

順送り訳データに基づく英日同時機械翻訳の評価

土肥 康輔¹ 胡 尤佳¹ 蒔苗 茉那¹ 須藤 克仁^{1,2} 中村 哲^{1,3} 渡辺 太郎¹

¹ 奈良先端科学技術大学院大学 ² 奈良女子大学

³ The Chinese University of Hong Kong, Shenzhen

{doi.kosuke.de8, ko.yuka.kp2, makinae.mana.mh2, sudoh, s-nakamura, taro}@is.naist.jp

概要

同時通訳では、原発話の語順を極力維持して訳出することで遅延を抑制する、順送り方略がよく用いられるが、オフライン翻訳では、流畅さを優先して原発話と語順が大きく異なる訳出がされることがある。オフライン翻訳を参照訳とする従来の評価手法では、同時機械翻訳モデルが出力する同時通訳らしい語順の訳出を十分に評価できていない可能性がある。そこで本研究では、語順差が大きい英語・日本語間のオフライン機械翻訳、および同時機械翻訳モデルを順送り訳データを用いて評価した。順送り訳データで同時機械翻訳モデルを評価すると、オフラインデータセットで評価したときに比べて高いスコアとなり、同時機械翻訳の自動評価で語順を考慮することの必要性が示唆された。

1 はじめに

同時通訳 (simultaneous interpreting; SI) とは、原発話の終了を待たずにリアルタイムに翻訳していくタスクである。SI は時間的制約が厳しいだけでなく、聞く・理解する・翻訳する・話すといった複数のタスクを同時に行う必要があるため、その遂行には高い認知的負荷がかかる。そのため通訳者は、発話をより短い単位 (チャンク) に区切りながら、それらのチャンク順に訳出していく順送り方略や、発話全体を理解する上で無くても支障が小さい箇所を省略・要約して訳出する短縮方略など、様々な方略を用いることで、認知的負荷を抑制しつつリアルタイムな訳出を実現している [1]。その結果、同時通訳文はオフライン翻訳文と異なる特徴を持つことが知られているが、利用可能な同時通訳文データが限られているため、同時機械翻訳モデルの学習と評価では主にオフライン翻訳コーパス (例: MuST-C [2]) が用いられている [3, 4, 5]。

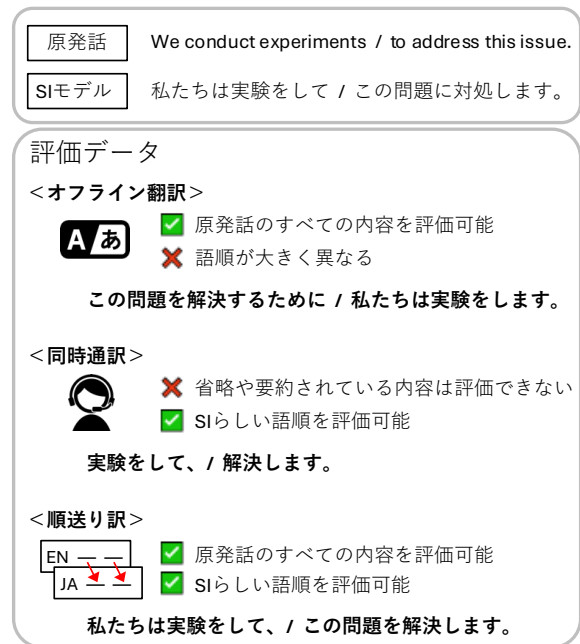


図 1 データによる翻訳品質の評価の違い

この課題に対して、英語・日本語間では複数の同時通訳コーパスが構築されている [6, 7, 8]。大規模な同時通訳コーパスである NAIST-SIC[8] のデータに対して、文単位の対応づけを自動で行なった NAIST-SIC-Aligned[9] が提案されたことで、同時機械翻訳モデルの学習に同時通訳データを用いることが可能となった。[10, 9] は実際に同時通訳データを用いて同時機械翻訳モデルを学習・評価している。

NAIST-SIC には、プロの同時通訳者が実際に同時通訳を行なったデータが収録されているが、実際の同時通訳データを同時機械翻訳モデルの評価に用いることは、モデル性能を過小評価してしまう可能性がある。同時通訳者は要約や省略といった短縮方略を用いるため、発話全体の理解に影響を及ぼさない語句は必ずしも翻訳されないことがある。図 1 の例では、同時機械翻訳モデルは “to address this issue” を “この問題に対処します” と翻訳しているのに対し

て、同時通訳者は「解決します」と翻訳している。このように、同時通訳者が訳出しなかった語句を同時機械翻訳モデルが訳出した場合、モデルが「正しい翻訳」を出力していたとしても適切に評価することができない。

この課題に対して、原発話の内容の要約や省略を含まず、順送り方略のみを適用して作成された、順送り訳評価データ [11] が提案された。[11] では、構築されたデータの品質を人手評価を通して検証していたが、同データを用いて同時機械翻訳モデルを評価する実験は行われていない。そこで本研究では、英日オフライン機械翻訳モデルと同時機械翻訳モデルの出力を順送り評価データを用いて評価し、既存のオフライン翻訳ベースの評価データ、および同時通訳ベースの評価データを用いた評価との比較を行った。順送り訳データを参照訳に用いると同時機械翻訳モデルの評価が高くなるのに対して、オフラインデータを参照訳に用いるとオフライン機械翻訳モデルの評価が高くなり、同時機械翻訳モデルの評価で語順を考慮することの必要性が明らかになった。また、同時通訳データを参照訳に用いると、モデル性能を過小評価してしまう恐れがあることが示唆された。

2 関連研究

2.1 同時通訳コーパス

同時通訳コーパスは、同時機械翻訳モデルの開発だけでなく、同時通訳の特徴を分析するのに有用な言語資源である。オフライン翻訳コーパスと比べるとデータ量が限られているが、いくつかの英日同時通訳コーパスが公開されている [6, 7, 8]。これらの同時通訳コーパスを用いて、訳出を遅延、品質、語順の観点から分析したり [12, 8]、同時通訳者が用いている方略や訳出パターンを明らかにする研究が行われている [13]。また、同時通訳データを用いた同時機械翻訳モデルも構築されている [10, 9]。

2.2 同時通訳における語順

英語と日本語のような、統語構造が大きく異なる言語対において、順送り訳データを用いることで同時機械翻訳モデルを構築する研究が行われている。原発話をチャンクに区切り、そのチャンクごとに翻訳を付与することで順送り訳データを作成する方法が提案されている [14]。また、原言語文や目的言語

文の文法規則に着目し、ルールに基づいて文を書き換えたり [15]、語を並び替えたりすることで [16]、言語間の語順差を小さくする手法が提案されている。[17] は、原言語および目的言語文中のチャンク境界情報に加えて、目的言語文中で省略可能な語句の情報を付与した GCP 同時通訳コーパスを構築した。The NAIST English-to-Japanese Chunk-wise Monotonic Translation Evaluation Dataset (NAIST-CMT-ED) [11] は、同様にチャンク境界情報を含む順送り訳データセットであるが、評価を目的とした比較的小規模なものである。

同時通訳と他の通訳モードやオフライン翻訳との間での語順の違いを調査した研究もあり、同時通訳では、原言語での語順により近い翻訳が行われていることが確かめられている [18]。また、[12] は同時通訳コーパスを分析することで、通訳者が訳出する語順を決める上で影響のある要因を明らかにした。

3 順送り訳

統語構造の差が大きい言語対では、同時通訳者は原発話をチャンクに分割し、チャンクの順番通りに訳出していくことで順送り訳を行っている [18]。そこで本研究では、チャンクごとの順送り訳データである NAIST-CMT-ED を用いる。

NAIST-CMT-ED は 511 文対から成る英日順送り訳データで、そのチャンク境界は [18] で提案された同時通訳者の方略に基づいている。原発話 (英語) は [10] の評価セットと同一で、TED talks の 8 つの講演の一部である。翻訳はチャンク分割された原発話の書き起こしをもとに、後ろのチャンクの情報を含めずに文頭からチャンクの順番どおりに訳出されている。ただし、文の流暢さを保つために、前のチャンクの情報が繰り返し訳出されていたり、翻訳が後ろのチャンクに先送りされている場合もある。

4 実験

順送り訳データを翻訳品質の評価に用いることの影響を検証するため、同時機械翻訳モデルとオフライン機械翻訳モデルを、NAIST-CMT-ED、同時通訳データ、オフライン翻訳データのそれぞれで評価する実験を行なった。

4.1 データ

本研究では、以下の 4 種類の評価セットを用いた。

- **n-cmt** NAIST-CMT-ED に収録されている順送り

表 1 評価に用いたデータセットの語数

Dataset	Sum	Per Sent.±SD
NAIST-CMT-ED	13,508	28.38±18.66
NAIST-SIC	8,914	18.73±12.08
NAIST-SIC-Aligned	8,072	16.96±11.52
Offline	9,907	20.81±12.62

訳文

- **si_hum** NAIST-SIC に収録されている同時通訳文、人手で原発話文と対応づけ
- **si_auto** NAIST-SIC-aligned に収録されている同時通訳文、自動で原発話文と対応づけ
- **offline** TED talks の字幕データに基づくオフライン訳文

評価に用いたデータセットの語数を表 1 に示す。si_auto は自動手法で原発話文との対応づけが行われているため、誤りを含んでいる可能性がある。また、自動手法に含まれるフィルタリング法の影響で、si_auto は si_hum と比べて訳文が短い傾向がある(表 1)。そのため、本研究では si_auto に加えて、新たに人手で対応づけを行なった si_hum も用いた。

4.2 音声翻訳モデル

音声翻訳モデルには、既存研究から以下の 3 種類を用いた。

- **ST_offline** オフライン翻訳データで学習した音声翻訳モデル [5]
- **simulST_offline** オフライン翻訳データで学習した同時音声翻訳モデル [10]
- **simulST_si_offline** 同時通訳データとオフライン翻訳データの両方で学習した同時音声翻訳モデル [10]

全てのモデルは、エンコーダに HuBERT-Large [19]、デコーダーに mBART50 [20] を用いており、原発話の音声を入力とし、対応する翻訳をテキストで出力する。音声翻訳モデル(ST_offline)は原発話の終了を待って翻訳を生成するのに対して、同時音声翻訳モデル(simulST_offline と simulST_si_offline)は原発話の途中で翻訳を生成し始めるモデルである。エンコーダーとデコーダーは inter-connection [21] と length adapter [22] によって結合されている。2 つの同時音声翻訳モデルでは、Bilingual Prefix Alignment [4] を用いてモデルを学習しており、デコーディングポリシーには local agreement [3] を用い

た。音声翻訳モデルは、checkpoint averaging を行なったモデルを使用した([5] 中の Inter-connection + Ckpt Ave. に対応)。同時音声翻訳モデルは、IWSLT2023 Evaluation Campaign¹⁾の simultaneous track の規定を満たすモデルを使用した([10] 中の Offline FT と Mixed FT + Style に対応)。

4.3 評価指標

翻訳品質の評価には、BLEU²⁾ [24]、BLEURT [25]、COMET [26]、BERTScore [27] を用いた。BERTScore は bert-base-multilingual-case を用いてスコアを算出した。4.1 節に示した 4 種類のデータを参照訳とし、これらの評価指標のスコアを算出した。

4.4 実験結果

表 2 は、音声翻訳モデルと同時音声翻訳モデルの翻訳品質の評価結果を示している。BLEU では ST_offline をベースラインに指定し、paired bootstrap resampling [28] を用いてスコアに統計的に有意な差があるかを検定した。それ以外の評価指標では、一元配置分散分析で検定し、チューキーの方法で多重比較を行なった。

n-cmt を参照訳として BLEU で評価すると、simulST_si_offline が最も高いスコアとなり、同時通訳ベースの評価データ(si_hum と si_auto)で評価したときも同様の結果となった。一方で、オフライン翻訳ベースの評価セットである offline を参照訳とすると、オフラインデータのみで学習したモデルのスコアが高くなった。これと同様の傾向が BLEURT と BERTScore の結果においても確認された。この結果は、同時通訳データとオフライン翻訳データの両方で学習した simulST_si_offline が、より同時通訳らしい訳出をしていることを示唆しており、そのようなモデルは同時通訳の特徴を備えた評価データを参照訳として評価する必要があることを示している。また、従来行われている、オフライン翻訳データを参照訳とする評価では、同時通訳データを用いて学習したモデルの性能を過小評価してしまう可能性があることを示唆している。

ここで n-cmt, si_hum, si_auto の結果を比較すると、評価スコアは n-cmt を参照訳としたときに最も高く、si_hum, si_auto の順となっている。si_hum は人間の同時通訳者が実際に同時通訳を行なった

1) <https://iwslt.org/2023/simultaneous>

2) BLEU は sacreBLEU [23] を用いて算出した。

表 2 翻訳品質評価の結果. †: ST_offline と有意差あり. *1: 他の 2 モデルより有意に高い. *2 他の 2 モデルより有意に低い. *3 ST_offline より有意に低い. 有意水準: $p < .05$.

Model	BLEU				BLEURT				COMET			
	n-cmt	si_hum	si_auto	offline	n-cmt	si_hum	si_auto	offline	n-cmt	si_hum	si_auto	offline
ST_offline	14.487	8.856	8.637	17.775	0.553	0.447	0.414	0.538	0.838	0.797	0.781 ^{*1}	0.833
simulST_offline	15.406 [†]	8.446 [†]	7.773 [†]	17.907	0.556	0.442	0.406	0.531	0.826	0.780	0.763	0.821
simulST_si_offline	15.982 [†]	12.031 [†]	11.020 [†]	13.191 [†]	0.567	0.493 ^{*1}	0.460 ^{*1}	0.519	0.807 ^{*2}	0.774 ^{*3}	0.761	0.789 ^{*2}

Model	BERTScore (Pre.)				BERTScore (Rec.)				BERTScore (F1)			
	n-cmt	si_hum	si_auto	offline	n-cmt	si_hum	si_auto	offline	n-cmt	si_hum	si_auto	offline
ST_offline	0.801	0.735	0.722	0.789	0.769	0.739	0.735	0.788	0.784	0.737	0.728	0.788
simulST_offline	0.799	0.730	0.717	0.783	0.770	0.738	0.734	0.786	0.783	0.734	0.725	0.784
simulST_si_offline	0.817 ^{*1}	0.764 ^{*1}	0.746 ^{*1}	0.759 ^{*2}	0.784 ^{*1}	0.766 ^{*1}	0.760 ^{*1}	0.757 ^{*2}	0.800 ^{*1}	0.764 ^{*1}	0.752 ^{*1}	0.757 ^{*2}

データに基づいているため、原発話中の内容が省略されていたり、十分に訳出されていない文が含まれている。si_auto はそのような人間の同時通訳データを自動手法で対応づけ、フィルタリングしているため、si_hum よりも原発話の内容が欠落したデータとなっていると考えられる。実際に、BERTScoreでの評価結果では、n-cmt を用いると precision が recall よりも高くなったのに対して、si_auto を用いると recall が precision よりも高くなり、si_hum を用いると precision と recall はほぼ同じ値となった。この結果は、実際の同時通訳データを参照訳に用いると、モデル性能を過小評価してしまう恐れがあることを示している。

しかしながら、COMET を用いた評価結果は、他の評価指標での結果と異なる傾向となった。COMET で評価したときは、オフラインの音声翻訳モデルである ST_offline が全ての評価セットにおいて最も高いスコアとなった。ついでスコアが高かったのは、オフラインデータのみで学習した同時音声翻訳モデルである SimulST_offline であり、学習時にオフラインデータのみを使用している 2 つのモデルが高いスコアを得るという結果となった。これは、COMET が他の評価指標と異なり、原発話文も評価に用いていることが影響している可能性がある。

原発話文の影響を検証するために、参照訳を用いない COMET-QE [29] を算出したところ、COMET を用いたときと同様の結果となった: ST_offline = 0.831, simulST_offline = 0.798, simulST_si_offline = 0.766。加えて、n-cmt と offline をオラクルデータとみなして COMET-QE を算出したところ、n-cmt は offline よりも高いスコアとなった (n-cmt = 0.832 vs. offline = 0.812)。offline の一部には、訳抜けがあるデータが含まれていることを踏まえると、COMET スコアは原発話文中の内容が翻訳文中によ

り多く含まれているときに高いスコアとなる傾向があることが示唆される。同時通訳においては順送り方略等の様々な方略が用いられているが、COMET のこの傾向は同時通訳文が持つ特徴と相性が悪く、同時機械翻訳モデルを COMET で評価した結果は、慎重に解釈する必要があることを示唆している。

5 おわりに

本研究では、オフラインの音声翻訳モデルと、同時音声翻訳モデルの翻訳品質を、順送り訳評価データと既存の評価データを用いて評価し、結果を比較した。BLEU, BLEURT, BERTScore での結果は、同時通訳データを用いて学習された同時機械翻訳モデルを評価するには、順送り訳データを用いることが必要であることを支持するものであった。しかし、COMET での評価結果はこれに反するもので、他の様々な評価指標や、別の同時機械翻訳モデルを用いたさらなる検証が必要である。

また、本研究では順送り訳データを評価に用いることの効果を検証したが、順送り訳データを同時機械翻訳モデルの学習に使うこと [30] の効果を検証することは今後の課題である。人間の通訳者による実際の同時通訳データと比べて、大幅な省略や要約が含まれていない順送り訳データを用いることで、[10] で問題となっていた訳抜けの問題が軽減される可能性がある。加えて、同時通訳文よりもより多くの情報を含む順送り訳が、通訳の聞き手や読み手に対してどのような影響を与えるのか、認知的負荷の面から分析することも今後の課題とする。

謝辞

本研究の一部は JSPS 科研費 JP21H05054, JST 次世代研究者挑戦的研究プログラム JPMJSP2140 の助成を受けたものである。

参考文献

- [1] He He, Jordan Boyd-Graber, and Hal Daumé III. Interpretation vs. translation: The uniqueness of human strategies in simultaneous interpretation. In **Proc. of NAACL**, pp. 971–976, 2016.
- [2] Mattia A. Di Gangi, Roldano Cattoni, Luisa Bentivogli, Matteo Negri, and Marco Turchi. MuST-C: a Multilingual Speech Translation Corpus. In **Proc. of NAACL**, pp. 2012–2017, 2019.
- [3] Danni Liu, Gerasimos Spanakis, and Jan Niehues. Low-Latency Sequence-to-Sequence Speech Recognition and Translation by Partial Hypothesis Selection. In **Proc. of Interspeech 2020**, pp. 3620–3624, 2020.
- [4] Yasumasa Kano, Katsuhito Sudoh, and Satoshi Nakamura. Simultaneous neural machine translation with prefix alignment. In **Proc. of IWSLT**, pp. 22–31, 2022.
- [5] Ryo Fukuda, Yuta Nishikawa, Yasumasa Kano, Yuka Ko, Tomoya Yanagita, Kosuke Doi, Mana Makinae, Sakriani Sakti, Katsuhito Sudoh, and Satoshi Nakamura. NAIST simultaneous speech-to-speech translation system for IWSLT 2023. In **Proc. of IWSLT**, pp. 330–340, 2023.
- [6] Hitomi Toyama, Shigeki Matsubara, Koichiro Ryu, Nobuo Kawaguchi, and Yasuyoshi Inagaki. CIAIR Simultaneous Interpretation Corpus. In **Proc. of Oriental COCOSA**, 2004.
- [7] 松下佳世, 山田優, 石塚浩之. 英日・日英通訳データベース (jnpc コーパス) の概要. 通訳翻訳研究への招待, Vol. 22, pp. 87–94, 2020.
- [8] Kosuke Doi, Katsuhito Sudoh, and Satoshi Nakamura. Large-scale English-Japanese simultaneous interpretation corpus: Construction and analyses with sentence-aligned data. In **Proc. of IWSLT**, pp. 226–235, 2021.
- [9] Jinming Zhao, Yuka Ko, Kosuke Doi, Ryo Fukuda, Katsuhito Sudoh, and Satoshi Nakamura. NAIST-SIC-aligned: An aligned English-Japanese simultaneous interpretation corpus. In **Proc. of LREC-COLING**, pp. 12046–12052, 2024.
- [10] Yuka Ko, Ryo Fukuda, Yuta Nishikawa, Yasumasa Kano, Katsuhito Sudoh, and Satoshi Nakamura. Tagged end-to-end simultaneous speech translation training using simultaneous interpretation data. In **Proc. of IWSLT**, pp. 363–375, 2023.
- [11] 福田りょう, 土肥康輔, 須藤克仁, 中村哲. 原発話に忠実な英日同時機械翻訳の実現に向けた順送り訳評価データ作成. 研究報告自然言語処理 (NL), 2024-NL-259(14), pp. 1–6, 2024.
- [12] Zhongxi Cai, Koichiro Ryu, and Shigeki Matsubara. What affects the word order of target language in simultaneous interpretation. In **Proc. of IALP**, pp. 135–140, 2020.
- [13] Hitomi Tohyama and Shigeki Matsubara. Collection of simultaneous interpreting patterns by using bilingual spoken monologue corpus. In **Proc. of LREC**, 2006.
- [14] 中林明子, 加藤恒昭. 同時機械翻訳のための文脈を考慮したセグメントコーパス. 言語処理学会第 27 回年次大会発表論文集, pp. 1659–1663, 2021.
- [15] He He, Alvin Grissom II, John Morgan, Jordan Boyd-Graber, and Hal Daumé III. Syntax-based rewriting for simultaneous machine translation. In **Proc. of EMNLP**, pp. 55–64, 2015.
- [16] Hyojung Han, Seokchan Ahn, Yoonjung Choi, Insoo Chung, Sangha Kim, and Kyunghyun Cho. Monotonic simultaneous translation with chunk-wise reordering and refinement. In **Proc. of WMT**, pp. 1110–1123, 2021.
- [17] 東山翔平, 今村賢治, 内山将夫, 隅田英一郎. Gcp 同時通訳コーパスの構築. 言語処理学会第 29 回年次大会発表論文集, pp. 1405–1410, 2023.
- [18] 岡村ゆうき, 山田優. 「順送り訳」の規範と模範 同時通訳を模範とした教育論の試論. 石塚浩之 (編), 英日通訳翻訳における語順処理 順送り訳の歴史・理論・実践, pp. 217–250. ひつじ書房, 2023.
- [19] Wei-Ning Hsu, Benjamin Bolte, Yao-Hung Hubert Tsai, Kushal Lakhotia, Ruslan Salakhutdinov, and Abdelrahman Mohamed. Hubert: Self-supervised speech representation learning by masked prediction of hidden units. **IEEE/ACM Transactions on Audio, Speech, and Language Processing**, Vol. 29, pp. 3451–3460, 2021.
- [20] Yuqing Tang, Chau Tran, Xian Li, Peng-Jen Chen, Naman Goyal, Vishrav Chaudhary, Jiatao Gu, and Angela Fan. Multilingual translation with extensible multilingual pretraining and finetuning. **arXiv**, 2020.
- [21] Yuta Nishikawa and Satoshi Nakamura. Inter-connection: Effective Connection between Pre-trained Encoder and Decoder for Speech Translation. In **Proc. of INTERSPEECH 2023**, pp. 2193–2197, 2023.
- [22] Ioannis Tsiamas, Gerard I. Gállego, Carlos Escolano, José Fonollosa, and Marta R. Costa-jussà. Pretrained speech encoders and efficient fine-tuning methods for speech translation: UPC at IWSLT 2022. In **Proc. of IWSLT**, pp. 265–276, 2022.
- [23] Matt Post. A call for clarity in reporting BLEU scores. In **Proc of the Third Conference on Machine Translation: Research Papers**, pp. 186–191, October 2018.
- [24] Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. Bleu: a method for automatic evaluation of machine translation. In **Proc. of ACL**, pp. 311–318, 2002.
- [25] Thibault Sellam, Dipanjan Das, and Ankur Parikh. BLEURT: Learning robust metrics for text generation. In **Proc. of ACL**, pp. 7881–7892, 2020.
- [26] Ricardo Rei, Craig Stewart, Ana C Farinha, and Alon Lavie. COMET: A neural framework for MT evaluation. In **Proc. of EMNLP**, pp. 2685–2702, 2020.
- [27] Tianyi Zhang, Varsha Kishore, Felix Wu, Kilian Q. Weinberger, and Yoav Artzi. BERTscore: Evaluating text generation with bert. In **Proc. of ICLR**, 2020.
- [28] Philipp Koehn. Statistical significance tests for machine translation evaluation. In **Proc. of EMNLP**, pp. 388–395, 2004.
- [29] Ricardo Rei, Ana C Farinha, Chrysoula Zerva, Daan van Stigt, Craig Stewart, Pedro Ramos, Taisiya Glushkova, André F. T. Martins, and Alon Lavie. Are references really needed? unbabel-IST 2021 submission for the metrics shared task. In **Proc. of WMT**, pp. 1030–1040, 2021.
- [30] Yusuke Sakai, Mana Makinae, Hidetaka Kamigaito, and Taro Watanabe. Simultaneous Interpretation Corpus Construction by Large Language Models in Distant Language Pair. **arXiv**, 2024. arXiv:2404.12299.

A 順送り訳データの概要

表 3 に同一の原発話文に対するオフライン翻訳 (offline), 同時通訳 (SI), 順送り訳 (CMT) の例を示す。オフライン翻訳は TED talks の字幕データ, 同時通訳は NAIST-SIC, 順送り訳は NAIST-CMT-ED にそれぞれ収録されていたものである。オフライン翻訳ではチャンクの順序が入れ替わっている箇所があるのに対して, 同時通訳と順送り訳ではチャンクの順番通りに訳出されている。また, 同時通訳では訳出されていないチャンクが存在している。

表 3 オフライン翻訳, 同時通訳, 順送り訳の比較. “/” はチャンク境界を示し, 原発話文の各チャンクの前に付いている番号は出現順序を表している. 各目的言語文のチャンクの前の番号は, 原言語文での番号と対応している.

Source	(1) The US Secret Service, / (2) two months ago, / (3) froze the Swiss bank account / (4) of Mr. Sam Jain right here, / (5) and that bank account / (6) had 14.9 million US dollars in it / (7) when it was frozen.
Offline	(1) 米国のシークレットサービスは / (2) 2 ヶ月前に / (4) サム・ジェイン氏の / (3) スイス銀行口座を凍結しました / (5) その口座には / (6) 米ドルで 1490 万ドルありました [The US Secret Service / two months ago / Mr. Sam Jain's / froze the Swiss bank account / that bank account / had 14.9 million US dollars]
SI	(1) アメリカのシークレットサービスが, / (3) スイスの銀行の口座を凍結しました。 / (4) サムジェインのものです。 / (5) この銀行口座の中には, / (6) 一千四百九十万ドルが入っていました。 [The US Secret Service / froze the Swiss bank account / it is Sam Jain's one / in this bank account / had 14.9 million dollars]
CMT	(1) アメリカ合衆国シークレットサービスは, / (2) 2 ヶ月前に, / (3) スイスの銀行口座を凍結しました, / (4) ここにいるサム・ジェイン氏の口座です, / (5) そしてその銀行口座には / (6) 490 万米ドルが入っていました, / (7) 凍結された時。 [The US Secret Service / two months ago / froze the Swiss bank account / the account of Mr. Sam Jain right here / and that bank account / had 14.9 million US dollars in it / when it was frozen]

表 4 は, NAIST-CMT-ED に収録されている順送り訳の統計量を示している。原発話文は平均で 3 つ程度のチャンクに分割されており, 各チャンクの語数は 15 語程度となっている。

表 4 順送り訳データの統計量. 各項目の標準偏差と, 目的言語文の語数は本研究で算出したものである。それ以外の値は [11] から引用している。

Data	Sum	Per Sent.±SD
# sentence pairs	511	—
# chunks	1,677	3.28±2.12
# source words	8,104	15.86±10.16
# target words	13,981	27.36±18.55