# 独話音声の言語処理技術とその応用

松原 茂樹 (名古屋大学)

Spoken Monologue Processing and Its Applications

Shigeki Matsubara (Nagoya University)

## 1 まえがき

話し言葉 (*spoken language*) は、一人の話者のみが連続して話す「独話 (*monologue*)」と複数の話者が交替で話す「対話 (*dialogue*)」に分類できる。これまでの話し言葉処理に関する研究は、対話文を対象としたものがほとんどであり、応用として、音声対話や対話翻訳などのシステムの実現が検討されてきた。

それに対して本稿では、講演や講義などの独話を対象とした話し言葉処理技術について概説する。テキストと同様、独話には、人間の知識や意見などが内在しており、それをデータとして蓄積することにより、貴重な知的資源としての利用が可能となる。

## 2 独話音声の特徴

独話文は、対話文に比べ、1 文の長さが長く文の構造が複雑であるといった特徴がある[1]。そのような文に対して処理を実行すると、一般に、解析時間が長くなるとともに、高い処理精度の達成が難しくなる。また、対話にはターンという物理的に明確な区切りが存在するのに対して、独話では文の区切りを認識するのは難しいという問題がある。

独話の処理単位として「節 (*clause*)」が有効である。節とは、述語をまとまりとした言語的単位であり、例えば、独話文

- 先日総理府が発表いたしました世論調査によりますと死刑を支持するという人が八十パーセント近くになっております

は、「先日総理府が発表いたしました」、「世論調査によりますと」、「死刑を支持するという」、「人が八十パーセント近くになっております」の 4 つの節から構成される。節境界解析により検出された節境界ではさまれた単位を節境界単位とよぶ[2]。

## 3 独話音声の言語処理

**3.1 節境界単位に基づく独話文解析**　独話文を一つ以上の節の連接とし、また、各節の文節は、節の最終文節を除き節内の文節に係るとする。これを仮定することにより、独話文の解析を、節レベルと文レベルの二段階の解析として実行でき、解析の効率化が期待できる [4]。

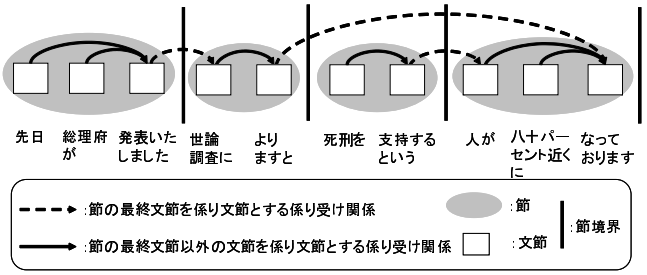


図 1: 節と係り受けの関係

この手法では、まず、独話文を節境界単位に分割する。分割には、節境界解析プログラム CBAP[2] を用いる。係り受け解析の手順は以下の通りである。

1. 節レベルの係り受け解析
   一文中のすべての節境界単位に対して、各節境界単位ごとにその内部の係り受け構造を解析する。
2. 文レベルの係り受け解析
   一文中のすべての節境界解析に対して、その最終文節の係り先を解析する。

NHK の解説番組「あすを読む」の文字化データに形態素解析、文節まとめ上げを施した 500 文を用いて実験した。節境界をまたぐ係り受け関係は 152 個存在した。一方、学習データとしては、5532 文を使用した。実験の結果、節境界分割を行わない文単位での解析と比較して、解析速度は平均して 3 倍向上し、分割による処理の効率化を確認した。係り受け正解率は 79.0%であり、文単位の解析 (76.3%) に優る解析精度を備えていることがわかった。

**3.2 独話文の同時的な解析**　同時通訳や字幕生成のように、独話を入力と同時に処理するようなシステムでは、話者による音声入力に従って順次、解析を実行する漸進的解析手法が必要である。節を解析単位とすることにより、独話の漸進的解析が可能となる [5]。

独話音声の漸進的解析では、節が入力されるたびにその節の内部の係り受け構造を作りあげればよく、最終文節の係り先のみ、適切なタイミングで決定を試みる必要がある。本研究では、係り先の決定を、後続するいくつかの文節との係り受けの尤度を考慮した動的なタイミングで実行する。これは、文節間の係り受け関係が文をまたぐことはなく、また、その距離が格段に長くなることはないという観察に基づいている。

「あすを読む」の 7 番組 (470 文) を用いて解析実験を

[1] もちろん、一文が長く複雑な構造をもった対話文も存在する。しかし、これまでの対話処理では、短くて単純な対話文からなる対話が処理対象として用いられてきた。

[2] 節には、主節に埋め込まれた従属節も存在するため、文を一次元的に分割することは一般には困難であるが、節の終端位置を検出することにより節に相当する単位を近似的に抽出することができる。

実施した。学習データとして、前節と同じ 5532 文を使用した。係り受け解析の正解率は、全体で 76.2%となった。全 7 番組の解析時間は 12.5 秒であり、本手法の実用性が示された。

## 4　独話音声の言語資源

独話音声研究を遂行するために、音声言語コーパス等の言語資源の整備が重要となる。そのような言語資源の一つとして、CIAIR 同時通訳データベースがあげられる [3]。これは、独話音声とその通訳音声を収集し、文字化したデータである。

**4.1　独話データの収録**　1 つの講演者発話ソースに対し、複数の通訳データを収録した。個人に特化しない通訳事例を幅広く収集したり、複数の通訳事例を比較することが可能であり、通訳経験年数の違いによる訳出の特徴分析などにも利用できる。20 時間余りの独話音声に対して、約 50 時間の通訳音声を収集している。

**4.2　収録音声のデータベース化**　音声データの文字化は、日本語話し言葉コーパス (CSJ)[1] の書き起こし基準に準拠した。以下に主な基準を列挙する。

- **発話単位**: 200msec 以上の無音（ポーズ）ではさまれた音声区間を発話単位とした。
- **タグ付与**: 発話単位に時間タグ（開始時刻、終了時刻）を付与した。また、話し言葉に特有の言語現象 (フィラー、言い淀みなど）に談話タグを与えた。

データの規模は、話者音声の約 17 万単語に対して、通訳音声が約 39 万単語となった。

音声データを視覚的に分析するために、話者と通訳者の発声タイミングを図式的に表示するツールを開発した (図 2 参照)。これにより、両発話の時間的重なりに関する様相を観察できる。また、話者音声と通訳者音声を対応づけるための作業支援環境を開発し、発話単位を最小単位とした細かいレベルでの対応付けを行っている (図 3 参照)。これらのデータは同時通訳者の訳出戦略の獲得などに利用している（例えば、[7])。

## 5　独話処理技術の応用

**5.1　音声翻訳**　独話音声の翻訳では、必然的に同時通訳の形態を採用することになる。音声言語文を単位とした逐次的な翻訳と異なり、話者発声途中のどの時点で訳出を開始するか（訳出タイミング)、また、どのようなフレーズを単位として翻訳処理を実行するか（通訳タイミング）といった問題に対処する必要がある。著者らは、これまでに漸進的な解析・変換・生成に基づき、動的なタイミングで処理を実行する同時翻訳機 *Linas* を開発しており [6]、簡単な独話データを用いた翻訳処理を試みている。

**5.2　音声要約**　講演や講義などの音声を再利用可能なデータして蓄積するためには、音声を要約するなどにより、適切に編集することが重要となる。その他に、リアルタイム字幕生成などを目的とした場合には、ユーザが追従可能な速度で文字を表示することが不可欠である。


図 2: 同時通訳音声の視覚化

| # | 講演者発話 | 通訳者発話 |
|---|---|---|
| 0 | 0001 – 00:05:264–00:09:399 N: The theme for this speech is going to be the American | 0001 – 00:06:440–00:08:207 I: (F え)次のテーマですが<br>0002 – 00:08:944–00:09:783 I: アメリカの |
| 1 | 0002 – 00:09:840–00:11:032 N: Presidential debate | 0003 – 00:10:296–00:12:775 I: (F え)大統領に関するディベート |
| 2 | 0003 – 00:11:424–00:13:391 N: and who would be the<br>0004 – 00:13:640–00:15:215 N: better president for America<SB> | 0004 – 00:13:096–00:14:424 I: そして誰が<br>0005 – 00:14:648–00:18:255 I: より良い大統領とアメリカのためになり得るかということですが |
| 3 | 0005 – 00:16:272–00:18:327 N: (F um) Let's see, today is | 0006 – 00:18:728–00:19:263 I: 今日が |
| 4 | 0006 – 00:18:640–00:20:400 N: December fifteenth | 0007 – 00:19:528–00:21:887 I: 十二月の十五日ですので |
| 5 | 0007 – 00:20:696–00:24:407 N: and it's been about a month and a half since | 0008 – 00:22:472–00:24:711 I: そして(F まあ)一ヶ月半ほど<br>0009 – 00:25:160–00:26:311 I: 経ってると思うんですが |

図 3: 同時通訳音声の対訳対応付け

そのために、音声を要約することが一つの方法であり、話者発声に対して同時的に処理する必要がある。

## 6　まとめ

本稿では、独話音声を対象とした話し言葉処理に関する研究動向として、解析技術、言語コーパス、及び、応用システムについて概観した。講演や講義など、独話には人間の知識や意見が含まれる極めて知的なコンテンツであり、それらを蓄積し、知識ベースとして整備することにより、共有可能な知的資産として利用できる。このような仕組みの実現に向けて、独話音声の言語処理技術の一層の高度化が望まれる。

**謝辞**　独話処理に関して日頃ご議論頂く、ATR の柏岡秀紀氏、名古屋大学の遠山仁美、笠浩一朗、大野誠寛の各氏に感謝いたします。本研究の一部は、総務省戦略的情報通信開発制度重点領域研究「講演など独話の知的構造化に関する研究開発」によります。

文献

(1) Maekawa, K., et al.: *LREC-2000*, 947-952 (2000).
(2) 丸山 ほか：自然言語処理, 11(3):39–68 (2004).
(3) Matsubara, S., et al.: *LREC-2002*, 1:153-159 (2002).
(4) Ohno, T., et al.: *COLING/ACL-2006*, To appear.
(5) Ohno, T., et al.: *Interspeech-2005*, 3449-3452.
(6) Ryu, K., et al.: *COLING/ACL-2006*, To appear.
(7) Tohyama, H., et al.: *LREC-2006*, 2564-2569 (2006).