

## 英語質問文の意味解析方式

4R-6

別所克人 岩瀬成人

NTT情報通信研究所

### 1. はじめに

英語の問合せ入力文から、店・企業等の情報を検索・案内するイエローページサービスシステムを研究している。本システムでは、入力文の意味構造からユーザの意図を表すキーワードを抽出し、そのキーワードを用いて、あらかじめシステムが保持する推論知識を活用して、入力文に対応するサービス名を推論する。

我々は、イエローページサービスシステムへ入力される文の意味構造をいくつかのパターンに分類し、そのパターン毎に、抽出するキーワードの種類とキーワードの意味構造における位置の関係を整理し、その結果に基づきキーワード抽出処理として、文の主辞となる動詞を始点にして、特定の深層格の語句を再帰的にたどっていく方式を提案した[1]。

英文の意味構造を得る処理すなわち意味解析の手法としては、拡張LINGOLのように構文解析に用いる文法規則中に意味処理関数を記述し、ボトムアップに意味解析を進める方法等があるが[2]、各文法規則毎に複雑な意味処理関数を作成しなければならず、保守者の負担が重くなるという難点がある。またボトムアップに解析を進めていくため、得られた意味構造を利用するアプリケーションが必要としない深層格の内部の意味解析までも行い、効率が悪くなる問題もある。

これに対し本稿では、トップダウン式に処理を進めていく意味解析の一方式を提案する。本方式の長所として以下のことが挙げられる。

- (1) 構文木をトップダウンにたどっていくため、意味解析後のキーワード抽出処理が必要とする最低限の範囲の意味構造を求めることができる。
- (2) 各句毎に、抽出すべき主辞や主辞に対する表層格、修飾句を構文木上の位置として規定するため、意味解析処理全体の流れを容易に把握できる。

### 2. 構文構造と意味構造

意味解析アルゴリズムを述べるにあたり、本節では意味解析処理の入力となる構文構造と出力となる意味構造について述べる。

本システムにおける構文解析は、拡張遷移ネットワーク(ATN) [3]を採用しており、正しい構文構造(構文木)を得るために、入力文中の各単語に対応する品詞の順序

のみならず、句や品詞の人称・数といった属性のチェックも行っている。得られる構文木の各ノード(句や品詞)は、後の意味解析でその情報を活用してチェックを行うために、構文木セルと呼ばれる形をとっている。その内容は、種別(句の細分類)や辞書意味、態といった句や品詞の属性とその属性値の組の集合である(図1参照)。

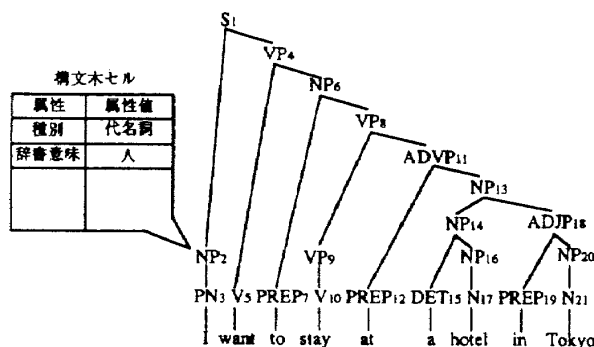
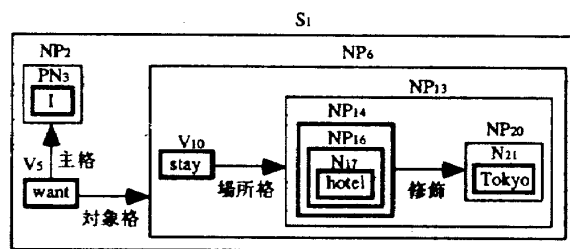


図1 "I want to stay at a hotel in Tokyo."の構文構造

意味解析処理結果の意味構造は、意味フレームと呼ばれるものの集合であり、各意味フレームはある構文木ノードと対応づけられている。後のキーワード抽出処理で意味フレーム間を適切にたどっていきけるようにするために、意味フレーム間は、主辞関係、深層格関係、修飾関係の3種類の方向づけられたリンクにより結合されている(図2参照)。



各枠線が一つの意味フレームにあたり、濃い枠線の意味フレームはそれを囲む一つ外側の意味フレームの主辞フレームである。主格、対象格、場所格のリンクは深層格関係のリンクであり、修飾のリンクは修飾関係のリンクである。主辞関係のリンクは省略している。

図2 "I want to stay at a hotel in Tokyo."の意味構造

### 3. 意味解析アルゴリズム

構文木は句の下にまた句があるといった入れ子構造をしているため、意味解析アルゴリズムは構文木を再帰的にたどる形になる。処理は構文木の根のノード(S)のフレームを設定し、それを入力として全体処理(入力フレームに対応する句内の意味構造を求める処理)を呼び出すことによって行う。ある句に対応するフレームを新規に設定したときに、それを入力として全体処理を再帰

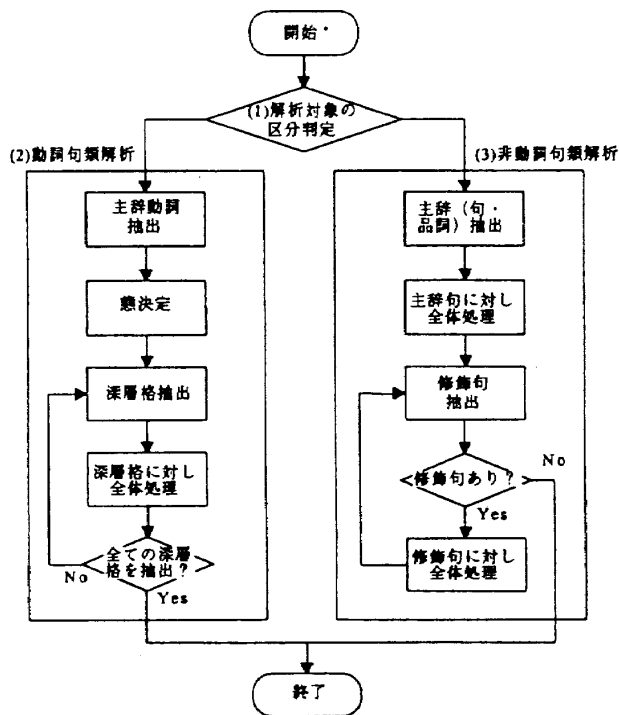


図3 意味解析の全体処理

呼び出しする(図3参照)。

意味解析は、動詞とその深層格との関係を解析する処理(動詞句類解析(2))と、名詞句と形容詞句間等の修飾関係を解析する処理(非動詞句類解析(3))とに大別される。以下、図3中の各モジュールの説明をする。

(1) 解析対象の区分判定

解析対象(入力フレームに対応する句)に対し、上記前者の処理か後者の処理のどちらを行うかを、句とその種別により決定する。表1にその分類の一部を示す。解析対象で、前者の処理を行うものを動詞句類、後者の処理を行うものを非動詞句類と呼ぶ。

表1 解析対象の区分への分類の一部

| 句          | 動詞句類                            | 非動詞句類           |
|------------|---------------------------------|-----------------|
| 文(S)       | 直下にVPを含む                        | 直下にVPを含まない      |
| 名詞句(NP)    | 種別:不定詞句<br>種別:動名詞句              | 種別:一般<br>種別:代名詞 |
| 形容詞句(ADJP) | 種別:不定詞句<br>種別:現在分詞句<br>種別:過去分詞句 | 種別:一般           |
| 副詞句(ADVP)  | 種別:不定詞句                         | 種別:一般           |

(2) 動詞句類解析

本方式では、各解析対象(非動詞句類を含む)の種類(句と種別の組)毎に、その主辞となる句・品詞や、主辞に対する各表層格、修飾句の構文木上の位置をあらかじめ決めておく(図4参照)。

動詞句類解析では、解析対象(例:図1のS1)の主辞動詞(同V5)をまず求め、解析対象フレームから主辞動

詞のリンクが張られた主辞動詞フレームを設定する。次に構文木セルを参照することにより、解析対象の態(能動態か受動態)を決定する(図1のS1の態は能動態)。

本方式では動詞の辞書意味は、各態毎の動詞の深層格と表層格の対応及びその深層格に格納される句の属性を定めた制約条件により分類されている(図5参照)。

態と主辞動詞の辞書意味の組合せにより、指定した深層格(例:図5の対象格)に対応する表層格(図1のNP6)を求め、それが制約条件を満たせば、主辞動詞フレームから深層格関係のリンクが張られた深層格フレームを設定し、それを入力として全体処理を呼び出す。

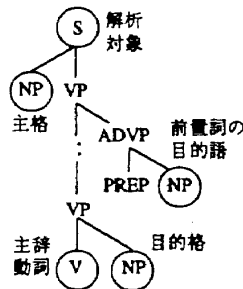


図4 文(S)の主辞とその表層格の位置

| 深層格 | 表層格 |        | 制約条件       |
|-----|-----|--------|------------|
|     | 能動態 | 受動態    |            |
| 主格  | 主格  | byの目的語 | 名詞句で辞書意味:人 |
| 対象格 | 目的格 | 主格     | 名詞句で辞書意味:事 |

図5 動詞の辞書意味の一例(図1のV5の辞書意味)

(3) 非動詞句類解析

解析対象(例:図1のNP13)の主辞となる句・品詞(同NP14)を求め、解析対象フレームから主辞関係のリンクが張られた主辞フレームを設定する。主辞が句の場合、主辞句フレームを入力として全体処理を呼び出す。次に、主辞の修飾句(同NP20)を求め、主辞フレームから修飾関係のリンクが張られた修飾句フレームを設定し、それを入力として全体処理を呼び出す。

4. おわりに

本稿では、英語入力文の意味解析の方式を提案した。本方式の処理の特徴として以下のことがあげられる。

- ・ATNに基づく構文解析結果中の構文木セルの情報を、意味解析時の種別、辞書意味、態等のチェックのため活用するようにした。
- ・解析対象が動詞句類とそうでないものとで処理を分け、構文木をトップダウンに再帰的にたどることによって意味構造を生成するようにした。

今後、本方式を組み込んだプロトタイプの推論システムの推論正解率の評価を行い、本方式の検証を進める。

参考文献

[1]別所,岩瀬,戸部:自然言語処理による分野名推論方式—英語版推論方式の実現—,情処第55回全国大会講演論文集(2),pp.93-94, Sep. 1997.  
 [2]石崎俊:自然言語処理,昭晃堂  
 [3]長尾真編:自然言語処理,岩波書店