

同時通訳コーパスにおける発話対応関係の推定

高木 亮†

松原 茂樹†

稲垣 康善†

†名古屋大学大学院工学研究科

†名古屋大学言語文化部

atakagi@inagaki.nuie.nagoya-u.ac.jp

1 はじめに

同時通訳は、人間の言語活動のなかでも極めて特殊かつ高度な言語行為であり、計算機による同時通訳の実現のために、同時通訳者が用いるテクニックを獲得し、活用することは有効な方法の一つである。このような観点に基づき、同時通訳音声の収集が行われており、これまでに、数十万語規模のデータベースの構築が進められている [4]。通訳テクニックを獲得するために、話者発話とその通訳者発話を細かく対応をとりながら観察する必要があり [1]、分析に用いる対訳コーパスが比較的小さな単位で対応づけられていることが望ましい。しかし、コーパスの規模が大きくなれば、それを人手で行うのは容易ではない。

本稿では、同時通訳コーパスにおけるフレーズの対応づけ手法について述べる。本手法では、対訳語の一致に関する語彙情報、ならびに、話者発話と通訳者発話の開始時間情報を利用して対応づけを行う。フレーズ対応関係の推定実験を通して、本手法の有効性を評価した。

2 同時通訳コーパス

名古屋大学統合音響情報研究拠点 (CIAIR) では、音情報処理技術の高度化を目指し、様々な音声・言語データベースの構築を進めている [5]。その一環として、同時通訳音声データを収集し、独話、及び対話音声の文字化作業を通して日英双方向の平行コーパスを作成している [4]。このコーパスでは、発話を 200ms 以上のポーズでフレーズに分割し、もしくは、50ms 以上 200 ms 未満のポーズでも文末であればそこでフレーズに分割している。さらに、フレーズには通し番号、発話開始時間、発話終了時間、及び話者 ID を付与している。図 1 に日本語独話とその同時通訳音声の書き起こしデータの一部を示す。左側が話者発話、右側が通訳者発話である。

3 フレーズ対応付け手法

本手法では、話者の発話と通訳者の発話との間のフレーズ対応関係を、語彙情報、及び時間情報を用いて推定する。 n 個のフレーズからなる発話 S と m 個のフレーズからなるその通訳発話 T を

$$S = s_1, s_2, \dots, s_n \quad (1)$$

$$T = t_1, t_2, \dots, t_m \quad (2)$$

Phrase Alignment of Simultaneous Interpreting Corpus: Akira Takagi, Shigeki Matsubara and Yasuyoshi Inagaki (Nagoya University)

0001 - 00:05:056-00:06:304 N: 英語講演のですね	0001 - 00:07:744-00:09:157 I: (F uh) The simultaneous
0002 - 00:06:768-00:07:984 N: 同時通訳システム	0002 - 00:09:358-00:10:944 I: (W a) translation system of
0003 - 00:08:432-00:09:488 N: いうことについて	0003 - 00:11:448-00:12:447 I: English lecture
0004 - 00:09:752-00:11:783 N: (F えー) 説明いたします (SB)	0004 - 00:12:896-00:14:895 I: is going to be elaborated (SB)
日本語話者発話	日英通訳者発話

図 1: CIAIR 同時通訳コーパスの書き起こしデータ (一部)

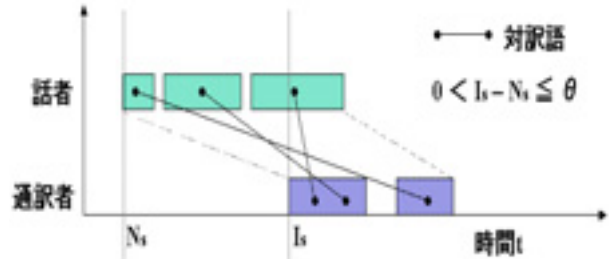


図 2: 発話の対応関係

とする。 S の連続する c 個のフレーズ $S_{a,c} = s_{a-c+1}, \dots, s_a$ が T の連続する d 個のフレーズ $T_{b,d} = t_{b-d+1}, \dots, t_b$ に対応するとき、その対応組 p を $\langle a, c; b, d \rangle$ で表す。 c または d が 0 である場合、すなわち、1 対 0、0 対 1 の対応も認める。発話全体におけるフレーズの対応関係を対応組の列 $P = p_1, p_2, \dots, p_k$ で表す。

$S_{a,c}$ と $T_{b,d}$ が対応しているとき、両者は以下に述べる 2 つの制約を充足していると考えられる (図 2 参照)。

時間的制約 通訳者は話者の発話を聞いてから、通訳を開始するので、通訳者の発話は話者の発話に先行しない。また、話者発話に追従するため、通訳者の発話開始時間は話者の発話開始時間より、ある一定時間 θ 以上遅れることはない。すなわち、

$$0 < I_s - N_s \leq \theta$$

N_s : 話者発話開始時間, I_s : 通訳者発話開始時間
語彙的制約 通訳者発話に含まれる単語は、話者発話に含まれる単語と対訳関係にある。

平行コーパスの文の対応関係を推定する手法 [2, 6] をもとに、ある対応組 p の対応度 $h(p)$ を式 (3) のように与える。

$$h(p) = \begin{cases} \frac{2f_{ST}}{f_S + f_T} & (p \text{ が時間的制約を満たす}) \\ 0 & (p \text{ が時間的制約を満たさない}) \end{cases} \quad (3)$$

$$f_S = (S_{a,c} \text{ に含まれる自立語数}) \quad (4)$$

$$f_T = (T_{b,d} \text{ に含まれる自立語数}) \quad (5)$$

$$f_{ST} = (S_{a,c} \text{ と } T_{b,d} \text{ での対訳語数}) \quad (6)$$

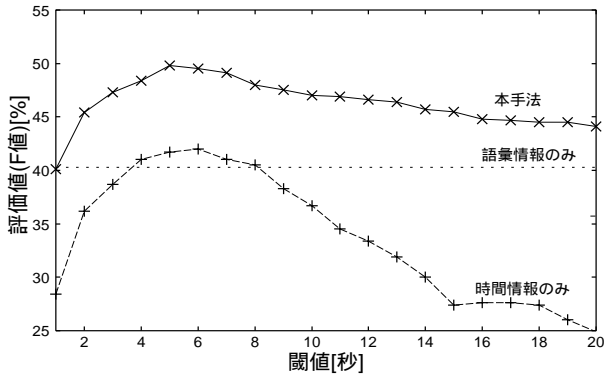


図 3: 16 講演平均の F 値と閾値の関係

表 1: 実験結果: 16 講演の評価値の平均 (閾値 = 5 秒)

手法	適合率(%)	再現率(%)	F (%)
本手法	67.7	39.8	49.8
語彙情報のみ	69.7	28.7	40.3
時間情報のみ	59.9	32.2	41.7

話者発話と通訳発話全体における対応 P の対応度 $H(P)$ を, P を構成する対応組の対応度の総和, すなわち

$$H(P) = \sum_{i=1}^k h(p_i) \quad (7)$$

とする. $H(P)$ が最大となるように対応組の列を動的計画法を用いて最適化する.

4 実験と評価

本手法の有効性を評価するために, 名古屋大学 CIAIR 同時通訳データベースに収録されている約 10 分間の日本語講演音声とその通訳音声からなるパラレル発話テキスト 16 組を用いて, フレーズ対応づけ実験を行った.

話者の発話開始時間に対する通訳者の発話開始時間の遅れの閾値を 1 秒間隔で設定し, 各閾値での対応づけ評価値 (F 値) を調べた. 比較のため, 語彙情報のみを用いた対応づけ方法, 及び, 時間情報のみを用いた対応づけ方法で実験を行った. 対応づけの評価には単語レベルでの評価方法 [3] を用いた. これは, フレーズに含まれる単語数によって対応度に重みをつけて評価をする方法である. なお, 正解フレーズ対応は手作業で作成し, 語彙情報として EDR 日英対訳辞書を用いた.

16 講演の閾値と F 値の平均の関係を図 3 に示す. 横軸は閾値の値 (秒), 縦軸は 16 講演の F 値の平均を示す. 図中の $F = 40.3\%$ の直線は, 語彙情報のみを用いた手法の結果を示す. F 値による評価では本手法が他の 2 つの評価よりも高い値を示しており, 語彙情報と時間情報を併用することの有効性が示された.

閾値が 5 秒のとき, F 値が 49.8% で最大となった. このときの適合率, 再現率, 及び F 値の平均値を表 1 に示す. しかし, あらゆるフレーズ対応に対して, 閾値を 5 秒で設定することが適切であるわけではなく, 実際, 話者に対する通訳開始時間の遅れは 16 講演平均で 4.1 秒である. 遅れ時間の分布を示す図 4 から分かるように, 発話開始時間の差

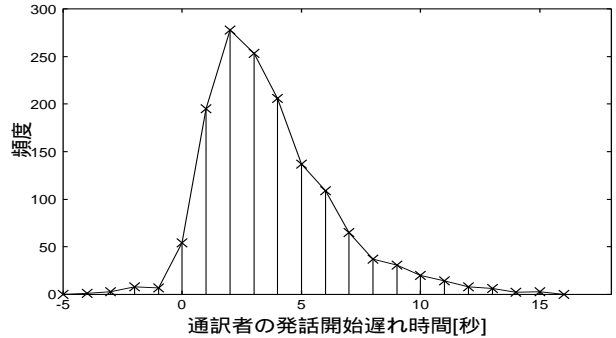


図 4: 通訳者の発話開始遅れ時間と頻度の関係

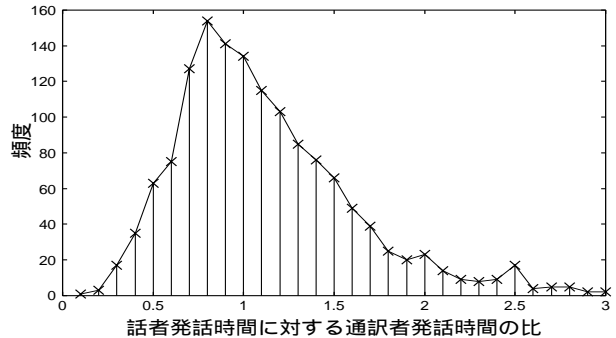


図 5: 話者発話時間に対する通訳者発話時間の比と頻度の関係

はフレーズ対応によってばらつきがある. 対応ごとの話者発話時間に対する通訳者発話時間の比の分布を図 5 に示す. すなわち, 発話時間の長さは対応する各フレーズで互いに関連する. 本手法では, 時間情報を対応づけに関する制約として利用したが, 時間情報を詳細に分析し, 活用することによって, より精度の高い対応づけが期待できる.

5 おわりに

本稿では, 同時通訳コーパスにおけるフレーズ間の対応づけの自動化手法について述べた. 今後はより細かな単位での対応づけを試みるつもりである.

参考文献

- [1] Y. Aizawa, S. Matsubara, N. Kawaguchi, K. Toyama, Y. Inagaki, "Spoken Language Corpus for Machine Interpretation Research", *ICSLP-2000*, Vol. III, pp. 398-401, (2000).
- [2] N. Collier, K. Ono, H. Hirakawa, "An Experiment in Hybrid Dictionary and Statistical Sentence Alignment", *COLING-ACL'98*, Vol. 1, pp. 268-274 (1998).
- [3] P. Langlais, M. Simard, J. Veronis, "Methods and Practical Issues in Evaluating Alignment Techniques", *COLING-ACL '98*, pp. 711-717 (1998).
- [4] 松原茂樹, 相澤靖之, 河口信夫, 外山勝彦, 稲垣康善, "音声翻訳研究のための話し言葉コーパスの構築", 情報処理学会第 61 回全国大会講演論文集 (2), pp. 181-182 (2000).
- [5] 武田一哉, 板倉文忠, "文部省 COE プログラム統合音響情報研究拠点 (CIAIR)- 音声・音響情報処理の多角的研究 -", 日本音響学会誌, Vol. 56, No. 11, pp. 748-751 (2000).
- [6] 宇津呂武仁, 松本裕治, "対訳辞書および統計情報を用いた二言語対訳テキスト照合", *コンピュータソフトウェア*, Vol. 12, No. 5, pp. 12-21 (1995).