Vol.2016-NL-226 No.3
Vol.2016-SLP-111 No.3
2016/5/16

対象ドメインの高頻出句に対する人手対訳追加による講義音声翻訳の検討

後藤 統興[1,a)]　山本 一公[1,b)]　中川 聖一[1,c)]

概要：自動音声認識と統計的機械翻訳を組み合わせた, 英語音声を日本語文書へ翻訳する音声翻訳システムについて, 統計的機械翻訳の改善を行った. 単語列を句とし, 翻訳対象ドメインの英語書き起こしから頻出する句を抽出した. 抽出の手法として n-gram 単位で抽出する手法と構文解析を用いて抽出する手法について比較を行った. 抽出した句のうちテストデータに出 現するものに対して人手によって翻訳を行うことで, 対象ドメインに頻出する英日の頻出フレーズ対を作成した. 作成した英日フレーズ対を翻 訳モデルへ学習させることで対象ドメインへの適応を行った. 学習の手法として学習コーパスとして追加する手法とフレーズテーブルに追加す る手法を比較した. 性能の比較を行うために評価基準 BLEU を用いた. 結果として書き起こしの英語文を日本語へ翻訳した結果, 構文解析によって抽出し, 作成した英日フ レーズ対をフレーズテーブルに追加する手法において BLEU10.5 を得ることができ, ベースラインのモデルの BLEU10.2 を 上回った. また, 英語を習熟した評価者三名に評価を依頼し, 本手法の有効性を確認した.

キーワード：音声翻訳, 音声認識, 機械翻訳, 頻出句, フレーズベース翻訳

Investigation of speech translation for Lectures by using manually extracted parallel phrases for frequently occurrences in-domain.

Norioki Goto[1,a)]　Kazumasa Yamamoto[1,b)]　Seiichi nakagawa[1,c)]

1. はじめに

近年, マサチューセッツ工科大学（MIT）をはじめとした有力な教育機関によるウェブサイトを経由した英語の講義映像の無料公開が行われている [1]. 現在は日本でも JOCW[*1] などの取り組みが行われており, より安価に学習を行うことが可能になっている. しかし, 外国で公開されている映像には日本語の音声や書き起こしが存在しないため, 日本人学生のように英語の未習熟な人間には, そういった講義への参加の意欲が失われたり, 学習効率の低下を招いてしまう. これを解決するために, 講義映像に日本語字幕を付与する方法があげられる. しかし高い品質の字幕を得るためには, 専門家による人手翻訳が必要となり, 高いコストがかかってしまう. 対照的に, 英語講義映像に対する自動的な日本語の字幕作成システムは低品質であっても有用であり, 低コストで英語音声を日本語文書へと変換できる. このため, 英語音声から英語文書の書き起こしを行う自動音声認識（Automatic Speech Recognition: ASR）と英語文書から日本語文書への翻訳を行う統計的機械翻訳（Statistical Machine Translation: SMT）を組み合わせることで音声翻訳システムを構築する試みがあり, 本稿でも同様に音声翻訳システムを構築した.

音声翻訳の要素である ASR と SMT を改善することで音声翻訳システム全体の精度が改善されることが明らかであり, 各要素について精度向上のための幾つかの先行研究がある. 音声翻訳の国際会議 IWSLT[2] では, 英語講演であ

[1] 豊橋技術科学大学
Toyohashi Univer, Toyohashi, Aichi 441–8580, Japan
[a)] ngoto@slp.cs.tut.ac.jp
[b)] kyama@slp.cs.tut.ac.jp
[c)] nakagawa@slp.cs.tut.ac.jp
[*1] http://www.jocw.jp/

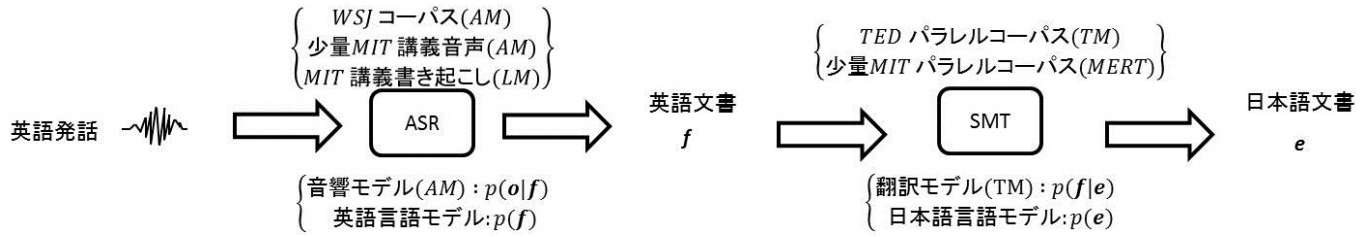

図 1 音声翻訳システム概要図

る TED Talk[*2] に対して, 音声認識と機械翻訳を組み合わせた音声翻訳タスクが存在し, フランス語, ドイツ語, 中国語などの英語への翻訳や, その逆などの話し言葉に対する音声翻訳の精度向上の研究が試みられている.

また, 文献 [3] では, TC-STAR プロジェクトの一環として英語-スペイン語, 英語-ドイツ語の講義音声翻訳システムを作成しており, それぞれ 18.6 の BLEU, 13.2 の BLEU を達成している.

文献 [4] では, 音声翻訳対象と同一ドメインである MIT の英日対訳コーパスと単語誤り率 (Word Error Rate: WER) が 18.8%の英語音声認識システムを用いて英日講義音声翻訳システムを作成し, 27.0 の BLEU を得ている.

文献 [5] では, 原言語コーパスに現れる頻度を基準に, SMT の学習コーパスでカバーしていない n-gram のフレーズを学習データへ順次追加していく手法が提案され, 我々も検討を行った [6]. しかし, n-gram 頻度基準を用いると, 複合句の一部が不完全なものとなり, 自動翻訳や人手による翻訳の質が低下するおそれがある. 文献 [7] では文献 [5] で追加するフレーズに対して, フレーズ間の重複問題や句構造の断片化問題等を指摘し, より少ないコストで翻訳精度の向上を図るため, 構文解析木の部分木を追加フレーズとして抽出し, 比較を行っている. その際, 句の長さや部分的な重複を考慮した句の極大性の概念を導入し, 追加するフレーズの選別を行っている.

本稿ではドメインへの適応のために, 構文解析木を用いた翻訳対象ドメイン内に存在する英語のフレーズの抽出及びその翻訳を作成し, 翻訳モデルに利用する検討を行った.

また, ドメイン適応のために専門用語のフレーズ対の登録方法も検討した.

2. 音声翻訳システムの概要 [6], [8]

音声翻訳システムは英語音声を英語文書へと変換を行う自動音声認識 (Automatic Speech Recognition: ASR) と英語文書を日本語文書に変換する統計的機械翻訳 (Statistical Machine Translation: SMT) を組み合わせることで構築される.

音声翻訳システムの概要図を**図 1** に示す. 本システムは英語講義映像から音声を抽出し, LIUM Speaker Dialization tools[9] を用いて自動的に発話単位 $\boldsymbol{o}$ に区切る. 発話 $\boldsymbol{o}$ は ASR に入力され, 英語文書として出力される. ASR からの出力文書 $\boldsymbol{f}$ を発話単位毎に SMT の入力とし, 発話単位の日本語文書 $\boldsymbol{e}$ として出力することで, 英語音声を日本語文書として出力する音声翻訳システムを構築した.

[*2] https://www.ted.com/talks?language=ja

3. 英語講義音声の日本語への翻訳

3.1 統計的音声翻訳

統計的音声翻訳問題は音響特徴を $\boldsymbol{o}$, 原言語を $\boldsymbol{f}$, 正解目的言語を $\boldsymbol{e}$ とすると $\boldsymbol{o}$ を与えられたときの, 尤もらしい目的言語 $\hat{\boldsymbol{e}}$ へと変換する問題として次式のように定式化される.

$$\hat{\boldsymbol{e}} = \arg\max_{\boldsymbol{e},\boldsymbol{f}} P(\boldsymbol{e},\boldsymbol{f}|\boldsymbol{o}) \quad (1)$$

$$= \arg\max_{\boldsymbol{e},\boldsymbol{f}} P(\boldsymbol{f}|\boldsymbol{o})P(\boldsymbol{e}|\boldsymbol{f}) \quad (2)$$

$$\approx \arg\max_{\boldsymbol{e},\boldsymbol{f}} P(\boldsymbol{o}|\boldsymbol{f})P(\boldsymbol{f})P(\boldsymbol{f}|\boldsymbol{e})P(\boldsymbol{e}) \quad (3)$$

ここで, $P(\boldsymbol{o}|\boldsymbol{f})$, $P(\boldsymbol{f})$ は音声認識システムにおける音響モデルと原言語の言語モデル, $P(\boldsymbol{f}|\boldsymbol{e})$, $P(\boldsymbol{e})$ は機械翻訳における翻訳モデルと目的言語の言語モデルと呼ばれ, これらの積を最大化する $\hat{\boldsymbol{e}}$ の探索問題を統計的音声翻訳問題とみなすことができる.

3.2 翻訳モデル

統計的機械翻訳には単語対応を得るための代表的なモデルとして, IBM の Brown らの報告した IBM モデルがある [11]. また, IBM モデルの単語単位の対応付けを原言語・目的言語のフレーズ間の対応付へと拡張する**フレーズベース翻訳**が提案されている [12]. しかしフレーズベース翻訳では対応付けされたフレーズの並び替えのみを行うため, 文法構造が大きく異なる二言語間の翻訳は困難である. この問題に対処するため, フレーズ間の対応付とその並び替えだけではなく, フレーズ内の単語の間に非終端記号を設け, フレーズ内に別のフレーズの挿入を許す**階層的フレーズ翻訳**の手法が提案されている [13]. また, 階層的フレーズベースで用いる非終端記号に原言語側の動詞句や名詞句などの句の構文情報を用いる **Tree-to-String** モデルが提案されており, 文法構造の異なる二言語の翻訳の精度向上を可能にしている [14]. 本稿では翻訳対象である MITOpenCourseWare (MITOCW) のテストデータに対

して, これらのモデルの翻訳精度の比較を行った.

3.3 構文解析を用いた追加フレーズの抽出

本稿で対象とする, 講義音声には特有の発話スタイルと講義ドメインに依存したフレーズがある. このため, 文献 [7] とは異なり, すでに学習コーパスによってカバーされているフレーズや, 被覆極大でないフレーズについても, 翻訳対象ドメインの書き起こしに高頻度で出現する場合は追加するフレーズとした.

翻訳対象ドメインである MITOpenCourseWare（MITOCW）に頻出するフレーズを翻訳モデルに優先して正しく翻訳させるため, 講義の書き起こしから英語のフレーズを抽出し, その日本語訳とのペアを追加フレーズ対として翻訳テーブルや学習コーパスに追加する.

先行研究 [8] では MITOCW で公開されている英語の講義書き起こしに対し, Text-NSP ツール [15] を用いて 3 から 6 の n-gram の単語組みの頻度をそれぞれ算出し, 上位の単語組みを n-gram それぞれについて抽出し, 英語側の追加フレーズとしていた. その後, 英語の追加フレーズに対してテストデータの正解の日本語訳を参照する **オラクルな条件** で翻訳を行い日本語側の追加フレーズとした.

本稿では, 3-gram の単語組みの多くは be 動詞や副詞のみで構成され, 日本語への翻訳が不可能であったため, 4 から 6 の n-gram 単語組みのみを追加フレーズとした. 抽出されたフレーズとその翻訳対は, 表 1 のような形式で, フレーズベースモデルの翻訳テーブルへ追加した. また, 自動的に日本語側のフレーズ訳を生成する試みとして翻訳サイトを用いて英語側フレーズを翻訳したものを日本語側フレーズとして用いた実験を行った.

n-gram 単位では句同士の一部が重なり, 翻訳句の量が増えてしまうことや, 意味的なまとまりになりづらく, 翻訳が困難になるという問題がある. 自動的に追加フレーズの翻訳を行う場合, 意味的なまとまりになっていることが望ましいと考えられる. また, 人手によって高品質な日本語訳を作る場合, 翻訳句の量が増えるにしたがってコストも増加してしまう.

この問題に対処するため, Berkeley Parser[*3] を用いて翻訳対象ドメインである MITOCW の書き起こしの構文解析木を作成し, 追加フレーズの抽出に利用する. 追加フレーズの抽出例を図 2 に示す. Berkeley Parser は与えられた英文に対して構文解析を行い, 解析木を生成する. この際, 句の品詞情報（動詞句, 名詞句等）も得ることができるが, 今回は使用していない. 次に, 生成された解析木の部分木の終端列が 4 単語以上のまとまりとなっている節を選択する.

表 1 翻訳テーブルへの追加フレーズ例

a piece of code ||| コード の 断片 ||| 0.3 0.3 0.3 0.3 ||| ||| 0 0 0 |||

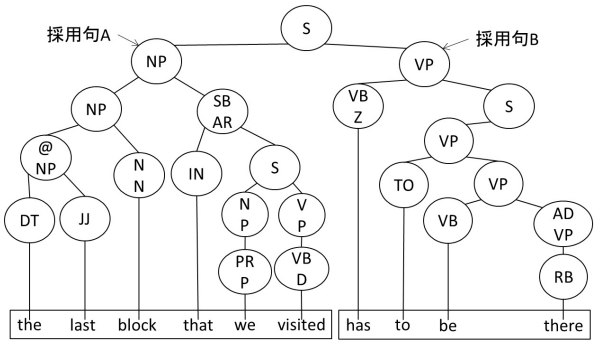

図 2 解析木を用いた英語の追加フレーズ抽出

この単語のまとまりをフレーズとみなし, MITOCW の講義書き起こしに対する頻度をフレーズの長さ（単語数）毎にそれぞれ算出し, 上位のものを英語側の追加フレーズとした.

導出された英語の追加フレーズに対して, 翻訳サイト Excite 翻訳 [*4] を用いて翻訳し, 人手による修正を行った. 比較のため, 日本語訳はテストデータの正解の日本語訳を参照するオラクルな条件でも行う. これを日本語側の追加フレーズとした. 高頻出フレーズ対の利用は次の二通りの手法について比較を行う.

(1) 翻訳テーブルへの登録
すでに話し言葉ドメインの英語–日本語パラレルコーパスによって学習された翻訳モデルの翻訳テーブルに対して, そのフォーマットに従い, 高頻出フレーズ対を登録する. その際, 翻訳確率や単語アライメント等のパラメータも登録する必要があり, 本稿では固定値を用いる.

(2) 学習コーパスへの追加
SMT を学習する英語–日本語パラレルコーパスに対して, それぞれに高頻出な英語, 日本語のフレーズを追加する. コーパスに追加する際には, 高頻出なフレーズを固定回数重複して追加することにより, 翻訳確率の調整を行っている.

3.4 翻訳ドメインに出現する専門用語への対応

講義というドメインには頻繁に専門用語が出現する. 専門用語は未知語となることが多く, また既知語であってもドメインに適していない日本語へ翻訳される場合が多い.

本稿では, ウェブ上で利用できる英日の翻訳サイトから参照することのできる専門用語の対訳用例から, テストデータに出現する専門用語を含む平均約 16 単語からなる対訳文をコーパス, あるいは一般的な講演ドメインによって学習された翻訳モデルに追加することで, 翻訳対象とする講義において出現する専門用語への対応を行った.

[*3] http://code.google.com/p/berkeleyparser/

[*4] http://www.excite.co.jp/world/

4. 実験条件

4.1 音声認識システム [6], [8]

4.1.1 データベース

音声認識システムに使用したデータベースとネットワーク構造を表 2 と表 3 に示す. 本稿では先行研究 [8] と同様のモデルを使用した. 多数の隠れ層を持つディープニューラルネットワーク（DNN）と, 隠れマルコフモデル（HMM）を組み合わせた DNN-HMM モデルを音響モデルとして用いている. 音響モデルは, 表 2 に示した書き言葉の音声コーパスである Wall Street Journal(WSJ) によって, 初期パラメータが学習される. その後, MIT 講義音声を追加学習することによって発話スタイルの適応と発話話者の適応を行っている.

英語の言語モデルの学習には, MIT 講義の英語書き起こし 20000 文と WSJ の英語書き起こし 49190 文を用いた. 認識語彙のサイズは 20000 語である. DNN の入力には左右 5 フレームの MFCC とエネルギー, その一次微分と二次微分の 39 次元の特徴量を 11 フレーム連結して用いた計 429 次元の特徴量を用いた. また出力はトライフォンの共有状態とし, 隠れ層 7 層の DNN を構築した.

音声認識のテストデータには MIT の男性話者 2 名の音声（各話者約 10 分, A:65 発話, B:94 発話）を用意し自動的に発話単位に区切った計 159 発話を使用した.

4.1.2 音声認識結果

本稿で用いる音声認識システムの認識精度を表 4 に示す. 本稿ではテストデータに対して単語誤り率が 21.0%（話者 A : 26.2%, 話者 B : 13.9%）の音声認識システムを使用する.

4.2 機械翻訳システム

4.2.1 データベース

機械翻訳システムに使用したデータベースを表 5 に示す. MIT の男性話者による英語講義の日本語翻訳システムを作成するため, 学習データとして TED Talk の英日対訳文 800PDF ファイル（約 100,000 文）を用いた. 翻訳モデルのパラメータの調整に, 認識対象と同一ドメインである対象話者 A の開発用講義書き起こし（54 文）の翻訳文を用いた. テストデータには音声認識に用いたテストデータの書き起こし（話者 A : 65 文, 話者 B : 94 文, 計 : 159 文）とその日本語訳を用意した.

日本語の言語モデルを作成するために TED Talk の書き起こし日本語翻訳文書約 100,000 文と日本語講義コーパスである CJLC[18] を約 1,000 文用いて, 5-gram の言語モデルを作成した.

本稿で扱うテストデータには, 音声を自動で区切った場合の書き起こしを発話とし, 翻訳対象とするため, 一発話に非常に多くの単語が含まれる場合がある. 表 6 にテストデー

表 2 音声認識データベース

コーパス	話者数	時間数	発話数	用途
WSJ	129	85 時間	49190	DNN 基礎モデル学習
	-	-	49190	英語言語モデル学習
MIT	23	4 時間	1672	発話スタイル適応
	2 （A, B）	10 分	125	話者適応
		10 分	159	テストデータ
	-	-	約 20000	英語言語モデル学習

表 3 DNN-HMM の仕様

特徴量	12 MFCCs+Δ+ΔΔ+energy +Δenergy+ΔΔenergy = 39 次元
入力層	左右 5 フレームコンテキスト [39*11 = 429 ノード]
隠れ層	7 層 [各 2048 ノード]
出力層	トライフォン共有状態数 [2001 ノード]
学習率（開始時）	0.01
学習率削減倍数	0.5
最少学習率	0.001
目的関数	Cross entropy criterion
活性化関数	Rectified linear

表 4 音声認識システムの単語誤り率（Word Error Rate;WER）

ASR	WER[%]		
	話者 A	話者 B	平均
DNN-HMM 発話スタイル& 話者適応	26.2	13.9	21.0

表 5 翻訳データベース

データ	データ数	用途
TED Talk	約 100,000 文	翻訳モデル学習（日英パラレル）
	約 100,000 文	言語モデル学習（日本語のみ）
MIT 講義	54 文（話者 A）	パラメータ調整（日英パラレル）
	159 文（話者 A:65 話者 B:94）	テストデータ（英語）
	約 45000 文	追加フレーズ抽出（英語のみ）
CJLC	約 1000 文	言語モデル学習（日本語のみ）
コンピュータ用語対訳例	9000	専門用語学習（英日パラレル）

表 6　テストデータに含まれる長い英文例

and so i joked badly i 'll agree at the end of last lecture that we can just stop now go straight to the final exam because this is all you need to know the point is yes it 's enough to start with but we want to add things to this that let us problem solve well

タに含まれる長文の例を示す. また, ASR からの出力を想定しているため, カンマやコロンといった区切り文字を除去し, 大文字小文字を小文字に統一している. テストデータの平均文長は 35.7 単語である. 実験で用いる翻訳モデルのデコーダは翻訳に広く用いられている Moses[16] を使用し, パラメータの調整には Moses に実装されている Minimum Error Rate Training（MERT）ツール [17] を用いている.

表 5 に記述したテストデータを人手で書き起こしたものと, ASR によって認識した結果を機械翻訳システムの入力として翻訳実験を行った. ベースラインは翻訳モデル学習コーパスを用いて翻訳モデルを学習し, パラメータ調整コーパスを用いて MERT を行ったモデルの翻訳精度である. 評価基準には機械翻訳で広く用いられている 4-gram の BLEU を用いた.

4.2.2　翻訳モデルの比較

フレーズベース翻訳モデル, 階層的フレーズベース翻訳モデル, Tree-to-String 翻訳モデルの作成には翻訳モデル学習データを用いた. また, モデル作成の際には Moses に実装されている各モデルの作成ツールを使用した. Tree-to-String 翻訳モデルの作成時には Moses にある Berkeley Parser のラッパーを使用した.

4.2.3　フレーズの追加条件

英語の追加フレーズ抽出のため, 表 5 に記述した MIT45000 文を追加フレーズ抽出コーパスとして用いる. 追加フレーズの翻訳には Excite 翻訳を人手で修正した翻訳結果を用いた. 翻訳テーブルへ追加する際は予め翻訳モデル学習コーパスによって学習された翻訳テーブルに対して, フレーズの翻訳確率を 0.3 に固定し, ワードアライメントを記述せずに追加した.

学習コーパスへ追加する際は翻訳モデル学習のためのパラレルコーパスに対して, 英語・日本語のフレーズをそれぞれ追加して翻訳モデルの学習を行った. 追加したフレーズを優先して翻訳されるようにするためコーパスへ同じフレーズを 1, 5, 10, 20, 30, 50 回それぞれ重複して学習を行った.

専門用語辞書の対訳例の利用について, 翻訳テーブルに追加する際にはあらかじめ 9000 文の対訳例のみを用いて SMT を学習する. その際に生成される翻訳テーブルを, Moses に付属する線形補完アルゴリズムを用いた翻訳テーブル統合ツールを利用して,TED で学習された翻訳テーブルとの統合を行った.

また, コーパスに追加する際には, 構文解析や n-gram の基準で導出した追加フレーズと同等の回数重複してパラレルコーパスに追加を行った.

翻訳テーブル・コーパスそれぞれについて追加をした後, パラメータ調整コーパスを用いて MERT ツールによるパラメータの重みの調整を行った.

4.2.4　人手による評価

本稿で用いたテストデータは, 正解訳が一通りだけであり, 評価基準 BLEU では言い回しや漢字の変換, 表記ゆれといった, 意味的には同じであっても翻訳の仕方が異なったためにその文の評価値も下がってしまう場合がある. フレーズを追加した翻訳モデルの出力にベースラインの BLEU スコアに対する大幅な改善は見られないとしても, 人が評価した時に翻訳文が改善している可能性は否めない. そのため, 人手による評価を行った.

被験者に原文の英語テストデータを確認させ, ベースラインと Excite 翻訳の人手修正を用いた追加フレーズ翻訳モデルの 2 つの翻訳モデルの翻訳結果のどちらがより原文の日本語訳に近いか, 内容が理解できるかというものを評価基準とした. また, 発話単位の長さによって翻訳の難しさが変わることを想定し, 一発話に 26 単語以上の英単語が含まれていた場合を長文, 26 単語未満の場合を短文とした. 更に長文の中には追加フレーズが複数含まれる場合に評価の差異が生まれることを考え, 追加フレーズが 1 回のみ出現する長文と 2 回以上出現する長文に分けて評価を行った. 評価者は英語の習熟度の高い 3 名である.

5.　実験結果

5.1　翻訳モデルの比較

フレーズベース, 階層的フレーズベース, Tree-to-String のそれぞれの翻訳モデルについてテストデータを翻訳した際の BLEU を表 7 に示す. 階層的フレーズベースの翻訳モデルを用いた場合, フレーズベースの翻訳モデルと比べて大きな精度の差は見られなかった. また, Tree-to-String の翻訳モデルを用いた際には精度の劣化が見られた. 翻訳対象である MITOCW のテストデータが話し言葉であるという点, 及び発話単位を自動的に区切っている点が構文解析へ悪影響を及ぼしたと考えられる.

これ以降の実験は, 翻訳精度の高い結果となったフレーズベース翻訳モデルに対する高頻出なフレーズ対を利用した結果を報告する.

表 7　翻訳モデルの比較結果
-テスト文：人手による書き起こし-

翻訳手法	話者 A	話者 B	平均値
フレーズベース翻訳モデル	11.7	8.1	10.2
階層的フレーズベース	11.0	8.7	10.1
Tree-to-String	9.2	7.4	8.6

5.2 テストデータにおける高頻出フレーズおよび専門用語の出現回数

n-gram, および構文解析によって導出した高頻出英語フレーズの, テストデータに出現する種類数, 延べ総数を表 8, 表 9 に示す. また, 専門用語の単語数, 専門用語がテストデータに出現する延べ総数を表 10, 表 11 に示す.

5.3 書き起こしの翻訳結果

作成した追加フレーズを翻訳テーブルへ追加した際の書き起こし文の翻訳実験結果を表 12 に示す.

追加フレーズの翻訳テーブルへの追加について, オラクルな条件（理想的な日本語訳を用いた場合のフレーズ対を追加した条件）で翻訳した場合と,Excite 翻訳を人手で修正したフレーズを用いた場合の両方に対して, n-gram と構文解析木共に翻訳精度の改善が得られた. また, オラクルな条件では n-gram の基準で抽出した追加フレーズは, 構文解析木の場合よりも大きな改善が得られている. これはテストデータに対するカバー率が高く, テストデータに対して高品質な翻訳の候補が翻訳テーブルへ追加されるためである. しかし, 翻訳サイトを参考に人手を用いて追加フレーズの日本語訳を作成した場合には, 構文解析木に基づくフレーズを用いて作成した追加フレーズ対の方が有効であった. これは上記した n-gram 基準の追加フレーズの翻訳の難しさが原因であると考えられる. また, 専門用語辞書を利用することで翻訳性能の改善がみられ, Excite 翻訳の人手修正を行ったフレーズの追加と併用することで得られる改善が増加した.

表 8 テスト内に出現する高頻出英語フレーズの種類数

導出手法	高頻出フレーズ種類数		
	話者 A	話者 B	合計
n-gram	393	182	570
構文解析木	63	29	92

表 9 テスト内に出現する高頻出英語フレーズの延べ総数

導出手法	高頻出フレーズ出現延べ総数		
	話者 A	話者 B	合計
n-gram	457	186	643
構文解析木	78	30	108

表 10 テスト内に出現する専門用語の種類数

専門用語フレーズ種類数		
話者 A	話者 B	合計
141	125	266

表 11 テスト内に出現する専門用語の延べ総数

専門用語出現延べ総数		
話者 A	話者 B	合計
629	657	1286

表 12 追加フレーズを翻訳テーブルへ追加した際の実験結果-テスト文：人手による書き起こし-

翻訳手法	話者 A	話者 B	平均値
ベースライン	11.7	8.1	10.2
Excite 翻訳+人手修正（n-gram）	12.1	6.9	10.0
Excite 翻訳+人手修正（構文解析木）	11.8	8.8	10.5
専門用語辞書利用	10.9	11.4	11.1
Excite 翻訳+人手修正（構文解析木）＋ 専門用語辞書利用	**12.5**	**10.0**	**11.6**
オラクル（n-gram）	22.9	18.4	20.9
オラクル（構文解析木）	18.0	13.2	15.9

表 13 Excite 翻訳+人手修正を用いた追加フレーズ翻訳モデルの翻訳結果の人手評価

文の種類	文数	ベースラインのほうが良い	Excite 翻訳+人手修正のほうが良い	判別不能
短文	22（x3 名）	9	25	32
長文 追加フレーズ 1 回出現	13 文（x3 名）	6	2	31
長文 追加フレーズ 2 回以上出現	22 文（x3 名）	20	15	31
合計	57 文（x3 名）	35	42	94

人手による翻訳結果の評価結果を表 13 に示す. 表 13 の 3 から 5 列目の数値はどちらの翻訳モデルが良いか, あるいは判別不能かを発話単位ごとに評価した 3 評価者の合計値を示している. 自動的に区切られた発話はしばしば長い発話となり, 翻訳に悪影響を及ぼしている. このため, 短文については翻訳テーブルへ追加フレーズを加えた場合のモデルが良い結果となったが, 長文については追加フレーズの出現回数にかかわらずほぼ同等の値となった. 理由としては, 表 6 に示したように長文の機械翻訳は難しすぎるためと思われる.

また, 追加フレーズを翻訳モデル学習のためのパラレルコーパスに追加した際の書き起こし文の翻訳実験結果を表 14 に示す. オラクルな条件でフレーズを追加した場合, 追加フレーズを 30 回重複して追加するまでは順に BLEU が改善されたが, 翻訳テーブルに追加する手法に比べて非常に精度が悪くなった. あまり性能の改善が見られない原因として, 独特な言い回しの日本語訳が単語単位で学習され, 翻訳テーブルに元々あるフレーズ内の単語に対して悪影響を及ぼしていると考えられる. Excite 翻訳+人手修正を用いた場合は回数を増やすことで精度が下がることはなく, 翻訳テーブルに追加する手法とほぼ同性能でベースラインよりも良い結果となった. ドメインに特有の発話スタイルに適応させる場合は単語単位のアライメントをせず, フレーズ単位のアライメントをすることが有効である事がわ

表 14 追加フレーズを学習コーパスへ追加した際の実験結果-テスト文：人手による書き起こし-

翻訳手法	追加倍数	話者 A	話者 B	平均値
ベースライン	-	11.7	8.1	10.2
Excite 翻訳+人手修正	1	9.5	7.1	8.6
	10	9.9	7.1	8.9
	30	11.6	7.9	10.0
	50	**12.5**	**7.9**	**10.6**
専門用語辞書利用	50	10.8	8.1	9.8
Excite 翻訳+人手修正 専門用語辞書利用	50	11.3	8.9	10.3
オラクル	1	9.3	6.4	8.0
	10	12.1	7.7	10.2
	30	**12.9**	**7.8**	**10.8**
	50	11.4	8.6	10.3

表 15 追加フレーズを翻訳テーブルへ追加した際の実験結果 -テスト文：ASR 出力-

翻訳手法	話者 A	話者 B	平均値
ベースライン	8.7	7.5	8.2
Excite 翻訳+人手修正（n-gram）	8.9	6.3	7.8
Excite 翻訳+人手修正（構文解析木）	8.5	7.7	8.2
専門用語辞書利用	9.3	10.1	9.7
Excite 翻訳+人手修正（構文解析木）＋ 専門用語辞書利用	**9.2**	**9.8**	**9.4**
オラクル（n-gram）	16.5	15.8	16.2
オラクル（構文解析木）	12.6	10.3	11.7

表 16 追加フレーズを学習コーパスへ追加した際の実験結果-テスト文：ASR 出力-

翻訳手法	追加倍数	話者 A	話者 B	平均値
ベースライン	-	8.7	7.5	8.2
Excite 翻訳+人手修正	1	6.9	6.7	6.9
	10	8.7	6.9	8.0
	30	9.1	7.6	8.5
	50	9.4	7.4	8.5
専門用語辞書利用	50	8.1	8.5	8.3
Excite 翻訳+人手修正 専門用語辞書利用	50	**9.3**	**8.3**	**8.9**
オラクル	1	7.6	7.0	7.2
	10	8.6	7.6	8.2
	30	8.3	7.6	8.0
	50	**8.4**	**8.2**	**8.4**

かった. また, コーパスに追加する場合では専門用語辞書を利用することによって得られる改善は少なかった. 直接的に確率の操作を行うことができるフレーズテーブルへの追加と異なり, 期待した専門用語への翻訳確率が低く学習されているためであると思われる.

5.4 音声認識結果の翻訳結果

ASR の出力を翻訳した実験結果を表 15 に示す． Excite 翻訳の人手修正による構文解析木を用いて抽出した追加フレーズ対を用いた翻訳モデルはベースラインとの差は見られなかった.

表 17 構文解析による導出フレーズ追加の有無によるテスト文の翻訳精度-テスト文：人手書き起こし-

翻訳手法	追加フレーズ含有文（102 文）	追加フレーズ非含有文（57 文）	全文（159 文）
ベースライン	10.4	9.9	10.2
翻訳テーブルへの追加（Excite 翻訳+人手修正）	11.2	9.7	10.5
学習コーパスへの追加（Excite 翻訳+人手修正）	11.8	8.9	10.6
翻訳テーブルへの追加（オラクル）	20.3	9.9	15.9
学習コーパスへの追加（オラクル）	10.5	9.8	10.3

追加フレーズを学習コーパスへ追加した際の ASR の出力を翻訳した実験結果を表 16 に示す. 翻訳テーブル, 学習コーパスともに追加フレーズを加えた際に, ASR 出力の翻訳結果に対しても同様の改善が見られた. このことから, 本稿と同じ条件でフレーズの追加を行う場合, 追加フレーズを人手で翻訳できる場合は翻訳テーブルへ追加し, 自動的に翻訳する場合は学習コーパスへ追加することでシステムの性能を改善できると考えられる.

5.5 高頻出フレーズを追加したことによる翻訳性能への影響

構文解析によって追加したフレーズを翻訳モデルに追加したことによって, テストデータ全体に対する BLEU の向上が見られた. これは追加したフレーズを含む発話に対する翻訳の精度が向上したためであると考えられる.

一方, 追加したフレーズを含まない分に対しては, 追加フレーズが悪影響を及ぼす可能性がある. そこで, 翻訳テーブルに追加する手法と学習コーパスで追加する手法について, 構文解析を用いて導出した追加したフレーズを含む発話と含まない発話について翻訳精度の比較を行った. 書き起こしの翻訳精度の比較結果を表 17 に示す. 追加したフレーズを含む発話に対しては, 各追加手法について改善が見られていることがわかる.

学習コーパスに追加する手法では Excite 翻訳の人手修正した追加フレーズを加えた場合, 追加フレーズとは関係のないテストデータに対する精度の劣化が見られる. これはたとえば英語の熟語”what’s go on” を Excite 翻訳では” オンの進行であるもの” というような訳し方をし, さらに単語アライメントを学習してしまうため, フレーズ追加前に学習されていた翻訳テーブルに対してノイズを与えてしまうからである. また, 追加フレーズに関係のあるテストデータに対してはオラクルな場合よりも Excite 翻訳を人手により修正した追加フレーズを加えたほうが良い結果となった. オラクルな場合では, 日本語の正解訳を参照して追加フレーズの日本語側が作成される. 日本語の正解訳は話し言葉であるため独特な翻訳であり, パラレルコーパスに追加

する形で学習することによって, 原言語と目的言語間のアライメントがうまく取れず, この場合も翻訳テーブルに対するノイズデータになってしまうからであると考えられる.

翻訳テーブルへの追加手法では, 追加したフレーズを含まない発話に対して大きな精度の減少は見られていない. これは追加したフレーズに対する単語アライメントを与えないことによって単語単位ではなく, フレーズ単位のスコアを重視した結果追加したフレーズに関係のない文章への影響がないということが考えられる.

学習コーパスへの追加手法を用いる場合は, 追加するフレーズと関係のないテストデータに対してはベースラインのモデルを組み合わせる手法 (10.9 の BLEU になりうる) などの検討が必要である.

6. まとめ

本稿では, 構文解析木を用いたフレーズの抽出及びその翻訳により作成した追加フレーズ対を用いることによる, 音声翻訳システムの改善の検討を行った. 追加フレーズ対を翻訳テーブルへ追加した場合, オラクルな条件では先行研究の手法である n-gram 基準の追加フレーズのほうが精度の向上が見込まれたが, 自動的に作成する場合は構文解析木を用いて作成した場合のほうが精度がよく, 人手による書き起こしを翻訳した場合にベースラインを上回る 10.5 の BLEU を得た. また, 専門用語の辞書を利用することでさらに 11.6 の BLEU を得ることができた.

学習コーパスへ追加した場合, オラクルな条件よりも Excite 翻訳の人手修正を追加した時に精度が良くなり, 書き起こしを翻訳した際にベースラインを上回る 10.6 の BLEU を, ASR の出力結果を翻訳した際にベースラインを上回る 8.5 の BLEU を得た. コーパスへの追加手法では専門用語辞書を利用することによる改善はあまり得られなかった.

謝辞 本研究は JSPS 科研費 25280062 の助成を受けた.

参考文献

[1] H. Abelson. *The creation of opencourseware at MIT.* Journal of Science Education and Technology. Vol. 17, No. 2, 2008, pp. 164–174.

[2] M. Cettolo, J. Niehues, S. stuker, L. Bentivogli, R. Cattoni and M. Federico. *The IWSLT 2015 Evaluation Campaign.* In proceedings of IWSLT 2015, 2015, pp. 2–14.

[3] M. Kolss, M.Wolfel, F. Kraft, J. Niehues, M. Paulik and A. Waibel, *Simultaneous German-English Lecture Translation*, In proceedings of IWSLT 2008, 2008, pp. 174–181.

[4] T.Hori, K.Sudoh, H.Tsukada, and A.Nakamura, *World-Wide Media Browser Multilingual Audio-visual Content Retrieval and Browsing System*, NTT Technical Review, Vol. 7, No. 2, 2009, pp. 1–7.

[5] Michael Bloodgood and Chris Callison-Burch. *Bucking the Trend: Large-Scale Cost-Focused Active Learning for Statistical Machine Translation.* In proceedings of ACL. 2010, pp. 854-864.

[6] 後藤 統興, 山本 一公 and 中川 聖一. 英語音声講義音声の認識と日本語への翻訳の検討. 日本音響学会論文集. 2015, pp. 181–184.

[7] 三浦 明波, Graham Neubig, Michael Paul and 中村 哲. 構文木と句の極大性に基づく機械翻訳のための能動学習電子情報通信学会技術研究報告書. Vol. 115, No. 347, 2015, pp. 109-115.

[8] Norioki Goto, Kazumasa Yamamoto and Seiichi Nakagawa. *English to Japanese Spoken Lecture Translation System by Using DNN-HMM and Phrase-based SMT.* In proceedings of ICAICTA, 2015.

[9] M. Rouvier et al. *An Open-source State-of-the-art Toolbox for Broadcast News Diarization.* In proceedings of INTERSPEECH, 2013, pp. 1477–1481

[10] 関 博史 and 中川 聖一. 音節単位 DNN-HMM による音声認識の検討. 研究報告音声言語情報処理. Vol. 2013-SLP-99, No. 4, 2013, pp. 1–6.

[11] P. Brown et al. *A statistical approach to machine translation.* In proceedings of Computational Linguistics. Vol. 16, No. 2, 1990, pp. 79–85.

[12] F. Och, P. Koehn and D. Marcu. *Statistical phrase-based translation.* In proceedings of HLT-NAACL. 2003, pp. 48–54.

[13] David Chiang. *A Hierarchical Phrase-Based Model for Statistical Machine Translation.* in proceedings of the 43rd Annual Meeting of the ACL. 2005, pp. 263–270.

[14] Yang Liu , Qun Liu , and Shouxun Lin. *Tree-to-String Alignment Template for Statistical Machine Translation.* in proceedings of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the ACL. 2006, pp.609-616.

[15] Satanjeev Banerjee and Ted Pedersen. *The design, implementation, and use of the ngram statistics package.* in proceedings of Computational Linguistics and Intelligent Text Processing. 2003, pp.370–381.

[16] P. Koehn et al. *Moses: Open source toolkit for statistical machine translation.* In proceedings of the ACL 2007. 2007, pp. 177–180.

[17] F. J. Och. *Minimum error rate training for statistical machine translation.* In proceedings of The 41st Annual Meeting of the Association for Computational Linguistics. 2003, pp. 160–167.

[18] S. Kogure et al. *Speech Recognition Performance of CJLC: Corpus of japanese Lecture Contents.* In proceedings of INTERSPEECH. 2008, pp. 1554–1561.