

## 6D-4

日本語質問文における格助詞'と'  
に関する一考察

篠原英彰, 絹川博之, 木山忠博

(日立製作所システム開発研究所)

## 1. はじめに

自然語インタフェースの代表例として、データベース検索を日本語文で指示するものがある。しかし、構文的曖昧性を含む日本語から、論理的な検索式を生成するには幾つかの問題がある。本報告では、格助詞'と'が持つ多義について、検索式生成の立場からの検討結果について述べる。

## 2. 格助詞'と'が持つ多義性

## 2.1 日本語から見た多義

質問文中に存在する助詞の'と'には、格助詞と接続助詞があり、さらに格助詞には、比較の基準、並列、引用、強調を示すものがある。ここで、質問文の意味解析において問題となるのは並列の'と'であり、これは名詞句の等位接続を表わしている。

例えば、'テレビとラジオを買った人'という文を考えてみる。'テレビとラジオ'という名詞句が動詞'買う'の目的格になることが解析できればよい、というシステムもあるが、データベース検索用質問文解析の場合は、'人'という検索対象の性質を限定するために'テレビ'と'ラジオ'があり、各々を条件の一つとして扱わなければならない。データベース検索の立場から見ると、この文は次のように解釈するのが自然である。

A: テレビを買った人の集合

B: ラジオを買った人の集合

とした場合、 $A \cap B$ を要求している。

ところが、同じ文体において、'横浜と品川に住んでいる人'という文に対しては、

A: 横浜に住む人の集合

B: 品川に住む人の集合

とした場合には、 $A \cup B$ を要求している。

と考えた方が自然である。

すなわち、構文的に同じ'と'であっても単語の持つ意味や文全体から伝わる常識から、多義が生じる。

## 2.2 データベースから見た多義

次に、入力文の係り受けの曖昧性ではなく、データベ

ースのデータ格納形式によって生じる多義について述べる。この多義は、意味的な多義ではなく、データベースの検索方法に関する多義である。例文として先程と同じ'テレビとラジオを買った人'を用いて説明する。

今、図1に示す2種類のデータベースを考える。図1(i)は氏名と購入品が1対1対応で格納されており、図1(ii)は購入品のカラムがマルチvalueになっている。これらのデータベースから'テレビとラジオの両方を買った人'を検索するには、図1(i)の場合は、'購入品=テレビ'を条件として得た氏名と、'購入品=ラジオ'を条件として得た氏名を比較し、両者に含まれる氏名を結果とする。図1(ii)の場合は図1(i)と同じ方法でも検索可能だが、'購入品=テレビ&ラジオ'を条件にした方が検索効率がよい。さらに、購入日に関するカラムがあり、同じ日に買った物をマルチvalue形式で格納するという条件でデータベースが作られている場合、'テレビとラジオを同じ日に買った人'を要求するときと、'とにかくテレビとラジオの両方を買った人'を要求するときでは検索式が異なる。

このように、データベースのデータ格納形式により、用いるべき検索式記述形式に多義が生じる。

## 3. 意味内容構造化表現方法の検討

我々は検索要求文の意味内容を木構造の有向グラフで表わす方法を試みている。初めに一般形について述べる。グラフは文ノード、テーブルノード、一般ノードの3種類のノードと、それらを結ぶアークから成る。文ノードは木構造のトップを示し、1入力文に対して1個存在する。テーブルノードはデータベースのテーブル名に対応する。一般ノードはデータベースのカラム名または実体名に対応する。また、論理演算子も一般ノードの一種類とする。そして、アークにはそのアークの始点に当るノードが持つ意味コード(データベース中の対応するカラム名)がアークラベルとして付与される。このようにして、テーブルノードの子ノードである一般ノードによって、そのテーブルに対する条件を表す。従って、入力文が示す検索条件式はテーブルノードと一般ノードから成

る部分木の組み合わせで表現できることになり、各部分木は対応するテーブルを持つ全情報の中の部分集合を表現している。

2. で述べた 'と' の多義をこの意味内容構造化表現方法で表すと図2のようになる。図2(i)は 'テレビを買った人' の集合と 'ラジオを買った人' の積集合を示しており、図2(ii)は和集合を示している。さらに、図2(iii)は '購入品' のカラムがマルチvalueのときの 'テレビとラジオ' が同一カラムにある集合を示している。すなわち、テーブルノード間の論理演算は集合の演算であり、一般ノード間の論理演算はカラム内の演算として解釈する。

4. 入力文の意味表現方法と解析手順

4.1 入力文の意味表現方法

3. では、2. で述べた多義を有向グラフで表わす方法について述べた。しかし、実際の処理に当っては、データベース形式に対する意味表現の独立性を保つという観点から、入力文の意味表現は全てカラム名と実体を対にしたものを項とする集合演算の形で表わすことにした。すなわち、図2(i)および図2(ii)の形式である。そのため、日本語の係り受けにより生じる多義は解析で対処し、データベースの構造に由来する多義(検索効率を意識したときに生じる検索式記述形式の多義)は検索言語生成処理で対処することを基本方針にした。

4.2 解析手順

2. で用いた文を例に解析手順を説明する。解析は全てノード操作により行なう。

- (1) 入力文の各文節をノードに割り当て、リスト状に接続する。
  - (2) 助詞等の解析により、表層的な係り受け関係を表わすように変換する(図3(i))。
  - (3) テーブル構成によって表現されている、意味モデルに基づく係り受け関係に変換する(図3(ii))。
  - (4) 全体を集合演算の形式に統一するため、'and' をテーブルノードの親の位置まで移動すると共に、部分木全体をコピーし、条件を振り分ける(図3(iii))。
- しかし、'and' を 'or' にするか否かは現在のところ、文体により一義に決定している。

5. まとめ

日本語質問文から検索条件式を抽出する際の、格助詞 'と' の扱い方に関する一方法を述べた。現在、我々はこの方式を自然語インタフェース HITNICE-Q に組み込み、評価中である。

6. 参考文献

- 1) 絹川博之: 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, pp499-509, (1986.5)
- 2) 篠原英彰 ほか: 自然語インタフェース(HITNICE-Q)における多義解消方式, 情報処理学会第36回全国大会, 7U-1, pp1315-1316, (1988.3)

氏名	購入品
A	テレビ
A	ラジオ
B	ラジオ
B	テレビ

氏名	購入品
A	テレビ, ラジオ
B	ラジオ
C	テレビ

図1 データ格納例(販売テーブル)

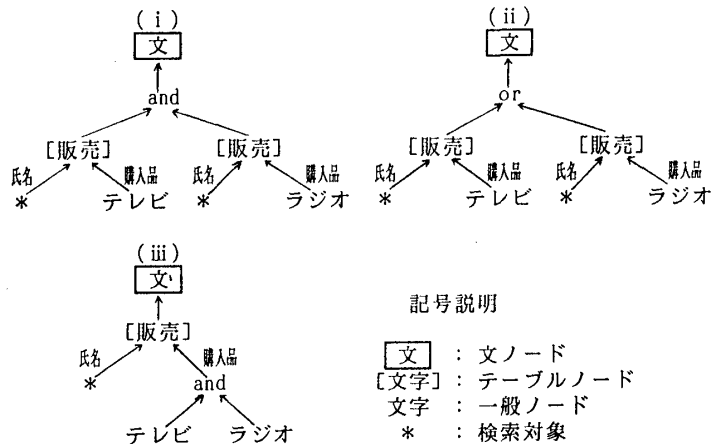


図2 意味内容構造化表現方法

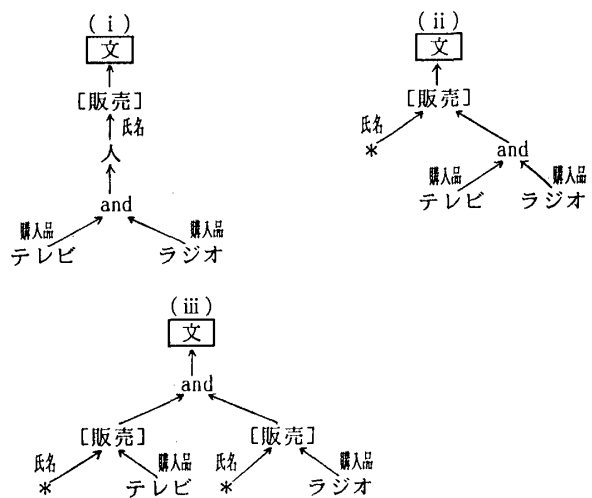


図3 解析例