

誤りパターンを考慮した同時音声翻訳のための文の分割法

王夢如^{†1} 浅野優^{†1} 岩山真^{†1} 小林義行^{†1}

概要：講演や生放送では、同時性の高い音声翻訳システムが求められている。これまでに、統計的機械翻訳システムの精度を最大化しつつ翻訳を始めるタイミングを選択する文分割法が提案されている。しかし、既存研究では翻訳しにくい文の誤りパターンを考慮していないため、適切に文が分割されない場合がある。本研究では、翻訳誤りを引き起こした文の文法構造に注目し、分割後のサブセンテンスの翻訳しにくさの予測値を考慮する文分割法を提案する。

キーワード：文分割, 系列パターンマイニング, 統計的機械翻訳

Untranslatable Patterns Sensitive Segmentation Strategies for Simultaneous Speech Translation

MENGRU WANG^{†1} YU ASANO^{†1}
MAKOTO IWAYAMA^{†1} YOSHIYUKI KOBAYASHI^{†1}

1. はじめに

講演, 生放送, テレビ会議の多言語対応や, 公共施設における訪日観光客対応などの場面では, 高い同時性を備えている音声翻訳システムへのニーズが高まっている。従来の機械翻訳は, 一文の発話が終わるまで待ってから翻訳を行うため, 翻訳の同時性が低い。一方, 同時音声翻訳は, 発話を分割し, 発話の途中で翻訳を開始することで遅延時間を削減する。従って, いかに翻訳精度を低下させずに文を細く分割するかが重要となる。文分割の既存手法には, 翻訳精度(BLEU)に基づく文分割手法[1]や句読点を推測する手法[2]等がある。

我々は日中同時音声翻訳システムの実現をめざしている。本システムには様々な活用場面があるが, 本研究では駅での駅員と訪日中国人観光客の会話を対象とする。我々はまず, 対象言語の違いはあるものの, 講演の英日同時翻訳で有効性が示されている[1]の手法を適用した。そこでの翻訳誤りを分析したところ, 文を翻訳しにくい単位(サブセンテンス)に分割してしまう場合が散見された。このような分割誤りを防ぐためには, サブセンテンスごとの翻訳しにくさを予測する必要がある。

本研究では, 翻訳しにくさを翻訳に失敗した文に頻出するパターンの有無とみなし, 誤りパターンを考慮する文の分割法を提案する。具体的には, 既存の文分割手法の目的関数に対し, 文の翻訳しにくさの指標を追加することで評価値を補正する。

本手法の有効性を検証するために, 駅における会話の日

中対訳文コーパスを作成し実験を行った。評価指標の一つとして BLEU を用いた結果, 分割後のサブセンテンスの平均長(形態素数)が 4 と 10 の場合, 有意水準 5% で提案手法が既存手法を上回ることがわかった。

2. 既存手法

分割後の翻訳精度を最大化する分割位置を貪欲法に基づいて選択する手法が提案された[1]。翻訳精度の評価関数として, 表記の類似度に基づく評価指標 BLEU を用い, 動的計画法により分割後の BLEU が最大となる分割位置を決めて行く。分割位置は前後の品詞で指定する。

既存手法は BLEU のみにより文の分割位置を評価し, 翻訳のしにくさは考慮していない。そのため, 文を翻訳しにくいサブセンテンスに分割してしまう可能性がある。例えば, 下記の場合, 名詞と名詞の間で分割した場合の BLEU が最大となる。しかし, この分割により, 「券の変更はできませんが, 乗車」という翻訳しにくいサブセンテンスが出来てしまい, 翻訳結果でも「寝台」と「券」の訳語(それぞれ, 「卧铺」「车票」)が離れてしまう不適切な訳となっている。

【分割結果】寝台/券の変更はできませんが, 乗車/駅の発車/時刻/前になります。

【翻訳結果】卧铺可以变更车票乘车车站发车时间仅限于前。

3. 提案手法.

3.1 概要

2 章で挙げた既存手法の問題点を解決するために, 本研究では, 翻訳しにくさを表す指標を既存の目的関数に加え

^{†1} (株)日立製作所 研究開発グループ
Hitachi Ltd., Research & Development Group

ることで評価値を補正し、分割位置を抽出する。翻訳しにくさを用いて分割位置の評価を補正する例を、図 1 で説明する。各分割位置候補の評価値は、BLEU による翻訳精度から分割後の文の翻訳しにくさ U を減算したものとす。例えば、BLEU の評価値は、名詞と名詞で分割する場合 0.57、助詞と記号で分割する場合 0.42 となり、既存手法だと評価値の大きい前者で分割する。翻訳しにくさ U の値は、前者の場合 0.3、後者の場合 0.14 となり、それらを減算して補正した結果、前者は 0.27、後者は 0.56 となり、提案手法だと後者で分割するようになる。

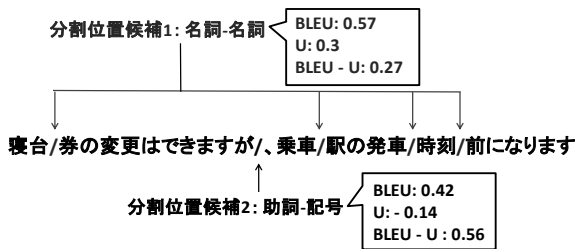


図 1 翻訳しにくさを考慮した文分割の例

提案手法を実現するシステムの全体を、図 2 に示す。まず、翻訳に失敗した文、および、翻訳に成功した文から、翻訳しにくさの予測モデルを学習する。詳細は 3.2 節で説明する。次に、対訳コーパスを用いて分割位置（前後の品詞）を抽出する。その際に、前述の予測モデルを用いて BLEU を補正する。詳細は 3.3 節で説明する。ここで抽出した分割位置を分割条件とみなし、実際の発話を分割する。

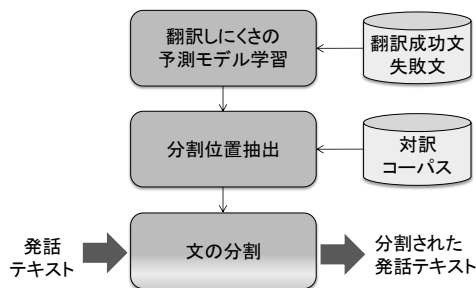


図 2 システムの全体

3.2 翻訳しにくさの予測モデル学習

翻訳しにくさの予測モデルでは、翻訳に失敗した文に頻出する系列パターンを特徴量として用いる。ここでの系列は、文の係り受け関係を考慮した品詞・表記混じりの系列である。係り受け関係に着目した理由は、係り受けによる複雑な構文構造が翻訳しにくさに影響していると考えたためである。また、品詞・表記混じりの系列に着目した理由は、品詞のみの場合に比べて細かい粒度の系列パターンも扱えるためである。上記の系列パターンの抽出方法と、それらを用いた予測モデルの学習手法について順に述べる。

まず、各文に対して形態素解析と係り受け解析を行い、

係り受け関係を考慮した品詞・表記混じりの系列に変換する。例えば、「友達大勢と一緒に乗車したいのですが割引のようなものはありますか」は次のような係り受け構造になる。各形態素は、(表記, 品詞) と表す。名詞は n 、動詞は v 、助動詞は av 、助詞は p と表す。

(友達, n) (大勢, n) (で, p) -D
 (一緒に, n) (に, p) -D
 (乗車, n) (し, v) (たい, av) (の, n) (です, av) (が, p) --D
 (割引, n) (の, p) (よう, n) (な, av) -D |
 (もの, n) (は, p) -D
 (あり, v) (ます, av) (か, p)

次に係り元と係り先の二項関係からなる系列パターンに変換する。上記の例の「友達大勢と一緒に乗車したいのですが」の部分からは、「友達大勢で--乗車したいのですが」と「一緒に--乗車したいのですが」に関する 2 つの係り元と係り先の二項関係を生成する。このような二項関係に頻出する品詞・表記交じりの系列パターンを高速に抽出するために、PrefixSpan (PREFIX-projected Sequential PAttern mining)[5] を拡張して用いた。具体的には、同じ系列と判定するための制約として、従来は表記の一致のみを用いていたのに対し、品詞の一致、係り元、係り先の種別の一致も考慮するように拡張した。これにより、上記の例に関しては、以下のような系列が抽出される。各形態素は、(表記, 品詞, 係り情報) と表す。アンダースコアは任意の要素に相当する。

(_, p , 係り元) (_, n , 係り先)
 (の, _, 係り元) (_, n , 係り先)

これらの系列から、翻訳に失敗した文に出現しやすい系列パターンを選んで予測モデル学習のための特徴量として用いる。本手法では、全ての文での出現頻度と、翻訳が失敗した文での出現頻度の比でソートし、値の大きいものから最大で 30, 50, 70 系列を選んで実験した。選んだ特徴量を用いて One-Class SVM[5] で予測モデルを学習する。学習した SVM の信頼度を、文 f に対する訳しにくさの予測値 $UT(f)$ とする。

3.3 分割位置の抽出

文の分割位置は、前後の形態素の品詞で指定する。分割位置の抽出には、既存手法を拡張したアルゴリズム (Algorithm 1) を用いる。まず、人手で分割後のサブセンテンスの平均長 (形態素数) μ を指定し、式(1)に基づいて訓練データ F 全体における分割回数 K を計算する。動的計画法により、現在の分割数 k に対して、評価値 w を最大化する分割位置の素性集合を求める。最後に分割位置集合 Φ_k を返す。 w は、式(2)に示すように、BLEU に基づく翻訳精

度の評価値 $EV(e_{hyp}, e_{ref})$ と 3.2 節の手法で算出した翻訳しにくさの予測値 $UT(f_n)$ からなる。 e_{hyp} は翻訳結果、 e_{ref} は参照訳、 f_n は n 番目の文である。 $MT(f_n)$ は翻訳結果であり、任意の翻訳エンジンで取得する。 $c(\phi|F)$ は分割位置 ϕ の学習データ F 全体における出現回数である。

$$K := \max\left(0, \left\lfloor \frac{\sum_{f \in F} |f|}{\mu} \right\rfloor - N \right) \quad (1)$$

$$w(\Phi_j \cup \{\phi\}) := \sum_{n=1}^N \frac{EV(MT(f_n | \Phi_j \cup \{\phi\}), e_n)}{-\alpha UT(f_n | \Phi_j \cup \{\phi\})} \quad (2)$$

Algorithm 1

```

 $\Phi_0 \leftarrow \emptyset$ 
for  $k = 1$  to  $K$  do
   $w^* \leftarrow -\infty, \Phi_k \leftarrow nil$ 
  for  $j = 0$  to  $k - 1$  do
    for all  $\phi \in \{\phi | c(\phi|F) = k - j \wedge \phi \notin \Phi_j\}$  do
       $w' \leftarrow w(\Phi_j \cup \{\phi\})$ 
      if  $w' > w^*$  then
         $w^* \leftarrow w', \Phi_k \leftarrow \Phi_j \cup \{\phi\}$ 
      end if
    end for
  end for
end for
return  $\Phi_k$ 
  
```

4. 実験

4.1 実験データ

実験のために日中対訳文コーパス、および、日中機械翻訳結果の評価データを作成した。 駅員と訪日中国人観光客の会話を想定して、中国語と日本語のネイティブスピーカーが 4465 件の中日対訳文対を作成した。 また、日本語の 4465 文を商用の機械翻訳エンジン[a]で中国語に翻訳し、その結果を著者の一人が評価した。 評価は、成功、部分的に成功、失敗の 3 種類とした。 内訳は、失敗が 2664 件、部分的に成功が 807 件、成功が 994 件である。

翻訳しにくさの予測モデル学習には、機械翻訳が失敗した 2664 件と成功した 994 件の計 3658 件を用いた。 分割位置の抽出には、残りから 500 件を選んで用いた。

4.2 実験結果

最初に、翻訳しにくさの予測モデル学習について説明する。 学習に最適な設定は 8-fold 交差検定で決めた。 特徴量を抽出する PrefixSpan では、系列の最大長を 1, 2, 3 に、最小支持率を 0.3, 0.4, 0.5 に設定して実験した。 予測モデル SVM のカーネルは、線形カーネル、多項式カーネル、RBF カーネルを用いた。 本研究では、翻訳しにくさをモデル化したいため、翻訳に失敗した文の再現率が最大になる設定を選択した。 実験の結果、最大長は 2、最小支持率は 0.3、カーネルは RBF カーネルを選んだ。 34 個のパターン

が抽出された。 この予測モデルでは、翻訳に失敗した文の再現率は 0.96、適合率は 0.73、F 値は 0.83 になる。 また、モデルの正解率は、0.71 になる。

次に、分割位置の抽出について説明する。 既存手法と提案手法で文を分割し、その結果を比較した。 実験では、分割後のサブセンテンスの平均長 μ を 4, 6, 8, 10 に設定して 5-fold 交差検定を行い、それぞれで、分割結果の翻訳精度を比較した。 翻訳精度の評価方法には様々な方法があるが、今回は自動評価指標の一つである BLEU を用いた。 提案手法における、 $UT(f_n)$ の重み α は、予備実験により 1.5 とした。

表 1 に結果を示す。 訓練データの BLEU は全ての平均長において、既存手法の方が高い。 これは、既存手法が BLEU を直接最大化しているのに対し、提案手法は BLEU を補正した値を最大化しているためと考えられる。 一方、テストデータの BLEU は、平均長が 6 の場合を除き、提案手法が既存手法を上回った。 差を bootstrap resampling[4]により検定した結果、平均長が 4 と 10 の場合は、有意水準 5% で提案手法が有意に上回ることがわかった。 逆に、いずれの平均長でも既存手法が提案手法を有意に上回る結果は出なかった。

表 1 翻訳精度

	μ	訓練	テスト
既存手法	10	208.74	50.87
	8	209.89	50.62
	6	211.06	50.06
	4	211.48	50.11
提案手法	10	206.15	51.42
	8	204.72	50.82
	6	207.22	50.01
	4	210.27	51.48

5. 考察

本章では、提案手法による(1)翻訳しにくいパターンの抽出、(2)翻訳しにくさによる補正の効果、(3)分割位置、(4)分割後のサブセンテンスの長さについて順に考察する。

まず(1)翻訳しにくいパターンの抽出について考察する。 本実験では、合計 34 個のパターンが抽出された。 その内訳は、品詞のみのパターンが 20 個、品詞と表記の両方からなるパターンが 11 個、表記のみのパターンが 3 個であった。

次に(2)翻訳しにくさによる補正の効果について考察を行う。 表 2 (UT:提案手法, BL:既存手法, TR:翻訳結果, U:翻訳しにくさの予測値) に示すように、既存手法(BL)で分割した文は、#1 の「行けなく」/「なりました」のような動詞のまとまりの間や#3 の「お得」/「な」のような修飾関係にある形態素の間で分割されているため、意味の

a) 株式会社フィートがサービス化した翻訳エンジン

まとまりが壊れてしまい、翻訳しにくさ (U 列の値) が提案手法(UT)に比べ高くなった。この翻訳しにくさを分割位置の評価関数に取り入れることにより、翻訳しにくい文になってしまう分割位置が選択されにくくなる。

表 2 翻訳しにくさによる補正例

#	提案手法 (UT) と既存手法 (BL) による分割と翻訳結果 (TR)	人手評価*	U
1	UT: 行けなくなりました TR:不能去了	○	0.04
	BL:行けなく/なりました TR:不能去经常去	×	0.09
2	UT:○○鉄道では/○○駅から○○駅までの往復乗車券と全席指定の往復特急券がセットになったお得な切符を販売しています TR:○○鉄路的从○○站到○○站的往返票和全座位指定的特快车票往返的优惠套餐出售.	△	0.03
	BL:○○鉄道では/○○駅から○○駅までの往復乗車券と全席指定の往復特急券がセットになったお得/な切符を販売しています TR:○○铁路是从○○站是全席指定的特急券划算组合了往返的车票和○○站的往返票卖票.	×	0.08

※人手評価: ○正しい, ×: 誤っている, △: その他

次に, (3)分割位置各手法により抽出された分割位置に対して考察を行う。既存手法と提案手法で選択された分割位置を表 3 に示す。既存手法と提案手法の両方で抽出された分割位置は「共通」の列に示す。提案手法のみで選択された分割位置には, 助詞-記号が選択された。それに対して, 既存手法のみで選択された分割位置には, 助詞-助動詞, 助詞-連体詞, 助詞-形容詞, 形容詞-助詞, 助動詞-副詞が含まれていた。これらの位置で分割する場合, 意味のまとまりが壊れてしまう可能性が考えられる。

表 3 選択された分割位置 (1 回目の場合)

既存手法のみ	共通	提案手法のみ
助詞-形容詞	助動詞-動詞	助詞-記号
助動詞-記号	記号-接続詞	
助詞-連体詞	接続詞-副詞	
記号-記号	名詞-連体詞	
接続詞-記号	名詞-接続詞	
形容詞-助詞	感動詞-記号	
助詞-助動詞	名詞-接頭詞	
助動詞-副詞	記号-形容詞	
記号-動詞	連体詞-助動詞	
	助動詞-連体詞	
	動詞-名詞	
	感動詞-接続詞	
	副詞-名詞	
	助動詞-接頭詞	

最後に(4)分割後のサブセンテンスの長さについて考察を行う。表 4 に示すように, 提案手法の場合, 分割後のサ

ブセンテンスの長さの分散が小さくなった。その原因は, 動詞-助動詞のような文末によく現れる分割位置が選択されなくなったためであると考えられる。BLEU に基づく既存手法の場合, 表 5 にあるように文末での分割位置が選択されやすい。それは, 文末に動詞と助動詞が現れる場合, 助動詞の前までの文のみでも BLEU が高くなるからである。

表 4 分割後のサブセンテンスの長さ

	μ	μ_{test}	σ_{test}
既存手法	10	9.67	66.75
	8	7.46	40.75
	6	5.87	26.43
	4	3.92	11.73
提案手法	10	9.49	53.30
	8	7.69	19.12
	6	6.02	22.51
	4	3.88	11.59

表 5 既存手法による分割例

1	入力: バスモで入ら/ない.
	結果: 用 PASMO 进来不.
2	入力: 使わなく/なりました.
	結果: 不不用经常去.
3	入力: 乗車料金と乗車券の金額は一致しませ/ん.
	結果: 您的金额乘车费用和车票一致.

6. おわりに

本研究は, 翻訳しにくさを考慮した文の分割法を提案し, それを日本語の発話文の分割に適用した。提案手法の翻訳精度を既存手法と比較した結果, 分割後の平均長を 4, 10 に指定した場合に有意水準 5% で上回った。今後の課題として, 提案手法を統計ベース以外の様々な機械翻訳エンジンに適用し, その有効性を検証することが挙げられる。

参考文献

- [1] 小田悠介, Graham Neubig, 清水宏晃, Sakriani Sakti, 戸田智基, 中村哲, 翻訳精度の最大化による同時音声翻訳のための文分割法, 言語処理学会第 20 回年次大会発表論文集, p.302-305, 2014.
- [2] 村田匡輝, 大野誠寛, 松原茂樹, 読点の用法的分類に基づく自動読点挿入, 研究報告音声言語情報処理 (SLP), vol.2010-SLP-81, no.8, p.1-8, 2010.
- [3] Pei, J., Han, J., Mortazavi-Asl, B. and Pinto, H., PrefixSpan: Mining Sequential Patterns Efficiently by Prefix-Projected Pattern Growth, proceedings of the 17th international conference on data engineering, p. 215-224, 2001.
- [4] Koehn, P., Statistical Significance Tests for Machine Translation Evaluation, EMNLP. 2004, p. 388-395, 2004.
- [5] Schölkopf, B., Platt, J.C., Shawe-Taylor, J. and Smola, A.J., Estimating the support of a high-dimensional distribution, Neural computation, vol.13(7), p.1443-1471, 2001.