

融合方式による機械翻訳システムの実験

田中 穂積, 井佐原 均, 安川 秀樹
(電子技術総合研究所) (ICOT)

はじめに

筆者等は[田中他 82]で, トランスファ方式に代わる新しい機械翻訳の方式として融合方式(Integral Method)を提案し, その詳細について論じた. この方式の大きな特徴は, 意味解析を行い意味構造を抽出する過程で訳語の選択を行うことであり, 訳語選択のためのいわゆるトランスファ(変換)過程が意味解析過程と融合していることにある.

英日機械翻訳の場合には更に, 意味解析の過程で, 翻訳文の語順を決定すると共に助詞の選択を行う. したがって融合方式によれば, 意味解析が終了するとともに日本語文が合成される.

以下では, 融合方式による機械翻訳のその後の検討結果と実験結果を, 英日機械翻訳実験システムを中心に述べる.

2 意味表現用言語 SRL の拡張

[田中他 82]では, 機械翻訳用の辞書項目の記述として, [田中 79]で提案した意味表現用言語 SRL を用いることを述べた. その後の検討により, 機械翻訳

用の辞書項目の記述として, SRL に 2つの拡張を施した. これは,

- (1) 1つのスロット中に, 複数個の制約条件とアクションの対がCASE文として書けるようにした.
- (2) self スロット中のプロトタイプ記述に訳語を並記できる様にした.

(1)の例を図1に示す. 上から7行目にCASE文がある. 8行目にば door なども主語になりうる事が書かれている. この場合には, 9~10行目のアクションが起動され, 以下の obj スロットが REMOVE されるので, open は目的語を欠く自動詞であると判断されることになる.

従って, The door opens the door に対する意味解析は失敗することになる.

(2)の例を図2に示す. 図2の上から5行目のプロトタイプ記述(a SYMBOL)には, "文字"という訳語が並記されている. self スロット中のプロトタイプ記述は, このユニットに対す

```
(OPEN V ((VI VT1 VP6A VP2A ...) 0)
  *((OPEN 開 K・U)
    unit
    (part-of)
    (self (a DEFAULT-CASE))
    (sf !+action)
    (subj (CASE (((OR (a HUMAN) ...)))
      (((OR (a DOOR) ...)))
      (@REM-S 'obj
        (*SUB-U THISONE))))))
    (obj ((OR (a DOOR) ...))
      (@YAKUQO ORIGIN *((開 KE・RU))
      (@G0BITSUKE FILLER 'WO))
    (prep ((a INSTRUMENT) WITH)
      (@G0BITSUKE FILLER 'DE)
      =instrument)))
```

図1 open の辞書記述

```
(CHARACTER N ((UC C) 0)
  *((CHARACTER 性質)
    unit
    (part-of)
    (self ((a SYMBOL) 文字)
      (a NATURE))
    (sf abstract)
    ...))
```

図2 character の辞書記述

```
{ I hate his character.
  The processor sends
    a character.
```

る様々な見方 (view) を表現したものであると見なす。このユニットをどのように見なすかによって、日本語の訳語が異なる場合が少なくない。そこで、プロトタイプ記述に訳語を並記しておくことにより、意味解析により CHARACTER という語を SYMBOL として見なすべきことが分かった場合には、(α SYMBOL) に並記した訳語“文字”を CHARACTER の訳語とすることかできるようにする。

用言の訳し分けは、スロットに付加されたアクションを起動することにより容易に実現されることは [田中(他) 82] ですでに述べた。(2) に述べた SRL の記法の拡張は、体言の訳し分けについての 1 つの手がかりを手えるものである。

3. 訳語の構造

日本語の用言は、様々な活用変化、音便変化をするだけでなく、助動詞の相互承接においても語尾変化を伴うものである。我々は、日本語の用言の訳語の構造をつぎの形式にした。

(α β · γ)
 └──┬──┘
 語幹 変化語尾

ここで、βとγの部分はアルファベット表記とすることとした。(ただし、βにはNILを許す)。図1ではすでに

(南 K · U)

が open の訳語であることが示されていた。その他の例をあげる。

(見 NIL · RU)

(食 BE · RU)

(破壊 S · URU)

(ヒカ R · U)

(寒 NIL · I)

以上の例から明らかな様に、そこから活用変化の種類が分かるだけでなく、βによって音便変化の種類が分かる。アルファベット表記のβとγは、最後にこれを平仮名のコードに変換するプログラムを介して、日本語表示されることになるが、そのためのプログラミングの労力は小さい。

4. ホーンビーの文型パターン

[高松(他) 81]では、ホーンビーの文型パターン [ホーンビー 77] を用いて英語の動詞の訳語選択が可能であることが説明されている。動詞 get の日本語訳語選択については、確かにホーンビーの文型パターンが利用可能である [高松(他) 81]。

ところがこの様な場合は極めて少なく、動詞 take に代表されるごとく、むしろ主語や目的語になる名詞がどのような意味を持つかによって、take の日本語訳が決まることの方が多い。

筆者等は、ホーンビーの文型パターンは、むしろ英語の文の基本的な構造を決めるための文法規則として利用すべきであると考え、そうした英語の文法規則を開発している。たとえば

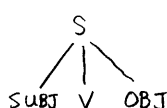
S → SUBJ V OBJ

などである。この規則を適用すべきかどうかは、Vがどのような文型パターンであるかに依存して決まる。たとえば図1の open の記述中の <advice>部 [田中 79] には、VP6A という文型パターンが書かれているので、この文法規則は適用可能となる。こうした文法規則は、SUBJ, V, OBJ の結び付きを平等に見なす構造を与える。これは以下に述べる理由により、意味解析上好ましい構造である。

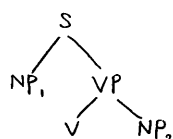
ところがこの規則は、

$$\begin{cases} S \rightarrow NP \ VP \\ VP \rightarrow V \ NP \end{cases}$$

の2つに分解して考える言語学者の考え方と一致しない。[田中82]述べた様に筆者等は、構文解析木の構造に沿った意味解



(a)



(b)

析を行う方式を採用している。この意味析の方式を用いるとすれば、(b)の構造は、VとNP₂との結び付き、すなわち動詞と目的語との結び付きを重視しすぎた構造であり、我々が文を先頭の語から順に受容して理解して行く過程が素直に表現できない。

特殊な仕かけを用いぬ限り、(b)の木の前は動詞と目的語の意味解析が行われてから、主語とVPとの意味的な結び付きが計算される構造になっているからである。(a)の木の形では、SUBJ, V, OBJは平等であり、こうした問題は生じない。

ホービーの文型パターンの利用と日本語訳語の選択に関しては、不定詞の用法(副詞的、形容詞的用法など)については、かなり利用できるのではないかと考えて、そのためのプログラム作りを進めている。これは、次の例を見れば理解できるだろう[ホービー 81]。

He stopped to rest [VP4A]

He awoke to find the house on fire [VP4B]

He agreed to come at once [VP4C]

He seemed to be surprised at the news [VP4D]

He appeared to have left already [VP4E]

He forgot to post your letter [VP7A]

He brought his brother to see me [VP16A]

He wants me to stay [VP17A]

He wishes me to stay [VP17B]

5. 強制翻訳モード

2. で述べた様に、素直な解釈を行う

とすれば *The door opens the door* は意味的に異常な文であり、図1の記述により、それが検出できる。ところがこの文も、こぢっけをしたり、特殊な文脈を与えれば理解可能である。そこで、これを翻訳不能として片付けるのではなく、何とかして訳を見い出す努力をすることも必要であろう。我々は、意味解析に失敗すると強制翻訳モードに入る。この場合には、syntactic な情報のみを用いて訳文を作り出す。先の例文からは、強制翻訳モードにより、“そのドアはそのドアを開ける”という訳文が出力される。

ここで syntactic な情報のみを用いるということの意味は、*<unsatisfied-slot>*

<slot-name>

<constraint>・<prep>・<action>

で[田中他82]、*<slot-name>*と*<prep>*のみを用いてユニット同会話を行うということである。*<constraint>*は用いないのがある。

6. おわりに

訳文抽出に際し、[高松(他)82]は変換規則を幾つか設定する興味ある方式を提案している。融合方式によればそれらの幾つかは、辞書項目の記述として解決できる。変換規則で対処すべきか、辞書項目の記述で対処すべきかは、現在の言語理論でも論争の種になっている。

最後に付録として、翻訳結果を示す。人間の介入はまったくないことに注意。

参考文献

- [田中他82] 意味表現用言語SRLの機械翻訳への応用—融合方式の提案、情報処理学会NL研究会、1982・5月。
- [田中 79] 計算機による自然言語の意味処理に関する研究、電総研研究報告 No.797, 1979。
- [高松(他)81] 動詞格パターンと格構造に基づく英日機械翻訳、信学論誌D, 9, 1981, 815-822。
- [高松(他)82] 英日翻訳における格構造変換、情報処理学会25回全国大会, 6K-6, 1982。
- [様垣 75] 日英比較表現論、大修館、1975。
- [ホービー 81] 英語の型と語法、オックスフォード大学出版局、1981。

- 1 \$ I TAKE A COLD. 私は風邪をひく。 13 \$ I HAVE A DANCE. 私は踊る。 16 \$ HE IS A BOY. 彼は少年である。
- 2 \$ I TAKE A BATH. 私は風呂に入る。 14 \$ I HAVE A DICTIONARY. 私は辞書を持っている。 17 \$ HE IS GAS. ヘリウムはガスである。
- 3 \$ I TAKE A BUS. 私はバスに乗る。 15 \$ I HAVE A COLD IN THE COLD. 私はその寒さのなかで風邪をひいている。
- 4 \$ I TAKE A COLD IN THE COLD. 私はその寒さのなかで風邪をひく。 18 \$ THE DOOR OPENS. そのドアは開く。
- 5 \$ I TAKE A DICTIONARY. 私は辞書を取る。 19 \$ ALL DOORS AND WINDOWS OPEN. すべてのドアと窓が開く。
- 6 \$ I TAKE MY DICTIONARY TO HER. 私は彼女のところへ私の辞書を持って行く。 20 \$ ALL OF DOORS AND WINDOWS OPEN. ドアと窓のすべてが開く。
- 7 \$ I TAKE MY CHILD TO HER. 私は彼女のところへ私の子供を連れていく。 21 \$ I OPEN THE DOOR. 私はそのドアを開ける。
- 8 \$ I TAKE MY DICTIONARY TO HER WITH HIM. 私は彼と彼女のところへ私の辞書を持って行く。 22 \$ I OPEN THE DOOR WITH A KEY. 私は鍵でそのドアを開ける。私は鍵のついたドアを開ける。
- 9 \$ I TAKE MY CHILD TO THE BUS-STOP WITH HER. 私は彼女とそのバス停へ私の子供を連れていく。 23 \$ I OPEN THE DOOR WITH HER. 私は彼女とそのドアを開ける。
- 10 \$ I TAKE MY DICTIONARY TO THE BUS-STOP WITH HER. 私は彼女とそのバス停へ私の辞書を持って行く。
- 11 \$ I TAKE A CAR, A BUS AND A PLANE. 私は車とバスと飛行機に乗る。 24 \$ SULPHURIC-ACID EATS IRON. 硫酸は鉄を侵す。
- 12 \$ I TAKE A CAR, BUS AND PLANE. 私は車とバスと飛行機に乗る。 25 \$ HE EATS THE COLDEST FRIED-CHICKEN. 彼は最も冷たい鶏の唐揚げを食べる。

26 \$ THIS IS A FILM WHICH HE DEVELOPS.

これは彼が現像するフィルムである。

27 \$ THE SYSTEM WHICH HE DEVELOPS IS A PARSER.

彼が開発するシステムはパーザーである。

28 \$ THE FILM DEVELOPS.

そのフィルムは現像される。

29 \$ THIS IS A SYSTEM WHICH HE DEVELOPS.

これは彼が開発するシステムである。

31 \$ I PRESENT MY BOOK TO HER.

私は彼女に私の本を贈る。

32 \$ I PRESENT MY CHILD TO HER.

私は彼女に私の子供を紹介する。

33 \$ I PRESENT A PAPER TO HIM.

私は彼に論文を提出する。

34 \$ I PRESENT AN OVERVIEW OF THE SYSTEM.

私はそのシステムの概観を述べる。

30 \$ THIS IS AN EXPERIMENTAL MACHINE-TRANSLATION-SYSTEM DEVELOPED AT ETL.

これは電子技術総合研究所で開発されている実験的な機械翻訳システムである。

35 \$ EXTENDED-LINGOL IS A PARSER USED IN AN ARTIFICIAL-INTELLIGENCE-SYSTEM
\$ FOR INTERPRETING DIALOGUES.

拡張し I N G O L は対話を解釈するために人工知能システムで使われているパーザーである。

拡張し I N G O L は対話を解釈するための人工知能システムで使われているパーザーである。

36 \$ THIS PAPER PRESENTS AN EXPLANATORY OVERVIEW OF A LARGE AND COMPLEX GRAMMAR, DIAGRAM,
\$ WHICH IS USED IN A COMPUTER-SYSTEM FOR INTERPRETING ENGLISH DIALOGUES.

この論文では大きくて複雑な文法すなわち D I A G R A M の (この D I A G R A M は英語の対話を解釈するためにコンピュータシステムで使われている) 説明的な概観を述べる。

この論文では大きくて複雑な文法すなわち D I A G R A M の (この D I A G R A M は英語の対話を解釈するためのコンピュータシステムで使われている) 説明的な概観を述べる。

(1) \$ WATASI GA ANI WO MITA.

I LOOKED AT A BROTHER .

(2) \$ WATASI WO ANI GA MIRU.

A BROTHER LOOKS AT ME .

(3) \$ WATASI WA IE KARA GAKKOU E ITTA.

I WENT TO A SCHOOL FROM A HOUSE .

(4) \$ GAKKOU KARA ANI GA IE E IKU.

A BROTHER GOES TO A HOUSE FROM A SCHOOL .

(5) \$ ANI GA WATASI WO IWAU.

A BROTHER CONGRATULATES ME .

(6) \$ ANI GA TANZYOUBI WO IWAU.

A BROTHER CELEBRATES A BIRTHDAY .

(7) \$ ANI GA KONYAKUSITA WATASI WO IWAU.

A BROTHER CONGRATULATES ME WHO WAS ENGAGED .
SOMEONE CONGRATULATES ME WHOSE BROTHER WAS ENGAGED .
SOMEONE CONGRATULATES ME TO WHOM A BROTHER WAS ENGAGED .

(8) \$ WATASI GA KONYAKUSURU IMOUTO WO IWAU.

I CONGRATULATE A SISTER WHO IS ENGAGED .
SOMEONE CONGRATULATES A SISTER TO WHOM I AM ENGAGED .

(9) \$ WATASI WA KAGI DE TO WO HIRAKU.

I OPEN A DOOR WITH A KEY .

(10) \$ KAGI DE TO WO HIRAKU.

A DOOR OPENS WITH A KEY .

(11) \$ TO GA HIRAKU.

A DOOR OPENS .

(12) \$ TO GA KAGI DE HIRAKU.

A DOOR OPENS WITH A KEY .