

電子化辞書管理のための自然言語インターフェース — システム依存構造の生成 —

9B-5

落合 尚良 森 義和 *奥井 伸司

(株)日本電子化辞書研究所 *NEC技術情報システム開発

1. はじめに

我々が開発中の電子化辞書は、新規語の登録、新しい解釈による情報の追加等、日々拡張更新されている。日々作業を行なう辞書開発の担当者からは、簡単に利用出来る管理システムが望まれている。そこで作業者が使い慣れている自然言語をインターフェースに用いた管理システムを試作した。自然言語インターフェースは対象を限定することで対象知識で文書中の省略を補完できることが文献[1]で報告されている。

本稿では、管理に対する要求を対象システム依存の意味構造に生成する機構について報告する。

本システム全体の概要については文献[2]を参照。

2. システム依存意味構造の生成

辞書開発者の管理に対する要求は多義にわたるが、文献[3]よりその文の構造は属性と属性値の条件を組み合わせた要求が多いこと、「形態素解析する」の様なプリミティブな要求があることが解かった。複雑な管理要求を一文で入力することはユーザにとって曖昧な表現になり大変困難なことと考え、そこで作業者の管理要求を対象システムが持つ機能ごとに書き下すかたちで記述し、書き下された文をまとめ、プログラム化し処理できる機構を設計した。しかし、たとえ書き下した文でも、対応システムに依存した概念へ置換するときの曖昧性、文をまとめてプログラムとして扱う場合の、文と文との関係付けに利用される指示代名詞による指示先の選択など解決する必要がある。そこで、3で曖昧性の解消、4で管理要求のプログラム化実現について説明する。

3. 曖昧性解消と省略補完

1) 自然言語の表現で、その概念が異なるものであっても対象システムに対する要求では同じ概念になるものもある。例えば、「検索する」、「教える」もデータベースに対しての要求は同じである。概念の対応の違い、概念のグループ構成の違いがシステム依存の部分である。

Natural Language Interface for Electronic Dictionary System Manager

- Generation of System Dependency Structure -

Takayoshi OCHIAI, Yoshikazu MORI,*Shinji OKUI

Japan Electronic Dictionary Research Institute,Ltd

*NEC Scientific Information System Development,Ltd

*本研究はEDR在任中に行なったものである。

● 文献[3]の分析された例文の語彙からシステムに対しての要求を分類し概念のグループ化を行ない、それを知識として蓄え置換処理を行なう。

2) 自然言語での入力は、省略を含む表現が多くシステムに対しての要求にも起こりえる。

例) 出力媒体を指定する場合

“出力する”の場合は、システムに接続されている媒体すべてが対象になる。

“表示”の場合は、媒体は画面と推測できる。

“書く”の場合なら、ファイルと推測できる。

このような判断が一般的に出来る。

● 省略のありえる語彙と、その省略を許す語彙の関係をあらかじめ知識として蓄え、省略された語彙を補完する。

3) データベースの情報に関して その規模(数の大小)を数値で入力することは難しく、曖昧な程度表現を使うことがある。

接続属性をたくさん持っている見出し語は。

● データベースの属性に対して程度を表わす語彙の数的な情報を知識として蓄え置き換える。

4. 管理要求のプログラム化

管理要求が書き下した複数文で入力され、1文ごとに解析された概念構造を保持しながら指示代名詞で構造を連結し最終的な構造は条件が組み合わさってプログラム化されたシステム依存構造になる。処理のなかで指示詞がどの構造を指しているか曖昧な場合は、ユーザに対して問い合わせを行ない決定する。結合に関しては、指示された先の構造と指示代名詞を含む文の構造とが整合性が取れているかを確認した上で構造を結合する。整合性のチェックは属性と属性値等の条件に矛盾がないかの検証である。

4.1 処理例1

次のような管理要求があった場合

- 1 読みが「あい」の見出し語と品詞を教えて。
- 2 その左連接属性は「JLV1」です。
- 3 それを見出し語順に出して。
- 4 それの読みが「あう」も教えて。

1文目の解釈は、他の構造を参照する必要はなく、概念構造をシステム依存構造に生成する。

2文目の解釈は、「その」の指す対象は1で、曖

昧性がなく条件をandし、1のactionを継承しシステム依存構造を生成する。

3文目の解釈は、actionに対する条件で曖昧性、整合性には問題なし、条件を追加しシステム依存構造を生成する。

4文目の解釈は”それ”の指す先が2通りの解釈が可能であり、この判断はユーザに対して以下の文を表示し問い合わせを行なう。

1文目を指す	読みが「あい」であり 左接続属性が「JLV1」である。 または 読みが「あう」である。 を検索し見出し語を出力します。
2文目を指す	読みが「あい」であり 左接続属性が「JLV1」である。 または 左接続属性が「JLV1」であり 読みが「あう」である。 を検索し見出し語を出力します。

4.2 処理例 2

以下のような管理要求があった場合

- 1 概念見出しが2個以上の見出し語を検索して。
- 2 それをAファイルに出力して。
- 3 それを多い順にならべて。
- 4 それとBファイルとでdiffを取って。

1文目の解釈は、他の構造を参照する必要はなく、構造をシステム依存構造に生成する。

2文目の解釈は、”それ”の指す対象は一意で、曖昧性がなく、システム依存構造に生成する。

3文目の解釈は、解析結果として”多い順”のactionを行なう場合、”それ”の指す対象に”順”を決定できる情報が必須となる。指す対象は1、2が考えられるが、”多い”が必要としている数の情報に対する条件が記述されているものが1しかなく曖昧性は解消される。1の構造を参照し生成処理を行なうが、参照する構造内に”順”を決定するための情報が無い。しかし、要求を満たす必須の条件であるため、対象になる構造の数に関しての条件が記述されている属性に対し、その数の情報を保存できるように条件を補完する。

4文目の解釈は、”それ”の指す対象に曖昧性があり、1、3の二つが推測できる。この場合の曖昧性の解消法は、4.1処理例1で述べた通りである。

ここでは、別の問題について説明する。ここでの要求であるdiffを実行する環境はUNIXを想定した。このため生成されるシステム依存構造は、環境がデータベースに対しての処理とUNIXのファイルに対する処理の二つの環境に対してになる。このため生成構造には、対象となる環境を識別できる機構を設定

した。対象が二つになることで、その対象ごとの処理を関連づけて処理できるように、環境ごとにシステム依存構造をグループ化し、そのグループを連結し処理できるように設計した。

4.3 不整合の例

- 1 左接続属性が「JLV1」の見出し語を教える。
- 2 それを「JRV1」も教えてください。

この場合、”それ”の指す先が1であることは曖昧性がなく、システムは構造を参照しシステム依存構造の生成処理をはじめ。

- 1の構造の条件に2の条件を加える処理をする。

- 1の条件 左接続属性が「JLV1」

しかし、左接続属性の属性値に「JRV1」の存在を確認できない（左接続属性の属性値は固定であるため判断可能である）

この時の判断は、

右接続属性の「JRV1」を検索します。

* 「JRV1」から”右接続属性”を類推できる

* ”検索”は、要求のデフォルトである

* このチェックは属性値の個数が有限なものに行なう

5. 管理要求のマクロ化

プログラム化された管理要求をマクロとして定義することにより辞書管理で頻繁に行なわれる情報の整合性チェック、大量の更新作業等の作業を効率的に行なうことが出来る。またマクロ化された管理要求を、書き下す文内で使用することで作業が高度化、複雑化しても入力作業が増えない利点もある。

6. おわりに

大規模な電子化辞書を開発する作業者の利用を考え自然で効率的なシステムを目指しシステムを試作した。開発された機能から順に辞書開発者に利用してもらい、評価改良を行なっている。機能全体の開発を済ませ統合的な管理システムとしての有用性を検証して行く。

謝辞

本研究の機会を与えて下さった横井所長に深く感謝いたします。

参考文献

- [1]宮部「日本語インターフェース文脈文法 ー解析手法ー」45回情処全大 2F-1
- [2]落合「辞書管理における自然言語インターフェースシステム概要ー」46回情処全大
- [3]松山「辞書管理における自然言語インターフェースーコーパスの分類ー」46回情処全大
- [4]EDR電子化辞書：EDR,TR-016 1989