

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

福田 りょう^{1,a)} 土肥 康輔^{1,b)} 須藤 克仁^{1,c)} 中村 哲^{1,d)}

概要: 同時通訳では、時間的制約に対応するためにしばしば順送り方略が使用される。これは発話をより短い単位であるチャンクに区切りながら順次訳出していく方略である。英日同時機械翻訳でも順送り方略を使用することで、語順の違いによる遅延を最小限に抑えることが期待される。順送り方略を常に適用して原発話に忠実な訳出を行う英日同時翻訳システムの実現を目指し、本研究では高品質な順送り訳評価データを作成した。英日同時通訳コーパス NAIST-SIC の英語テキストの一部を通訳の方略に基づき自動でチャンキングし、日本語の順送り訳テキストを手で作成して付加した。本稿では、本データの作成手順と品質評価、及び分析結果を報告する。

1. はじめに

同時機械翻訳（同時翻訳）は、通常の翻訳よりも同時性の高い翻訳を行うための機械翻訳技術であり、言語の壁を超えた即時性の高い情報伝達を実現する技術としてシステムの実用化が期待されている。同時翻訳システムの主要部分である機械翻訳モデル（Machine Translation model; MT）の学習には主に翻訳コーパスが使用されるが、時間に制約されないオフライン翻訳の事例（例：表 1a）を学習した MT は、原言語の発話（原発話）から大きく遅れた訳出を行う場合がある。特に、英語と日本語のように語順が大きく異なる言語対でこの問題が顕著である。より低遅延な同時翻訳システムを実現するためには、通訳者による同時通訳の技法を取り入れることが有効であると考えられる。これまでに、同時通訳コーパスの整備が進められており [2], [8], [12], [17], それらを用いた同時翻訳システムの構築も検討されている [4], [6], [7]。これらの手法では、同時通訳の事例（例：表 1b）から同時通訳の方略を獲得することを目的として MT の学習が行われる。

遠山ら [14] は、英日同時通訳における同時性を重視した訳出方略を「順送りによる訳出（順送り方略）」と「短縮による訳出（短縮方略）」に大別した。順送り方略は、発話をより短い単位であるチャンクに区切りながら、それらを文

頭から順次訳出していく方略である。順送り方略においては、適切なチャンク境界の認定（チャンキング）と、つなぎ言葉を用いて各チャンクの翻訳を適切に接続することが重要である。短縮方略は、要約や重要でない箇所の省略などの技法を用いて遅延を低減する方略である。

前述したように、これまで同時通訳コーパスを用いたシステム構築が検討されてきた [4], [6], [7]。これらの先行研究において、システムの目標は順送り方略と短縮方略を使いこなす同時通訳者の訳出であり、そのため MT の評価には同時通訳の事例が用いられた。これに対して我々は、順送り方略だけを使用する英日同時翻訳システムの構築を検討している。そのようなシステムを作成する動機は主に以下の 2 点である。

- (1) **学習の難しさの緩和** 同時通訳において、短縮は訳出遅延の減少に寄与する重要な方略であると同時に、文脈等も考慮しながら欠落させても問題ない語句を見極める必要がある難易度の高い操作である。現時点では、MT が通訳者並みの適切さで短縮を行うことは実現されていない。近年の研究では、同時通訳コーパスを用いて学習した MT が重要な語句も含めて過剰に省略する問題が指摘されている [4]。順送り方略に限定すれば、短縮のための語句の重要性の判断が不要であるため、現在の MT の枠組みで実現しやすいと考えられる。
- (2) **原発話に忠実な同時翻訳の実現** 英日同時通訳で順送り方略を適用する際、通訳者は原発話を聞きながらチャンキングを行い、英語の文章構造に近い日本語訳を作成する。時間的制約がある中でこの順送り方略を常に適用することは難しく、通訳者は状況に応じて順送り

¹ 奈良先端科学技術大学院大学
NAIST, Takayama-cho, Ikoma, Nara 630-0192, Japan
a) fukuda.ryo.fo3@is.naist.jp
b) doi.kosuke.de8@is.naist.jp
c) sudoh@is.naist.jp
d) s-nakamura@is.naist.jp

表 1 英日翻訳におけるオフライン翻訳, 同時通訳, 順送り訳の比較. “/” はチャンクの境界を示す. 原発話のチャンクの先頭の数字はチャンクの通し番号である. 訳文のチャンクの先頭の数字は対応する原発話のチャンクの番号である. **赤字**は交差したチャンクを示す.

原発話	1 Groups like Anonymous / 2 have risen up / 3 over the last 12 months / 4 and have become a major player / 5 in the field of online attacks.
a. オフライン翻訳	1 Anonymous というグループは / 3 この 12 ヶ月ほど / 2 活気づいていて / 5 オンライン攻撃において / 4 大きな存在になってます
b. 同時通訳	1 アノニマス集団ですね, こちらのほうが, / 5 オンライン攻撃の, / 4 主要な, プレイヤーになっています.
c. 順送り訳	1 アノニマスのようなグループが / 2 台頭してきています, / 3 過去 12 ヶ月にわたって, / 4 そして主要なプレイヤーになっています, / 5 オンライン攻撃の分野において.

と逆送りによる訳出を使い分けられていると考えられている [15]. 常に順送り方略を適用できない状況で, 短縮方略は原発話に追従するために有効である一方, 原発話の重要な情報を欠落させる恐れがある. MT は人間と比べ非常に高速な処理が可能のため, 常に順送り方略を適用することで原発話に忠実な同時翻訳を実現できる可能性がある.

順送り方略だけを使用する英日同時翻訳システムの構築はこれまで検討されておらず, システム構築に必要な評価データも公開されたものは存在しない. そこで本研究では初期検討として, MT を評価するための英日順送り訳データを作成した. 英日同時通訳コーパス NAIST-SIC の英語テキストの一部を通訳の方略に基づき自動でチャンキングし, 日本語の順送り訳テキストを人手で作成して付加した (例: 表 1c) *1. 本稿では, 順送り訳の作成手順と品質評価, 及び分析結果を報告する.

2. 関連研究

文法構造が大きく異なる言語間の同時翻訳システムを実現するために, 順送り訳データの作成を行なった先行研究が存在する. 日英同時翻訳のための取り組みとして, Murata ら [5] は, 日本語発話を最大長 4.3 秒という制約のもとで人手でチャンキングし, 翻訳を付与して順送り訳データを作成した. He ら [3] は, 英語の翻訳文を日本語の語順に近づけるための構文変換ルールを定義した. 英日同時翻訳のための取り組みとして, 二又ら [13] は, 事前並び替えとスタイル変換モデルを用いて, 日本語の翻訳文を英語の語順に近づける自動言い換え手法を提案した. この手法では, 変換前後で文の意味が大きく異なることや, 変換により流暢さが失われるといった問題があった. 中林ら [16] は, 文より短いチャンク単位の対訳ペアを作成し, チャンク単位で文脈を考慮した翻訳を行う MT を構築した. この手法では, チャンク単位の翻訳を適切に接続し自然な訳文を生成することが課題として残された. 東山ら [18] は, 原発話テキストに人手の順送り訳テキストを付与した同時通訳コーパスを作成した. [18] と比較し, 本研究で作成し

た順送り訳データは評価を目的とした小規模なものであるが, 異なる以下の特徴がある. (1) 一貫性のあるチャンキングと人間の作業負担の軽減のために, チャンキングを自動化する. (2) 同時通訳との比較を可能にするために, 同時通訳コーパスに対して順送り訳を付与する. (3) 作成したデータを公開する.

3. 順送り訳の定義

本稿では, 順送り訳の定義を「チャンク単位で交差量が 0, かつ省略が 0 になるような訳出」とする. 交差量は, 原文のチャンクの位置を基準にしたとき, 訳文のチャンクの位置が右方向に移動したチャンクの数である [10]. 省略は, 対応する訳文のチャンクが存在しない原発話のチャンクの数である. 例えば表 1 のそれぞれの訳出の交差量と省略は以下の通りである.

- **オフライン翻訳** 交差量=2, 省略=0.
- **同時通訳** 交差量=1, 省略=2.
- **順送り訳** 交差量=0, 省略=0.

大きい交差量は同時性の低下に繋がる. オフライン翻訳の例では, 原発話の 2 番目のチャンクを訳出する前に 3 番目のチャンクの完了を待つ必要がある. 4 番目と 5 番目のチャンクについても同様である. また省略が多いほど原発話の情報が欠落する. 同時通訳の例では, “have risen up over the last 12 months” が省略されており, 聞き手に原発話の情報を全て伝えることができない. 一方で, 交差量と省略が 0 である順送り訳は, 同時性と原発話への忠実さの観点で理想的な訳出だと考えられる. ただし訳文全体として自然になるようにつなぎ言葉を多用するため, 訳出が長くなりやすい. 訳出が長すぎる場合, 原発話に追従するために発話速度を上げる必要があり, 聞き手が一度で理解しづらい, 聞き疲れしやすいといった問題が生じる可能性がある. こうした懸念を踏まえた順送り訳の有用性の検証は今後の課題とし, 本研究ではその前段階として順送り訳の作成を行った.

4. 順送り訳データ作成

本研究では [4] で評価データとして使用された, TED

*1 チャンキングのプログラムと順送り訳データ: <https://dsc-nlp.naist.jp/data/NAIST-SIC/Aligned-Chunk-Mono-EJ>

表 2 チャンキングしたデータの統計情報

文数の合計	511
チャンク数の合計	1,677
単語数の合計	8,104
文あたりの単語数	15.86
文あたりのチャンク数	3.28
チャンクあたりの単語数	4.83

talks の 8 トーク 511 文 *2 に対して順送り訳を付加した。このデータは、TED talks に基づく英日同時通訳コーパス NAIST-SIC [2] を MT の学習のためにアライメント手法を用いて対訳コーパスとして整備した NAIST-SIC-Aligned [9] の一部である。

4.1 チャンキング

岡村・山田 [10] は、具体的で簡潔かつ安定的に実行できるルールとして、4 つのチャンキングのルールを定めた (以下, I-IV)。このルールに 1 つのルール (以下, V) を追加した 5 つのルールを適用した。

- I 節を導く接続詞、関係詞 (主語を修飾する場合は除く) の前。
- II to 不定詞、前置詞、動名詞の後に 3 語以上が続く場合。
- III 主語が 3 語以上で長い場合。
- IV コンマ (1 語ずつの羅列は除く)、セミコロン、ハイフン等の記号の前後。
- V 文頭 (または節を導く接続詞、関係詞の直後) の前置詞句、副詞句の後。

まず入力文に対してルール I, III, IV, V をそれぞれ適用した後、全てのチャンク境界を保持する形で 4 つのルールの結果を統合した。その後追加でルール II を適用した。ルール II だけ後から適用したのは、「連続する 3 語 A, B, C に対し B と C の間にチャンク境界がある場合は、A と B の 2 語が続くと考える」といったように、他のルールの適用結果に影響されるためである。また、極端に小さいチャンクの出現を防ぐために、チャンクに含まれる最小単語数が 3 以上になるように、それよりも小さいチャンクを後ろのチャンクと一体化した。

チャンキングは spaCy の英語モデル *3 による構文解析情報を用いて自動化した。自動チャンキングの精度を評価するために、一部のデータ (88 文) を用いて著者らによるチャンキングとの一致度を測定した結果、F1 値 = 75.9%、適合率 = 72.5%、再現率 = 79.7% であった (自動チャンキングの最小単語数が 1 の場合)。

チャンキング後のデータの統計情報は表 2 の通りである。岡村・山田 [10] の分析による、実際の同時通訳におけ

*2 対象データは自動アライメントに基づきセグメント化されており、各セグメントが厳密に文単位に対応する訳ではないが、簡単のためここでは文と表記する。

*3 https://spacy.io/models/en#en_core_web_sm

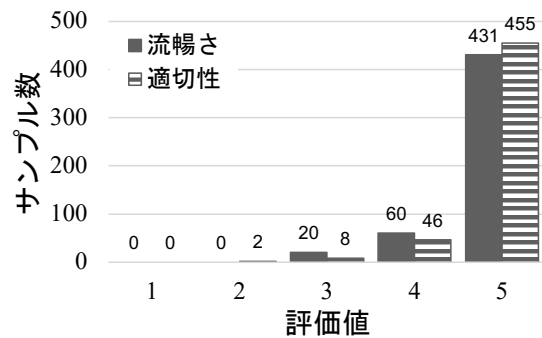


図 1 流畅さと適切性の評価。

る文あたりの単語数=36.96、文あたりのチャンク数=6.24 という値と比べると、チャンクの平均長は 1 単語程度短く、やや積極的にチャンキングを行っていると考えられる。ルール V の追加がこの原因の一つとして考えられる。

4.2 順送り訳の作成

表 1 に例示するような順送り訳を作成するため、翻訳者にチャンク境界 (“/”) を挿入した英文を提供し、以下の基本ルールで作業を依頼した。

- (1) 文頭からチャンクごとに翻訳する。
- (2) 文全体を見てチャンク間の繋がりが自然になるように翻訳する。
- (3) 後ろのチャンクの情報を含めずに翻訳する。

また、文の流畅さを保つために以下の操作も許容したが、下の操作ほど慎重に適用してもらった。

- 前のチャンクを繰り返し訳出すること。
- 翻訳すべき情報を後ろのチャンクに先送りすること。
- 不要な情報を省略すること。

5. 分析

5.1 品質評価

作成した順送り訳データの品質評価を行った。通訳と翻訳の実務経験を持つ 1 名のアノテータに、翻訳の評価基準である流畅さ (Fluency) と適切性 (Adequacy) [11] をサンプルごとに評価してもらった。

- **流畅さ (Fluency)** 翻訳文がその言語を母語とする人にとって、その言語の文章としてどの程度自然な表現であるかを以下の 5 段階で評価する: 1 理解不能, 2 不自然, 3 非母語的, 4 良い, 5 まったく問題ない。
- **適切性 (Adequacy)** 原文の情報がどの程度翻訳文に含まれているかを以下の 5 段階で評価する: 1 情報なし, 2 少しの情報, 3 多くの情報, 4 ほとんどの情報, 5 すべての情報。

品質評価結果として、評価値の分布を図 1 に示す。流畅さと適切性の平均±標準偏差はそれぞれ 4.8 ± 0.49 , 4.9 ± 0.41 と高く、511 件中 409 件は流畅さと適切性がともに 5 であった。概して自然かつ原文の内容に忠実な順送り訳で

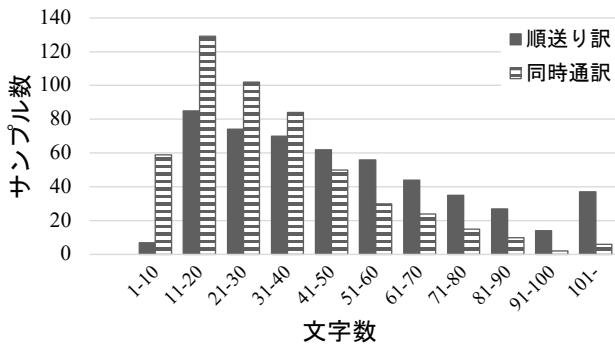


図 2 順送り訳と同時通訳の長さの度数分布.

表 3 同時通訳と順送り訳の比較.

例 1. 同時通訳の訳抜けと順送り訳の繰り返し	
原発話	Its flexible design / allows for deep squats, / crawls and highagility movements.
同時通訳	これでスクワットをしたりすることができるわけ です.
順送り訳	その柔軟な設計は / 深いスクワットを可能にし、 / クロールや高い敏捷性の動きを可能にします.
例 2. 同時通訳の省略と順送り訳のつなぎ言葉	
原発話	They use online attacks / to make lots of money / and lots and lots of it.
同時通訳	(直前の訳：オンラインの犯罪というのは、要するに お金を稼ごうとしているわけです.) これ攻撃をして、お金をたくさん稼いでると.
順送り訳	彼らがオンライン攻撃を用いるのは、 / 多くのお金 を稼ぐためです、 / それも山ほどのお金を.

あり、システム構築のための評価データとして適切であると考えられる。また流暢さ、または適切性が4以下のサンプル102件の中で、チャンキングが不適切というコメントを受けたものが37件あった(詳細は§5.4で述べる)。作業全体を通してのコメントは、「チャンクの位置のために訳出困難、意味ずれなどが多発した。それ以外は全体にとっても美しい順送り訳になっていると感じた」という内容だった。

5.2 同時通訳との比較

作成した順送り訳の長さをNAIST-SICの同時通訳と比較した。訳出の文字数のサンプル平均±標準偏差は、同時通訳が31.6 ± 21.63に対し順送り訳は50.3 ± 33.37であった。図2に長さの度数分布を示す。また文字数の合計は同時通訳が16,166、順送り訳が25,719であった。オフライン翻訳を含む音声翻訳コーパスMuST-C [1]から、今回作業対象にしたトークの日本語訳を抽出し、文字数を合計すると、22,289であった。オフライン翻訳と比べて、同時通訳は短く、順送り訳は長く訳出する傾向があると言える。訳出の長さの違いは、同時通訳は短縮方略により訳出が短くなりやすい、順送り訳は繰り返しやつなぎ言葉により長くなりやすい、といった特徴を反映していると考えられる。

表3に、特徴的な同時通訳と順送り訳の事例を示す。例1で、同時通訳は“flexible design”を省略し、“deep squats, crawls and highagility movements”を“スクワットをしたりすること”と大胆に要約している。時間的制約の中では、原発話に追従するためにこのような短い訳出が生じやすい。また例2で“online attacks”を単に“攻撃”と訳しているのは、“オンラインの...”という直前の訳出を考慮した結果であると考えられる。このような文脈を考慮した省略も、同時通訳の短縮方略の一つである。一方で順送り訳には、省略と先送りを避けつつ流暢さを保つための工夫が多く見られた。例えば、例1では“allows”を繰り返し訳出しており、例2では“のは、”、“です、それも”といったつなぎ言葉でチャンクを接続している。

5.3 流暢さや適切性が低い翻訳

順送り訳の中で流暢さや適切性が低い訳出の例を表4に示す。適切性が低い翻訳として、意識(例3)や過大な訳出(例4)などがあった。流暢さの低い翻訳は、例5のように原発話の長い文であることが多かった。このようなケースでは、文の分割(“...として。その構想...”)や繰り返し(“構想を... その構想... 構想です。”)といった文としての自然さを保つための工夫が多く見られた。

5.4 不適切なチャンキング

流暢さ、または適切性が4以下で、チャンキングが不適切というコメントを受けた37件のサンプルについて、各チャンク境界が5つのチャンキングのルールに従っているかどうかを手で評価した。37サンプルに含まれるチャンク境界163個の中で、チャンキングのルールに従っていないもの、つまり自動チャンキングのエラーが29個あった。表5の例6では、代名詞thatの前で分割する、節を導かない接続詞andで分割する、といったチャンキングのエラーが生じて訳出困難であった。例7では句動詞“running into”が誤って分割されたが、翻訳者は先送りで対処した。このような誤った境界での分割の他に、分割すべき場所で分割しないという誤りも存在することが、人手チャンキングに対する再現率=79.7% (§4.1)から伺える。

チャンキングのエラーを含まないにも関わらず、チャンキングが不適切というコメントを受けたサンプルは37件中15件あった。例8で、“one of”と“most of”はルールIIに従っているが分割位置としては不自然である。評価者は「チャンク箇所のためこのような過剰訳にせざるを得ない可能性がある」とコメントしている。

自動チャンキングは作業負担の軽減に有用であるものの、精度面や、チャンキングのルール自体にも改善の余地があると言える。

表 4 流暢さや適切性が低い訳出の例.

例 3. 意識	
原発話	To show you now / what we are working on / by starting out talking / about the American soldier, / that on average / does carry about 100 lbs.
順送り訳	これからお伝えする上で、 / 私たちの取り組みをお伝えする上で、 / 初めに話すのは、 / アメリカの兵士についてです、 / 平均して / 彼らは約 100 ポンドの荷物を運びます。
評価	To show you now を「これからお伝えする上で」は意識しすぎ。(流暢さ 5, 適切性 3)
例 4. 過分な訳	
原発話	And that is actually / for the last 500 years, / since its very successful introduction, / I must say.
順送り訳	そしてそれは実際、 / 過去 500 年間にも渡って、 / その導入が大成功してから今日まで続いていると / 言わなければならない。
評価	「今日まで続いていると」が過分な訳ではないか。(流暢さ 4, 適切性 3)
例 5. 流暢さが低い訳	
原発話	And getting this idea during winter, / I knew that I had several months / to plan it, / to find the different locations for the pieces / of the puzzle basically.
順送り訳	そしてこのアイデアを冬に思いついたので、 / 私には数ヶ月あることがわかっていました、 / 構想を練る期間として。 / その構想は、様々な場所を見つける為の、 / 基本的にパズルのピースとなるような場所を見つけるための構想です。
評価	なし(流暢さ 3, 適切性 5)

表 5 不適切なチャンキングの例。“/I,V” はルール I と V に該当すると判断されたチャンク境界.

例 6. チャンキングのエラー (1)	
原発話	But not only / I,V that it will integrate / I and network you to the universe / I and other devices out there.
順送り訳	しかし、まだあります、 / それは統合し、 / あなたを宇宙に統合しつなげるのです、 / 宇宙と周囲の他のデバイスに。
例 7. チャンキングのエラー (2)	
原発話	It was running / II into bankruptcy last fall, / I,IV because they were hacked into.
順送り訳	それは、 / 昨秋、破産寸前でした、 / ハッキングされたためです。
例 8. エラーではないが不自然なチャンキング	
原発話	And so I consider myself one / II of these people, / IV along with most / II of the other experimental quantum physicists, / I,IV who need a good deal of logic / II to string together these complex ideas.
順送り訳	そして私が自分自身について考えているのは、 / 私はこのような類の一人である、 / 大半の / 他の実験的量子物理学者たちと同じような類の一人であるという事です、 / この類の人たちは多くの論理を必要とします、 / これらの複雑なアイデアを結びつけるためにです。

6. おわりに

本稿では、英日同時翻訳システムの構築に向け順送り訳評価データを作成した。品質評価により、作成した順送り訳の流暢さと適切性が高いことを確かめた。一方で、チャンキングに改善の余地があることも明らかになった。今後の主要な課題は、順送り方略を常に適用する同時翻訳システムが実現した場合に、それがどのような場面で有用かを明らかにすることである。原発話に忠実な同時翻訳はすべての情報を伝達できる一方で、訳出が長くなるため、聞き手にとって一度で理解しづらい、聞き疲れしやすいといった問題が生じる可能性がある。今後は、同時翻訳の聞き手の視点から作成したデータを評価する実験を行う予定である。その後、本評価データを用いた同時翻訳システムの構築に取り組む。

謝辞 本研究は JSPS 科研費 JP21H05054, JP21H03500, JP23KJ1583 の助成を受けた。

参考文献

- [1] Di Gangi, M. A., Cattoni, R., Bentivogli, L., Negri, M. and Turchi, M.: MuST-C: a Multilingual Speech Translation Corpus, *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)* (Burstein, J., Doran, C. and Solorio, T., eds.), Minneapolis, Minnesota, Association for Computational Linguistics, pp. 2012–2017 (2019).
- [2] Doi, K., Sudoh, K. and Nakamura, S.: Large-Scale English-Japanese Simultaneous Interpretation Corpus: Construction and Analyses with Sentence-Aligned Data, *Proceedings of the 18th International Conference on Spoken Language Translation (IWSLT 2021)* (Federico, M., Waibel, A., Costa-jussà, M. R., Niehues, J., Stuker, S. and Salesky, E., eds.), Bangkok, Thailand (online), Association for Computational Linguistics, pp. 226–235 (2021).
- [3] He, H., Grissom, II, A., Morgan, J., Boyd-Graber, J. and Daumé, III, H.: Syntax-based Rewriting for Simultaneous Machine Translation, *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing* (Márquez, L., Callison-Burch, C. and Su, J., eds.), Lisbon, Portugal, Association for Computational

- Linguistics, pp. 55–64 (2015).
- [4] Ko, Y., Fukuda, R., Nishikawa, Y., Kano, Y., Sudoh, K. and Nakamura, S.: Tagged End-to-End Simultaneous Speech Translation Training Using Simultaneous Interpretation Data, *Proceedings of the 20th International Conference on Spoken Language Translation (IWSLT 2023)* (Salesky, E., Federico, M. and Carpuat, M., eds.), Toronto, Canada (in-person and online), Association for Computational Linguistics, pp. 363–375 (2023).
 - [5] Murata, M., Ohno, T., Matsubara, S. and Inagaki, Y.: Construction of Chunk-Aligned Bilingual Lecture Corpus for Simultaneous Machine Translation, *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)* (Calzolari, N., Choukri, K., Maegaard, B., Mariani, J., Odijk, J., Piperidis, S., Rosner, M. and Tapias, D., eds.), Valletta, Malta, European Language Resources Association (ELRA) (2010).
 - [6] Ryu, K., Mizuno, A., Matsubara, S. and Inagaki, Y.: Incremental Japanese spoken language generation in simultaneous machine interpretation, *Asian Symposium on Natural Language Processing to Overcome language Barriers* (2004).
 - [7] Shimizu, H., Neubig, G., Sakti, S., Toda, T. and Nakamura, S.: Constructing a speech translation system using simultaneous interpretation data, *Proceedings of the 10th International Workshop on Spoken Language Translation: Papers* (Zhang, J. Y., ed.), Heidelberg, Germany (2013).
 - [8] Tohyama, H., Ryu, K., Matsubara, S., Kawaguchi, N. and Inagaki, Y.: CIAIR Simultaneous Interpretation Corpus, *O-COCOSDA 2004 : proceedings of the 7th Oriental COCOSDA workshop*, pp. 72–74 (2004).
 - [9] Zhao, J., Ko, Y., Doi, K., Fukuda, R., Sudoh, K. and Nakamura, S.: NAIST-SIC-Aligned: Automatically-Aligned English-Japanese Simultaneous Interpretation Corpus (2023).
 - [10] 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, 英日通訳翻訳における語順処理 (浩之石塚, 編), ひつじ書房, pp. 217–250 (2023).
 - [11] 隅田英一郎, 佐々木裕, 山本誠一: 機械翻訳システム評価法の最前線, 情報処理, Vol. 46, No. 5, pp. 552–557 (2005).
 - [12] 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, No. 22, pp. 87–94 (2020).
 - [13] 二又航介, 須藤克仁, 中村 哲: 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, 言語処理学会 第 26 回年次大会 発表論文集 (2020).
 - [14] 遠山仁美, 松原茂樹: 同時通訳コーパスを用いた通訳者の訳出パターンの分析, 電子情報通信学会技術研究報告 = IEICE technical report : 信学技報, Vol. 103, No. 487, pp. 13–18 (2003).
 - [15] 蔡 仲熙, 笠浩一郎, 松原茂樹: 同時通訳における順送り方略の適用に影響する起点言語の特徴, 通訳翻訳研究, Vol. 20, pp. 25–40 (2020).
 - [16] 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 (2021).
 - [17] 松原茂樹, 相澤靖之, 河口信夫, 外山勝彦, 稲垣康善: 同時通訳コーパスの設計と構築, 通訳研究, Vol. 1, pp. 85–102 (2001).
 - [18] 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 (2023).