

自然言語文の解析と生成の統一文法に関する一方式

4T-1 瀧塚孝志・杉村領一・赤坂宏・佐野洋・重永信一

新世代コンピュータ技術開発機構 東芝 松下電器

1. はじめに

談話における理解モデルやプラン・ゴールなどの実験的な自然言語処理の研究を行って行く上で、構文解析と文生成は計算機の入出力となる基本的な機能である。そのため、共通に使える基本的な自然言語処理ツールを逐次型推論マシン (PSI) 上で提供するための開発を昨年度より行っている。自然言語処理ツールの開発方針は、機械翻訳で行われている程度の量に対処できる辞書と文法を構築することである。本稿では、自然言語処理ツールの概要及び解析と生成の両ツールにおける辞書と文法の統一方式について報告する。

2. 自然言語処理ツール

ICOTにおける自然言語処理の研究を含め、理論的研究の多くが文脈理解を目標に進められている。しかし従来の文法の殆どは、一文内の言語現象のみを説明するものである。照応関係や接続詞の選択などは、文脈としての意味が一文の表層に属する現象として知られている。工学的に使用できる文脈や発話環境をも考慮した文法や単語選択規則は皆無に等しい。

また文の意味に関し、機械翻訳では表層の係り受け構造を格パターンとか概念ネットワークとして表現する方式が採られている。しかし文章理解においては、もっと規格化されたプリミティブを用いる表現形式でないと、一般的な推論規則が記述できない。例えば、次文を同義としたい。

(1) 私は、頭が割れそうに痛い。

(2) 私は、ひどい頭痛がする。

また、次文は異議に翻訳したい。

(3) 風で、頭が痛い。 (肉体的)

(4) 彼の事で、頭が痛い。 (精神的)

この様に単純化したときに、プリミティブな意味に関与しない多くの言語情報が失われ、自然な文を生成することが難しくなる。これらの言語情報のうち辞書から復元できる情報と談話の環境として管理すべき情報の切り分けに関しても殆ど議論がなされていない。

上述の課題 (照応・談話構造・プリミティブ翻訳・話題管理) を処理するモジュールが、将来上位階層に位置するものとして自然言語処理ツールの設計を行った。自然言語処理ツールは、以下の4つのモジュールからなる。

- | | |
|--------------|------|
| ① 記述言語 (CIL) | ③ 解析 |
| ② 辞書 | ④ 生成 |

昨年度は、意味の記述法や解析と生成に必要な文法規則が不明確であったため、それぞれ独自にモジュールを

試作することにした。そのため、辞書は構文情報を一切持たない概念意味記述であり、解析と生成は別々の辞書と文法を作成し使用していた。

量に対処できるツールとするには、細かな構文的素性の記述された大規模辞書と多数の文法規則が必要となる。そこで本年度は、解析と生成で使用する品詞の統一化作業から始め、意味の記述法と合成法の規格化作業を行い、辞書及び文法の作成基準を作ることにした。

2. 1. 解析ツール

解析は、およそ次のような手順で行われる。

- ① 形態素解析
- ② 構文解析
 - i 分節解析
 - ii 係り受け解析

形態素解析で品詞を切出し、分節解析で自立語と付属語からなる分節を切出す。係り受け解析で構文的に可能な依存関係を見付け出し、意味解析処理を呼ぶ。意味解析は、意味的に係りうるかを辞書に登録されている素性情報の単一化により検査する。

解析ツールで採用している格は、構文的必須格 (主格・対格・与格) と自由格である。構文的必須格に関しては、受動態のとき能動態の主格が「に／から」でマークされるのかを動詞分類により判定し、能動態の表層格に直す。表層格と深層格の対応は、辞書の情報により行うことにしている。

2. 2. 生成ツール

生成ツールは、一文の意味表現から自然言語文を生成する。意味表現には、記述言語 CIL の特性を利用して、深層格でラベル付けされた格フレームの表現を採用している。生成は、およそ次のような手順で行われる。

- ① 単語決定
- ② 格成分の語順決定
- ③ 助詞変換
- ④ 格成分の生成

単語決定では、構文的に期待される品詞及び語義と格情報から辞書を引き表層の単語を決める。辞書の深層格と表層格の対応、及び自由格の表層格と語順を決定する規則を用いて格成分の語順を決定する。派生名詞化や態変化などの構文的素性に基き表層格を助詞などに変換する。そして、格成分に関し同様のことを繰り返す。

解析及び生成で用いる構文情報に関し、格などの情報に加

え、述語には次の情報が付与されている。

- ・派生 (名詞化)
- ・態 (能動/受動)
- ・相 (単純/状態/始動/進行/反復/終結/…)
- ・時制 (現在/過去/完了/未来)
- ・様相 (推量/意志/義務/許可/禁止/伝聞/…)
- ・態度 (断定/疑問/命令/勧誘/…)
- ・極性 (肯定/否定)

格成分には次の情報が付与されている。

- ・指示 (この/その/あの)
- ・話題 (は/も/こそ/さえ/…)
- ・省略 (有/無)

生成では、語順や助詞を指示・変更出来るように更に細かな情報も決められている。また、文章理解のモジュールが細かな構文情報を考慮しなくとも文生成出来るよう、それぞれの既定値を決めている。

3. 品詞分類

計算機処理を行いやすいよう、新たな基準で品詞を再分類することにした。一般の品詞分類と大きく異なる点は、用言の活用形を細分し、活用形が決まれば一意的に去層の語尾が決まるようにしていることである。すなわち、助詞や助動詞とされていた「た/て/たり」などを活用語尾の中に入れ、音便処理を不要にした。活用形は、現在13個に分類されている。決定された品詞とその概要を述べる。

① 動詞

- i 強変化動詞 (五段動詞の多く)
- ii 弱変化動詞 (一段動詞の多く)
- iii その他
 - ・サ/カ変動詞
 - ・ある/いる/行く/なされる動詞

副詞との共起や相を検査するための素性や深層格と表層格の語義毎の対応、態変化規則、派生語・反意語など多くの情報が付与される。

② 名詞

形容動詞と言われている単語も名詞に含めている。構文的属性値として、「な」や準用詞の「の」の後方接続、連用性 (副詞的用法の機能)、動作性 (サ変性)、状態性 (格助詞の接続)、複合名詞性、語尾性 (さ、的、性…)、数量詞 (本、台…) が付く。意味的属性値として用言の格要素に入るかを検査するための情報が与えられる。

動詞から派生した名詞や名詞文で多用される名詞などには格を与える。

③ 形容詞

- i 普通形容詞
- ii ナ型形容詞

形容できる体言を検査するための素性や「がる・げ/まる・める」などの派生語の素性が付与される。形容詞も評価者や評価物などの格を持つ。

④ 副詞

構文的属性値として、「に/と/する」の後方接続性が付く。意味的属性値として「状態/程度/数量」などの素性が付く。主体めあての副詞と呼ばれる格を持つ副詞が

あり、表層の態を拘束する。

- (1) 太郎は、次郎をわざと殴った。(太郎)
- (2) 次郎は、太郎にわざと殴られた。(次郎)

数量副詞は、修飾する体言を持っており、動詞の素性により形態的にその体言が決まる。

- (3) それには、問題がたくさんある。
- (4) それには、たくさんの問題がある。

その他、次のものが品詞として採用し、辞書に登録する内容を決めた。連体詞、接続詞、準用詞、準体詞、助詞 (格助詞・とりたて詞・接続助詞・並立助詞・終助詞)、助動詞、補助動詞、補助形容詞、補助名詞、感動詞。

4. 深層格

深層格は、修飾のされかたにより次の4つに分類される。

- ① 格関係：行為者や対象、材料など主に格助詞でマークされる関係で、20個ほどの深層格を設定している。
- ② 事象関係：原因や理由、目的など接続詞や接続助詞でマークされる深層格である。
- ③ 意味的關係：格関係とならない「の」(私の本、像の鼻)の関係で、所有や部分、数量などがある。
- ④ 限定的関係：程度や様態、状況、時間など川言を修飾する深層格である。

5. 統一文法

文生成で問題となる語順や助詞の選択の基本的情報を辞書に記載することにより、解析と生成の文法統一を容易にした。辞書も文字列と語義の両方から検索できるように、一つの辞書から解析と生成用のインデックスを自動作成することになっている。

現在解析ツールでは構文解析までを行い、意味解析は辞書(ユーザ)まかせになっている。慣用表現の処理や多義語の解釈など解析と生成の完全な統一文法を考えたときには、意味構成規則から統一化しなければならない。

解析において、意味辞書を用いても一文内で減らせる構文の曖昧性には限度があり、たくさん構文木が生成される。機械翻訳では、最も妥当な構文木を選択するための文法規則が組み入れられている。文脈処理により係り受けの曖昧性を減らすとき、複数の構文木の中から適当な解釈を選択するより、曖昧性を表現する一つの構文木のほうが、問題点が明確になり処理がしやすい。そのため、係り受けの曖昧性を含めた構文木表現方式と妥当な構文木を選択するための文法規則を検討中である。結果がまとまりしだい、一つの意味構成記述から解析と生成用の文法を作成するための仕様をまとめる予定でいる。

6. おわりに

これらの自然言語処理ツールは、ICOTで現在開発中の談話理解システムDUALS-IIIで使用する計画である。そして、談話理解システムの構築の経験に基づき、照応関係や接続詞の選択などの文脈的言語現象や、発話しなければならない一群の意味を複数の文の意味として展開することを支援するモジュールを追加する予定でいる。