

機械翻訳に関する1つのアプローチ

市山俊治 村木一幸

(日本電気(株) C&Cシステム研究所)

1. はじめに

本稿は、Prologを用いた英日機械翻訳実験について報告する。

近年機械翻訳に関する研究が再び活発化し、多くの機械翻訳実験システムが開発されている。

筆者等は、ACOS概説書マニュアル(英語版)を対象とした英日機械翻訳実験システムTRAP(TRANslation by Prolog)を開発した。機械翻訳にはジョイント方式・トランスファー方式の2方式があるが、TRAPではトランスファー方式を採用している。

トランスファー方式の機械翻訳システム開発では、解析手法、変換手法、生成手法、及び辞書構成法が問題となる。ここを処理の観点から見直すと、解析・変換・生成は本変形(換)操作という基本演算が繰返し適用されるが、METEOのQ-System¹⁾、Ariane 80のROBRA²⁾以外では、3つの処理は全く異なるシステム記述言語をインポートされているのが通常である。しかし、システムの統一性、資源の均質性を保つことは困難であり、ひいては開発管理のコスト増大を招く。

他方、Q-System、ROBRAでは、そうした非均質性を排除することに成功していると言える。しかし両システムは、本変換規則適用順を開発担当者が明示的に規定する必要があり、規則の再帰呼び出しが許されない。よって規則数が膨大になったり、制御の複雑さによって規則開発・管理が困難になる。

こうした欠点を排除すべく、TRAPにおいては、解析・変換・生成全てをPrologによって記述している。Prolog

を用いる利点は以下のものである。

- i) 表現力の豊かによる規則数爆発の回避。
- ii) 制御の簡潔さによる容易な管理
- iii) 可逆性を利用した双方向翻訳システム開発コストの減少

i), ii)は上述の問題点を解決し、iii)は、Prologを使うことにより新しく期待される効果である。これは、TRAPで日本語構文生成に、日本語解析モジュールを全く手直ししないうえに用いていることから明らかになる。すなわち1つの言語ペアに対する双方向翻訳システムを開発する際の大巾なコスト低下を計ることが出来る。

以下、TRAPシステムの概要を大筋にその基本的処理について説明する。

2. TRAPシステムの概要

TRAPの翻訳方式はトランスファー方式であり、現在は形態素解析は含まない。図1に示すように、英語文を解析する英語解析フェーズと、英語中間表現(EIL)を日本語中間表現(JIL)へ変換する変換フェーズと、JILから日本語を生成する生成フェーズからなる。各フェーズは、Prologプログラムとしての規則と辞書によって実現される。

変換は構文レベルでのトランスファーであり、各フェーズは各々の辞書内容によって関連付けられる。

また、各々のモジュールはプロセス名PARSE, TRAN, GENと呼び出され、最初にGENから出力される結果が翻訳文である。

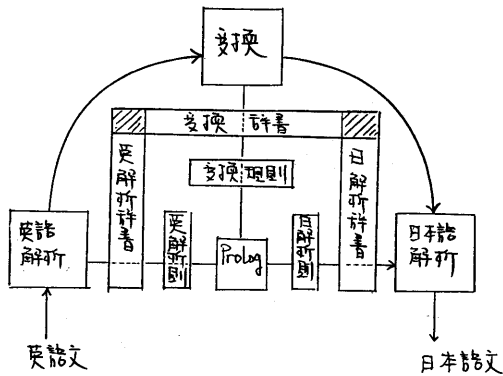


図1. TRAP 構成図

3. TRAP システムの基本動作

TRAP はその翻訳対象として ACOS6 General Description を採用している。本マニコアルは、日本語解説書を日本人が訳し、米人校正者が下巾着書き直しを行ったものであるが、日本人による英作文に交通に視られる特徴を残す。

- i) 語彙数、語の用法、文型のバリエーションが少なく、自然な制約付き英語となっている。
- ii) 関係詞節を用いた長文が頻繁に現れるが、構文的には固い。
- iii) 説明書の性格から平叙文に限られる。
- iv) 専門的用語が多く、多くはターム(専門熟語)となる。

22, Prolog による文法記述として、DCG¹⁾, Metamorphosis Grammar¹¹⁾ という CFG, CSG に対応するものが提案されている。TRAP の解析規則は DCG を基本としたが、その理由として、本二節と DCG の符合による扱い易さにある。つまり、DCG は Prolog の本二節と CFG の構文的・機能的(命題領域に於いて)同一性に基づいた文法記述手法であり、CFG の制約はそのまま受

け継ぐが、Prolog による記述能力を高めることは容易である。

つまり、TRAP において用いる制御変数の導入と、制御用述語の導入によつて、文脈処理が可能となる。解析の骨格を成す述語は下に示す4項述語である。

PRED(*A, *B, *C, *D, *E)

*A は入カストリング用ターム、*B はバッファ用ターム、*C は中間構造保持用ターム、*D, *E は制御用タームである。実際、*D は前文脈の継承、*E は後への文脈の伝達に使用される。こうした *D, *E 等は ATN パーサ⁹⁾ のレジスター、LINGOL パーサ¹⁰⁾ のメッセージに対応している。

辞書形式は上記述語と同じであり、

DIC([string*]*B], *B, *C, *D, *E)

上記の形式をもつ。こうした形式の利点は、string* によつて熟語登録が非常に簡単になるという点にある。

こうした形式の解析については第4節を詳述し、ここには、交換を含めた全体の流れについて説明する。

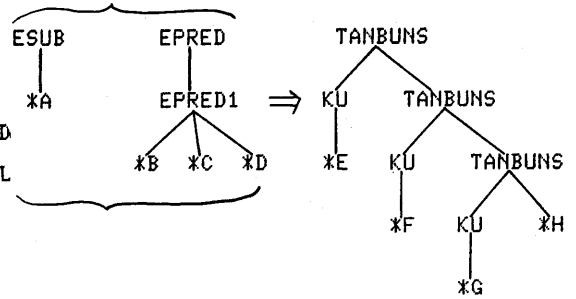
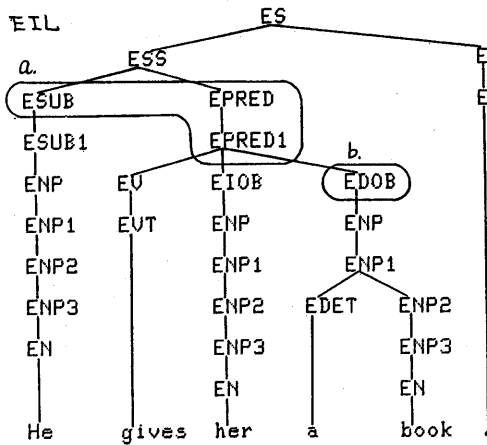
英語解析・日本語解析は、孤張 DCG 文法を用い、意味検証を同時に実行しながら英語内部表現(EIL)、日本語内部表現(JIL)を生成する。次項に示す図2は、入力英文「He gives her a book.」の翻訳例である。以下本図に従って説明する。

図2.aの入力文は英文解析モジュール PARSE に渡され、図2.bの EIL を出力する。EIL は TRAN に渡され、図2.cの JIL に変換される。JIL は日本語解析 GEN に入力され、図2.dの翻訳結果を出力する。

図2.bから図2.cへの変換は図3に

a. He gives her a book.

b. EIL



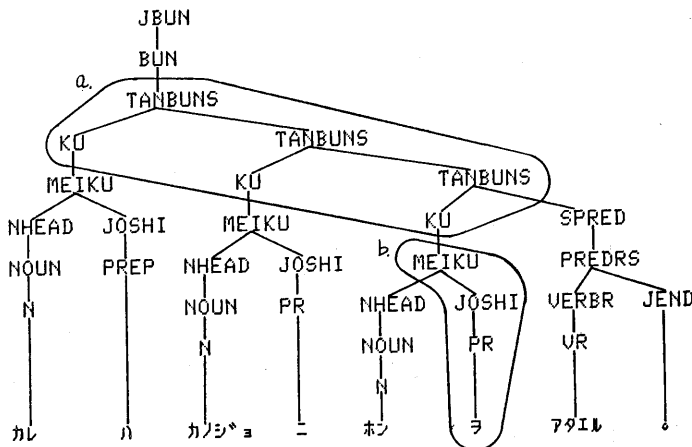
a. 変換規則

TBUN([ESUB,*A],[EPRED,[EPRED1,*B,*C,*D]],
[TANBUNS,[KU,*E],[TANBUNS,[KU,*F],
[TANBUNS,[KU,*G],*H]]])
:-TKU(*A,*E),TKU(*C,*F),TKU(*D,*G),
TSPRED(*B,*H).

b. Prolog規則

図3. 変換規則例

c. JIL



d. カレ ハ カノシヨ ニ ホンヲ アタイル。

e. He gives her book.

図2. 変換例

示すような変換規則をトップダウンに適用することで行なわれる。図2.bのEIL中のaを指定する構造から、図2.cのaを指定する構造への変換を行う規則が図3.aであり、そのProlog規則が図3.bに示される。

前述したように、可逆性を用いると図2.dをGENの入力として図2.eの出

力を得る。この場合、冠詞「a」の脱落が起こるが、その理由はEIL→JILの変換で、冠詞「a」の訳は無視されているからである。一般に、変換は多対1の写像とならざるを得ないため、逆変換は誤った変換が起こり得る。しかし、構文生成と解析との関係から、変換の入力、出力構造に十分な構文的情報を付加することによって前提すれば、可逆性を利用し、解析規則・辞書の生成への利用が可能となる。

4. 解析の意味の扱い

単に構文情報だけでなく文解析を行なえば、一般に各詞句の認識や係り受け関係に多くの曖昧を生じることが、その大半は簡単な意味検証によりみまされることが可能である。特にTRAPのような、トップダウン・デファースト・バックトラフバーズの制御方式を除くシステムでは早期に誤りを発見し、追

は代替解析を起動することが、大まか初歩向上につながる。

TRAPの意味検証は現在、以下の5つのタイプが用意されている。

1. 主・述の数的一致
2. 各々の動詞格構造
3. 名詞の意味カテゴリ
4. 係り受け意味カテゴリ
5. 冠詞・名詞の数的一致

TRAPを用いた意味情報の表現はフレーム構造を持ち Visé と呼ばれる。 Visé はデータベース項内文の意味検証、抽出とその有効性が実証されている。 Visé 上に定義される基本演算 MERGE, SELECT, RENAME 等を用いた意味検証を行う。

図4は、各語に付加される意味情報の例であり、各品詞の持つ意味の属性と取り得る値の例を図5に示す。

```
EN([SYSTEM:*A],*A,[EN,SYSTEM],*I,
    [CTY, SYS], [NUM, SIN], [CNT, THING])).
```

名詞辞書形式

```
EUT([PERMIT:*A],*A,[EUT,PERMIT],*I,
    [EACT,PERMIT], [NUM, SIN, PL, UNC],
    [SUB, SYS, HUM], [TENS, PR, TLES]),
    [17,[IOB, SYS, HUM], [DOB, ING, THING]],
    [14,[DOB, ING, THING], [ADVP, TO], [SYS, HUM]],
    [13,[DOB, ING, THING]],
    [23,[DOB, HUM, SYS], [COMP, TO-DO]])).
```

動詞辞書形式

図4. 辞書形式

名詞

TY(セマンティックカテゴリ): SYS, HUM, THING, ING, etc.

NUM(単複可算): SIN, PL, UNC

MNT(修飾可能な名詞のセマンティックカテゴリ) 値はTYと同じ

形容詞

MNT

冠詞類

NUM(数詞の単複不可算)

動詞類

NUM(主語単複不可算)

SUB(主語セマンティックカテゴリ)

TENS(時制): PR, PA, PP, ING, TLES

可能な文型(1~28)

目的語, 副詞句となるセマンティックカテゴリ

図5. 意味情報

基本演算について説明すると、MERGEは、1対1の意味情報中の属性に対し、

集合演算 JOIN に似た演算を施す。

JOIN との違いは、両方に同一属性名が存在しないが、一方の属性値が NIL の場合にも明らかとなる。例えば、

```
:-MERGE([ENUM, PL, SIN], [CTY, SYS],
        [ENUM, PL], [CMNT, HUM], *A).
```

に對して、

```
*A=[ENUM, PL], [CTY, SYS], [CMNT, HUM]
```

が得られる。 SELECT は指定された属性の属性値を返し、 RENAME は属性名変更を行う。

次に例をもとに意味検証について説明する。 次項図6、図7は、単文レベル、名詞句レベルでの意味検証を示したものである。

図6に於いては、まず主語 "the techniques" の意味 [NUM, PL], [SUB, SYS] が動詞句に伝えられ、主動詞 "permit" の意味中の [

NUM, PL, SIN, UNC], [SUB, SYS, HUM] との間で MERGE され、数的一致、セマン

ティックカテゴリの一一致がとられる。 次に "permit" 辞書中の "permit" が取り得る文型に従って規則を起動し、文型指示以下に記載される目的語、補語に関する意味制約を

文法カテゴリの一一致を EPRED に伝える。 以後、発見される名詞句等に対する制約を分配し、単文レベルでの意味検証が行われる。 本例では、"permit" に係る語と、各々の路中の意味カテゴリの一一致を検証される。

図7に於いては、名詞句内の修飾関係の検証の流れを示す。 つまり名詞 "a" により幹となる名詞が単数 (可算) であり、 "distributed processing" により、幹の意味が system であることが要請され、幹名詞 "system" の意味と MERGE され、

[CTY,SYS],[NUM,SIN]]

22, 現 TRAP システムの実行例を次
 項 付図 1 に示す。 TRAP は, 当研究
 所獲得の α に よつて試作された C-Prolog⁴
 に よつてインポートされ, メイン
 プロセッサ TRANS に より英文入力モ
 ードとなる。 実行時間は ACOS700 に お
 いて CPU-time 28 秒である。 現在,
 先読み, cut を利用して, 高いコストか
 ら今後効率化が可能である。

5. おわり K

ここに提案した南発手法は、計算機マニュアルという制約下で、右対象文の翻訳システム南発法として、ユニバクトで、初果的であることが確かめられ

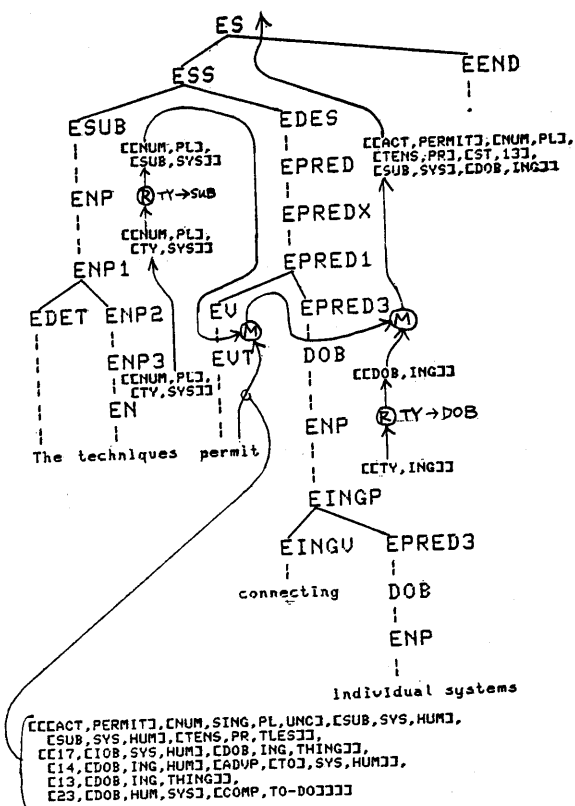


図6. 単レベルの意味処理例

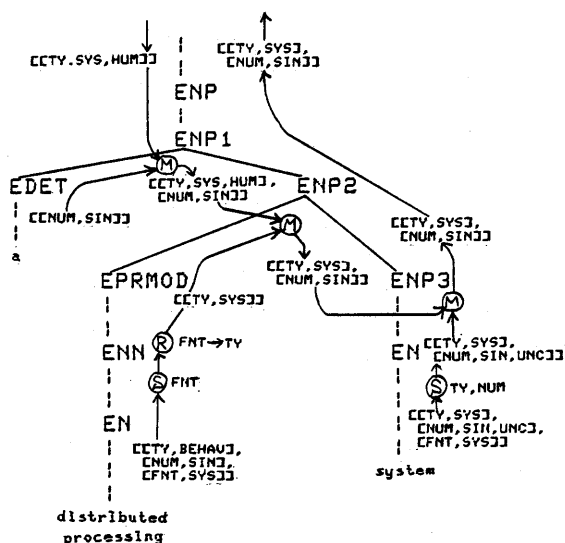


図7. 名詞句レベルの意味処理例

た。この際、可逆性を利用して、解析
モジュールを生成モジュールを使うと
いう新しい手法は、多言語パーキ対
する翻訳システム開発のコスト低減に
大いに意義をもつと考えられる。

また、TRAPではいくつかの残欠が
問題がある。

- i) 意味変換及び意味抽出の導入
- ii) より効率的な辞書、規則管理法
- iii) Prolog 機能の拡張、高速化

現在 i), iii) 点を検討中であるが、i) の意
味の積極的利用により、正しい翻訳文生
成を行う際には、構文解析・意味解析
・意味抽出を統合した意味抽出プロセス
を作ることにより、変換モジュールの極
小化を計ることが考えられる。この
際、Visé 構造に似た構造表現が重要な
働きを持つことになる。

付図 1. 翻訳例

WITH THE ADVANCEMENT OF DATA BASE / DATA COMMUNICATION SYSTEMS, THE COMPU
TER, WHICH ONCE WAS USED MERELY AS A COMPUTING DEVICE, HAS COME TO TAKE
A LEADING PART IN THE COMPUTERIZED SOCIETY.

IL
CES, CESS, CESUB, CEPREPP, CEPREP, WITHJ, CENP, CENP1, CEDET, THEJ, CENP2, CENP3, CEN,
ADVANCEMENTJJJJ, CEPOSTMOD, CEOFF, CEOF, OFJ, CENP, CENP1, CENP2, CENP3, CEN, DATA, B
ASE, /, DATA, COMMUNICATION, SYSTEMSJJJJJJJJJJJ, CECOM, ; J, CESUB1, CENP, CENP1, CEDET
THEJ, CENP2, CENP3, CEN, COMPUTERJJJJ, CEPOSTMOD, CECOM, ; J, CERELWHICH, CEWHICH, W
HICHJ, CEPREDPP, CEADU, ONCEJ, CEPRED2, CEPU, CEBE, WASJ, CEPP, USEDJJ, CEADU, MERELY
JJ, CEPREPP, CEPREP, ASJ, CENP, CENP1, CEDET, AJ, CENP2, CEPREMOD, CEAPJ, CEAPJ1, CEING
U, COMPUTINGJJJJ, CENP3, CEN, DEVICEJJJJJJJJJJ, CECOM, ; JJJJJ, CEPRED, CEPRED
1, CEV, CEHAVE, HASJ, CEPP, COME, TOJ, CEVT, TAKEJJ, CEDOB, CENP, CENP1, CEDET, AJ, CENP
2, CEPREMOD, CEAPJ, CEAPJ1, CEAPJ1, LEADINGJJJJ, CENP3, CEN, PARTJJJJ, CEPOSTMOD, CEP
REPP, CEPREP, INJ, CENP, CENP1, CEDET, THEJ, CENP2, CEPREMOD, CEAPJ, CEAPJ1, CEPP, COMPU
TERIZEDJJJJ, CENP3, CEN, SOCIETYJJJJJJJJJJJJJJJJJJ, CEEND, CECOL, ; JJJ

JL
CJBUN, CBUN, CTANBUNS, ECU, CFUKUKU, CNHEAD, CRENKU, CNHEAD, CNOUN, CN, テー
タコミュニケーション システム JJJJ, ERJOSHI, CNO, JJJJ, CNHEAD, CNOUN, CN, ハッ
テン JJJJ, CJOSHI, CPRE
P1, トモ JJJJ, CTANBUNS, ECU, MEIKU, CNHEAD, CEM, ECU, CFUKUKU, CAD1, タニ JJJJ, CEM, ECU
CFUKUKU, CNHEAD, CAJNOUN, EMODR, CAJR, コンピューティング JJJJ, CNOUN, CN, テ
ハイス JJJJ, CJOSH
IF, CPREP1, トシ JJJJ, CEPRED, CPREDYE, CVERBY, CPUY, ツカワレ JJJJ, CAVERB, CAU3R, タ
ニ JJJJ, CNHEAD, CNOUN, CN, コンピュータ JJJJ, CJOSHI, CPR, カ JJJJ, CTANBUNS, ECU, MEIKU, CNHEAD, C
RENKU, CNHEAD, CAJNOUN, EMODR, CAJR, ショウホウカ JJJJ, CNOUN, CN, シ
カ JJJJ, CJOSH
IR, CPREP2, ニオイ JJJJ, CNHEAD, CAJNOUN, EMODR, CAJR, シ
ウケ JJJJ, CNOUN, CN, タ
ニ JJJJ, CJOS
HI, CPR, ヲ JJJJ, CPRED, CPREDYS, CVERBY, CVY, トル JJJJ, CAVERBS, CAUY, ヨウニナ
ツテ JJJJ, CAU3R, キタ
JJJJ, CEJEND, ; JJJJJJJJJJJ

出力 テー
タハイス& テー
タコミュニケーション システム ノ ハッ
テン トモニ タニ コン
ピューティング テ
ハイス トシ ツカワ
レ タ
コンピュータ カ
ショウホウカ シ
カイ ニオイテ シ
トウケ タチハ
ヲ トル ヨウニ
ナツテ キタ

最後に、本研究の機会を与えられ
たコンピュータシステム研究部三上部長、箱崎
課長に感謝します。また、牧野主任
栗野氏、宮下氏には討論、開発への支
援に感謝致します。

参考文献

- 1) F. C. N. Pereira and D. H. Warren, "Definite Clause Grammar
for language Analysis", *Artificial Intelligence*, Vol. 13, 1980.
- 2) 市山, 他, "Prolog を用いた自然言語入力変換処理システム, 情報処理, 1982
- 3) 村本, 他, "自然言語理解におけるプロログの利用, 情報処理, 1981.
- 4) 橋田 "C-Prolog の研究", 〒7=カキモ, 1981
- 5) D. G. Bobrow and T. Winograd, "An Overview of KRL, a Knowledge Repre-
sentation language", *Studies in Cognitive Science*, Vol. 1, No. 1, 1977.
- 6) B. Vauquois, "The Evolution of Second Generation approach and the
main features of the Grenoble System", 1979.
- 7) A. Colmerauer, "Metamorphosis Grammar", in *Natural language Com-
munication with computers*, Springer-Verlag, 62, 1978
- 8) A. Colmerauer, et al., "Les aptitudes d'un formalisme pour analyser
et synthétiser des phrases sur ordinateur, Publication interne,
Département d'Informatique, Université de Montréal, 1970.
- 9) W. A. Woods, "Experimental Parsing System for Transition Network
Grammar", in *Natural language Processing*, Algorithmic Press, 1971
- 10) 田中, 他 "自然言語処理のためのプロログモックシステム - 拡張 LINGAL
について -", 信管学会論文誌, J60-D, 12, 1977