

同時通訳コーパスを用いた通訳者発話の話速変動の分析

于海貝^{*}, 笠浩一郎, 松原茂樹 (名古屋大学)

A Corpus-based Analysis of Simultaneous Interpreter's Utterance Speed

Haibei Yu, Koichiro Ryu, Shigeki Matubara (Nagoya University)

1 はじめに

同時通訳システムの実現においては、訳出タイミングが重要となる。小野らは、同時通訳者の訳出遅延について詳細な分析を与えているが [1], 発声速度に関する知見についてはこれまでほとんど解明されていない。そこで本稿では、同時通訳データを用いて、通訳者の発話速度の変動を分析する。講演者の発話中と発話後との間で通訳者の発話速度を比較する。

2 同時通訳データベース

名古屋大学同時通訳データベース [2] では、講演の同時通訳音声データが収録されている。音声は、200 ミリ秒以上のポーズで単位分割され、各発話単位には、発話 ID, 開始・終了時間, 対訳対応情報が付与されている。講演者発話と通訳者発話の対訳対応が与えられたデータの例を Fig. 1 に示す。

3 形態素の時間情報の付与

発話単位の音声データと書き起こし文に対して、以下の手順で形態素単位の時間情報を付与する。

- 形態素解析器「茶筌」[3] を用いて書き起こし文を形態素に分割し、形態素の境界と各形態素の読みを求める。
- 音声認識エンジン Julius[4] を用いて、形態素とその音声信号区間とを対応付ける。
- 形態素の発話開始・終了時刻を、対応する音声信号の時間情報から推定する。

本手法が付与した時刻と人手で付与した時刻の誤差が 120 ミリ秒以下を正解とすると、正解率は 99.2%であった。

4 通訳者の発話速度の分析

同時通訳者は講演者の発話状況に応じて、発話速度を柔軟に制御していると思われる。特に、講演者の発話が完了する前と後とでは、訳出内容が定まっていなかったり、あるいは、定まっているという違いがあり、それは発話速度に影響することが予想される。そこで、通訳者の発話中と発話後という 2 つの状態に着目し、同時通訳コーパスを用いて、その発話速度を比較した。Fig. 2 に、対訳対応関係にある講演者発話と通訳者発話の時間的関係を示す。この例では、通訳者発話のうち、00:12:056 から 00:12:592 までは講演者が発話中、それ以降は発話後であり、この場合、両者の話速が比較対象となる。

分析には、名古屋大学同時通訳データベースに収録されている英日通訳者の発話データを用いた。データ量は 660.35 秒

	講演者発話	対訳対応	通訳者発話
0	001-00:05:120-00:08:204 N:For my final topic I will speak about		001-00:06:328-00:09:304 I:私の最後のお話として
1	002-00:08:616-00:12:592 N:the little differences between Tokyo and Kansai		002-00:12:056-00:16:255 I:東京と関西の違いについてお話をしたいと思います

Fig. 1: 対訳対応済みの同時通訳データの例

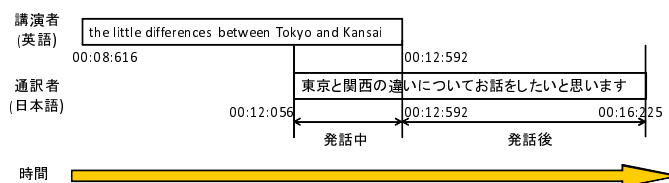


Fig. 2: 同時通訳の模式図

Table 1: 発話速度の比較

	発話中	発話後
モーラ数	701	6313
発話時間 (s)	86.99	573.36
話速平均 (mora/s)	8.05	11.01

の発話であり、発話単位数にして 483, 形態素数にして 3100, モーラ数にして 7014mora に相当する。モーラは、形態素解析器 [3] の解析結果を利用して定めた。話速は、発話時間あたりのモーラ数として計算した。

分析結果を Table 1 に示す。発話中の話速は 8.05(mora/s) であり、発話後の話速 11.01(mora/s) に比べて 26.9%低下している。

5 おわりに

本稿では、名古屋大学同時通訳データベースを用いて、通訳者の発話速度の変動を分析した。講演者の発話中における通訳者の話速は、発話後の話速と比べ 26.9%低下することがわかった。今後は、通訳者の発話状態を細分化し、発話速度の推移をより詳細に分析する予定である。

文献

- [1] 小野他: 通訳研究, No.7, pp. 49-64, 2007
- [2] <http://sidb.el.itc.nagoya-u.ac.jp/>
- [3] <http://chasen.naist.jp/hiki/ChaSen/>
- [4] <http://julius.sourceforge.jp/>