

PIVOT E-J:英語解析部の構成

2B-5

飯野 香*

石岡 隆光**

奥村 明俊*

亀井真一郎*

小山 剛弘*

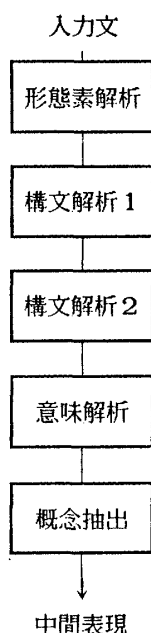
*日本電気株式会社

**日本コンベンションサービス株式会社

■はじめに:

機械翻訳システムPIVOTの英語解析部について説明する。PIVOTの英語解析部は英文一文を入力とし、それを、汎言語性を目指して設定された独自の中間表現に変換する。本稿ではこの変換過程の概要、およびその特徴点を述べる。

■構成:



■各部の構成:

1. 形態素解析

次の4個のフェーズからなる。

- ①形態素分割
- ②辞書引き
- ③接辞のまとめ上げ
- ④語幹と語尾のまとめ上げ

各フェーズは入力文を左から右へ一回走査する。出力は、形態・統語・意味情報の束からなる語の並びである。

2. 構文解析1

次の約10個のフェーズからなる(数個省略)。

- ①括弧・引用符の認定
- ②数量表現の認定とまとめ上げ
- ③コンマ用法の推定
- ④隣接語による品詞・用法曖昧性削減
- ⑤従属接続詞かもしれない語の品詞・用法曖昧性削減1
- ⑥動詞・形容詞・名詞等の品詞・用法曖昧性削減および名詞連続のまとめ上げ
- ⑦that省略の推定
- ⑧従属接続詞かもしれない語の品詞・用法曖昧性削減2
- ⑨等位接続の範囲推定
- ⑩文全体の構造に関する情報の収集

各フェーズは入力文を左から右へ一回走査し次のような処理を行う。

文の大きな区切れや入れ子の認定(①③)、局所的な品詞・用法曖昧性削減(④⑤⑧)、および局所的な係受け関係の認定と構造化(②⑥)。

また、(1)文は節からなり、節にはただ一つの定動詞が出現すること(2)動詞の必須格パターン情報。この二つを主要な手がかりとして、文全体の整合性を意識しながら品詞・用法の曖昧性を削減する(⑥)。係受け構造は、ごく局所的なものしか作らないが、ここで一度、文全体の構文解析を行っていると言ってもよい。

構文解析1での曖昧性解消は、不可能とわかった品詞・用法を削除する形で行われ、最終的な決定は次の構文解析2に委ねられる。

最後に、定動詞の形と位置、従属接続詞・文の等位接続詞・カンマや括弧による挿入の位置など、文全体の構造に関わる情報を収集して構文解析2に渡す(⑩)。

出力は形態素解析部と基本的に同じで、一部構造化された語の並びである。

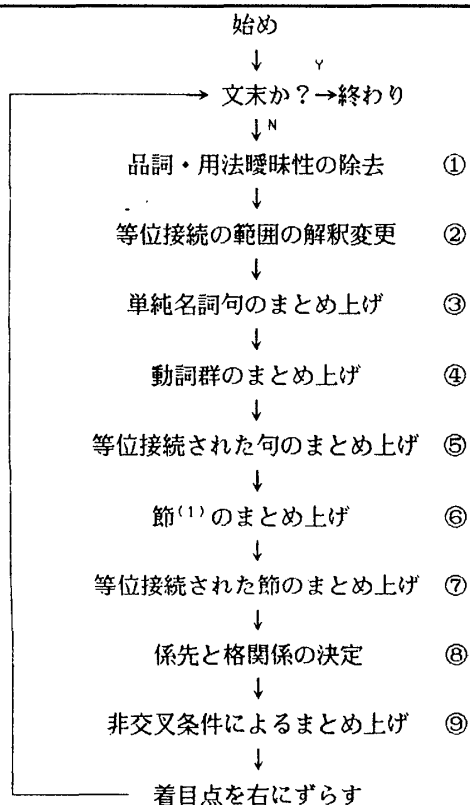
3. 構文解析2

一つのフェーズからなる。文頭から始め、着目した各語に対して次の処理を加えながら右へ移動し、文末に至って構文解析を終了する。

PIVOT E-J: The structure of its English analysis module

Kaori IINO*, Takamitsu ISHIOKA**, Akitoshi OKUMURA*, Shin'ichirou KAMEI*, Takahiro KOYAMA*

* NEC Corporation, ** Japan Convention Services Inc.



注(1) 節の前半のみ。節の終わりはまだ判らない。

まず、語の品詞・用法曖昧性の除去(①②)を行う。構文解析2は、構文の解釈を一意に決定しながら進むので、ある時点で着目している語より左側の部分の係受け関係は(間違っている場合も含めて)完全に定まっている。そこでまず、これと矛盾しないように、構文解析1が残した曖昧性を削減する(①②)。それでも残った場合は、構文解析1の⑩が収集した文構造情報、品詞の頻度情報(辞書情報)、および経験則的ルールによって、品詞・用法を最終的に決定する。次に、句・節のまとめ上げを行う(②~⑦)。次に、今まとめ上げた句・節の係先と格関係を必須格パターン情報または自由格の意味関係の知識ベースによって決定する(⑧)。必須格の場合は文法格のみ与える。最後に、係元と係先の間にはさまれた部分があれば、その部分を構造化する(⑨)。

4. 意味解析

次の2個のフェーズからなる。

- ①意味選択
- ②文法格から深層格への変換

各フェーズは構文解析が出力した木構造をそれぞれ一回走査する。意味解析はまず、次の3つの操作によって意味選択を行う。

- a. 用言と格要素との間の意味的な制限により取り得ない意味を取り除く。
- b. 熟語や数量表現などのパターンを捉えてその意味を選択する。
- c. 用言と格要素との間の関係の強さやその意味の一般的な妥当性から点数を計算し、最も点数の高い意味を選択する。

そして、意味選択の結果、用言の文法格と深層格の対応が一意に求まり、文法格を深層格に変換する。

5. 概念抽出

1つのフェーズからなる。

意味解析が出力した木構造を一回走査し、原言語に依存した構造から汎言語的な中間構造に変換する。構造変換には次の3つがある。

- ①自立語で表現されている関係の概念を格関係に変換。
(have「もつ」を所有者と所有物の関係に変換する)
- ②格関係の兄弟間に関係が成立するような用言において兄弟間の関係を考慮した構造に変換。
(I believe him to be honest. を I believe that he is honest. の構造に変換する。)
- ③原言語に依存した複合的な概念を展開及び縮退する
(hose を「ホースで」、「水を」、「まく」という3つの概念に展開する。)

■ 特徴 :

1. 多段の解析

構文解析1と2は、基本的に同じ言語現象を対象とするルールセットを備えており、原則として1では可能性の絞り込み、2では決め付けを行なう。また、1の中でも同様の能力を持つルールが複数回適用を試みられる場合がある。つまり、種々の処理を積層させて繰り返しトライすることや、解釈の見直しにより、相互に依存して決まる語の構文属性の可能性の中から、慎重に正解を導き出す。

2. 文の全体的構造を踏まえた解析

構文解析2は、1が概ね把握した文の全体的構造を参照しながら解析を行う。それが、人間の構文把握と比較して不足している意味情報を、不完全ながら代替する。

謝辞:

本システム構築に多大な貢献をされた Steve Cudhea 氏と西川知子氏に感謝します。

参考文献:

- 1) 飯野・奥村『PIVOT E-J:英語形態素解析』
- 2) 小山・奥村・亀井『PIVOT E-J:英語構文意味解析』ともに昭和63年電子情報通信学会秋期全国大会