に当たる「男」に設定する．最終的な翻訳結果を表 2 に示す．

3.2　右確率を用いた訳出タイミングの調整

前節で述べた手続きで，並べ替えを許さない翻訳を行うことが可能となるが，高精度の翻訳を行うために，並べ替えを許す必要がしばしばある．並べ替えを行わないと正確な翻訳が得られない例を表 3 に示す．この例では,「プレー でき ます　か」を「can we play」と翻訳したいが，この文全体がフレーズテーブルに存在せず，言語間の語順が異なるため「プレー」と「でき ます か」を個別に翻訳した際，正しい語順を実現することができない．従って，このフレーズの

表 3　フレーズテーブルのみで得られた翻訳単位で翻訳が失敗する例

単位	結果
何 時 から	from what time
プレー	play
でき ます か	?

並べ替えを許すような翻訳単位を利用すれば，単純にフレーズの境界で翻訳の単位を確定した場合より高い翻訳精度が実現できると考えられる．

この問題をシンプルかつ効果的に解決するために，各フレーズの並べ替えの起きにくさを表す「右確率」に着目した訳出タイミング決定法を提案する．フレーズベース機械翻訳システムの並べ替えモデル [6] における右確率は，現在のフレーズが原言語単語 f_j と目的言語単語 e_i で終わる場合，f_{j+1} から始まるフレーズが目的言語文においても e_{i+1} 以降に始まる確率である．つまり，両言語でフレーズの並ぶ順番が同じである確率を表すため，語順に影響しない訳出のタイミングを発見するのに適している確率である．

この確率に基づいて訳出のタイミングを決定するために，まず 3.1 節の手法に基づいてフレーズを仮確定する．次に，仮確定されたフレーズの右確率を，閾値 t と比較する．右確率が t を超えた場合，未処理の単語を全て翻訳し，t に満たなかった場合は翻訳せずに保持しておく．例えば，$t = 0.5$ と設定した際,「プレー」の右確率が 0.5 に満たないため，すぐに翻訳せずに,「でき ます か」が全て入力されてから「プレー」も併せて翻訳する．この枠組みでは，$t = 1.0$ の場合は通常の 1 文ごとの翻訳システムとなり，$t = 0.0$ の場合は 3.1 節のフレーズごとの翻訳となる．

3.3　言語モデル適応

上記の 2 手法では，訳出タイミングの決定に着目した．しかし，言語モデル (LM) の学習法にも注意する必要もある．文全体で学習された言語モデルを文より短い単位の翻訳に用いると，不要な句読点が大量に生成されるなどの悪影響が見られる．

この問題を解決するために，本研究では文より短い単位の翻訳に適した言語モデルの学習法を提案する．具体的には，3.2 節の翻訳単位決定法を原言語ではなく，目的言語の文に対して施し，翻訳単位に分割された文で言語モデルを学習する．予備実験では，3.1 節で提案したフレーズごとの翻訳を行った際，言語モデルを適応しなかったベースラインで 34.04，言語モデルを適応したシステムで 38.46 の BLEU スコアが得られたことから，この手法がシンプルでありながら，効果的に翻訳単位と言語モデルの差を補えることが分かる．

表 4　データの文と単語数

ja-en	文	単語 (ja)	単語 (en)
学習	162k	1.38M	1.19M
テスト	1,018	8,782	7,496
テスト (11+)	217	3,092	2,234
fr-en	**文**	**単語 (fr)**	**単語 (en)**
学習	44k	1.02M	880k
テスト	960	26,753	22,717

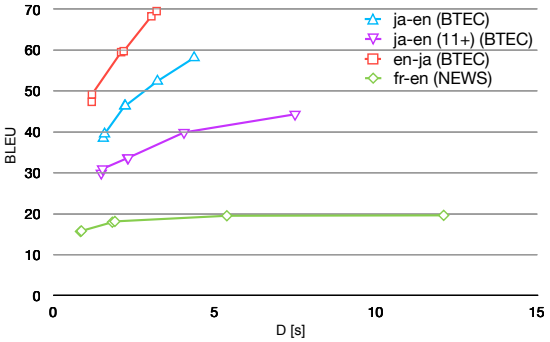

図 1　人手書き起こしに対する翻訳の精度と遅延．各線上の点は左から，右確率が 0.0, 0.2, 0.4, 0.6, 0.8, 1.0 であった場合の結果である．

4　実験的評価

4.1　実験設定

本研究の最終目的は音声翻訳のスピード向上であるが，訳出のタイミングが翻訳のスピードと精度に及ぼす影響に着目するために，まず人手や音声認識により書き起こされたテキストに対して実験を行う．右確率の閾値 t を，0.0, 0.2, 0.4, 0.8, 1.0 で変動させ，1.0 は文ごとの翻訳を行うベースラインに値する．言語モデルの適応は翻訳時と同じ閾値で行う．言語対として日英 (ja-en) を主とするが，翻訳の方向の影響を調べるために英日 (en-ja)，語順の差が少ない言語対における効果を調べるために仏英 (fr-en) 翻訳の実験も行う．音声認識には Julius[7]，機械翻訳には Moses[8]，日本語の形態素解析には Mecab[9] を利用する．

表 4 に実験データの諸元を示す．日英・英日翻訳において Basic Travel Expression Corpus (BTEC, [10]) を利用し，仏英において NEWS[11] データを利用する．BTEC の文が比較的短いため，11 単語以上からなるテストデータにおいての性能評価も行う．

翻訳精度の評価に BLEU[12] を利用し，日英では参照文 12 文，仏英では参照文 1 文を利用する．また，0-5 許容性評価 [13] を人手で行う．翻訳の遅延を $D = A + T$ と定義し，A は認識結果を待つことによる遅延であり，T は機械翻訳の処理時間である．

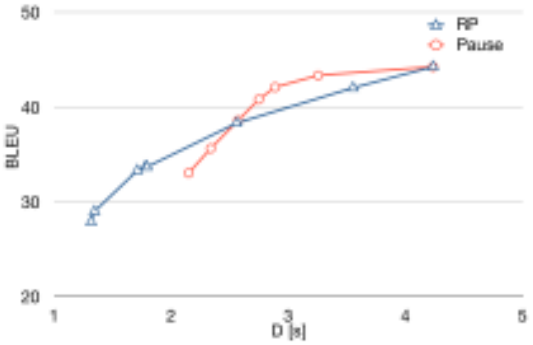

図 2　音声認識結果に対する，ポーズ (Pause) と右確率 (RP) をそれぞれ用いて訳出のタイミングを決定した際の翻訳精度と遅延

4.2　翻訳の精度と遅延

まず，人手により書き起こされた文章に対する実験結果を図 1 に示す．この結果から分かるとおり，全てのデータに置いて，閾値を減らすと遅延の大幅な減少が見られるが，翻訳の精度も若干減少する．

また，BTEC の日英翻訳において，全ての文と 11 単語以上からなる文の結果を比較すると，スピードと精度の曲線はどちらのデータも同様の形状を示した．しかし，長い文の方が短い文より，遅延 D の減少が顕著であり，提案手法は長い文に対して特に効果的であることが分かる．

次に言語対による差について考察する．日英と英日翻訳に関しては，全体的に少し精度が異なるが，スピードと精度のトレードオフはほぼ同様の傾向を示し，翻訳の方向による影響は大きくないと考えられる．しかし，日英と仏英翻訳を比較すると，仏英において，右確率の閾値を 1.0 から 0.8 へ変更した際，翻訳精度を保ったまま遅延が 12.1 秒から 5.40 秒へ減少したことが分かる．閾値を更に減らしても日英や英日翻訳に比べて精度の減少が小さいことから，提案法は語順の近い言語で特に有効であることが分かる．

4.3　認識結果に対する実験

図 2 に，音声認識の結果を出力として，提案法とポーズ情報 [5] を用いて翻訳の単位を決定した際の遅延と翻訳精度を示す．右確率の閾値として 0.0, 0.2, 0.4, 0.6, 0.7, 0.8, 0.9, 1.0 を利用し，ポーズ情報の閾値として 1, 2, 3, 4, 5, 10 フレームの無音区間を利用する．ポーズ情報を用いた手法に対しても，3.3 節に従って言語モデルの適応を行い，右確率の閾値を予備実験に基づいて 0.8 と設定する．

まず，認識結果と人手による書き起こしの結果を図 1 と図 2 で比較すると，誤りを含まない書き起こしを用いた方が翻訳精度が高いという自然な結果となった．また，スピードと精度のトレードオフは両方の設定において同様の傾向を示した．次に，図 2 においてポーズ情報を用いた翻訳単位の決定と比較すると，

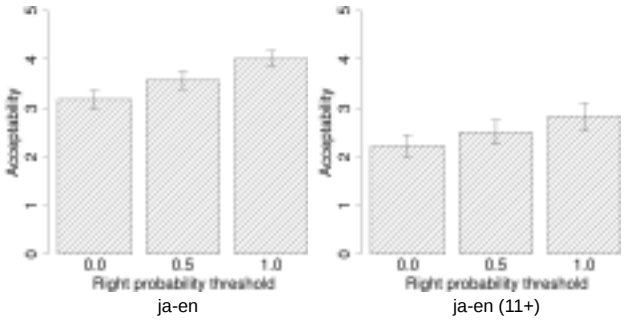

図 3　人手による許容性の評価

表 5　「サーフィン に いい 場所 を 教え て 下さい .」を翻訳した際の結果

RP	結果	許容性
0.0	for surfing / tell me a good place /	5
1.0	please tell me a good surfing place ? /	5

提案手法は遅延が少ない領域において比較的高い精度を実現しており，より遅延の多い領域においてポーズに基づく翻訳単位の決定がより高い精度を実現している．また，閾値を低く設定することで，ポーズ情報で切れない箇所においても訳出することが可能であることも分かる．無音区間のフレーム長を 1 に設定した際 2.14 秒の遅延で 33.0 の BLEU スコアを実現しているのに対して，右確率閾値 $t = 0.4$ に基づく単位決定は 1.71 秒の遅延と 33.3 の BLEU を実現していることから，同等の精度を保ちながら遅延を約 20.0%減らせたことが分かる．

4.4　人手評価

右確率の閾値を 0.0, 0.5, 1.0 に設定したシステム出力に対して人手による許容性評価を行った結果について述べる．BTEC 日英の 300 文，BTEC(11+) 日英の 160 文を計 5 人の評価者に点付けしてもらった．図 3 にその結果を示す．

この結果から，閾値が小さくなるのに連れて人手評価も減少することが分かるが，減少の幅は BLEU と比較してさほど顕著ではない．この理由として考えられるのは，語順が多少不自然であっても意味が問題なく通じる翻訳結果が見受けられるからである．その一例を表 5 に示す．この結果から，語順が異なっても自然性を判断できる自動評価尺度が必要であると言える．

5　まとめ

本研究では，言語的特徴を取り入れながら，容易に多くの言語対に適応できる訳出タイミング決定法を提案した．実験的評価において，提案法は音声翻訳における遅延を減少させることができることが分かった．また，右確率に対する閾値を調整することで，翻訳のスピードと精度のバランスを調整できることが分かった．今後の課題として，模擬実験から実際の同時通訳への展開，同時通訳のための自動評価尺度，文法や韻律などの情報との組み合わせなどが考えられる．

謝辞　本研究の一部は，JSPS 科研費 24240032 の助成を受け実施したものである．

参考文献

[1] E. Matusov, A. Mauser, and H. Ney, “Automatic sentence segmentation and punctuation prediction for spoken language translation,” in *Proceedings of IWSLT*, 2006, pp. 158–165.

[2] F. Goldman-Eisler, “Segmentation of input in simultaneous translation,” *Journal of Psycholinguistic Research*, vol. 1, no. 2, pp. 127–140, 1972.

[3] P. Koehn, F. Och, and D. Marcu, “Statistical phrase-based translation,” in *Proceedings of NAACL-HLT*, 2003, pp. 48–54.

[4] K. Ryu, A. Mizuno, S. Matsubara, and Y. Inagaki, “Incremental Japanese spoken language generation in simultaneous machine interpretation,” in *Proceedings of Asian Symposium on Natural Language Processing to Overcome Language Barriers in Hainan Island China*, 2004.

[5] S. Bangalore, V. K. R. Sridhar, P. K. L. Golipour, and A. Jimenez, “Real-time incremental speech-to-speech translation of dialogs,” in *Proceedings of NAACL*, 2012.

[6] P. Koehn, A. Axelrod, A. Mayne, C. Callison-Burch, M. Osborne, and D. Talbot, “Edinburgh system description for the 2005 IWSLT speech translation evaluation,” in *Proceedings of IWSLT*, 2005.

[7] A. Lee, T. Kawahara, and K. Shikano, “Julius—an open source real-time large vocabulary recognition engine,” 2001.

[8] P. Koehn, H. Hoang, A. Birch, C. Callison-Burch, M. Federico, N. Bertoldi, B. Cowan, W. Shen, C. Moran, R. Zens *et al.*, “Moses: Open source toolkit for statistical machine translation,” in *Proceedings of ACL*, 2007, pp. 177–180.

[9] T. Kudo, K. Yamamoto, and Y. Matsumoto, “Applying conditional random fields to japanese morphological analysis,” in *Proceedings of EMNLP*, 2004, pp. 230–237.

[10] T. Takezawa, E. Sumita, F. Sugaya, H. Yamamoto, and S. Yamamoto, “Toward a broad-coverage bilingual corpus for speech translation of travel conversations in the real world,” in *Proceedings of LREC*, 2002, pp. 147–152.

[11] J. Civera and A. Juan, “Domain adaptation in statistical machine translation with mixture modelling,” in *Proceedings of WMT*, 2007, pp. 177–180.

[12] K. Papineni, S. Roukos, T. Ward, and W. Zhu, “BLEU: a method for automatic evaluation of machine translation,” in *Proceedings of the 40th annual meeting on association for computational linguistics*, 2002, pp. 311–318.

[13] I. Goto, B. Lu, K. Chow, E. Sumita, and B. Tsou, “Overview of the patent machine translation task at the NTCIR-9 workshop,” in *Proceedings of NTCIR*, 2011, pp. 559–578.