

# 係り受け関係を含む問合せが可能な日本語質疑応答システムの研究

五十嵐 大介 沖 龍太

## 1. 背景と目的

モバイルコンピューティングの急激な普及に伴って、音声による問合せシステムの必要性が増している。現在、音声から文字列への変換が可能になっているが問合せ文の意味解析については十分解明されたとはいえない。特に、Yahoo、gooなどの検索エンジンは“キーワード(単語)”での検索が可能であるが係り受けを含む自然言語”での検索はできない。このような検索が可能になれば、音声を自然言語に変換した問合せで、目的の情報を探し出すことも可能になる。

本研究では、係り受けを含む自然言語による問合せ文から意味構造を抽出し、抽出した意味構造からデータベースにアクセスする形式的な問合せ言語、構造化クエリ言語(Structured Query Language: SQL)形式に変換するシステムを試作し、意味構造抽出と形式記述への変換の妥当性を評価することを目的とする。なお、問合せ文は店舗検索時によく使用される願望文(例えば、～したい、～食べたい、～行きたい)などの日本語に限定する。

## 2. 試作システム概要

### 2.1. 主な機能

#### (1) 自然言語解析

問合せ文を形態素解析 構文解析をおこない、語彙機能文法の F 構造を生成する。

#### (2) SQL 生成

F 構造から SQL を生成する。

#### (3) データベース結果セット表示

生成された SQL を使用しデータベースの結果セットを取得、表示する。

### 2.2. 試作システムの特徴

試作システムの特徴は、以下の点である。

- (1) 係り受け文を文の末尾から解析する方法を採用している。
- (2) 語彙機能文法に従って意味構造を抽出している。
- (3) 概念、述語辞書およびコーパスを併用している。

### 2.3. 試作システムの入出力画面

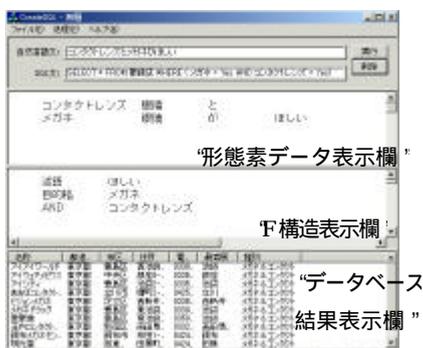


図 1 入出力画面

## 3. 評価とまとめ

### (1) 問合せ文中の格助詞の分布

機械工学科、電気電子工学科、システム工学科、情報工学科の1~4年生からアンケートを取り、77個の文から問合せ文

を集計した。その結果、一つの文中に格助詞が二つ使用されている文の出現頻度は42.9%で最も多く、格助詞は“で、が、の、を、に”が多用されていた。出現割合数239個の内、各格助詞の出現頻度は以下のようであった。

表 1 各格助詞の出現頻度

格助詞	で	が	の	を	に
出現頻度	24.7%	22.6%	18%	14.2%	11.7%

### (2) NTT の i タウンページ文章 (自然語) 検索と比較

(1)の出現頻度にしたがってテスト文をアンケートから選択し、意味的に適合した検索が行えるかを NTT i タウンページと比較する実験を行った。その結果は表 2 で示すとおりである。

表 2 NTT と試作システムの比較表

テスト文	出現頻度	NTT	試作
池袋でメガネを買いたい	24.2%		
新宿で演劇を観たい	15.2%		
渋谷にある薬局が知りたい	15.2%		
群馬の温泉に行ってみたい	15.2%		
携帯電話の電池がほしい	12.1%	x	
車の免許を取りたい	3.0%		
魚が釣れるルアーがほしい	3.0%		
コンタクトレンズとメガネがほしい	3.0%	x	
パソコンを新宿で買いたい	3.0%		
映画を見に行きたい	3.0%		
新宿のデパートへ行きたい	3.0%		

出現頻度から見た NTT と試作システムの意味的な適中率は、以下のとおりである。

$$\text{NTT 適中率} = (0.85 / 1) \times 100 = 85\%$$

$$\text{試作適中率} = 100\%$$

上記結果より試作システムは NTT と比較して係り受けを含む自然問合せ文の適中率を 15%向上させることができた。この結果から、係り受け文を末尾から解析し意味構造を抽出する方法と意味構造から SQL 形式への変換の妥当性が確認できた。

## 4. 課題

今後、コーパスから自動で構文情報を抽出し、構文解析で使用される格助詞辞書を自動で生成 拡充されることが望ましい。

## 参考文献

- [1] 田中穂積 (著) 自然言語解析の基礎 産業図書株式会社 (平成 5 年)
- [2] ピーター・セルズ (著) 郡司 隆男・田窪 行則 石川 彰 (訳) 現代文法理論 GB理論、GPSG、LFG 入門 産業図書株式会社 (1988 年)

