

講演の同時通訳データの分析

柏岡 秀紀, 田中 英輝
ATR 音声言語通信研究所

1 はじめに

近年、音声翻訳システムや音声ワープロ等の商品がでてきており、音声言語に対応したシステムへの期待が高まってきている。これまで ATR では、“対話”を中心に音声翻訳システムの基礎技術についての研究を進め、その実現可能性、有効性について検証してきた [1]。旅行会話を中心とした“対話”は、比較的文が短く、非文法的な表現や省略された表現が多く現れるという特徴をもつ。

音声翻訳システムの応用を考えたとき、旅行会話の様な特徴をもつ“対話”とともに、講演やニュース等の“独話”が重要な位置を占める。“独話”は、1文が長く、比較的複雑な構造の表現が現れるという特徴がある。本稿では、ニュースや講演を対象とした“独話”の音声翻訳システムの基礎研究として、同時通訳的な技術に焦点をあて、その第1段階として、人間による同時通訳データの収集、および分析を行った。対象とした“独話”は、NHKで放送されている「あすを読む」という解説番組である。実際に同時通訳データを収集し、同時通訳のタイミングについて、通訳者の発声に含まれるポーズに着目し、通訳の発話(同時通訳者の発話、英語)をはじめのタイミングと原発話(解説者の発話、日本語)の関係を調べた。さらに、同時通訳のために予め拾い出していたキーワードが、どの程度遅れて訳出されているかを調べた。その結果、同時通訳者の発話のタイミングは、原発話の述部に深く関連していることが分った。また、キーワードはほぼ全てが訳出されており、遅れは長くても10秒であった。このような結果を踏まえた上で、同時通訳の追従性について考察し、同時通訳的な音声翻訳システムへの検討を行う。

2 同時通訳データ

前節でも述べたとおり対象とした“独話”は、NHKで放送されている番組「あすを読む」¹である。放送された番組の録画ビデオに対して、同時通訳専門の通訳者に同時通訳をしてもらった。同時通訳データは、解説者の原音声と通訳者の音声を左右2チャンネルに分けて収録しており、両者の発話の時間的な関係を見る

ことができる。

本節では、我々が収集している同時通訳データについて述べる。

2.1 対象データ

「あすを読む」は、1人の解説委員が、主に時事問題について解説をおこなっている番組であり、²話し言葉としての“独話”、特に講演としての特徴が見受けられる。番組での解説のために、原稿は作成されるが、それを単に読み上げているものではない。解説委員の発話についての詳細な特徴分析は、文献 [2] に示されている。

2.2 同時通訳データの収集

今回収録した同時通訳データは、ビデオに対して同時通訳を行ったものである。1人の通訳者に、「あすを読む」の録画ビデオを見てもらい、通訳するという形態をとった。通常、講演や会議の同時通訳 [3] では、2,3名で1つのグループとなり、グループ単位で、半日、あるいは、1日の同時通訳業務を行う。グループで業務を行うのは、実際に同時通訳の発話をしている通訳者のサポートを行うためである。サポートとしては、聞き取り難い単語を筆記したり、対訳を調べたりという直接、通訳につながるものから、手元にはない資料が配布された場合など、早急に資料を取り寄せたり、飲み物の確保まで多様である。

今回の同時通訳データ収集では、このようなグループを組む形態はとらず、他からのサポートは一切おこなっていない。1番組が10分と比較的短く、テレビの視聴者に、資料が配布される訳ではないので、手元の資料や配布資料についてのサポートは必要ないと判断した。また、咄嗟に対訳の出にくいと思われる専門用語や話題をつかむ上での重要語については、予め我々がキーワードリストを作成し、各番組の同時通訳の収録に際して通訳者に、事前に渡した。事前に渡していたキーワードリストの例を表1に示す。

また、1番組の収録が終るまでは、休憩をはさまず収

¹1曜日から金曜まで毎日1番組10分で放送されている解説番組

²ATRではNHKとの共同研究により使用許諾を得ている。

表 1: 番組「定期借家制度導入へ」のキーワード例

定期借家 家賃滞納 法務省 老朽借家 正当事由制度	賃貸アパート 立ち退き料 不動産市場 貸し渋り 遮二無二	貸事務所 議員提案 住宅建設市場 既存契約	明け渡し 新規契約 活性化 住宅事情	借地・借家法 中途解約 弱者切り捨て 住宅政策の貧困	正当事由 定期借地権 推進 福祉政策の歪み
---------------------------------------	--	--------------------------------	-----------------------------	-------------------------------------	--------------------------------

録している。同時通訳収録中は、ビデオの一時停止も禁止した。“巻戻してもう一度”ということも許していない。通訳者にもよるが、ほとんどの場合3番組ないし5番組を連続して通訳し、10分程度の休憩を取り、次の3番組ないし5番組を取るという形態で進められている。通常、同時通訳を行う場合にも、1時間程度で通訳者の交代をしており、そのペースとほぼ同じである。約半日で10番組程度の収録となる。

表 2: 番組数と語彙の関連（日本語）

番組数	語彙数（異り）
1	475
5	1561
10	2482
20	3939
50	6731

2.3 同時通訳者のアンケート

収録後、通訳者にアンケートをとった。通訳者にとって解説者の発話速度がかなり速く、難しいという印象がかなり見受けられた。専門の解説者がテレビ放送の番組で話しているためと考えられる。ただし、興味あるコメントとして、発話速度の問題ではなく、解説者の発話スタイルの問題として捉えているものがあつた。それは、解説者が内容を理解して自分の言葉で話しているか否か、つまり、原稿を丸暗記した読み上げ的な発声であるのか自発的な発声であるかにより、難しさが変わるという意見である。

また、通訳者の専門としている分野と異なるトピックを取り扱っている番組の場合、難しいと感じており、通訳者として納得できる訳出が出来ているわけではない。

3 同時通訳データの分析

本節では、50番組分の同時通訳データについて分析を行った結果について述べる。

3.1 文数、文長、語彙数など

解説者の原発話は、一番組に平均で約2000形態素、約60文（一文に平均約33形態素）であつた³。1形態素の短い文から、178形態素からなる長い文も見られた。文の認定は、書き起こし作業者による判断である。一番組で利用されている語彙は平均475語であり、その語彙数は番組数により表2のような増加をしている。

³形態素の単位はATRで開発した翻訳システムTDMTに依存した体系を利用している。

表 3: 原発話と通訳の発話との文数の比較

番組ID	1	2	3	4	5
原発話の文数	71	67	85	59	47
通訳者の発話文数	79	67	86	85	83

一方、通訳者の発話は、一番組あたり平均で1135形態素、70文であり、利用している語彙は平均416であつた。

通訳者の発話文数と解説者の原発話の文数とを比較（表3）すると、平均的には余り差がないようにみられるが、どちらかと言えば通訳者の発話文数が多い。また、1番組あたりの形態素数は、解説者に比べ通訳者の方が少くなっている。このことから、原発話では1文で表現されていることを、通訳者は複数の文で表現しつつ、ある程度の情報を削除していると思われる⁴。

また、間投詞、フィラー（解説者の日本語では、「えー」「ま」など、同時通訳者の英語では「e」「ah」など）は、1番組あたり、解説者は平均120回、通訳者は平均164回出現していた。

3.2 翻訳開始のタイミング

収集した同時通訳データでは、原発話と通訳の発話の時間的なズレを見ることができる。しかし、通訳の発話には、非文法的な表現や省略、語順の入れ替わりなど話し言葉の特徴が顕著にあらわれており、発話内容の正確な対応を取ることは非常に困難である。まし

⁴但しこの削除される割合は、今回の収録方法に依存したものと考えられ、本来の同時通訳で、同じ状況になるとは限らない

て、文ごとの対応をとることはできない。そこで、以下の2点に着目して分析した。

1. 同時通訳の発話を開始するタイミングと同時通訳者のポーズの関係
2. 予め与えたキーワードの訳語の出現するタイミング

3.2.1 ポーズによる訳出のタイミング

まず、通訳者がどのようなタイミングで原発話に対する通訳を開始するかは、通訳者が話しはじめる時間に、原発話がどのような表現であるかを調べることで、推測することが可能とおもわれる⁵。そこで、通訳者がポーズを置いて話しはじめている時刻の原発話を分析した。この分析では、原発話の単語に着目し、同時通訳者が発話をはじめるトリガーとして、原発話の品詞の特徴を調べた。

表 4: 同時通訳者の発話開始と原発話の関係

通訳者のすべてのポーズ		8895
ポーズに関連		
直前がポーズ	1120	3087
そのタイミングがポーズ	1614	
直後がポーズ	353	
動詞に関連		
直前が動詞	1129	2905
そのタイミングが動詞	1243	
直後が動詞	533	
その他		2903
通訳者の1秒以上のポーズ		1151
ポーズに関連		
直前がポーズ	133	373
そのタイミングがポーズ	56	
直後がポーズ	183	
動詞に関連		
直前が動詞	152	382
そのタイミングが動詞	157	
直後が動詞	73	
その他		397

単語あるいは音素単位の詳細な時刻データは、書き起こしテキストと音声データとのアラインメントを取ればできるが、簡略化して、書き起こしテキストを全て平仮名に置き換え、全ての仮名が同じ長さであると仮定して、時刻データを得た。

また、実際に同時通訳者が発話を開始するトリガー

⁵ただし、通訳者自身も発話しており、通訳者自身の持つ発話のリズムもあるため、あくまで推測ということになる。

となる単語は、発話開始時刻より少し前の単語⁶、あるいは、その単語に隣接して直後に発声されている単語を推定していると考え、それらの品詞を調べた(表4)。その結果、同時通訳者の全てのポーズのあと、発話を開始するトリガーとなっているのは、全体の34.7%が原発話者のポーズに関連するタイミングで、32.7%が動詞に関連するタイミングで行われているという結果が得られた。残りの32.6%には、名詞、サ変名詞、他の用言、外来語(片仮名表記されるような語)が、含まれていた。

また、同時通訳者のポーズが一秒以上ある場合は、発話のリズムから生じるポーズというより、原発話を理解、あるいは、翻訳するのに必要な原発話を待って発話していると考えられる。この場合でも、原発話のポーズに関連するタイミングで通訳をはじめる割合が32.4%、原発話の動詞に関連するタイミングで通訳をはじめる割合が33.2%、その他が34.4%となっている。

より詳細な分析が必要であるが、「原発話のポーズ」あるいは「いくつかの動詞に関連する発話」が、「あるまとまった内容を持つ原発話の部分」の区切りとみなせる。この区切られた部分は、節ということができる。つまり、「同時通訳者が通訳を開始するのは、原発話があるまとまった内容を持つ節として理解できる部分をトリガーとする」と考えることができる。

3.2.2 キーワードの訳出タイミング

次に、通訳者に提示しているキーワードに対応する訳語がどれくらいおくらせて、同時通訳に出現しているかを調べた。その結果を表5に示す。

時刻情報として、キーワードの含まれる発声単位(ポーズ、あるいは、間投詞等により区切った単位)の開始時間を比較している。この時刻情報では、発話の開始時刻に5秒前後のずれが見られた。長いものでは、10秒程度のずれが見られるが、原発話の終了時刻と通訳の開始時刻を比べると、2秒程度であった。我々が予め用意したキーワードについては、そのほとんどが訳出されており、出現する時刻のずれは、長くて10秒であることから、通常、同時通訳において重要な語句は10秒以内に訳出されると考えられる。

4 同時通訳者の翻訳単位

同時通訳的な翻訳システムを実現するためには、翻訳の出力に対して、追従性、正確さ、理解しやすさなどが求められる。特に、追従性という観点では、同時通訳的

⁶聴覚の認識時間を考慮する必要がある。認識時間については、幾つかの研究報告がなされているとおもわれるが、この分析では、もとの時刻情報も大雑把な推定であることから、200msの遅延があると仮定している。

表 5: キーワード対応リスト (例)

/定期借家/fixed term rental housing/

J:2990 - 6477 皆さんは定期借家という制度をご存じでしょうか。

E:7996 - 10099 fixed term rental housing?/

/定期借家/fixed term rental housing/

J:25864 - 31810 定期借家という制度の内容、あるいはこの問題、課題について今夜は考えてまいりたいというふうに思います。

E:28672 - 31206 the program, fixed term rental housing

/賃貸アパート/rental houses/

J:7231 - 9263 賃貸アパート、貸事務所など

E:11225 - 15478 for rental offices or rental houses at the expiry of the contract

な翻訳システムに固有の特徴であり重要である。前節の分析から、同時通訳者は「あるまとまった内容を持つ部分」を聞き取った時点で通訳をはじめている。この「まとまった内容を持つ部分」が、同時通訳者が翻訳を行う単位であり、“文”ではなく“節”であるといえる。独話では、3.1 節で示したように一文が従属節を伴い非常に長くなる傾向がある。翻訳システムの追従性を高めるには、翻訳の単位として、“文”という単位ではなく“節”の単位が適切といえる。

また、解説者、通訳者の発声文数(書き起こし時点で認識された文数)を比べると、その間には大きな隔はなく、ほぼ同数、あるいは、通訳の文の方が少し多いぐらいである。このことから、通訳者は、文をまとめて訳出しているというよりは、節毎に分割し訳出していると考えられる。

現状の多くの翻訳システムでは、一文を処理単位としており、入力にも、文として整った形態を期待している。しかし、文として整った形態を入力するためには、講演などの独話では、一文が長いために時間的なズレが大きくなり、追従性が大きく損われてしまう。これを回避するために、一形態素ごとに処理を行うシステムも提案されている[4]。このシステムでは、処理機構の漸進性を重視しており、対象は対話文となっている。以上のことから、頻繁に長い文が現れる特徴をもつ独話などを対象とした翻訳システムには、随時、入力列を見ながら、適切な節となる部分で分割し、翻訳する機構が望まれる。

5 まとめ

本稿では、ニュースや講演を対象とした“独話”の音声翻訳システムの基礎技術の研究として、同時通訳的な技術に焦点をあて、実際に、同時通訳データを収集

し、同時通訳のタイミングについて分析した。同時通訳データは、NHKの番組「あすを読む」を対象に、その収集の方法と語彙数、文数などの特徴について述べた。また、同時通訳の発話のトリガーを、同時通訳者のポーズを手掛りに分析し、キーワードの対応を調べることで、同時通訳的な翻訳システムの処理単位として、“節”の単位の可能性が大きいことを示した。

今後は、より正確に時刻情報を取り出し、データの分析を進めるとともに、実験用の翻訳システムを構築し、節を翻訳単位とした時の利点、問題点をあらいだし、同時通訳的な技術の確立をめざす。また、同時通訳的な翻訳システムの目指すべきシステムと同時通訳者との比較を通して、情報の削除、追加のメカニズムを提案していきたい。さらに、翻訳結果の評価についても、単なる訳出された1文としての翻訳評価だけでなく、全体としての理解しやすさの指標などを盛り込んだ評価基準の確立を目指す。

参考文献

- [1] 菅谷、竹澤、横尾、山本: “ATR-MATRIX と人間との音声翻訳能力比較実験”、(日本音響学会春季研究発表会、2000 年)
- [2] 丸山、熊野、柏岡: “日本語における独話の特徴と文分割”、(言語処理学会 第 7 回年次大会、2001 年)
- [3] 実吉典子: 「通訳」、(実業之日本社、1992 年)
- [4] 渡辺、松原、外山、稲垣: “英日同時翻訳のための漸進的日本語生成”、(言語処理学会 第 6 回年次大会、2000 年)