

同時通訳システムの実現のための日本語-英語話し言葉翻訳に関する研究

代表研究者 松原茂樹 名古屋大学言語文化部助手
 共同研究者 稲垣康善 名古屋大学大学院工学研究科教授

はじめに

音声、及び言語処理技術の進展にともない、近年、話し言葉翻訳がますます重要な研究テーマとなっている。すでに、対話翻訳の実験システムがいくつか提案されており [3]、特定タスクドメインでの異言語間対話の実現可能性が明らかになりつつある。ただし、対象となる対話は、発話者だけでなく通訳者の発話権をも保障することが前提となっており、今後はより自然なクロスリンガルコミュニケーションを目指し、同時通訳技術を高度化することが望まれる。このような目的に対して、同時通訳音声データを収集し、分析することは、効果的な方法の一つである。

本報告では、我々が現在、構築している同時通訳コーパスについて述べる。このコーパスの特徴として、

- ・異なる言語間での対話だけでなく、講演通訳も収録した英語と日本語の平行コーパスである。
- ・同一の講演に対して経験度が異なる複数の通訳者を用意し、複数の講演通訳データを収録している。
- ・話者、及び通訳者の発話をポーズで分割し、各々を一発話として収録している。各発話には開始時間、終了時間を付与している。

などがあり、話し言葉翻訳技術の向上はもちろん、通訳会話の分析や通訳理論の構築のための基礎データとしても活用できる。

一方、日本語から英語への同時通訳装置の実現のために、実発話環境下において利用可能なロバストな日本語音声理解技術が不可欠である。特に、音声認識ソフトウェアの実用化が進んでいる現状においては、音声認識によって文字化されたテキスト、すなわち、トランスクリプトの解析が重要である。

そこで本報告では、日本語音声トランスクリプトの係り受け解析手法について述べる。本手法では、日本語文の係り受けに関する構文的制約として、係り先の唯一性に関する制約を緩和する。すなわち、フィラーや言い淀みなどの文節については、その係り先は存在しないとして解析し、解析結果を部分的な係り受け構造によって表す。これにより、音声認識エラーを含むテキストに対しても正しい係り受け関係を同定することができる。また本手法では、尤度の高い係り受け構造を獲得するために、音声対話コーパスに付与された係り受け情報を活用する。各文節間の係り受け確率を統計的に獲得し、尤度の低い係り受けで構成される構造を枝刈りすることによって正しい係り受けを取り出す。

本手法の評価のため、CIAIR 車内音声対話コーパス (CIAIR-HCC) [5] に収録されたドライバー発話200文の音声データに対して、係り受け解析実験を行った。実験では、コーパスの83対話から獲得した統計的係り受け情報を用いて、大語彙連続音声認識ソフトウェア Julius [6] によって生成されたトランスクリプトの係り受け解析を行った。その結果、自然な対話音声のロバストな解析における本手法の利用可能性を確認した。

2 同時通訳コーパスの構築

2.1 データの収集

我々は、対話だけでなく、通訳者を介した講演音声も収録している。コーパスの内容を表1に示す。円滑な異言語間コミュニケーションを実現する上で、通訳者の発声タイミングが重要である。話者と通訳者との間の発声の同時性が高いほど、

- ・独話では、聴衆が、通訳者の発声と話者の振舞いを結び付けて理解すること
- ・対話では、話者が、効率的で結束性の高いインタラクションを遂行すること

が可能となる [2]。同時通訳システムの実現のための基礎データの提供を目指し、同時通訳者を介した英語独話、日本語独話、及び英語日本語間対話を収録している。

独話では、英語話者の発声と通訳者による日本語発声、ならびに、日本語話者による発声と通訳者による英語発声、をそれぞれ収録している。同時通訳は、豊富な訓練を必要とする高等技術であり、通訳者の熟練度により、結果に大き

な違いが生じる。熟練度の影響を調査するために、同一の独話に対して経験年数が異なる複数の同時通訳者を用意し、その通訳結果を収録している。社会的にみて頻度の高いテーマ及び内容（表2参照）を採用した。話者には、原稿の読み上げでなく、できるだけ独話調で話すように依頼している。

表1 コーパスの内容

会話様式	講演, 対話
使用言語	英語, 日本語
発話様式	原稿あり発話, 模擬対話
通訳様式	同時通訳
データ様式	音声, 書き起こしテキスト

表2 講演のテーマと対話のトピック

講演	政治, 経済, 技術, 言語, 都市, 環境 など
対話	旅行 (入国審査, 空港, ホテル, 電話予約など)

一方、対話では、英語話者と日本語話者の通訳者を介した対話を収集している。通訳の品質を高めるために、英日及び日英の二人の同時通訳者を用いている。対話ドメインとして、これまでに構築されているいくつかの対話データベース [4] と同様、「旅行会話」を採用し、表2に示すように、空港やホテルなど、海外旅行において想定されるトピックをいくつか定めた。なお、対話様式は模擬的であるものの、できる限り自由な発話を収集するため、対話タスクと話者役割のみを事前に設定している。

実音響環境下での音声データを収集するために、教室レベルの録音環境で収録を行っている。同時通訳では、対象となる音声だけでなく、その発声者の表情や振舞いもまた、重要な情報となるため、通訳者は、話者をガラス越しに観察可能な専用のブースに入り通訳を行う。すべて同一のスタンドマイクを使用し、話者とその通訳者の音声は、サンプリング周波数 16kHz、16ビットでデジタル化し、デジタルオーディオテープに複数チャンネル環境で収録している。

2.2 データの分析

収集した音声データの文字化は、人手によって行っている。データの言語学的分析として、話し言葉の特徴的現象である、フィラー、言い淀み、言い誤り、言い直しなどにタグを付与している。現在までに文字化が完了しているデータ量を表3に示す。

同時通訳技術の実現にあたり、システムが生成する内容とその出力タイミングが重要なポイントである。原言語と目標言語との間の語の生起順序の違いのため、通訳文の品質を確保し、かつ、入力に対する出力の同時性を満たすことは難しい。我々は、通訳者が、どのような発話をどのようなタイミングで発声するのかを調査するため、話者、及び通訳者の発話をポーズで分割し、各々を発話単位と定め、その開始時間及び終了時間を記録している。

データの文字化作業で付与した発声時間情報をもとに、話者と通訳者の発声タイミングを視覚的に表示するツールを作成した。対話に対する音声の視覚化の例を図1に示す。図中、右側の発話が、左のグラフ内に記された時間に行われたことを表している。グラフの左から順に、英語話者、英日通訳者、日本語話者、日英通訳者の発声時間であり、話者と通訳者の発声の重なり具合がわかる。

表3 現在までに文字化されたデータの量

時間数 (延べ)	講演	40 時間 (通訳者: 24 時間)
	対話	32 時間 (通訳者: 16 時間)
	合計	72 時間 (通訳者: 40 時間)
総発話数	講演	13,553 文 (通訳者: 8,656 文)
	対話	8,676 文 (通訳者: 3,813 文)
	合計	22,229 文 (通訳者: 12,469 文)

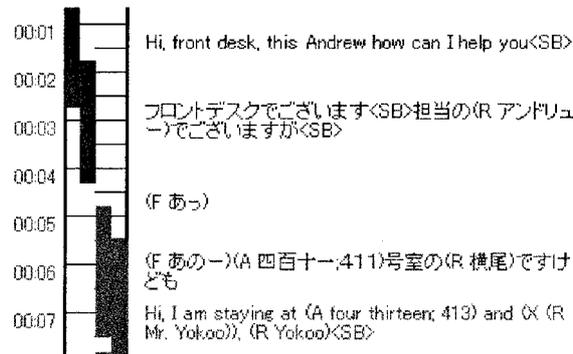


図1 対話音声と通訳音声の視覚化

3 日本語話し言葉の分析

日本語 - 英語話し言葉翻訳システムの実現のために、日本語音声を入力に解析するための方法が必要である。そこで本研究では、名古屋大学 CIAIR で構築されている音声コーパス (CIAIR-HCC) [5] を用いて話し言葉の特徴を分析し、自然な対話音声の解析方法について検討を行った。このコーパスでは、道案内や店情報検索などをタスクとするドライバーとナビゲータとの実走行環境下での対話音声を収録している。

3.1 音声言語コーパスの概要

CIAIR-HCC は、走行車内特有の言語現象や発話の重なり具合の分析を通して車内対話をモデル化するための基礎資料として活用することを想定し、作成されている。コーパスでは、収集の第一段階として「人間対人間」の対話を収録している。収集した音声データの書き起こし作業は、人手によって行っており、データの分析にあたっては、日本語話し言葉コーパス (CSJ) の音声書き起こし基準 [8] に準拠したタグ付けを行っている。データの言語学的分析として、フィラー、言い淀み、言い誤りなどにタグを付与するとともに、発話をポーズで分割し、各々を発話単位と定め、その開始時間及び終了時間を記録している。図2に書き起こしテキストの例を示す。各発話単位の開始・終了時間の右側に、性別 (男性/女性)、話者役割 (ドライバー/ナビゲータ)、対話タスク (道案内/情報検索など)、雑音状況 (有/無) に関する情報を付与している。

```

0001 00:01:543-00:10:148 M:DRIVER:
  ちょっと & チョット
  小腹<H>が & コバラ<H>ガ
  すいたんだけど<H> & スイタンダケド<H>
  この & コノ
  近くに & チカクニ
  ファーストフード店で<H> & ファーストフードテンテ<H>
  あるのかなあ<SB> & アルノカナア<SB>
0002 00:10:5683-00:13:969 F:DRIVER:
  はい & ハイ
  マクドナルドと & マクドナルドト
  モスバーガーが & モスバーガーガ
  ございますが<SB> & ゴザイマスガ<SB>
0003 00:14:156-00:17:905 M:DRIVER:
  (F あっ)じゃ & (F アッ)ジャ
  マクドナルドの & マクドナルドノ
  場所を & バシヨオ
  教えてくださいんだけど<SB> & オシエテホシンダケド<SB>
0004 00:18:092-00:21:136 F:DRIVER:
  はい & ハイ
  マクドナルドは & マクドナルドワ
  ドライブスルーされますか<H><SB> & ドライブスルーサレマスカ<H><SB>

```

図2 書き起こしテキストの例

3.2 話し言葉に特有な言語現象

収録された対話音声データの書き起こしテキストに基づいて話し言葉に特有な現象の対話データの特徴分析を試みた。現在までに書き起こし作業が完了している195対話のドライバー発話を分析の対象とした。対話の収録時間、ドライバーの発話時間、発話単位数、発話文数、形態素数を表4に示す。対話収録時間に対するドライバー発話時間は約2割であり、実走行車室内で行われるため、疎らな対話となっている。話し言葉に特有な言語現象として、フィラー、言

い淀み、及び、言い誤りを取り上げ、その出現頻度について調べた。諸現象の出現総数と1発話単位あたりの出現個数(出現率)を表5に示す。

ドライバー発話に出現したフィラーの総数は6,171個であり、1発話単位あたり0.34個出現している。出現位置については、全体の74.3%のフィラーが発話単位の文頭に現われている。言い淀みは、ドライバー発話に952回、6.5%の発話単位に出現し、言い誤りは、526回、3.6%の発話単位に出現した。一秒あたりの出現回数はそれぞれ、0.034回、0.010回であった。

表4 調査に使用した対話データ

項目	数値
対話数	195
収録時間(秒)	158,214
ドライバー発話時間(秒)	35,216
ドライバー発話単位数	18,073
ドライバー発話文数	33,076
ドライバー形態素数	259,453

表5 話し言葉に特有な言語現象の出現回数と出現率

項目	ドライバー発話	
	出現回数	出現率(%)
フィラー	6,171	34.14
言い淀み	1,201	6.65
言い誤り	638	3.53

4 統計情報に基づく日本語係り受け解析手法

係り受け解析では、係り受け規則を用いて文を解析し、その結果を係り受け構造によって表現する。日本語の係り受け規則の例を表6に示す。日本語係り受け解析では、一般に、上述した係り受け規則に対して以下の3つの構文的制約に従う。

- ・係り受けの非交差性 係り受けは互いに交差しない。
- ・後方修飾性 文末の文節を除き、必ず後方に位置する文節に係る。
- ・係り先の唯一性 文末の文節を除き、係り先は必ず一つ存在する。

しかしながら、日本語音声トランスクリプトには、前節で示したように話し言葉に固有の言語現象が頻出する。さらに、現状の音声認識技術では、自然な対話音声に対する認識性能は十分でなく、認識誤りの発声が解析精度に及ぼす影響が大きくなる。本研究では、フィラーや言い淀み、言い誤りが多数出現することに着目し、「係り受けの唯一性」に関する制約を緩和する。すなわち、文末以外の箇所でも係り先が存在しないとする係り受け分析を許容する。

表6 係り受け規則の例

係り文節	受け文節
名詞 + 格助詞「が」	動詞
名詞 + 格助詞「を」	動詞
動詞 連用形	動詞, 形容詞, 形容動詞
動詞 連体形	名詞
副詞	動詞, 形容詞, 形容動詞
連体詞	名詞

4.1 解析の概要

図3に本手法の解析の流れを示す。各処理の概要は以下に示す通りである。

1. 音声認識 入力された日本語音声に対し、音声認識ソフトウェアを用いて形態素分析済みのトランスクリプトを得る。以下の処理では、最尤の認識結果を用いる。
2. 文節まとめ上げ 文節の基本単位にしたがって、文節をまとめ上げる。ただし、文節は、0個以上の接頭辞および1個以上の自立語、0個以上の付属語または接尾辞からなるものを基本単位とする。文節系列は、形態素解析結果に基づいて一意に定める。
3. 係り受け解析 得られた文節列に対し、統計的に獲得した文節間係り受け確率と構文的係り受け制約に基づいて係り受け構造を作成する。

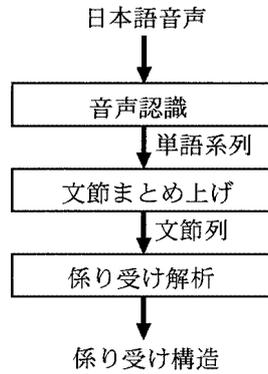


図3 解析の流れ

4.2 統計情報の獲得

対話コーパスに付与された係り受けデータから、係り受け関係に関する統計情報を獲得する方法について述べる。統計情報として、係り文節では自立語の原形 h_i と係りの種類 r_i を、また、受け文節では自立語の原形 h_j を利用する。ここで係りの種類とは、活用形や助詞など、適用可能な係り受け規則を決定するための係り受け文節の構成要素である。係り文節とその係りの種類の例を表7に、また、係りの種類と係り先の品詞の例を表8に挙げる。係りの種類と文節の自立語に関する統計データを用いて、各文節間の係り受け確率を以下のように計算する。

$$P(h_i \xrightarrow{r_i} h_j | h_i, r_i) = \frac{C(h_i \xrightarrow{r_i} h_j, h_i, h_j, r_i)}{\sum_x C(h_i \xrightarrow{r_i} x, h_i, x, r_i)} \quad (1)$$

ここで、 $x C(h_i \xrightarrow{r_i} x, h_i, x, r_i)$ はコーパス中における、自立語の原形が h_i で、その係りの種類が r_i である文節の出現頻度を、また、 $C(h_i \xrightarrow{r_i} h_j, h_i, h_j, r_i)$ は、自立語の原形が h_i で、その係りの種類が r_i である文節と自立語の原形が h_j の文節との係り受け関係の出現頻度を表す。

表7 係り文節と係りの種類の例

係り文節	係りの種類
本を	格助詞「を」
静かな	連体形
ゆっくり	副詞

表8 係りの種類と係り先の品詞（一部）

係りの種類	係り先の品詞
連用形	動詞，形容詞
連体形	名詞
接続詞	動詞，形容詞
連体詞	名詞
副詞	動詞，形容詞，副詞
感動詞・フィラー	なし
格助詞「が」	動詞，形容詞
格助詞「に」	動詞，形容詞
格助詞「まで」	動詞，形容詞
格助詞「を」	動詞
格助詞「へ」	動詞

4.3 係り受け解析の手続き

文節まとめ上げによって得られた文節列に対し、各文節間の係り受け関係を求める。この中から、以下の制約を満たす最も多い係り受け関係をもつ構造を求める。

- ・文節間係り受け確率があらかじめ設定された閾値を超える。
- ・文末を除き、各文節は後方に高々一つの係り先文節をもつ。
- ・係り受けが非交差である。

本手法では、係り受け構造の尤度を各文節間係り受け確率を用いて判定する。

5 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 車内音声対話データベース (CIAIR-HCC) [5] を用いて解析精度に関する実験を行った。実験では、大語彙連続音声認識ソフトウェアを用いてテスト用音声データをトランスクリプトに変換し、学習用テキストデータを用いてその係り受け解析を行った。

5.1 実験に使用したデータ

係り受け確率を計算するための学習用データとして、CIAIR-HCC の83対話分のドライバー発話を使用した全7,985発話単位に対して人手で係り受け分析データを付与した。データの品詞体系や係り受け文法については、原則として京大コーパス [7] の作成基準に準拠することとした。ただし、話し言葉に特有な部分については、

- ・ フィラーや言い淀みは未定義語とし、その係り受け先は存在しない。
- ・ 述部など、受けとなる文節が省略された場合、係り受けが存在せず、係り文節が単独で係り受け構造を形成する。
- ・ 話し言葉に固有の表現については、新たな辞書項目を追加して対応する。

などの基準を設けて作業を行った。

また、テストデータとしては、CIAIR-HCC の音声認識評価用テストセットの音声データ200文を使用した。発話単位の平均長は12.21語である。表9に、テストセットにおける話し言葉に特徴的な表現の出現頻度を示す。

表9 200発話単位における言語現象の出現頻度

項目	出現回数	発話単位あたりの頻度
フィラー	146	0.73
繰り返し	10	0.05
言い淀み	23	0.12
言い直し	15	0.08
係り受け倒置	18	0.09
述部省略	15	0.08

5.2 解析実験

音声認識には、日本語ディクテーション基本ソフトウェアJulius (99年度版) [6] を使用した。言語モデルは、CIAIR-HCC の29,829発話単位を使用して作成した (語彙数5,120、パープレキシティ19.27)。音声認識の結果、単語正解率 (*Corr*) で59.99%、単語正解精度 (*Acc*) で45.26%を得た。

また、テスト用音声データの書き起こしテキストに対して、人手で係り受け解析を行い、それを正解データとした。係り受けの総数は292個 (発話単位あたり1.46個) である。解析の精度を評価するために、係り受けの適合率 (precision) 及び再現率 (recall) を求めた。文節間係り受け確率の閾値を様々にかえた場合の適合率と再現率を図4に示す。閾値が0.10のときに取り出した201個の係り受けのうち、正しい係り受けは75個であり、適合率で37.3%、再現率で25.7%を得た。必ずしも高い解析精度が得られたとはいえないが、この値は実環境下における自由対話音声に対する認識精度に大きく依存する。すなわち、この実験結果は、高い音声認識結果の獲得が望めない状況においても、正しい係り受け関係が得られる可能性があることを示している。係り受け確率の閾値が0.10のときの解析例を表10に示す。

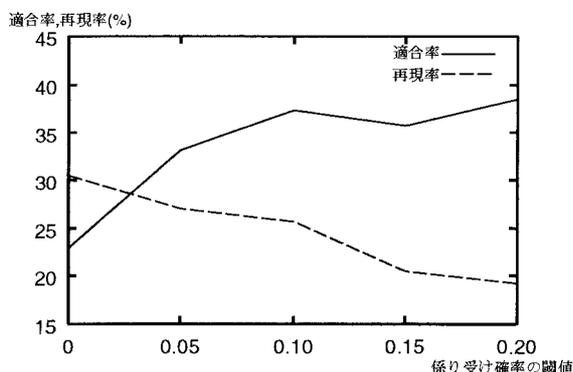


図4 音声トランスクリプト解析実験の結果

表10 音声トランスクリプトと係り受け構造の例

音声 1:	ああじゃあ京都行くまでにえーっと通行止めとかないのかなあ
トランスクリプト 1:	あー京都行くまでに減っ寄っずっと止めとかないのかな
係り受け構造 1:	((京都)(行くまでに))((止めとか)(ないのかな))
音声 2:	ええーこの辺の近くでコンビニはないかなあ
トランスクリプト 2:	え方も辺の近くでコンビニはないかな
係り受け構造 2:	((辺の)(近くで))((コンビニは)(ないかな))

6 まとめ

本報告では、作成した同時通訳コーパスについて述べた。このコーパスは、同時通訳者を介した独話、及び対話データから構成されており、話し言葉処理のための言語データとして、また、同時通訳に有用な知見を獲得するための基礎資料として利用できる。引続き、コーパスの収集を進めるとともに、言語処理に有用な各種タグ付け、ならびに、対訳アライメント作業の実施を予定している。

また本報告では、日本語音声の係り受け解析手法について述べた。係り先の種類による構文的制約や統計情報をもとに、解析可能な部分について部分的な構造を作成することにより、認識誤りを含むトランスクリプトに対してロバストに解析することができる。CIAIR 音声コーパスのドライバー発話音声200文を用いて解析実験を行った結果、自然な対話音声の解析における本手法の利用可能性を確認した。

参考文献

- [1] 江原、小倉、篠崎、森元、樽松: 電話またはキーボードを介した対話に基づく対話データベースADDの構築, 情報処理学会論文誌, 33 (4), pp. 448-456 (1992).
- [2] Matsubara, S. and Inagaki, Y.: Incremental Transfer in English-Japanese Machine Translation, *IEICE Transactions on Inf. & Sys.*, E80-D (11), pp. 1122-1129 (1997).
- [3] Takezawa, T. et al.: Japanese-to-English Speech Translation System: ATR-MATRIX, *Proceedings of ICSLP-98*, pp. 957-960 (1998).
- [4] 浦谷、竹沢、田代、森元、匂坂: ATRの新音声言語データベース、情報処理学会第48回全国大会講演論文集 (3), pp. 79-81 (1994).
- [5] Kawaguchi, N., et al.: Construction of Speech Corpus in Moving Car Environment, *Proceedings of ICSLP-2000*, Vol.III, pp. 957-960 (2000).
- [6] 河原、李、小林、武田、峯松、嵯峨山、伊藤、伊藤、山本、山田、宇津呂、鹿野: 日本語ディクテーション基本ソフトウェア (99年度版) の性能評価、情処研報、SLP31-2 (2000).
- [7] 黒橋、長尾: 京都大学テキストコーパス・プロジェクト、言語処理学会第3回年次大会発表論文集, pp.115-118 (1997).
- [8] 前川、籠宮、小磯、小椋、菊池: 日本語話し言葉コーパスの設計、音声研究, 4 (2), pp. 51-61 (2000).

< 発 表 資 料 >

題 名	掲載誌・学会名等	発表年月
Efficient Incremental Dependency Parsing	<i>Proceedings of 6th Int. Workshop on Parsing Technology</i> , To appear	2001年10月
Incremental Parsing for Interactive Natural Language Interface	<i>Proceedings of IEEE Conf. on SMC Workshop on NLPKE</i> , To appear	2001年10月
Multimedia Data Collection on In-Car Speech Communication	<i>Proceedings of 7th European Conf. on Speech Communication</i> , To appear	2001年 9 月
Interactive Incremental Parsing and Its Application to Multilingual Chat Assistant System	<i>Proceedings of 2th Int. Conf. on Software Eng., Networking and Parallel Comp.</i> , 112-119	2001年 8 月
統計データを用いた話し言葉音声の係り受け解析	情報処理学会研究報告、Vol. 2001, No.54, 63-68 / No.55, pp.23-28	2001年 6 月
Spoken Language Corpus for Machine Interpretation Research	<i>Proceedings of 6th Int. Conf. on Spoken Language Processing</i> 3:398-401	2000年10月
Construction of Speech Corpus in Moving Car Environment	<i>Proceedings of 6th Int. Conf. on Spoken Language Processing</i> 3:362-365	2000年10月
Spoken Language Parsing based on Incremental Disambiguation	<i>Proceedings of 6th Int. Conf. on Spoken Language Processing</i> 2:999-1002	2000年10月
複数語ごとの入力に対する漸進的構文解析	情報処理学会第61回全国大会講演論文集、Vol. 2, pp.153-154	2000年10月
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、Vol. 2, pp.181-182	2000年10月
Simultaneous Japanese-English Interpretation based on Early Prediction of English Verbs	<i>Proceedings of 4th Symposium on Natural Language Processing</i> pp.268-273	2000年 5 月