

# 同時通訳のストラテジー

## ー訳出セグメントの分割位置とその再構成手法の分析ー

中林 明子

加藤 恒昭

東京大学大学院 総合文化研究科

akiyasumi@hotmail.com

kato@boz.c.u-tokyo.ac.jp

### 1 はじめに

同時通訳者は原文の発話を聞きながら、ほぼ同時に発話の内容を他の言語に訳出する。特に英語・日本語間など語順の異なる言語間の同時通訳においては、処理の負荷を制御し、同時性を保つために、一文を「意味の単位 (unit of meaning)」で分割して訳出単位とし、訳出セグメントごとに前から順に処理する。そして訳出セグメントの訳出結果を再構成、簡素化、省略することで自然な訳出を実現する [3]。本稿ではこのような同時通訳の枠組みを前提に、訳出セグメントの分割と再構成を英日同時通訳者がどのように行っているかというストラテジーを分析する。このストラテジーを同時機械翻訳に応用することを考えており、本稿をその実現に向けた基礎データの収集分析と位置付ける。

### 2 動機と枠組み

同時通訳を実現する同時機械翻訳を検討している。同時機械翻訳においても、同時性を保つために入力文を訳出セグメントに分割し、それらを順次翻訳していくのが一般的な枠組みである。訳出セグメントの分割位置については、様々な研究が行われてきているが [1, 2, 4, 7]、明確な定義は得られていない。また、例えば主語の後で文が分割される場合は、名詞句からなる訳出セグメントと動詞句からなる訳出セグメントが生成されるが、このように統語単位が文ではない訳出セグメントの翻訳品質を高めることも重要な課題である。

本稿では、同時通訳者が行っている訳出セグメントへの分割と、訳出セグメントを翻訳した結果の再構成について分析を行い、同時性を保った自然な訳出を実現するこれらストラテジーから知見を得る。ここで、再構成とは統語的に要素を変形させることを指し、広

義では訳出セグメントへの分割や主節と従属節の入れ替え、能動態と受動態の入れ替えも含んで用いられる [3] が、本稿においては句と節の入れ替えを再構成と呼ぶこととする。

分析対象のコーパスは、CIAIR 同時通訳コーパス [5] を使用する。2000 年に収録された独話データ (24 スピーチ) は、1 スピーチにつき 4 人の通訳者が通訳を行っている。これから、10 スピーチ (計 510 文、12,859 単語) を分析対象とした。

### 3 訳出セグメントの分割位置

#### 3.1 分割位置

同時通訳者が判断した訳出セグメントの分割位置を分析するにあたり、その手掛かりは原文ではなく、通訳文に現れることが多いことに着目する。例えば、以下のように同時通訳例を見ると、原文だけでは分割位置の同定が難しい。

原文: If you do that, the ups and downs seem to level out and you build more. It's a natural way of making money. (SXPSX006.NX02.ETRANS)

通訳: その間にはいろいろな上下があると思いますがその長い期間の間にそういったものは平均が取れて最終的にはお金が儲かっていくと思います。 (SX-PSX006.L.IA08.JTRANS)

上記の原文において第一文の条件節と主節の間、第一文と第二文の間で訳出セグメントが分割されると想定されるが、通訳例では、*the ups and downs* に対応する“いろいろな上下があると思いますが”の後で区切られると考えることができ、それは訳出の単位とは対応していない。

「意味の単位」とは聞き手の心の中の認知的表現 (cog-

nitive representation) である [3] との考えに基づき、「意味の単位」は通訳文に現れると仮定し、分割位置の同定に通訳文を参考にすることを提案する。具体的には、日本語の通訳文における節を「意味の単位」であると考え、それと対応する原文の節や句を原文における「意味の単位」、つまりそれぞれの通訳者が判断した分割位置と考える。上記例では、原文において、以下の位置で「意味の単位」を認識していると考えられる。

原文：If you do that, the ups and downs seem / to level out / and you build more. It's a natural way of making money. /

同時機械翻訳においては、原文から訳出セグメントの特徴を明らかにすることが必要である。ここで提案した方法で訳出セグメントを同定し、それを注釈づけたスピーチを教師データとする機械学習を行うことで、原文から訳出セグメントを決定することが可能である。

### 3.2 分析結果

提案した枠組みに基づいて、訳出セグメントの同定を行った。1つのスピーチに対して4人の通訳者の通訳文があるため、パイロットとして原文ファイル SXPSX005.NX02.ETRANS に対応する4人の通訳者のパフォーマンスから、訳出セグメントをそれぞれ同定した。分割位置の分布を表1に示す。延べ441セグメントのうち、364セグメント(82.5%)において分割位置が一致した。

表1 訳出セグメントの分割位置が一致した通訳者の数(ファイル SXPSX005.NX02.ETRANS)

一致した通訳者数	セグメント数(延べ数)
4	64 (256)
3	36 (108)
2	24 (48)
1	29 (29)
合計	153 (441)

併せて、分割された通訳文のセグメントの順序が原文の順序とほぼすべての箇所一致し、順序が異なる箇所は2箇所だけであることを確認した。これらのことから、ここで提案した通訳文の節が「意味の単位」を判断する基準となりえると判断した。以下の分析では、4人のうち3人の通訳者のデータが一致した分割位置を最終的な訳出セグメントの分割位置とする。

表2に分析対象とした10スピーチ全体における文

およびセグメントの合計とそれぞれに含まれる単語数の平均を示す。原文における文は510であるのに対し、通訳文の平均は769.75であることから、通訳者は原文を分割して通訳文を構成していることがわかる。翻訳そのものに要する時間を無視すると、同時通訳者はセグメント単位で翻訳することにより、文単位で翻訳する場合に比べて待機時間を7秒削減していると考えられる。

表2 文および訳出セグメントの数と含まれる単語数

		原文	通訳文平均
スピーチ	単語数合計	12,859	
	時間合計	41:49	
文単位	文数総数	510	869.75
	平均単語数	25.2	
	平均時間	0.12	
セグメント単位	セグメント合計	1,124	1,333
	平均単語数	11.4	
	平均時間	0.05	

分割された訳出セグメントを構成する、節と句の分布を表3に示す(複数の異なる節や句から構成される21セグメントを除く)。訳出セグメントは、様々な統語単位で構成されている。このうち、関係詞節から構成される訳出セグメントは、次節で述べるように通常は名詞修飾節を含む名詞句として訳出されることが多いが、同時通訳においては連用節に再構成されることが多い。また句から構成される訳出セグメントは、名詞句と前置詞句が主に句から節への再構成の対象となっている。なお、動詞句はそのほとんどが主語が省略された節であるため、再構成は必要としない。10%を超える訳出セグメントに対して再構成が行われていることがわかる。

表3 訳出セグメントを構成する節と句

関係節	39 (3.5%)
不定詞節	23 (2.1%)
分詞節	21 (1.9%)
その他の節	853 (75.9%)
節合計	946 (84.2%)
名詞句	65 (5.8%)
動詞句	56 (5.0%)
前置詞句	32 (2.9%)
形容詞・形容動詞句	4 (0.4%)
句合計	157 (14.0%)

## 4 関係節の訳出ストラテジー

関係節は通常関係節が先行詞を修飾して名詞句を構成し、主節に埋め込まれるように訳出される [8]。しかし同時通訳においては同時性を保つために下記の 2 例のように順送りに訳出する場合がある。

原文：Also in San Francisco, you can go to Hayes Valley / another small neighbourhood / that has the shops (SXUSX021.NX05.ETRANS)

通訳：あとサンフランシスコヘズバリーもいいですよ。/こちらも小さな町なんですけど/店がいろいろ並んで (SXUSX021.K.IA06.JTRANS)

原文：I'd like to speak about an analogy / that a professor had said to his class once. / (SXUSX012.NX03.ETRANS)

通訳：まずアナロジーを話したいと思います。/私の教授の一人が一度授業で話をしてくれたアナロジーです。/ (SXUSX012.L.IA10.JTRANS)

上記例では共に、関係節が訳出セグメントとなって、順送りに訳出され、節として構成されている。これらのように関係節が先行詞に係らず、独立した節になっている場合を順行訳と呼ぶ。また、関係節が先行詞に係って埋め込まれるような訳出を逆行訳と呼ぶ。訳出結果を見てみると、第一の例では関係節に先行詞は現れず、主節と関係節は複文的な関係となっているのに対し、第二の例では、関係節において先行詞“アナロジー”が反復されている。また関係節は名詞修飾節の形をとって反復された先行詞に係り名詞句を作る。それに助動詞“です”を付加することで節に再構成されている。以下では、どのような関係節が訳出セグメントとなって順行訳を選択するか、またその場合にどのような再構成方略がとられるかを分析する。

### 4.1 訳出順序の選択にかかわる要因

どのような場合に訳出セグメントが関係節の単位で構成され、順行訳が選択されるか、訳出順序の選択に関わる要因を分析した。

**■限定用法と非限定用法** 関係節には限定用法と非限定用法があり、日本語において関係節にあたる名詞修飾節も限定的な場合と非限定的な場合に分類される。通常の翻訳において限定用法では名詞修飾節をとる逆行訳が選択されるが、非限定的名詞修飾節はほかの従

属節を用いた複文に置き換えることが可能である [6] ため、非限定用法では順行訳が選択されることがある。

同時通訳における限定・非限定の用法と訳出順序の関係を図 1 に示す。非限定用法の翻訳データは 158 件、限定用法の翻訳データは 160 件であった（どちらの用法とも取れる 4 件のデータを除く）。同時通訳においても、非限定用法は順行訳が選択される傾向が強い。一方、通常逆行訳として翻訳される限定用法においても、46% のケースにおいて順行訳が選択される。

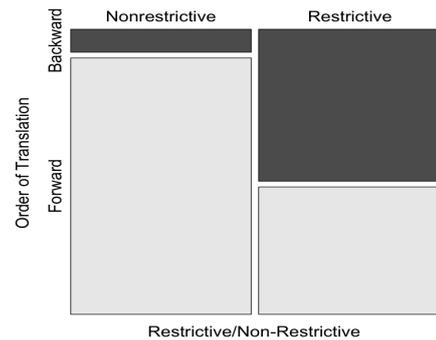


図 1 限定用法と非限定用法の訳出順序

**■関係節における先行詞の役割との関係** 上記限定用法においてどのような場合に順行訳・逆行訳が選択されるかを調べるために、関係節における先行詞が果たす統語的役割を調べた。関係節において先行詞が主語の役割を持つデータ (SBJ: 35 件)、目的語および修飾語の役割を持つデータ (OBJ: 18 件、MOD: 4 件) に対して逆行しやすさ B を以下のように定義し、比較した。値が 1 に近いほど、逆行訳を選択する通訳者が多く、逆行しやすいことを示す。

$$B = \text{逆行訳数} / (\text{逆行訳数} + \text{順行訳数})$$

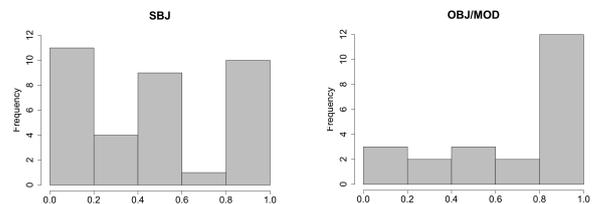


図 2 関係節内の先行詞の役割ごとの逆行しやすさ B の分布

関係節における先行詞の役割が目的語、または修飾語の場合は、主語である場合に比べて逆行訳をとりやすいと考えられる (p 値=0.023 < 0.05, t 検定)。つまり限定用法において順行訳が選択されるのは、関係節における先行詞の役割が主語である場合が多いと言える。

## 4.2 訳出のパターン

前出の例で示したように、限定用法の訳出パターンは、関係節における先行詞の反復、名詞修飾節の有無、訳出セグメントの翻訳結果への情報の付加の組み合わせで説明することができる。これらの利用頻度や利用を決定する要因について調査した。

**■関係節における先行詞の反復** 関係節において先行詞は様々な形で反復される。限定用法の順行訳で使用されている先行詞反復のタイプを非限定用法の場合の先行詞反復のタイプと比較し、表4に示す。限定用法で

表4 反復される先行詞

	限定	非限定	合計
先行詞	32 (43.8%)	37 (25.5%)	69
指示語	14 (19.2%)	37 (25.5%)	51
抽象語	11 (15.1%)	9 (6.2%)	20
その他	3 (4.1%)	3 (2.1%)	6
なし	13 (17.8%)	59 (40.7%)	72

は先行詞をそのまま使用する場合、“これ”や“それ”といった指示語を使用する場合、“こと”や“人”など抽象化した語を使用する場合の割合が全体の78.1%である。一方非限定用法では先行詞を省略する場合は全体の半数近くに上る。

**■名詞修飾節の有無** 順行訳においても名詞修飾節をとる場合ととらない場合がある。関係節内の先行詞の役割ごとの傾向を図3に示す。関係節において先行詞

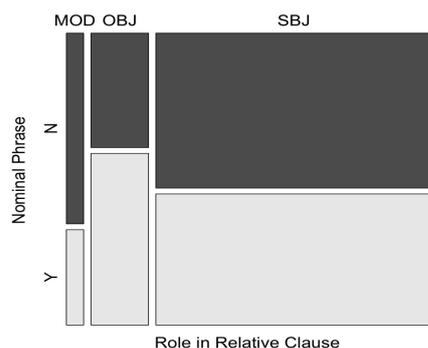


図3 先行詞の役割と名詞修飾節の有無の関係

が目的語である場合は、名詞修飾節をとること(Y)が多く、主語である場合は、名詞修飾節をとらない場合(N)がわずかに多いが、大きな違いは見られない。

**■訳出セグメントの翻訳結果への情報の付加** 名詞修飾節を作る場合、関係節は名詞句となる。同時通訳においては、名詞句に情報を付加し、節として再構成することが多く見られる。表5に付加される情報を示す。11

表5 付加される情報

パターン	件数
動詞句の付加	11
です(けれども)の付加	10
その他	6

件のケースで動詞句が付加され、10件のケースで助動詞の“です”が付加されて節に再構成される。付加される動詞句のうち前文の動詞句の繰り返しが9件、その他文脈に沿った動詞句を付加する場合は2件であった。

## 5 おわりに

同時機械翻訳で課題となる訳出セグメントの分割位置と不完全な訳出セグメントの訳出方法を、同時通訳者のパフォーマンスをもとに分析した。

訳出セグメントの分割位置においては、通訳者の通訳文から「意味の単位」を判断する方法を提案した。また上記方法によって分割した訳出セグメントをもとに、関係節をとりあげ、どのような再構成が行われているかを分析した。今後は、名詞句や前置詞句についても分析を進める。さらに今後の展開として、得られた知見を同時機械翻訳に応用していく。

## 参考文献

- [1] Grissom II, A., He, H., Boyd-Graber, J., Morgan, J., & Daumé III, H. (2014). Don't until the final verb wait: Reinforcement learning for simultaneous machine translation. *Proceedings of the 2014 Conference on empirical methods in natural language processing (EMNLP)*, 1342-1352.
- [2] Gu, J., Neubig, G., Cho, K., & Li, V. O. (2016). Learning to translate in real-time with neural machine translation. *arXiv:1610.00388*.
- [3] Jones, R. (1998). *Conference interpreting explained*. Routledge.
- [4] Kashioka, H., Maruyama, T., & Tanaka, H. (2003). Building a parallel corpus for monologue with clause alignment. *Proceedings of the 9th Machine Translation Summit*, 216-223.
- [5] 松原茂樹, 相澤靖之, 河口信夫, 外山勝彦, & 稲垣康善 (2001). 「同時通訳コーパスの設計と構築」『通訳研究』(1), 85-102.
- [6] 日本語記述文法研究会 (2008). 『現代日本語文法 6 第11部・複文』. くろしお出版.
- [7] Oda, Y., Neubig, G., Sakti, S., Toda, T., & Nakamura, S. (2015). Syntax-based Simultaneous Translation through Prediction of Unseen Syntactic Constituents. *ACL* (1), 198-207.
- [8] 遠山仁美, & 松原茂樹 (2003). 「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」『信学技報』103(487), 13-18.