自然言語処理 44－6
（1984. 7. 26）

エスペラントを仲介言語とする機械翻訳の試み

勝守 寛　　福田 基
（中部大学・工）　（中部大学・情報処理センター）

1．はじめに

近年コンピュータを用いた自動翻訳の研究は，わが国でも欧米諸国でも広く行われている。最近では日本語と英語の間の自動翻訳システムは，大企業等による大がかりな研究開発が進んで実用化の段階に入って来たといわれる。一方ヨーロッパ共同体ＥＣのような多種言語を取り扱う国際機関における必要性から，オランダのＢＳＯ社その他が最近開発した翻訳システムＤＬＴ（distributed language translation）[1] が注目されている。ＤＬＴでは入力言語ＳＬ（source language）を出力言語ＴＬ（target language）に変換する中間段階で，内部言語ＩＬ（interlingua）としてエスペラントをベースに開発したものが使われている。ＥＣ諸国間の多種言語相互翻訳で，高品質の訳文が得られている。（Fig. 1）

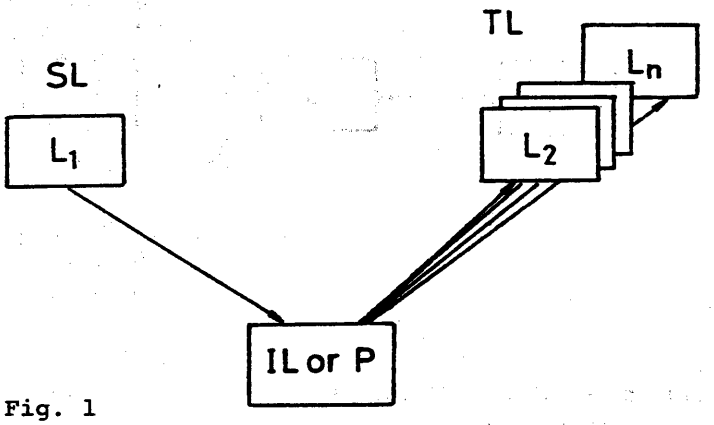


Fig. 1

よく知られているように，一般にｎ通りの異なった自然言語Ｌ1，Ｌ2，．．Ｌｎの間で必要な相互翻訳プログラムＬｉ⟷Ｌｊはｎ(ｎ-1)通りあるが，もしChomsky [2] の言うような深層構造の立場で，個々の言語によらない普遍的意味表現ピボット（pivot）Ｐが中間言語として実現できると，Ｌｉ⟷Ｐ（ｉ=1，2，．．ｎ）の２ｎ通りのプログラムを作ればよいことになる。上記ＤＬＴシステムにおける内部言語ＩＬは理想的なピボットに一歩近づこうとするものと考えられる。

ＤＬＴで内部言語にエスペラントをベースとして限定修正されたものが使われているのは，エスペラントが機械処理をするのに他のどの自然言語にもない優れた特長をもつためと考えられる。エスペラントは１００年近く前（1887）にポーランドのZamenhofによって発表され，現在も国際補助語として期待されている人工語である。当然のこととしてコンピュータで処理することを前提に作

られた言語ではないが，人工的に合理的な構成で作られ，規則的な文法をもっているので，比較的少ない手間をかけて機械処理に適する形にできる利点があると思われる。

数年前から，われわれもこのような点に着目して，エスペラントが多種言語間の相互翻訳に際し仲介の役をする橋渡し言語(bridge language, pontlingvo) ＬＯとして有効な働きをする可能性があるのではないかと考え，検討してきた。[3,4,5] われわれが考えている橋渡し言語ＬＯは，ユーザーが訳文として取り出せるものであって，ＤＬＴにおける内部言語ＩＬやピボットＰのように，一般にはユーザーにわからないブラックボックスではない。（ＤＬＴでも必要ならＩＬを出力できる。）この場合も翻訳プログラムは Ｌｉ⟷ＬＯ (i=1, 2,.. n) の２ｎ通り作ればよいことになる。(Fig. 2)　つまりＬｉとＬＯの間には明確な中間表現がないか，または低レベルの中間表現で直接変換方式に近い簡単なものをつくり，任意の自然言語Ｌｉ，Ｌｊ間の翻訳ではＬＯが中間表現的な働きをする，といったシステムができないものか検討してみた。

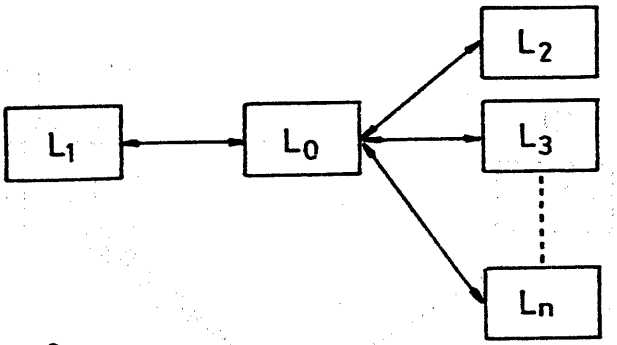


Fig. 2

この目的のために文献３）ではまず日本語とエスペラント間で，形態素解析から合成への単語直接変換に近いものに，低レベルの構文トランスファーを考慮したプログラムを作成した。文献４）では英語とエスペラント間での同様なプログラムを作成したが，構文解析を多少レベルアップしたのが文献５）である。

このような方式でわれわれが目標としている翻訳のレベルはごく初歩的でしかも小規模のものである。近年増加する学術情報に対処するための二次情報として，論文題名，抄録，キーワードなどの膨大な資料の翻訳の必要性が増しているが，ごく限られたせまい範囲の専門分野で手軽に機械翻訳でき，僅かの後処理ですむようなものができないだろうか。

翻訳する文書の内容について分野を限定すれば，対応する単語の多義性，文法上の任意性などのあいまいな多重性をかなり減らすことができるはずであるが，入力文または出力文の一方の言語がエスペラントである場合，変換の際のあいまいさをさらに少くすることが期待できる。

利用したコンピュータは中部大学情報処理センターのＦＡＣＯＭ　Ｍ１６０ＡＤで，使用言語はＦＡＣＯＭ　ＯＳIV／Ｆ４　ＰＬ／Ｉ　である。文字列データ処理の容易さを考えてＰＬ／Ｉを使用したが，現在ＬＩＳＰの使用を検討中である。

２．構造解析

文献５）で行った方法について，エスペラントから英語に翻訳する場合を例にとって解析のプロセスを説明する。

１）形態素分解：語と語の間のスペースによって単語単位に分解し，単語の語尾の形などによって品詞の区別をして種々のパラメータを認識抽出する。

２）文章の構造の簡略化：単語間の係り結びがエスペラントでは比較的分り易いので，たとえば形容詞＋名詞を１まとめにして名詞相当語としたり，前置詞＋名詞を１まとめにした修飾語として１語扱いをする。

３）単文，重文，複文の判断と処理：節と節を結ぶ接続詞や関係詞がないと単文，等位接続詞で結ばれていると重文，それ以外は複文と判断して，それぞれつぎのように処理する。

（ａ）単文：並べられた単語のうち最初にくる主格（名格）の名詞，代名詞，疑問代名詞，不定詞を主語とする。エスペラントでは英語の形式主語に相当する場合は文頭に動詞がきて主語はない。語尾がａｓ，ｉｓ，ｏｓ，ｕｓ，ｕ，ｉの単語を動詞とし，語尾で法および直説法の時制がわかる。補語，目的語の判断は動詞の種類，対格の名詞相当語の有無，不定詞の存在などによる。文中にｎｅがあれば否定文とし，ｎｅをとって肯定文と同様の処理をする。疑問文はĈｕ...？の場合，Ĉｕをとって肯定文と同様の処理をするが，訳文生成の際にｂｅ動詞または助動詞と主語の順序の入かえをする。疑問代名詞，疑問副詞などを含む場合は疑問詞をとって肯定文と同様の処理をし，訳文生成のときに疑問詞をつける。

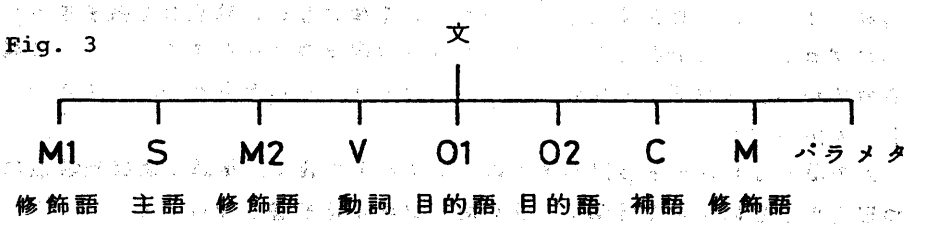
（ｂ）重文：等位接続詞をはさんで，前後の二文にわけ，それぞれを単文として処理する。

（ｃ）複文：従属接続詞または関係詞の位置により，主文と従属文にわけてそれぞれを単文として処理する。

４）修飾語の処理：主語，動詞，目的語，補語以外で文頭に来るものＭ１，文尾に来るものと時を表わす副詞をＭ３，それ以外の副詞はＭ２とする。

５）構造表現：各要素の処理が終ったらＦｉｇ．３，Ｆｉｇ．４に示すような構造表現をする。

Fig. 3



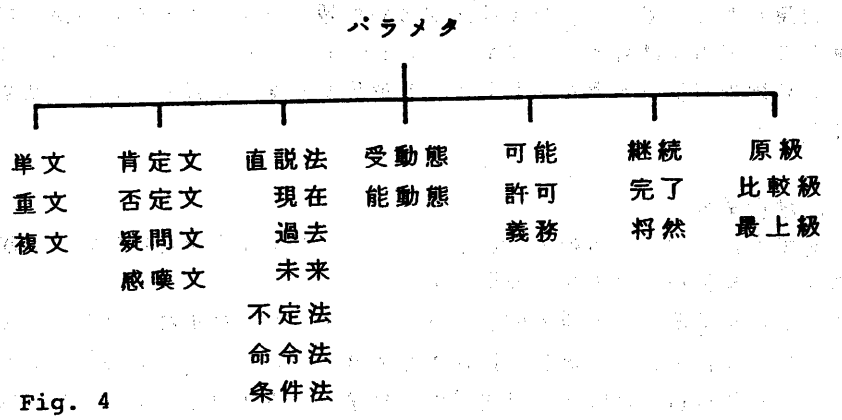


Fig. 4

3．　生成文法

エスペラント文を生成する場合の各要素に対する処理の概略を説明する。

1）主語：複数名詞は語尾にｊをつける。形容詞も数を一致させる。

2）動詞：不定法，現在時，過去時，未来時，条件法，命令法，に対応して語尾はつぎのとおり，ｉ，ａｓ，ｉｓ，ｏｓ，ｕｓ，ｕ

受動態はｅｓｔｉの適当な形式＋必要な動詞の受動分詞でつくる。

英語の助動詞に対応するものはエスペラントにはないので，つぎのような関係で処理する。

可能　ｃａｎ など → ｐｏｖｉの適当な形式

許可　ｍａｙ など → 条件法にする

義務　ｍｕｓｔ など → ｄｅｖｉの適当な形式

単純未来　ｗｉｌｌ など → 未来形にする

意志未来　ｗｉｌｌ など → ｖｏｌｉの適当な形式

疑問　ｄｏ など → ĉｕで始める

3）目的語：単数対格は語尾にｎを，複数対格は語尾にｊｎをつける。

4）補語：主語と同様

5）修飾語：副詞には品詞語尾のないものと，品詞語尾ｅ，ｅｎのある派生副詞があることを考慮して処理する。

4．翻訳システムプログラム

1）システムの概要

英文を読み込んでエスペラント文を出力する場合の翻訳システムを例として説明する。Fig.5 に示すように，システム全体は互いに独立した機能をもつ５つのプログラムに分割され，各プログラムは階層構造をなすモジュールで構成されている。プログラム全体のロード・モジュールは９６Ｋバイトである。

2）入出力文

入力原文はレコード長80バイトのファイル上に書く。単語および句読点は１つ以上の空白によって区切る。文の終りは終止符を書き改行する。

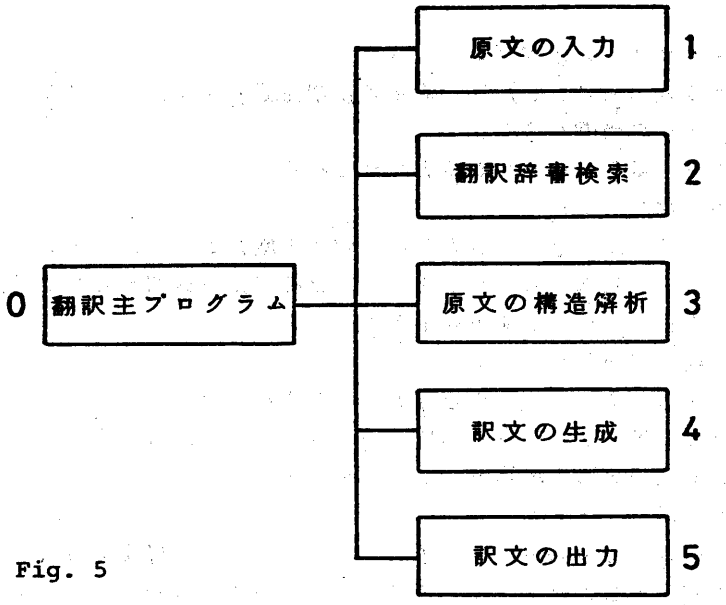


Fig. 5

訳文はレコード長80バイトのファイルに出力される。文は第1カラムから始まり，単語および句読点は1つの空白によつて区切られ，1文が1行に書き切れないときは改行する。

3）単語翻訳辞書と連語翻訳辞書

単語翻訳辞書は英単語に対応するエスペラントの単語とそれに付随する情報を含むデータセットであって，レコード長100 バイトの非ブロック化順編成となっている。現在1780語が登録されて，磁気ディスク上に3シリンダの容量をもつ。

さらによく使われる連語(Collocation)について連語翻訳辞書ができている。レコード長120バイトの非ブロック化索引順編成データセットで，現在163語が登録されて，磁気ディスク上に1シリンダ の容量をもつ。

単語辞書，連語辞書の内部フォーマットを Fig.6 に示す。 付随情報欄は品詞の種類によって1～6個までのデータをもっている。たとえば名詞ならば人間か物か，単数か複数か，主格か目的格かなどの情報を示す。

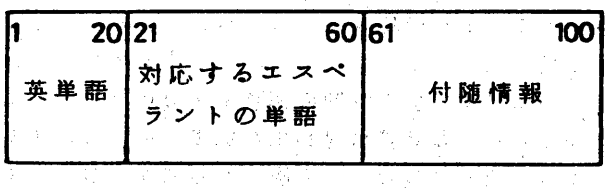
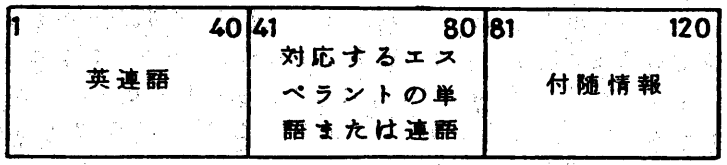


Fig. 6



4）翻訳主プログラムと副プログラム

主プラグラム（ＭＤＬ０）は，変数の初期設定および副プログラム（ＭＤＬ１～ＭＤＬ５）の管理をする。

副プログラムはそれぞれつぎのとおりである。

原文入力（ＭＤＬ１）：

入力ファイルから語，句読点などの要素を順次とり出し，１レコードづつ入力バッファに転送，入力バッファ内の語を空白を区切りとして単語に分解する。終止符がくると１文の終了とする。

翻訳辞書検索（ＭＤＬ２）：

原文の単語をキーとして単語辞書を検索する。見つからないときは語尾を変化させて再検索する。さらに見つからないときは固有名詞として扱う。連語辞書を原文の単語を総称キーとして検索する。

構造解析（ＭＤＬ３）：

単語に付随する情報および語順をもとにして原文の構造を解析する。前後の品詞が同じ単語または１まとめにしてできる単語の組をまとめて編集する。文が平叙文，疑問文，命令文，感嘆文，の何れであるかを解析し，平叙文型に編集する。接続詞または関係詞で分けられる文は節に分解する。各節について肯定，否定，時制，態，法などの判定をする。述語動詞を決定し，前後の単語がそれぞれ文のどのような構成要素であるかを解析する。

訳文生成（ＭＤＬ４）：

従属節の時制を主節の状態によって処理するが，英語とエスペラントとの文法上の相違に注意する。疑問文は疑問詞または Ĉu を文頭につける。

訳文出力（ＭＤＬ５）：

単語を連結して生成された文を出力バッファに入れる。出力バッファがいっぱいになるか，訳文がすべて入ったら，出力バッファの内容を出力ファイルに書き出す。

5．訳文例

日本語はローマ字で入出力する。現在は入出力ともヘボン式にしてあるが，はねる音「ん」はｎとし，つまる音は最初の子音字を重ねて表わしてある。現状では，日本語からエスペラントへの場合と，エスペラントから日本語への場合で，助詞の表し方が不統一になつている。（「は」が入力文では”ＨＡ”，出力文で”ＷＡ”になつているのは，担当者間の調整不十分のためである。）　また日本語ローマ字でかくとき，長音符は使わずかな書きのようにする。

（例　応用　おうよう　ＯＵＹＯＵ，　大阪　おおさか　ＯＯＳＡＫＡ　）

エスペラントの入出力で supersigno （上つき符）の問題があるが，これはあまり本質的な問題ではないので，現状ではアポストロフィまたは１重のクォーティションマーク ' で代用することにした。すなわち Ĉ，Ĝ，Ĥ，Ĵ，Ŝ，Ŭ　はそれぞれ Ｃ'，Ｇ'，Ｈ'，Ｊ'，Ｓ'，Ｕ' で入出力する。

つぎにいくつかの訳文例を示す。

1）日本語からエスペラントへ

```
INPUT DATA
ﾜﾀｼ ﾊ ｷｮｳｼﾞｭ ｰ ﾃﾞｽ
WATASHI HA
KYØUJU -
DESU
MI ESTAS PRØFESØRØ
```

```
INPUT DATA
ｶﾉｼﾞｮ ﾉ ﾎﾞｳｼ ﾊ ｱｶｲ ｰ ﾃﾞｽ
KANØJØ NØ
BØUSHI HA
AKAI -
DESU
S'IA C'APELØ ESTAS RUG'A
```

2）エスペラントから日本語へ

```
KIU VI ESTAS ?
ANATA WA DARE DESU KA ?
```

```
LI SØLVØS LA PRØBLEMØN .
KARE WA MØNDAI Ø KAIKETSUSURUDESHØU .
```

3）英語からエスペラントへ

```
PLEASE INPUT ENGLISH
: IN WHICH RØØM DØ YØU SLEEP ?

**************************************
**   EN KIU C'AMBRØ VI DØRMAS ?
**************************************
```

4）エスペラントから英語へ

```
PLEASE INPUT ESPERANTØ
: LI MØNTRIS AL MI LA VØJØN AL LA STACIDØMØ .
```

```
HE SHØWED ME THE WAY TØ THE STATIØN .
```

6．おわりに

以上のようなごく初歩的な段階での試みから結論らしいものを引き出すことは無理である。しかし当然予想されることながら，つぎのような点は明らかと思われる。

1）　エスペラントはその人工語としての特長から，文の構造解析は比較的容易なので，エスペラントから日本語，エスペラントから英語への翻訳は，これらの逆よりも処理し易い。

2）　エスペラントは公平な国際語といわれるが，やはりヨーロッパ的言語であるから，英語との相互翻訳の方が日本語との相互翻訳よりは処理し易い。

3）　最も問題が多いのは，日本語からエスペラントへの翻訳であるが，これは日本語の構造解析が複雑なためと考えられる。すでに日英間の機械翻訳のために広汎にわたって日本語の構文研究がなされているから，それらの成果を借用するのが賢明と思われる。

謝辞

中部大学情報処理センター長・中村嘉平教授，同主任・水島章次助教授ほかセンター職員各位に大変な御協力を戴いたこと，および日本エスペラント学会理事・永瀬義勝氏に有益なご助言を戴いたことに対し厚くお礼申し上げる。また工学部学生・名知克頼君，若山淳一君の助力に感謝する。

参考文献

1）A. P. M. Witkam, Distributed Language Translation, BSO, Utrecht (1983)

2）N. Chomsky, Aspects of Theory of Syntax, MIT Press (1965); 安井稔訳，文法理論の諸相，研究社（1970）

3）林昌寛，箕浦寛人，渡辺司雄，中部工大1981年度卒論（勝守研）
——日本語とエスペラント間の機械翻訳システム

4）田島久之，田中常則，村松琢巳，中部工大1982年度卒論（勝守研）
——英語とエスペラント間の機械翻訳システムI

5）垣内孝弘，田中成和，八木しのぶ，中部工大1983年度卒論（勝守研）
——英語とエスペラント間の機械翻訳システムII