

2F-1 日本語インタフェース文脈文法 - 解析手法 -

宮部 隆夫

NEC 関西C&C研究所

1. はじめに

日本語インタフェースシステムの重要な利点の一つとして、ユーザの簡略な入力表現をシステム側が文脈を利用して補完する機能が挙げられる。ユーザ・システム間対話は、特定タスク(対象領域と話題の範囲が制限される)を遂行する目的指向の協調的対話に限定される(場合が多い)。このため、通常の対話に必要な多量の常識や高度な推論機能などを利用しなくとも、この特徴(制限)を優先規則として利用する簡便な方式が有意である。本稿では、上記特徴を利用した文脈処理手法(機構と文法)を提案する。日本語インタフェース処理機構[1]を基本枠とし、対話の前文脈情報を焦点スタックで管理する機構[2]を備えたと共に、以前提示した文脈処理の問題点・必要機能[3]を解消・実現する文脈文法を有する。

2. 特徴と問題点

インタフェース内対話では、話題やその推移過程がタスク関連の対象領域知識により強く規制・限定される場合が多い。「省略表現が多い」、「接続表現が少ない」、等の特徴があり、陽に記述・表現される情報は少ない。

(2-1). 対象領域知識利用による省略補完

(前文脈) NECの本社の住所を教えてください。

(入力) 関西C&C研も。

(解釈) 更に、関西C&C研の住所も教えてください。

従って、表示されない知識を補うために、対象領域に関する詳細な知識が必要となる。必要な知識として、以下の情報が挙げられる[3](図1)。

類義語/下位語	会社と企業, 会社とメーカー, 等
対象と属性	会社と経常利益, 会社と社長, 表示と順序, 等
多義語 (派生語)	会社(組織と名前), 平均(動作と結果), 等

図1. 必要な対象領域知識

また、文脈文法に以下の機能が必要となる。

(2-2) 先行詞候補の選択1 焦点の重視

先行詞の候補として、前文脈中の焦点を優先する。

(前文脈) N社の株を1万株以上もつ会社は?

(入力) そこの社長は?

(解釈) N社の株を1万株以上もつ会社の社長は?

候補「N社」と「会社」の内、焦点は「会社」

(2-3) 先行詞候補の選択2 照応表現の分類(値と属性)

「属性」か「値」かの差が照応相手を限定する。

(前文脈) 徳島に支社のある会社の名前は?

入力が「属性」ならば、前文脈中の「属性」と対応

(入力) 本社住所は? (属性)

(解釈) 徳島に支社のある会社の本社住所は?

入力が「(属性)値」であれば、「値」と対応

(入力) 大阪は? or N社のある地区? (値)

(解釈) 大阪に支社のある会社の名前は?

or N社のある地区に支社のある会社の名前は?

(2-4) 派生語の曖昧さの保存

派生語に関して、文中で別の意味で利用する場合がある。

(入力) 本社が大阪の会社とその営業利益は?

(解釈) 本社が大阪の会社の名前と、本社が大阪の会社の営業利益?

会社は、「会社名」と「会社(組織)」の曖昧さをもつ

(2-5) 構文構造と意味表現の差の吸収

(前文脈) NECの本社の電話番号は?

(入力) 手紙の宛先は?

(解釈) NEC(に出す手紙)の本社住所(宛先)は「(本社への)手紙の宛先」と「(本社)住所」を同義と解釈する機能

なお、入力はキーボードを仮定し、また、対象システム例としてDB検索システムを取り上げた。

3. システムの設計方針

システム設計に際しては、2章で提示した問題・機能を実現するために、以下の項目を基本方針としてシステムを設計した。

1). 対象領域(システム)知識の一元化[4]

対象領域知識は、文脈処理に不可欠である(2-1)。この知識は意味(タスク)解析でも利用する大容量の知識であり処理の整合性の保持するために一括管理する。

2). 複数(2重)文脈モジュール構成

通常の文脈処理を行う構文文脈処理モジュール(文脈推定部:優先解釈を行う)と,(2-5)に記したように意味表現への変換結果に対しての文脈処理を行う意味文脈モジュール(文脈確定部:意味の確認と確定)との2重構成とする。

3). 前文脈との結束性を重視した優先解釈(preference)

入力文中の照応表現や省略表現に対応する前文脈中の先行詞候補に曖昧さがある場合の選択基準(2-2, 2-3)

3-1). 焦点の選択

先行詞として,前文脈中の焦点(及び,それと関連する候補)を優先する(2-2)。文の焦点候補として,対象格・対象相当格をとる。

3-2). 属性・値の区別による照応関係推定

先行詞選択に際して,照応詞が属性・値かにより,それと同じカテゴリを先行詞として選択する(2-3)。

4). 派生語による曖昧さの保存

通常多義語の曖昧さと派生語間の曖昧さとを区別する。派生語については(2-4),処理中・処理後もその曖昧さを保持し,同じ文(章)中であっても異なる意味での解釈を許容する。

4. システム・文法構成

前章で提示した方針に従った文脈処理機構(図2)と文法(図3)を提示する。

システム(機構)は,データベースシステムを対象としたインタフェースシステム内の文脈解析モジュールとして実装する。意味(タスク)解析モジュールを間にはさんで,構文文脈処理(文脈推定)と,意味文脈処理(文脈確定)との2モジュールとして構成する。両文脈処理モジュールとも,文脈文法規則を介して,前文脈を管理する焦点スタックを参照する。(図2)

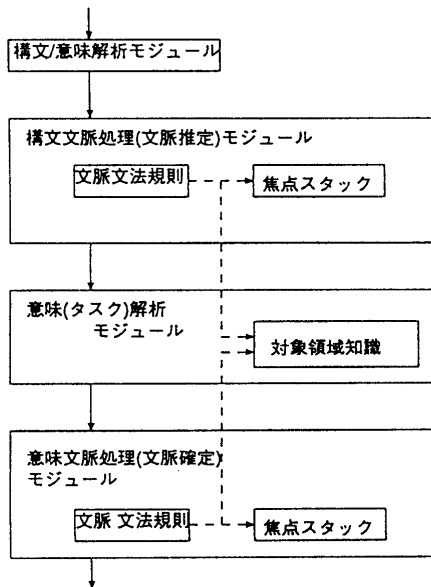


図2. モジュール構成図

文法は,構文文脈文法を基本枠とし,そこで扱えないような機能(2-5)等を,意味文脈文法として規則化する。

構文文脈文法:入力文の分析・分類・先行詞の探索・前文脈との関係付けの4部分規則で構成する。入力文中の焦点・照応詞などを検出した後(分析),話題の関連(結束性)と対象領域知識を局所的に利用して照応・省略補完処理をし(分類・探索),補完した構造の関係を推定する(関係づけ)。(図3)

入力文の分析	
・焦点の抽出	
・照応(指示)表現の抽出	指示詞など
・省略部位の抽出	欠落必須格など
入力文(部分)の分類	
・対象領域内の表現	対象領域知識と対比
・対象システム内の表現	対象システム知識と対比
・メタ質問	キーワード利用
先行詞(前文脈中)の探索(複数候補)	
・探索規則	
焦点優先	object など
属性値と属性の区別	instance, 修飾
・意味制約	対象領域・システム知識
前文脈との関係づけ	対象モデルへの部分マッピング
・置換	処理条件 or 処理対象の置き換え
・追加	追加
・緩和(削除)	処理条件の緩和 or 処理対象の削除
・進行	(前文脈)処理対象の処理条件化

図3. 文脈文法規則 基本構成

意味文脈解析: 対象システム上の意味モデルへマッピングされた結果の意味表現に対して,意味解析部で矛盾が生じた場合や複数の意味解釈結果が生じた場合,及び,(2-5)に記したように構文文脈推定が困難であった場合に,文脈処理の観点から意味の再解釈・選択を行う。なお,処理結果に対して,ユーザへの確認を行う。

5. おわりに

日本語インタフェースシステムにおける文脈処理上の問題点とその解決案を提示した。更に,その手法を実現するシステムの構成と文脈文法も示した。本システムは,自然言語インタフェースキット I F - K i t [3]の文脈処理モジュールとして構築し,評価・改良を進めている。

参考文献

[1] 李・市山「データベースの日本語インタフェースにおける文脈処理」43 回情処全大 5H-9 (1991)
 [2] Sidner, C.L. "Focusing in the Comprehension of Definite Anaphora" Computational Models of Discourse
 [3] 宮部 「ヒューマンマシン対話における照応解析」44 回情処全大 5P-5 (1992)
 [4] 谷,市山 他 「自然言語インタフェース構築キット: I F - K i t」 信学研 NLC91-62