

7P-5

計算機言語文法書からの知識獲得

今井 達也 , 永田 守男

(慶応義塾大学理工学部)

1. 始めに

コンピュータで問題を解決しようとする場合に、いままでは、必要となるべき知識を人間があらかじめ予測してシステムが利用できる形に変換して、メモリ上に展開していた。

ところがコンピュータに、人類が全く知らない新しい事実を発見させたり、探索的な作業を行わせようとする、必要な知識をいかにしてシステム内に取り込むかが問題となる。その知識は、膨大かつ予測が困難なものである事が多い。そこで、人間の介入なしに自動的に情報を収集しておいて、これを知識として利用できる形に変換、格納する機能が必要である。一方、有史以来、人間の持つ技術についての知識のほとんどは文書の形で記録されているので、そのような情報源としては文献が有効である。そこで本研究は、文献から知識を獲得する技術研究の一端として、計算機言語の文法書から知識を獲得する事を考える。そして、計算機言語の文法書から、言語のシンタックスに関する知識を獲得し、その知識に沿って、入力される文字列のシンタックスチェックを行うシステムについての研究を行う。

2. 文法書の特徴

知識源として見た計算機言語の文法書は次のような特徴を持っている。

- 1 抽象形容詞がなく、概念は全て名詞の結合によって表現される。
- 2 名詞句により示される概念の種類が限られる。
- 3 知識の内容が手続き的である。
- 4 知識の矛盾がない。
- 5 知識のほとんどが本の中で閉じている。
- 6 文書が定型的である。

3. 関数型知識表現

2.で述べたような文法書の特徴から、本研究のシステムでは知識の内部表現形式として関数型の知識表現を採用している。すなわち、一つの名詞句の意味を一つの合成関数の関数値で表現し、単語の意味知識として、その単語を説明している名詞句の内部表現を知識ベースに登録している。

一つの名詞句は、一つ以上の名詞または名詞句を用言で結合したものとみなす事ができるので、ある名詞句の内部表現は、それを構成する名詞または名詞句の内部表現をパラメータとして、用言を表現する関数に与えたもので生成できる。すなわち、ある名詞句(一つ概念を表現している)は、それを構成する各名詞句を表す関数値(概念)をなんらかの関数で結びつけることによりその意味を表現できる。

例えば、

「英字の後に英字または数字を付けた物」

という名詞句は、

「英字」という名詞句と

「英字または数字」という名詞句を、

「Xの後にYを付けた物」という用言で結合したものであり、その内部表現は、

「英字」を表す (match0 eiji) と

「英字または数字」を表す

(table (match0 eiji) (match0 suuji))

をパラメータとして

「Xの後にYを付けた物」を表す

(order t X Y)

という関数に適用したもの、すなわち

(order t (match0 eiji)

(table (match0 eiji)

(match0 suuji))

)

となる。

4. 知識表現の生成

文章から3.で述べた様な内部表現を生成するのに、本研究のシステムでは意味属性別の用言のパターンによる構文解析と、その用言のパターンと内部表現のパターンを一対一対応で変換することを同時に行っている。

システムは Fig.1 の様に用言のパターン(構文パターン)とそれに対応する内部表現(関数)のパターンをその意味属性ごとに分類して持っている。システムは文章が入力されると、それを文(*BUN)の構文パターンのどれかにマッチさせる。マッチングは以下の様に行われる。すなわち、構文パターン中の終端記号を左から見てゆき、入力文字列中でそれを探していく。構文パターン中の全ての終端記号がその出現順に入力文字列中で発見できれば、構文パターン中の終端記号以外の部分(すなわち、非終端記号の部分)にマッチすべき入力文字列の部分列が切り出せる。つぎにこの部分文字列を、対応する非終端記号により示される意味属性を持つ構文パターンにマッチしようとする。ただし、非終端記号 *cn は名詞を示すので、これにはどんな文字列もマッチする(辞書にない新しい名詞を獲得できるようにするために、*cn にはどんな文字列でもマッチし、その文字列が辞書にない場合、新しい名詞として辞書に登録する。)入力文字列の全ての部分が終端記号または非終端記号 *cn に対応づけられた時にマッチングが成功する。ある構文パターンでのマッチングが失敗すると、システムはバックトラックし、他の構文パターンのマッチングを試みる。全ての構文パターンでのマッチングが失敗したときに、その入力文字列と非終端記号とのマッチングに失敗し、その文字列を切出してきた親のマッチングが失敗する。

この場合、ある構文パターンでのマッチングが成功すると、システムはそのパターンにマッチした文

字列の内部表現を生成する。マッチングが成功した時点で、その構文パターン中の非終端記号に対応する部分文字列の内部表現が決定されているので、その表現で構文パターンと対応する内部表現パターン中の、適当なスロット（構文パターン中の対応する非終端記号と同じ記号で示される）を置換することによって内部表現を生成することができる。

Fig.1 のようなパターンがあったとき、

「英字の後に英字または数字を書いたもの」

という名詞句は、Fig.1 の規則4

(*sn-1 の後に *sn-2 を書いたもの)

とマッチし、「英字」と「英字または数字」という二つの名詞句に分解される。各名詞句はそれぞれ

*sn (構造を表す名詞句) のパターンにマッチングが成功し、その結果として内部表現

(match0 eiji)

.....*sn-1 「英字」

(table (match0 eiji) (match0 suuji))

.....*sn-2 「英字または数字」

が決定される。

そこで、最初の文字列「英字の後に.....」とマッチした規則4 の内部表現パターン

(order t *sn-1 *sn-2)

のスロット *sn-1 *sn-2 を各内部表現で置換する。結果、

(order t (match0 eiji)

(table (match0 eiji)

(match0 suuji)))

)

が最初の文字列「英字の後に.....」の内部表現として得られる。

5. 考察

本研究で採用した二つの手法、についてその特徴を考察してみた。

「関数型知識表現」について、

- 2.1 で述べたように、文法書の文章は、一つの概念と一つの名詞句（または名詞）が一对一対応しているので、これを一つの値として考える事ができる。また、一般的に名詞句は入れ子構造（すなわち複数の名詞句が集まってさらにひとつの名詞句を構築する）になっているといえる。この事は、関数が一つの値を表現し、合成関数という形で入れ子構造を形成できるという性質にマッチしていて、一つの名詞句と一つの関数値を一对一対応させることにより、文章から内部表現への変換が極めて素直にできる。一つの概念と一つの関数値を対応させようと

すると、ある種類の概念を表すための具体的な関数値をどう決めればよいかが問題となるが、2.2 の特徴により文法書の場合、このような問題は比較的少なくてすむ。

- 2) 動的な内容（手続き、規則等）の知識を表現し易い。文法書をシンタックスチェックの目的でみると、その内容をシンタックスチェックのための手続きとみなすことができる。関数型の内部表現の場合、表現中にでてくる基本関数の定義に推論方法を組み込むことによって、それらを合成してできる名詞句に対応する内部表現と、文字列をマッチさせる方法を内部表現自体に持たせることができる。

内部表現の生成法について、

- 3) 文章解析時に、意味属性別に分類されたパターンを使うことによって、単なる構文的文章解析よりも正確な解析ができる。また、文章中の一部分の意味が逆に推定できる。
- 4) 構文パターンと内部表現のパターンを一对一対応に変換することは、内部表現が関数型表現の様に入れ子構造になっている場合に有効である。すなわち、ある知識表現が、他の知識表現を構成する要素としてそのまま埋め込める場合に、文の解析過程を逆にたどることにより知識表現を生成することができる。

6. おわりに

本研究は文章情報から知識を獲得、利用するための研究の一端として位置づけることができる。本研究と趣旨を同じとする研究としては西田他[1]の研究があるが対象や手法は本研究と異なるものである。また、本研究で採用した手法と似た考え方を持つものとしては、モンタギュー文法[2]がある。本研究の手法はモンタギュー文法とは全く異なるアプローチにより考えられたものであり、一般性を持ったモンタギュー文法の枠組みと、より特定の文献に対象を絞ったシステム（文法）との中間に位置するものである。

参考文献

- [1] 西田, 川村, 堂下: LSI の動作記述からの知識獲得について, 情報処理学会, 論文誌, Vol.26-No.6 (1985)
- [2] 長尾, 淵: 論理と意味, 岩波書店, (1973)

