

## 6D-5

## 名詞句の同一性の理解と応用

野垣内 出 飯田 仁

ATR自動翻訳電話研究所

## [1] はじめに

自然言語理解の立場から、名詞/名詞の照応関係の理解を考察すると、処理の目的によっては、“個としての同一”まで理解できなくても“集合としての同一”で充分、処理が行えることも多い。機械翻訳を例にすれば、名詞の訳語の選択、知識ベースとのリンクなどの処理は、“集合としての同一”で充分であり、名詞の訳語などの処理を行うことができる。また、日本語のように名詞が冠詞などを伴わない裸の名詞で使われる言語では、“個としての同一”の理解が必要とされる場合でも、“個としての同一”の理解の前提条件として、“集合としての同一”を設けておけば、処理上有効である。本報告では、模擬対話<sup>[Arita87]</sup>における名詞/名詞の照応関係の理解の手法と実験システムについて、“集合としての同一”を中心に報告する。また応用例についても述べる。

## [2] 集合としての同一の理解

発話に現れた名詞Aと名詞BがどちらもCと呼べるものを指示しているとき、“集合としての同一”と呼ぶことにする。例えば、“申込み書”と“用紙”がどちらも“参加申込み書”と呼べるものを指示しているとき、“集合としての同一”と呼ぶことにする。また、これらが1枚の“参加申込み書”を指示しているとき、“個としての同一”であるとする。

現在、以下に示す“集合としての同一”の理解のための条件は、エキスパートシステム<sup>[ART87]</sup>上にまとめられている。

## [2.1] 同一のための条件

照応についての条件を明らかにするために、特定の目的を持った会話の中から照応関係にあるものを選び、これらが照応するための条件を考察した。“申し込み用紙”の例では、“参加申込み書”と“用紙”のように、意味分類上の上下関係にあるもの、“参加の申込み用紙”、“登録の申込み書”における“参加”、“登録”のように、前提となることからの関係にある語を用いている例が多い。これらの条件は、名詞とそれに対応する知識ベースの関係として記述できる。例えば、前者の条件“概念的に上位/下位関係にある”については、対応するドメイン知識<sup>[Nogaito88]</sup>(国際会議の登録に関する知識ベース)の要素が、is-a relationにあるかどうかで判断する。実験システムではfig. 1に示すルールとなっ

ている。このルールでは、対話上の2つの名詞がドメイン知識上でis-aの関係であれば、照応の可能性があるととして、weak-identと呼ぶ関係を与えている。

```
(defrule identify100
  "setting is-a-weak-ident"
  (schema 名詞1) ;発話中の名詞1
  (schema 名詞2) ;発話中の名詞2
  (schema 名詞1 (is-a 名詞2))
  =>
  (bind ?rel 'is-a)
  (assert
   (gensym "weak-ident")
   ...)))
```

fig. 1 ルール化した同一のための条件の例

## [2.2] 制限的な条件

この同一のための条件は、多くの候補を示すためこれらから適切に選び出す必要がある。このため、候補に関して制限的な条件<sup>[Nogaito87]</sup>を与えている。これらの内、代表的なものを以下に示す。

- ・一貫性 ある語があるものを指示していれば、その語は別のものを指示できない。
- ・別のものを指示するマークが存在する。(別の、他の、他に、...)
- ・共起しえない属性がある。(赤い用紙、青い用紙)

これらの制限的な条件は、主に同一発話内の条件であるが、これらの制限だけでは、解消されない例のある。次の例は、最も近くにある候補が正しくない例である。

申し込み者: 会議で使われる言葉は日本語ですか? 英語には自信がないのですが。 [d-1-1]

事務局: 今回の会議は国際会議で、外国からの出席者も多く使用言語は英語となっています。 [d-1-2]

この例では、会議で使われる“言葉”の候補として、“日本語”がでてくるが、名詞と名詞の関係についての疑問文であるために、応答まで発話内の名詞の処理を行っていない。“言葉”と“使用言語”が

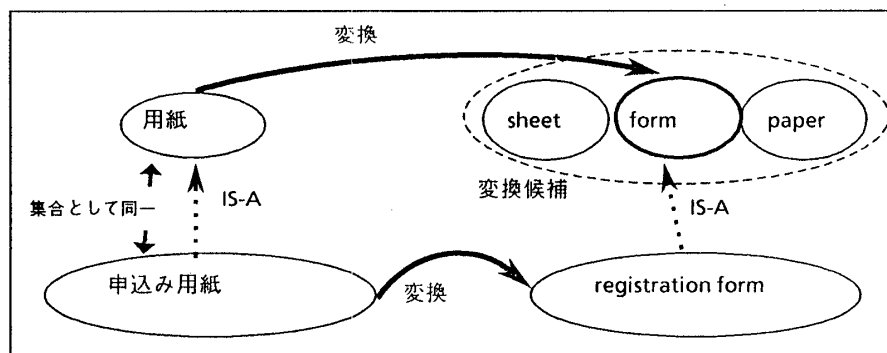


fig. 2 照応関係にある名詞の変換

同じであることと、事務局側の回答により、会議で使われる“言葉”が“英語”だけであることが理解できる。

### [2.3] 推論による同定

対話のなかに同定のためのマーカがなく、解析範囲を限定できない場合もある。以下に示した2例では、“ホテルの宿泊料金”の話題に続く会話の例である。同じの話題に続く対話であるが、“料金”は、それぞれ“タクシー料金”、“ホテルの宿泊料金”となる例である。

事務局: 1泊シングルで、6000円から10000円です。 [d-2-1]

申し込み者: わかりました。京都駅からタクシーを使った場合、料金はいくらぐらいかかりますか? [d-2-2]

この例では、“料金”は、直前の“ホテルの宿泊料金”ではないが、この発話の代わりに以下の発話があると“料金”は“ホテルの宿泊料金”となる。

申し込み者: わかりました。割引を使った場合、料金はいくらぐらいかかりますか? [d-2-2']

これらの同定には、推論による同定を行う。同じ発話内の名詞に対して条件を示す部分がある場合には、発話内に名詞は発話外の名詞を指示できるとして推論を行っている。例えば、“タクシー”を使えば、“タクシー料金”がかかる。“タクシー料金”は、“料金”と言えるので、“料金”は、“タクシー料金”の可能性がある。会議参加者は、“割引”を使うことができる。“割引”の対象の中に“ホテルの宿泊料金”がある、などの知識から推論を行い、料金は、それぞれタクシー料金、ホテルの宿泊料金であることを得る。

### [3] 同一性の理解の機械翻訳への応用

機械翻訳において、言語間で名詞が示す概念上の差があるとき、訳語選択のために、同一性の理解が必要となる。特に対話上の名詞がその単語自身では、訳出のための十分な情報を持たない場合、訳出の際、目標言語側で、照応関係が成り立たなくなることがある。例えば、“用紙”は、ある目的のための紙であり、その用紙の目的、形態などによって訳語が決まる。原稿用紙などは、paperで

あり、登録用紙は、formである。“用紙”自身には、訳出のための情報はない。しかし、登録用紙を指示している“用紙”をpaperと訳せば、英語側の聞き手は、paperの先行詞となるもの、例えばmanuscript paper (原稿用紙)を連想して、対話上に混乱を生じさせることとなる。

照応関係を変換後も維持するための手法について述べる。原言語側で、“用紙”と“登録用紙”が照応関係にあるとする。先にも述べたように変換対象である“用紙”には、情報がないので、変換対象である“用紙”の先行詞を検索する。次に、先行詞の変換を行い、この対象の変換候補と先行詞の変換を知識ベースで解析を行い、照応関係をもつものを選ぶ。先の“用紙”の変換候補としては、sheet, form, paperがある。一方、“用紙”の先行詞である“登録用紙”の訳は、registration formであり、registration formと“用紙”の変換候補と照応関係にあるかを知識ベースの階層関係を調べることにより得る。そして、これら変換候補の内、照応関係にあるformを適切な訳とする。すなわち、訳語であり、対象言語側でも照応可能であるものである。[fig. 2 参照]

### [4] おわりに

対話における名詞/名詞の照応の理解の手法について述べた。ここで示した手法は、エキスパートシステム上でルールの形で実験システムとして実現されている。また、その応用を示した。

### 謝辞

本研究の機会を与えて下さるとともに、適切な助言を述べられたATR自動翻訳電話研究所 榎松 明 社長、同言語処理研究室 相沢輝昭 室長、また議論に参加してくれた言語処理研究室の方々に感謝します。

### 参考文献

- [Arita87] 有田,飯田,「日本語におけるタスクオリエンテッドな対話の構造」,電子情報通信学会,NLC-87-10,1987
- [Nogaito87] 野垣内,「照応・指示関係の同一性の分類・解析」,人工知能学会研究会資料SIG-FAI-8701-1,1987
- [Nogaito88] NOGAI,“Noun Phrase Identification in Dialogue and its Application”, 2nd International Conference on Machine Translation, CMT,CMU
- [ART87] ART Reference Manual, Inference Corporation, 1987