

英日機械翻訳における補文の変換と生成

鈴木 雅実 橋本 和夫 野垣内 出 榎 博史
(KDD 研究所)

1. はじめに

異なる2言語間の翻訳を行なう際に問題となる点の一つに、補文表現の変換の問題がある。補文の表現形式については、英語と日本語で差異が見られるが、その間の変換を行なって適切な翻訳文を生成するために考えるべき点を挙げ、その検討を行なう。また筆者等の試作英日機械翻訳システム VALANTINE-EJ上での実現状況を説明する。

2. 補文とその形式

一般に、文の構成要素である主語や目的語の位置に文(節・句)が埋め込まれている状況を補文という。英語と日本語とは、その形式上の特徴に次のような違いがある。

英語 補文:

統語カテゴリーと 統語マーカによる分類

- ・節(clause)補文 that-clause, wh-clause
- ・不定詞句補文 to-不定詞, 原形不定詞
- ・動名詞句補文 所有格マーカの有無
- ・現在文詞句補文
- ・過去文詞句補文

日本語補文:

接続形式と補文標識による分類

- ・形式名詞による名詞化 -こと, -の
- ・引用表現 ~と, ~よう(に)
- ・「~て」接続
- ・助動詞文 使役, 希望, 被害の受身等

表2.1 英語補文の例

- 1e) He admitted that he was wrong.
- 2e) They saw her leave the house.
- 3e) She believes it to have been a fact.
- 4e) We can't understand your coming back.
- 5e) I want this work finished by Monday.
- 6e) They made me do it.

注) アンダーライン部分が補文,

表2.1, 2.2に、それぞれ英語と日本語の対応する補文の例を示す。

このように英語・日本語において広く用いられる補文表現の形式と意味については、幾つかの理論的考察が提出されている[1~5]。本稿では、機械翻訳の立場から、英語の補文表現を適切な日本語表現に変換する方法について検討を試みたい。

英語補文表現を日本語に翻訳・変換する際に問題となるのは次のような点である。

A. 英語補文の認定と構造の確定

英文の解析時に、補文構造を正しく認識して然るべき中間表現を得なければならない。ここで問題なのは、述語がどの種類の補文を取り得るかということと、その局面において正しい構造を把握する方法である。

B. 補文構造のトランスファーに関わる問題

英語の補文表現を日本語にトランスファーする時に、構造をそのまま保持して変換するだけでは不完全な場合がある。それは補文内部の要素と外部との相互作用の違いが影響していることによる。

C. 日本語補文生成における適切な補文形式の選択

日本語中間表現から実際に翻訳文を生成する時に、適切な補文形式を得るためには、種々の条件チェックを必要とする。本稿では、このうち補文標識の選択と使役文の生成の項目を具体的に扱う。

表2.2 日本語補文の例

- 1j) 彼は自分が誤っていたことを認めた。
- 2j) 彼らは彼女が家を出て行くのを見た。
- 3j) 彼女はそれを事実だったと信じている。
- 4j) あなたが戻って来ることが理解できない。
- 5j) この仕事が月曜日までに終わって欲しい。
- 6j) 彼らは私にそれをやらせた。

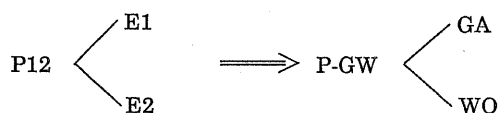
太字が統語マーカ・補文標識

筆者等が開発中の、結合価に基づく機械翻訳システムVALANTINEでは、述語成分と6種類の結合価成分の組み合わせによる依存構造AVSを中間表現として用いている。E1~E6の各結合価成分(ロール)は、名詞句、形容詞等の他、述語のタイプによる選択制限の下で、各種の補文構造をとることが許されている。その例を以下に示す。

表2.3 結合価ロールが補文となる例

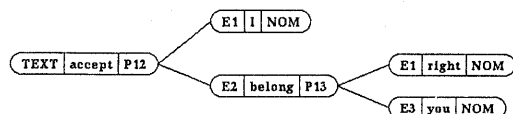
補文の結合価 ロール	例文(イタリック部分が補文)
E1	<i>To know everything</i> is impossible.
E2	I believe <i>you will come tonight</i> .
E3	He persuaded me <i>that it was true</i> .
E4	John gave <i>who was present</i> beer.
E5	You can call me <i>whatever you like</i> .
E6	John seems <i>what he is</i> .

また、日本語AVSでは、が・に・を等の格助詞の標準形(表層の変形を受ける以前の抽象的な格標識)が、結合価成分のカテゴリーを指示している。従ってE-AVSの述語-結合価成分のパターンは、変換辞書内の規則によってトランスファーを受ける。例えば



ここで、E1またはE2が補文であった場合、それに対応するGAまたはWOの成分は、英語側から内部の統語情報を受け取るとともに、それ自身がP-の述語タイプを持つことになる。

図2.1 E-AVS 例



この場合の主述語P-GW および補文中の述語P-~には、変換辞書で指定された、述語の意味カテゴリ-情報(英日共通)が付加される。このようなトランスファーが行なわれるE-AVSとJ-AVSの対応例を図2.1と2.2に示す。原文は次の通りである。

7) I accept that the right belongs to you.

補足になるが、AVSの各ノードは、表示部分以外にも各種の統語的・意味的な内部情報をフレーム形式で持ち、それらはトランスファー・生成を通じて、適宜参照される。

3. 英語補文の認定と構造の確定

鈴木 他[6]でも例に挙げた通り、expect, promiseという2つの動詞は、その後に名詞句とto不定詞を伴って、表層上は同じ統語パターンをとりながら、次の2つの文を比較すれば分かるように、意味上は異なる構造に解釈される。

8) I expect you to call on her.

9) I promise you to call on her.

8)では、下線部aの補文が expect に対してE2となっているのに比べ、9)では、promise に対して下線部bがE4, 下線部cがE2に該当する。この違いは、補文中の述語 call の意味上の主語(E1)が、両者で異なっていることによる。従って、上の例では、述語(動詞)固有の性質を踏まえた文の解釈によって、意味構造の違いを反映させなければならない。そこで、補文を結合価成分としてとり得る動詞の場合、補文の形式の統語的な特徴の他に、結合価パターンを表わすコードを予め付与しておくことが望ましい。

図2.2 J-AVS 例

英語には、前の例文8), 9)に掲げたような、名詞句+to不定詞が後に続く動詞だけでもかなりの数が見られ、使用頻度も高い。Hornby [7]の動詞パターンによると、to不定詞部分が、副詞的な意味を表わすVP16Aの動詞、目的補語に相当するVP25の動詞を別にすると、このパターンはすべてVP17として扱われている。不定詞部分を名詞句(前置詞句)に置き替えた構文が成立し、その構文との意味的等価性が保たれるか等のチェックをすることにより、VP17の動詞パターンに対しては、少なくとも4種類の結合価コードが対応することが分かる(Emons [8]による)。

このことは、より深い意味解釈を行なう時に、言語表現とそれが指す内容との対応関係を明確にするという点で重要である。即ち、格文法で言うところの深層格ラベルのようなものを用いるにしても、どの文要素に対して格を割当てればよいかという判断を経なければならない。辻井・山梨[9]によれば、格の判定も一種の認知プロセスであり、視点(基準の設定)によって変化するものである。また、補文全体に対して名詞句と同様な深層格を与えること自体、議論が分かれるところであろう。先の例に限らず、補文表現については、深層の意味関係と実際の表現形式の間に複雑な対応が存在するので注意を要する。

現在のVALANTINEシステムでは、いわゆる任意格については意味的な役割を指定しているが(可能な場合)、結合価成分(必須補足語)については、すでに紹介した通り、統語構造を抽象化した結合価ロールを使用している。機械翻訳の立場から言えば、利用できる情報が多いに越したことはないのであるが、できるだけ不確定な要素は取り除いてトランスファーしたいという要求もある。となると、語彙に依存した規則を重視することになるが、効率良く翻訳を行なうための工夫として、語彙のグループ化(clustering)をしておく必要がある。そのために我々は、種々の用例をもとに、英々辞典の語釈義文等を参考にしながら、語彙の特性を記述する作業を試みている。その方法としては、中核となる意味と示差的な特徴に分けて分析を行なう、一種の成分分析のような手順をとっている。

後の5章で触れることになるが、日本語における補文の形式を決定するのに、主文の述語の意味的特性が関与することが知られている。ただ、問題は

次章に関連することであるが、形式自体も意味を持ち得るということであり、そこから生まれる微妙な意味の違いを抽出することは非常に困難である。このような領域については、まだ機械翻訳を前提とした規則化の及ばないところであろう。

4. 補文構造のトランスファーに関わる問題

英語でも日本語でも、補文はそれを支配する主文に支えられて、文中で一定の役割を担っているが、その形式と意味の關係に両言語間のギャップがあるように思われる。本章では、統語的な性質によって英語の補文を2種類のタイプに分け、各々について機械翻訳上の問題を検討する

英語の補文表現のうち、that-clause に代表される節タイプの補文と、to不定詞に代表される句タイプの補文では、次のような違いがある。

節タイプの補文: 述語が定形となっており、テンス、モーダリティ等が明確になっている。補文の述語が支配する結合価成分も、単文中と統語的特徴が変わらない。

句タイプの補文: 述語が不定形となっており、テンス、モーダリティ等がexplicitに与えられていない。補文の述語が支配する結合価成分は、主文と膠着して統語的に変形を受ける場合がある。

日本語補文を英語補文との対比で捉えようと、引用表現等で用いられる、助詞「と」でマークされる補文が英語の節タイプの補文に近く、使役や願望等の助動詞が用いられる補文は句タイプの補文に近いと思われるが、一致しない部分もある。また、形式名詞が補文標識となる補文は、両タイプの中間的な性格を持っている。

節タイプの補文は、英日それぞれの結合価成分の対応付けを行なうことにより、比較的容易にトランスファーができるが、形式上の問題として典型的なのは、時制の一致である。これについては、特にここでは論じないが、トランスファーまたは生成時にテンスを変換することにより、日本語の補文形式を整えることができる。

これに対し句タイプの補文は、種々の要因から意味構造を抽出すること自体が難しい場合があることと、その意味を日本語にトランスファーしようとする時に、構造以上のものを付加しなければならないこともあること等から、節タイプの補文よりも難しい問題を抱えている。

先ず第一の要因は、主文の述語が補文の述語に対して持つ含意が、節タイプの補文のように明示されないことである。次の3つの例文は同じ意味内容を表わしていると思われるが、10aの that-clause 補文で explicit に表わされていることが、10b, 10c の動名詞・不定詞句補文では表面的には示されていない。

10a) I propose that we should start early.

10b) I propose starting early.

10c) I propose to start early.

この propose のように、補文内容に対する主体の意思が関与する述語について、同様の現象が見られる。日本語では、上のような含意を明示しないで、補文標識「こと」で形式名詞化することもできる。しかし、それだけでは不自然な場合、述語の意味カテゴリーによって、それが支配する補文の述語に対する含意を間接的に付加することも可能である。そこで VALANTINE-EJ では、トランスファー辞書の中で英日共通に指示される述語の意味カテゴリーを条件にして、生成段階で補文の述語にモーダリティ等を付加することもできる枠組みを採っている。

次に第二の要因として、主文の述語とはあまり関係がなく、補文の内部または外部の名詞句と補文の述語との意味的關係によって、補文全体の意味が定まる現象が挙げられる。端的な例は、BE + to 不定詞の構文で、形式的な主述語 be 自身には copula としての機能しか無く、文の主語と to 不定詞の意味的な繋がりによって、補文は、予定される事柄を示したり、義務を表わしたり、様々な意味を持つ。この曖昧性の解消には、恐らく静的な言語知識だけでは不十分で、談話に係る動的な知識が必要であろう。この点については、文を理解するための背景となる知識の記述に関する研究成果を待たねばならない。

さらに第三の要因は、前頁で述べたように、補文自体の独立性が節タイプの補文に比して弱いため、統語的な変形を受ける要素が存在することである。これについては統語的に明らかになっている部分も多い(助詞の交替等)。ただし、言語使用の常として、個人による許容度の差により、規則化が難しいものもある。このうち、次章では使役文の生成方法について述べることにする。

5. 日本語補文生成における適切な補文形式の選択

本章では、中間表現から日本語補文を生成する際の問題点のうち、補文標識の付加と使役文における助詞の生成に関わる項目について検討したい。

5.1 適切な補文標識の選択

日本語の格助詞「を」または「に」が実際に使用される場合、その直前に入る補足語(結合価成分)は、単純名詞句か、形式名詞によって名詞化された補文でなければならない。例えば、

11a) 彼は私に再検討を約束した。

11b) 彼は私にその問題を再び検討することを約束した。

ところが、名詞句で表わしていたものを、補文化しようとする、必ずしも形式名詞の付加が適切であるとは限らない。

12a) 私は彼女に原稿の清書を頼んだ。

NP1

12b) 私は彼女に原稿を清書することを頼んだ。

CS2

12c) 私は彼女に原稿を清書するよう(に)頼んだ。

CS3

上の例の場合、12b)と12c)では微妙な違いがあるが(この場合、12cを default とする)、補文CS2、CS3の役割は、aの名詞句NP1と等しいと考えられる。従って、これらの文を抽象化した中間表現のレベルでは、述語「頼む」に対して、NP1, CS2, CS3を同一の補足成分として扱うことができる。この補足成分を意味的に厳密に定義しようとすると、「対象格」のような曖昧な格ラベルでは不足で、「依頼内容を表わす行為」とでも決めることになる。

これに対し、VALANTINE システムでは、「WO」という抽象的な結合価カテゴリーを設定しているので、NP1, CS2, CS3とも結合価コード P-GNW の述語「頼む」に対するロール「WO」が与えられる。この中間表現から表層文を生成する時に、名詞句ならば格助詞「を」を付け(default)、補文ならば、主述語「頼む」の意味カテゴリーをチェックした上で、後に「よう」を付けるという指示がなされる。実際の規則は次のように、助詞生成規則と補文標識選択規則から成っている。

助詞生成規則は、日本語結合価ロール GA, WO, NI等をもつ補足語に対して、実際にどの助詞を付けて生成するか(助詞が省かれる場合も有り得る)を指定する規則である。個々の規則の条件部には、補足語自身の統語的な特徴の他、上位の述語の結合価および、意味的特性(意味カテゴリー, アスペクト等を含む)の組み合わせからなる条件が書かれている。また実行部には、補文標識の有無と、生成すべき助詞が指示されている。表5.1に結合価ロール WOからの助詞生成規則の一部を示す(注を参照)。

表5.1 結合価ロールWOからの助詞生成規則

```

if voice[superior] = passive
  → (RULE GA) (SIGN)
if feature[superior] > +causative, infl[]
  → (JOSHI nil)
if feature[superior] > +instruction, infl[]
  → (JOSHI nil) (SIGN)
if feature[superior] > +wish +aspect, infl[]
  → (JOSHI nil)
if feature[superior] > +wish +attitude
  → (JOSHI が) (SIGN)
if syntax-node[] = wh-clause
  → (JOSHI か)
if marker[] = what
  → (JOSHI か) (SIGN)
else
  (JOSHI を) (SIGN)

```

注) feature: 意味カテゴリー
 +instruction: 勧告動詞(頼む, 薦める等)
 +wish: 願望を表わす(たい等)
 +aspect: アスペクト動詞(始める等)
 +attitude: 態度を表わす(好きだ等)
 syntax-node: 統語カテゴリー
 marker: 統語マーカ
 infl: 活用をもつ
 [superior]: 上位ノード, []: 自分自身
 =: 左側の値は右側に等しい
 >: 左側の値は右側のどれかである
 ,: AND結合
 (RULE X): 結合価ロールXの規則を参照
 (JOSHI x): 助詞xを生成せよ
 (SIGN): 述語タイプならば補文標識を付けよ

一方、補文標識選択規則の方は、助詞生成規則で補文標識有りにマークされたものについて参照される規則で、やはり幾つかの場合分けがある。規則の条件部では、補文の上位の述語の意味カテゴリーや、補文自身の統語カテゴリー等がチェックされ、それに従って実行部で補文標識が選択される。表5.2にその概略を示す。

表5.2 補文標識選択規則

```

if feature[superior] > +perception +affect;
  syntax-node[] = wh-clause
  → (FR の)
if feature[superior] > +instruction
  → (FR よう)
if marker[] = what
  → (FR べき)
if marker[] = how
  → (FR 方法)
else
  (FR こと)

```

注) +perception: 知覚動詞(見る, 感じる等)
 +affect: 影響動詞(助ける, 妨げる等)
 ;: OR結合
 (FR z): 後に補文標識 z を付けよ

以下に、表5.1および5.2の規則を用いて生成された翻訳結果を原文とともに例示する。

- 13e) It was clarified that the trouble of the pressure bulkhead caused the accident.
 13j) 圧力隔壁のトラブルが事故を引き起こしたことが明らかにされた。
 (補文は受動態の述語に対してロールWOである)
- 14e) I heard her repeat the announcement.
 14j) 私は彼女がアナウンスを繰り返すのを聞いた。
 (+perception の例)
- 15e) The CCITT recommends the carriers to adopt the following standards.
 15j) CCITTはキャリアに以下の規格を採用するよう勧告する。
 (+instruction の例)

16e) The committee finished reporting the result of the investigation.

16j) 委員会は調査の結果を報告し終えた。

(+aspect の例)

17e) They can't think what to do next.

17j) 彼らは次に何をすべきか考えることができない。

(「べき」を補文標識として補っている)

18e) I will tell you how to use this machine translation system.

18j) 私はあなたにこの機械翻訳システムを使う方法を教えるつもりだ。

(ここでは「方法」をあえて補文標識として訳出している)

5.2 使役文の変換と生成

英語では make, let 等の使役動詞と原形不定詞を用いて、また日本語では使役の助動詞を用いて使役文が作られる。使役文の意味構造については、英語と日本語で異なる形式を仮定することもできるが、VALANTINEでは、次のように使役動詞(助動詞)を英語・日本語とも使役文の主述語(結合価 = 2)として扱っている。言語学的な根拠の他、トランスファーの容易さを勘案したのがその理由である。

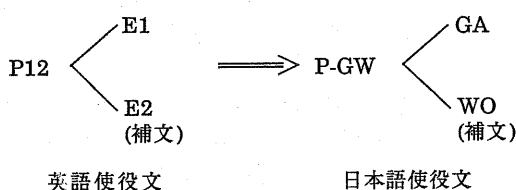
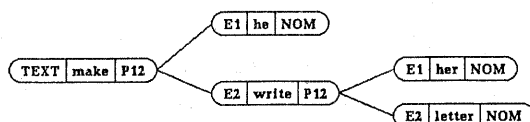


図5.2a 使役文の E-AVS



実際のトランスファーを次の例文について説明する。

19) He made her write the letter.

この英文の解析結果として得られるのが、図5.2aのE-AVSである。それがトランスファーされて、図5.2bのJ-AVSとなる。

さて、この後の問題は、J-AVSから翻訳文を生成するに当たって、補文中の名詞句に対する格助詞を適切に変換することである。4章でも述べたように日本語で助動詞を用いる構文では、補文の意味上の主語に付く格助詞は、外部からの影響を受けるので、それを考慮した生成を行わなければならない。使役文では、補文の意味上の主語には「に」または「を」が付くことになるが、その決定にはいわゆる表層フィルターの現象が関係するので、その条件が必要となる。[10]

VALANTINE-EJでは、この場合、5.1と同様に結合価ロールGAからの助詞生成規則(表5.3)が使用される。

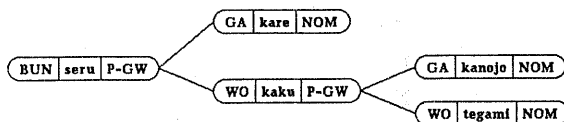
表5.3 結合価ロールGAからの助詞生成規則
(使役に関係する部分のみ)

```

if mode[superior] > +resultative
→
  if type[superior] > P-G P-GN
  → (JOSHI を) (SIGN)
  else
    (JOSHI に) (SIGN)
  
```

注) mode: 構文的特性
+resultative: 使役の対象となる述語の意、E-AVSを作る時点で与えられる。
type: 結合価成分のタイプ

図5.2b 使役文の J-AVS



さらに、補文の述語自体は、表5.1の結合価ロールWOからの助詞生成規則によって、上位の述語が使役を表わす(featureとして+causativeをもつ)ことから、助詞が省かれる。この場合は未然形で後に接続するのであるが、接続関係規則は別途参照される。

以上の生成規則を用いて生成された、使役文の例を次に示す。

20e) They made her walk slowly.

20j) 彼らは彼女をゆっくり歩かせた。

21e) I will make the people believe that the defendant is innocent.

21j) 私は被告が無実だと人々に信じさせるつもりだ。

22e) You should not let him depend on his father.

22j) あなたは彼を彼の父親に頼らせるべきでない。

5.1では補文標識の選択について、また5.2では使役文の生成について、VALANTINEシステムでの規則化の方向を示すことをねらいとした。即ち結合価構造を中間表現とした場合の、補文の変換と生成に関する定式化を、具体的に述べた訳である。これらの規則によって、基本的な述語のかなりの部分についてdefaultの処理が行なえることが分かったが、条件を排他的に決められないものもあるので評価は難しい。また、さらに多くの例文の処理を通じて、規則の改良を図る予定である。

6. まとめ

本稿では、英語の補文表現から日本語の補文表現への変換に関する問題点の整理を試み、結合価に基づく機械翻訳の立場から、幾つかの検討すべき点を指摘した。また、その一部については、現在行っている規則化の状況を紹介した。

今回は、主に統語論に近い所を扱ったので、補文形式と談話の関連について立ち入る余裕はなかったが、この問題は文を理解した上での翻訳に欠かせないテーマであり、新たな角度から探究を進めたい。

謝辞

本稿を執筆するに当たっては、日本IR^(株)の若菜氏、井倉氏との討論が大いに参考になった。また、規則をシステムにインプリメントする上で、^(株)SCCの向平氏、田中氏にはたいへんお世話になった。ここで厚くお礼申し上げる。

最後に、日頃ご指導頂くKDD研究所の鍛冶所長、野坂副所長、樽松次長、武田第一特別研究室長に感謝する。

参考文献

- [1] 久野. 1973. 日本文法研究. 大修館書店
- [2] Josephs, L. S. 1976. Complementation. *Syntax and Semantics 5: Japanese Generative Grammar*, ed. by Shibatani, M., New York: Academic Press
- [3] 柴谷方良. 1978. 日本語の分析, 大修館書店
- [4] 國廣哲弥 編. 1980. 日英語比較講座 第2巻 文法, 大修館書店
- [5] 村田勇三郎. 1982. 機能英文法. 大修館書店
- [6] 鈴木 他. 1985. 結合価に基づく機械翻訳の試み. 自然言語処理研究会 49-2
- [7] Hornby, A. S. 伊藤健三 訳. 1977. 英語の型と語法, オックスフォード大学出版局
- [8] Emons, R. 1974. *Valenzen englischer Prädikatsverben*, Tübingen: Niemeyer Verlag
- [9] 辻井・山梨. 1985. 格とその認定基準. 自然言語処理研究会 52-3
- [10] 柴谷 他. 1982. 言語の構造 意味・統語篇, くろしお出版