

自然言語インターフェースのための 構文解析用ツールの試作

4L-9

藤山一敏*, 吉光宏*, 島田ひとみ**, 近藤省造**, 太細孝**, 岩下安男***

* 株式会社 三菱総合研究所 ** 三菱電機株式会社

*** 財団法人 新世代コンピュータ技術開発機構

1. はじめに

自然言語インターフェース [1] を構築するには、応用分野に依存する単語および文型の収集が必要である。我々は、そのための構文解析用ツールを逐次型推論マシン P S I 上で開発中である。

なお、自然言語インターフェースが扱う対話として、データ検索等の目的指向型のを前提に考えている。

2. 特徴

子供が言葉を覚える時、文法を意識することはなく、相手の発話とその子供が置かれている状況から間接的に受け取られる発話の意味とを対照することによって覚えていくのであろう。

この推測に基づき、我々は、システムに最初から文法を与えるのではなく、文型テンプレート (図-1) を用いて実現した。文型テンプレートとは、文の構成要素の順番および、それらの満たすべき制約条件と意味を記述したものである。

文型テンプレートを用いた理由としては、さらに次のことが挙げられる。

- (1) 未知文型の収集が容易である。
- (2) 意味の抽出が容易に実現できる。
- (3) データ検索等の自然言語インターフェースで扱う文は比較的単純である。

構文解析用ツールは、この文型テンプレートにより、入力文の構文解析および意味抽出を行うものである。

入力文に未知の単語が存在する場合、ユーザへの問い合わせによりその意味属性を取り込み、単語辞書に反映させる。

適合する文型が存在しない場合、ユーザへの問い合わせにより文型テンプレートを生成し、文型辞書に反映させる。

3. 辞書

3.1 単語辞書

単語辞書は、次の3種類の情報により構成される。

- (1) 単語見出し
品詞、意味属性情報を検索するためのキー
- (2) 品詞情報
品詞名称

文型テンプレート:

S → X の Y を Z

制約条件および意味:

X 品詞: 名詞
分類: 車の属性値
意味: value_of_condition
Y 品詞: 名詞
分類: 車名(車の属性値)
意味: object
Z 品詞: 動詞
ラベル: 買う
意味: request

上の文型テンプレートにより解析可能な文例:

2ドアセダンの **** を買いたいなあ。

図-1 文型テンプレートの説明

(3) 意味属性情報

名詞における分類

活用する品詞におけるラベル

(文型テンプレートとのリンク情報)

他の単語との関係 など

本構文解析用ツールでは、応用分野に依存しない全ての助詞、連体詞、接続詞と一部の名詞、動詞、形容詞、形容動詞、副詞の辞書を初期辞書として持つ。

動詞などの活用する品詞においては、それに接続する助動詞、助詞を含めて1つの単語として扱う。

3.2 文型辞書

文型テンプレートが格納された辞書である。

図-2に文型テンプレートの具体例を示す。

入力文:

akai kuruma ga hoshii na.

(赤い車がほしいな。)

文型テンプレート:

s(.,) → adje(.,value_of_condition),
noun(mono,object), [ga],
adje(hoshii,request).

図-2 文型テンプレートの具体例

A Parsing Tool for Natural Language Interfaces.

Kazutoshi FUJIIYAMA*, Hiroshi YOSHIMITSU*, Hitomi SHIMADA**, Shozo KONDO**, Takashi DASAI**, Yasuo IWASHITA***

* MITSUBISHI RESEARCH INSTITUTE, ** Mitsubishi Electric Corporation, *** Institute for New Generation Computer Technology

4. 処理概要

処理としては、文の入力、構文解析、文型生成・登録、および単語辞書登録の4つからなる。

ユーザから入力された文は構文解析部に渡され、構文解析および意味抽出が行われる。

構文解析法としては、BUP [2] を参考にした。文型テンプレートを次の単語の予測情報として用い、入力文と単語辞書とのマッチングを行いながらボトムアップに解析する。その際、文型テンプレートの予測が覆されたり、それと単語辞書の意味属性情報がマッチしない場合、その文型テンプレートは適合不可能として、次の文型テンプレートに対し適合性を調べる。

適合する文型テンプレートが存在しない場合、文型生成および登録処理に移る。

また、単語が存在しない場合、単語登録処理に移る。

5. 文型テンプレートの統合・一般化

文型テンプレートは、文例の増加に伴い、その数が膨大なものになることが予想される。これを解決するために文型テンプレートの統合・一般化を考える。

方法としては、論理和による一般化処理 [3] が考えられる。この方法は、共通部分を持つ文型テンプレートに対して次の処理を行うものである。

それぞれの文型テンプレートの共通でない部分を、新しく非終端記号を導入して置き替え、共通の文型テンプレートを生成する。共通でない部分については、先程の非終端記号に結びつけ、別の文型テンプレートとして、それぞれ登録する。

このようにして、複数の文型テンプレートを統合できる。

6. おわりに

現在、文例数と文型テンプレート数の関係、処理速度、未知単語収集能力等に関して、構文解析用ツールの評価を行っている。

今後、さらに検討すべき項目として次のものが考えられる。

- (1) 文型テンプレートの統合・一般化の強化
- (2) 単語登録、文型生成の場合のユーザ入力情報の類推機能の強化

- (3) 活用のある品詞の辞書の持ち方

なお、本研究は ICOT の委託研究である“知的対話システム”の一環として行うものである。

<参考文献>

- [1] 島田、他：知的対話システムの対話管理機能、信学会、人工知能と知識処理研究会、A186-8, 1986
- [2] 松本、田中：Prologに埋め込まれたbottom-up parser: BUP、自然言語処理研究会34-6, 1982
- [3] 田中、淵：人工知能ハンドブック（第Ⅲ巻）、共立出版、昭和59年、642-664

(a) 構文解析された例

```
** 文を入力して下さい **
kuruma ga hoshii.
** 構文解析結果 **
s (noun ([kuruma], mono, object,
          [is_a ([norimono])], "車"),
  adje ([hoshii], hoshii, request,
        [], "ほしい")
)
```

(b) 未知単語が存在する場合、単語登録処理を行う。

```
** 文を入力して下さい **
siguma ga hoshii.
** 構文解析結果 **
s (unknown ([siguma], noun, mono),
  adje ([hoshii], hoshii, request,
        [], "ほしい")
)
```

** sigumaは未知単語です **

** 単語登録処理 **

sigumaを登録しますか? (1:登録する 2:しない)

1

sigumaの分類は正しいですか? (1:はい 2:いいえ)

1

sigumaの上位概念を入力してください。

norimono.

sigumaの全体を表わす概念を入力して下さい。

.

sigumaの集合を表わす概念を入力して下さい。

shamei.

表示用テキストをメニューより選択して下さい。

(c) 適合する文型がない場合、文型生成・登録処理を行う。

** 文を入力して下さい **

akai kuruma ga hoshii.

** この文章に対する文型は登録されていません **

** 文型生成処理 **

カーソル位置で分割ならば1を、そうでないならばリターンを入力して下さい。

akai kuruma ga hoshii

1 1 1

分割は正しいですか? (1:はい 2:いいえ)

----> akai kuruma ga hoshii (ブランクは分割位置)

1

** 単語登録処理 **

akaiの品詞番号を入力して下さい。

1. 名詞 2. 動詞 3. 形容詞 4. 形容動詞

5. 副詞 6. 登録せず

3

akaiの終止形を入力して下さい。

akai.

否定形ですか? (1:否定形 2:いいえ)

2

表示用テキストをメニューより選択して下さい。

** 文型の生成・登録処理 **

akaiの意味を入力して下さい。

kurumaの格に対応する番号を入力して下さい。

1. 動作主格 2. 目的格 3. 源泉格 4. 傍着格

2

hoshiiの意味を入力して下さい。

request.

図-3 構文解析用ツールの実行例