

翻訳における多義性を例表現選択により解消する方式

5B-3

山口 昌也 滝口 伸雄 小谷 善行 西村 恵彦
(東京農工大学 工学部 電子情報工学科)

1. はじめに

計算機で自然言語処理を行う際、問題となることの一つに、多義語の処理がある。自然言語における単語の曖昧性を解消するために、これまで多くの研究がなされてきている^[1]^[2]。

現在、我々が作成している翻訳支援システムでは、システム利用者が知っている言語から知らない言語へ翻訳するという状況を想定し、自然言語の多義性を人間との対話によって解消する。本システムでは、まず、文を句の集まりと考え、句の分割をユーザとの対話により確定する。次に、翻訳対象言語で表した例表現を提示することで単語の意味を確定する。

従来の自然言語処理システムの多義語の処理では、処理された結果を見ないと、入力した文が正しく処理されたかどうか確かめることができない。例えば、日本人が日本語から英語に翻訳する場合を考えてみる。この場合、出力された結果を入力した人が理解できないとすると、システムがどのように処理したのかを推定することは難しい。しかし、システム利用者が入力文の意味を知っている場合、その意味をシステムに説明することができる。さらに、システムの出力が人間にとて理解できるものであれば、システムの処理した内容を確かめることができる。そこで、本稿では、源言語の例表現を用いて、対話的に多義語の問題を解消する方式(図1)を提案する。

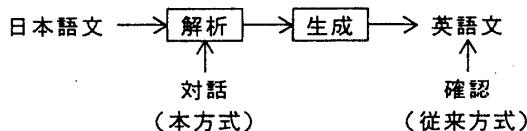


図1 本方式による対話の位置づけ

2. 例表現選択による説明

例えば、「太郎は、クリームを落とした」という文を考えてみる。この場合、「落とす」という単語の意味は、文脈情報を用いないと特定することができない。少なくとも次の三つの意味にとれるだろう。

- 意味1 “クリームを（床などに）落とした”
- 意味2 “（服などについた）クリームを落とした”
- 意味3 “クリームを（何かの選択候補から）落とした”

以上のような条件で、もし、システムが“花瓶を落とした”という例表現を出力し、人間がそれを選択したとする。この場合、意味1の“落ちる”的意味を選択したことになる。一方、もし、“汚れを落とした”という出力を選択した場合は、意味2の意味になる。

このように、利用者は、提示された例表現を選択することによって、自分の意図した意味を伝えることができる。これは、図2のAの部分の選択にあたる。

これと同様のことを、出力側の言語表現を選択する場合にも行うことができる。例えば、英語の場合、“笑う”的意味に対して、複数の表現がある。その一部を次に示す(日本語の表現は、例表現である)。

laugh : 大声で笑う
smile : にっこりと笑う

このとき、たとえ利用者が入力文の中で“笑う”とだけ表現していても、システムが例表現を提示して、利用者が自分の意図にあった表現を選べば、それに対応した表現を選択することができる。このことは、図2のBの部分の選択にあたる。このように、出力側の言語を知らないても、翻訳源言語の例表現を選択することによって、利用者の意図する意味を説明することができる。よって、入力側、出力側の言語の違いを意識することなく対話できる。

さらに、例表現は、システムの選択した意味を利用者に説明する場合も有効な手段である。例えば、なんらかの制約によって、対話の必要なく意味を決定できたとする。この場合、利用者がその意味を尋ねたとき、システム側が選択した意味の例表現を出力することによって、説明することができる。

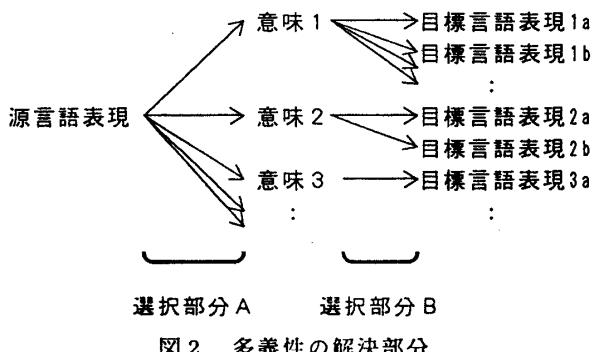


図2 多義性の解決部分

3. 対話処理

3.1. 対話の流れ

ここで示す対話の流れは、動詞の格支配構造を基本としている。まず、格支配構造に基づいて動詞の意味を限定する。そして、残った意味に対して、例表現を使ってその多義性を解消する。次に、動詞に係るもののが意味を特定していく。

このような流れにしたのは、動詞の先に解消するこ

によって、複数の語の意味を特定できる可能性があるからである。これによって、対話の回数を減らすことができる。

次に、大まかな対話処理の流れを次の例文を使って説明する。

例：「太郎 が 扱っている 機械 を 調べる」
 (1) 入力文中に、最初に現れる動詞(V1)の格要素を特定する。そのために、部分的に構文解析を行う。つまり、動詞V1の格要素が得られるまで構文解析する。例の場合は、"太郎 が 扱っている 機械" までである。

そして、得られた格要素を画面に提示する(格要素は、網掛けして示す)。

例： 太郎 が 扱っている 機械 を 調べる
 (2) 提示した格要素が不足である場合は、利用者がそれを提示する。一方、格要素でないものが含まれていた場合は、それを利用者が削除する。

例："太郎" が "調べる" に係るのであれば、"扱っている" の格要素ではない。したがって、そのことをシステムに伝える。

(3) まず、動詞V1の格支配構造に合わない意味は削除する。合うものだけ、例表現を生成し、それを提示する。生成の方法は、3.2で示す(意味の分類は、参考文献[4]にしたがった)。

提示例： 片手でハンドルを扱う(操作するの意)
 事件を扱う(取り扱うの意)

(4) 利用者は、提示された例表現が自分の入力した単語の意味と同じ場合は、提示された例表現を選択する。利用者が同じ意味を示しているかうまく判断できない場合、システムは、さらに異なる例表現を提示する。複数の例表現により、表そうとしている意味が特定されていくと考えられる。

(5) 動詞V1の意味を特定した後、V1の格要素の意味を特定していく。ただし、それぞれの要素が多義でなければ、対話は行わない。ただし、意味を出力するように要求されたときは、例表現を作成し、出力する。

(6) すべての格要素を特定し終わったとき、動詞V1の述語素から意味構造を作り出す。そして、文中で次に現れる動詞V2, V3...について同様の処理を行っていく。

(7) 入力文のすべての意味構造を作成し終わったとき、知識ベースへ格納する。

3.2 例表現生成方法

例表現は、知識ベースにすでに格納されているインスタンスをもとに作成される。一度に複数の意味の例表現を提示するので、それぞれの意味が明確になる例表現でなければならない。ここでは、動詞と名詞の多義性を解消する際に用いられる例表現について示す。

3.2.1 動詞

動詞の意味の例表現は、動詞の格支配構造に基づいて生成する。このとき、それぞれの意味において特徴的な格支配構造を強調する。その要素として、次のことが考えられる。図4のような格支配構造(参考文献[4]による)を持つ動詞"投げる"を例として説明する。

- 他の意味にはない格要素を選択する。
例：マウンドから投げる(意味1, カラ格 LOC)
- 格要素の意味属性の違いに着目する。
例：チームは試合を投げた(意味2, ガ格 ORG)
疑問を投げた(意味3, ヲ格 MEM)
- 例表現に含まれる格要素は、多義性の少ないものを選ぶ。

| 意味 | 格支配構造 |
|-----|--------------------------------------|
| 1 : | 方格 HUM ヲ格 CON カラ格 LOC |
| 2 : | 方格 HUM/ORG ヲ格 ACT |
| 3 : | 方格 HUM/LIN/ACT ニ格 HUM/ABS ヲ格 ACT/MEN |

図3 "投げる" の格支配構造

3.2.2 名詞

名詞の例表現を提示する部分は、3.1の(5)である。名詞の例表現は、動詞の格要素として表す。名詞の場合、図4のように、それぞれの意味ごとに異なった動詞を用いる。

ここでは、3.2で述べたように知識ベースのインスタンスとして存在していることを前提としている。名詞の例表現を作成する際には、次のような動詞を事項に基づき、選択する。

- 対象となる名詞が置かれる格は、できるだけ複数の意味属性をとらないもの
- 名詞の持っている意味と格支配構造のとる意味が複数重ならないもの
- なるべく多義でない動詞

ベースを弾く(楽器のベース)
 ベースを担当する(バンドのベース)
 走者がベースを回る(野球のベース)

図4 "ベース" の提示例

4. 終わりに

本稿では、例表現によって、多義語の問題を解決する方式を示した。本方式では、翻訳対象言語を用いて例表現を行う。このことにより、翻訳目標言語に関する知識を利用者に要求することなしに、多義語の問題を解決できる。今後、プロトタイプのシステムを作成し、対話における手間及び作成される例表現の評価を行う予定である。

参考文献

- [1] 長尾確, 丸山宏, 自然言語処理における曖昧さとその解消, 情報処理 Vol. 33 No. 7, 1992
- [2] Hiroshi Maruyama, Hideo Watanabe, An Interactive Japanese Parser for Machine Translation, COLING-90, pp. 257-262, 1990
- [3] 情報処理振興事業協会技術センター, 計算機用日本語基本動詞辞書 I P A L (Basic Verbs) 説明書, 情報処理振興事業協会, 1987
- [4] 情報処理振興事業協会技術センター, 計算機用日本語基本動詞辞書 I P A L (Basic Verbs) 辞書編, 情報処理振興事業協会, 1987