

## ヘルプデスク支援システムにおける用語獲得

1N-1

伊藤 山彦 高山 泰博 鈴木 克志

三菱電機（株）情報技術総合研究所

### 1. はじめに

ヘルプデスク業務を支援する目的で、過去に蓄積した事例を検索し、オペレータの問題解決を支援するシステムを開発中である。特徴として、対象分野に依存した概念関係の知識(オントロジ)を利用し、事例の細かな記述の違いを弁別可能とした言語事例分類/推論技術を用いる。ヘルプデスク業務では、新製品や新機能など日々新しい語が生まれるため、人手によるオントロジの構築・維持に多大な労力を要する。このためオントロジを自動構築する技術が必要である。

本ヘルプデスク支援システムが利用するオントロジは、用語辞書、IS-A 辞書、HAS-A 辞書、格辞書、言い換え辞書の5つである。本稿ではオントロジ獲得における1ステップとして、ヘルプデスク業務においてオペレータが記録した障害対応記録から、辞書に登録されていない新しい語彙(以下、単に「用語」と呼ぶ)を獲得する手法について述べる。

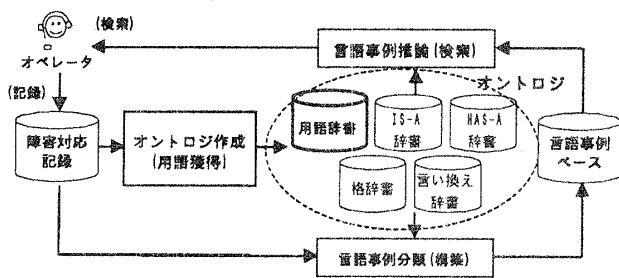


図1 用語獲得の位置づけ

### 2. 用語獲得の課題

従来、用語の抽出は、(1)既知語を組み合わせて複合語を生成する処理<sup>[1]</sup>や、(2)辞書に登録されていない語が現れ解析が失敗した際に、字種や形態素長を手がかりに用語の範囲を推定する処理<sup>[2]</sup>によって行っていた。

Acquisition of Vocabulary for the Helpdesk Support System

Takahiro ITO, Yasuhiro TAKAYAMA, Katsushi SUZUKI  
Mitsubishi Electric Corporation.

5-1-1 Ofuna, Kamakura, Kanagawa 247-8501, JAPAN

(1)の処理で獲得される用語は、解析辞書に予め登録されている既知語を組み合わせた複合語に限られるため、複合語以外の新規用語が頻出するヘルプデスク業務において十分な用語獲得ができない。また、(2)の処理で抽出できる用語は、分野に依存しない汎用的な枠組みで、未知語周辺から一般的に推定できるものに限られる。

本稿では上記課題を踏まえ、分野に特定の表現パターンを利用し、既知語の組み合わせでない新規の未知語に関しても獲得可能な用語獲得方法について述べる。

### 3. 用語獲得処理

本用語獲得処理では、獲得すべき用語のパターンを用語パターン辞書に定義し、入力テキストの形態素解析結果と照合する。

#### 3.1 処理の流れ

図2に用語獲得処理の流れを示す。

- (1)入力テキストを形態素解析する。形態素解析では基本語辞書と既存の用語辞書を参照する。
- (2)形態素解析の出力結果を用語パターン辞書と照合し、用語の候補を獲得する。
- (3)獲得した候補を表示し登録すべき用語を手で選択する。人による判定を支援するために、KWIC(KeyWord In Context)表示によって用語が現れる文脈の確認を可能とする。
- (4)獲得した用語を保存する。獲得した用語は解析用の用語辞書に追加する。

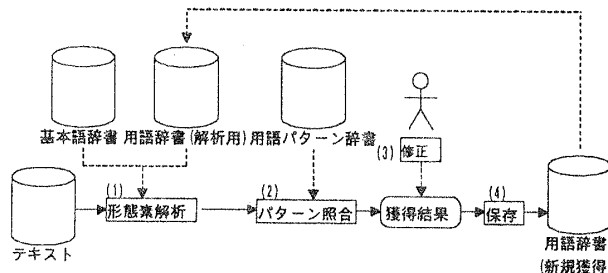


図2 用語獲得処理の流れ

#### 3.2 用語パターンの記述

用語パターン辞書に記述するルールの記述仕様

を図3に示す。用語パターンは、パターン定義部と用語決定部からなる。パターン定義部では、テキストの形態素解析結果と照合する形態素パターンを指定する。用語決定部では、パターン定義部における照合が成功したとき、獲得する用語のパターン定義部における範囲、品詞、及び概念を指定する。

用語パターン	::=	パターン定義部 → 用語決定部
パターン定義部	::=	形態素記述(,形態素記述)*
形態素パターン	::=	表記パターン 品詞パターン  字種パターン 文字列長パターン  概念パターン
表記パターン	::=	s(string)
品詞パターン	::=	p(string)
字種パターン	::=	t(string)
文字列長パターン	::=	l(number)
概念パターン	::=	c(string)
用語決定部	::=	用語範囲決定部,用語品詞決定部, 用語概念決定部
用語範囲決定部	::=	range(number, number)
用語品詞決定部	::=	pos(string)
用語概念決定部	::=	concept(stiring)
形態素記述は形態素パターンを以下の演算で組み合わせたもの		
AND:		形態素パターン1 && 形態素パターン2
OR:		形態素パターン1    形態素パターン2
NOT:		!形態素パターン
0回以上の繰り返し: 形態素パターン*		
1回以上の繰り返し: 形態素パターン+		

図3 用語パターン辞書の記述仕様

次に示す R1 は用語パターンの例である。

R1:(p(名詞)||p(固名)),(s(事)||s(支))  
→ range(1,2),pos(固名),concept(組織名)

R1は『品詞が「名詞」または「固有名詞」である形態素の後に表記が「事」または「支」である形態素が続いたとき、全体の品詞は固有名詞であり「組織名」という概念を持つ』という意味である。rangeの引数である1と2は、それぞれ獲得する用語のパターン定義部の中における形態素記述の先頭の位置と最後の位置を表わす。

### 3.3 IS-A 階層の利用

既存の IS-A 辞書がある場合は、ルールに直接記述されていない関係を推論して用語の獲得を行う。例えば、図4のような IS-A 階層がある場合、次のパターン R2 による照合について述べる。

R2:c(試験),s(で),t(カタ),p(を),s(発見)  
→ range(3,3),pos(名詞),concept(障害)

R2では、形態素記述であるc(試験)は「試験」の下位概念を持つ全ての語と照合する。その結果、

折り返し試験,で,ビットエラー,を,発見

という形態素解析結果と、R2のパターン定義部の照合は成功し、「ビットエラー」は「障害」という概念を持つと判定できる。

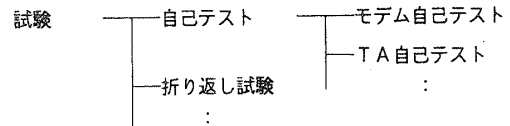


図4 IS-A 階層

## 4. 実験

ネットワークに関する障害対応記録(2,109 KByte)に対して、用語獲得実験を試みた。以下に、獲得のためのルールと獲得した語彙の例を記す。

R1:(p(名詞)||p(固名)),(s(事)||s(支))  
→ range(1,2),pos(固名),concept(組織名)

「長崎事」「東北支」など、8語獲得

R3:(p(名詞)||p(未知)),(s(断)||s(劣化))  
→ range(1,2),pos(名詞),concept(障害)

「CD断」「電流劣化」など、56語獲得

R4:(p(名詞)||p(未知)),(s(線)||s(網))  
→ range(1,2),pos(名詞),concept(回線)

「HD線」「公衆網」など、17語獲得

R1の「事」は「事業所」の略であり、「支」は「支店」の略である。略語は未知語になりやすく規則性があるので、本手法による獲得が有効である。また「断」が障害を表わすことは、ネットワークの分野に特有である。このように、分野に特定の表現パターンから、その分野における特定の意味を持つ用語を獲得することが可能である。

## 5. おわりに

本稿では、ヘルプデスク支援システムへの適用を目的とした用語獲得について述べた。今後、IS-A辞書や格辞書など、他のオントロジ構築方法を確立すると共に、用語パターンの学習や自動獲得について検討する予定である。

## 参考文献

- [1]小川ほか:複合語キーワードの自動抽出法,情報処理学会NL研,Vol.93, No.79, pp.103-110 (1993).  
[2]脇田ほか:日本語校正支援システムFleCS-未知語登録作業の省力化,情報処理学会第52回全国大会,Vol.3,pp.27-28 (1996).