

命題情報とモダリティ情報の個別抽出に基づく文解析手法の提案

4G-7

熊本忠彦 伊藤 昭 海老名 毅
郵政省 通信総合研究所 関西先端研究センター

1. はじめに

我々は、対話によってユーザの計算機利用を支援するシステム(以下、支援システムと呼ぶ)を開発している^[1]。ユーザは、何らかの障害/問題が発生したとき、支援システムに助けを求めることができる。支援システムは、ユーザの発話(話し言葉)を理解し、そのときの計算機の状態に合わせて適切な応答を生成する。

ある発話意図を話し言葉で表現しようとするとき、その表現方法は多様である。しかしながら、その多様性の多くは、命題情報そのものではなく、モダリティ情報の表現方法に起因しているものと考えられる。従来の意味解析手法は、発話文の命題情報とモダリティ情報を混在して取り扱うため、モダリティ情報の表現において観測される多様性の影響を受けやすい。

本稿では、発話文から命題情報とモダリティ情報を個別に抽出し、それらを並列に解析する手法を提案する。但し、抽出されたモダリティ情報を解析する手法はすでに別稿^[2]で提案しているため、本稿では、命題情報とモダリティ情報の個別抽出法、および抽出された命題情報の解析手法を提案する。

2. モダリティ表現の多様性

我々は、支援システムの代わりに人間コンサルタントが支援するという疑似的な「対話による支援」実験を行い、支援対話データベースを作成した^[3]。この支援対話データベースからユーザ発話文を抽出し、ユーザ発話文コーパスを作成した。ユーザ発話文には、日常会話文と同様、間投表現や言い直し、言い淀み、言い換え表現、倒置などが現れていたが、これらはコーパス作成にあたって取り除かれた。このような表現の取り扱いは今後の課題である。

ユーザ発話文コーパスによれば、ユーザは、「あるメッセージを別のフォルダに移す」という行為の手順を尋ねるとき、表1に示されるような表現を用いた。

表1 メッセージの移動方法に関する質問文の例

メッセージの移し方がわかりません
メッセージの移動ってどうやるんですか
メッセージの移動の仕方がわかりません
メッセージの移動の方法を教えてください
メッセージを移動するにはどうしたらいいんですか
メッセージを移動するにはどうすればいいんですか

これらの発話文はすべて同じ発話意図を有するが、そのモダリティ(「ある行為の手順を尋ねる」)は多様な表現となって現れている。

3. 文の意味(発話意図)の記述^[2]

発話文は命題とモダリティから成ることが知られている^[4]。我々は、命題(タスク領域上で定義可能な「もの」や「行為」、「状態」)を表現する記述を「発話命題」と呼び、フレーム表現で記述する。また、モダリティ(発話命題に対するユーザの態度(attitude))を表現する記述を「発話意図タイプ」と呼び、シンボルで記述する。従って、一つの発話意図は、

(発話意図タイプ 発話命題)

という形で記述される。

4. 命題情報とモダリティ情報の個別抽出

入力された発話文はまずJUMAN^[5]によって形態素解析される。JUMANは、入力発話文を適当な形態素に分解し、各形態素に原形や品詞名、品詞細分類名、活用形といった情報を付加する。

命題情報とモダリティ情報は、JUMANの出力である形態素列から個別に抽出される。その様子を表2に示す。モダリティ情報には、「機能語」や品詞情報(品詞名/品詞細分類名)、活用形情報、原形情報の4種類がある。機能語は、本論文では発話意図タイプの推定に影響を与える意味情報の種類と定義される。また、命題情報には、原形情報と品詞情報があり、「もの」や「行為」、「状態」を表現する語(名詞や動詞、形容詞、判定詞など)、およびそれらの関係を表現する語(助詞)から抽出される。

On a Semantic Analysis Method Extracting Propositional Information and Modality Information Separately from a Spoken Japanese Sentence

Tadahiko KUMAMOTO, Akira ITO, and Tsuyoshi EBINA

Kansai Advanced Research Center, Communications Research Lab., MPT of Japan

表2 命題情報とモダリティ情報の個別抽出の例

モダリティ情報	対応する入力形態素	命題情報
普通名詞←	メッセージ	→メッセージ, 普通名詞
	を	→を, 格助詞
サ変名詞←	移動	→移動, サ変名詞
する←	する	→する, 動詞
には←	には	
HOW句←	どうしたら	
OK述語←	いい	
	んです	
疑問助詞←	か	

5. 命題情報の意味解析手法

抽出された命題情報は意味主導で解析され、発話命題が生成される。例えば、発話文「メッセージを移動するにはどうしたらいいんですか」から抽出された命題情報「メッセージ(普通名詞)」および「を(格助詞)」、「移動(サ変名詞)」、「する(動詞)」に対して、普通名詞「メッセージ」からは、ものフレーム「メッセージ」が生成される。但し、そのフレームの格情報属性には格フレーム「を」が束縛される。また、サ変名詞「移動」からは、行為フレーム「リソース移動」が生成される。その行為フレームの対象属性に前述のものフレーム「メッセージ」が束縛され、その結果、行為フレーム「リソース移動」は行為フレーム「メッセージ移動」に特殊化(specialize)される。この行為フレーム「メッセージ移動」が発話命題である。

6. 評価

命題情報とモダリティ情報の個別抽出を行う場合と行わない場合とでは、意味解析の対象となる節^[6]の数が大きく異なる(図1参照)。1文あたりの節数は、個別抽出を行うことによって1.9節/文から1.2節/文に減少した。このような現象は、例えば表2に示されるように、「メッセージを移動するには/どうしたら/いいんですか」という複文(節数3)が「メッセージを移動する」という単文(節数1)として解析されるため、生じる。解析対象となる節数の減少は、文解析の安定化および効率化を意味する。

図1は、命題情報なし(節数0)の場合があることを示している。このような場合には、モダリティ情報

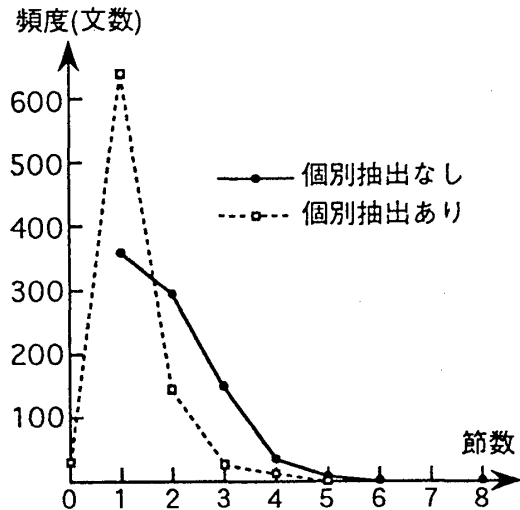


図1 意味解析の対象となる節の数

の解析結果からデフォルトの発話命題を生成できる。

7. まとめ

本稿では、発話文から命題情報とモダリティ情報を個別抽出することによって、意味解析の安定化/効率化を行う手法を提案した。

今後の課題は、ユーザ発話文コーパス作成にあたって取り除かれた現象(言い直し、言い淀みなど)にどう対処するかである。

参考文献

- [1] 熊本忠彦, 伊藤昭, 海老名毅: “電子メール処理プログラム(XMH)利用支援システムの試作”, 第7回人知全大, 20-3, pp. 555-558 (Jul. 1993).
- [2] 熊本忠彦, 伊藤昭, 海老名毅: “支援対話におけるユーザ発話意図の認識”, 信学論, J77-D-II, 6, pp. 1114-1123 (Jun. 1994).
- [3] Kumamoto, T., Ito, A., and Ebina, T.: “Construction of a dialogue database from advisory dialogues between novice computer users and an expert consultant”, *Proc. of International Workshop on Sharable Natural Language Resources*, Nara, Japan (Aug. 1994).
- [4] 益岡隆志: “モダリティの文法”, くろしお出版, 東京(1991).
- [5] 妙木裕, 松本裕治, 長尾真: “汎用日本語辞書および形態素解析システム”, 情処第42回全大(1991).
- [6] 益岡隆志, 田窪行則: “基礎日本語文法”, くろしお出版, 東京(1991).