特定研究「機械処理のための言語構造の言語間対照研究」
昭和63年度　研究発表会　要旨集

日時：昭和63年12月9日(金)　10時～17時10分
場所：国立国語研究所

1　語の意味用法の対照研究
（石綿敏雄・竝木崇康・田中茂範）
2　機械翻訳における訳語選択に関する研究
（田中穂積・徳永健伸）
3　語結合における文法的・意味的特徴に関する言語間対照
（村木新次郎）
4　法情報に関する日英語対照
（草薙裕・中石実・堀口純子・野田尚史）
5　機械翻訳前処理自動化のための基礎研究
（成田一・村田忠男）
6　自然言語の統語論および意味論の形式化と機械処理の試み
（白井英俊）
7　連語構造における意味素性の適合に関する言語間比較
（中道真木男）
8　英日機械翻訳のための会話文の文意構造記述に関する研究
（河口英二）

研究代表者　石綿　敏雄（茨城大学）

機械翻訳における訳語選択に関する研究

田中　穂積，徳永　健伸（東京工業大学工学部）

1. はじめに

機械翻訳において困難な問題の一つに訳語選択がある。適切な訳語選択を行うためには，文の統語構造を解析することだけでなく，意味解析，文脈解析を行う必要がある。現在これらの技術はまだ解決されたとは言い難く，様々な問題が山積している。そこで昭和６１年度は，訳語選択に関する実験を行うためのツールとして，Lang LABと呼ばれるシステムを開発した。Lang LABは論理型言語Prologをベースにしており自然言語解析用の様々な機能が組み込まれている。第２章ではLang LABの概要を説明する。

最近の機械翻訳の研究によれば，意味解析，文脈解析の基礎となる辞書に，どの様な知識をどの様な形式で蓄えておくかが訳語選択には重要であることが明らかにされている。そこで昭和６２年度は，訳語選択に特に重要であるとされている概念の上位／下位関係シソーラスを作成した。これについては第３章で説明する。

最終年度の昭和６３年度は，前年までに得られた成果を総合化する目的で，最も困難であると考えられている文脈を考慮した訳語選択アルゴリズムの研究を進めた。そして照応指示関係を決定することが訳語選択に重要であることを指摘し，Sidnerの焦点モデルが利用可能なことを見いだした。これを用いて指示代名詞を含む文の訳語選択を行うプログラムをLang LAB上に作成し，その有効性を検討した。この訳語選択の過程では概念の上位／下位関係シソーラスが使われる。文脈を考慮した訳語選択の計算モデルについては第４章で説明する。

2. 自然言語処理のためのソフトウエアシステム Lang LAB

2.1 Prologによる自然言語処理

ロジックプログラミングの枠組みと自然言語処理との整合性の良さは，フランスのマルセイユ大学のColmerauraの開発したMetamorphosis Grammar [Colmeraure 78] において初めて明確に示された。ColmerauraがMetamorphosis Grammarで示した基本的な考え方は次の様なものである。まず，拡張文脈自由文法の一形式であるMetamorphosis Grammarで書かれた文法規則を，それと一対一に対応するPrologプログラム（Horn節）に変換する。変換後のPrologプログラムに自然言語の文を与えて実行すると，トップダウンで縦型探索（depth first）による統語解析を行うことができる。

この方法によれば，従来，自然言語を統語解析する場合に作成しなければならなかったパーザの作成が不要になる。Prologのインタープリタに組み込みの機能をそのまま用いて，パーザを代用することが可能になるからである。Metamorphosis Grammarのもう一つの重要な特徴は，統語解析と意味解析とを融合させた自然言語解析が可能になるということである。Metamorphosis Grammarの形式に自然な拡張を施すことで，それが可能になるのであるが，これは認知心理学の観点からも好ましいと考えられる。

その後イギリスのエジンバラ大学のPereira等は，Metamorphosis Grammarの考え方をさらに洗練したDefinite Clause Grammar（DCG）を開発している[Pereira 80]。Pereira等のDCGもMetamorphosis Grammarと同様，統語解析と意味解析を融合したトップダウンで縦型探索の自然言語処理を行うPrologプログラムに変換されるが，この方式には一つの大きな問題がある。それは，左再帰的な文法規則があると，統語解析過程で無限ループに陥るという問題である。この問題はボトムアップな統語解析を行うことにより避けることができる。

電子技術総合研究所の松本等の開発した方法によると，DCGの形式で書かれた文法規則を，ほぼそれと一対一に対応するPrologプログラムに変換し，このPrologプログラムに自然言語の文を与えると統語解析と意味解析とを融合したボットアップで縦型探索の自然言語処理を行うことができる[Matsumoto 83]。これは，BUPシステムと呼ば

英題　Word Selection in Machine Translation

れている。その後東京工業大学の今野等は，BUPシステムに左外置変形が扱えるように拡張を施したBUP_XGと呼ばれているシステムを開発している[今野 86]。本稿で説明する自然言語処理システムLang LABは，このBUP_XGをベースにし，熟語処理機能を内臓した自然言語処理のためのソフトウエアシステムである[徳永 88]。

2.2では，DCGによる文法記述形式とLang LABシステムで採用されているDCG_XGと呼ばれる文法記述形式を説明する。Lang LABシステムにはBUP_XGと比較しておよそ6.5倍ほど高速な統語解析アルゴリズムが組み込まれている。2.3ではLang LABの実行速度を実験結果とともに示す。文献[松本 83]では，再度計算を避けるアルゴリズムを導入することによって，統語解析速度がおよそ1桁改善可能なことが報告されているから，再計算を避けるアルゴリズム導入以前の，最も初期のBUPシステムに比べてLang LABシステムは，統語解析に関しておよそ60倍以上の高速化がはかられていることになる。筆者はLang LABシステムによる統語解析については，十分満足できる速度が得られたと考えている。なおLang LABシステムを機能の面から見るとBUP_XGシステムのスーパーセットになっている。

2.2 文法記述形式DCGとDCG_XG

Lang LABシステムでは，文法規則の記述は，BUP_XGシステムで用いた形式，すなわちDCGに左外置が扱えるよう拡張を施したDCG_XGの形式を用いる。説明を簡単にするために，DCGに意味解析を行うプログラムが補強されていない最も単純な記述例を図2－1に示す。

```
(a1)  s --> np, vp.
(b1)  np --> pron.
(c1)  pron--> [yon].
(d1)  vp --> [walk].
```

図2－1　DCGによる文法規則と辞書の記述例

統語解析過程で意味解析を行いたければ，例えば以下に示すようにDCGのボディ中の任意の場所に，意味解析用のPrologプログラムを中括弧で括って挟み込めば良い。意味解析結果を憶えておくために，各非終端記号に対応する述語に引数をもたせておくこともできる。たとえば，

```
s(Ssem) --> np(NPsem), vp(VPsem),
    {seminterp(NPsem, VPsem, Ssem)}.
```

これは最終的に，次のPrologプログラムに変換され実行される。

```
s(Ssem, X, Z) :-np(NPsem, X, Y),
                vp(VPsem, Y, Z),
                seminterp(NPsem, VPsem,
                          Ssem).
```

但し述語seminterpは，NPsemとVPsemとを用いて意味解析した結果をSsemに返すものとする。

英語の関係代名詞節の埋め込み文は，平叙文中の名詞句が一つ欠落した構造をしているが，これは先行詞が関係代名詞の左方に移動してできたと考えられる。この様な語句の移動を左外置と呼んでいる。この場合，移動した跡には痕跡を残すと考え，この痕跡をシステムに発見させることを前提にして文法規則を記述すると，そうでない場合に比べて，文法規則の数や文法カテゴリーの数を減らすことが可能になり，文法の見通しが良くなる。

トップダウンの統語解析システムのATNG[Woods 71]やXG[Pereira 81]には，予測を利用した痕跡発見機構が組み込まれている。純粋にボトムアップの統語解析システムの場合には予測が行えないから，解析すべき文の単語と単語の間のほとんど全てに痕跡があるとして処理を進めなければならないので効率が悪い。幸い，我々のLang LABシステムははボトムアップを基本としているものの，トップダウンの予測をも利用している　今野はこの点に着目した痕跡発見機構を考案しており[今野 86]，それはLang LABシステムに組み込まれている。Lang LABにおける文法記述形式はDCG_XG　と呼ばれている。

DCG_XGによる文法記述はDCG　のスーパーセットになっており，次の記法が許されている：

```
np--> det, noun, srel../np.
```

記号"../"（スラッシュとよぶ）に後続するカテゴリをスラッシュ・カテゴリとよぶ。この記法はGPSGを参考にしたものである[Gazdar 82]。"srel../np"は，srelの中に痕跡となるスラッシュ・カテゴリnpが一つ存在することを表している。文法記述者は，DCG_XGの形式で文法を書きさえすれば，Lang LABシステムによって，自動的にスラッシュ・カテゴリと関係代名詞節中の痕跡との対応がとられる。

DCG_XGには，以下に示すような"〈"と"〉"で囲んだ記法がある。

```
np--> det, noun, 〈srel../np〉.
```

この場合，srel中の（スラッシュ・カテゴリnpに支配される）痕跡は，detとnounとからなる名詞句

以外とは対応付けができない。したがって英語の場合，複合名詞句制約に違反した非文法文の統語解析に失敗する。ここで〈，〉で囲んだ記法は，Pereira [Pereira 81] の open, close カテゴリで括ることに対応しているが，LangLAB システムで採用した今野らの DCG_XG の記法の方が簡潔であると思うがどうだろうか。

LangLAB には熟語を扱う機能が備えられている。記述例を示す。

(1) vp --> [kick, the, backets].

(2) np --> [not, only], np, [but, also], np.

(1)の記述では，動詞 kick が kicks, kicked, kicking のように変化する可能性があるが，これらは LangLAB により自動的に処理される。(1), (2)の記述とも通常の辞書の一部をなす。熟語用の辞書が出来るわけではない。複合語の記述例を次に示す。

(3) n --> [computer, system, manual].

(3)の記述では最後の manual が複数形に変化することがあるが，これも LangLAB により自動的に処理される。

最後に DCG_XG で記述された文法がどの様に実行されるかを図２－２に示す。LangLAB 使用者は DCG_XG の形式で文法と辞書を記述しトランスレータを起動すれば，それらが Prolog プログラムに変換され，このプログラムに解析すべき文を与えると，ボトムアップで縦型探索による自然言語処理を行うことが出来る。

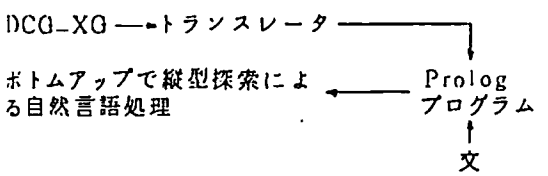

図2－2 LangLAB システムによる自然言語処理

2.3 LangLAB の実行速度

LangLAB の実行性能を評価するために，以下に示す例文の解析に要した時間（全ての解析結果を得るまでの時間）を示す。使用計算機は SUN3／260 で Quintus-Prolog を用いた。

(1) She was given more difficult books by her uncle.
解析結果＝14個，解析時間＝4.1 sec.

(2) Be careful not to break the vase when you put it in.
解析結果＝4個，解析時間＝1.3 sec.

(3) This paper presents an explanatory overview of a large and complex grammar that is used in a sentence.
解析結果＝4個，解析時間＝4.5 sec.

(4) The annotations provide important information for other parts of the system that interprets the expression in the context of a dialogue.
解析結果＝9個，解析時間＝14 sec.

(5) For every expression it analyzes, diagram provides an annotated description of the structural relations holding among its constituents.
解析結果＝2個，解析時間＝4.4 sec.

以上から自然言語処理ツールとして LangLAB により妥当な処理速度を得ていることが分かる。

3. 訳語選択と上位／上位関係シソーラス ISAMAP

適切な訳語選択を行うためには，文の統語構造を解析することだけではなく，意味解析，文脈解析を行う必要がある。現在これらの技術は解決されたとは言い難く様々な問題が山積している。最近の研究によれば，意味解析，文脈解析の基礎となる辞書に，どの様な知識をどの様な形式で蓄えておくかが，訳語選択には重要であることが明らかにされている。訳語選択には従来用いられてきた意味マーカ以上に微細な知識の分類体系が必要になることを以下で説明する[林 86][石綿 86][荻野 87]。

ソース言語のレベルでみると，語源が同じであるという理由で１つの辞書項目に納められているものが，ターゲット言語では意義が異なるとして，複数個の訳語が対応している場合の問題を，動詞 take の訳し分けを例にして考えてみよう[田中 87b]。

(1) I take a plane.

文(1)の場合には，take の目的語が「乗り物」であれば take を「乗る」と訳すという知識を用いることになるだろう。そのためには plane が「乗り物」であるという知識の利用，言い替えると plane の上位「乗り物」が位置するような上位／下位関係シソーラスを用いた推論が必要になる。我々は，訳語選択には意味マーカは充分ではないと考えている。訳し分けに当たって，意味マーカ以上に微細な概念が必要になることが多いからである。このことを見るために，次の(2)と(3)に含まれる take の訳し分けを考えてみよう。

(2) I take a cup of water.

(3) I take medicine.

文(2)の場合には，目的語が＋liquidという意味のマーカを持てばtakeを「飲む」と訳す，という知識を記述することは可能だろう。しかし文(3)については，目的語に貼るべき意味マーカが用意されていないとみるのが妥当であろう。そこで，目的語を直接指定して，takeの目的語がmedicineならtakeを「飲む」と訳すという翻訳用知識を組み込むことも考えられる（語直接指定方式）。ところが，目的語にはpill（－liquidという意味マーカを持つことに注意）とか薬品名がくるなど際限がない。以上の理由から語直接指定方式と意味マーカを用いた訳語選択方式は，いずれも好ましい方法であるとはいえない。

この問題を解決するためには，50前後の意見マーカ［長尾 85］による粗い意味分類ではなく，もう少し微細で大量の概念の分類体系が必要になる。文(2)と(3)の例では，目的語が「薬」の場合にtakeを「飲む」と訳す，という訳語選択用の一般的な知識と，medicineやpillの上位に「薬」が位置する上位／下位関係シソーラスを利用した推論を行い，適切な訳語選択を行うことができる。

これまで，大量の知識の分類は困難であるという理由から，意味マーカによる粗い意味分類と語直接指定方式とがとられてきた。しかし，より高度な機械翻訳システムを開発するためには，大量の知識を意味分類したシソーラスの作成が不可欠であろう［内田 88］。そのようなシソーラスの階層の上位には，意味マーカの体系が位置することになる［堺 88］，［田中 87a］。

訳語選択で用いる概念の微細さの目安になるもの

1－－－－－2－－－－－3－－－－－4－－－－－5－－－－－6－－－－－7－－－－－

具体物　　無生物《・物，・物体》

生産物　　$品物《・$物品，製品，》

具体的道具

・機械類

知能機械

人操縦物

情報記録伝達機器

工農具類

武器《・兵器》

生活用具

・家庭用具

・学習研究用具

遊具

・装身具

医療器具

情報伝達具

・灯火《$ライト》

紐類

針類

・棒竿

・目盛

部品

・衣類

紙類

布類

帽子

履物

・食べ物《・食品》

・飲物

・調味料

嗜好品

1 - - - - - 2 - - - - - 3 - - - - - 4 - - - - - 5 - - - - - 6 - - - - - 7 - - - - -

酒類
収納器具
・容器
収納具
・薬品類
$ 情報伝達媒体
建設物 ・建物類
・部屋
・施設 交通施設
・標識
象徴物
金券類
証明書類
生理現象物
気象物
・天体
原材料《・資源》
土石
・鉱物
・物資 ・石材
・木材
・鉄材
燃料
・紙類！
$ 皮革
・物資 ・元素
混合物・化合物
液体
気体
粘性体
粉体
農水産物
農作物！
畜産酪農物
水産物 ・魚介類！
海草！
$ 獲物 人間以外の動物！
・物理的実体
・生物 ・植物 ・樹木
・草
・花
農作物
・動物 人間！
人間以外の動物
・魚介類
・虫類

図３－１ ISAMAPの一部

はないだろうか。英語基本動詞辞典［小西 80］には，英語の基本動詞約350に対して結合価が与えられている。一つの動詞には一般に複数個の結合価が与えられており，各結合価毎に日本語の訳語が対応づけされている。結合価には，主語が「人」で目的語が「仕事」ならacceptを「引き受ける」と訳せ，といった知識が書かれている。したがって英語基本動詞辞典の結合価の記述で，主語や目的語に書かれている概念は，動詞の訳語選択に必要な概念についての手がかりを与える［石綿 75］。

我々は手作業で英語基本動詞辞典からそれらを抜きだし，968個の概念を得た。これらは英語の基本動詞を適当な日本語に翻訳するために必要な概念の微細さに対する一つの目安を与える。我々の開発した約4000の概念を，上位／下位関係に基づき階層化したシソーラスISAMAPの一部を図3－1に示すが，そこには英語基本動詞辞典から抽出された968個の概念が配置されている。ISAMAPの詳細は文献［山中 87］を参照されたい。

訳語選択には図3－1に示すようなシソーラスと語直接指示方式とを組み合わせる方法が有効だろう。ただし，語直接指示方式で与える訳語選択の知識は特殊なものに限るべきであり，それらは一般に熟語と呼ばれているものになるはずである。これらは次章の文脈解析も考慮した訳語選択アルゴリズムでも使われる。

図3－1で使われている記号を説明する。

(1) $\alpha\langle\beta, \gamma, \rangle$のように，〈 〉で囲まれた$\beta$, γなどは，記号〈の左の概念αとほぼ同義と考えられる概念である。

(2) $\alpha\langle\beta, \gamma, \rangle$のように，〈 〉で囲まれた$\beta$, γなどは，記号〈の左の概念αを属性とした時，その値であることを示している。

(3) 紙幅の関係で，上位／下位関係がインデンテーションをつけて表示できなかったものは，( )を用いて階層構造が示されている。なお，( )の中が(α, β(γ,),)の様に，「,)」で終了しているのは，まだ枚挙が完結していないことを意味している。

4. 文脈情報を用いた訳語選択

これまでに，実験システムを含め多くの機械翻訳システムが作成されているが，そのほとんどは，文を翻訳の単位とするシステムであり，文脈情報を使用したシステムは数少ない。数少ない文脈情報を利用した翻訳システムの例として電総研で開発されたCONTRAST［石崎 86］がある。CONTRASTでは，パラグラフ単位で翻訳する文章の解析をおこない，文章の内容を中間表現に変換し，そこから常識知識による推論を用いて目標言語を生成している。しかしながら，領域を特定しない常識知識を計算機上に実現するには多くの課題が残されている。また，そうした常識知識を用いた推論を現時点の計算機でおこなうには効率が悪いという問題もある。本章では，問題を代名詞の照応解決にしぼり，照応解決が訳語選択におよぼす影響について考察し，実際に英日翻訳の実験システムを構築することによってその効果を検証する。照応解決には，Sidnerの焦点モデルを用い，できるだけ浅い解析で照応を決定する［Sidner 83］。

4.1 照応と訳語選択

代名詞の照応解決が訳語選択におよぼす影響について具体例をあげて考察する。英語を日本語に翻訳する場合，代名詞は翻訳されずに省略されるか，その代名詞の指示している名詞が明示的に繰り返し用いられることが多い［Sidner 83］。これまでの文毎の翻訳を行うシステムでは，代名詞の翻訳は目標言語の対応する適当な代名詞に翻訳し，実際にそれが何を指すかは，出力された文章を読む人間の推論能力にまかせるしかなかった。たとえば，

1. (a) Taro bought a new camera.
（太郎は新しいカメラを買った。）
(b) He likes it.
（彼はそれが気に入っている。）

において，「彼」が太郎であり，「それ」が前文のカメラを指すことは，文単位の翻訳である限り認識できない。例文1では，he, itをそれぞれ，「彼」，「それ」と翻訳しても人間が読めば，「彼」，「それ」が何を指しているかが解るのでそれほど不自然な訳とはならない。しかしながら，代名詞が何を指示しているかによって，格関係が変わり，それにともない訳語が影響を受けることがある。たとえば

2. (a) Taro bought a new camera.
（太郎は新しいカメラを買った。）
(b) He asked Hanako to take a picture of him with it.
（彼はそれ（カメラ）で花子に写真を撮ってもらった。）
3. (a) Taro had a little dog.
（太郎は小さい犬を飼っていた。）
(b) He asked Hanako to take a picture of

him with it.
（彼はそれ（犬）と一緒に花子に写真を撮ってもらった。）

例文２ｂと例文３ｂはまったく同じだが，先行する文によってｉｔの指示する対象が異なり，with it が文中で果たす格関係が異なる。例文２ｂではｉｔがカメラを指示しているため，ｉｔは主動詞 take の道具格となるが，例文３ｂではｉｔが犬を指示しているので随伴格になる。この例では，指示対象によって格関係が異なるためその翻訳がまったく異なってしまう。このような場合は，代名詞ｉｔを単に「それ」と訳したのでは不十分で，照応関係を正しく解析しないと２(b)，３(b)のような訳し分けができない。

別の例をあげよう。

4. (a) Taro bought a film.
（太郎はフィルムを買った。）
(b) He took a picture.
（彼は写真を撮った。）
(c) He developed it.
（彼はそれ（フィルム）を現像した。）

例文４では４ｃのｉｔが何を指示しているかによって，主動詞 develop の訳語が異なる。先行する文脈からｉｔが指示するものが，camera であったら，この develop は「現像する」ではなく「開発する」と訳さなければならない。

Carter によれば［Carter 87］，このような代名詞の照応の多くは浅い解析で解決できる。文脈情報の利用の第１段階として照応解析を翻訳システムに組み込むことは有意義なことである。

4.2 照応解析を含む翻訳システム

本節では，Sidner の焦点モデルを用いて，照応解析をおこなう機構を組み込んだ翻訳システムについて述べる。

4.2.1 Sidner の焦点モデル

Sidner は焦点を用いて局所的な情報から照応を解決するモデルを提案している［Carter 87］［Sidner 83］。このモデルでは，話者が談話中で中心におく要素（焦点）は，照応詞の先行詞になりやすい，という仮定がなされている。Sidnerのモデルでは以下の６つのレジスタを用いて焦点の状態を表わす。

- DF（Discourse Focus），AF（Actor Focus）
  現在，焦点となっている要素がはいる
- PDF（Potential Discourse Focus），PAF（Potential Actor Focus）
  最後に読まれた文で焦点になれなかった要素がはいる
- DFS（Discourse Focus Stack），AFS（Actor Focus Stack）

焦点の移動が起こった時，古い焦点を積む

照応解析アルゴリズムは以下のようになる。

1. 最初の文に対して焦点候補をDFに設定する。
2. 焦点を使って照応を解決する。
3. 解決した照応により焦点の状態（各レジスタの値）を更新する。

１は最初の文にのみ適用され，２と３は文が読み込まれるたびに適用される。焦点候補は次の優先順位によって決定する。

1. 主語（be 動詞文法，there 挿入文）
2. それ以外の文法要素
   (a) 対象格
   (b) 動作主格を除く要素
   (c) 動作主格
   (d) 動詞句

なお，焦点になれなかった要素は PDF にて登録される。AF，PAF，AFS はレジスタの内容が動作主になりうる要素であるという点を除き，DF，PDF，DFS と同様に扱われる。

4.2.2 システムの概要

本節では Sidner の焦点モデルを組み込んだ実験的な機械翻訳システムについて述べる。本システムは Lang LAB をベースに実装されており，システムはすべて Prolog で記述されている。図４－１にシステムの処理の概要を示す。

入力された文章は Lang LAB 上で一文づつ統語解析，意味解析がおこなわれるが，同時に照応詞が含まれる場合は先行詞の決定もおこなう。照応詞の解決は照応詞の意味解析をおこなう時点で起動する。１文が解析されると前節で述べた Sidner のアルゴリスムを適用して焦点の管理をおこなう。解析の結果として文の意味を表す中間表現が得られる。この中間表現は概念のレベルに近いもので，この時点で概念レベルの曖昧さは解決されている。したがって，この段階で多義語の訳語選択の問題は解決されていることになる。この中間表現からトップダウンに出力文を生成する。

例文４について中間表現を得る様子を見てみよう。以下に，各文が解析された後の中間表現と焦点レジスタ　内容を示す。中間表現の中で＊がついているものは，照応解決によって確定したものであること

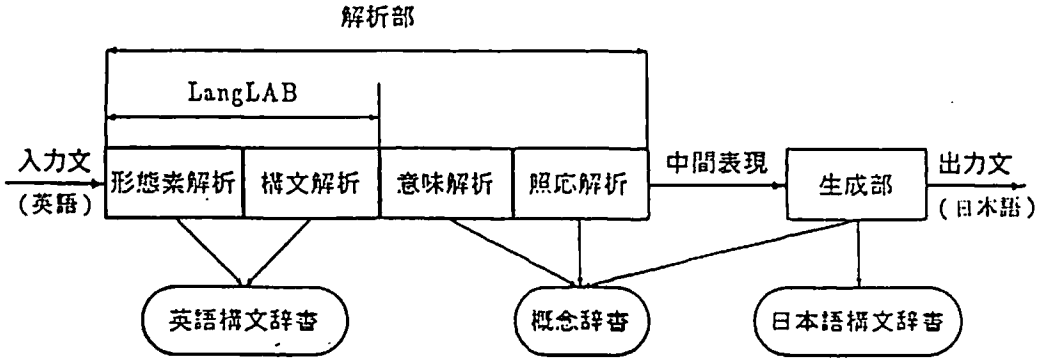

図４－１　システムの概要

を示している。

Taro bought a film.

DF：film(expected)　AF：taro(expected)
PDF：buy　PAF：[ ]
DFS：[ ]　AFS：[ ]
中間表現：[物を買う，AGENT：[太郎]，
OBJECT：[フイルム，SPEC：[不定]]]

He took a picture.

この文のHeは人間なのでAFをみて太郎であることを知り，太郎がAF上でconfirmされる。DFは更新されず前文のものがそのまま残る。

DF：film　AF：taro(confirmed)
PDF：picture, take　PAF：[ ]
DFS：　AFS：[ ]
中間表現：[(写真などを)撮る，AGENT：[・太郎]，OBJECT：[写真，SEPC：[不定]]]

この段階で目的語がpictureだから，takeが「撮る」，pictureが「写真」という概念に対応させられている。

He developed it.

ここでitの指すものが調べられるが，この例では前文のDFのトップにfilmがあり，itがそれを指すことが分かる。

DF：film(confirmed)　AF：taro
PDF：develop　PAF：[ ]
DFS：[ ]　AFS：[ ]

itがfilmであることを理解したのでシステムは動詞developを，「現像する」という概念に対応させ，次の中間表現を得る。

中間表現：[現像する，AGENT：[・太郎]，
OBJECT：[フィルム]]

この段階では，先に述べたように，中間表現の概念レベルで曖昧さが解消しているので，日本語の訳語選択は容易である。

4.3　考察

本章では正しい翻訳を得るためには１文内の情報だけでなく，文脈情報も利用する必要があることを指摘し，照応の解決と翻訳との関係およびSidnerの焦点モデルによる照応解析を組み込んだ機械翻訳の実験システムについて述べた。Sidnerの焦点モデルは，局所的な文脈情報に基づいて照応を決定するための処理が比較的浅いという利点があるが，反面，より大域的な情報を考慮することが必要なこともある。常識的なより深い推論を必要とする場合には本方式による照応解決には限界がある。たとえば，

5. (a) He took a picture.
(彼は写真を撮った。)
(b) He developed it.
(彼はそれ(フィルム)を現像した。)

例文５の場合(例文４と異なり)itの先行詞が文章中に明示的に現れないので，「写真を撮る」という動作に「フィルム」が関係し，「写真を撮ったら普通フィルムを現像する」という常識的な知識を用いて深い推論をおこなわないと照応を解決することができない。対象とする領域をあらかじめ限定して大域的な談話構造をモデル化する研究はあるが[Grotz 77][木下 88]領域を限定しないモデルについては今後の課題である[Schank 77]。

最後に，翻訳システムを考える場合，高品質の翻訳結果を得るためには，照応の解決だけでなく生成における照応詞の使用も重要な問題である。原言語で代名詞となっていたものを目標言語でも代名詞で表現しなければならないというわけではない。特に，日本語の場合は代名詞化より省略される場合が多い。人間が翻訳した文章などを調べ，原言語と目標言語の照応詞の使い方などを調査する必要がある[田中 84]。

参考文献

[Aitchison 87] Aitchison, J.: Words in the Mind — An Introduction to the Mental Lexicon, Basil Blackwell (1987).

[Carter 87] Carter,D.: Interpreting Anaphors in Natural Languge Texts, Ellis Horwood LTD. (1987).

[Colmeraure 78] Colmeraure, A.: Metamorphosis Grammar, in Bolc.(ed.): Natural Language Communication with Computers, Springer-Verlag, 133-190 (1978).

[Gazdar 82] Gazdar, G., Pullum, A. F.: Generalized Phrase Structure Grammar: A Theoretical Synopsis, Indiana University Linguistics Club, (1982).

[Grotz 77] Grotz, B.: The Representation and Use of Focus in a System for Understanding Dialogs. IJCAI '77, 67-76 (1977).

[林 66] 林 大：分類語彙表 国立国語研究所，秀英出版 (1966).

[石綿 75] 石綿敏雄：日本語の生成語彙論的記述と言語処理への応用，国立国語研究所報告54，電子計算機による国語研究Ⅶ，秀英出版 (1975).

[石綿 86] 石綿敏雄：日本語の特質，高橋延匡(他)：日本語情報処理，近代科学社，1-61 (1986).

[石崎 86] 石崎俊，井佐原均，横山晶一，内田ユリ子：文脈情報を用いた機械翻訳システム contrast の特徴，情報処理学会第32回全国大会，1599-1600(1986).

[情報処理振興事業協会 87] 計算機用日本語基本動詞辞書 IPAL (Basic Verbs) -解説編-，情報処理振興事業協会 (1987).

[木下 88] 木下聡，佐野洋，浮田一男，天野真家：文脈理解のための知識の表現と推論，Proc. of Programming '88, 205-212, ICOT, 1988.

[金田 81] 金田一京助，見坊豪紀，金田一春彦，柴田 武，山田忠雄：新明解国語辞典(第三版)，三省堂 (1981).

[国研 72] 国立国語研究所報告43：動詞の意味・用法の記述的研究，秀英出版 (1972).

[国研 73] 国立国語研究所報告44：形容詞の意味・用法の記述的研究，秀英出版 (1973).

[小西 80] 小西友七(編)：英語基本動詞辞典 研究社 (1980).

[今野 86] 今野聡，田中穂積：左外置を考慮したボトムアップ構文解析，日本ソフトウエア科学会編，コンピュータソフトウエア，3, 2, 115-125 (1986).

[Lakoff 80] Lakoff, G. and Johnson, M.: Metaphors We Live By, The University of Chicago Press (1980).

[Lenat 86] Lenat, D., Prakash, M. and Shepherd, M.,: CYC: Using Common Sence Knowledge to Overcome Britlleness and Knowledge Acquisition Bottenecks, The AI magazine, 6, 4, 65-85 (1986).

[Lyons 68] Lyons, J.: Introduction to Theoretical Linguistics, Cambridge University Press (1968).

[Matsumoto 83] Matsumoto, Y. et.al.: BUP-- A Bottom-up Parser Embedded in Prolog, New Generation Computing,1, 2, 145-158 (1983).

[McArthur 81] McArthur, T: Longman Lexicon of Contemporary English, Longman (1981).

[長尾 85] 長尾 真(他)：科学技術庁機械翻訳プロジェクトの概要，情報処理，26, 10, 1203-1213 (1985).

[大野 81] 大野 晋，浜西正人：角川類義語新辞典，角川書店 (1981).

[荻野 83] 荻野綱男：シソーラスについて，ソフトウェア文書のための日本語処理の研究-5，1-61，情報処理振興事業協会 (1983).

[荻野 87] 荻野孝野：日本語の意味分類試案，計量国語学会第31回大会発表資料 (1987).

[Pereira 81] Pereira, F.: Extraposition Grammar, American Journal of Computational Linguistics, 7, 4, 243-256 (1981).

[Pereira 80] Pereira, F. and Warren, D.: Definite Clause Grammar for Language Analysis--A Survey of the Formalism and a Comparison with Augmented Transition Networks, Journal of Artificial Intelligence, 13, 231-278 (1980).

[Roget 82] Roget's Thesaurus (New Edition) (First edition by Peter Mark Roget 1852) Longman(1982).

[堺 88] 堺　和宏：自然言語の意味処理のための辞書に関する研究．東京工業大学修士論文．1988

[Schank 77] Schank, R. C. and Abelson, R.P. : Scripts, Plans, Goals and Understanding, Hillsdale, N.J. : Erlbaum (1977).

[Sidner 83] Sidner, C. L. : Focusing in the Comprehension of Definite Anaphora. In M. Brady and R. C. Berwick, editors, Computational Models of Discourse, 267-330. MIT Press (1983).

[田中 84] 田中穂積（他）：より自然な翻訳へのアプローチ［1］… 英日翻訳における表現の対応，ICOT TR-073, ICOT (1984).

[田中 87] 田中穂積，仁科喜久子：上位下位関係シソーラスISAMAPの作成．，情報処理学会自然言語処理研究会，25-44, November (1987).

[田中 87b] 田中穂積：機械翻訳における訳語選択，日本語学，明治書院，49-55 (1987).

[徳永 88] 徳永健伸，岩上　真，上脇　正，田中穂積：自然言語解析システムLangLAB，情報処理学会論文誌，29, 7, 703-711 (1988).

[鶴丸 76] 鶴丸弘昭，藤田　毅，首藤公昭，吉田将：日本語の機械処理，電子通信学会オートマトンと言語研究会，AL76-43, 41-50 (1976).

[鶴丸 86] 鶴丸弘昭，日高　達，吉田　将：単語間の上位－下位関係の自動抽出，情報処理学会情報学基礎研究会報告，86, 3, 1-8 (1986).

[Woods 71] Woods, W. A. : Experimental Parsing System for Transition Network Grammar, in Rustin (ed.) : Natural Language Processing, Algorithmic Press (1971).

[内田 88] 内田裕士：電子化辞書の開発，［自然言語処理技術］シンポジウム論文集，89-98，情報処理学会 (1988).