

# 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

遠山仁美<sup>†</sup> 松原茂樹<sup>‡</sup>

<sup>†</sup> 名古屋大学大学院国際言語文化研究科 〒464-8601 名古屋市千種区不老町

<sup>‡</sup> 名古屋大学情報連携基盤センター 〒464-8601 名古屋市千種区不老町

E-mail: {hitomi, matubara}@matubara.net

**あらまし** 本論文では、英語発話に追従して日本語訳を生成するという英日同時通訳の時間制約に関して、プロの同時通訳者の訳出方略を解明することを目的に、実際の英日同時通訳データに出現する訳出パターンについて分析した。分析には、名古屋大学独話同時通訳コーパスの英日通訳データ 4,578 対訳対応を使用し、同時性を重視するための方略である、順送りによる訳出、及び、短縮による訳出に該当する対訳対応を抽出した。出現頻度が高く、汎用性のある訳出パターンについて分析した結果、両訳出技法ともに同時通訳者が実際に活用していること、ならびに、同一の英文パターンに対して複数の訳出パターンが存在することが明らかになった。

**キーワード** 英日同時通訳、対訳コーパス、同時通訳者、訳出手法、多言語コミュニケーション

## Corpus-based Analysis of Simultaneous Interpreting Patterns

Hitomi TOYAMA<sup>†</sup> Shigeki MATSUBARA<sup>‡</sup>

<sup>†</sup> Graduate School of Languages and Cultures, Nagoya University, Furo-cho, Chikusa-ku, Nagoya-shi, 464-8601, Japan

<sup>‡</sup> Information Technology Center, Nagoya University, Furo-cho, Chikusa-ku, Nagoya-shi, 464-8601, Japan

E-mail: {hitomi, matubara}@matubara.net

**Abstract** This paper provides an investigation of simultaneous interpreting patterns in a bilingual spoken monologue corpus. 4,578 English-Japanese aligned utterances in the simultaneous interpretation corpus, which has been constructed at CIAIR, Nagoya University, were used for the investigation. Simultaneous interpreters have to generate Japanese translations simultaneously with English speeches. Therefore, they have various kinds of strategies to raise simultaneity. In this investigation, the interpreting patterns with high frequency and high flexibility were extracted from the corpus. In a result, it turns out that characteristic strategies are utilized frequently.

**Keyword** English-Japanese simultaneous interpretation, bilingual corpus, simultaneous interpreter, interpreting technique, multilingual communication

### 1. まえがき

自然で円滑なマルチリンガルコミュニケーション支援環境の実現を目指し、近年、同時通訳機に関する研究が進められるようになった[1,2]。同時通訳は、人間にとっても極めて高度な言語処理活動である。高い性能を備えた同時通訳技術を開発するために、プロの同時通訳者による振舞いを観察し、そこで得られた知見を利用することは効果的な方法の一つである。

同時通訳における重要なポイントとして、通訳が話者の発話に追従して遂行されなければならないという訳出タイミングに関する制約、いわゆる when-to-say の問題を挙げることができる。特に、英語と日本語のように構造上の差異が大きい言語間の通訳では、致命的となる。これに対してプロの通訳者は、膨大な訓練を通して蓄積した同時通訳技法を用いるこ

とにより、このような制約に対処しており、なかでも、原発話に対してどのような訳を生成するかという how-to-say に関する方略を駆使することが、その重要な部分を占めると指摘されている[5,6,7,8]。同時通訳に有用な訳出パターンを収集し、それを通訳ルールとして利用することにより、より同時進行性の高い通訳処理の実現が期待できる。

そこで本論文では、通訳者による発話を大量に収集し、調査することにより、同時通訳において利用可能な訳出方略について分析する。英語から日本語への独話通訳を分析対象とし、英語文とその標準的な日本語訳文との間で語の生起順序が大きく異なる場合の訳出方略について、通訳例を提示することにより考察する。

本分析では、名古屋大学独話同時通訳コーパス[3]を用いた。これは、名古屋大学統合音響情報研究拠点(CIAIR)によ

て構築された音声対訳コーパス[4]の一つであり、文化や政治等をトピックとする講演音声とその通訳音声から構成されている。本調査では、このうち、4,578 個の発話対訳対応を用いており、これは、同時通訳データの観察的分析の対象としては最大規模のものである。

本論文の構成は以下の通りである。まず、次の 2 章で、英日同時通訳における時間的問題について述べ、その関連研究を示す。3 章で同時通訳における訳出方略について議論し、4 章で分析に用いたコーパスについて概説する。5 章では、コーパスの詳細な分析によって獲得された訳出パターンを示す。

## 2. 英日同時通訳

### 2.1 英日同時通訳における時間的問題

同時通訳では、話者の発話に対して同時進行的に訳出を進める。韓国語と日本語のように構文的性質が比較的近い言語間の通訳に比べ、英語と日本語のように互いに違った性質をもつ言語間の場合には、対応する語の生起順序(word order)が異なり、それは訳出タイミングにおいて困難な問題を引き起こす。

英日同時通訳における問題の代表例として、英語は SVO 言語であるのに対し、日本語は SOV 言語であることを挙げることができる。英語話者発話の O の部分が長くなると、通訳者は O の発話が完了するまで V の訳出を待たなければならないが、しかしその一方で、文末まで待つ訳出を開始しては話者の発話に追従することができなくなる。他にも、英語における後置修飾節が日本語では前置修飾として現われるなど、同様のケースは数多く存在し、それは英日同時通訳において致命的である。

### 2.2 関連研究

プロの同時通訳者は、様々な訳出方略を駆使することによって上述の問題を克服していると考えられる。その方略の一つとして、できる限り同時的な訳出が可能となるような訳出方略を生成することがある。特に、英語から日本語への通訳の場合、日本語における語順の柔軟性等を利用することにより、いろいろな工夫を施すことができる。このような訳出方略について、実際の通訳データを用いた研究がこれまでに行われている。

通訳学的観点からの分析では、まず、同時通訳における訳出方略を具体的に示したものとして文献[5]がある。これは、通訳者養成のための教本として通訳協会によってまとめられたものであり、同時通訳で利用可能な訳出パターンが詳細に体系化されている。その他に、言語学的アプローチに基づく同時通訳処理単位の考察[6]や、通訳処理に関する認知モデルの実例を用いた検証[7]、同時通訳の訳出方略を応用した翻訳

文生成技法の提案[8,9]などがある。これらはいずれも実データを使用し定性的な説明を与えているものの、使用したデータの規模は比較的小さく、与えられた分析の汎用性や利用頻度については明らかではない。

一方、工学的観点からの分析も最近になって進められるようになりつつある。これらの多くは、通訳システムへの応用を指向しており、解説番組や模擬講演データを用いて通訳単位や訳出タイミング等に関する定量的な分析が与えられている[10, 11]。しかしながら、これまでのところ言語処理ツールを活用した大規模データの統計的分析が主であり、観察的アプローチによる定性的分析が望まれる。

## 3. 同時通訳方略の分類

プロの同時通訳者は、話者の発話に追従して訳出するという時間制約に対して、「いつ訳出するか(when-to-say)」及び、「どのように訳出するか(how-to-say)」の両側面に関する方略を駆使することにより対処していると考えられる。

このような観点に基づき、本研究では、同時通訳者の方略について以下のような分類を設定し、その方略における訳出パターンをコーパスを用いて調査する。

### A. 同時性を重視した訳出(when-to-say)

原発話にできる限り追従することを目指した訳出方略である。ただし、英語と日本語では語順が大きく異なるため、標準的な日本語をそのまま生成したのでは同時性を満たすのは難しい。以下のような方略にしたがって訳文の生成を工夫する。

#### A - 1. 順送りによる訳出(how-to-say)

話者発話の生起順序にしたがって文頭から訳出を進める方略である。話者の発話を文末まで待つ必要がないため、同時性を保ちながら話者の発話に追従できる。この場合、生起順序を構成する話者発話の分割単位を認定すること、及び、各単位の訳を適切に接続することがポイントとなる。訳文全体として自然になるようにつなぎ語を生成することになり、結果として発話量が多くなる場合がある。

#### A - 2. 短縮による訳出(how-to-say)

全体として通訳者の発話量が少なくなるような訳文を生成する方略である。話者発話よりも短い時間で生成できるため、訳出の遅れが小さくなり結果的に同時進行性の高い通訳が可能となる。技法として、訳文の言い換えや要約、重要でない箇所の省略などがある。

### B. 訳文の品質を重視した訳出(when-to-say)

話者発話を文末まで待ち、十分な情報を収集してから訳文

を整えたうえで訳出する方略である。原発話文の長さが短いときや文間の休止時間が長いとき、また、同時性を重視した方略を適用できない場合に利用される方略である。

以上の分類を図1に示す。あらゆる通訳文を厳密に分類することは、実際には困難であるが、これを本研究のコーパス分析における指針とする。

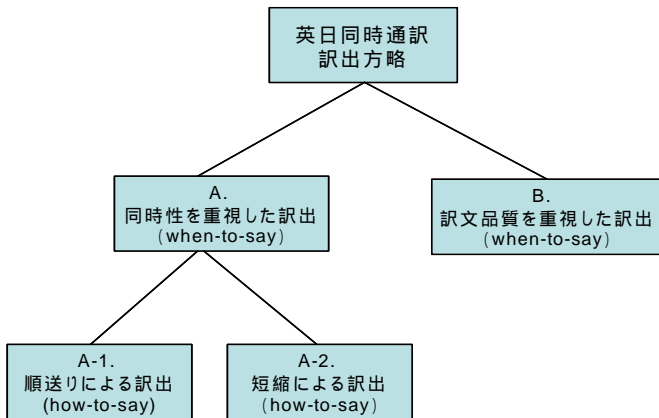


図1. 英日同時通訳方略

#### 4. コーパス分析の概要

本研究では、前章で示した分類のうち、同時性を追求した訳出方略(方略A)に着目することとし、特に、標準的な訳文を生成したのでは、訳出の同時性が大きく損なわれる発話文を中心に分析を実施した。また、将来のシステム開発への応用可能性を考慮し、出現頻度が比較的高く、また、汎用性を備えた訳出パターンを対象とした。

分析には、名古屋大学 CIAIR で構築している独話同時通訳コーパス[3]を使用した。これは、文化や政治など日常的話題に関する講演とその通訳の音声データ、ならびに、文字化データから構成されており、プロの同時通訳者の協力のもとに収録されている。文字化データには、話者情報や言語情報、時間情報などが付与されており、データ分析のための計算機支援環境も提供されている。

調査では、英語講演及びその英日通訳の発話対訳対応データを用いた。使用したデータの規模を表1に示す。26講演に含まれる4,578対応から上述の訳出パターンに該当する対応を抽出した。なお、表1のうち、形態素(単語)数については、英語は単語数とし、日本語は茶筌[12]の解析結果をもとに形態素数を計算した。異なり形態素数では、英語は表記が異なる単語を、日本語は基本形が異なる単語を数えた。

表1 使用した独話同時通訳コーパスのデータ規模

項目	英語話者	英日通訳者
収録時間(秒)	15627	15627
発話時間(秒)	12780	11001
形態素(単語数)	38473	44399
異なり形態素(単語)数	1880	2895
発話単位数	1494	1977

#### 5. コーパスを用いた調査結果

本章では、同時通訳コーパスを用いた分析結果について述べる。3章で示した「順送りによる訳出」(A-1)と「短縮による訳出」(A-2)に該当する訳出パターンを対象とする。なお、以下では、取り上げる英文パターンごとに、まず、一般的な英文法テキスト[13]に掲載された標準的な訳出パターンを示し、その上で、同時通訳において有用なパターンについて、実際に出現した通訳例を用いて論じる。なお、以下の通訳例は通訳対訳対応の列で示すものとする。N:は英語話者発話、I:は通訳者発話を表す。

##### 5.1 順送りによる訳出のパターン

複文構造を持つ長い文を同時通訳する場合、話者発話を文末まで聞いてから訳出を開始すると、通訳者は多くの情報を記憶していなければならない。そのため、長い文の場合は、意味のまとまりごとに文頭から分割し、順送りに訳出することが望ましい。本節では、複文として、従属節、あるいは、関係節が文の後半に出現する発話文を対象として、順送り訳の訳出パターンを分析した。

##### 5.1.1 従属節をもつ複文の訳出パターン

従属節が主文の前に来る場合は、容易に順送りに訳すことができる。しかし、従属節が主文の後ろに来る場合は、訳出を工夫する必要がある。本調査では、because, even if, beforeを対象とした。

##### because

従属接続詞 because を含む文は、通常、以下のように訳される。

英文: A because B

訳文: B だから A

この場合、Bの発話が終わるまで全体の訳出を待たなければならない。一方、同時通訳では、becauseの前で文を分割し、英語の文体に合わせて順送りに訳出していく手法が頻りに用いられる。調査では、順送りに訳されたケースが23件、後ろから前に訳されたケースが4件あった。順送り訳では主に以下の3つの訳出パターンが見られた。

<パターン1> (例1参照)

### 訳文：A, なぜならば B

まず主節 A を訳し,続けて「なぜならば～」と because 節を続ける。

#### <パターン 2> (例 2 参照)

### 訳文：A, B なのです

主節 A と because 節を順送りに訳し,最後に because の訳にあたる「なのです」を訳出する。

#### <パターン 3> (例 3 参照)

### 訳文：A, B なのです

主節 A と because 節 B を順送りに訳すが,その 2 つの節を結び because 自体の訳出を省略する。このように,主節 A と従属節 B の意味的結びつきが強い場合, because 自体の訳出を省略できる。これは,訳出量を減らし,時間を節約する方略 A-2 にも該当する。

#### 例 1

N: But you don't see the pedal **because**

I: あ ペダルはあまり見ていないと。なぜなら

N: you don't understand the pedal.

I: ペダルの役割を あー 理解していないからです。

#### 例 2

N: So when I came to Kansai, it was a really big shock **because** these people look so normal.

I: ところが関西に来たときにまたびっくりしました。普通なんです。

#### 例 3

N: I was completely shocked **because**

I: 私はすごいショックを受けてしまいました。で

N: it was a train that I had never seen, a train like I had never seen before.

I: 電車は前私が見たこともないような電車でした。

### even if

even if は,通常,以下のように訳される。

英文: A even if B

訳文: たとえ B であるとしても A

because と同様に,英語・日本語間で A, B の生起順序が逆になる。そのため,同時通訳では,以下の訳出パターンが用いられる。コーパス中に even if は 4 件あり,4 件とも全て順送りで訳されていた。

#### <パターン 4> (例 4 参照)

### 訳文：A, たとえ B であるとしても

順送りに訳すために, even if の前で文を分割し,「た

とえ」を生成する。

#### 例 4

N: And another way to start is always be willing to work **even if** you're a student.

I: えー それからまた始めるにあたりましていつもです。ね。へー 喜んで仕事をしなさい。学生でもです。ね。

### before

一般には, before は以下のように訳される。

英文: A before B

訳文: B する前に A

しかし,この訳出に従うと, B を聞き終わるまで,訳出を開始できない。本調査では,以下の訳出パターンが見られた。(3 件)

#### <パターン 5> (例 5, 6 参照)

### 訳文：A してから(した後に) B

#### 例 5

N: Okay, you should pay back ahh most or all of your loans **before** you invest a lot of money.

I: そして えー 皆さんの借金のほとんどを返してから投資をするということにしたほうがいいと思います。

#### 例 6

N: you should pay back ahh most or all of your loans

I: クレジットカードでの借り入れを

N: **before** you invest a lot of money.

I: おー 返済した後 お 投資を考えなさいということがあります。

### 5.1.2 関係節 を持つ文の訳出パターン

従属節の場合と同様,関係節が主文の後ろに来る場合,一般に,関係詞の前で文を分割し,順送りに訳出すると同時性が高くなる。本調査では,特に,出現頻度が高く,訳出に何らかの工夫を必要とする,関係代名詞 **which**, 関係副詞 **where** についてその訳出パターンを示す。

### which

関係代名詞の文を訳す場合,通常,関係節が主文にかかるように訳す。

英文: A <先行詞> which B

訳文: B <先行詞> A

コーパス中に見られた順送りの訳出パターンは主に次の 2 つである。

<パターン6> (例7参照)

訳文: A これに関してはB

関係代名詞の文法的役割「代名詞+接続詞」を反映するように訳すパターンである。

<パターン7> (例8参照)

訳文: A, B A

主節のAに続いて、関係節Bを訳し、再びBの関係節がAの主節にかかるように言い直すパターンである。言い直しを用いるため、結果として、通訳者の発話量は長くなる。

例7

N: One egg exception might be your home **which** you pay off little by little over the years, but

I: まあ 一つ例外はこれは えー 住宅ですね。これに関しては まあ 長い時間をかけて少しずつ返すということですけども。

例8

N: exception might be your home **which** you pay off little by little over the years,

I: 例外はですね えー 家であります。数年かけてえ 少しずつ分割で払っていく家の場合は例外でありますけれども お ローンとか

where

関係副詞 where 節が主文の後ろに来る場合、通常、関係節が主文にかかるように訳す。

英文: A <先行詞> where B

訳文: B <先行詞> A

しかし、文全体が長い場合など、関係節を聞き終わってから訳出すると同時性が失われる。本調査では、where の前で文を分割し、文頭から順送りに訳出するケースが見られた(9件)。

<パターン6> (例9参照)

訳文: A B

例9では、主文と関係節が、そのまま2文で訳され、それらをつなぐ where の訳出は省略されている。この訳出パターンは短縮による訳出という方略 A-2 にも該当すると考えられる。

例9

N: set about collecting broken toys from the garbage

I: 壊れたおもちゃをお ごみの中から集めてきて

N: and bringing them back to his little shoe shop

I: え そして それを自分の靴工房に持ち込んで

N: **where** he would repair them by hand and repaint them and make them look new.

I: えー 手で一つ一つ うー 修理をしたり色を塗り直したりしてそして新しくきれいに

5.2 短縮による訳出のパターン

訳出の発話量を少なくすることにより時間を節約し、同時性を高める方略(1-B)について論じる。本研究では発話量を減らす手法として、「省略」と「圧縮(要約)」の2つに分けて調査を行った。省略とは、ある単語を訳出しなかった場合であり、圧縮(要約)は、英語の文体よりむしろ、日本語の文体に即して訳し、結果的に文全体を短くする方略である。

5.2.1 省略

同時通訳における時間的切迫、省略は話者の発話に追従するための重要な方略の1つである[14]。本研究では、出現頻度の高い思考・知覚に関する動詞のうち、think を対象に調査した。

思考・知覚動詞の主語

“think + 名詞節” について調査を行った。

英文: A think B

訳文: A はB と思います。

調査の結果、主語(多くは“I”)は訳出されず、think の訳は文末に現われる傾向があった。

<パターン7> (例10参照)

訳文: B と思います

think の部分が「**なんでしょう**」(例11)や、「**～ではないでしょうか**」(例12)という疑問形になるパターンも見られた。

例10

N: **I think** that is the key, I think that is the pedal of our bicycle here.

I: まあ これはとても重要だ**と思います**。これこそが あー さきほどの自転車のペダルではないか**と思います**。

例11

N: I think that's part of maintaining um a proper appearance

I: 身なりを正しく整えなければということ**なんでしょう**。

例12

N: **I think** that has a very major affect on your mind.

I: こういったことが あー 住んでる人の心に大きな影響を与える**のではないでしょうか**。

## 5.2.2 要約(圧縮)

要約(圧縮)は、英語の文体を反映して訳出するのではなく、通常の日本語の文体に基づいて訳すことにより、結果的に無駄な言い回しを省き、発話自体を短くする方略であるとする。

### 受動態の変換による短縮

日本語の特徴の1つに、話者の主観により文章を構成する点が挙げられる[8]。逆に、英語では受動態が頻繁に用いられる。このとき、通常の翻訳文は次のようになる。

英文: A is <動詞の過去分詞形> by B

訳文: AはBによって<動詞>される

同時通訳では、英語受動態の文構造を反映して訳出するよりも、能動態の文体で訳出する手法が頻繁に用いられる[5]。本調査では、英文の受動態表現における訳出時の圧縮(要約)現象に着目し、調査を行った。

### <パターン8>(例12参照)

訳文: Bを<動詞>する

#### 例12

N: We have places in North America, South America, Europe and Africa that where the land can be developed better.

I: えー北アメリカ あ南アメリカヨーロッパそしてアフリカでそういう うー土地をですねえよきゅよく開発をしていくことができます。

## 6. まとめ

本論文では、英語発話に追従して日本語訳を生成するという時間制約について、プロの同時通訳者の訳出方略を解明することを目的に、実際の英日同時通訳データに出現する訳出パターンについて分析した。分析には、名古屋大学独話同時通訳コーパスの英日通訳データ4,578対訳対応を使用し、同時性を重視するための方略である「順送りによる訳出」及び、短縮による訳出に該当する対訳対応を抽出した。出現頻度が高く、汎用性のある訳出パターンについて分析した結果、両訳出技法ともに同時通訳者が実際に活用していること、ならびに、同一の英文パターンに対しても複数の訳出パターンが存在することが明らかになった。

本研究では、同時性を重視するための方略として、主に複文を対象として分析を実施したが、それ以外の英文についても同様の方略が存在し、様々な訳出パターンとしてコーパスに現われる。また、複数の訳出パターンが選択的に用いられ、その用いられ方にはいくつかの要因が存在すると予想される。コーパスを用いたこれらの分析は今後の課題である。

## 謝辞

同時通訳データベースの使用にあたり数々のご教示をいただいた名古屋大学稲垣研究室の諸氏、ならびに、コーパス分析にあたり重要なお示唆をいただいた本学大学院生の野誠寛氏、大原誠氏、笠浩一朗氏に深く感謝いたします。本研究の一部は、文部科学省科学研究費補助金C O E形成基礎研究費(課題番号11CE2005「多元音響の統合的理解」代表:名古屋大学 板倉文忠 教授)、ならびに(財)中島記念国際交流財団 研究助成によります。

## 文献

- [1] S. Matsubara, K. Toyama, Y. Inagaki, "Sync/Trans: Simultaneous Machine Interpretation between English and Japanese," *Advanced Topics in Artificial Intelligence, Lecture Notes in Artificial intelligence*, Vol.1747, pp.134-143, 1999.
- [2] H. Mima, H. Iida, O. Furuse, "Simultaneous Interpretation Utilizing Example-based Incremental Transfer," *Proc.17<sup>th</sup> International Conference on Computational Linguistics and 36<sup>th</sup> Annual Meeting of the Association for Computational Linguistics*, pp.855-861, 1998.
- [3] S. Matsubara, A.Takagi, N. Kawaguchi, and Y. Inagaki, "Bilingual Spoken Monologue Corpus for Simultaneous Machine Interpretation Research," *Proceedings of 3<sup>rd</sup> International Conference on Language Resources and Evaluation*, Vol. I, pp.135-159, 2002.
- [4] 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*, No.1, pp.85-102, 2001.
- [5] 松本 謙太郎, 向 謙治郎, 中沢 弘雄, 日本通訳協会編, "英語通訳への道," 大修館書店, 2000.
- [6] 船山 伸他, "同時通訳における処理単位," *通訳理論研究*, No.10, 1996.
- [7] 水野 的, "同時通訳動態モデルの展開( )," *通訳理論研究*, No.8, pp.9-26, 1995.
- [8] 亀井 忠一, "頭からの翻訳法," 信山社, 1994.
- [9] 近藤 正臣, "英語における後置修飾節の邦訳," *大東文化大学紀要*, 第30号, pp.277-294, 1992.
- [10] 柏岡 秀紀, 田中 英輝, "講演の同時通訳データの分析," *言語処理学会第7回年次大会発表論文集*, pp.433-436, 2001.
- [11] A. Takagi, S. Matsubara, N. Kawaguchi and Yasuyoshi Inagaki, *Corpus-based Analysis of Simultaneous Interpretation*, *Proceedings of 5th Symposium on Natural Language Processing(SNLP^2002)*, pp. 167-174, 2002.
- [12] 松本 裕治, 北内 啓, 山下 達雄, 平野 善隆, 松田 寛, 高岡 一馬, 浅原 正幸, "日本語形態素解析システム 茶釜 version 2.2.0," 奈良先端科学技術大学院大学松本研究室, 2000.
- [13] 河上 道夫, 広田 成章, "高校新基礎英語新訂版," 桐原書店, 1992.
- [14] 小林 淳夫, "通訳の極意: 達人のテクニックとトレーニング方法," 南雲堂フェニックス, 2001.