英日同時通訳システムのための疑似同時通訳コーパス自動生成手法の提案

二又 航介　須藤 克仁　中村 哲
奈良先端科学技術大学院大学

futamata.kosuke.fg6@is.naist.jp

1　はじめに

異なる言語を話す人々のコミュニケーションを支援する技術として，同時通訳システムの研究開発が行われている [1]．同時通訳システムは，原言語の入力文の終了を待たずに目的言語への訳出を開始する翻訳システムである．同時通訳システムは遅延を抑えつつ正確に部分訳出を行うため，講義や講演等の文が長くなる傾向にある場面において円滑なコミュニケーションを可能とする．また，同時通訳システムの学習に，原言語と近い語順で翻訳された文章で構成される同時通訳コーパスを利用することで，遅延を最小限に留められると期待できる．しかし，現在利用可能な同時通訳コーパスの量は非常に小さいため，大量の同時通訳コーパスを利用して同時通訳システムを学習させることは現実的ではない．本研究では，事前並び替えとスタイル変換によって，対訳コーパスから擬似同時通訳コーパスを自動生成する手法を提案する．実験の結果，文節単位で事前並び替えを適用した文から口語文へスタイル変換を適用することにより，実際の同時通訳文に近い文を生成できることが明らかになった．

2　同時通訳の遅延と順送り訳

英語と日本語のように語順が大きく異なる言語間の同時通訳では，訳出開始までの遅延が大きな問題となる．これは英語が主辞前置型 (head-initial) 言語，日本語が主辞後置型 (head-final) 言語であり，語順が大きく異なることに起因する．図 1 及び図 2 に，それぞれ訳出開始までに大きな遅延が発生する例，遅延が少ない例を示す．

図 1 の英語原文では”A brand-new computer on the desk”と修飾部が長くなっているため，訳出開始までに大きな遅延が発生している．一方で図 2 に示すように原言語の語順に近い形で訳出を行うことで，遅延を少なくすることができる．また図 2 の訳出例は語順の洗練性は無いが，助詞等により致命的な間違いはほとんど無い．このような通訳方略を「順送り」[2] という．

以上の例のように，英日翻訳では「順送り」を用いて訳出を行っても，決定的な問題は生じにくいと言える．これは，日本語が膠着語に分類され，文節における順序の入れ替えを比較的許容する言語であるからである．本研究ではこのような「順送り」方式の疑似同時通訳コーパスを自動生成する手法を提案する．

3　疑似同時通訳コーパスの作成

疑似同時通訳コーパスの作成するため，文節単位の事前並び替えとスタイル変換を適用した．事前並び替えにより，対訳コーパスにおける日本語文を英語原文との語順が近くなるように並び替え，スタイル変換により実際の同時通訳文に近い自然な文へと変換する．

3.1　事前並び替え

事前並び替えは統計的機械翻訳 (SMT) において，語順が大きく異なる言語対の翻訳精度を向上させるために使用される手法である．事前並び替えにより，原言語における文の語順が目的言語における文の語順に近づくように並び替えられる [3]．図 3 に事前並び替えの適用例を示す．事前並び替えにより，英語参照文と並び替え文における単語間の交差が無くなっていることがわかる．

対訳コーパスにおける日本語文と「順送り」の訳文では語順が大きく異なるため，事前並び替えを適用することにより,「順送り」の訳文と語順が類似するようにする．しかし，実際の同時通訳文は図 2 の訳出例に示したようにある程度大きな文節を保つ傾向にある．そこで，対訳コーパスにおける日本語文に文節単位で

英語原文: A brand-new computer on the desk which **my father fave me** on my birthday doesn't work now.

日本語訳文: **父から誕生日に貰った、**机上にある新しいコンピュータは今故障しています。

訳出タイミング: (待ち時間 . . .) 父から誕生日に貰った，机上にある新しい . . .

図 1: 訳出開始までに遅延が発生する例

英語原文: A brand-new computer **on the desk** which my father fave me on my birthday doesn't work now.

日本語訳文: **机上にある**新しいコンピュータですね、これは父から貰ったものです、ですが今故障しています。

訳出タイミング: (待ち時間 . . .) 机上にある新しいコンピュータですね 、これは父から貰ったものです . . .

図 2: 訳出開始までの遅延が少ない例

事前並び替えを適用することにより，極端な並び替えを無くし「順送り」の同時通訳文に近い語順にする．

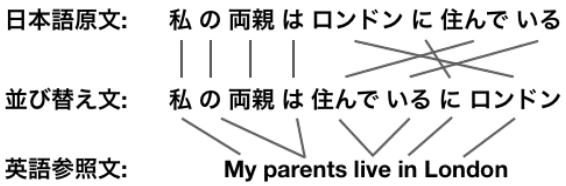

図 3: 事前並び替えの適用例

3.2 スタイル変換

事前並び替えを適用した文は，語順の変化により流暢ではなくなるため疑似同時通訳文としては不十分である．そこで，事前並び替えを適用した文にスタイル変換を適用することにより，流暢な文へと変換する．

スタイル変換とは，言い換え生成の一種である．この技術により，文体 (スタイル) を自動的に制御することができる．ここでスタイル変換の学習に用いるコーパスをそれぞれ $\{D_i\}_{i=1}^{K}$ と定義する．各コーパス D_i は特定のスタイルを含む文から構成される．また，各 D_i に対応するスタイルを $S^{(i)}$ と定義する．このときスタイル変換の目的は，入力文 $\mathbf{x}$ と変換先スタイル $\hat{\mathbf{s}} \in \{\mathbf{s}^{(i)}\}_{i=1}^{K}$ が与えられたとき，入力文 $\mathbf{x}$ の意味情報を保持しつつ，スタイル情報 $\hat{\mathbf{s}}$ を含む文章 $\hat{\mathbf{x}}$ に変更することである．このような変換を可能にする関数 $f_\theta(\mathbf{x}, \mathbf{s})$ を得ることにより，スタイル変換を行う．スタイル変換の手法として本研究では，教師なしニューラル機械翻訳 (Unsupervised Neural Machine Translation: UNMT) による手法 [4, 5] を利用した．

4 実験

提案手法により作成した疑似同時通訳コーパスの品質を評価するため評価実験を行った．UNMT によるスタイル変換の学習アルゴリズムとしては，Lample らの手法 [4] と同様の設定を採用した．

4.1 データセット

文節単位で事前並び替えを適用したコーパスを作成するにあたって，事前並び替えの学習には ASPEC を使用した．ASPEC は中規模のコーパスで，比較的長文かつ専門用語が多く複雑な文章から構成されている．事前並び替えには，Nakagawa[3] の Bracketing Transduction Grammar(BTG) による手法を用いた．

スタイル変換には，表 1 に示す各コーパスのうち，原言語側と目的言語側に異なる 2 つのコーパスを使用した．ASPEC(preordered) は ASPEC に対して文節単位の事前並び替えを適用したものである．TED コーパスは英語で行われた講演を日本語へと翻訳したもの (TED(Caption)) と同時通訳したもの (TED(SI)) から構成されており，TED(SI) は独自に収集したものである．CSJ(日本語話し言葉コーパス) は日本語で行われた講演音声を集めたコーパス，OS(OpenSubtitles) は多言語から構成される映画字幕コーパスである．

4.2 自動評価による実験

自動評価の指標には，スタイル変換の評価尺度として一般的に使用される分類器の正解率 (ACC), Perplexity(PPL)，BLEU に加えて RIBES を使用した．TED(Caption) と TED(SI) を使用して学習させた 2 値分類器の正解率により，疑似同時通訳文が実際の

表 1: 各コーパスの統計情報

Corpus	Number of sentences		
	Train	Val	Test
ASPEC	1000000	1790	1812
ASPEC(preordered)	1000000	1790	1812
TED(Caption)	21239	-	-
TED(SI)	32990	-	-
TED	54229	-	-
CSJ	97687	-	-
OS	693194	-	-
TED(SI)+CSJ	130678	-	-
TED(SI)+CSJ+OS	823872	-	-

同時通訳文らしい文であるか評価する．Perplexity により，疑似同時通訳文が実際の同時通訳文として流暢なものであるかどうか評価する．Perplexity の計測には同時通訳文である TED(SI) を利用して言語モデル (LM) を学習させたものを使用した．表 2 に分類器の精度及び言語モデルの Perplexity を示す．またスタイル変換前後の文ペアの BLEU スコアにより，疑似同時通訳文が意味情報を保持しているかを計測し，疑似同時通訳文と ASPEC(preordered) との RIBES スコアにより,「順送り」方式の翻訳文のような語順になっているか否か計測する．

スタイル変換を適用するコーパス対及び自動評価指標による結果を表 3 に示す．1)，2) はスタイル変換適用前のコーパスを表し，ベースラインとして用いた．3) は対訳文から同時通訳文へのスタイル変換，4)，5) は対訳文から事前並び替え文へのスタイル変換を表す．6) から 8) は事前並び替え文から口語文へのスタイル変換を表し，口語文の量がどのような影響を及ぼすか分析を行った．また疑似同時通訳文は対訳コーパスの日本語文から生成する必要があるため，3) から 5) では ASPEC，6) から 8) では ASPEC(preordered) のテストデータを用いて評価を行った．

表 3 に示す実験の結果，8) のコーパス対において最も同時通訳文らしい傾向がみられた．3) はコーパスサイズが非常に小さいため，全ての評価値が悪かったと推測される．4)，5) は，ACC と PPL の値に関して 2) と大きな差が表れず，2) のような文が生成される傾向がみられた．一方 6) から 8) は，口語文のコーパスサイズが大きくなるに伴い ACC，RIBES スコアが向上したが，PPL に関しては悪化する傾向が見られた．これは，口語文のコーパスとして TED(SI) だけでなく，様々なドメインに及ぶ口語文を使用したからであると推測される．

以上の結果から，文節単位で事前並び替えを適用した文から口語文へスタイル変換を行うことにより，同時通訳らしい文を生成できることが明らかになった．また，口語文の量とコーパスのドメインが，疑似同時通訳文の品質を大きく左右することも明らかになった．

表 2: 分類器の精度及び言語モデルの Perplexity

Classifier		LM
Precision	Recall	PPL
97.43	97.43	55.00

表 3: 疑似同時通訳文の自動評価結果

	Corpus pair	ACC	BLEU	PPL	RIBES
1)	ASPEC	38.70	-	2995.06	-
2)	ASPEC(preordered)	60.08	-	3680.84	-
3)	TED(Caption) → TED(SI)	0.05	1.32	37.03	43.07
4)	ASPEC → ASPEC(preordered)	59.56	72.21	3645.41	80.07
5)	ASPEC+TED(Caption) → ASPEC(preordered)+TED(SI)	63.46	**74.05**	3738.11	82.56
6)	ASPEC(preordered) → TED(SI)	0.05	3.27	**35.30**	59.23
7)	ASPEC(preordered) → TED(SI)+CSJ	35.10	30.25	450.95	83.94
8)	**ASPEC(preordered) → TED(SI)+CSJ+OS**	**76.34**	51.10	934.32	**91.33**

4.3 人手評価による実験

疑似同時通訳文は英語参照文と同様の語順であることが望ましいため，ASPEC(preordered) との RIBES スコアを計測するだけでは不十分である．そこで，人手評価により擬似同時通訳文の品質を計測した．人手評価には最も同時通訳文らしい傾向が見られた 8) の結果を対象とした．

人手評価には 5 人の被験者に，英語原文，日本語原文，疑似同時通訳文を提示し，”英語原文と疑似同時通訳文の単語対応関係が一致しているかどうか (Word-order)”，”疑似同時通訳文は日本語として自然であるかどうか (Fluency)”，”疑似同時通訳文は口語的であるかどうか (Colloquial)”，”日本語原文と疑似同時通訳文が意味的に同一であるかどうか (Identity)” の 4 つの指標を満たすか否か 5 段階で，合計 30 文評価してもらった．

人手評価による実験結果を表 5 に示す．’Word-order’，’Fluency’，’Colloquial’，’Identity’ のそれぞ

表 4: 生成された疑似同時通訳文例

Example(1)	(Word-order=4.2, Fluency=4.4, Colloquial=4.6, Identity=4.6)
ASPEC(EN)	As one method for judging the pain mechanism, drug challenged test (DCT) is described.
ASPEC(JA)	痛みの機序を判定する 1 つの方法として，薬理学的疼痛機序判別試験（ＤＣＴ）について述べた。
ASPEC(preordered)	1 つの方法として，痛みの機序を判定する薬理学的疼痛機序判別試験 (ＤＣＴ) について述べた
ASPEC(style-transferred)	一つの方法として，痛みの機序を判定するための，薬理学的疼痛機序判別試験（ＤＣＴ）について説明していきます 。

Example(2)	(Word-order=2.6, Fluency=2.6, Colloquial=2.4. Identity=2.4)
ASPEC(EN)	Therefore, an approach to the problem from the art side by participation of artists was deceived, and the cooperative research was carried out by participation of an artist.
ASPEC(JA)	このため，アーティストが参加してアート側から問題にアプローチすることを考え，アーティストを加えて共同研究を行った。
ASPEC(preordered)	このため, 考え, アーティストをことをアプローチする問題にアーティストが参加してアート側から共同研究を行った, 加えて。
ASPEC(style-transferred)	このため, アーティストを作ることにアプローチすることができますが,問題にアーティストが参加してアート側から共同研究を加えています 。

表 5: 人手評価による実験結果の平均値

	Num samples	Word-order	Fluency	Colloquial	Identity
All samples	30	3.59	3.49	3.53	3.48
Word-order $\geq$ 3.0	25	3.78	3.66	3.46	3.80
Fluency $\geq$ 3.0	21	3.82	3.92	3.77	3.90
Colloquial $\geq$ 3.0	19	3.86	3.99	3.91	3.96
Identity $\geq$ 3.0	22	3.88	3.84	3.69	4.03

れの値は被験者 5 人の平均値を表す．また'All samples' は全 30 サンプルの平均値を表し，'Word-order $\geq$ 3.0，Fluency $\geq$ 3.0，Colloquial $\geq$ 3.0，Identity $\geq$ 3.0 はそれぞれの値において平均値が 3.0 以上であったサンプルを表す．

表 5 に示すように，全サンプルから計算された平均値は全ての評価指標において 3.0 を超えていた．また，それぞれの基準において平均値が 3.0 を超えたサンプル数は 20 前後であり，多くの疑似同時通訳文が実際の同時通訳文に現れる特徴を持っていると言える．

表 4 に疑似同時通訳文の例を示す．表 4 における Exampe(1) は各評価指標の平均値が高かったサンプル，Example(2) は．平均値が低かったサンプルである．Example(1) では ASPEC(EN) と ASPEC(preordered) の単語対応関係が類似しているものの，Example(2) では大きく異なっている．その結果 Example(1) では'Word-order' の評価値が高く．Example(2) では低かったと推測される．したがって，事前並び替えの精度を向上させることにより実際の同時通訳文により近い文を生成できると考えられる．

5　おわりに

本稿では，事前並び替えとスタイル変換により対訳コーパスから擬似同時通訳コーパスを自動生成する手法について提案した．実験の結果，文節単位で事前並び替えを適用した文から口語文へスタイル変換を適用することにより,「順送り」の同時通訳文に類似した文を生成することが可能であると明らかになった．しかし，事前並び替えの結果に誤りが含まれると疑似同時通訳文も英語原文の語順と異なる傾向がみられた．

今後の展望として，事前並び替えを適用した文がより同時通訳文らしい語順となるように事前並び替えの手法を改善する．また疑似同時通訳コーパスを用いて機械翻訳システムの学習を行うことで，実際の同時通訳文に対する翻訳精度の向上を目指す．

謝辞　本研究は JSPS 科研費 JP17H06101 の助成を受けたものである．

参考文献

[1] Tomiki Fujita, Graham Neubig, Sakriani Sakti, Tomoki Toda, and Satoshi Nakamura. Simple, Lexicalized Choice of Translation Timing for Simultaneous Speech Translation. In *Proceedings of Interspeech*, pp. 3487–3491, 2013.

[2] 水野的. 同時通訳の理論—認知的制約と訳出方略. 朝日出版社, 2015.

[3] Tetsuji Nakagawa. Efficient top-down BTG parsing for machine translation preordering. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pp. 208–218, Beijing, China, July 2015. Association for Computational Linguistics.

[4] Guillaume Lample, Myle Ott, Alexis Conneau, Ludovic Denoyer, and Marc'Aurelio Ranzato. Phrase-based & neural unsupervised machine translation. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pp. 5039–5049, Brussels, Belgium, October-November 2018. Association for Computational Linguistics.

[5] Sandeep Subramanian, Guillaume Lample, Eric Michael Smith, Ludovic Denoyer, Marc'Aurelio Ranzato, and Y-Lan Boureau. Multiple-attribute text style transfer. *CoRR*, Vol. abs/1811.00552, , 2018.