Vol.2015-SLP-105 No.8
2015/2/28

# 講義音声の前処理と2段階翻訳に基づく日英音声翻訳

川口 亮[1,a)]　山本 一公[1,b)]　中川 聖一[1,c)]

**概要**：日本語の講義を対象とした日英音声翻訳を行う．機械翻訳を行う際に必要となる翻訳する言語間のフレーズの対応のとれた大量のパラレルコーパスは，通常，書き言葉のコーパスである．講義音声のような話し言葉の機械翻訳においては書き言葉のコーパスはそのままでは利用できない．本稿では，書き言葉で構成されているコーパスを講義音声の機械翻訳でも利用可能にするために，講義音声を話し言葉から書き言葉に整形する手法を提案する．話し言葉から書き言葉への整形はフィラーの除去及び，人手で定めたルールに基づいた整形の手法を検討する．また，日本語と英語では語順が大きく異なるため翻訳性能が語順の近い言語間と比較し低下することがある．そこで日本語の語順を並び替えてから英語に翻訳する方法と日本語から英語に機械翻訳を行った後に，再度機械翻訳で英語に翻訳を行う方法で語順の違いに対処する手法を検討した．

# Lecture Speech Translation based on Preprocessing and 2-Step Translation.

Kawaguchi Ryo[1,a)]　Yamamoto Kazumasa[1,b)]　Nakagawa Seiichi[1,c)]

## 1. はじめに

近年，国内外の大学などでウェブサイト上で講義映像を配布し，多くの人が容易に様々な講義を受けることが可能となった [1] [2]．しかし，現状では多くの講義が英語であり，英語が母国語ではない日本人にとっては理解し難いものが大半となっている．我々の研究室の Veri の先行研究 [3] において，英語講義映像に対して日本語字幕を付与することにより，例えば重要句の字幕表示だけでも日本人学生の講義の理解度が向上するということが分かっている．英語講義映像に対して日本語字幕を付与するためには，英語講義の音声を聞き取り，内容を日本語に翻訳する必要があるが，人手による翻訳はコストが高く，増え続けている講義全てに対して行うのは非現実的である．そこで，コンピュータによる機械翻訳を利用する．

音声翻訳を行うには音声認識と機械翻訳を組み合わせる必要があるが，音声認識りにより機械翻訳に悪影響を与えてしまうことがある．Saon ら [4] や，Casacuberta [5], Bertoldi ら [6] はワードラティスや，コンフィージョンネットワークといった音声認識結果の複数候補を使用することにより，音声認識誤りに頑健な翻訳を行う方法を提案している．また，日本語と英語のように文法構造が大きく異なる言語間の翻訳では，文法構造が近い言語間に比べると翻訳精度が低くなる傾向がある [7]．この文法構造の違いを解決するために福田ら [8] は，ルールベース翻訳と，統計的な情報を用いた翻訳の２つの翻訳手法を組み合わせる手法を提案している．星野ら [9] は，事前に日本語の文章を文レベルおよび，句レベルで英語に近づけた後に翻訳を行うという手法を提案している．

機械翻訳を行う際には，翻訳する言語間のフレーズの対応のとれた大量のパラレルコーパスが必要となる．講義音声ではニュースのように原稿を読み上げているような書き言葉ではなく，話し言葉と呼ばれるものであり，機械翻訳の学習で使用することができるコーパスの多くが両言語で発行されている新聞記事などの書き言葉のものであるため，そのままでは利用できない．そのため，話し言葉を翻訳しやすくするために，Hany ら [10] らは会話音声の分割と，言い淀みの除去を行う手法を提案している．また，Zhang ら [11] らも句読点の予測をし，会話音声の分割を行うという手法を検討している．本稿では，書き言葉で構成されているコーパスを，講義音声の機械翻訳でも利用可能にするために，講義音声を話し言葉から書き言葉に整形を行う手法を提案する．話し言葉を書き言葉に整形する研究は講演等を対象にして既に行われている．下岡 [12] らは話し言葉から書き言葉への整形を翻訳問題とし，統計的機械翻訳の手法を用いている．また，堀ら [13] は整形を行うためのモデルを WFST として表現する手法を提案している．Neubig [14] らは log-linear フレームワークの枠組みによる整形処理を WFST を用いて実現した．また，日本語と英語のように語順が大きく異る言語間の機械翻訳では，語順の近い言語間の機械翻訳と比べると性能が低くなることがある．そこで，日本語の語順を並び替えてから英語に翻訳

[1] 豊橋技術科学大学
Toyohashi University of Technology
[a)] kawaguchi@slp.cs.tut.ac.jp
[b)] kyama@slp.cs.tut.ac.jp
[c)] nakagawa@slp.cs.tut.ac.jp

する方法と，一度日本語から英語に機械翻訳を行ったのちに，再度英語-英語の機械翻訳を行う方法で，語順の違いによる翻訳性能の低下を防ぐ手法を提案する．

講義音声を対象とした音声翻訳を Waibel ら [15] は行っている．また，堀ら [16] はマサチューセッツ工科大学の講義を大量のパラレルコーパスを用いて英語から日本語に翻訳する研究を行っている．本稿では講義音声の大量のパラレルコーパスがない場合の日本語講義を英語に翻訳する方法を検討する．

## 2. 講義音声の認識

### 2.1 音声認識

音声認識は，特徴ベクトル系列 $O = o_1^T$ が与えられたとき，単語系列 $W = w_1^M$ が意図されたとする確率 $P(W|O)$ を最大化する $\hat{W}$ を求める問題として，次式のように定式化される．

$$\hat{W} = \underset{W}{\operatorname{argmax}} P(W|O) \quad (1)$$

$$= \underset{W}{\operatorname{argmax}} P(O|W)P(W) \quad (2)$$

ここで，$P(O|W)$ は音響モデル，$P(W)$ は言語モデルである．

#### 2.1.1 音声分析と特徴パラメータ

対数パワースペクトルに対し，各帯域に相当する三角窓関数をメル尺度で乗じることでフィルタバンクの出力を得る．得られたメル化対数パワースペクトルに対し逆コサイン変換を行い, 時間波形化することによって得られるケプストラム係数をメル周波数ケプストラム係数 (Mel-Frequency Cepstrum Coefficients: MFCC) と呼ぶ．10ms ごとに得られる 1〜13 次の MFCC は音声認識において一般的に使用される特徴パラメータとなっている．

#### 2.1.2 音響モデル

音響モデルは音声信号の音素・音節らしさをモデル化したものとなっており，一般的に非定常な時系列からなる混合ガウス分布を出力確率とした隠れマルコフモデル (Hidden Markov Modell: HMM) によって構築される．各状態から出る特徴パラメータを混合ガウス分布でモデル化する方法を GMM-HMM と呼ぶ．一方，ディープニューラルネットワークで各状態における特徴パラメータの事後確率を求めるモデルを DNN-HMM と呼ぶ.

#### 2.1.3 言語モデル

言語モデルは，単語の並びが妥当であるほど高い確率を与えるモデルとなっている．言語モデル $P(W)$ には，単語 $w_i$ の生起確率を，直前の $N-1$ 個の単語 $w_{i-N+1}^{i-1}$ の条件付き確率 $P(w_i|w_{i-N+1}^{i-1})$ として表現する $N$-gram モデルを用いるのが一般的である．$N$-gram モデルでは，単語列 $W$ の生起確率は次式によって求められる．

$$P(W) = \prod_{i=1}^{M} P(w_i|w_{i-N+1}^{i-1}) \quad (3)$$

## 3. 音声翻訳

統計的機械翻訳 (Statistical Machine Translation:SMT) [17] は，近年機械翻訳の手法なかでも注目されている手法であり，原言語と目的言語の対訳の存在する大量のコーパスを用いることによって，統計的な特徴により翻訳規則を学習し翻訳を行う手法である．

統計的機械翻訳は，原言語文 $f$ が与えられたとき，目的言語文 $e$ が意図されたとする確率 $P(e|f)$ を最大化する $\hat{e}$ を求める問題として，次式のように定式化される．

$$\hat{W} = \underset{e}{\operatorname{argmax}} P(e|f) \quad (4)$$

$$= \underset{W}{\operatorname{argmax}} P(f|e)P(e) \quad (5)$$

ここで，$P(f|e)$ は翻訳モデル，$P(e)$ は言語モデルである．

### 3.1 翻訳モデル

翻訳モデルは原言語の単語列から目的言語の単語列へ翻訳される確率を計算するモデルであり，翻訳の確からしさを表すモデルである．翻訳モデルを構築するためには，翻訳言語間の単語または，フレーズの対応を取る必要がある．フレーズの対応アライメントを $\alpha$ とすると翻訳モデルは

$$P(f|e) = \sum_{\alpha} P(f, \alpha|e) \quad (6)$$

$$= \sum_{\alpha} P(f|e, \alpha)P(\alpha|e) \quad (7)$$

と表すことが出来る．対応を考慮した $P(f, \alpha|e)$ の計算方法として EM アルゴリズムによる IBM モデルを用いて推定を行う．IBM モデルにはモデル 1 からモデル 5 まで存在し，対応付けの特徴をモデル化するもである．高次になるほど複雑なモデルとなり，精度が上がる．

IBM モデルは単語単位の対応付けである．これを複数の単語からなるフレーズ間の対応付けへと適応する必要がある [18]．つまり，原言語と目的言語のフレーズを 1 対 1 で対応付けを取る必要がある．翻訳モデル $P(f|e)$ における $f$，$e$ はそれぞれお $l$ 個のフレーズに分割され以下の式で表される．

$$P(f_1^l|e_1^l) = \prod_{i=1}^{l} \phi(f_i|e_i)d(start_i - end_{i-1} - 1) \quad (8)$$

ここで $\phi(f_i|e_i)$ はフレーズの翻訳確率を表しており，$d(x)$ は並び替えのコストを表している．

### 3.2 言語モデル

言語モデルは翻訳された目的言語の単語列が目的言語らしくなるように使われる．言語モデルは音声認識で一般的に用いられる $N$-gram 式 (3) と同様のものを使うのが一般的である．音声認識では，通常 3-gram，機械翻訳では 5-gram を使用する．

### 3.3 評価尺度

機械翻訳における自動評価手法として，BLEU(Bilingual Evaluation Understudy) [19] が広く使われており標準的な評価手法ととなっている．BLEU は $n$-gram の幾何平均で評価を行う適合率ベースの評価手法であり，以下の式で計算することができる．

$$BLEU = BP \cdot exp(\frac{1}{N}\sum_{n=1}^{N} logP_n) \quad (9)$$

$$P_n = \frac{\text{翻訳文と正解文で一致した } n-gram \text{ 数}}{\text{翻訳文中の } n-gram}$$

$$BP = min(1, exp(1 - r/c))$$

式中の $c$ は翻訳文の単語数，$r$ は正解文の単語数である．

また $N$ は $n$-gram である．$BP$ は翻訳文が正解文より短い単語数の場合にかかるペナルティである．正解文に現れている単語の多くが翻訳文に現れていなくても，翻訳文と正解文の単語数の差が大きく，なおかつ翻訳文の単語数が正解文に現れていれば，不正に $P_n$ の値は高くなってしまうため，これを防ぐためにペナルティを設けている．BLEU の $n$-gram 数が大きいほど流暢さが重要視されていることになり，逆に値が小さいと適切さが重要視されていることになる．本稿では 3-gram および，4-gram を用いた．なお，日本語と英語のように語順が大きく異る言語間では BLEU は適切な尺度ではない点に注意を要する必要がある [20].

## 4. 話し言葉 - 書き言葉整形

一般的に音声翻訳を行う際には，音声認識を行った後に原言語の話し言葉コーパスと目的言語の書き言葉コーパスを用いて話し言葉を考慮した翻訳を行うことが多い．しかし今回のタスクでは，話し言葉である講義と合った大量のパラレルコーパスを入手するのは困難である．そこで，本稿では，話し言葉-書き言葉整形を行うことにより，原言語の書き言葉コーパスと目的言語の書き言葉コーパスの対訳コーパスを利用できるようにすることを目的とする．

### 4.1 フィラーの削除

話し言葉特有の発話の種類として多いのが「まぁ」,「えー」,「えっと」などの発話中に単語間に挟み込まれるフィラーであり，話者に依存せずに見られる現象である．従って，このフィラーを削除することにより，音声翻訳の精度を向上させることができると考えられる．フィラーの削除には我々の研究室での先行研究 [21] を用いた．この手法では文境界が未知の話し言葉系列に対して，文境界検出と依存構造解析を同時に行うアルゴリズムである Improved-SDA [22] を用いてフィラーの除去を行っている．Improved-SDA によって係り先が存在しなかったフィラーが削除されるようになっている．

### 4.2 ルールに基づく話し言葉整形

話し言葉特有の表現として，音便により本来の発音と異なる発音へと変化する，繰り返し同じ単語を発音してしまうといった特徴がある．こういった特徴をルールとして作成することで，話し言葉を書き言葉に整形することができると考えられる．変換ルールの例を図 1 に示す．講義音声を人手により書き起こしたものに対して話し言葉-書き言葉整形を行った例を図 2 に示す．

話し言葉特有の表現 例:「じゃなく」→「ではなく」,「やることは」→「すること」
言い直し 例:「それは それは違います」→「それは違います」

**図 1** 話し言葉-書き言葉整形ルールの例

## 5. 翻訳言語間での語順の対応

統計的機械翻訳を行う際には，日本語と英語といった文法構造上，主語と述語の並び順が大きく異なっていると，翻訳性能に悪影響を及ぼすことがある．

### 5.1 述語構造解析器を用いた語順の並び替え

英語は主語の後に述語という語順であるが，日本語では

書き起こし
じゃあつづきをやります
まあ音声のお一発声器官はどうなっているかという
模式図ですけどもえーまあ先ほど言ったように

整形後
それではつづきをやります
音声の発声器官はどうなっているかという
模式図ですけが. 先ほど言ったように

**図 2** 整形例

必ずしも主語の直後に述語が来るとは限らないので，この語順を合わせることにより，翻訳の性能が向上すると考えられる．並び替えの例を図 3 に示す．上の文がもともとの文で，下の文が英語に語順を合わせたものになっている．

範疇というのはカテゴリーというカテゴリカルな知覚であります
範疇というのは知覚でありますカテゴリーというカテゴリカルな

**図 3** 述語構造解析を用いた語順の並び替え例

また，日本語は主語が抜けることが多いが，主語がないときには述語を先頭に持ってくることとした．述語構造解析には SynCha[*1] を使用した．実行例を図 4 に示す．「酷使していますの」ga=”1”は id=”1”の形態素をガ格としていることを示しているので，「心臓は」という id=”1”の部分に掛かっていることが分かるので，「心臓は」の直後に「酷使しています」を移動させる．

```
* 0 2D 0/1
心臓　名詞,一般,*,*,,,心臓,シンゾウ,シンゾー,,　O　id="1"
は　助詞,係助詞,*,*,,,は,ハ,ワ,,　O
* 1 2D 0/0
もっと 副詞,一般,*,*,,,もっと,モット,モット,, O
* 2 -1D 1/4 0.000000
酷使　名詞,サ変接続,*,*,,,酷使,コクシ,コクシ,,　O
し　動詞,自立,*,*,サ変・スル,連用形,する,シ,シ,,　O　ga="1" type="pred"
て　助詞,接続助詞,*,*,,,て,テ,テ,, O
い　動詞,非自立,*,*,一段,連用形,いる,イ,イ,,　O
ます　助動詞,*,*,*,特殊・マス,基本形,ます,マス,マス,, O
```

**図 4** SynCha の実行例

### 5.2 2 段階統計的機械翻訳による語順の並び替え

原言語から目的言語への翻訳を行った後に，その翻訳結果の語順を統計的機械翻訳を用いて並び替える事によって語順を原言語と目的言語へと合わせることができると考えられる．原言語を日本語，目的言語が英語の場合を例に上げると，まず一回目の翻訳で日本語を語順が日本語に近くなってもよい英語に翻訳する．その後に不正確な英語の語順と正確な英語の語順のパラレルコーパスにより学習した機械翻訳システムで，再度翻訳を行い，英語本来の語順に近づけるという手法である．

## 6. 音声翻訳の実験及び評価

### 6.1 実験データ

評価実験では，日本語から英語への翻訳実験を行った．日本語講義音声コンテンツコーパス (Corpus of spoken Japanese Lecture Contents: CJCL) [23] を対象として実験

[*1] http://www.cl.cs.titech.ac.jp/~ryu-i/syncha/

を行った．このコーパスには大学における講義音声と音声データを人手で書き起こした発話内容が含まれている．テストを行う際には，このコーパスから抜き出した二人の話者の講義内容と音声の書き起こし，それぞれ 131 文と 104 文を使用した．また，開発セットとしてテストセットのうちの一人の話者のデータ，99 文を使用した．テストセット，開発セットともに，二人の専門家に翻訳を依頼したものを翻訳の正解結果として用いた．この際の BLEU スコアは，文毎のスコア計算時に高くなる方の正解を採用することで計算を行う．

テストセットと開発セットの両方のデータが有る話者を話者 A, テストセットしか存在しない話者を話者 B とする．モデルの学習には，パラレルコーパスとしては新聞記事 (JENAAD, REUTERS)，講義で使われていた教科書の日英版, および TED を使用した．加えて，英語の言語モデルの学習には MIT オープンコースウェアの書き起こしのものを使用した．コーパスの諸元を表 1 に示す．

**表 1** コーパスの詳細

| | 単語数 | 異なり単語数 | 文数 |
|---|---|---|---|
| 教科書 (英語) | 63858 | 4192 | 2294 |
| 教科書 (日本語) | 77233 | 3286 | 2294 |
| 新聞記事 (JENNAD,REUTERS)(英語) | 5390252 | 75227 | 205684 |
| 新聞記事 (JENNAD,REUTERS)(日本語) | 6879266 | 48051 | 205684 |
| TED(英語) | 2001844 | 40592 | 100324 |
| TED(日本語) | 2253233 | 33902 | 100324 |
| MIT(英語) | 5850652 | 42320 | 82202 |

### 6.2 実験条件

機械翻訳のデコーダとしては，Moses を用いた．言語モデルの学習には SRILM [24] を使用し，$Kneser-Ney$ スムージング [25] による前向き 5-gram の言語モデルを作成し，翻訳モデルの学習には，アライメント対応として，GIZA++ [26] で学習を行った．また，開発セットを使用した特徴関数の重みの調整には MERT [27] を用いた．

音声認識のデコーダとしては, GMM-HMM において認識実験を行う際には本研究室で開発された SPOJUS++ [28] を用いた．DNN-HMM において認識実験を行う際には，デコーダに SPOJUS++WFST 版を用いた．音響分析条件は表 2 に，GMM-HMM の条件は表 3 に，DNN については表 4 とした [30]．音響モデルとして，日本語話し言葉コーパス [29] の学会講演，模擬講演から学習した左コンテキスト依存音節モデル（left-to-right 型 HMM，5 状態 4 出力分布，4 混合全共分散行列，男性話者モデル）を用いた．モデル数は 928 である．

言語モデルの学習には NHK ニュース音声コーパス (単語数 1564848, 異なり単語数 32975) を使用し，Palmkit[*2] を用いて学習した 3-gram モデルを用いた．

評価尺度には BLEU を使用した．また，MERT によるパラメータ調整は話者 A の開発セットを用いた．

### 6.3 翻訳実験結果 - 人手書き起こし -

提案手法の有効性を確認するために，まず人手による書き起こしに対しての翻訳実験を行った．

#### 6.3.1 ベースラインの結果

ベースラインとして翻訳モデル，言語モデルの学習に新聞のみ，及び新聞と講義で使用した教科書の日英版のみを使用した実験結果を表 5(a) に示す．パラメータ調整には，モデルの学習に使用した新聞記事のうちの一部を利用した．

[*2] http://palmkit.sourceforge.net/

**表 2** 音響分析条件

| | |
|---|---|
| サンプリング周波数 | 16kHz |
| プリエンファシス | 0.98 |
| 分析窓 | Hamming 窓 |
| 分析窓長 | 25ms |
| フレームシフト | 10ms |
| 特徴パラメータ (38 次元) | MFCC + △MFCC + △△MFCC + △Pow + △△Pow |

**表 3** GMM-HMM の条件

| | |
|---|---|
| 単位 | 左コンテキスト依存音節 (928 種) |
| 状態数 | 5 状態 4 出力分布 |
| 混合数 | 4 混合全分散行列 |

**表 4** DNN の条件

| | |
|---|---|
| 入力層 | 429 ユニット (11 フレーム) |
| 中間層 (3 層) | 2048 ユニット |
| 出力層 | 5097 ユニット |
| 誤差関数 | CrossEntropy |
| 活性化関数 | Rectified Linear Unit |

**表 5** ベースラインの実験結果

(a) パラメータ調整が新聞記事-話者 A によるテスト結果-

| 整形の種類 | 翻訳モデル | 言語モデル | BLEU-3 | BLEU-4 |
|---|---|---|---|---|
| 整形なし | 新聞記事 | 新聞記事 | 5.82 | 2.48 |
| 整形なし | 新聞記事+教科書 | 新聞記事+教科書 | 5.94 | 2.61 |

(b) パラメータ調整が新聞記事-話者 B によるテスト結果-

| 整形の種類 | 翻訳モデル | 言語モデル | BLEU-3 | BLEU-4 |
|---|---|---|---|---|
| 整形なし | 新聞記事 | 新聞記事 | 5.94 | 2.56 |
| 整形なし | 新聞記事+教科書 | 新聞記事+教科書 | 6.01 | 2.84 |

#### 6.3.2 話し言葉-書き言葉整形

音声の人手による書き起こしに対して整形を行い，翻訳を行った結果を表 6 に示す．表 6(a) はパラメータ調整の開発セットとテストセットで話者が同じ時の実験結果，表 6(b) は開発セットとテストセットで話者が異なる際の実験結果である．パラメータ調整は少量であっても同じドメインのコーパスの利用が良かった．しかし，パラメータ調整が話者独立の場合は話者依存の場合と比較して，その効果は小さい．パラレルコーパスについては利用可能なものは全て利用した際の結果が良い結果が得られた．また，ルールベース，フィラーの除去ともに性能の向上を確認することができた．更に 2 つの組み合わせることにより，より良い性能が得られた．

また，開発セットとテストセットの話者が同一の場合については，翻訳モデル，言語モデルの学習に使用するコーパスの組み合わせを複数試したが，翻訳モデルについては「新聞記事+教科書+TED」を，言語モデルについては「新聞記事+教科書+TED+MIT」を使用したものが最も BLEU 値が高くなったため，以後の実験でもこの条件で実験を行った．

本大学の福島ら [31] の研究では，統計的機械翻訳により，話者 A の話し言葉-書き言葉の翻訳を行うことにより整形を行う方法を提案しており，講義音声に対して BLEU の 4-gram で値 5.64 を得ている．

#### 6.3.3 述語構造解析器を用いた語順の並び替え

語順の並び替えを行うために, 係り受け解析器 CaboCha[*3]

[*3] https://code.google.com/p/cabocha/

**表 6** 整形による翻訳性能の実験

(a) パラメータ調整が話者依存の話者 A によるテスト結果

| 整形の種類 | 翻訳モデル | 言語モデル | BLEU-3 | BLEU-4 |
|---|---|---|---|---|
| 整形なし | 新聞記事+教科書 | 新聞記事+教科書 | 7.63 | 4.61 |
| 整形なし | 新聞記事+教科書+TED | 新聞記事+教科書+TED+MIT | 7.85 | 4.78 |
| ルールベース | 新聞記事+教科書+TED | 新聞記事+教科書+TED+MIT | 8.12 | 4.81 |
| フィラーの除去 | 新聞記事+教科書+TED | 新聞記事+教科書+TED+MIT | 8.53 | 5.21 |
| ルールベース+フィラーの除去 | 新聞記事+教科書+TED | 新聞記事+教科書+TED+MIT | 8.65 | 5.64 |

(b) パラメータ調整が話者独立の話者 B によるテスト結果

| 整形の種類 | 翻訳モデル | 言語モデル | BLEU-3 | BLEU-4 |
|---|---|---|---|---|
| 整形なし | 新聞記事+教科書 | 新聞記事+教科書 | 6.08 | 3.10 |
| 整形なし | 新聞記事+教科書+TED | 新聞記事+教科書+TED+MIT | 6.15 | 3.21 |
| ルールベース | 新聞記事+教科書+TED | 新聞記事+教科書+TED+MIT | 6.21 | 3.39 |
| フィラーの除去 | 新聞記事+教科書+TED | 新聞記事+教科書+TED+MIT | 6.43 | 3.61 |
| ルールベース+フィラーの除去 | 新聞記事+教科書+TED | 新聞記事+教科書+TED+MIT | 6.45 | 3.67 |

原文 :長さニセンチぐらいの筋肉がこのように閉じたり開いたりします
並び替え :長さニセンチぐらいの筋肉が閉じたり開いたりしますこのように

原文 :色はほんとうは無限個の色に連続的に変わっているのです
並び替え :色は変わっているのですほんとうは無限個の色に連続的に

**図 5** 並び替えの成功例

**表 7** 述語構造解析器を用いて並び替えを行った実験結果

| テストセット | 翻訳モデル | 言語モデル | BLEU-3 | BLEU-4 |
|---|---|---|---|---|
| 話者 A | 新聞記事+教科書+TED | 新聞記事+教科書+TED+MIT | 8.75 | 5.79 |
| 話者 B | 新聞記事+教科書+TED | 新聞記事+教科書+TED+MIT | 6.51 | 3.73 |

と，述語構造解析器 SynCha を使用した．SynCha により，述語がどの形態素に掛かっているのかが分かるため，この情報を元に並び替えを行った．並び替えの成功例を図 5 に示す．上のものが並び替え前，下のものが並び替え後となっている．

上記のようにして英語と並び替えた日本語のコーパスから翻訳モデルの学習を行い実験を行ったものを表 7 に示す．前節の話し言葉-書き言葉整形で最も性能の良かった，ルールベース+フィラーの除去も行った．トレーニングデータも前節と同じデータを利用している．

パラメータ調整が話者依存，話者独立どちらのテストセットも共に，0.1 程度 BLEU が向上した．これは並び替えが上手くいき，英語と日本語の対応が取りやすくなったためだと考えられる．

#### 6.3.4 2 段階統計的機械翻訳による語順の並び替え

学習に用いたコーパスは日→英，英→英ともにパラレルコーパスとしては，「新聞記事」「教科書」「TED」を使用し，言語モデルはパラレルコーパスで使用したコーパスの英語コーパスに加え，MIT のコーパスを使用した．学習の方法としては，通常の方法で言語モデル，翻訳モデルともに学習を行って一回目の翻訳を行い英語に翻訳する．その後，機械翻訳された英語と，パラレルコーパスの日本語の正解翻訳結果の英語とのパラレルコーパスで，再度翻訳モデルの学習を行い，学習されたモデルを用いて語順の正しい英語に翻訳するという手順である．実験結果を表 8 に示す．翻訳手法の行の日→英というのは日本語から英語に翻訳するということを表していて，これまでと同じ翻訳の仕方である．日→英→英というのが日本語から英語に翻訳した後に，もう一度機械翻訳を用いて語順の並び替えを行うということを示している．

どちらのテストセットでも日→英だけの場合と比べると，性能が 0.1 程度向上するという結果が得られた．

述語構造解析器を用いた語順の並び替えと 2 段階統計的機械翻訳による語順の並び替えを組み合わせて使用した際の実験結果を表 9 に示す．組み合わせることにより少しではあるが，性能の向上が見られた．

**表 8** 2 段階翻訳による翻訳性能の実験

(a) パラメータ調整が話者依存の話者 A による日→英→英テスト結果

| 整形の種類 | 翻訳手法 | BLEU-3 | BLEU-4 |
|---|---|---|---|
| 整形なし | 日→英 | 7.85 | 4.78 |
| 整形なし | 日→英→英 | 7.93 | 4.84 |
| ルールベース+フィラーの除去 | 日→英 | 8.65 | 5.64 |
| ルールベース+フィラーの除去 | 日→英→英 | 8.71 | 5.72 |

(b) パラメータ調整が話者独立の話者 B による日→英→英テスト結果

| 整形の種類 | 翻訳手法 | BLEU-3 | BLEU-4 |
|---|---|---|---|
| 整形なし | 日→英 | 6.15 | 3.21 |
| 整形なし | 日→英→英 | 6.22 | 3.33 |
| ルールベース+フィラーの除去 | 日→英 | 6.45 | 3.67 |
| ルールベース+フィラーの除去 | 日→英→英 | 6.54 | 3.72 |

**表 9** 述語構造解析，2 段階統計的機械翻訳による語順の並び替えの組み合わせのテスト結果

(a) パラメータ調整が話者依存の話者 A による日→英→英テスト結果

| 整形の種類 | 翻訳手法 | BLEU-3 | BLEU-4 |
|---|---|---|---|
| 整形なし | 日→英 | 7.91 | 4.82 |
| 整形なし | 日→英→英 | 7.93 | 4.84 |
| ルールベース+フィラーの除去 | 日→英 | 8.75 | 5.79 |
| ルールベース+フィラーの除去 | 日→英→英 | 8.78 | 5.82 |

(b) パラメータ調整が話者独立の話者 B による日→英→英テスト結果

| 整形の種類 | 翻訳手法 | BLEU-3 | BLEU-4 |
|---|---|---|---|
| 整形なし | 日→英 | 6.17 | 3.22 |
| 整形なし | 日→英→英 | 6.25 | 3.35 |
| ルールベース+フィラーの除去 | 日→英 | 6.51 | 3.73 |
| ルールベース+フィラーの除去 | 日→英→英 | 6.59 | 3.76 |

**表 10** パラメータ調整が話者依存の話者 A の音声認識結果による日→英→英テスト結果

(a) GMM-HMM

| 整形の種類 | 翻訳手法 | BLEU-3 | BLEU-4 |
|---|---|---|---|
| 整形なし | 日→英 | 3.44 | 1.48 |
| 整形なし | 日→英→英 | 3.49 | 1.50 |
| ルールベース+フィラーの除去 | 日→英 | 3.68 | 1.62 |
| ルールベース+フィラーの除去 | 日→英→英 | 3.69 | 1.67 |

(b) DNN-HMM

| 整形の種類 | 翻訳手法 | BLEU-3 | BLEU-4 |
|---|---|---|---|
| 整形なし | 日→英 | 4.78 | 2.32 |
| 整形なし | 日→英→英 | 4.81 | 2.37 |
| ルールベース+フィラーの除去 | 日→英 | 5.01 | 2.52 |
| ルールベース+フィラーの除去 | 日→英→英 | 5.05 | 2.57 |

### 6.4 翻訳実験結果 - 音声認識結果 -

人手による書き起こしに対しての実験結果で性能の良かった，統計的機械翻訳による語順の並び替えを行った手法を用いて実験を行った．GMM-HMM を用いた際の実験結果を表 10(a) に示す．テストセットは，話者 A のものを使用した．音声認識結果は Corr=0.387,Acc=0.329 である．

整形あり，なしともに並び替えを行うことで BLEU は少し向上したが，人手書き起こしの場合と比較するとあまり改善されていない．理由としては，音声認識誤りが多すぎることによって，うまく並び替えることができなくなるからだと考えられる．

そこで，音声認識性能の良い DNN-HMM を用いて行った実験結果を表 10(b) に示す．テストセットは，話者 A のものを使用した．音声認識結果は Corr=0.498, Acc=0.442

で，認識精度はまだ低い．BLEU 値は向上したが，まだ書き起こしの翻訳結果よりも相当悪い．

## 7. おわりに

新聞のパラレルコーパスよりも TED のパラレルコーパスが講義音声の翻訳に有効であることがわかった．また，パラメータ調整に少量でも同じドメインのコーパスを利用すること良いことも分かった．整形と語順の並び替えを行うことにより，人手書き起こしに対しては BLEU スコアで 0.7 程度，音声認識結果に対しては，0.2 程度向上させることができた．音声認識結果に対して性能があまり向上しなかった理由としては，現状の音声認識結果は，正解率が 50%程度しかないためだと考えられる．今後の課題としては，倒置表現を整形するといったものが残っている．また，本研究で使用した評価尺度 BLEU はグローバルな語順についてはうまく評価できず，日本語・英語の翻訳のような語順が大きく異なる言語間の翻訳では，人間による評価とは相関が低い．RIBES [32] のような評価尺度を用いることにより，本研究の語順の並び替えによる性能の評価を行うことも課題である．

### 参考文献

[1] H. Abelson. The creation of opencourseware at mit. Science Education and Technology, 17:164-74, 2008
[2] Take the world ’s best vourses, online, for free.(coursera.org.). http://www.coursera.org/.
[3] Veri Ferdiansyah. Effect of captioning lecture videos for learning in foreign language. 音声言語情報処理研究会, SLP-97-4, 2013
[4] G. Saon and M. Picheny. Lattice-based viterbi decoding techniques for speech translation. IEEE Workshop on, Automatic Speech Recognition and Understanding, pp. 386-389,2007.
[5] Casacuberta, Francisco. Some approaches to statistical and finite-state speech-to-speech translation, Computer Speech Language 18.1, pp. 25-47, 2004.
[6] Nicola Bertoldi and Marcello Federico. A new decoder for spoken language translation based on confusion networks. IEEE Workshop on Automatic Speech Recognition and Understanding, pp. 86-91, 2005.
[7] H.Schwenk, Marta R.Costa-jussa, and Jose A.R.Fonollosa. Continuous space language models for the IWSLT 2006 Task, IWSLT, pp 166-173. 2006.
[8] 福田智大, 村上仁一, 徳久雅人, 池原悟. ルールベース翻訳を前処理に用いた統計翻訳, 言語処理学会第 16 回年次大会, PB2-12, pp. 676-679, 2010.
[9] 星野翔, 宮尾祐介，須藤克仁，永田昌明. 日英統計的機械翻訳のための述語項構造に基づく事前並べ替え. 言語処理学会第 19 回年次大会，pp. 394-397, 2013.
[10] Hany Hassan, Lee Schwartz, Dilek Hakkani-Tür, Gokhan Tur. Segmentation and Disfluency Removal for Conversational Speech Translation, Segmentation and Disfluency Removal for Conversational Speech Translation, Proc. of Interspeech, pp. 318-322 2014.
[11] Dongdong Zhang, Shuangzhi Wu, Nan Yang, Mu Li. Punctuation Prediction with Transition-based Parsing, ACL (1) pp. 752-760, 2013.
[12] 下岡和也, 南條浩輝, 河原達也. 講演の書き起こしに対する統計的手法を用いた文体の整形. 自然言語処理, Vol. 11, No. 3, pp. 67-83, 2004.
[13] T. Hori, D. Willet, and Y. Minami. Paraphrasing spontaneous speech using weighted finite-state transducers. In Proc. SSPR, pp. 210-222, 2003.
[14] G. Neubig, S. Mori, and T. Kawahara. A WFST-based Log-linear Framework for Speaking-style Transformation. Proc. Interspeech, pp. 1495-1498, 2009.
[15] Fügen, Christian, Kolss, Muntsin, Bernreuther, Dietmar, Paulik, Matthias, Stüker, Sebastian, Vogel, Stephan and Waibel, Alex. Open domain speech recognition & translation: lectures and speeches, Machine Translation 21.4 209-252, 2007.
[16] 堀貴明, 須藤 克仁，塚田 元，中村 篤．世界メディアブラウザ」－音声認識と統計翻訳に基づく多言語動画コンテンツ検索／閲覧システム, 第 2 回音声ドキュメントワークショップ講演論文集, pp.59-66, 2008.
[17] Philipp Koehn, Franz Josef Och, and Daniel Marcu. Statistical phrase-based translation. Proc. the North American Chapter of the Association for Computational Linguistics on Human Language Technology-Volume 1, pp. 48-54, 2003.
[18] P. Brown et al. A statistical approach to machine translation. Computational Linguistics, 16(2), 79-85, 1990.
[19] K. Papineni, et al. Bleu: a method for automatic evaluation of machine translation. Proc. the 40th annual meeting on association for computational linguistics, pp. 311-18., 2002.
[20] 高地 なつめ，磯崎 秀樹. スクランブリングを考慮した和訳の自動評価法の NTCIR-9 データによる検証. 情報処理学会研究報告 NL-219 No.2, 2014.
[21] 藤井 康寿, 山本 一公, 中川 聖一. 文レベル情報と複数仮設を用いた音声認識結果の自動整形の評価, 日本音響学会春季講論集，2-6-10, 2010.
[22] T. Oba, T. Hori, and A. Nakamura. Improved sequential dependency analysis integrating labeling-based sentence boundary detection. IEICE, Vol. E93-D, No. 5, pp. 1272-281, 2010.
[23] 土屋雅稔, 小暮悟, 西崎博光, 太田健吾, 山本一公, 中川聖一. 日本語講義音声コンテンツコーパスの作成と分析. 情報処理学会論文誌, Vol. 50, No. 2, pp. 448-59, 2009.
[24] A. Stolcke, et al. Srilm-an extensible language modeling toolkit. Proc. ICSLP, Vol. 2, pp. 901-904, 2002.
[25] Reinhard Kneser and Hermann Ney. Improved backing-off for m-gram language modeling. ICASSP, Vol. 1, pp. 181-184, 1995.
[26] F.J. Och and H. Ney. A systematic comparison of various statistical alignment models. Computational linguistics, Vol. 29, No. 1, pp. 19-51, 2003.
[27] F.J. Och. Minimum error rate training in statistical machine translation. Proc. 41st Annual Meeting on Association for Computational Linguistics-Volume 1, pp. 160-167, 2003.
[28] Y. Fujii, K. Yamamoto, and S. Nakagawa. Large vocabulary speech recognition system: Spojus++. Proc. MUSP-11, pp. 110-118, 2011.
[29] 前川 喜久雄.『日本語話し言葉コーパス』の概観 ver.1.0, 3, 2004.
[30] 関 博史，山本 一公，中川 聖一．年齢・性別に依存しない DNN-HMM による音声認識法の検討．音声言語情報処理学会研究報告 SLP-104-29，2014.
[31] 福島 太喜, 秋葉 友良，講義音声翻訳における話し言葉の整形と翻訳の同時最適化法の検討，日本音響学会春季研究発表会 1-3-8, 2014.
[32] 平尾 努，磯崎 秀樹，Duh,K.，須藤 克仁，塚田 元，永田昌明．RIBES: 順位相関に基づく翻訳の自動評価法，言語処理学会年次大会，pp. 1115-1118, 2011.