自然言語処理 143－9
音声言語情報処理 36－4
（2001. 6. 1）

統計データに基づく話し言葉音声の係り受け解析

松原 茂樹†⋆ 佐藤 利光‡ 河口 信夫§⋆ 稲垣 康善‡

†名古屋大学言語文化部 §名古屋大学大型計算機センター
‡名古屋大学大学院工学研究科計算理工学専攻
⋆名古屋大学統合音響情報研究拠点
matubara@lang.nagoya-u.ac.jp

概要

実対話環境下におけるロバストな音声理解の実現のために，音声認識結果に対する言語解析が重要となる．本稿では，統計情報を用いた日本語音声の係り受け解析手法を提案する．本手法では，統計情報として音声対話コーパスから獲得した文節間係り受け確率を用いる．日本語係り受け文法，及び，係り受け確率を活用することにより，誤りを含む音声認識結果から正しい係り受け関係を抽出することができる．大語彙連続音声認識ソフトウェアを用いた日本語音声解析実験を行った結果，自然な話し言葉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．

Robust Parsing of Spoken Language based on Statistical Dependencies

Shigeki MATSUBARA†⋆, Toshimitsu SATO‡, Nobuo KAWAGUCHI§⋆ and Yasuyoshi INAGAKI‡

†Faculty of Language and Culture, Nagoya University
‡Department of Computational Science and Engineering, Nagoya University
§Computation Center, Nagoya University
⋆Center for Integrated Acoustic Information Research, Nagoya University
matubara@lang.nagoya-u.ac.jp

ABSTRACT

This paper proposes a method of dependency parsing of the Japanese speech using statistical lexical information. The method utilizes the probability of the dependency between *bunsetsu* and *bunsetsu*, which is acquired from a spoken dialogue corpus. With the Japanese dependency grammar and the statistical dependencies, it can extract the correct dependency relation from the Japanese speech. An experiment with a speech recognition software has shown the availability of the method for a parsing parsing of spontaneously spoken language.

1 はじめに

音声を用いたヒューマンコンピュータインタラクションの実現のために，実対話環境下において利用可能なロバストな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．しかし，自然な対話音声のトランスクリプトには，フィラーや言い淀み，言い直しなどの書き言葉にはみられない言語現象が頻出するとともに，音声認識誤りも含まれるため，それを従来の言語解析手法を用いて処理することは難しい．

本稿では，日本語音声トランスクリプトの係り受け解析手法を提案する．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．近年，大規模テキストコーパスから獲得した統計情報を用いた係り受け解析について盛んに研究されている [9, 2]．実際，これまでに，話し言葉の係り受け解析手法についてもいくつか提案されているが (例えば，[1])，その多くは人手によって作成された転記テキストを対象としている．しかし，対話音声に対して高精度の認識性能を期待することが難しい現状では，エラーを含む認識結果に対してロバストに言語解析する技術が求められる [10]．本手法では，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，実走行車内音声対話コーパス (CIAIR-HCC)[4] 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，CIAIR-HCC 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 話し言葉の解析

名古屋大学 CIAIR で構築されている車内音声対話コーパス (CIAIR-HCC)[4] 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を収録している．

2.1 車内音声対話コーパスの概要

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．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統合的に参照することができるため，走行・運転状況とドライバー発話との関係の分析に有用である．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．ただし，十分に訓練されたナビゲータがシステムの役割を担当し，ジェスチャ応答を行わない，ドライバーと目線を合わせて会話をしない，など，WOZ システムとの対話に近い形態を採用している．コーパスの収録方法や収集システムの詳細については，文献 [3] を参照されたい．

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 1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:D:N:
    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て <H>          & ファーストフードテンテ <H>
    あるのかなあ <SB>                 & アルノカナー <SB>
0002 00:10:683-00:13:969 F:O:N:
    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:D:N:
    (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                            & バショオ
    教えてほしいんだけど <SB>         & オシエテホシンダケド <SB>
0004 00:18:092-00:21:136 F:O:N:
    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 1: 書き起こしテキストの例

2.2 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 1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 2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

表 1: 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 2: 話し言葉に特有な言語現象の出現回数と出現率

	ドライバー発話	
項目	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

りの出現回数はそれぞれ，0.034 回，0.010 回であった．

2.3 音声トランスクリプトの係り受け解析

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 3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，係り先は必ず存在し，かつ，二つ以上存在しない．

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．すなわち，音声トランスクリプトに対しては，ロバストに解析する必要があり，そのために上述した構文的制約を緩和することは一つの方法である．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．フィラーなどの言語表現については，単独で係り受け構造を形成するとして解析する．特に，言い淀みについては，それ自体で単語を形成しないため，別の単語として認識されることになるが，そのような表現についても係り先が存在しないとして解析する．

表 3: 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

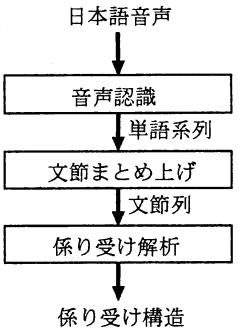

図 2: 解析の流れ

3 統計情報に基づく係り受け解析手法

本稿では，音声対話コーパスから文節間係り受け確率を獲得し，それを用いて日本語音声を係り受け解析する手法を提案する．音声認識では考慮されにくい，距離が離れた文節間依存性に関する統計情報を活用することにより，意味的に正しくない部分を取り除くことができる．

3.1 解析の概要

図 2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

表 4: 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 5: 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

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．

次節以降では，統計的係り受け情報の獲得方法とそれを用いた係り受け解析について説明する．

3.2 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 4に，また，係りの種類と係り先の品詞の例を表 5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，$\sum_x C(h_i \xrightarrow{r_i} x, h_i, x, r_i)$ 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 6: 文節列「どっか 売っ この 近くに ある」の係りの種類と係り先の品詞

見出し	係りの種類	係り先の品詞
どっか	名詞	動詞
売っ	連用形	動詞，形容詞
この	連体詞	名詞
近くに	格助詞「に」	動詞，形容詞

表 7: 文節間係り受け確率

	売っ	この	近くに	ある
どっか	0.01	×	×	0.26
売っ	-	×	×	0
この	-	-	0.19	×
近くに	-	-	-	0.62

3.3 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．

3.4 解析例

言い淀み「も」及び「な」を含む日本語音声「どっかもな近くにある」に対して，音声認識によって生成されたトランスクリプト「どっか 売っ この 近く にある」の解析例を示す．ここでは，文節間係り受け確率の閾値を 0.1 とする．文節のまとめ上げにより，文節列「(どっか)(売っ)(この)(近くに) (ある)」が得られる．これらの文節の係りの種類と係り先の品詞を表 6に，また，係り受け可能な文節間の係り受け確率を表 7に示す．この表は，例えば，「どっか」が「売っ」に係る確率が 0.01 であることを表している．「どっか」の受け文節として，係り受け規則からは「売っ」と「ある」が可能であるが，係り受け確率の閾値か

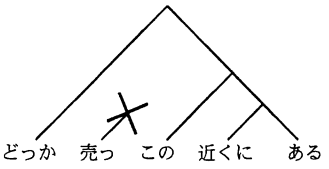

図 3: 「どっか 売っ この 近く に ある」の係り受け構造

ら「ある」に係ると判定される．また，「売っ」の受け文節の候補は「ある」であるが，受け文節は存在しないことが統計情報からわかる．同様に，「この」と「近くに」の受け文節は，それぞれ「近くに」と「ある」となる．

トランスクリプト「どっか 売っ この 近く に ある」の係り受け解析結果を図3に示す．言い淀み部分「も」の認識結果である「売っ」に対しては，係り先がないとする係り受け構造を生成する．音声認識において単語 n-gram ベースの言語モデルを使用した場合には，単語連接の確率に従ってトランスクリプトが生成されるものの，この例が示すように，言語解析において係り受け確率を使用すれば，言い淀みなどに相当する部分を取り除ける可能性がある．一方で，言い淀み「な」に対する認識結果「この」については，「近く」との間で係り受け関係がなりたつことになる．これは，統計データの中に「この」が「近く」に係る場面が多く見られたためである．しかし，同時に，「この」と「近く」が頻繁に連接していれば「この近く」といった単語系列がトランスクリプトに出現しやすい．このような場合には，認識エラーの文節間からも係り受け関係が抽出されることになる．

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 車内音声対話データベース (CIAIR-HCC)[4] 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．

4.1 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[1]．全 7,985 発話単位に対して人手で係り受け

[1]車内音声対話データベースにおける転記フォーマットは，日本語話し言葉コーパス (CSJ)[8] の書き起こし基準に準拠している．2.0 秒以上の休止で分割された発話を発話単位と定めている．

表 9: 話し言葉に固有の表現の例

話し言葉での表現	書き言葉での表現	品詞
こっ (から)	ここ (から)	名詞形態指示詞
どん (くらい)	どの (くらい)	連体詞形態指示詞
(どう) しょう	(どう) しよう	サ変動詞
ちゃう	ちがう	ワ行動詞
いっ (かな)	いい (かな)	形容詞
(食べ) れる	(食べ) られる	動詞性接尾辞
(吸い) てえ	(吸い) たい	形容詞性述語接尾辞
(友達) ん (ところ)	(友達) の (ところ)	接続助詞

表 10: 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

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．
- 表 9に示すような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．これは，CIAIR-HCC の異なる男性 10 名と女性 10 名がドライバー役割を担った 20 対話それぞれについて，ドライバー発話 10 文の音声を抽出することによって作成されている．発話単位の平均長は 12.21 語である．表 10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．

4.2 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

表 8: 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)))

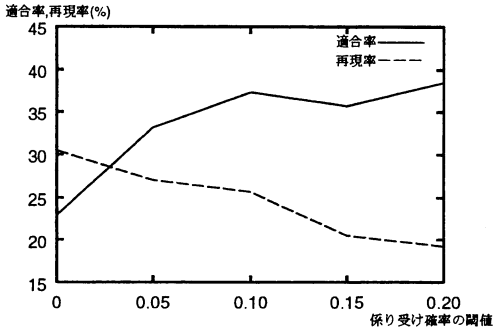

図 4: 音声トランスクリプト解析実験の結果

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，人手で係り受け分析を行い，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 8に示す．

5 おわりに

本稿では，日本語音声トランスクリプトの係り受け解析手法を提案した．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 車内音声対話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。

謝辞　本研究の一部は文部省科学研究費補助金 COE 形成基礎研究費 (課題番号 11CE2005) の補助を受けて行われた。

参考文献

[1] 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).

[2] 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).

[3] 河口，松原，岩，梶田，武田，板倉: 実走行車内における音声データベースの構築，情処研報，SLP30-12，pp.57-62 (2000).

[4] Kawaguchi, N., et al.: Construction of Speech Corpus in Moving Car Environment, *Proceedings of ICSLP-2000*, Vol.III, pp. 957–960 (2000).

[5] 河口，松原，若松，梶田，武田，板倉，稲垣: 実走行車内音声対話コーパスの設計と特徴，信学技報，NLC2000-57，pp. 61-66 (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).

[9] 内元，関根，井佐原: 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).

[10] 脇田，河井，飯田: 意味的類似性を用いた音声認識正解部分の特定法と正解部分のみ翻訳する音声翻訳手法，自然言語処理，Vol.5，No.4，pp.111-125 (1998).