# 1F-2 文書全体の情報を利用した翻訳方式

田中 克己, 野上 宏康, 平川 秀樹, 天野 真家
(株)東芝 総合研究所

## 1.はじめに

自然言語処理のひとつの応用としての機械翻訳システムは、現状では基本的に入力文書を文単位に区切り、一文ごとに翻訳処理を行う仕組みになっている。しかし、現行方式で行われている一文単位の翻訳処理では、自然言語が生来持っている、文脈に依存した曖昧性のために一文のみの翻訳結果としては適切であっても、文書全体として見たときには不適切な翻訳結果を生じてしまう可能性がある。ここで実際には一文単位の処理だけからは解消できない曖昧性を、文書の他の部分の情報から解消できる場合がある。本稿ではこの考えに基づいた翻訳処理方式の構想と現行の翻訳システムへの適用手法について述べる。

## 2.翻訳処理における曖昧性

翻訳処理においては入力文一文単位の処理では解決できないと思われる曖昧性が各段階において存在する。その中から重要であると思われる曖昧性を挙げる。

(1)品詞の曖昧性

同一の単語は、複数の品詞を持つ場合が多い。ここで適切な品詞の選択を誤ると次の構文解析処理で構文解析に失敗するため品詞の選択のやり直しが必要になり、処理効率が低下する。

例1)英語文"Output Primitives"の解析において、単語"output"には名詞、動詞の現在形、過去形、過去分詞形の4つの品詞の候補が考えられる。この部分だけではそのうちのどれを選択するかを適切に判断することは不可能である。

(2)単語の区切りの曖昧さ

日本語のような膠着語については、形態素解析の際に文を単語の列に分解する処理が必要である。この処理の時に単語の区切り方に曖昧性が生じることがある。また、英語のような単語が分かち書きされる言語についても単語をイディオムの一部ととるか単独の単語ととるかどうかによる曖昧性が存在する。

例2)「今日本人に会った。」という日本語の文には、

今日/本人/に/会った。

今/日本人/に/会った。

の二つの単語の句切りの可能性が存在する。

このうちどちらを選択するかについては、システム側で用意された単語と単語の接続情報を用いて優先度を決定することになるが、この一文だけでは適切に判断することは不可能でる。

(3)構文的曖昧さ

構文解析処理では、構文的に適切な構文構造を生成するこの際、入力情報に対して構文的、意味的にも妥当な複数の構造が生成される場合がある。

例3)係り受けの曖昧さ

英文"The same header file defines some interesting macros on rectangles."においては前置詞句"on rectangles"は動詞"defines"と名詞"macros"のどちらにも係る可能性があり、この一文からは構文解析処理時にはそのうちのどちらかを判定することはできない。

例4)並列関係の曖昧さ

英語の名詞句"Painting panels and individual items"においては、単語"item"の意味が漠然としているために、意味的な解釈における

[Painting panels] and [individual items]

Painting [panels and individual items]

("[","]"で囲まれた部分は句の範囲を示す。)

の二つの可能性のうちどちらを選択するかをこの部分のみから決定することは不可能である。

## 3.曖昧性の解消方法

第2章で説明したように翻訳処理の各段階でそれぞれ曖昧性が生じる可能性がある。翻訳処理全体としてみた場合の曖昧性は各段階における曖昧性を乗算したものになり、処理が進むにつれて曖昧性が増加していくことになる。

しかし人間が実際に文書を読み理解する過程においては曖昧性が生じていない場合が多い。そのような場合の例を図1、2、3に示す。ここでは下線部Aの部分によって下線部1の部分の曖昧性が解消されている。これは下線部Aの部分の解析には曖昧性が生じることがなく解析結果が一意に決定され、それと同様な部分の解析に影響を与えて同様の解析結果を生じるような解釈をさせているためだと思われる。ここではこの仮

Output primitives
1

The system provides function for drawing geometical output primitives(for example,polygons,circles,and ellipses)as well as functions for performing raster operations. The coordinates of output primitives are specified in X space(with the exception of some raster functions),Geometical output primitives include rectangles,polymarkers,circlar and elliptical arcs.

| 1. Output primitives | A. of output primitives |
|---|---|
| ○ 名詞 | ○ 名詞 |
| 動詞(現在形) | ~~動詞(現在形)~~ |
| 動詞(過去形) | ~~動詞(過去形)~~ |
| 動詞(過去分詞形) | ~~動詞(過去分詞形)~~ |

図1 品詞の曖昧さの解消例

---

Machine translation system using infomation retrieved from the whole document
TANAKA Katumi, NOGAMI Hiroyasu, HIRAKAWA Hideki, AMANO Shin-ya
TOSHIBA Corp.

説に従って、第2章での問題の解決手法として翻訳対象の文書全体から情報を抽出し、それを用いて翻訳処理の各段階の曖昧性を解消する方法を提案する。

基本的な考え方としては、図4に示すように2段階の翻訳処理を考え、第1フェーズで文書全体から曖昧性を解消または減少させるための情報を抽出し、その情報を第2フェーズの翻訳処理の際に参照しつつ各段階の処理を行なうことである。具体的に言うと、まず第1フェーズの翻訳処理を行って曖昧性がない、あるいは少ない結果が得られた場合、そのときの処理内容を記憶しておく。第1フェーズの処理終了後に第2フェーズの翻訳処理を行い、そこでは処理中に曖昧性が生じた場合は第1フェーズの処理により得られた曖昧性を解消するための情報を検索する。その結果両者の処理内容に一致するものが存在した場合、それを優先するようなに選択を行なう。これにより曖昧性の解消または減少がなされることになる。

この方法により同一文書中においては同一の解釈を優先するということになる。また処理の対象となる文の前方だけでなく後方の情報も利用する事が可能になる。

## 4.おわりに

現在、形態素解析処理における妥当な品詞の選択について着目し、そのために必要な情報の抽出、利用の方法について検討を進めている。また意味解析処理における係り受けの決定は訳文の品質に係わりが深く、その曖昧性を解消することは重要な課題となっている。これについても引続き検討を行っていく予定である。

## 参考文献

1)稲垣他: 意味連結パターンを用いた係り受け解析, 情報処理学会研究会資料, NL67-5

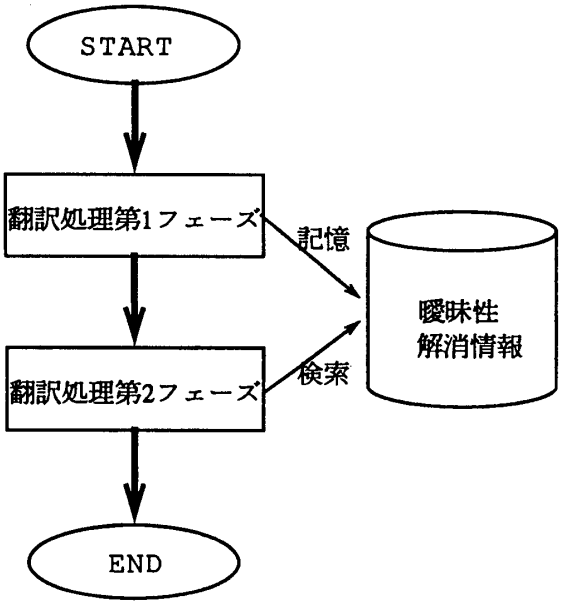


図4 文書全体情報利用のための処理方式

Macros on rectangles
A

The same header file defines some interesting macros on rectangles. To determine an edge not given explicitly in the rect:
1

```
#define rect_right(rp)
#define rect_bottom(rp)
Rect *rp;
```

returns the coordinate of the last pixel within the rectangles on the right or bottom,respectively.

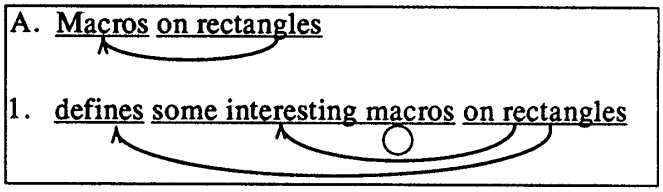


図2 前置詞句の係り受け関係の曖昧さの解消例

Painting panels and individual items
1

To paint a panel and an item,use:
A

```
panel_paint(panel_object,paint_behavior);
```

paint_behavior should be either PANEL_CLEAR, which causes the rectangle occupied by the panel and item to be cleard prior to painting,or PANEL_NO_CLEAR,which causes painting to be done without any prior clearing .

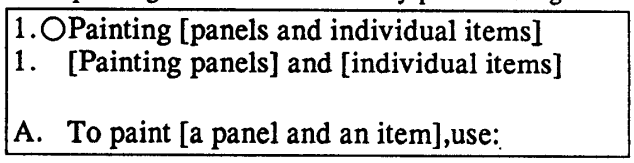


図3 並列関係の曖昧さの解消例