

自然語インタフェースにおける語句学習方式

7P-6

木山 忠博

絹川 博之

(株)日立製作所 システム開発研究所 関西システムラボラトリ

1. はじめに

自然語インタフェースにおいて、利用者固有の語句を受理することは必須要件の一つである。その手段として、これまで利用者が使用すると予想される語句をシステム管理者が予め辞書登録する方法や、語句が理解できない場合に利用者が辞書エディタを使用し登録する方法が用いられてきたが、登録作業は容易ではなかった。本稿では、データベース検索を目的とした入力文を対象に、解釈不可能な単語が存在する場合に、その単語が持つべき単語情報(品詞、操作指示対象情報など)を推定し学習する方式を提案する。

2. 語句学習方法

2.1 限定表現による語句学習

特定の文型に限って入力文中の未知語の単語情報を推定し推定結果が正しいか否か利用者に問い合わせ、正しいと認められた場合に、推定した単語情報を辞書登録する(図1)。語句学習の対象とする表現は次の(1)~(2)の表現である。

(1) 「AがBである」「AがB以外である」等の
AとBの従属を表わす表現

この場合のA・Bに相当する語は、データベースのカラム名・カラム値を表わす語である。「メーカーがA衣料である商品は」という検索文を用いて説明する。例えば、「メーカー」が未知語であり「A衣料」は図2の「製造元」に対応するカラム値とすると、AとBの従属関係が認識できるので、「メーカー」は「A衣料」というカラム値のカラム名に相当する語であると推定する。即ち「メーカー」は「製造元」を表わす語であると推定する。

(2) 「AがBまたはCかつD」のBとCとDの
論理和・論理積を表わす表現

「メーカーがA衣料またはB紡績の製品は」という検索文を用いて説明する。例えば、「メーカー」と「B紡績」が未知語であるとする。「メーカー」は前記(1)により「製造元」を表わす語であると推定される。次に、BとCの論理和関係が認識できるので、「B紡績」は「A衣料」と同じカラム名のカラム値を表わす語であると推定する。

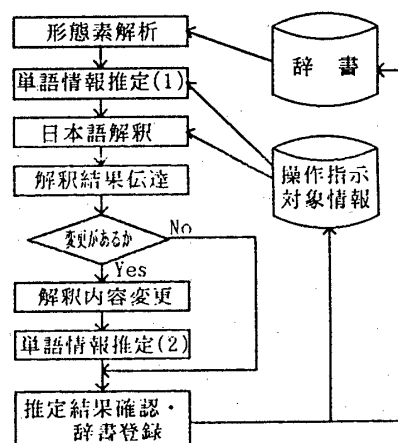


図1. システム構成

製品名	価格	製造元
スカート	25000	A衣料
スラックス	13000	B紡績

図2. 対象データベース例

2. 2 利用者による変更情報を利用した語句学習

未知語や解釈不可能な語を無視し解析を進め、その解釈結果を利用者に提示し、解釈結果に満足でない場合は、利用者が解釈結果を変更する。この時の利用者による変更情報と未知語や解釈不可能な語の特徴とが一致する場合に、図1の単語情報推定部(2)の変更情報を基に未知語や解釈不可能な語が持つべき単語情報を推定し、推定結果確認・辞書登録部で推定結果が正しいか否か利用者に問い合わせ、正しいと認められた場合に、推定した単語情報を辞書登録する。

122文の例文を収集し、この例文から検索文の特徴を検出し、本方式の語句学習の対象とした。検索文の特徴を次の(1)~(4)に示す。

(1) 連体修飾を受けない名詞はカラム値を表わす語であることが多い。

また、「に」や「の」等の助詞を伴う。

[例文1] スラックスの価格は？

(2) 文の最も後方に現われる連体修飾を受ける名詞はカラム名であることが多い。言い換えれば、文末には検索対象が現われることが多い。

[例文2] スカートの価格は？ 5万円より高い商品の製造元は

(3) 疑問指示代名詞が修飾する語または修飾される語はカラム名であることが多い。

[例文3] スカートを作っている会社はどこ？

(4) 形容詞または形容動詞は数値による比較表現を受けることが多い。

[例文4] 5万円よりも高価な商品は

上記(1)~(4)の検索文の特徴と、利用者による変更情報が一致する場合、即ち、例文の2重下線部に相当する語が未知語や解釈不可能な語であり、検索文の特徴に合った解釈内容の変更を利用者が行った場合に、単語情報を推定する。

また、解釈内容の変更機能として対話型のメニュー処理による変更機能を用意している。

本方式による単語学習の例を図3に示す。

3. おわりに

本方式に基づいた語句学習機能をデータベース用自然語インタフェースに適用した。未知語の学習は自然語インタフェースの性能を向上させる重要な課題の一つである。今後は、語句学習の対象となる表現を拡大し評価を行うとともにツール化を図る''''予定である。

参考文献

1) 間瀬、木山、絹川：自然語インタフェースにおけるインタラクティブ型多義解消方式の開発、情報処理学会第44回全国大会(1992)
 2) 西森、木山、絹川：汎用日本語形態素解析ツールの開発、情報処理学会第44回全国大会(1992)

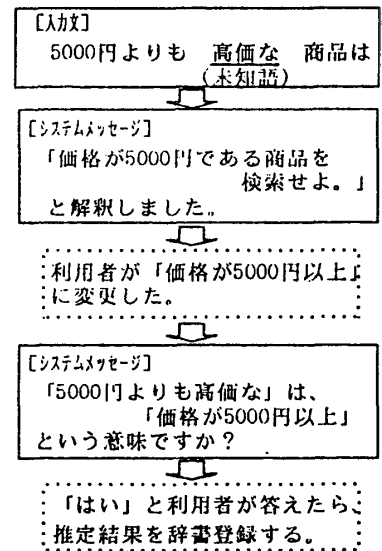


図3. 単語学習の例