

機械翻訳における概念変換について

内田裕士 増山顕成
(富士通研究所)

1. はじめに

2. 機械翻訳における交換

ある。換そ
る。變、
あります。
すと換て
こ變し
すに麻
直文内
しのを
現語規
で標換
語目變
表言規
の文は、
他の器
を語換
味言變
意原のす
た、こ力
されは、
さる。出
現テキシ
る械す力
あ機な入
でシが変
語訳とを
言翻こ文
し器れ
翻たでに

來て対式と
出せ換
がか変
い備文
連をが
の則の
式規つ
方換立
で般両
翻，端
訳變に
翻，機
訳といく
かもぶ
傷數呼
でのと
こ則現
ど規表る
を換間あ
則交中で
規な式を
換要現方
変必表式
のにるツ
こ訳なボ
翻」とビ

式だ記現方し間すのと中換こ則な規確變。規明文あ變、力でをに、出式係般ま方関一まる応。そのい対る。それでなれ文に文さの大用力で限入採べ不入にすばよど、數呼ばよどは則とたなは持機に規式を詰め換式を現電を接続する、直間ゆなりを中わ行式を方こなくい方、いをが方はは誤要誤式れ翻必經式換。確おな變る正てた文あ、しもでと憶を

式數不可方の現い換則実扱變規り、換は取る変ムを。要ス則じ必シ規同に訳換はめ翻変質たる、のるなは翻を得に大めは果限たに結無る的訳が現理の則実も質規をてじ換ムつ同変テと。スルをはるシモ方のあ訳が式のあ訳が方の翻訳なくけ、少うてだでよつるのえの連くる抑どててあにだつてで數たよ達能なが可能

間と、つ式前
中向よ方前
方に換
しると交る
析めこ語あ
解高る单で
ををすはの
文)換例も
力る交なる
入すに端え
に々極換
る現別もき
あ表を最置
がじどのに
向同な者語
方を文後单
の現構の
つ表とする語
2な語あ言
は様單で標
て多を向目
し(加方で)
と度増す語
向象のら單
ある
方抽則減の
るの規を語
えそ換言方
抑、交則原ト
をしる規はッ
數出よ換れ
則りに交こ
規作せ、
換を合える
表現も抑
表組でて
の

抑式言うきたを方標いべけ
數ア目とぶ避則フる呼を規ス接すとと
換ン直換式こ
変ラは変方るてト語にアする
し文单造フ加ス構の構ス増
クル語文シが
ッあ言構ラ則
ミで原のト規
を流、語文換
向主ち言構変
方の持標・て
つ研究を目換つ
2の辞造接に
の訳構直せ
記翻対文語合
上械に構單み
ム現は原正構
テのはと
スばもしれ語
シえる換こ単
訳例い変り
翻てにるりる
のるれ語あああ
際いわ单ででで
実ていの式のの
えと語方もも

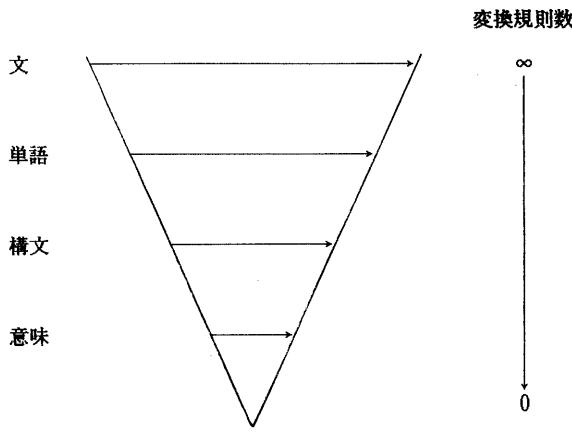


図1

3. 概念変換

取り扱い可能な変換規則の数が一定であるとすると、一般には中間表現の入力文に対する抽象度をあげるほど、質の良い翻訳結果が得られることになる。入力文を解析するという行為はまさに、この抽象度を上げるために行なわれるものである。

本論文で紹介する変換方式は機械翻訳システムATLAS/IIの変換部に採用されている変換方式である。ATLAS/IIの処理概要を図2に示す。

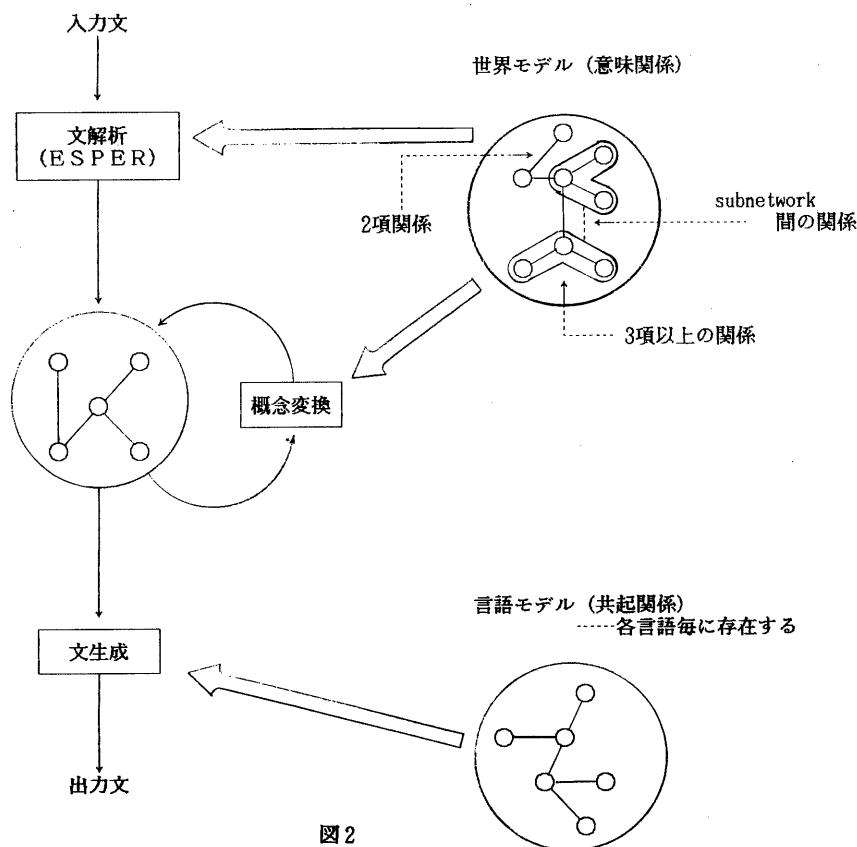


図2
(2)

成るの規世エ対にたれ
たりな逸造か造しさ成
生うも換はチに等れさ
しこ的念構う構回義生
用起則概念ど念い定ら
使に規る概か概言でか
を間タすのとの式造
ル念メ介こいこ語形構
デ標した紹。て、言な念
モ、しでるる様様概
語し義これなれ目同て
言とこさにさはとっ
とドをがわスナル従
析一係部表スルみデテ
解ノ閑てレとモギル
たののしカい語界デ
し念間係とオし言世モ
用標ト闇造の正。て語
使なッの構つはるし言
世の内間念1果れとの
ル語ブト標の結わ係こ
デ単サツはと折行関が
モはとネ果こ解が起ト
界ルトブ結るば換共ス
世デッサ折いれ変がキ
はモネ。解てあ念係テ
I界味るのれで標闇の
I世意い文さうばの語
／。たて力義それ間言
Sるしれ入定。あ素標
Aい義さ。によるが態目
して定成るルれ要形。
Tっを構あデさ必で。く
A行係らでモクトすのい
を閑か則界ツし閑もて

4. 概念構造

れす場も
に造報
る構情
ある時念の
がの概ど
クどたな
ーなきる
<結文
特殊<PAST>
ア>である
特し、た身
いでしけ
なの析受
たも解や
もすを文
をわ文令
ド表力命
一を入
ノ報、徵
の情る。特
手加れ的れ
相付体さ
になわ文わ
か的表の表
ほ味が文で
の意相力く
こる様入一
すのにア
は対々かの
に種ほら
造ドやのれ
構一報ここ
念ノ情はた
概はる合ま

5. 概念變換規則

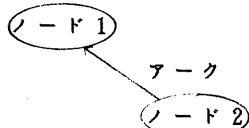
概念変換ルールは部分ネット間の関係として記述される。

(部分ネット, 部分ネット, relation, condition)

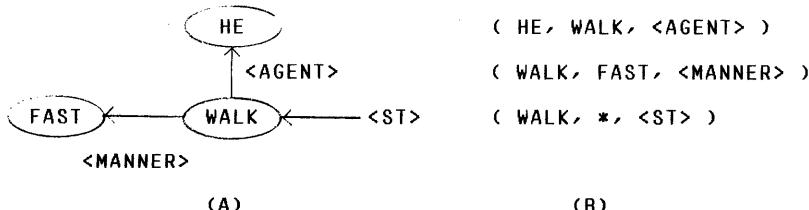
部分ネットは2項関係の集合で表現される。2項関係とは次のようなものである。

(ノード 1 , ノード 2 , アーク)

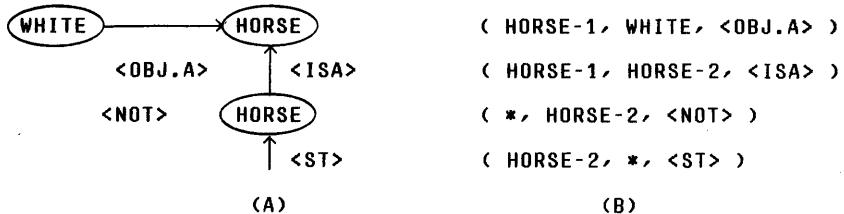
これで次のようなネットワークを表現する。



この 2 項関係の集合で部分ネットを表わすのである。たとえば、"彼は早く歩く。" という文の構造は下図(A) のように表わされるが、これを 2 項関係の集合で表現すると下図(B) のようになる。



クも念はつ
一で櫻馬2
ワジ、白が
ト同は「ド
ト」が合、一
ド係現たを
ネ) 場ばノ
一関表、念
ノ間をる櫻
つ念ドすう
も(ノ区と
ルるて HORSE
号) べなけ
記ラ異付

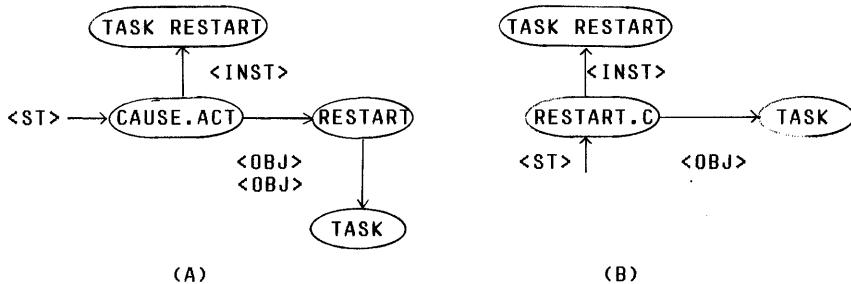


また、1つの変換規則の適用範囲を広げるために、シソーラスにおける上位の概念記号や変数を記述できるようになっている。変数は*と数字で表わす。
 relationは部分ネット間の関係を表わす。
 conditionは変数の値域を指定したり、ある特別な関係が世界モデル中に存在するかどうかを調べたりする。

次に、日本語から英語への概念変換の例を紹介する。

英語では主語・他動詞・目的語という構文パターンはきわめて基本的である[2]ため、他動詞を中心にして生成した方が、生成しやすく、かつ自然な文が生成されることが期待できる。

日本語で自動詞+させるの形で表現されているものも、もし英語に他動詞があるならば、一語で表現した方が自然である。たとえば、"タスククリスターはタスクを再実行させる。"という文の概念構造は下図(A)のようになる。



日本語においては再実行という単語は自動詞であるが、英単語のrestartは他動詞にもなる。自動詞と他動詞では、当然、表わす意味が異なっている。英単語restartの自動詞の表わす概念をRESTART、他動詞の表わす概念をRESTART.Cとする。

このとき、上図のCAUSE.ACTとRESTARTはまとめて他動詞概念であるRESTART.Cにした方が良い。これによれば、上図(A)の概念構造は上図(B)の概念構造に変換される。このときの変換規則は次のようなものである。

```

(*0, CAUSE.ACT, <OBJ>) (*1, CAUSE.ACT, <INST>) (*2, CAUSE.ACT, *3)
(CAUSE.ACT, *4, *5),
(*1, *6, <INST>) (*2, *6, *3) (*6, *4, *5),
(J->E, J<-E),
(C.V, *0, <>) (*6, *0, <CAUSE TO>) (ENG.C, *6, <>)
  
```

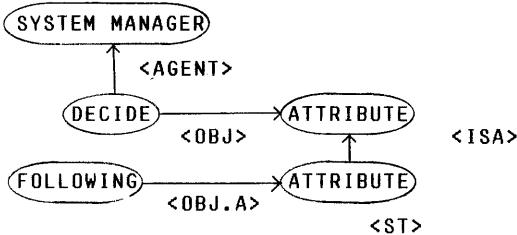
condition部で、(*6, *0, <CAUSE TO>)という関係は*6が*0を他動詞化したことを示す。この例文の場合、*6がRESTART.Cに、*0がRESTARTに当たる。次の(ENG.C, *6, <>)は*6が英語の概念に存在することを示す。

この変換された概念構造から次のような英文が生成される。
 Task restart restarts a task.

日本語の叙述表現についても同様の変換が必要になる。
 次の例文について考えてみる。

"システム管理者が定める属性は次の属性である。"

この文の概念構造は下図のようになる。



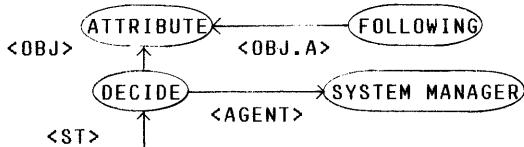
この概念構造を他動詞中心の構文に変換する変換規則は次のようなものである。

```

(*0-1,*1,*2) (*0-1,*0-2,<ISA>) (*0-2,*,<ST>) (*0-2,*3,*4)
(*5,*0-2,*6),
(*0-1,*1,*2) (*1,*,<ST>) (*0-1,*3,*4) (*5,*0-1,*6),
(J->E,J<-E),
(C,V,*1,<>) *2=<AGENT> | <INST> | <OBJ>
}

```

この規則によって下のように変換される。



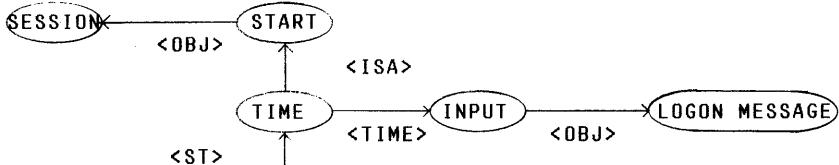
この概念構造によれば、次のような英語が生成される。

The system manager decides the following attributes.

この概念変換に似た例として形式名詞が述語になった場合がある。たとえば、次のような文の場合である。

"セッションの開始はログオンメッセージを入力したときである。"

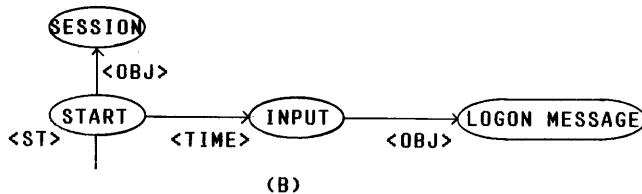
この概念構造は下図のようになる。



この概念構造を下図(A)のルールにより下図(B)のような概念構造に変換する。

```
( (*0,*1,<ISA>) (*2,*1,*3) (*1,*4,*5),
  (*2,*0,*3) (*0,*4,*5),
  (J->E,J<-E),
  (C,V,*0,<>) (C,EN,*1,<>) *3=<TIME> | <PLACE> | <COND> )
```

(A)



生成される英文は、

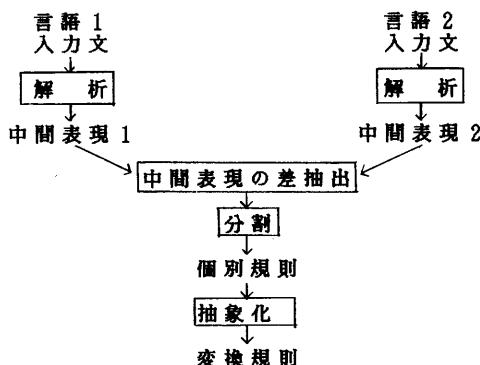
The session starts when a logon message is input.

となる。

6. 自動学習の可能性について

高品質の機械翻訳を目指すには上記のような変換ルールを大量の文章について調べ、収集する必要がある。これを入手して行うのは、コストのかかる自動学習を構念レベルで行なうので、変換ルールの自動学習を行える可能性が高い。

その方法の一案を図に示す。



学習のための入力データとしてはある言語対の対訳文が必要である。図において言語1入力文と言語2入力文が対訳となっている。これらはの入力文に対して解析が行われ、それぞれ、中間表現1と中間表現2が得られる。変換規則はこの中間表現の差を埋めるものであるので、中間表現の差をとるこれが変換規則となる。ただし、このままでは不十分なので差を生じた部分を限定化する必要がある。これが分割処理になる。これで個別規則が得られるわけであるが、この個別規則をより一般的なものとし、適用範囲を広げるために抽象化を行う必要がある。

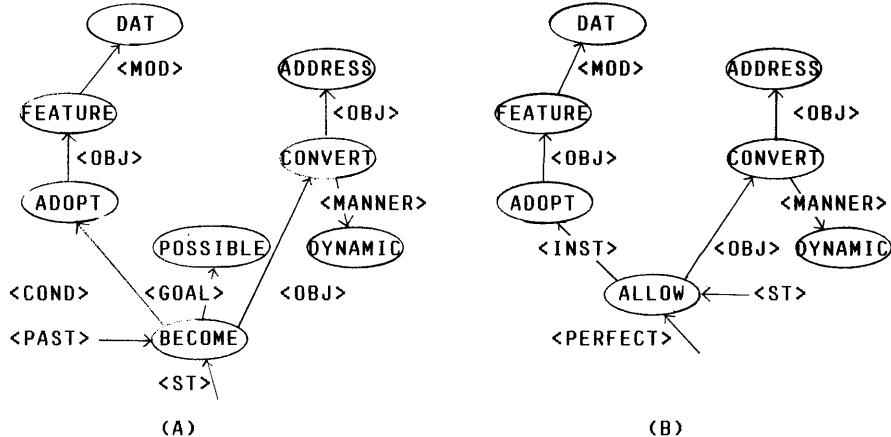
次に具体的な例にしたがって説明する。

以下の対訳文が入力されたとする。

(A) DAT 機構の採用によりアドレスを動的に変換することが可能になった。

(B) Adoption of the DAT feature has allowed to convert addresses dynamically.

上の 2 文のそれぞれの概念構造は次のようになる。



これを 2 項関係集合に直すと次のようになる。

| | | |
|-----|---|---|
| (A) | (DAT, FEATURE, <MOD>) (FEATURE, ADOPT, <OBJ>) (ADOPT, BECOME, <CAUSE>) (ADDRESS, CONVERT, <OBJ>) (DYNAMIC, CONVERT, <MANNER>) (CONVERT, BECOME, <OBJ>) (POSSIBLE, BECOME, <GOAL>) (*, BECOME, <PAST>) (BECOME, *, <ST>) | (DAT, FEATURE, <MOD>) (FEATURE, ADOPT, <OBJ>) (ADOPT, ALLOW, <INST>) (ADDRESS, CONVERT, <OBJ>) (DYNAMIC, CONVERT, <MANNER>) (CONVERT, ALLOW, <OBJ>) (ALLOW, *, <ST>) (*, ALLOW, <PERFECT>) |
| (B) | | |

上図の(A), (B) で互いに等しい 2 項関係を取り除くと以下のようなになる。

| | |
|--|--|
| (ADOPT, BECOME, <CAUSE>) (CONVERT, BECOME, <OBJ>) (POSSIBLE, BECOME, <GOAL>) (*, BECOME, <PAST>) (BECOME, *, <ST>) | (ADOPT, ALLOW, <INST>) (CONVERT, ALLOW, <OBJ>) (ALLOW, *, <ST>) (*, ALLOW, <PERFECT>) |
|--|--|

これで 変換規則が得られた。ただしこれだけでは全く等しい概念構造をサブネットとしてもつ解析結果が得られたときしか有効でない。これを汎用の規則とするためにはさらに抽象化の処理が必要である。抽象化は次のように行う。

1. 個別規則を集める。
2. 個別規則をネットワークの等しいものについてグループ化する。
3. 個別規則中の対応するノードで概念の異なるものをシソーラスの上位概念で置き換える。

上記の 3. の操作には人間の介在が必要であるかもしれない。

7. おわりに

概念変換はネットワークマッチングによって行われるので、規則数が増えた場合にはこのスピードが問題になってくるとおもわれる。概念変換の果たす役割は翻訳家のもっているノーカウに属するような知識であり、意訳を可能にするなど、高品質の翻訳結果を得るにはなくてはならないものであるので、その速度向上が今後に残された課題である。

謝辞： 日頃より、概念構造変換規則の作成に多大なる協力をしていただいている FIE の文屋俊介氏に感謝いたします。

[参考文献]

- [1] 内田裕士ほか：“日英機械翻訳システムATLAS/U”，自然言語
處理 29-3, 1982
- [2] E. G. サイデンステッカーほか：“日本語の翻訳”，大修館書店，
1983