Vol.2012-NL-209 No.13
2012/11/23

同時性を考慮した音声翻訳システムの検討

藤田 朋希　Graham Neubig　Sakriani Sakti　戸田 智基　中村 哲

概要：従来の音声翻訳システムでは発話から合成までに生じる遅延が大きく，ニュースや講演の内容をリアルタイムで理解することが困難である．そこで，本研究では音声翻訳システムの遅延の原因として考えられる機械翻訳モジュールの開始時間および処理時間の改善を図る．具体的にはフレーズベース機械翻訳のフレーズテーブルを用いて，翻訳単位を短くする方法を提案し，翻訳の精度と同時性への影響を調査する．実験的評価により，翻訳開始および翻訳処理の時間が減少させることができることを確認した．

キーワード：音声翻訳システム，同時通訳，遅延

Towards a Speech Translation System Considering Simultaneity

Abstract: In conventional speech translation systems, it is difficult to understand the content of news and lectures in real time, because of the large delay between the begining of the utterance and the end of synthesis. In this research, we try to improve translation start time and translation processing time of the machine translation module. Concretely, using the phrase table of phrase-base translation, we propose a method of shortening the translation unit and examine the effect on translation speed and accuracy. Through experimental evaluation, we confirmed the fact that translation start time and translation processing time decrease.

Keywords: speech translation system，simultaneous translation，delay

1. はじめに

世界各国において，留学，出張，海外旅行など，多言語でコミュニケーションをする場面が増加している．また，国内においても海外のニュースや講演を見て情報を収集することで，知識を深め，日常生活や業務に活かすことができる．このような場面で内容をスムーズに理解するためには，発話に含まれている情報を正確に把握するだけでなく，話の流れについていけるようにリアルタイムで情報を得る必要がある．これを音声翻訳システムで支援するためには，同時通訳者のようなリアルタイムかつ高精度な翻訳を行う技術が必要不可欠となる．

音声翻訳システムの翻訳精度改善については，精力的な研究がなされており，多大な研究成果が得られている一方で，同時通訳者のようなリアルタイムでの翻訳結果出力については，未だ数多くの問題が残されている．従来の音声翻訳システムは，発話内容からテキストへの書き起こしを行う音声認識 (ASR) モジュール，書き起こされたテキストから目的言語への翻訳を行う機械翻訳 (MT) モジュール，翻訳結果を再生する音声合成 (TTS) モジュールから構成される．通常，ASR モジュールが終了するまで MT モジュールを開始せず，MT モジュールが終了するまで TTS モジュールを開始しない．このため，発話から合成までに生じる遅延が大きく，ニュースや講演の内容をリアルタイムで理解することが困難である．

音声翻訳システムの遅延の主な原因として，MT モジュールの開始時間および処理時間が考えられる．通常の機械翻訳単位として文が用いられるため，1 文の発話が終了するまで MT モジュールを開始することができず，待ち時間が長くなる．また，翻訳単位が長くなればなるほど，処理時間は増加する．この問題に対して，従来の音声翻訳システムよりも翻訳単位を短くとることができれば，MT モジュールの開始時間を速くし，処理時間を短縮できると考えられる．

本研究では，音声翻訳システムの遅延の原因として考えられる MT モジュールの開始時間および処理時間の改善

[1] 奈良先端科学技術大学院大学
Nara Institute of Science and Technology

を図る．具体的には，フレーズベース機械翻訳のフレーズテーブルを用いることで，フレーズベース機械翻訳が持つ汎用的な言語非依存性を保持しつつ，翻訳単位を短くする方法を提案する．まず，原言語のフレーズパターンを利用することで，翻訳単位を決定する手法を提案する．さらに，翻訳単位が短くなりすぎて翻訳精度が劣化することを防ぐために，両言語の語順が同等である確率 (right 確率) を利用して、翻訳単位の長さの調整を行う手法も提案する．提案法における翻訳の精度と同時性への影響を調査するために，旅行対話のデータを利用して，翻訳単位の長さと翻訳精度の関係を検証する．フレーズテーブルと right 確率を用いた翻訳単位の決定方法により，翻訳開始および翻訳処理の時間を減少させることができることを示す．

2. 関連研究

音声翻訳システムの同時性を向上させる研究として，ASR モジュールで検出される無音区間を利用して翻訳単位を決定する手法 [1] がある．この手法では，音響特徴量のみを考慮しているため，同時通訳者のように文脈によって適切な翻訳単位を選択することが困難である．

この他に，構文ルールを適用した手法 [9] もある．この手法では各言語間の構文ルールを人手で構築しているため汎用性に欠け，多言語の翻訳に適用することが困難である．

3. 提案法

本説では，同時性の高い，かつ多言語に対応した音声翻訳システムを実現するために，対訳コーパスから得られたフレーズテーブルを利用した翻訳単位の決定方法を提案する．2 節で述べた従来手法に比べて，提案法では同時通訳者と同じように発話内容を考慮した翻訳単位で翻訳を行うことができ，汎用性の高い手法を用いているため多言語に対応できる．

表 1　フレーズテーブルの例

原言語フレーズ	目的言語フレーズ	right 確率
私	I	0.8
私 は	I	0.3
男	man	0.9
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	0.4
プレー	play	0.3
でき	can	0.7
でき ますか	?	0.6

3.1 フレーズベース機械翻訳

機械翻訳はルールベース機械翻訳と統計的機械翻訳に分類される．現在，統計的機械翻訳が世界的に研究の主流であり，フレーズベース機械翻訳 [6] もその 1 種である．

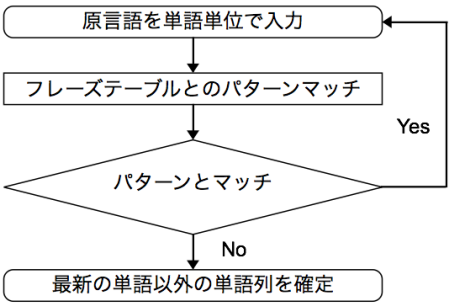

図 1　フレーズテーブルによる翻訳単位の決定

フレーズとは原言語および目的言語を翻訳しやすい単語列に区切った単位のことであり，単語単位で翻訳をするよりも翻訳精度が良いとされている．表 1 にフレーズテーブルの例を示す．フレーズテーブルは原言語と目的言語の対訳コーパスに対して，単語アライメント [7] を行うことで，自動的に抽出される．[6] なお，表中の 3 列目の right 確率は第 3.3 節で詳しく説明する．

そして，このように算出されたフレーズ対にスコアを付与する翻訳モデル，目的言語の文にスコアを付与する言語モデル，フレーズ間の並びにスコアを付与する並べ替えのモデルの 3 つを組み合わせて翻訳候補のスコアを算出する．

3.2 フレーズテーブルを利用した翻訳単位の決定

フレーズテーブルは対訳コーパスさえあればどの言語対でも取得することができるため，多言語に容易に対応できる．このフレーズベース機械翻訳の汎用性を保つために，まず，翻訳単位の決定にフレーズテーブルの原言語のパターンを利用する方法を提案する．

図 1 はフレーズテーブルを利用した翻訳単位の決定の処理を示す．原言語文の入力を $\mathrm{F} = f_1 \ldots f_j$，現在，マッチの対象となっている単語列を $\mathrm{G} = g_1 \ldots g_k$ とする．まず，F を単語単位で入力していき，G に追加していく．G に入っている単語がフレーズテーブルの原言語のパターンとマッチする場合は，G をそのまま保持する．G がフレーズテーブルにマッチしなくなった場合，$g_1 \ldots g_{k-1}$ を翻訳単位として確定し，$\mathrm{G} \leftarrow g_k$ とする．これにより，フレーズテーブルにある原言語の最長フレーズを翻訳単位として選択する．

この動作の具体例として，原言語の入力を式 (1) に示す．

$$\mathrm{F} = \text{“私” “は” “男” “です”} \tag{1}$$

この場合，f_1 = “私”，f_2 = “は”，f_3 = “男”，f_4 = “です” である．まず，f_1 を G に追加する．すると，G = “私” となり，これはフレーズテーブルの原言語側のフレーズとして存在するため G = “私” として保持する．今度は，f_2 を追加し，G = “私 は” となり，これもフレーズテーブルに存在するため G をこのまま保持する．今度は，f_3 を追加

する．すると，G =“私 は 男”となり，これはフレーズテーブル中の原言語側のフレーズパターンとマッチしないため，$g_1 \ldots g_{k-1}$ に当たる“私 は”を翻訳単位として確定し，G を g_k に当たる“男”に置き換える．この作業を原言語の入力終了まで繰り返す．

最終的に，F は表 2 のような翻訳単位と翻訳結果となる．

表 2　フレーズテーブルのみ

翻訳単位	翻訳結果
私 は	I
男 です	am a man

3.3　right 確率を利用した翻訳単位の長さ調整

前節で紹介したフレーズテーブルの原言語パターンを利用した翻訳単位の決定のみでは，翻訳単位が短すぎることもあり翻訳精度の高い翻訳結果を得ることができない．

F =“何”“時”“から”“プレー”“でき”“ますか” (2)

例えば，式 (2) の入力に対し，フレーズテーブルのみを利用した翻訳単位の決定による翻訳では表 3 のようになる．

表 3　フレーズテーブルのみ

翻訳単位	翻訳結果
何 時 から	from what time
プレー	play
でき ますか	?

この問題の原因として，“プレー”と“でき ますか”を 1 句の単位で翻訳すると“play”と“can”のように得たい翻訳結果と逆順になってしまうことが考えられる．そのため，逆順になりそうなフレーズを 1 つの単位として翻訳することが出来れば，翻訳精度は向上すると期待される．

本報告では，逆順になりやすい句を判別する手法として，right 確率を用いる手法を提案する．right 確率とは原言語を目的言語に翻訳する際に両言語の語順が同等である確率であり，フレーズベース機械翻訳の並べ替えモデルに利用される．図 2 に示すように，Monotone と Discontinuous-right の確率の合計であり，この確率が高いほど翻訳の際に並べ替えが不要となる確率が高い．

right 確率を利用した翻訳単位の調整の処理を図 3 に示す．まず，図 1 の処理を行って，翻訳単位を暫定的に確定する．その後，その翻訳単位の right 確率と閾値を比較して，閾値未満の場合は次の翻訳単位と結合して，閾値以上の場合は翻訳単位を決定する．例えば，閾値を 0.5 に設定すると，表 4 のように right 確率が閾値を下回る“プレー”を翻訳単位とせず，次の“でき ますか”と合わせて翻訳単位を長く取得する．これにより，自然な英訳が得られる．この枠組みにおいて，閾値を 1.0 に設定すると通常の文単位での翻訳処理と等価となり，閾値を 0.0 に設定すると 3.2 節の方法と等価となる．

表 4　フレーズテーブルと right 確率

翻訳単位	翻訳結果
何 時 から	from what time
プレー でき ますか	can we play ?

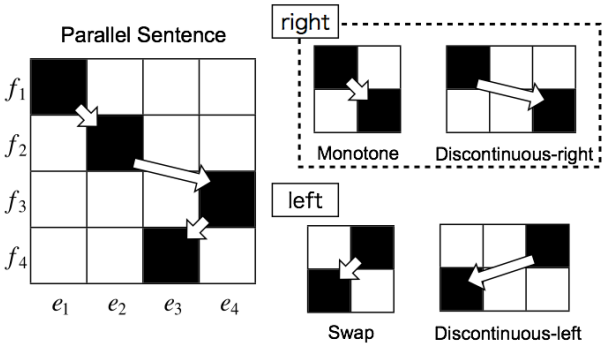

図 2　right 確率

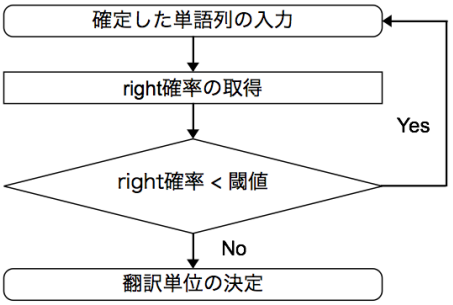

図 3　翻訳単位の調整

4.　実験

翻訳単位の長さ変えた場合の翻訳速度と翻訳精度の関係について調べるため実験を行う．さらに，閾値別の翻訳結果文を見てスコアを付与してもらう主観評価実験も行う．

4.1　実験設定

本論文では，MT モジュールの改善による，翻訳速度と翻訳精度の関係性に焦点を絞るために，ASR モジュールの代わりに書き起こしたテキストデータを使用し，TTS モジュールににについては考慮しない．ここで，1 文単位の翻訳を従来システムとし，1 句単位に区切ったテキストの翻訳を提案システムとする．

表 5 は実験データを示しており，タスクは BTEC[10] コーパスの旅行対話文である．また，テストデータの 1018 文は 1 文あたり 8.6 形態素しかあらず，比較的短い文からなっている．長い文に対する本手法の有効性を確認するために，テストデータから 11 形態素以上のみの文を使用しての実験も行う．このテストデータは表 5 の中でテスト 11+ と記載している．

また，機械翻訳エンジンとして Moses[5] を用いる．設定はデフォルトと語彙化された並べ替えモデル [4] を使用する．なお，原言語から目的言語に翻訳する際の並べ替え

表 5　実験データ

	文	形態素	形態素/文
学習データ	162,318	1,380,817	8.5
テスト	1018	8782	8.6
テスト 11+	217	3092	14.2

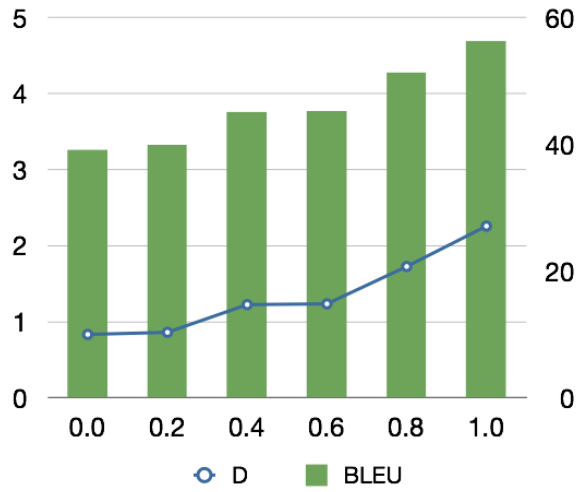

図 4　実験結果

の制限 (Distortion Limit) を予備実験により，従来システムで速度と精度のバランスを取った 12 に設定する．また，言語モデルは閾値に合わせて変更した．原言語文の翻訳単位の決定の際に設定した閾値と同じ値を使用して，目的言語文の学習単位を変更する．

4.2　翻訳の評価尺度

翻訳精度の評価尺度として BLEU[8] を使用する．同時通訳における同時性に対する確立した評価尺度がないため，遅延 D を翻訳速度の尺度として提案する．遅延 D は発話開始から翻訳終了までの平均時間と定義する．D の算出には式 (3) を用いる．

$$D = F \times U + T \tag{3}$$

F は翻訳する形態素数の平均を表し，テストデータに対する形態素数を翻訳回数で除算して算出する．U は 1 形態素あたりにかかる発話時間を表す．[3] の調査において，1 形態素当たりの平均発話時間が約 0.21 であったため，この値を U として利用する．T は機械翻訳システムの処理時間の平均を表し，実行時間を翻訳回数で除算して算出する．

4.3　実験結果

実験では，まず，テストに対して翻訳単位の長さにより，翻訳速度と翻訳精度の関係性を調査する．図 4 はテストのデータに対しての実験結果を示しており，閾値 0.0 の場合の D は約 0.8，閾値 1.0 の場合の D は約 2.25 で 3 倍違う．これに対して，BLEU は閾値 0.0 の場合は 39.13，閾値 1.0 の場合は 56.32 である．このことより，翻訳速度と翻訳精度の関係はトレードオフであることが言える．

また，長い文に対しても翻訳速度と翻訳精度の関係性を調査するため，実験を行う．図 5 はテスト 11+のデータに対しての実験結果を示しており，テストのデータと同様に，翻訳速度と翻訳精度の関係はトレードオフである．テストと比べ，BLEU が全体的に 10 点から 12 点ほど減少している．また，閾値を 0.8，1.0 にした際に遅延が極端に上がり，平均形態素数の多い文では 1 文単位の翻訳に時間がかかるが閾値を小さくすることで翻訳速度を大幅に改善できる．そのため，提案法が長い文に対して特に有効であることが分かる．

最後に，主観評価結果を表示する．この実験では入力文と閾値 0.0，0.5，1.0 にした際の翻訳結果文を見てもらい Acceptability 評価基準 [2] に沿って 0 から 5 のスコアを付与してもらう．なお，被験者は 5 人で各閾値の翻訳結果を

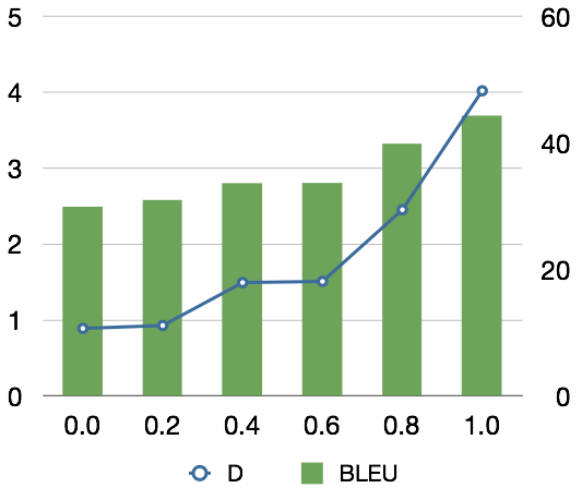

図 5　実験結果 (11 形態素以上)

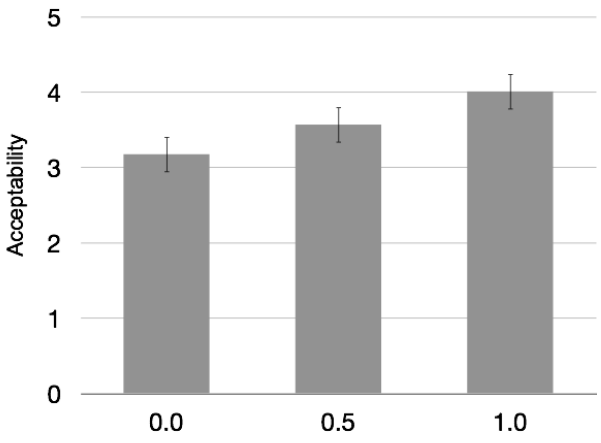

図 6　主観評価結果

300 文ずつ評価する．結果は図 6 に示す．この結果から，平均的に閾値が高いほど主観評価の結果が良い．

表 6，7，8 は閾値により翻訳結果および主観評価に差が出ていない例，翻訳結果および主観評価に差が出た例，翻訳結果に差が出たが，主観評価に差が出ていない例を示している．

表 6 「コーラ を ください」の主観評価例

閾値	翻訳結果	Acceptability
0.0	coke , please . /	5
0.5	coke , please . /	5
1.0	coke , please . /	5

表 7 「もっと 手頃 な ホテル は あり ませ ん か」の主観評価例

閾値	翻訳結果	Acceptability
0.0	more / reasonable / is there a hotel ? /	3
0.5	more / reasonable ? / is there a hotel ? /	4
1.0	do you have a more reasonable hotel ? /	5

表 8 「サーフィン に いい 場所 を 教え て ください」の主観評価例

閾値	翻訳結果	Acceptability
0.0	for surfing / tell me a good place /	5
0.5	for surfing tell me a good place /	5
1.0	please tell me a good surfing place ? /	5

まず，表 6 について，短い形態素の文であるため，翻訳結果に差が出ず，Acceptability も変わらない，対して，表 7 は長い形態素の文であるため，翻訳結果に差があり，Acceptability にも差が生じる．しかし，表 8 は翻訳結果に差があるが，Acceptability に差がない．これは「いい 場所 を 教え て ください」を 1 句として翻訳できているからである．このように長い句をフレーズテーブル定形句として学習されている場合は訳出が悪化しない．

5. おわりに

本研究では，同時性が高く，汎用性のある音声翻訳システムのための翻訳単位の決定方法を提案した．同時性については従来システムの 1 文単位の翻訳ではなく，1 句単位で翻訳する方法を提案することで，翻訳の開始時間および処理時間の減少を図った．汎用性についてはどの言語対でも自動的に計算可能なフレーズテーブルと right 確率のみを利用した．その結果，従来システムに比べ，翻訳時間を減少することが実験結果より確認できた．ASR モジュール，TTS モジュールを考慮してシステムの実現を図るのが今後の課題である．

参考文献

[1] Srinivas Bangalore, Vivek Kumar Rangarajan Sridhar, Prakash Kolan Ladan Golipour, and Aura Jimenez. Real-time incremental speech-to-speech translation of dialogs. In *Proceedings of NAACL*, 2012.

[2] I. Goto, B. Lu, K.P. Chow, E. Sumita, and B.K. Tsou. Overview of the patent machine translation task at the ntcir-9 workshop. In *Proceedings of NTCIR*, volume 9, pages 559–578, 2011.

[3] Shigeki Matubara Haibei Yu, Koichiro Ryu. A corpus-based analysis of simultaneous interpreterfs utterance speed. 2008.

[4] P. Koehn, A. Axelrod, A.B. Mayne, C. Callison-Burch, M. Osborne, and D. Talbot. Edinburgh system description for the 2005 IWSLT speech translation evaluation. In *International Workshop on Spoken Language Translation*, 2005.

[5] 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 *Annual meeting-association for computational linguistics*, volume 45, page 2, 2007.

[6] P. Koehn, F.J. Och, and D. Marcu. Statistical phrase-based translation. In *Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology-Volume 1*, pages 48–54. Association for Computational Linguistics, 2003.

[7] F.J. Och and H. Ney. A systematic comparison of various statistical alignment models. *Computational linguistics*, 29(1):19–51, 2003.

[8] K. Papineni, S. Roukos, T. Ward, and W.J. Zhu. BLEU: a method for automatic evaluation of machine translation. In *Proceedings of the 40th annual meeting on association for computational linguistics*, pages 311–318. Association for Computational Linguistics, 2002.

[9] Koichiro Ryu, Atsushi Mizuno, Shigeki Matsubara, and Yasuyoshi 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.

[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*, volume 1, pages 147–152, 2002.