

音声対話処理における係り受け解析への統計情報の利用

村瀬 隆久†1

松原 茂樹†2

河口 信夫†3

稲垣 康善†1

†1名古屋大学大学院工学研究科

†2名古屋大学言語文化部 / CIAIR

†3名古屋大学大型計算機センター / CIAIR

{murase,matu,kawaguti,inagaki}@inagaki.nuie.nagoya-u.ac.jp

1 はじめに

近年、音声認識技術の進歩に伴い、音声対話処理システムの実現が大きなテーマとなっている。対話音声を理解するためには、人の自然発話を解析する必要がある。しかし、話し言葉にはフィラーや言い淀み、言い直しなどが頻出するため、それを従来の言語解析手法を用いて処理することは困難である。そのため、話し言葉に対するロバストな解析手法が必要とされている。

本稿では、大規模音声対話コーパスを用いた対話音声の係り受けに関する特徴分析について述べ、それに基づく日本語話し言葉の係り受け解析手法を提案する。従来の係り受け解析手法では、係り受けの非交差性、後方修飾性、係り先の唯一性の3つの制約が用いられてきたが、本研究における分析の結果、対話音声では、倒置現象や係り先がない文節などを含む発話が無視できない頻度で出現することが分かった。そこで本手法では、後方修飾性の制約及び係り先の唯一性に関する制約を緩和する。すなわち、倒置を許し、フィラー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する。解析結果は、部分的な係り受け構造によって表す。

また、本手法では、音声対話コーパスに付与された係り受け構造から各文節間の係り受け確率を統計的に獲得し、それを用いて係り受け構造の尤度を計算する。近年、大規模テキストコーパスから獲得した統計情報を利用した係り受け解析手法が盛んに研究されている。藤尾ら [6] は、文節間の共起頻度を用いて係り受け確率を推定し、係り受け解析を行っている。内元ら [1] は、最大エントロピー法によって学習した係り受け確率モデルから係り受け確率を求め、その積により1文に対する係り受け確率を計算している。これらの手法はいずれも、書き言葉に対する解析手法である。また、話し言葉の係り受け解析手法についても提案されており、伝 [5] は、言い直しや言い淀みなどを語と語の間のある種の係り受け関係と捉え、従来の係り受け解析を拡張した解析手法を提案しているが、扱える話し言葉の言語現象は限られている¹。

一方、本手法では、自然発話文に対して、統計情報を利用することにより、最尤の構造を作成する。本手法の有効性を評価するために、名古屋大学 CIAIR 車内音声対話コーパス [3] に収録されたドライバー発話に対して、係り受け

```
0009 - 00:41:960-00:50:228 F:0:I:R:
この先 & コノサキ
三百メートルほど先に & サンビャクメートルほどサキニ
左手に & ヒダリテニ
サンクス & サンクス
五百メートルほど先に & ゴヒャクメートルほどサキニ
セブンイレブンが & セブンイレブンガ
ございます<SB> & ゴザイマス<SB>
0010 - 00:51:888-00:52:100 F:D:P:C:
(Fん) & (Fン)
0011 - 00:54:633-00:55:394 F:D:P:R:
(? 次左かな) & (? ツギヒダリカナ)
0012 - 01:02:750-01:03:993 F:0:I:R:
どちらに & ドチラニ
なございますか<SB> & ナザイマスカ<SB>
```

図 1: 車内音声対話コーパスの書き起こしテキストの例

解析実験を行った。実験では、車内音声対話コーパスの 81 対話から得られた統計情報を用いて、日本語音声の書き起こしデータに対して係り受け解析を行った。評価実験の結果、本手法の有効性を確認した。

2 対話音声の分析

名古屋大学 CIAIR で構築されている車内音声対話コーパス [3] を用いて、実走行車内対話音声の言語分析を行った。コーパスには、車両運転中におけるドライバーとナビゲータ (人間, WOZ システム, 音声対話システム) との対話が収録されている。

2.1 CIAIR 車内音声対話コーパスの概要

CIAIR における車内音声対話の収録は 1999 年から現在まで続けられている。専用の収録車を構築し [2]、音声や画像、車両状態などといったマルチモーダル情報を大量に収録している。これらは、実走行車内対話の分析に有用である。

収集した音声データの書き起こし作業は人手により行っている。書き起こしテキストの例を図 1 に示す。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している。さらに、各対話をポーズで分割し、各々を発話単位としてその開始時間、終了時間を記録している。また、各発話ごとに話者の性別 (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) に関する情報も付与されている。

¹山本ら [8] は、対話文の分析を通して、助詞省略と倒置に関するヒューリスティックを作成し、それを用いて助詞落ち文や倒置文の解析を行う手法を提案している。しかし、単純なヒューリスティックであるため、自由度の高い自然発話の解析への利用可能性は明らかではない。

表 1: 係り受け分析データ

項目	数値
対話数	81
総発話単位数	7,781
総ターン数	6,078
総文節数	24,993
総係り受け数	11,789
係り受け数 / 発話	1.52
係り受け数 / ターン	1.94

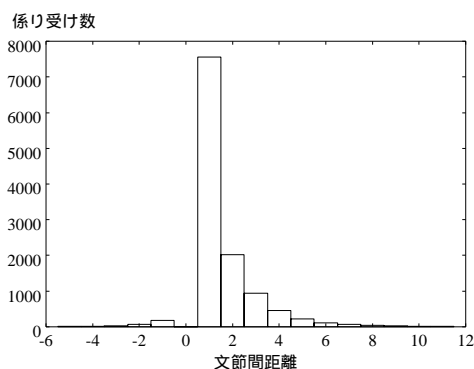


図 2: 文節間距離と係り受け数との関係

2.2 対話音声における係り受けの分布

対話音声における係り受けの特徴を明らかにするために、書き起こしデータに係り受け分析を与えた。車内音声対話コーパスの 81 対話からドライバー発話を取り出し分析の対象とした。係り受けの付与は人手で実施し、データの品詞体系や係り受け文法については京大コーパス [4] の作成基準に準拠した。ただし、話し言葉特有の現象については以下の作成基準に従った。

- フィラーや言い淀みは、係り先が存在しない。すなわち、単独で係り受け構造を形成する。
- 受け文節が省略された文節の係り先は存在しない。
- 話し言葉特有の言い回し表現（「こっから」、「食べてえ」など）については、新たな辞書項目を設けて、形態素ごとに品詞を定めた。
- 対話ターンを 1 つの修正対象とする。発話単位を越える係り受けも認めるが、対話ターン間にまたがった係り受けは認めない。

係り受け分析対象に関するデータを表 1 に示す。24,993 文節²に対して、11,789 個の係り受けが存在した。1 発話単位あたりの平均係り受け数は 1.52 個であり、1 ターンあたりでは 1.94 個であった。1 文あたりの係り受け数が平均 10 程度である新聞などの書き言葉と比べて、著しく少ない。しかし、このデータは話し言葉の係り受け解析が書き言葉のそれに比べ簡単であることを必ずしも意味しない。なぜなら、話し言葉では、あらゆる文節が受け文節をもっているわけではなく、係り先がない文節の特定も必要だからである。実際に、全文節数中の約 52.8% が係り先をもたない文節であった。

次に、倒置、及び、発話単位にまたがる係り受けについて調査した。倒置は 230 個存在した。全体の 3.8% のターンに出現して、1 ターンあたり約 0.04 回出現しており、無視できる頻度ではない。その出現位置について調査したところ、倒置全体のうちの約 86.5% がターンの最後の文節に出現している。一方、発話単位にまたがる係り受けは、73 個存在した。複数発話単位から構成されている 1,362 ターンの 5.4% に出現しており、発話単位を話し言葉解析の処理単位とすることは、必ずしも十分でないことがわかる。

²このうち、フィラー単独の文節は 3,049 個である。

また、係り文節と受け文節の間の距離についても調べた。文節間距離と係り受けの数の関係を図 2 に示す。ただし、文節間距離は受け文節の位置と係り文節の位置の差で測定する。全体の、約 64.3% が隣接する後方の文節への係り受けであり、また全体の約 8 割が後方の 2 文節以内に係っている。

3 日本語話し言葉の係り受け解析

本手法では、従来、係り受け解析で用いられてきた構文的制約を緩和し、また、音声対話コーパスから獲得した統計情報を用いることにより最尤の係り受け構造を作成する。なお、係り受けが発話単位にまたがることもあるという分析結果に基づき、本手法では 1 ターンを解析の単位とした。

3.1 構文的制約

従来の日本語係り受け解析では、一般に以下の 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。

係り受けの後方修飾性 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、係り先は必ず存在し、かつ、二つ以上存在しない。

しかし、日本語話し言葉には、前節で述べたように話し言葉に固有の現象や係り受けが頻出するため、これらの制約が完全に成り立つことを前提とすることは難しい。そこで本手法では、ロバスタな係り受け解析を実現するために、上記の構文的制約を緩和した。すなわち、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、係り先が存在しない文節を認め、その係り先をその文節自身に係るとした（係り先の唯一性の緩和）。また、倒置に対処するため、前方の文節に係ることを認める（係り受けの後方修飾性の緩和）³。

³係り受けが交差することは稀であるため、係り受けの非交差性は遵守する。

表 2: 係り文節と係りの種類の例

係り文節	係りの種類
電話が	格助詞「が」
レストランで	格助詞「で」
遅く	連用形
大きい	連体形
買える	連体形
ちょっと	副詞

表 3: 文節と係りの種類

見出し語	係りの種類
えーと	なし
コンビニ	名詞
ないかな	終助詞「かな」
そ	なし
その	指示詞
近くに	副詞

3.2 統計情報の利用

本手法では、前節で述べた、係り受け構造が付与された音声対話コーパスから、係り受けに関する統計情報を獲得し、これを活用して最尤の係り受け構造を作成する。統計情報として利用する属性は以下の通りである。

- 係り文節，受け文節の自立語の原形 h_i, h_j
- 係り文節，受け文節の自立語の品詞 t_i, t_j
- 係りの種類 r_i
- 文節間距離 d_{ij}
- 文節間のポーズの数 p_{ij}
- 係り文節の位置 l_i

ここで、係りの種類は、係り文節が付属語を伴うときはその付属語の語彙，品詞，活用形であり，そうでない場合は一番最後の形態素の品詞，活用形と定義される。表 2 に例を挙げる。また、係り文節の位置は、その文節が入力ターン内で一番最後の文節か否かを表す。前節で述べたように、倒置の多くは入力文の最後の係り文節に出現する傾向があり、倒置の確率を推定するためにこの属性を設けた。

以上の属性を用いて、各文節間の係り受け確率を以下のように計算する。

$$P(i \xrightarrow{rel} j|B) = \frac{C(i \rightarrow j, h_i, h_j, t_i, t_j, r_i)}{C(h_i, h_j, t_i, t_j, r_i)} \times \frac{C(i \rightarrow j, r_i, d_{ij}, p_{ij}, l_i)}{C(r_i, d_{ij}, p_{ij}, l_i)} \quad (1)$$

ただし、 C は共起頻度関数、 B は入力文節列である。

式 (1) の右辺の第 1 項は自立語の共起確率、第 2 項は文節間の距離確率を表している。これらの 2 つの係り受け確率を独立とみなして分離することにより、データの過疎性の問題を軽減することができる [6]。また、式 (1) では、係り先のない文節はそれ自身に係る (すなわち $i = j$) とみなすことにより、係り先をもたない場合の確率も計算することができる。

式 (1) によって得られた係り受け確率を用いて、入力文節列 $B (= b_1 \dots b_n)$ に対する係り受け構造 S の確率を以下の式によって計算する。

$$P(S|B) = \prod_{i=1}^n P(i \xrightarrow{rel} j|B) \quad (2)$$

$P(S|B)$ の値を文節列 B に対する係り受け構造 S の尤度とする。

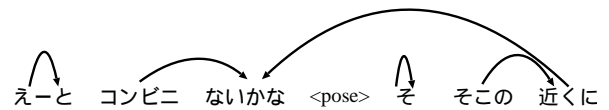


図 3: 「えーとコンビニないかな <pose> そその近くに」に対する係り受け構造

3.3 係り受け解析の手続き

文節をまとめ上げることによって得られた文節列に対して、以下の条件を満たす係り受け構造を作成する。

- 係り受けが非交差である。
- 各文節は高々一つの係り先文節をもつ。

作成された係り受け構造の中で、式 (2) によって得られた確率が最大のものを最尤の係り受け構造として出力する。

3.4 解析例

ポーズを含む日本語文「えーとコンビニないかな <pose> そその近くに」に対する解析例を示す。この文を文節に区切ると「(えーと)(コンビニ)(ないかな)(そ)(その)(近くに)」となる。これらの文節の係りの種類、及び、各文節間の係り受け確率をそれぞれ表 3, 4 に示す。表 4 は、例えば、「コンビニ」が「ないかな」に係る確率が 0.40 であることを表している。また、「えーと」が「えーと」に係る確率とは、「えーと」がどの文節にも係らない確率を意味する。表から得た各係り受け確率を用いて、最大の確率をもつ構造を作成した結果、図 3 のような係り受け構造が得られる。

4 評価実験

本手法の有効性を評価するために、名古屋大学 CIAIR 車内音声対話コーパス [3] を用いて係り受け解析の実験を行った。

表 4: 文節間係り受け確率

	えーと	コンビニ	ないかな	そ	そのの	近くに
えーと	1.00	0.00	0.00	0.00	0.00	0.00
コンビニ	0.00	0.01	0.40	0.00	0.00	0.00
ないかな	0.00	0.00	0.88	0.00	0.00	0.00
そ	0.00	0.00	0.00	1.00	0.00	0.00
そのの	0.00	0.02	0.00	0.00	0.00	0.75
近くに	0.00	0.00	0.80	0.00	0.00	0.02

表 5: 係り受け解析実験の結果

項目	(a)	(b)	(a)+(b)
適合率	82.0%	88.5%	85.5%
再現率	64.3%	83.3%	73.8%

(a) 係り先が存在する文節に対する効果

(b) 係り先がない文節に対する効果

4.1 実験に使用したデータ

実験用データとして、2.2節で述べた81対話分のドライバー発話の書き起こしテキストを使用した。100ターンをテストデータとして、残りの5978ターンを係り受け確率推定用学習データとした。テストデータにおける1ターンあたりの平均文節数は4.81個であった。

4.2 解析実験

正解データは、書き起こし文に人手で係り受けを付与することにより作成した。係り受けの総数は481個である。

実験結果を表5に示す。解析の精度を評価するために、係り受けの適合率、及び、再現率を求めた。本手法により得られた係り受けの総数415個のうち、正解データの係り受けと一致したものが355個あり、適合率は85.5%、再現率は73.8%となった。このことから、フィラーなどを含む自然発話文においても、本手法により、書き言葉を対象とした他の係り受け解析手法[1, 6]と同等の高い精度で係り受けを抽出できることを確認した。

また、正解係り受け481個のうち240個は係り先がない文節であり、そのうちの200個を本手法によって特定することができた。係り先をもたない文節の多くはフィラーや言い誤りなどであるが、そのような現象の特定のために、係り受け確率の利用が有効であることが分かった。

5 おわりに

本稿では、日本語話し言葉に対する係り受け解析手法を提案した。従来の係り受け解析で用いられてきた構文的制約を緩和し、統計情報を利用することによりロバストな解析を行うことができる。CIAIR車内音声対話コーパスを用いて解析を行った結果、本手法の利用可能性を確認した。本

稿では、音声認識率が100%であると仮定して実験を行ったが、音声対話処理システムでは音声認識によって生成されたテキストを処理することになる。そこには、多数の認識誤りが含まれるため、極めてロバストな解析技術が要求される[7]。本稿で提案した統計的係り受け解析の実用性を検証するために、今後は、音声認識システムを用いた評価を予定している。

参考文献

- [1] 内元 清貴, 関根 聡, 井佐原 均: 最大エントロピー法に基づくモデルを用いた日本語係り受け解析, 情報処理学会論文誌, Vol.40, No.9, pp.3397-3407 (1999).
- [2] 河口信夫, 牛窪誠一, 松原茂樹, 梶田将司, 武田一哉, 板倉文忠: 走行車室内音声対話収録システムの開発, 電子情報通信学会論文誌, Vol.J84-D-II, No.6, pp.1122-1129 (2001).
- [3] Kawaguchi, N., Matsubara, S., Takeda, K. and Itakura, F.: Multimedia Data Collection of In-Car Speech Communication, *Proc. of 7th European Conf. on Speech Communication and Technology*, pp. 2027-2030 (2001).
- [4] 黒橋 禎夫, 長尾 真: 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).
- [5] 伝 康晴: 話し言葉解析のためのコーパスに基づく優先度計算法, 自然言語処理, Vol.4, No.1, pp.41-56 (1997).
- [6] 藤尾 正和, 松本 裕治: 語の共起確率に基づく係り受け解析とその評価, 情報処理学会論文誌, Vol.40, No.12, pp.4201-4211 (1999).
- [7] 松原 茂樹, 佐藤 利光, 河口 信夫, 稲垣 康善: 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, NL-143, pp.63-68 (2001).
- [8] 山本 幹雄, 小林 聡, 中川聖一: 音声対話文における助詞落ち・倒置の分析と解析手法, 情報処理学会論文誌, Vol.33, No.11, pp.1322-1330 (1992).