

6N-4

キーワード抽出に基づく意味解析による音声対話システム

荒木 雅弘, 西田 豊明, 堂下 修司

京都大学工学部情報工学教室

1 はじめに

本稿では我々の研究室で現在開発中のスケジュール管理に関する音声対話システムにおける文解析手法と対話レベルの知識の利用について述べる。非文法的な発話にも対処できる文解析手法として、スコアの低い少数のキーワードから出発して、増進的に意味候補を作成する意味主導型のisland-drivenバージョンを提案する。意味候補の導出には意味解析用ネットワーク上でのパス解析の結果を利用する。また、オートマトンで表現した対話モデルからの次発話の予測を、ネットワーク上の概念ノードをマーキングすることで文解析に利用する方法を説明する。

2 抽出キーワードからの意味解析

2.1 手法の概要

音声対話システムにおいては省略表現などの断片的な発話、単語の欠落・無意味語の挿入などによる非文法的な発話が現われることから[1]、従来のような文法中心の文解析手法ではなく、意味・対話レベルの知識を利用した方法が必要である。

本システムの文解析(図1)では、まず単語ラティス中から重なりのないように抽出した3単語(3単語以下でラティスを覆ってしまう場合はそれ以下)間の意味ネットワーク(図2)上のパスから意味候補を作成する。単語間の修飾関係が意味候補の中に現われた場合はプレートマッチングによって局所的な検証を行う。この検証は通常のisland-drivenの場合のislandの拡張に相当する。さらにそれぞれの意味候補に関して、次の単語候補とその前後の単語から得られる意味候補を順次統合して、最終的に発話の意味表現を作成する。この手法では、通常のisland-driven解析の探索空間を意味を導入することによって削減しているだけでなく、無意味語の挿入、言いよどみなどを含む非文法的な発話にも対処できると考えられる。

単語ノードは音声処理部でキーワードスポットティングの対象となる単語である。単語ノードは、場合によってはいくつかの上位概念を経由(is-a アーク, instance_of アーク)して単語概念ノードへ至る。単語概念ノードで表されるものは日・時・場所など一般的なものの他に、述部をあらわす登録型動詞・変更型動詞や、また本タスクで対象とするイベントを細分化したもの(参加型・

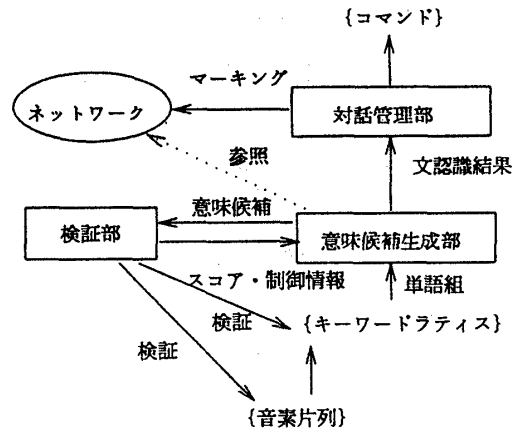


図1: 文解析部の構成図

縮切型など)がある。概念間で修飾関係を持ち得るものに対して limit アーク, attribute アークが張られる。また、文の意味を構成する開始時刻格・場所格・目的格などが格ノードとして用意され、それらが関連する単語概念と element_of アークで結合される。ネットワークの最上位には文の解釈に対応する登録文・問合せ文・回答文などの文ノードがあり、それぞれ構成要素となる格と case アークによって結合される。

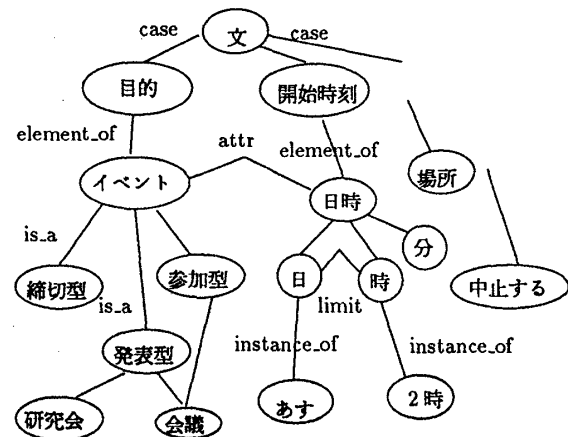


図2: 解析用ネットワークの一部

2.2 パスから意味表現への変換

例えば単語組 [あす, 2時, 会議] が入力された時の処理を以下に説明する。ノード「あす」とノード「2時」,

⁰A Speech Understanding System based on Multiple Keyword Spotting and Network-path Analysis

表 1: 意味候補の例

バス	意味候補
(1)-(a)	[concept(kaigi),attr([day(asu),hour(2))]]
(1)-(b)	[sent,[start_time,[day(asu),hour(2)],[obj,kaigi]]]
(1)-(c)	[sent,[end_time,[day(asu),hour(2)],[obj,kaigi]]]
(2)-(a)	[sent,[start_time,[day(asu)]]] & [concept(kaigi),attr([hour(2)])]
(2)-(c)	[sent,[end_time,[day(asu)]]] & [concept(kaigi),attr([hour(2)])]

ノード「2時」とノード「会議」のバス解析を行うと、以下のようなバスが得られる。ただし、入力文はユーザからの第一発話とし、簡単のためすべての文の解釈を文ノードとしてまとめたもので説明する。

「あす」と「2時」

- (1) あす→日→(limit)→時←2時
 (2) あす→日→日時→開始時刻格→文←終了時刻格
 ←日時←時←2時

「2時」と「会議」

- (a) 2時→時→日時→イベント←参加型←会議
 (b) 2時→時→日時→開始時刻格→文←目的格←イベント
 ←参加型←会議
 (c) 2時→時→日時→終了時刻格→文←目的格←イベント
 ←参加型←会議

ここで(1),(2)からと(a),(b),(c)から1つずつ組み合わせる文の意味候補を作るのであるが、(2),(b),(c)の場合のように抽出された単語が1文中での格を構成する部分であるとき(格仮説)は、格仮説どうしの矛盾がないような組合せのみが意味候補として残る。また、後で述べる対話構造に関する知識などから、特別にマーキングされたノードを通るバスが存在する場合はそれ以外のバスは除外される。このようにして、表1のような意味候補が得られる。

2.3 意味候補の検証

(1)のようにバスにlimitアークが存在する場合や、(a)のようにattrアークが存在する場合はその2単語を埋める規則は少数であるので、局所的に構文チェックを行い、仮説のスコアをつける。局所的なチェックにはテンプレートをを用いる。また、助詞などはキーワードラティスに現われていないので、音素片列に戻って該当範囲に対して予測単語の抽出を行う。

局所的な検証後、最初の単語組に重ならず、またその単語と隣接する単語とのバス解析の結果が意味候補と矛盾のないものをあらたに単語組に加えてゆく処理を文解析の終了まで行う。文解析の終了はラティスが単語連接の許容範囲を残して覆い尽くされるか、または単語組の隙間に存在する単語(それまでの意味候補と矛盾なく共存するもの)のスコアが閾値より低くなった時点で判定

する。全ての意味候補の検証が終了した時点で、単語認識のスコアの合計とそれらの単語がラティスをカバーしている度合を掛け合わせたものを文のスコアとし、一番文スコアの高いものを入力文の意味表現とする。

3 対話の場を表現したオートマトンによる対話管理

本システムでは対話の場を表現したオートマトンによってシステムの返答の決定およびユーザの次発話の予測を行っている。ユーザの次発話の予測には以下の情報が用いられる。

1. ユーザゴール ユーザの第一発話により対話が行われるオートマトンを特定し、それに対応してネットワーク中の主に動詞型概念をマークする。ただし、ユーザゴールは他のオートマトンの呼び出しによって入れ子構造になることもある。
2. 対話の構造 オートマトンの現在の状態から出ているアークに対応する解釈が予測される次発話の解釈である。その解釈に対応したネットワーク中の文ノードをマークする。
3. イベントの種類 対話の中でイベントの種類が特定されると、それと関連する単語集合が絞られる。イベントの種類とは参加型・縮切型・旅行型・休暇型であり、それらが定まると次発話でのキーワードスポッティングの対象となる時間や場所のスケール、関連イベントなどが絞られる。絞られた概念を表すノードをマークする。

これらの情報を使って意味ネットワーク上のノードにマークをつける。文解析部で意味候補を導出する時にこのマークがついているバスを優先的に採用する。また、次単語をラティスから抽出する時、および第2発話以降のスポッティングする時の対象となる単語は上位概念がマークされている単語のみとなる。

4 おわりに

キーワード抽出からの意味主導型のパーズング手法および音声対話システムにおける対話レベルの知識の利用法について述べた。本システムはキーワードラティスの生成、対話管理部との結合が終了すれば松本ら[2]と同じタスクで評価を行う予定である。

参考文献

- [1] Ward W. Understanding spontaneous speech: The PHOENIX SYSTEM. In *ICASSP-91*, pp. 365-367, 1991.
- [2] 松本真治, 河原達也, 堂下修司. 語彙-構文-意味制約を統合したA*探索による会話音声認識. 信学技報 91-SP, 1991.