

談話処理における 日本語ゼロ代名詞の扱いについて

吉本啓
(NTT基礎研究所)

日本語においては、必須格成分が文中に陽に表現されていなくとも理解可能なことが多い。この一般にゼロ代名詞とよばれる現象について、表現されていない情報を顕在化させるための方法を述べる。日本語のテキストが主題を軸とする構造をなすことに注目して、これを手掛かりとしてゼロ代名詞の同定が行えることを示す。また複合文の中の主語がゼロ代名詞化される場合、陽に表現されたほうの主語と指示対象が同一かどうかは接続詞の種類および主題の有無により決定されることを述べる。後者の考えを日英翻訳システムに応用して、分析の有効性を確かめる。

Study of Japanese Zero Pronouns in Discourse Processing

Kei YOSHIMOTO

Basic Research Laboratories, Nippon Telegram and Telephone Corporation

Zero pronouns, i.e. obligatory cases not expressed explicitly in sentences, play an important role in Japanese anaphora system. Topic structure in the discourse is shown to be useful in identifying what zero pronouns refer to. Syntactically, types of conjunctives and existence or nonexistence of a topic provide a key to the identification of the zero pronoun with the subject expressed explicitly in a complex sentence. The latter idea is verified by being applied to a Japanese-English translation system.

1. はじめに

ここで「ゼロ代名詞」と呼ぶのは、日本語の単位文の根幹をなす格構造の中の必須格成分のうち、陽に表現されていないが文脈および発話の状況・知識により理解可能なもののことである。ただし、関係節のように補文中に格成分が欠けていることが統辞論的に義務的で、しかも同一複合文中の他の語句としてあらわれる場合を除く。

ゼロ代名詞英訳の過程は次の3つに分かれる。

- ①日本語動詞の必須格成分が欠けていることを知る。
- ②欠けている必須格成分が何であるかを文脈・発話の状況・知識により理解し、当該の文の意味構造を作り出す。
- ③意味構造にもとづき、英語の規則に従って適切な英文として生成する。

①は解析用辞書に蓄えられた、動詞・形容詞の必須格成分についての情報と現実の文との照合により行われる。何を必須格成分とするかは問題だが、ここでは解決ずみのことと仮定する。③について、当該の格成分を英語のどの照応表現で表すかも問題である。日本語のゼロ代名詞は英語の人称代名詞 (he, she, it, they) にほぼ対応することのみを指摘して、詳しくは他の機会に論じる。ここでは②を中心に考察する。

2. 指示対象理解のプロセス

日本語のゼロ代名詞の指示対象を理解するための手掛かりを統辞論的・意味論的・談話文法的要因に分けて検討する。

2.1. 統辞論的要因

英語の人称代名詞の解釈が先行 (precede) と統御 (command) という統辞論的要因によって制約されることはすでに論じられており(1)、これを日本語に応用しようとする試みもある。英語の人称代名詞に対応すると認められる日本語のゼロ代名詞 (訳語としての「彼」「彼女」は十分に代名詞的性格をもつものとは認められない) がこれらの制約に従っているかどうかの議論は煩雑になるのでここでは論じない。ただ、先行と統御だけではゼロ代名詞にかかわる照応関係が説明でき

ない例が存在することを指摘するにとどめる。日本語では、先行詞がゼロ代名詞を統御するか否かに関わらず、先行詞が主語または主題であるという事実が照応関係の成立にとって重要である(2) (先行詞とはテキスト中にあらわれる、ゼロ代名詞と指示対象を同じくする語句。ここではゼロ代名詞に後続するものも含める。主語は、格助詞「ガ」が付加された名詞句。主題は、係助詞「ハ」が付加された名詞句)。実際、ゼロ代名詞の解釈にあたっては、ゼロ代名詞と同一複合文中の主語または主題、あるいはそのあらわれる直前の文の主語または主題を先行詞として解釈しようとする強い傾向が見られる。

このことは実際のテキストの調査によっても確かめられた。土居健郎「「甘え」の構造」(1971年、弘文堂)の最初の100文にあらわれるゼロ代名詞を全て調査したところ、先行詞を有するゼロ代名詞138例のうち、先行詞が主語または主題の少なくともどちらかであるものは117例で、85%にのぼった。先行詞をもつ例のうち108例(78%)の先行詞がゼロ代名詞と同一文中にあらわれ、29例(21%)がゼロ代名詞のあらわれる文の直前の文に出現した。以上のことは「ゼロ代名詞のあらわれる文と同一か、またはその直前の文の主語または主題を優先的に先行詞の候補とする」という簡単な方略に基づくアルゴリズムにより大部分の先行詞が決定できることを示す。

複合文中にゼロ代名詞があらわれる場合、従属文をしめくくる語(接続助詞の類。関係節を受ける主名詞も含む。以後、単に「接続詞」と呼ぶ)の種類によって、主語・主題との同定の正否が左右される(3,4)。これはゼロ代名詞の同定にあたって、重要な手掛かりとなる。

従属文が「ナガラ」「ママ」「タリ」「ツツ」(これをA類とする)といった接続詞によりしめくられる時は、従属文の主語は、それが主題化されるか否に関わらず、つねに主文の主語と一致する。しかも、2つの主語のうちどちらかは必ずゼロ代名詞化されなければならない。従って、ゼロ代名詞である従属文または主文の主語の先行詞は主文または従属文の主語となる。

1. むギターを弾きながら太郎_iは次郎をはらはらとして見ている。

2. 太郎_iはその噂を次郎から聞いたまま_i話した。

従属文が「タラ」「バ」「ト」といった接続詞によりしめくくられる時、および関係節となる時(これをB類とする)には、従属文または主文の主語が主題化されている時は、ゼロ代名詞化された主文または従属文の主語は、陽に表現された従属文または主文の主語と同じでなければならない(例文3、4)。

3. 太郎_iは上着を脱ぐと_i洋服掛けに掛けた。

4. 次郎_jはスイッチをつけても_jテレビを見ていなかった。

5. 太郎_iが上着を脱ぐと_{i,j}洋服掛けに掛けた。

6. 次郎_jがスイッチをつけても_jテレビを見ていなかった。

3では主文「洋服掛けに掛けた」の主語は従属文と同じ「太郎」だが、5では同じとの解釈はむずかしく(限定された状況で現象文—文全体が新しい情報を導入する文—としてなら許されるが)、「太郎」と異なる指示対象と解するのが一般的である。例文4と6についても同じ関係が成り立つ(6では主文・従属文の主語が同じとの解釈はますます難しくなる)。なお、関係節により従属文を形成する場合も同じことが成り立つので、接続助詞ではないが、これもB類に含める。

第三に従属文が「ノニ」「ノデ」「ガ」の類(C類とする)によりしめくくられる時は、主文と従属文の主語・主題の一致に特に制約はない。

7. 母親はしきりに男の子を押し出そうとするが、男の子は今にも泣き出しそうだ。

8. 母はその点、楽天的な性格なので、_i何事にも勇氣と夢がありました。

7のように接続詞の前後で異なる主題を取ることでもできるし、8のように従属文から主文(あるいはその逆)へ主題がひき続き係っていくこともできる。

以上に挙げたA~Cは、日本語の従属文の3つの類型である。A類の従属文においては、主語は

従属文内部に含まれることはなく、本来主文に係るのだと解される。この意味でA類は従属文というより従属句であり、内部にゼロ代名詞を認めることは定義と矛盾する。しかしここでは他との対比の必要上、また解析の便宜上、ゼロ代名詞として扱う。B類の従属文の中には主語を含むことはできるが主題は含まれず主文にはみ出してしまふのだと解することができる。C類の従属文は独立文と同じ性質をもっていて、その内部にあらわれる「ハ」「ガ」の係り方はディスコース的要因による。

以上を、他の接続詞の例も加えて表1に示す。

	例	ゼロ代名詞が陽に表現された主語と一致するか
A類	ナガラ(継続)、ツツ、ママ、タリ、テ(状態)、連用形反復、形容詞・形容動詞連用形(状態)	一致
B類	タラ、ナラ、バ、ト、テモ、マデ、テ(継起)、テ(理由・原因)、テカラ、マデ、ナガラ(逆接)、連用形(継起)、ズズニ、トタン、関係節主名詞(トキ、コト、ノを含む)	主題の場合一致、非主題の場合不定
C類	ノニ、ガ、ケレド、ノデ、カラ、タメ、シ、テ(中止)、連用形(中止)	ディスコース的要因による

表1. 接続詞の種類とその特徴

2.2. 意味論的要因

意味論的要因はゼロ代名詞の先行詞の候補が適切かどうかをチェックするのに役立つ。格成分と動詞との間の意味的適合性、さらにスクリプト的知識もこれに含まれる。2.3および2.4で詳しく述べる。

2.3. 談話文法的要因

ゼロ代名詞が何を指示するかの理解は、究極的には談話文法的要因にもとづいてなされる。とりわけ、「主題」の概念によってほとんどの場合が解決できる。

2.3.1. 主題とゼロ代名詞

主題(あるいはトピック)とは、テキストの中でそれについて語られる話題のことである。テ

クストは主題を軸として進展する。主題は日本語においては係助詞「ハ」によって示される。主題は次のような構造をなしてテキストの中にあられる。

- (1) 1つの主題に対して、つねに1つの解説(コメント)があらわれる。1つの解説には複数の述語・文を含みうる。
- (2) 主題 m に対する解説 m 中に主題 m_1 , 主題 m_2 , ... 主題 m_n があらわれうる。これらは主題に対し副主題(サブトピック)の関係にある。
- (3) テキストの進展につれて、主題 m にとって代わる主題 $m+1$, 主題 $m+2$ 以下があらわれる。

(2)によって、主題はテキストにおいて再帰的構造をとることになる。次を例として、主題のあらわれとゼロ代名詞の関係を見ていく。

9. 1千葉市のある海岸に白鳥が四羽姿を見せました。2一羽は成鳥ですが三羽はまだ灰色の生毛に包まれた幼鳥で、 \emptyset 親子のようです。3親鳥はときどき羽ばたきをして \emptyset 周囲を警戒しています。4幼鳥はトラックが近くを走っても騒音を気にする様子はありません。5 \emptyset 干潟で好物の水草を食べています。

この例文の主題-解説、さらに副主題の関係を簡略に示すと図1のようになる。まず、このテキストの中で、ゼロ代名詞化されているのはすべて主題であることが注目される(これは2.1に述べたことと一致する)。

最初の文(以下、「文1」のように呼ぶ)のように主題を示す「ハ」をもたず、「(姿を)見せる」のように出現の動詞、あるいは「アル」のような存在の動詞をもつ文(国語学でいう「現象文」)は主語(この場合「白鳥」)をテキストへと導入し、以下のテキストの主題として確立されるべきことを示す。(以下、文を逐次的に理解するにつれて、図1のような、主題を軸とする意味ネットワークを記憶内に作っていくものと考え)こうして、文2以下には最上位の主題として(陽には表現されていないが)「白鳥」がある。それゆえ、文2の「一羽」「三羽」は「白鳥」の副主題で、それぞれ「白鳥のうちの一羽」「白鳥のうち三羽」と解すべきである。文2の最後の単位文中のゼロ代名詞を、陽にあらわれる最も近

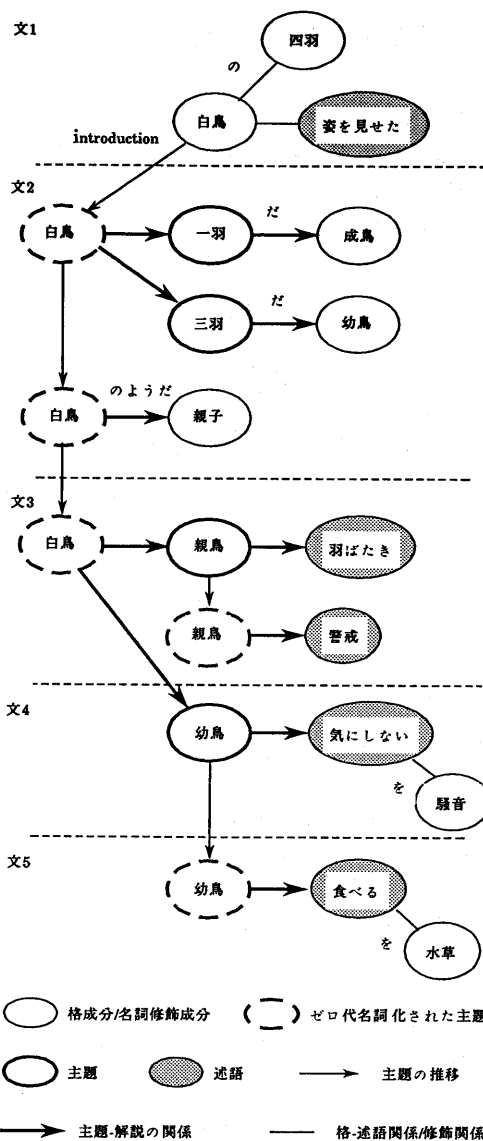


図1. 例9における主題の推移

い主題である「三羽」でなく「(四羽の)白鳥」と正しく理解するためには、以上の主題-副主題の構造の理解と、主題に対する解説の意味関係のチェック(「幼鳥」である「三羽」と「親子のようです」とは矛盾する)を正しく行うことが必要である。また、文3-5のゼロ代名詞には、それぞれの時点の副主題である「親鳥」「幼鳥」が各々補われる。以上のような、主題を軸とする意味モ

デルに基づいてゼロ代名詞の同定を行うには次の点が問題となる。

- ①主題-解説間の意味関係のチェック
- ②主題から副主題への推移の把握
- ③主題から他の主題への推移の把握

①は2.2で「意味論的要因」と呼んだもので、主題になる名詞の属性(それまでの解説の内容も属性として付け加えられるものとする)と解説になる述語の属性との無矛盾性のチェックを行う。

②の主題と副主題の関係としては、例9の「白鳥」と「幼鳥」のような<集合-要素>の外に、<全体-部分>(eg. その鳥は…羽は)、<原因-結果>(eg. 火事は…原因は)、<実体-属性>(eg. 部屋は…広さは)、<格成分-動作名詞>(eg. 選手は…動きは)、<動作名詞-格成分>(eg. 話し合いは…相手は)がある。いずれにしても、主題との間に(少なくとも話し手と聞き手にとって)十分な意味的連関が認められることが副主題出現の条件である。この意味的連関を用いて、主題が他の主題に推移する時、それらが主題-副主題の関係にあることを理解する。

③の主題から他の主題への推移には②のような制限はない。しかしテキストは意味的に連続していなければならないので、首尾一貫性(coherence)の観点からの制限が可能である。

②と③の区別が図1のように主題-副主題の推移を構造的に理解する上で重要である。これができれば、ゼロ代名詞の同定は容易である。ゼロ代名詞が認められた時、それを現行の、主題-副主題関係で最下位の主題(文2の \emptyset では「三羽」)に置き換える。①の意味関係チェックを行い、矛盾が発見されれば1ステップ上の主題を候補とする。矛盾のない主題を見つけるまでこれを繰り返す。

2.3.2. 感情移入-重要度とゼロ代名詞

以上の主題の進展に注目する手法によってゼロ代名詞の同定のほとんどは解決される。しかし、主題とならなくても、次のような場合にはゼロ代名詞化が行われることがある。

第一に、感情移入の程度の強い名詞句はゼロ代名詞となりやすい。感情移入の程度の最も強いのは話し手自身、その次は聞き手であるので、

これらはよくゼロ代名詞化される。感情移入にかかわる動詞(行く/来ル、ヤル/クレル/モラウ)および補助動詞(テイク/テクル、テヤル/テクレル/テモラウ)の場合は、主語や間接目的語がゼロ代名詞化されていても、話し手または話し手に身近な者か、あるいはまたそれ以外の者のどちらかに決まっているので容易に同定できる。

10. 太郎が[私に、私の弟に、etc.]お金をくれた。

第二に、テキストの中で重要と判断される名詞はゼロ代名詞化されやすい。

11. ジャガイモを洗って皮をむき、 \emptyset やわらかくなるまで \emptyset ゆでなさい。

上の例ではジャガイモをどうするかが問題になっているので同定が容易である。

以上の2つは広義の主題と解することもできる。いずれにせよ、主題、感情移入、重要度の3つの概念の整理が望まれる。

2.4.3つの要因の間の関係

以上では、ゼロ代名詞同定のための手掛かりを統辞論的要因・意味論的要因・談話文法的要因の3つに分けて考察した。そのうち決め手となるのは談話文法的要因である。このことを例文5を例として見てみる。

5. 太郎_iが上着を脱ぐと \emptyset _{i,j}洋服掛けに掛けた。

2.1で述べたように、上の文だけが与えられた場合の解釈は、ゼロ代名詞の指示対象は「太郎」とはべつのも、というのが普通である。とりわけ次のように2つの文が連続すると、

5'. 花子_iが着替えを手伝った。太郎_jが上着を脱ぐと \emptyset _i洋服掛けに掛けた。

ゼロ代名詞は「花子」との読みが優勢である。これは前の文の主語が「花子」であるという統辞論的・談話文法的理由に加えて、<女が男の着替えを手伝う>というスクリプトの知識が一般に認められていることによる。しかしまた、

5". 花子_iが着替えを手伝った。太郎_jが上着を脱ぐと、何洋服掛けに掛けた。いつもものぐさな太郎_jがめずらしいことをする、と花子_iは思った。

のように続くとゼロ代名詞の指示対象は「太郎」となる。これは、2番目の文が現象文との解釈を受けること、2番目の文と3番目の文とが首尾一貫すること(ものぐさなことと洋服掛けに服を掛けることとは相互矛盾の関係にあり、「めずらしいこと」に属する)による。

テキストの理解は究極的には首尾一貫性(coherence)が保たれるようになされる。2.1で述べた統辞論的要因は1ないし2文程度の狭い範囲の、2.3.1で述べた主題はそれよりもやや広い範囲のデフォルト的解釈を与える。それらはテキストの首尾一貫性と矛盾しないかぎりではゼロ代名詞の同定として正しく、またそうなることが実際多いが、もっとも上位にあってゼロ代名詞の同定を最終的に決定するのは首尾一貫性である。すなわち、我々は文を、それ以前の文に対して対比、原因、結果、推論、詳述、並列、例示、継起といった関係のどれかに当てはまるように理解し、ゼロ代名詞をはじめとする照応表現の同定もその範囲内で行おうとする(5)。この方面の研究が望まれる。しかし、統辞論的要因や主題だけを手掛かりとする方法でも大部分の場合について有効である。

3. 日英翻訳システム

3.1. 処理のあらまし

日本語から英語への翻訳の際に、2.1で述べた従属文の種類に応じてゼロ代名詞の先行詞を決定する実験システムを構成した。システムの全体は標準的なトランスファ方式に基づいている。解析時に、複合文中にゼロ代名詞が出現するのであれば、接続詞の種類と主題の有無を参照して、陽にあらわれない必須格成分(ここでは主として主語)を顕在化した意味構造を作り出す。処理の流れを図2に示す。

解析部で、格構造モデルを使って入力日本語文から日本語意味構造を得る。その際、述語としてあらわれる動詞の必須格成分が文中にあらわれなければゼロ代名詞と認定し、その部分のヌ

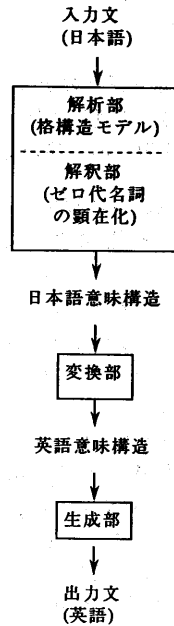


図2. 翻訳処理の流れ

ロットの値として dummy-noun-n を与えておく。解析部の後半部分をなす解釈部で、接続詞が A、B、C 類のどれか、B 類の場合はゼロ代名詞化されていないほうの主語が主題かどうかという 2 つのパラメータを参照して、dummy-noun-n に対し、その先行詞に関する情報を付け加える。パラメータの値と、ゼロ代名詞の代わりに顕在化される名詞句との関係を表2に示す。

接続詞	陽に表現された主語は主題か?	何を顕在化するか?
A類 (ナガラ, ママ)		陽に表現された主語
B類 (ト, テモ)	主題	陽に表現された主語
	非主題	不定
C類 (ガ, ノテ)		ディスコース的要因による

表2. ゼロ代名詞を顕在化するパラメータ

例文1の解析が終わった段階の日本語意味構造を図3に示す。主文の主語がゼロ代名詞であるので、意味構造の対応する部分には dummy-noun-1

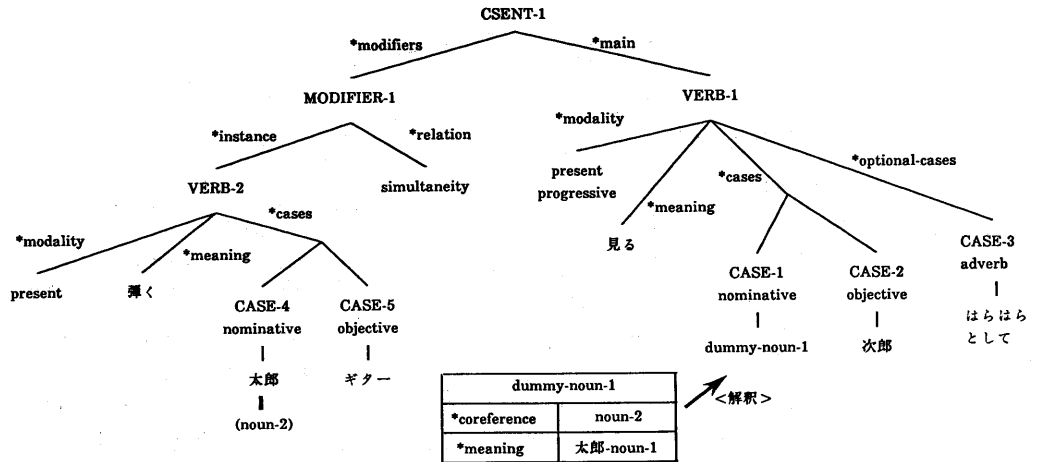


図3. 例文1に対する日本語意味構造

が入っている。解析部の中の、解釈部に入る直前の段階では、品詞・意味・修飾成分など訳文生成に必要な情報は一切入っていない。

例文1. 太郎がギターを弾きながら \emptyset 次郎をはらはらとして見ている。

解析部の中の解釈部で dummy-noun-1 に対し、その指示対象が従属文の主語と同じ「太郎」という情報が付加される。変換部でこれは 'while VERB-ing' 構文に合わせるために削除されて英語意味構造を得る。それが生成部に引き渡され、次の訳文を生じる。

訳文1. Taro looks at Jiro anxiously while playing the guitar.

例文1の「ガ」を「ハ」に置き換えても同じ解析過程を経て同じ訳文を生じる。例文2についても同様の処理が行われ、訳文2を得る。

例文2. 太郎はその噂を次郎から聞いたまま \emptyset 話した。

訳文2. Taro told the rumor as he heard it from Jiro.

B類の接続助詞である「ト」については、陽に表現されるほうの主語が主題か非主題かで異なる解析処理を行う。

例文3-1の日本語意味構造は図4のようになる。

例文3-1. 太郎は上着を脱ぐと \emptyset 洋服掛けに掛けた。

意味構造の、主文の主語・目的語に相当する部分に各々 dummy-noun-2, dummy-noun-3 が入っている。解釈部においては、入力文中の主語が主題化されているので、これを dummy-noun-2 の指示対象とするという情報を dummy-noun-2 に付け加える。dummy-noun-3 に対しても、指示対象が従属文の目的語と同じだという情報を付加する。これが変換・生成部を経て次の訳文を得る。

訳文3-1. Taro took off his jacket and he hung it on the hook.

例文3-2の日本語意味構造は、従属文の動詞のスロット *topic に値が入っていない(これは例文3-2が非主題文であることに対応する)ことを除けば図4と同じである。

例文3-2. 太郎が上着を脱ぐと \emptyset 洋服掛けに掛けた。

この場合、解釈部を終えても、主文の主語に相当するゼロ代名詞 (dummy-noun-4 とする) に対して特に指示対象に関する情報が与えられていない点が例文3-1の場合と異なる。これは、入力文が非主題であるため、dummy-noun-4 の指示対象が

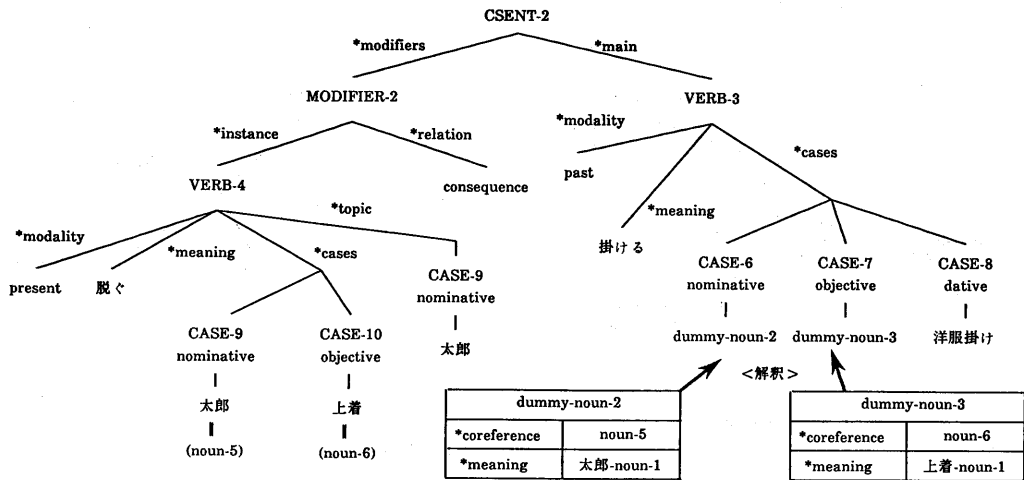


図4. 例文3-1に対する日本語意味構造

主文の主語「太郎」であるとの操作を解釈部において行わないからである。それゆえ、そのまま生成部に引き渡されれば、訳文は「Taro took off his jacket and NIL hung it on the hook.」となる。例文3-2の前の文に太郎以外の人物が主語または主題としてあらわれるなら、それを指示対象とすることになる。(2.1を参照のこと)たとえば

例文3-2'. 花子が着替えを手伝った。太郎が上着を脱ぐと洋服掛けに掛けた。

の文では、dummy-noun-4を「花子」を指示する代名詞に置き換えて、「Taro took off his jacket and she hung it on the hook.」の訳文を得る。

C類接続詞を伴う文については、A・B類について行ったようなゼロ代名詞の顕在化は一般的には行えない。例文4について見ると、主文の主語の位置のゼロ代名詞の先行詞を従属文の主語とする、ということではできない。

例文4. 彼女は先生に直接頼むので、 \emptyset とても困る。

しかし、この場合、主文の述語「困る」はたまたま感情をあらわす動詞であり、平叙文現在形で「ノグ」等の説明のムードを伴わず、「ガル」が後続しない形では主語は一人称単数に限られるという制約が日本語にはあるので、変換時にこれを補えば次のように適切に訳すことができる。

訳文4. I feel very annoyed because she asks the master directly.

4. むすび

本稿では日本語文の特徴であるゼロ代名詞についての情報をいかに顕在化するかについて、主題を主たる手掛かりとして述べ、その一部を日英翻訳に応用して分析の有効性を確かめた。ゼロ代名詞は照応表現の一種として談話処理の中心課題であり、また意味と直接に係わる問題であるので、その研究の意義は大きい。

<引用文献>

- (1) Reinhart, 1983, *Anaphora and Semantic Interpretation*. Croom Helm.
- (2) 佐藤, 1981, 「日英照応表現の対照研究」『英語学』24
- (3) 三上, 1970, 『文法小論集』くろしお出版
- (4) 南, 1974, 『現代日本語の構造』大修館
- (5) Hirst, 1981, *Anaphora in Natural Language Understanding*. Springer-Verlag.