

エスペラントを仲介言語とする機械翻訳システム I

勝守 寛*, 牧野 悟**, 平田 満***

A Machine Translation System by the Use of Esperanto as a Bridge Language I

Hiroshi KATSUMORI*, Satoru MAKINO**, and Mitsuru HIRATA***

(Received April 15, 1986)

A machine translation program written in LISP has been designed to translate Esperanto into Japanese. Any arbitrarily chosen Esperanto sentence is usually expected to be grammatically regular, lexically unambiguous, and deterministic, and so an analysis tree can be easily obtained. Thus the usefulness of Esperanto as a bridge language in constructing a machine translation system is emphasized.

Several examples of translation of declarative, interrogative, and exclamative sentences, as well as a compound and complex sentence, and an inverted order sentence are given, together with a few analysis trees.

1. はじめに

われわれは 1980 年ごろから、人工的国際補助語であるエスペラントの合理的構成や規則正しい文法の特長を生かして、任意の 2 つの自然言語間で、エスペラントを仲介言語 (bridge language) とする比較的手軽な自動翻訳システムが作れるのではないかと考えて種々の試みを行ってきた²⁾。

一方ヨーロッパでは、1972 年に西ドイツの Saarland 大学で, SUSY (Saarbrücker Übersetzungs-System) が開発されて、当初ロシア語からドイツ語への翻訳を目的としたが、その後ドイツ語を仲介として英, 独, 仏, 露, エスペラントなどの言語間の自動翻訳が試みられ、後になってエスペラントも仲介言語に使われた³⁾。また 1983 年にオランダの BSO 社で発表されたシステム

DLT (Distributed Language Translation) では、エスペラントをベースに開発された内部言語 (interlingua) が使われている⁴⁾。

最初われわれはコンピュータ言語として PL/I を使用して、日本語とエスペラント間、英語とエスペラント間の翻訳プログラムを作成したが、1984 年秋から本学のコンピュータが FACOM M-360 AP に置き換えられ、LISP が使用できるようになった。そこでまずエスペラント文から日本文への翻訳処理システムについて構文解析のレベルアップを計り⁵⁾、さらにかんりの改良を行ったので報告する。

2. エスペラントの特長と仲介言語としての有用性

一般に n 種類の自然言語 $L_1, L_2, \dots, L_n$ 間では $n(n-1)$ 通りの相互翻訳プログラム $L_i \leftrightarrow L_j$ を必要とするが、普遍的意味表現、ピボット (pivot) が共通の中間言語として実現すれば Fig.1 に示すように $2n$ 通りのプログラムですむことになる。

最近、大手メーカーなどが開発し、すでに販売している

この論文は、情報処理学会・自然言語処理研究会において、1986 年 3 月 28 日に行なった報告¹⁾を補足改訂したものである。

*工業物理学科 (Department of Engineering Physics)

**工業物理学科学生 (Senior Student, Engineering Physics)

***コンピューターサービス株式会社 (Computer Service Corporation)

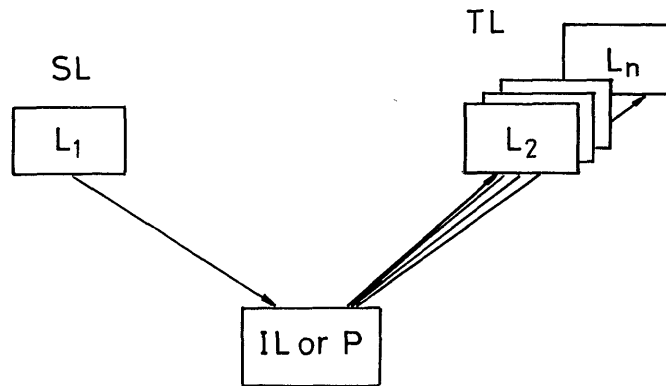


Fig.1 内部言語またはピボット

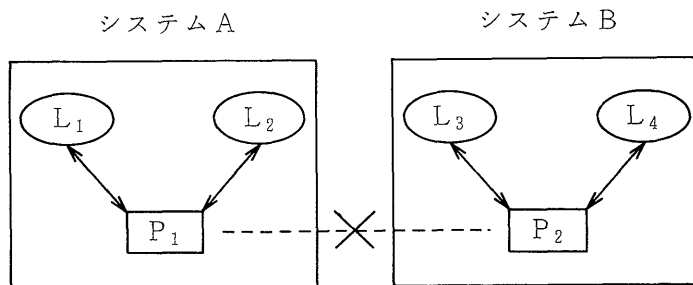


Fig.2 システムに依存する中間言語の問題

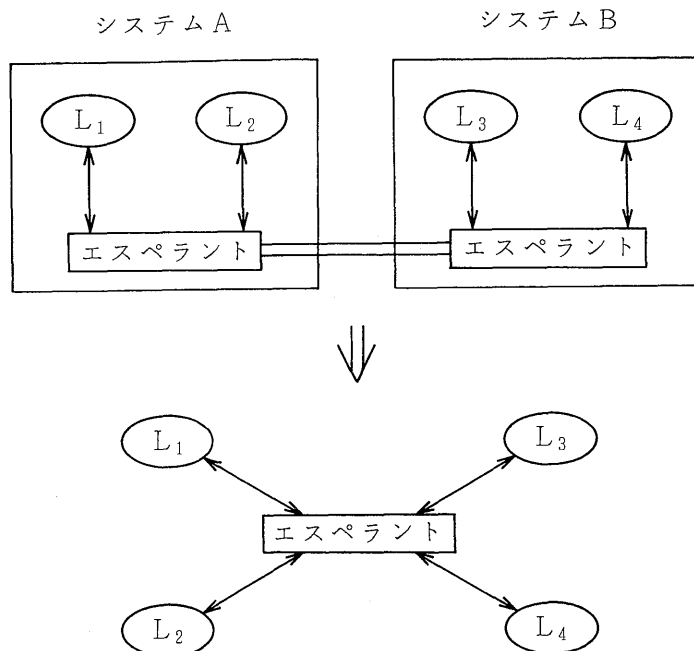


Fig.3 エスペラントを仲介言語とするシステム

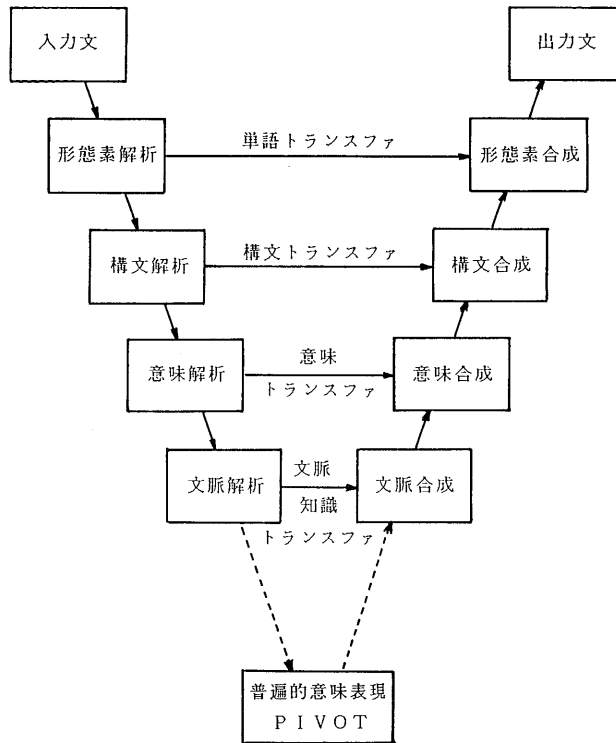


Fig.4 トランスファのレベル

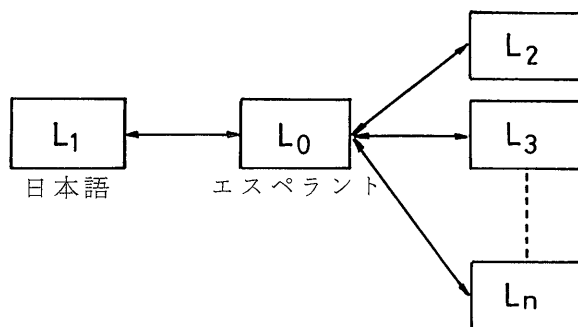
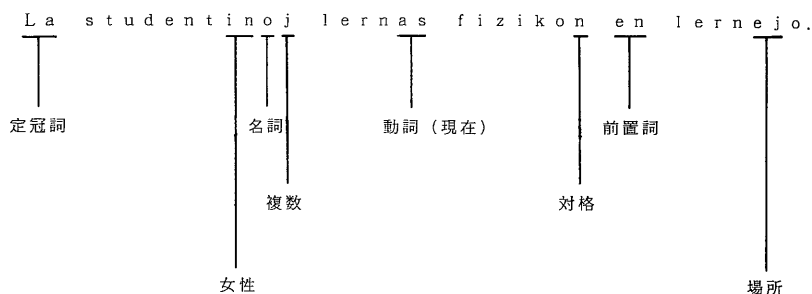
実用的な日英間翻訳システムでは、たとえば、富士通の ATLAS や日本電気の PIVOT など、意味トランスファからピボットへ一歩近づこうとするものと考えられるが、本格的なピボットの理想に到達するにはまた時間を要するであろう。しかも将来ピボットが実現したとしても異なるメーカーのシステムで別々のピボットが開発されるだろうから、それが本来もっているはずの共通性が保証されるかどうか疑問である。たとえば Fig.2 に示すように中間言語はシステムに依存するものになり、システムを超えての L_1 と L_3 との相互翻訳などは不可能となる。

このようにピボットに到達するのは容易ではなく、また共通性の保証が問題となるが、一方その代りにエスペラントのような人工的国际補助語を共通の仲介言語として橋渡しに使う機械翻訳システム (Fig.3) を考えると、やはり n 種類の自然言語間で $2n$ 通りのプログラムを作ればよいことになる。しかもこの場合エスペラントの合理的構成や規則正しい文法の特長を生かして、エスペラント文の解析、エスペラント文の生成の部分が非常に簡単であいまさが少ないため、深いレベルまで下がってピボットに近づくことなく、構文トランスファに意味トランスファを多少加味したレベル (Fig.4) で比較的質

の良い訳文を手軽に得ることが期待される。

エスペラントを仲介言語 L_0 とするシステムでは、たとえば Fig.5 で L_1 を日本語とすると、日本語 L_1 とエスペラント L_0 の間の翻訳プログラムを作っておけば、他の任意の自然言語 $L_i (i=2, 3, \dots, n)$ とエスペラント L_0 間の翻訳プログラムがどんな方式のものであっても、 L_0 を仲介として L_1 と L_i 間の ($L_1 \leftrightarrow L_0 \leftrightarrow L_i$) 訳文を求めることが容易である。もちろん日本語と英語間の翻訳だけに限れば、すでに実用化されている大がかりなシステムが市販されているし、また近い将来翻訳の需要が多くなると見込まれる言語に対しては大手メーカーによって日本語との機械翻訳システムが開発されるであろう。しかし少数民族の言語や国際性の少ない民族語については、同様の翻訳システムが市販されるのを期待することは当分無理であろう。このような少数派言語に対してもエスペラントとの間の翻訳システムさえ作っておけば、エスペラントを仲介として日本語との相互翻訳だけでなく、ヨーロッパで開発されているシステムと組み合わせて他の任意の言語との相互翻訳が容易になるものと思われる。

さて一般にある言語を仲介とする 2 重翻訳、たとえば日本語 $\leftrightarrow$ 英語 $\leftrightarrow$ ロシア語のようなことを行くと、人間が行う場合でも多義性などあいまいさが重なって誤訳のお

Fig.5 仲介言語 L_0 の役割

その 女学生たち は 学校 で 物理学 を 学ぶ。

Fig.6 単語中の文法情報

それが増大することは当然なので、エスペラントを仲介にする場合に対しても同様の懸念が生ずるかもしれない。しかしエスペラントの場合は規則的な体系をもち、例外のない文法、単語に多義性が少ないことなど機械処理にすぐれた特長をもっているため、誤訳のおそれはむしろ減少すると思われる。もちろん完全に近い訳文を得るには前処理、後処理などの人間の介入は各段階である程度必要であるが、エスペラントについては、いわゆる民族語に比べてはるかに少ない労力で済むものと思われる。

エスペラントの語彙論的特長は、Fig.6に示すように単語中に文法情報 (grammatical information: 文中における単語の形態・機能の記述) のラベルを含むことである。名詞、形容詞、動詞は語尾が決定しており、動詞は時制、態、相が決定している。また複数、対格も語尾によって示される。このように品詞は確定的であり、また単語には多義性が少ないことが特長である。

エスペラントは、膠着言語 (agglutinative language) なので、解析において単語は、はっきりした一連の形態素に分析可能であり、語に含まれる種々の情報を容易に抽出できる。また生成においては単純な語幹から規則的

な接辞添加によって語を形成できる。このように既存の概念から新しい概念が容易に作られ、意味が正確になり、基本的な語は少なくてもよい。また不規則変化がないので辞書に独立した項目を与える必要がない。

エスペラントの文法的特長は意識的に合理化された規則正しい体系であり、文法のきまりは、どんな場合にも例外なく通用する。文法以外の習慣的用法はごく限られたものだけである。語彙論的にも品詞が確定的であり、多義性がないことから、エスペラントは文の構造にあいまいさが無い。

このようにエスペラントは民族的な自然言語に近い形で実用性がありながら、人工的に作られた規則的な体系を持っていて、機械処理にはきわめて適した言語であることがわかる。もちろんヨーロッパの言語にもとづいて作られているので、特にヨーロッパ諸国における多国語間翻訳の仲介言語として有利なのは当然である。

3. エスペラントから日本語への翻訳システム

このシステムは、本学情報処理センターの大型コンピュータ FACOM M-360 AP 上に FACOM OS IV LISP で実現している。

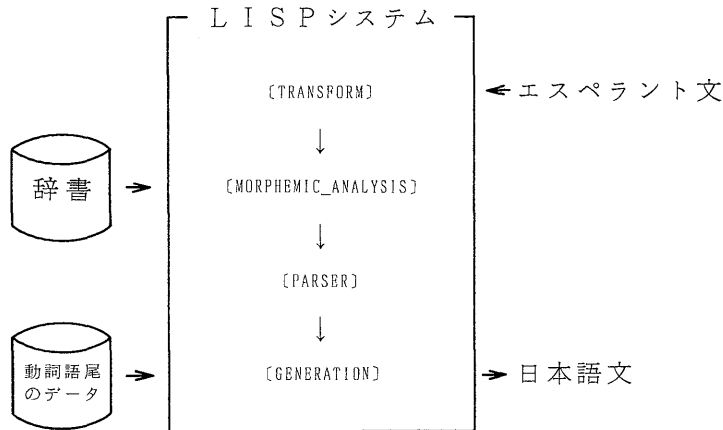


Fig.7 システム構成

システムは約 150 の関数より構成され、大きく 4 つの処理に分れる。システムの処理の流れは次のようになっている。以下関数名を [] で表すことにする。システム構成を Fig.7 に示す。

1. 感嘆文・疑問文の処理 [Transform]
 2. 形態素解析 [Morphemic_Analysis]
 3. 構文解析 [Parser]
 4. 訳文生成 [Generation]
- (1) 感嘆文・疑問文の処理 [Transform]

エスペラントでは感嘆文・疑問文は文法機能の位置関係が平叙文と同じであり、文頭に相関語を付加した形で示される。このため感嘆文・疑問文の要素を取り除いた文への変形は非常に容易である。

感嘆文の処理 — 入力文の最後が感嘆符「!」ならば、この文が感嘆文であると識別する。そして、入力文の最初の 2 つの単語 (Kia+形容詞, Kiel+副詞など) を感嘆文の要素として Short_Term_Memory (以下 STM と略す) に記録し、感嘆符「!」と共に入力文より削除す

```
(AKTIVA (APOS A) (MEAN1 活動的な) (MEAN2 活動的))
(ALFABETO (APOS N) (MEAN1 アルファベット))
(AMI (APOS VT) (MEAN1 愛) (EOAW する))
(ĈU (APOS Q) (MEAN1 か))
(DE (APOS P) (MEAN1 の) (MEAN2 に) (PASSIVE によって))
(EN (APOS P) (MEAN1 の中で) (MEAN2 の中) (GOAL の中に))
(ESPERANTO (APOS N) (MEAN1 エスペラント))
(ESTI (APOS VB) (MEAN1 で) (EOAW す))
(FIZIKO (APOS N) (MEAN1 物理学))
(INFANO (APOS N) (MEAN1 子供) (PLURAL 子供たち))
(KAJ (APOS CNJ) (MEAN1 と) (CO-CNJ そして))
(KE (APOS Q) -(MEAN1 と言うことを) (REL VP) (REL-MEAN のを))
(KIA (APOS Q)
  (MEAN1 どんな)
  (INTERJECTION なんて)
  (REL NP)
  (REL-MEAN NIL))
(KIEL (APOS Q)
  (MEAN1 どのように)
  (INTERJECTION どんなに)
  (REL VP)
  (REL-MEAN のように))
(LA (APOS DET) (MEAN1 その))
(MAŝIN-TRADUKO (APOS N) (MEAN1 機械翻訳))
(SINTAKSO-ANALIZO (APOS N) (MEAN1 構文解析))
(TUTE (APOS ADV) (MEAN1 全く))
```

Fig.8 辞書の記述例

る。

疑問文の処理 — 入力文の最後が疑問符「？」ならば、この文が疑問文であると識別する。そして、入力文の最初の単語 (Cu, または疑問詞) を疑問文の要素として STM に記録し、疑問符「？」と共に入力文より削除する。

STM に記録した情報は訳文生成の際、動詞語尾の選択や関連語の訳語選択に利用される。

(2) 形態素解析 [Morphemic Analysis]

入力された文の各単語によって辞書を引き、品詞、訳語などを得る。辞書には基本形が登録されているので、基本形への変形が必要である。この変形の過程で種々の形態素情報が抽出される。

辞書は見出し語と連想リストで表現される内容からなるリスト構造の形をとり、Fig.8 のように記述されている。項目数は約 2000 で、その内約 400 が動詞である。

辞書の記述中、APOS は品詞 (A Part of Speech), MEAN 1 は基本的な訳語を示す。形容詞 (A) と前置詞 (P) には MEAN 2 があるが、これは、この語が補語となったときの訳語である。名詞の訳語で複数形 (plural) の訳語が選択できる場合はそれを用いる。その他特別な記述がある。

辞書は翻訳システム立ち上げ時に、LISP システム内にローディングされ、属性リストの形で展開される。

エスペラントの単語は規則的な語尾を持ち、不規則変化がまったくなく、Fig.9 のようにたいへん簡単なアルゴリズムで辞書引きができる。たとえば英語の場合の複雑な語尾分析⁹⁾と比較すると、いかに簡単であるかわかるであろう。また単語に多品詞性がないのでこのプロセスにおいて品詞が確定される。ここで得られた情報はすべてその単語の属性リストに付加される。

(3) 構文解析 [Parser]

エスペラントは各単語に多義性がなく、文の構造に決定性がある。このためエスペラント文の解析は構文レベルで行い、利用する知識も非常に限られたものでよい。ここで作成したパーザ (構文解析プログラム) は、文法規則としては文脈自由型句構造文法による、一方向的解析手法の拡張遷移網 (ATN) にもとづくものである。文の構造に決定性があるので、横型探索のように候補を全部同時に調べる必要もなく、バックトラッキングも行わない。構文解析の流れを Fig.10 に示す。

パーザはまず重文の処理を行う。並列の文として扱えるものを分割する。そしてその個々の文を ATN パーザ

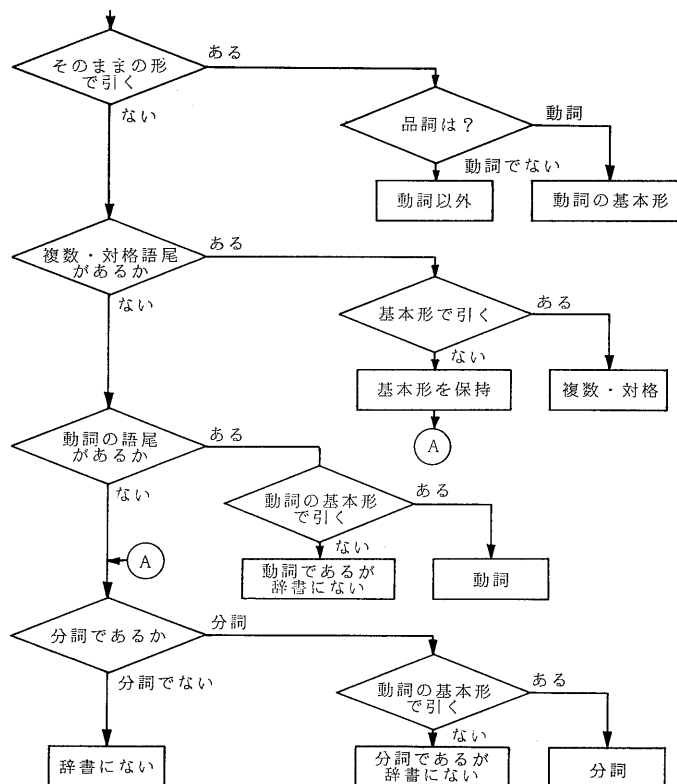


Fig.9 辞書引きアルゴリズム

が処理する。ATN パーザは基本的にトップダウンの動作であるが、一部ボトムアップによる部分がある。ATN パーザは関数 [Transition] を介して S, NP, VP の各 [Transfer] を関数呼び出しして、再帰的に書き替え規則を適用する。たとえば埋め込み文が現われた場合は、

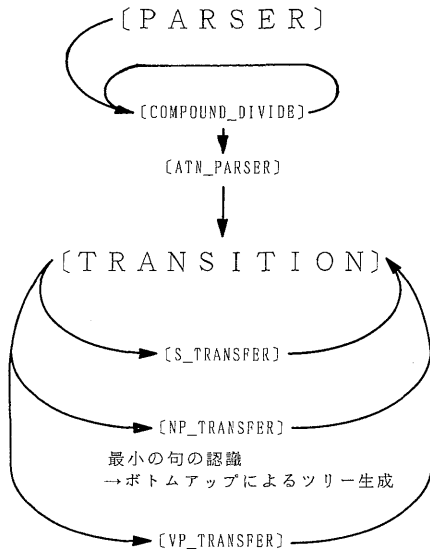


Fig.10 構文解析の流れ

再帰的に [S_Transfer] が呼び出される。

エスペラント文の名詞句の構造は、その中に含まれる句の位置的自由度が高いことが特徴である。このため英語のように語順の体系による構文解析ではなく、名詞句は特別の処理をする。まず、名詞句は並列に並ぶ最小の句を認識する。ここで単語の品詞名からアークを選ぶのであるが、このとき初期状態においてアークの選択に制約をつけない。このように認識された句から部分的にボトムアップにより句を再構成する。またこの解析途中で訳語選択に関する情報を単語の属性リストに付加する。

(4) 訳文生成 [Generation]

訳文生成は次の二つのプロセスによって行われる。訳文生成の流れを Fig.11 に示す。

[Generator] — 構文解析の結果からエスペラントの語順を日本語の語順になるように生成する。[Generator] はデータ駆動によって行われる。構文解析で得た句構造標識の範疇記号から生成規則を選び、日本語の語順になるように生成する。このとき NP と VP の間に適切な助詞を付加する。エスペラントではしばしば倒置が用いられる。倒置であると判断すると、関数 [Inversion] が起動され、助詞を付加する位置を適切な位置へ移動する。このように語順によるニュアンスを日本語に反映するようにした。STM に何らかの情報があれば、そのような情報を処理する関数を起動する。

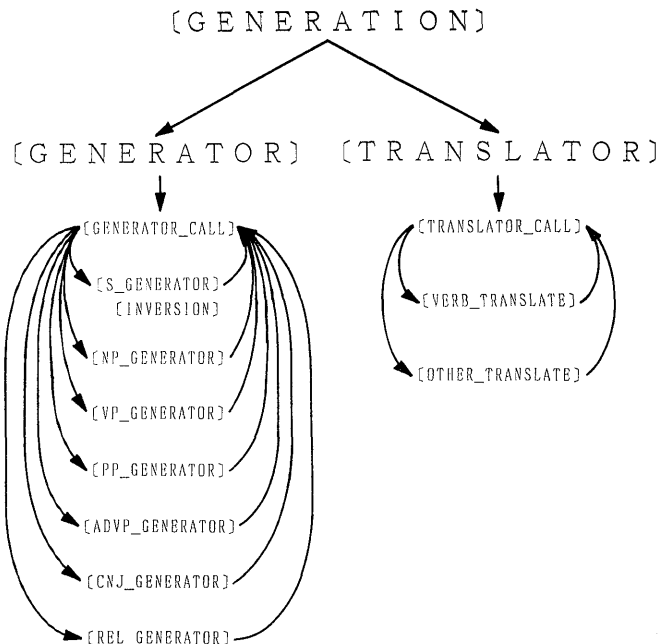


Fig.11 訳文生成の流れ

[Translator] — 日本語の語順になったエスペラントの各単語を、動詞の支配のもとで日本語の訳語で置き換える。多くの場合、すでに訳語は選択されており、単語の属性リストによって訳語を得る。動詞の語尾は動詞語尾専用のデータから得る。ここで、対格には適切な助詞

を付加する。

4. 訳文例

訳文例を Fig.12 (1), (2)に示す。これは実際のコンピュータによる出力結果である。

平叙文の例

```
((Esperanto estas tre utila lingvo por maŝin-traduko))
= S ==== NP === NP === (ESPERANTO)
      = VP === V ==== (ESTAS)
            = NP === NP === (TRE UTILA LINGVO)
                  = PP === P ===== (POR)
                        = NP === (MAŜIN-TRADUKO)
= (エスペラント は 機械翻訳 にとって たいへん 有用な 言語 です。)
```

```
((Mi ne vizitis lin hieraŭ)) *
```

```
= S ==== NP === NP === (MI)
      = VP === V ==== (NE VIZITIS)
            = NP === NP === (LIN)
                  = ADVP = (HIERAŬ)
= (私 は 彼 を きのう 訪問 しませんでした。)
```

```
((Estas bela vetero)) *
```

```
= (すばらしい 天気 です。)
```

```
((Mi iris tien multfoje)) *
```

```
= (私 は そこに 何度も 行 った。)
```

```
((Ŝi estas ĉiam maltrankvila pri la sanstato de sia edzo)) *
```

```
= (彼女 は 自分の 夫 の その 健康 について いつも 不安 です。)
```

```
((La gasto alvenos post tridek minutoj)) *
```

```
= (その 来客 は 三十 分 あとで 到着 するでしょう。)
```

```
((Mi ofendiĝis de liaj vortoj)) *
```

```
= (私 は 彼の 言葉 に 侮辱を感じ った。)
```

```
((Li kaj ŝi estas preskaŭ samaĝaj)) *
```

```
= (彼 と 彼女 は ほとんど 同年配 です。)
```

```
((Ŝi estas tute naiva)) *
```

```
= (彼女 は 全く 無邪気 です。)
```

感嘆文の例

```
((Kia bela floro ĝi estas !))
= (それは なんて すばらしい 花 でしょう。)
```

疑問文の例

```
((Kio estas ŝia laboro ?))
= (彼女の 職業 は 何 ですか。)
```

```
((Ĉu la fabliko ĉiam akceptas vizitantojn ?)) *
```

```
= (その 工場 は 訪問者 を いつも 受け入 れますか。)
```

```
((Kion vi havas en tiu skatolo ?)) *
```

```
= (あなた は その 手箱 の中に 何 を 持っ ていますか。)
```

Fig.12(1) 訳文例 (*印は阪氏⁷⁾の作文テキストから)

いくつかの例では、構文解析木を出力している。最初の例を見やすく書いたものが Fig.13 である。ここには比較のためエスペラントに対応する英文も示した。また * つきの文は阪氏のエスペラント作文テキスト⁷⁾から借用したものである。

エスペラントでは字上符(supersigno, Ĉ, Ŭ など)の問題があるが、入力エスペラント文の NLP(日本語ラインプリンター)への出力では、字上符付きの文字を出力できるようにした。

5. おわりに

以上、エスペラントから日本語への翻訳については、エスペラントの特長を生かして作った簡単なシステムで、比較的よい訳文が得られることを説明した。ヨーロッパとくに東欧圏で研究されているエスペラントを出力する自動翻訳システムと組み合わせれば、たとえばハンガリー語とかポーランド語などから日本語への翻訳が容易

に行えるであろう。

ところで日本語からエスペラントへの翻訳については、日本語の解析が複雑であるが、すでに日英翻訳などで十分に研究されている日本語の言語処理の手法を借用することにすれば、エスペラント文の生成については非常に容易にできるはずである。従来われわれが検討してきた日本語からエスペラントへの翻訳プログラムを、この立場から大幅に作り直す作業が、現在進行中である。また翻訳システムの記述に用いた処理系、FACOM OS IV LISP は、たいへん古典的な処理系なので、UTILISP (University of Tokyo Interactive LIST Processor) への移行を考えている。

謝 辞

この研究の内容について、ご親切なご助言を賜った大阪府立大学工学部・西田富士夫教授に厚く御礼申し上げます。また本学情報処理センターの利用に種々便宜を与え

倒置の例

```
((Li ĉiam povis alparoli al ŝi en la biblioteko))
= (彼 は その 図書館 の中で 彼女 に いつも 話しかけ る ことが でき た 。)

((En la biblioteko ĉiam li povis alparoli al ŝi))
= (その 図書館 の中で 彼 は 彼女 に いつも 話しかけ る ことが でき た 。)

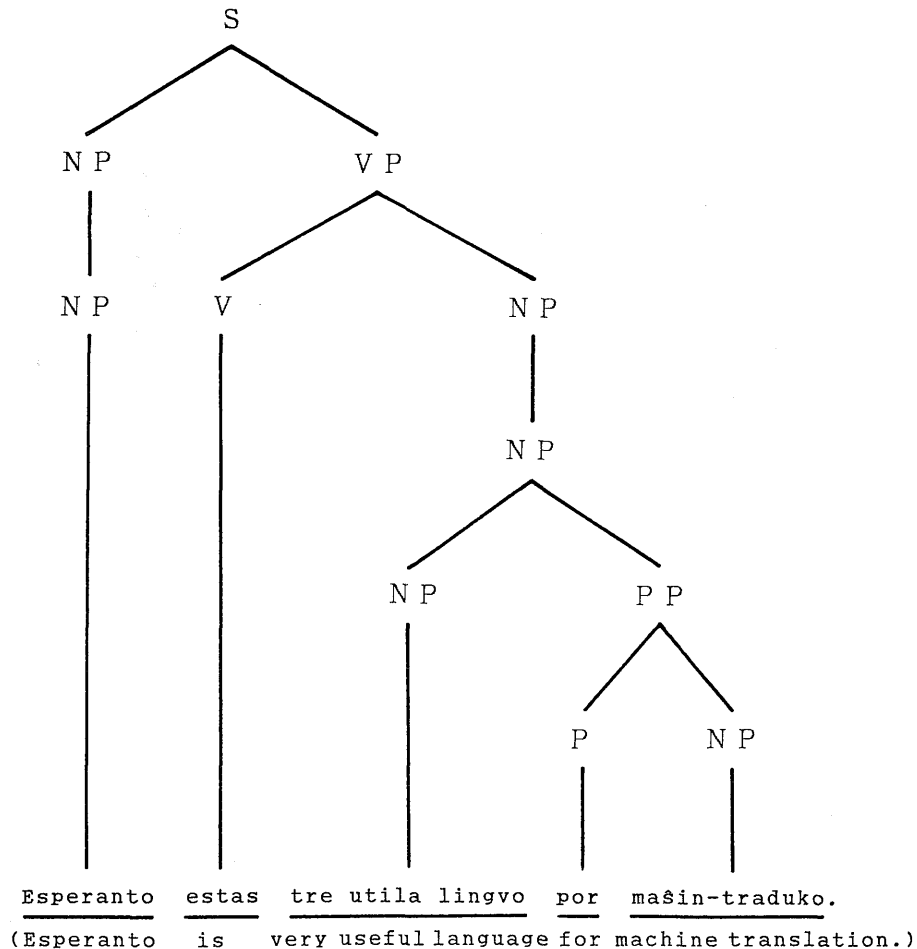
((En la biblioteko al ŝi li ĉiam povis alparoli))
= (彼女 に その 図書館 の中で 彼 は いつも 話しかけ る ことが でき た 。)
```

重文・複文の例

```
((Roza kaj Maria estas fratinoj ,))
(( kaj iliaj gepatroj estas instruistoj))      *
= S ==== NP ==== ? ==== (ROZA)
                        = CNJ == (KAJ)
                        = ? ==== (MARIA)
= VP ==== V ==== (ESTAS)
                        = NP ==== NP ==== (FRATINOJ)
= S ==== CO-CNJ (KAJ)
= S ==== NP ==== NP ==== (ILIAJ GEPATROJ)
                        = VP ==== V ==== (ESTAS)
                        = NP ==== NP ==== (INSTRUISTOJ)
= (ROZA と MARIA は 姉妹 です 。 そして (彼, 彼女)らの 両親 は 教師 です 。)

((Ĉu mi povas legi tiun leteron de via amiko kiu ĵus alvenis ?)) *
= S ==== NP ==== NP ==== (MI)
= VP ==== V ==== (POVAS LEGI)
= NP ==== NP ==== NP ==== (TIUN LETERON)
                        = PP ==== P ==== (DE)
                        = NP ==== (VIA AMIKO)
= NP-REL (KIU)
= S ==== NP ==== ADVP = (ĴUS)
= VP ==== V ==== (ALVENIS)
= (私 は たった今 着 いた あなたの 友達 の その 手紙 を 読 む ことが でき ますか 。)
```

Fig.12(2) 訳文例 (*印は阪氏⁷⁾の作文テキストから)



エスペラント は 機械翻訳 にとって たいへん 有用な 言語 です。

Fig.13 例文の解析木 (比較のため, エスペラントに対応する英語の単語を括弧内に示す。不定冠詞は省略)

て頂いた, 同センター長・中村嘉平教授, 同センター主任・水島章次助教授および同センター職員各位に深く感謝する。

参 考 文 献

- 1) 牧野 悟, 平田 満, 勝守 寛: エスペラントを仲介言語とする機械翻訳, 情報処理学会, 自然言語処理研究会資料 54-9 (1986).
- 2) 勝守 寛, 福田 基: エスペラントを仲介言語とする機械翻訳の試み, 情報処理学会, 自然言語処理研究会資料 44-6 (1984). 1981~1983 年度中のわれわれの研究報告はこの文献に示してある.
- 3) H. D. Maas: Aŭtomata Tradukado en kaj el Esperanto, Lingvo-Kibernetiko, Tübingen, p.75 (1982).
- 4) A. P. M. Witkam: Distributed Language Translation, BSO, Utrecht (1983).
- 5) 勝守 寛, 平田 満, 牧野 悟, 名知 克頼, 若山 淳一: 電気関係学会東海支部連合大会論文集 p.395 (1985).
- 6) T. Winograd: Understanding Natural Language, Academic Press Inc. (1972). または 洲 一博(ほか訳, 言語理解の構造, 産業図書 (1976) p.97 図 20 を参照.
- 7) 阪 直: エスペラント初級・中級の作文, 日本エスペラント学会 (1984).