

2F-7 自然語インタフェースにおける解釈結果確認文生成方式の開発

間瀬久雄 木山忠博 辻 洋 絹川博之
(株)日立製作所 システム開発研究所

1. はじめに

本稿では、DB検索の作業効率向上を目的とした問合せ日本語文の解釈結果を表す日本語文(確認文)生成方式について報告する。確認文により、(1)解釈結果の確認、(2)解析能力の限界の把握、(3)入力文への流用、(4)解釈結果の音声出力が可能となる。これらの実現のためには、「一義性」と「読みやすさ」が確認文に要求される。本方式では、(a)多義でない語句の選択、(b)省略表現の補完により一義性を実現し、(c)管理者の定義した語彙・構文の利用、(d)入力文の記述順序に対応させた検索条件の記述により読みやすさを向上する。なお、本方式はDBの内容には依存しない、汎用性を保っている。

2. 確認文生成の処理概要

図1に確認文生成の処理フローを示す。本処理は、木構造内部表現を入力とする。まず、検索対象を決定し、次に、図3の検索条件句・検索対象部・処置部に対応する日本語をそれぞれ生成する。次に、生成した検索条件句の記述順序を決定し、この順序に基づいて確認文を組み立てる。

3. 確認文生成方式

以下では図2のDBを想定し、次の例文を採り上げる。

(例1)「日立の男性の買ったテレビの平均価格は?」

確認文の構成を図3にBNFで示す。確認文は「検索条件部」「検索対象部」「処置部」の順で記述する。検索条件部は、いくつかの検索条件句からなる。(例1)では、検索条

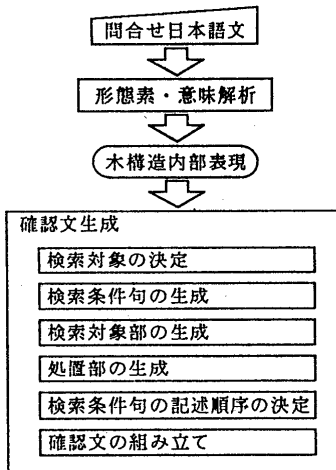


図1 確認文生成の処理フロー

Generation of Sentence for Confirming Analysis Result for Natural Language Interface
HISAO MASE, TADAHIRO KIYAMA, HIROSHI TSUJI,
HIROSHI KINUKAWA
Systems Development Laboratory, Hitachi, Ltd.

<購買テーブル>			<個人テーブル>					
氏名	型式	支払	氏名	性別	年齢	趣味	勤務先	住所

<商品テーブル>				<会社テーブル>			
型式	商品名	分類	製造元	価格	会社名	所在地	業種

図2 DBの構成

件句が「日立」「男性」「買った」「テレビ」であり、検索対象部が「価格」であり、処置部が「平均」である。

(1) 検索条件句の記述(一属性内の関係の記述)

検索条件句「日立」は、属性「勤務先」の属性値に相当し、この検索条件句はこの属性内で閉じている。この場合には、DBの属性名に相当する語句と属性値に相当する語句を、助詞「が」を介した対表現で記述する。属性名に相当する語句が省略されている場合、各属性に対応する語句を定義したDB定義情報を参照して省略を補完する。「日立」は属性値に相当するので、属性名「勤務先」を補完し、「勤務先が日立」と記述する。「男性」「テレビ」についても同様である。

(2) 検索条件句の記述(複数属性間の関係の記述)

検索条件句「買った」は、二つの属性「人」「商品名」の間の関係を表す。このような関係はつねに明示されるとは限らないが、確認文を生成する際には、自然な文を生成するための語彙・構文情報として不可欠である。ここでは、予めDB管理者に定義させた情報を参照して属性間の関係を記述する(図4)。図4に示すように、DB管理者の定義できる情報は、①属性間の関係を表す語句(「氏名」が購入した「商品名」)、②文の自然さを保つために属性名を別表現で表した語句(「氏名」→「人」)、③各テーブルが全体として何に関するデータかを表す語句(「商品テーブル」は「商品」に関するデータ)等である。検索条件句「買った」については、その直前の単語「男性」が「人」に関する属性値であり、直後の「テレビ」が「商品」に関する属性値であるので、「人」と「商品」との関係に置き換え、図4の定義情報を参照して、「人が購入した商品」と記述できる。

(3) 語彙の選択

語彙の持つ多義性が原因で確認文が誤解釈されることを防ぐために、対象DBにとって一義である語句のみを用いる。例えば、「日立」は、図2のDBでは「日立市」とも「日立製作所」とも解釈できるので、システムが一義に解釈できる語を用いて、「勤務先が日立製作所」と記述する。

(4) 検索条件句の記述順序

(例1)の四つの検索条件句をどの順序で記述するかは、入力文における出現順序に従う。すなわち、「日立」「男性」「買った」「テレビ」の順で記述する。上述のように、各検索条件句はそれぞれ、「勤務先が日立製作所」「性別が男性」「人が購入した商品」「分類がテレビ」と記述できるので、これらを付属語等を用いて結び付けることにより検索条件部

として、「勤務先が日立製作所で、性別が男性である人が購入した商品で、分類がテレビである」を生成する。

(5) 検索対象の決定

検索問合せ文から検索対象を正確に決定することは難しい。ここでは、検索問合せ文が、検索条件、検索対象の順で記述されることが多いという言語的な傾向から、文末に最も近い属性名に相当する単語であり、かつ、その属性に対応する属性値を表す単語が対になって存在しない単語を検索対象とする。検索対象と並列表現された単語も検索対象に含める。

(例1)では、「価格」を検索対象として決定する。なお、この条件を満たす単語が存在しない場合、デフォルトとしてDB管理者が予め指定した属性を検索対象とする。

(6) 処置部の記述

処置部は、検索結果に対する演算・編集を表すもので、平均・最大値・件数・ソート等がある。ここでは、図3に示すように各処置内容に対応させて記述方法を一義に固定することで充分に対処できると考える。(例1)では、処置部「平均」について、「価格の平均値を求めよ」と記述できる。

(7) 確認文の組み立て

図3の記述に従い、検索条件部、検索対象部、処置部をこの順序で結び付けることによって確認文を生成する。すなわち、「勤務先が日立製作所で、性別が男性である人が購入した商品で、分類がテレビである商品の価格を検索し、価格の平均値を求めよ。」を得る。

<p><確認文> ::= <検索部> を検索せよ。 <検索部> を検索し、 <処置部>。</p> <p><検索部> ::= <検索条件部> <検索対象部></p> <p><検索条件部> ::= (<検索条件句> で、) <検索条件句> である</p> <p><検索条件句> ::= <検索条件句 1> <検索条件句 2> <検索条件句 3> <検索条件句 4> <検索条件句 5></p> <p><検索条件句 1> ::= <属性名> が <属性値> <関係指示語> (<論理演算語> <属性値> <関係指示語>)</p> <p><検索条件句 2> ::= <属性名> が <属性値> から <属性値> まで</p> <p><検索条件句 3> ::= <属性名> が <処置種類句 3></p> <p><検索条件句 4> ::= <検索条件部> <定義属性名*> <定義構文*> <検索条件部> <定義属性名*> <定義構文*> <定義属性名*></p> <p><検索条件句 5> ::= <検索部> <関係指示語> である <属性名></p> <p><関係指示語> ::= 以上 以下 未満 より以上 以外</p> <p><論理演算語> ::= かつ または</p> <p><検索対象部> ::= <定義属性名*> の (<属性名> と) <属性名></p> <p><処置部> ::= その <属性名> の <処置種類句 1> を求めよ。 その <属性名> を <処置種類句 2> と 並べ替えよ。</p> <p><処置種類句 1> ::= 平均値 最大値 最小値 件数</p> <p><処置種類句 2> ::= 昇順に 降順に</p> <p><処置種類句 3> ::= 最大 最小</p> <p>注：* はシステム管理者が定義した語句である</p>

図3 確認文の構成

属性の別表現		テーブル内容を表す語句	
属性名	別表現	テーブル	内容
氏名	人	個人テーブル	人
商品名	商品	商品テーブル	商品
会社名	会社	会社テーブル	会社

属性間関係定義		
属性名 1	関係	属性名 2
氏名	が購入した	商品名
会社名	に勤務する	氏名
製造元	が製造した	会社名

図4 DB管理者定義情報(太字部分)一覧

表1 評価結果

評価対象	正確に生成できた文	検索条件句の記述順序が入力文と異なる文	正確に生成できない文
43文 (100%)	35文 (81%)	7文 (16%)	1文 (3%)

4. 評価と考察

図2のDBを対象として本方式を評価した。ここでは、利用者から収集した例文を対象とした。なお、本方式は、日本語解析処理で生成された内部表現を入力とするので、内部表現を正確に生成できた43例文を評価対象とした。

実験結果を表1に示す。確認文を正確に生成できた例文は43文中35文(81%)であり、検索条件の記述順序が入力文と異なるが意味的には正確に生成できた7文を含めると42文(98%)となる。確認文を正しく生成できなかった1文(例2)の原因は、内部表現で認定した検索対象を表す単語の選択ミスである。

(例2) どの製造元の商品を写真好きの人は購入したのか?(確認文) 商品を購入した人で、趣味が写真である人の氏名を検索せよ。

(例2)では検索対象として「どこ(製造元)」ではなく「人(氏名)」と認識してしまっている。この例では検索対象が疑問代名詞であるので、これを考慮すれば検索対象を正確に認定することができる。検索対象が疑問代名詞でなく、かつ、文末にない場合については今後の課題である。

5. おわりに

問合せ日本語文の解釈結果を表す日本語文(確認文)生成方式を提案し、評価した。その結果、確認文を生成することが有効である見通しを得た。

今後は、DBの検索指示のみでなく、さまざまな操作指示に対しても確認できるような汎用的な確認文生成方式を開発していく予定である。

参考文献

- 1) 絹川：表階層モデルに基づく自然語インタフェース処理方式，情報処理学会論文誌，Vol.27, No.5 (May, 1986)
- 2) 木山，辻，絹川：自然語インタフェースにおける対話型解釈内容変更方式，情報処理学会第45回全国大会(1992)
- 3) 森木，絹川：自然語インタフェースのための音声ガイド，情報処理学会第44回全国大会(1992)