

自然言語仕様書に基づくソフトウェア開発支援システム

4Q-3

— 仕様文解析部における複合名詞の解析 —

池田 朋男 田中 克己

(株)東芝 関西研究所

1 はじめに

我々は、自然言語仕様書に基づくソフトウェア開発支援システムの研究開発を行なっている [1]。

現在、仕様書に含まれる文中に現れる、未知語を含む複合名詞の構造を解析し、ユーザと対話を行うことにより、単語と対象知識の対応を決定し、解釈の曖昧性を解消する方式の研究を行なっており、今回は、複合名詞の構造解析手法の概要について報告する。

2 目的

ソフトウェアの仕様書を自然言語を用いて記述する場合、自然言語文の持つ曖昧性や省略などにより、明確な仕様書を記述する事が困難であるという問題が存在する。

我々が研究開発を行なっている自然言語仕様書に基づくソフトウェア開発支援システム [1] は、自然言語仕様書を解析し、曖昧な部分や不明確な部分をユーザに指摘することにより、設計者による明確な仕様書の作成を支援する。家電製品の分野を対象とするシステムを試作し、実際の利用に基づく評価実験を行なった結果、解析が失敗した場合にシステムが出力する情報が不十分であるという問題点が明らかになった [2]。

特に、文中に辞書・知識への未登録語が含まれている場合、格助詞などが省略されている場合には、1単語で表現されるはずの部分が複合名詞として表現されるため、対象知識との対応が一意に決定できず、文の解釈に曖昧性が生じる。このため、ユーザに曖昧性解消のための十分な情報を提供するためには、複合名詞の構造を解析することが必要となる。

3 複合名詞解析

3.1 基本方針

一般に自然言語文では、文の格パターンと、それぞれの格に出現する単語の意味の間に関係がある。特に本システムで扱う仕様文は、書式通りに記述されている場合、それぞれの格の種類と対象知識中の知識の種類が一对一の対応関係を持つ。例えば例1のような表現では、「の」格と対象知識中の機器、「が」格と属性、「に」格と状態がそれぞれ対応している。

例1) 「機器の属性が状態になれば～」

しかし、書式通りに仕様文が記述されていない場合には、上記のような対応関係が得られず、正しく解析を行

名詞	名詞 (冷凍室、温度...)
接頭辞	接頭辞 (最高、最低...)
接尾辞	接尾辞 (以上、未満...)
数値	数値を表す表現 (0秒、1度...)
アルファベット	アルファベット (A、B...)
数字	数字 (0、1、2...)
未知語	辞書に登録されていない単語

図1: 複合名詞解析の入力要素

機器	対象知識で機器として登録 (冷凍室...)
属性	対象知識で属性として登録 (温度...)
状態	対象知識で状態として登録 (動作中...)
数値	状態を表す数値表現 (0秒、1度...)
接頭辞	状態を修飾する接頭辞 (最高、最低...)
接尾辞	状態を修飾する接尾辞 (以上、未満...)
番号	同種の知識を区別する番号 (冷凍室1...)

図2: 複合名詞解析の出力要素

なう事ができない。特に格助詞の省略、対象知識中の機器や属性に対応する単語の省略は、しばしば起こり得る。

例2-1) 「冷蔵庫温度が設定温度以上になれば～」

例えば上記の例では、「冷蔵庫温度」と機器が対応するという解析結果が得られるが、

例2-2) 「冷蔵庫(の)温度が設定温度以上になれば～」

の様に助詞「の」が省略されている可能性もあり、この場合は「冷蔵庫」が機器、「温度」が属性に対応するという解析結果を得る事が必要になる。

このように、格助詞が省略された場合、また格に出現する単語が未知語である場合は、格に出現する表現が複合名詞として表される。この場合、それぞれの格と対象知識との対応関係は一对一であるとは限らず、文中の格の出現パターンにより対応関係が変化する。例えば、例1の表現では一対一対応になっているのに対し、例3の「の」格が存在しない表現では、「が」格に対し3種類の知識が対応する可能性がある。

例3-1) 「機器が状態になれば～」

例3-2) 「機器+属性が状態になれば～」

例3-3) 「属性が状態になれば～」

今回提案する方式では、複合名詞の入力パターンを格の出現パターンによって分類し、得られた入力パターンから解析結果の出力パターンへのマッピング処理を行な

A Software Development System Based on Natural Language Specifications

IKEDA Tomoo and TANAKA Katsumi

TOSHIBA Kansai Research Laboratory

E-mail: {tommy,tanaka}@krl.toshiba.co.jp

非数 ::= 名詞 数値 アルファベット 接頭辞 接尾辞 未知語
単語 ::= 非数 数字
入力 ::= 非数 * 非数 非数 * 数字 単語 * 数字 単語 * 非数 単語 * 数字 単語 * 数字

図 3: 入力パターンルールの一例

う事により、複合名詞と対象知識の対応関係を得る。

3.2 入出力

入力となる複合名詞は、図 1の入力要素の接続により構成される。入力要素は辞書に登録されている品詞によって分類される。

複合名詞の解析結果は、それぞれの入力要素、あるいは入力要素の組合せと図 2の出力要素の対応関係として得られる。出力要素は、対象知識中のノードの種類、数値など対象知識中には陽に登録されないもの、および接辞、番号などの修飾表現からなる。

3.3 解析方式

以下に実際の解析処理について説明する。

1. 複合名詞抽出

入力文の構文解析・意味解析を行ない、その解析結果をもとに、それぞれの格に出現する複合名詞を抽出する。それぞれの複合名詞について以下の解析処理を順に行なう。

2. 入力パターンの選択

入力文の格の出現パターンによって、入力パターンルールを選択する。入力パターンルールは、複合名詞を入力要素の並びのパターンによって分類したルールで、複合名詞は入力パターンルール中の一つのパターンに対応するように設計されている。図 3に、「の」格がない場合の「が」格の入力パターンルールを示す。選択された入力パターンルールと入力の複合名詞とのマッチングを行ない、入力パターンルール中の対応するパターンを得る。

3. 出力パターンの獲得

マッピングルールに基づき、出力パターンルール中から入力パターンと対応する出力パターンを得る。出力パターンは、出力を出力要素の並びのパターンによって分類したルールである。また、マッピングルールは、それぞれの入力パターンに対応する出力パターンの候補を示すルールである。一つの入力パターンに対し、一つ、あるいは複数の出力パターンが対応する。

4. 実際の入力と出力パターンとの対応

得られたそれぞれの出力パターンについて、実際の入力単語列と出力パターン中の要素との対応関係を得る。

3.4 実際の例

例 2 の文の「冷蔵庫温度が～」の部分について、上記の処理を実際に行なった例を以下に示す。ただし、「冷蔵庫」と「温度」は辞書に登録されているものとする。

<ul style="list-style-type: none"> • <u>機器+属性</u> が出力される場合 出力 ::= 機器 属性 機器 番号 属性 機器 属性 番号 機器 番号 属性 番号 • <u>機器</u> が出力される場合 出力 ::= 機器 機器 番号 • <u>属性</u> が出力される場合 出力 ::= 属性 属性 番号
--

図 4: 出力パターンルールの一例

非数 * 非数 → 機器 属性 機器 属性 非数 * 数字 → 機器 属性 番号 機器 番号 属性 番号 単語 * 数字 単語 * 非数 → 機器 番号 属性 単語 * 数字 単語 * 数字 → 機器 番号 属性 番号
--

図 5: マッピングルールの一例

まず「が」格の複合名詞として、「冷蔵庫(名詞)/温度(名詞)」が抽出される。この表現には「の」格が無いので、入力パターンルールとして図 3が選択され、その中からパターン「非数 * 非数」が得られる。図 4の出力パターンルール、および図 5のマッピングルールより、入力パターンに対応する出力パターンは「機器 属性」「機器」「属性」のいずれかである事が得られる。実際の入力である「冷蔵庫 / 温度」と上記の出力パターンの対応関係より、解析結果として以下の 3 通りの結果が得られる。

- 冷蔵庫(機器)、温度(属性)
- 冷蔵庫温度(機器)
- 冷蔵庫温度(属性) [機器が省略されている場合]

4 おわりに

自然言語仕様書に含まれる文中に現れる未知語を含む複合名詞の構造を解析する手法について述べた。現在、複合名詞解析の結果をもとにユーザと対話を行ない、解釈結果を一意に決定する方式について研究を進めており、ソフトウェア開発支援システムに実装し、実験・評価を行なう。

参考文献

- [1] 田中 克己, 他. 自然言語仕様書に基づくソフトウェア開発支援システム - システム構成とプログラム合成手法 -. 情報処理学会 第 48 回全国大会論文集 (3), 1994.
- [2] 池田 朋男, 他. 自然言語による仕様記述支援システム - 仕様書作成過程の分析 -. 情報処理学会 第 47 回全国大会論文集 (3), pp. 119-120, 1993.