

講演同時通訳データのライメント

柏岡 秀紀

ATR 音声言語コミュニケーション研究所

1 はじめに

現在、音声翻訳の技術は、主に旅行対話を対象として、実用可能な技術になりつつある [1]。この音声翻訳技術を適用する課題の拡張を考えると、講演やニュース等の通訳が重要な課題と考えられる。

講演の通訳では、講演者の発話とほぼ同時に訳出していく「同時通訳」が一つの典型的なシステムのモデルと考えられる。このようなモデルでは、同時性を保つために発声の早い段階から訳出していく技術 [4] が求められる。旅行対話では、翻訳対象となる一文が比較的短く、発話毎に聞き手により授受される情報が解釈され、不明な点は聞き直しが可能なため、発話毎に適切に訳出することで、同時性を確保しつつ、対話の目的を達成することができる。これに対して、講演やニュース等では、従来からの翻訳処理の単位である文の境界を判定することが難しく、また、一文が比較的長いことが知られている [2]。このことから、文を翻訳処理の単位とするシステムモデルでは、比較的長い一文の発声時間の訳出の時間的な遅延が生じ、同時性を保つことが困難である。さらに、翻訳処理精度と文の長さとの関係を見ると、比較的短い文では質の良い訳出が可能となってきているが、長い文の訳出は多くの問題を抱えている。つまり、講演の音声翻訳システムにおいては、まず、翻訳処理の単位となる文の長さが、同時性(追従性)の面と解析精度の面から大きな課題となる。

本稿では、講演のデータとして NHK の解説番組である「あすを読む」を用い、収録された同時通訳のデータを利用して、原発話と通訳発話の対応を取り、その時間的なズレや対応の取れる単位について考察を加える。

2 同時通訳データ

本稿では、NHK の解説番組「あすを読む」¹を講演データとして利用した。本節では、我々が収集した同時通訳データについて述べる。

¹月曜から金曜まで毎日1番組10分で放送されている解説番組

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行っている番組である。解説委員は、番組での解説のために原稿を作成しているが、単に読み上げているのではなく、言いづまり、言い誤りなどの話し言葉に特有の現象が観察される。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

同時通訳の収録は、録画されたビデオに対して行ったものである。収録では、通訳者1人が「あすを読む」の録画ビデオを見ながら同時通訳をするという形態をとった。収録した同時通訳のデータは解説者の原音声および通訳者の音声を左右2チャンネルに分けて、DATにより収録した。これにより、両者の発話の時間的な関係を見ることができる。

通常、講演や会議の同時通訳では、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 [3]。我々の行った同時通訳データの収録では、このようなグループを組む形態はとらず、通訳者一人で同時通訳を行い、他からのサポートは一切行っていない。一般に訳出が困難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、予めキーワードリストを作成し、通訳者に事前に渡している。

また、1番組の収録が終わるまでは、休憩をはさまず収録している。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た。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、そのペースとはほぼ同じである。約半日で10番組程度の収録となる。

収録されたデータに対して、原発話の日本語、通訳発話の英語共に、ポーズ等により発話をフレーズに分割し、時刻情報を付与した書き起こしを行った。

また、原発話の日本語に対しては、ATRで開発された旅行対話翻訳システム TDMT で利用されている形態

素体系による形態素情報を付与した。さらに、ATRで開発されている音声認識システム SPREC のモジュールを利用し、音声データと書き起こしデータの音素アライメントを取り、形態素単位での時刻情報も付与した。通訳発話の英語に対しても、ポーズ等によりフレーズに分割し、各々のフレーズに時刻情報を付与した。

3 アライメント

一般的なテキストにおける文アライメントでは、原言語と目的言語の文の対応関係を推定し、アライメントされたデータは、翻訳知識などの獲得に利用される。対応関係の推定には、各文に含まれる単語の対応関係やテキスト内の文の出現順序などの情報が利用される。

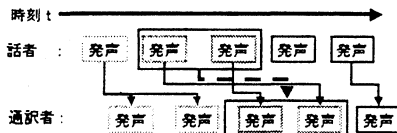


図 1: 同時通訳のアライメント

同時通訳データでのアライメントは、図 1 に示すような対応を取ることである。この時、テキストのアライメントにおける出現順序の制約は、時間的により強い制約として設定することが考えられる [5]。以下に本手法で利用する制約を示す。

- 語彙的制約
「通訳発話に含まれる単語は、原発話に含まれる単語と対応関係を持つ」ということを前提とする。
- 時間的制約
通訳者の発話は、原発話を通訳者が聞いた後に訳出しているので、発話時刻の順序関係として、対応する原発話に先行しない。また、話題の展開等に追従しなければならぬために、ある一定時間以上遅れて発話されることは非常に稀であると考えられる。

本稿では、アライメントをとる単位として、原発話、通訳発話共に、ポーズおよび間投詞により分割したフレーズ単位を用いる。

語彙的制約

原発話のフレーズ単位に対応する通訳発話のフレーズ単位は、対応する可能性のある通訳発話の単位の中

で最も多くの対応する対訳を含むフレーズ単位であると考え、以下の式により判断する。

$$\text{Pair(原発話単位、通訳発話単位)} = \text{argmax}_{\text{通訳発話単位}} F(\text{原発話単位、通訳発話単位})$$

$F(\text{原発話単位、通訳発話単位})$ は、対応する単語数とする。しかし、単純な単語数では、一つの見出し語に対して、多く数え上げてしまうことがある。そのため、原文の見出し単語の訳語が複数の通訳フレーズにある場合、1 個の単語数を各々に等分に分配し、 $1/$ 「見付かった対応する単語数」だけカウントする。

例えば、原発話のフレーズ phj に現れる日本語の見出し J1 に対して、通訳発話のフレーズ phea に J1 の訳語 E1、pheb に J1 の訳語 E2 の二つが訳語として対応しているとする。E1 は、e11、e12 の 2 語から、E2 は、e21、e22、e23 の 3 語から構成される複合語である。
日本語見出し 英語見出し (構成単語)

$$J1 \rightarrow E1(e11 e12) \\ E2(e21 e22 e23)$$

ただし、英語のフレーズ phea に、E1 を構成する単語の内、e11 のみが、また、pheb には、E2 を構成する単語の内、e21、e22 のみが含まれている場合、

$$F(\text{phj, phea}) = 1/(2 * 2) \\ (2 \text{ 種類の訳語、2 単語の一方のみ対応})$$

$$F(\text{phj, pheb}) = 2/(2 * 3) \\ (2 \text{ 種類の訳語、3 単語のうち 2 単語が対応})$$

となり、それぞれ、 $1/4$ 、 $1/3$ の対応があると考えられる。

また、原発話及び通訳発話に含まれる全ての単語について対応関係を調べることは効率的ではないため、原発話の内容語に限定して対応関係を調べた。

時間的制約

同時通訳において通訳者は、原発話に追従した発話を行うために、原発話と対応する通訳発話は、図 2 あるいは図 3 のような関係にあると考えられる。

図 2 は、原発話の開始より後、原発話の終了より前に、通訳発話を開始している状態である。このように対応する発話単位が重複している時は、図中の T2 の値が、一定値を下回ると仮定した。これは、原発話を聞いてから一定時間内に通訳発話を行うという前提から、重なっている部分に含まれる原発話の単語は、T3、T4 に含まれる部分で訳出されるものであり、訳出に際して予測の効く、あるいはほとんど訳出に影響のない単語で、時間的にも非常に短時間であると考えたためである。

一方、図 3 は、重複の無い訳出であるが、同時通訳の追従性の確保を考慮すると、図中の T2 の値は、一定値を下回ると仮定する。

上記二つの条件は、以下の式で考慮できる。

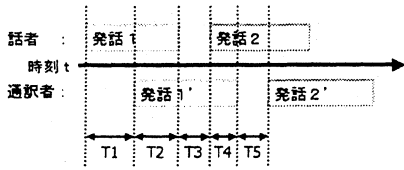


図 2: 発話時刻の対応関係 (重複あり)

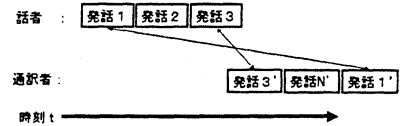


図 4: 対応をまとめる関係

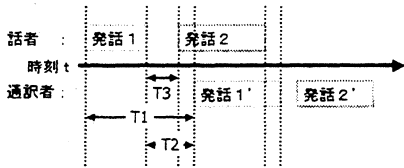


図 3: 発話時刻の対応関係 (重複なし)

$DT1 < \text{通訳発話の開始時刻} - \text{原発話の終了時刻} < DT2$
 $DT1$ は負の数で図 2 で示される制約、 $DT2$ は正の数で図 3 で示される制約を示す変数といえる。

アライメント実験

アライメントの対象には、2 節で示した「あすを読む」の同時通訳データ 48 番組を用いた。

語彙の対応を取るために、市販の和英辞書 (学研 ニューアンカー和英辞典)[7] から取り出した訳語対を利用した。見出し語と対応する訳語とを抽出し、ATR で開発した翻訳システム TDMT で利用している形態素体系に半自動的に変換したものを利用した [6]。抽出した見出し語数は約 3 万語であり、一つの見出し語に対して平均約 1.7 個の対訳が記述されている。全ての対応をみると、約 5 万対訳を有する対訳辞書で調べたことになる。

しかし、実際にアライメントに利用出来た語数は、非常に少なく表 1 のようになった。

表 1: 利用した語数

	全体	／番組	割合
検索対象とした単語数	37730	786	
辞書引き成功数	29714	619	79%
対応単語数	6446	134	17%

本稿で示すアライメントでは、全ての発話を対応付けることはせず、対応の取れる部分のみの対応付けを行った。これは、同時通訳では、原発話に追従するため、通

訳者が重要でない判断される内容については、時間的な追従性を考慮して省略したり、言い換えてしまうことがあり、全ての発話に対して対応する発話を見つけることができないと判断した為である。

また、ポーズ等により分割された原発話の単位と通訳発話の単位が、同じ内容ごとに分割されていることは期待できない。そこで、通訳発話の単位は複数の原発話の単位に対応することを許し、対応関係を取った。

上記の制約に基き、時間制約の許す範囲の全ての対応関係の中から、語彙的制約により、最も単語の対応関係の取れた発話単位の対に対応する発話単位と考える。ただし、後処理として、対応する発話単位間で、図 4 に示す様な時間的にクロスしている場合は、原発話の、発話 1、発話 2、発話 3 をまとめた発話と、通訳発話の、発話 3'、発話 N'、発話 1' をまとめた発話が対応の取れた発話とした。

4 対応の分析

ここでは、同時通訳データ「あすを読む」48 番組を対象に、前節で示した制約を利用したアライメントの結果について分析を加える。実験では、時間的制約として、 $DT1$ を -1 秒とし、 $DT2$ を 10 秒と設定している。言い換えれば、フレーズの重なりを 1 秒まで許し、発話開始の遅延を 10 秒まで認める条件下での対応である。48 番組全体で、ポーズ等による原発話のフレーズ単位は 17,065 個あり、その中で対応のとれたフレーズ単位は 5,218 個であった。

対応付けられた単位に対して、表 2 に示す基準による評価を行った。評価結果を表 3 に示す²。

この結果から、フレーズ単位で“○”となる適切な対応を取ることは困難であるが、“-”や“+”のような部分的な対応を取り出すには役立つことが解る。適切と判断された対応において、原発話の平均の長さは 7.8 単語、2.0 秒に対して対応する通訳発話の平均の長さは 7.1 単語、2.5 秒であった。

²対応の取れたペア毎のカウンタであり、フレーズ単位数とは異なる。

表 2: 評価判断基準

○	過不足のない適切な対応関係にある
+	通訳発話に対応する原発話に無い情報が含まれている
-	通訳発話に対応する原発話の一部の情報しか含まれてない
×	全く対応していない

表 3: 評価結果

○	+	-	×
870	619	1782	785

また、原発話と通訳発話の開始の時間的なズレは平均 3.9 秒、原発話の終了と通訳発話の開始の間には平均 1.8 秒の遅れがあった。また、“○”と評価された対応の中で重複のある発話は 143 個、重複のない発話は 727 個であった。

図 2 あるいは図 3 に示したように、通訳発話が、原発話に重なる場合と遅れて間の空く場合がある。最も大きな重なりは、1 秒(制約による)であり、最も大きな間の空きは、9.3 秒(ほぼ制約どおり)であった。平均的には、1.8 秒の間が空いている。ただし、これらの値は、時間制約に用いた DT1、DT2 の値の影響を大きく受けている。より深い考察は、時間制約による DT1、DT2 の値を変化させた実験結果を見なければならぬが、“通訳発話の開始時刻 - 原発話の終了時刻”を 0.5 秒毎で頻度をみると図 5 のようになる。

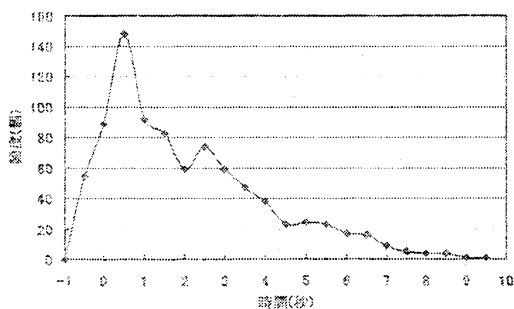


図 5: 時間のズレ

5 まとめ

本稿では、講演として「あすを読む」48 番組を利用して、市販の辞書を利用し、同時通訳のデータのアラ

イメントを行った。その結果、市販の辞書では、アライメントを精度よく行えるだけの単語対応を十分に取ることができなかったが、一定数の対応は得ることができた。対象とした「あすを読む」が時事問題の解説という番組の性格上、出現する語彙が広範囲に広がっており、通訳者の訳語も直接的な対訳を利用していないことが、高精度でアライメントを行えるだけの単語対応を取れなかった一因であろう。

対応の取れた範囲では、1 秒程度の遅延で訳出が始まっており、対応の取れたフレーズの長さは、日本語、英語ともに 7 語程度であった。

今後、対訳辞書を拡充し、より精度の高い対応付けを目指すとともに、対応付けされたフレーズ単位と、同時通訳者の翻訳単位との関係を調べ、翻訳システムとの応用について検討していきたい。また、対応の取れる単位の意味的なまとまり(主語、節、等のまとまり)との関係についても調べていくつもりである。

謝辞

本研究は通信・放送機構の研究委託「大規模コーパス音声対話翻訳技術の研究開発」により実施したものである。

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).
- [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, pp.429-432, (2001).
- [3] 実吉典子 「通訳」 実業之日本社, (1992).
- [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, pp.272-275, (2000).
- [5] 高木, 松原, 稲垣 “同時通訳コーパスにおける発話対応関係の推定”, 情報処理学会第 63 回(平成 13 年後期) 全国大会, pp.2-249 - 2-250, (2001)
- [6] 鷹尾, 下畑, 今村, 柏岡 “和英/英和辞書のカバレッジの比較とその応用”, 言語処理学会第 7 回年次大会, pp.58-61, (2001).
- [7] “ニューアンカー和英辞典データベース”, 学習研究社, (1999).