

音声言語処理のための部分木併合手法

田代敏久 竹澤 寿幸 森元 暉

ATR 音声翻訳通信研究所

概要

発話の部分的な統語解析結果から発話全体の構造を作成する手法(部分木併合手法)を提案する。部分木併合処理は、1) 音声認識部と言語処理部との効率的なインターフェース、2) 不適格な文を含む話し言葉の頑健な言語解析、の二つを同時に実現するために必要である。本稿では、部分木併合手法の概要及び本手法の有効性について議論する。

Handling Partial Trees for Speech and Language Processing

Toshihisa Tashiro, Toshiyuki Takezawa, Tsuyoshi Morimoto

ATR Interpreting Telecommunications Research Laboratories

Abstract

We propose a method of combining partial parses. This method is necessary to develop an efficient interface between the speech recognition component and the natural language component, and a robust parser which can analyze ill-formed sentences. In this paper, we discuss the outline and the characteristics of this method.

1 はじめに

実用的な音声言語システムを構築するためには、効率的で頑健な音声認識モジュール及び言語解析モジュールが必要である。認識モジュール及び解析モジュールの効率を向上させるためには、探索空間を適切に絞り込めるようなインターフェースを用意する必要がある。一方、頑健な認識・解析モジュールを構築するためには、自由発声されたユーザ発話から必要な情報を適切に抽出するメカニズムを用意する必要がある。

本稿では、効率的で頑健な音声認識・言語解析を実現するための部分木併合処理について述べる。まず、部分木併合処理の必要性について、

- ポーズ単位に基づく音声認識システムとのインターフェース
- 頑健な構文・意味解析

という二つの観点から論じる。続いて、部分木併合処理の基本的な考え方を説明する。さらに、本稿で提案する部分木併合手法の特徴を効率性、頑健性、領域

(domain)や課題(task)の変更の容易性(portability)等の観点から考察する。

2 部分木併合処理の必要性

2.1 ポーズ単位に基づく音声認識システムとのインターフェース

音声認識を、ポーズ(休止)を単位として行なう手法がいくつか提案されている[1, 3, 2, 4]。ポーズ単位に基づく音声認識を行なう利点としては、

- 人間が一息で発声できる長さには限度がある。そのため、ポーズ単位毎に認識を行なうことにより、探索空間が極端に長くなることが防げる[1]。
- ポーズを機械的に検出する機構を用意すれば、音声認識システムのユーザは、“文節”や“文”等の統語的な単位を意識せず、自由に発声することが可能となる[2]。
- 旅行会話データベースで調査したところ、ポーズ単位内の言語現象は、文脈自由文法で記述することが可能な場合が多く、文脈自由文法を言語制約として用いる音声認識システムにとって都合がよい[3]。

等が挙げられる。

我々は、ポーズ区間毎に部分木構造を出力できる音声認識システムを開発した[4]。このシステムの出力を言語処理部でそのまま処理できれば(文字列や形態素列に一旦戻すことなく処理できれば)、音声認識を利用する音声言語システム全体の処理効率を向上させると同時に、言語知識の保守・管理が容易になる。

2.2 頑健な構文・意味解析

文法的に不適格な発話、及び文法的に適格であっても機械的な構造解析が難しい発話(箇条発話や複雑な並列句等)を処理するために、発話を部分的に統語解析し、全体の意味の解釈を後続の処理(部分木併合処理)に委ねるといった手法は、かなり古くから提案されている手法である[5]。また、実世界(real world)のテキストを処理し、定量的に評価することが要求される分野においては、多くのシステムに採用されている手法でもある[6]。

我々は、既に、自然に発声された旅行会話データベースの言語現象の調査を行ない[7]、ほとんどの言語現象を部分木解析・併合の手法で処理できる見込みを得た。また、音響レベルの情報であるポーズと、統語レベルの情報である句の単位(部分木)の切れ目とが一致する場合が多いことを確認した[3]。

3 部分木併合処理の実際

3.1 基本的な考え方

自然言語の解析手法は、背景となる文法理論により次の2つに分類できる。

- 句構造解析:
語順を基本的な情報源とし、構成要素(constituent)をノードとする木構造(句構造)を作成することを主眼とする解析手法である。句構造文法(文脈自由文法)を効率的に処理するアルゴリズムが存在するため、多くのシステムで採用されている解析手法である。しかし、語順を手がかりに解析を行なうため、倒置や省略等を含むいわゆる文法的に不適格な文を処理することは困難である。
- 依存構造解析:
語及び語と語の関係を基本的な情報源とし、語をノードとする木構造(依存構造)を作成することを主眼とする解析手法である。依存文法は、形式言語理論が登場する以前の伝統的な文法理論に近い考え方を元にしており、語順が自由な言語の統語的な分析が可能である。しかし、計算機上で効率的に処理するための手続きは、文脈自由文法の解析アルゴリズムほど明確・一般的ではない。完全に語順の情報を無視して解析を行なうためには、入力中のすべて語と語の組合せの依存関係をチェックする必要があり、計算コストの点で問題がある。

部分木解析・併合処理とは、上記の2つの解析手法の実用的な統合手法と考えることができる。つまり、比較的語順が安定している部分は文脈自由文法を用いて効率的な解析を行ない、部分的な句構造を最大限活用しながら(語順の情報を完全に無視することなく)、依存構造解析を行なうことにより、処理の効率と頑健さを両立させることを狙うわけである。

この部分木解析・併合処理の考え方は、従来の日本語解析で頻繁に用いられてきた文節係り受け解析に極めて近いものである。しかし、従来の文節係り受け解析には、

- 部分木の単位が文節に固定されているため、連続した文節をひとまとめにして扱うことや、文節内に語の欠落や誤りがある場合を処理することは難しい。
- 依存構造解析の対象は、文節と文節に固定されているため、語の語の直接的な依存関係の抽出が不可能である。

という問題点があった。

そこで、我々は部分木解析・併合処理を、

- 部分木の単位は任意の句 (単独の語も部分木とみなせる) とする。
- 依存構造解析の対象は、ある部分木のトップノードと他の部分木のすべてのノードとする。

ことにより、より柔軟な解析を可能にした。

3.2 部分木併合処理の動作

部分木併合処理の動作を、具体的な例を用いて説明する。

3.2.1 入力

部分木併合処理の入力は、文 (発話) の部分的な句構造解析結果である。我々は、部分木併合処理の入力として、文脈自由文法を言語制約として用いる音声認識システムの出力を利用することを想定している [4]。この音声認識システムは、

具体的に▼これはどういう内容の▼ものなんですか (▼はポーズの位置を示す)

というような発声に対して、図 1 のような部分木リストを出力することができるようになっていく¹。

3.2.2 依存構造への変換

まず、複数の部分木の一つ一つを、部分的な依存構造に変換する。句構造から依存構造への変換は、句構造規則にタイプを付与することにより簡単に行なうことができる [8]。また、以降の処理を容易にするため、仮想的なルートノード (全体の依存構造のトップとなるノード) を用意する。図 2 に部分依存構造の例を示す。

3.2.3 部分依存構造のリンク

図 2 に示す仮想的なルートノードの以外の部分依存構造のトップノードを、他の部分木のノードと接続することにより、部分依存構造から全体の依存構造を復元することができる。つまり、部分木併合処理とは、各部分木の適切なリンク先 (係り先) を決定する探索問題と考えることができる。

我々は、部分木併合処理を、以下のような欲張り法 (greedy search) を利用して行なうことにした。

- まず、部分木 1 のリンク先を決定する。部分木 1 のトップノードと他の部分木のすべてのノードを対象に依存関係 (単語の係り受け関係) のチェックを行ない、もっとも整合性のよい組合せを決定する。整合性の良否は、品詞間の関係、単語間の関

¹実際には、音声認識システムは N-best 結果 (部分木リストのリスト) を出力し、部分木併合処理を含む言語処理で再順序付けを行なう必要があるが、本稿では説明を簡単にするため、音声認識システムで適切な 1-best 出力が得られると仮定する。

```

部分木 1:
(<連用修飾>
  (<形容名詞> 具体的)
  (<助動詞> に))

部分木 2:
(<後置詞句>
  (<名詞句>
    (<代名詞> これ))
  (<係助詞> は))

部分木 3:
(<連体詞句>
  (<名詞句>
    (<連体詞句>
      (<連体詞> どういう))
    (<名詞句>
      (<普通名詞> 内容)))
  (<連体助詞> の))

部分木 4:
(<動詞句>
  (<動詞句>
    (<名詞節>
      (<連体修飾節>
        (<動詞>
          (<名詞句>
            (<普通名詞> もの))
          (<助動詞> な)))
        (<準体助詞> ん))
      (<助動詞>
        (<助動詞語幹> で)
        (<語尾> す)))
    (<終助詞> か))
  
```

図 1: 部分木併合処理の入力

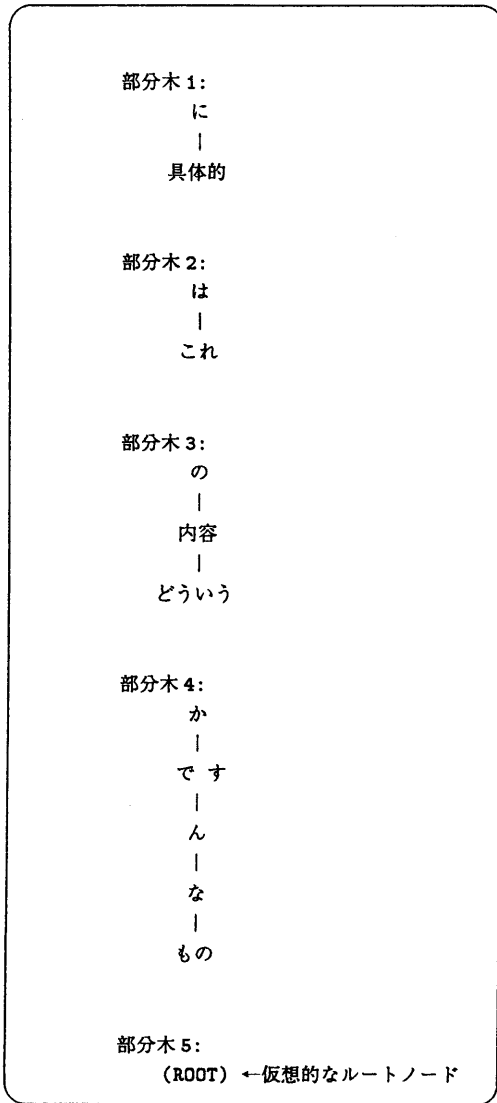


図 2: 部分依存依存構造

係等を総合的に判断して決定する。図 3 に部分木 1 のリンク先の決定後の探索空間を示す。

- 同様に、仮想的なルートノード以外の部分木のリンク先を決定する。図 4,5,6 に、部分木 2 及び部分木 3 の処理後の探索空間の様子、及び最終的な結果を示す。

この手法においては、表層の語順の制約は一切用いていない²。つまり、通常の文節係り受け解析で用いられるような制約(前の語が後の語を修飾するとか、係り受けは交差しないとといったような制約)を利用しない。そのため、いわゆる倒置や交差が生じている発話も問題なく解析できる。例えば、

これはどういう内容の▼ものなんですか▼具体的に
(▼はポーズの位置を示す)

のような倒置を含む発話からも、図 6 と同じ解析結果を得ることが可能である。

3.3 知識の抽出について

係り受けの整合性を判定するための品詞や語に関する知識は、いわゆる辞書的な知識として人間が記述することも可能であるが、辞書的な知識には、

- 開発・保守の手間が大きい。
- 探索時に必要なスコアを適切に付与することが難しい。

という問題がある。

我々は、係り受けの整合性を判定するための知識をあらかじめパーザと人手により解析した依存構造木データベース [8] から抽出することにした。データベースからは、係り受け関係のある語と語の知識だけではなく、その頻度(統計情報)も得ることができるため、探索時に必要なスコアを機械的な手法で決定することが可能になる。

4 議論

4.1 疎結合と密結合

音声認識処理と言語解析(理解)処理の統合方式としては、

- それぞれの処理を独立したモジュールで行なう疎結合方式。
- それぞれの処理を独立したモジュールには分割せず、知識を集中させて処理を行なう密結合方式。

²あくまで制約として用いないだけであり、リンク先の決定時に語順情報を選好として用いることは有効である。

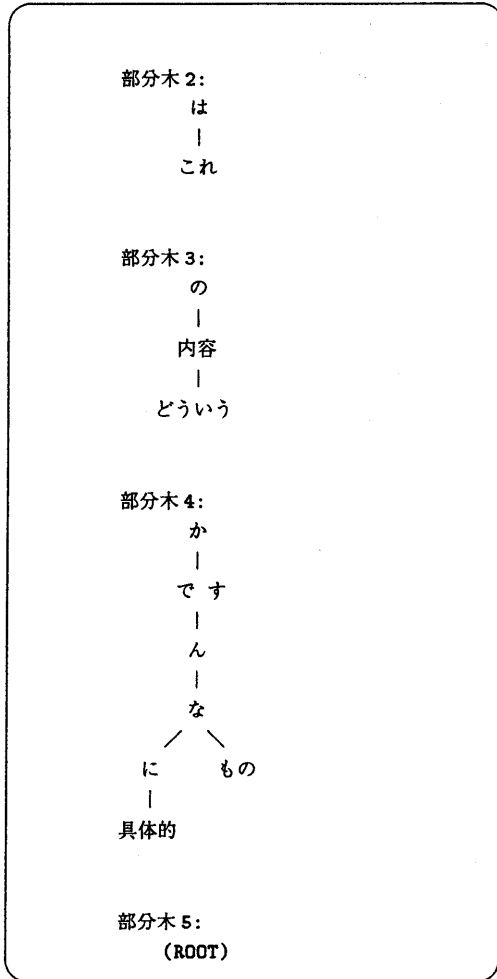


図 3: 部分木 1 の処理後

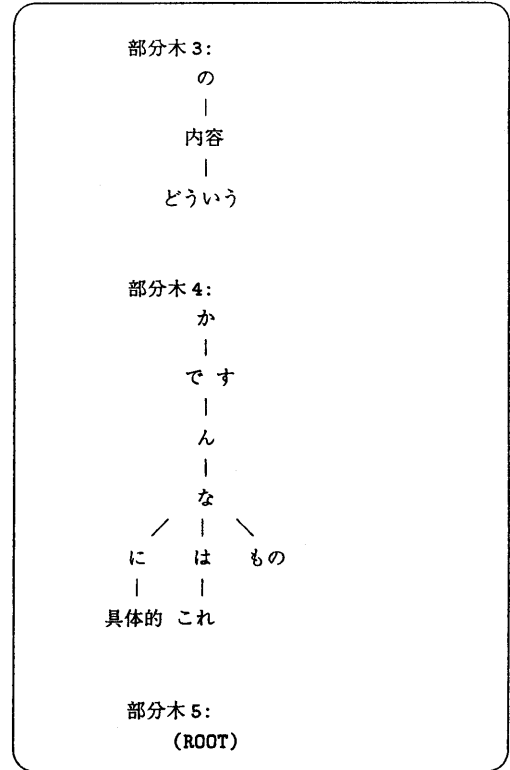


図 4: 部分木 2 の処理後

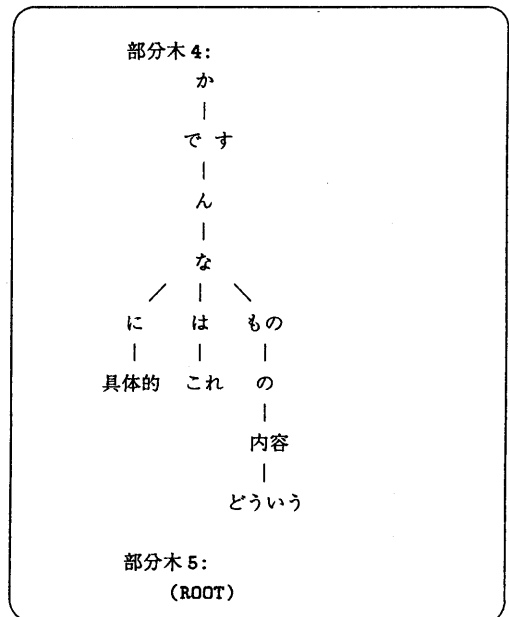


図 5: 部分木 3 の処理後

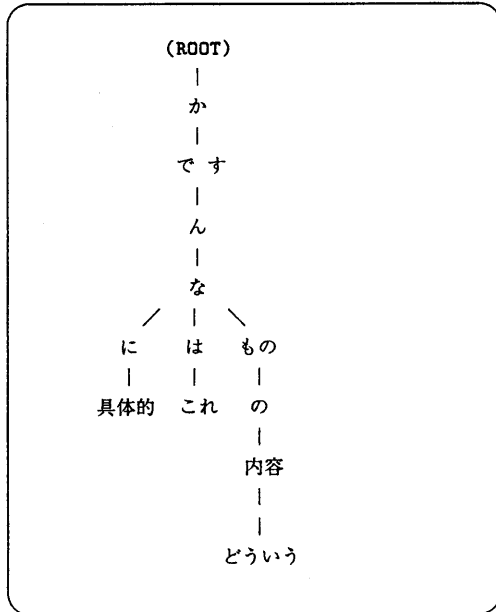


図 6: 部分木 4 の処理後 (最終状態)

の二つに分類でき、それぞれに長所と短所がある。

本稿で提案した部分木併合処理による音声認識処理と言語解析処理の統合方式は、基本的に疎結合方式である。しかし、音声認識処理と言語解析処理は部分木という単位で密接な関係を持っており、音声認識結果の文字列を改めて解析し直すような純粋な疎結合方式の欠点(冗長な処理を行ってしまうという欠点)は解消されている。

4.2 統語主導と意味主導

本稿で提案した手法は、基本的に統語解析を中心とする音声認識・言語処理手法である。一方、キーワードスポッティングやテンプレートマッチング、意味文法 (semantic grammar) 等、対象領域 (domain) の知識を積極的に認識・解析に利用する手法も存在する。

統語主導と意味主導のどちらが優れているかは一概には言えない。領域 (domain) や課題 (task) の変更の容易性 (portability) の点では統語主導方式が、ある特定領域での処理性能は意味主導方式に分がありそうである。

従って、将来の研究としては、

- どうやってポータブルなキーワードや意味文法ベースのシステムを作成するか。
- どうやって統語主導システムに特定領域の知識を組み込むか。

の二つの方向を検討する必要がある。

4.3 併合する部分木のレベル

既に述べたように、部分木併合に基づく処理自体は既に多くのシステムで採用されている手法である。しかし、本手法や文献 [9] のように、依存構造という比較的表層レベルで処理する手法もあれば、文献 [10] のように、領域依存性のある意味構造レベルで処理を行なう手法もあるように、様々なレベルの部分構造で併合処理を行なうことが可能である。

この問題も、統語主導対意味主導の問題と同様の観点で論じることができる。つまり、処理の頑健さという観点では、深いレベルで併合処理を行なった方が強力な曖昧性解消知識を利用できるという点で都合がいい。一方、移植性の観点からは表層レベルでの併合処理が有望であろう。

5 おわりに

本稿では、効率的で頑健な音声認識・言語解析を実現するための部分木併合処理を提案し、処理手法の概要を説明した。また、本手法の特徴を他の手法との比較により明確にした。

今後は実際の音声認識結果データを収集し、実験を繰り返しながら必要な知識の整備、より効率的な探索手法の検討等を行なう予定である。

参考文献

- [1] Takezawa, T. Morimoto, T.: "An Efficient Predictive LR Parser Using Pause Information for Continuously Spoken Sentence Recognition," ICSLP-94 pp.1-4(1994-09).
- [2] 伊藤, 秋葉, 上條, 田中: 休止を区切りとした対話処理, 情報処理学会音声言語情報処理研究会資料, 7-21(1995-07).
- [3] 竹澤, 田代, 森元: 自然発話の言語現象と音声認識用日本語文法, 情報処理学会音声言語情報処理研究会資料, 6-5(1995-5).
- [4] 竹澤, 森元: 部分木を単位とする構文規則を利用した連続音声認識の効果, 日本音響学会講演論文集 (1995-09).
- [5] Jensen, K. and Heidorn, G.E.: "The Fitted Parsing: 100% Parsing Capability in a Syntactic Grammar of English," ANLP83(1983).
- [6] 日本電子工業振興協会: 自然言語処理技術の動向に関する調査報告書, 第2章3.5節及び第5章4節 (1994).
- [7] 竹澤, 田代, 森元: 音声言語データベースを用いた自然発話の言語現象の調査, 人工知能学会言語・音声理解と対話処理研究会資料, 94-03-3(1995-2).
- [8] 田代, 森元: 日本語会話文の言語解析実験, 情報処理学会自然言語処理研究会資料, 108-18(1995-07).
- [9] Nasukawa, T.: "Robust Parsing Based on Discourse Information: Completing partial parses of ill-formed sentences on the basis of discourse information," ACL-95(1995-06).
- [10] Stallard, D., Bobrow, R.: "The Semantic Linker - A New Fragment Combining Method," ARPA HLT-93(1993-03).