

エスペラントを仲介言語とする機械翻訳 II

牧野 悟, 川出和秀, 竹内健二, 土井岳人, 西尾伸也, 山盛淳子, 勝守 寛
中部大学・工学部・工業物理学教室

数年来エスペラントを仲介言語とする多種言語間の機械翻訳の可能性を検討してきたが、その一環として、LISPで書かれた、日本文からエスペラント文への翻訳プログラムについて報告する。これはN進木拡張LINGOLによって日本語を解析し、その結果から対応するエスペラント文を生成する機械翻訳システムである。エスペラントの文法構造は完全に規則的であり、語彙論的にあいまいさが少ないことなど、その特長を生かして比較的簡単で効率のよいエスペラント文の生成システムができることを説明する。

コンピュータから出力したままの、後処理のしていない訳文例をいくつか示し、単純なシステムの割には良質の訳文が得られることを述べる。

A Machine Translation System by the Use of Esperanto as a Bridge Language II(in Japanese)

by Satoru MAKINO, Kazuhide KAWADE, Kenji TAKEUCHI, Takehito DOI, Shin'ya NISHIO, Junko YAMAMORI, and Hiroshi KATSUMORI (Department of Engineering Physics, Chubu University, Kasugai, Aichi 487 Japan)

A possibility of constructing a multilingual machine translation system by the use of Esperanto as a bridge language has been investigated since 1980. A machine translation program written in LISP, which transfers Japanese into Esperanto, is proposed.

Japanese sentences are analyzed by means of the N-ary Extended LINGOL. As Esperanto has completely regular structure and is lexically less ambiguous, Esperanto sentences can be easily generated.

Several examples of output sentences without post-editing show rather good quality of translation, considering a simple level of the system.

1. はじめに

われわれの研究室では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。1) 2) 3) 本報告ではN進木拡張LINGOL^{4) 5)}による日本語文の解析結果からエスペラント文を生成する翻訳システムについて述べる。

われわれが目指すものは、Fig. 1 に示すようなシステムに依存する内部構造を用いる中間言語方式ではなく、システム外へ取り出した場合にも自然言語として通用するエスペラントをFig. 2のように異なったシステム間の橋渡しとして利用しようとする多種言語間の翻訳システムである。

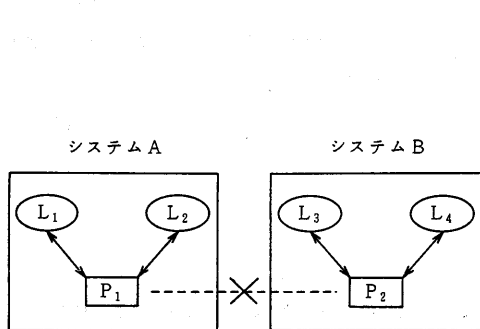


Fig.1 システムに依存する中間言語の問題

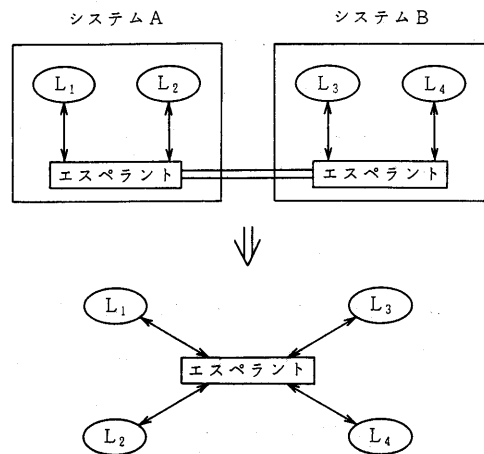


Fig.2 エスペラントを仲介言語とするシステム

エスペラント文の生成を主として説明するため、エスペラント文法の要点をまとめたものを、付録に示す。⁶⁾

エスペラントの語彙論的特長は、単語中に文法情報（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、相が規則的に決定している。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、また単語に多義性が少ないことも特長である。

エスペラントは、膠着言語なので、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ことは文献3)で述べた通りである。この報告では、エスペラント文の生成において、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との特長を強調する。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ことも大きな利点である。

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、ほとんど例外なく通用する。語彙論的にも品詞が確定的であり、多義性が少なく、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と云える。

2. 日本語からエスペラントへの翻訳システム

本システムは、中部大学情報処理センター FACOM M-360AP 上に U t i L i s p で実現している。処理は、日本語解析とエスペラント文の生成に別れている。システム構成を Fig. 3 に示す。

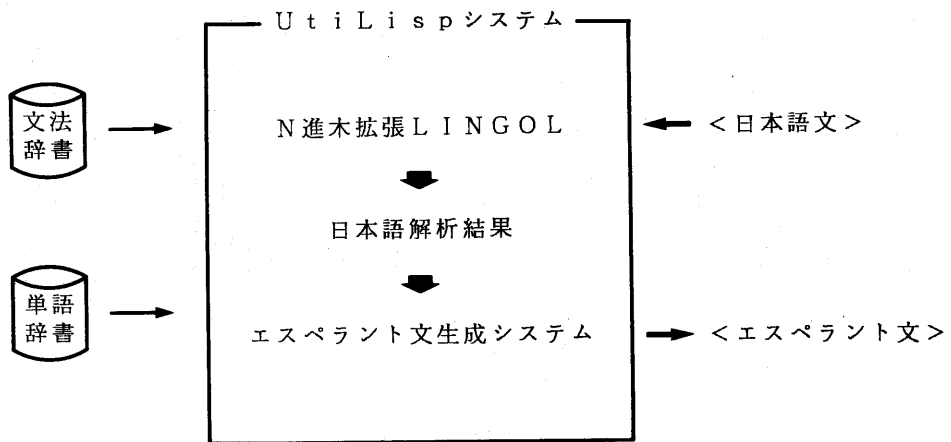


Fig. 3 システム構成

(1) 日本語解析

本システムの日本語解析部分は、拡張LINGOLを利用している。われわれの研究ではエスペラント文の生成に重点をおいているので、拡張LINGOLによる解析結果だけに興味を持ち、日本語解析の問題については議論しない。

なおN進木拡張LINGOLは Mac L i s p で記述されているため、利用に先立ち、Mac L i s p から U t i L i s p への移植が必要であった。

日本語解析に用いた文法規則の例を Fig. 4 に示す。また辞書の記述例を Fig. 5 に示す。

(2) エスペラント文の生成

エスペラント文の生成は、データ駆動によって行われる。拡張LINGOLによる構文解析で得た句構造標識の範疇記号から Fig. 6 に示すような生成規則を持つ関数を呼び出し、語順の変換と訳語生成を同時に行う。

例えば、「水素の原子は」という句の場合、拡張LINGOLに解析結果として
 (CASE ((CASE (水素 の)) (PP (原子 は))))

が得られ、各部を処理すると (hidrogeno (de)), (atomo) となり、これらの語順を入れ換え、前置詞「de」によって接続して (atomo de hidrogeno) となる。

また派生形を求める場合、「有名な」を例に上げると、解析結果
 (ADJ34 ((NP ((NOUN 有名))) (NA NA)))

から NP の「有名」を 에스ペラント「famo」に置き換え、これを形容詞の形「fama」に変換する。

多くの自然言語では、単語の派生形を作り出すのに膨大なデータを必要とするが、エスペラントでは不規則変化がないので基本的な規則だけでよい。

| | | | | | | | | | | |
|----------|---|----------|---|---------|--|-------|---|-------|---|-------|
| BUN | → | SENTENCE | + | END | | NOUN | → | NOUN | + | NOUN |
| SENTENCE | → | NPP | | VPP | | | | NOUN | + | LABEL |
| | | VPP | | | | | | NOUN | + | QUANT |
| NPP | → | CASE | | | | NPOS2 | → | NPOS2 | + | NOUN |
| VPP | → | CASE | | VP34 | | | | QUANT | + | NO |
| | | CASE | + | NP | | VHEAD | → | LABEL | + | NO |
| | | CASE | | AP | | | | VHEAD | + | SAUX |
| | | VP34 | | | | | | VHEAD | + | RAUX |
| | | NP | | | | | | VHEAD | + | WA |
| | | AP | | | | | | VHEAD | + | A |
| CASE | → | CASE | + | PP | | | | VHEAD | + | I |
| | | PP | | | | | | VHEAD | + | T |
| NP | → | NLNK | + | NP | | | | VHEAD | + | N |
| | | AP | + | NP | | | | NOUN | + | SHI |
| | | NLNK1 | + | NP | | COND5 | → | NOUN | + | SU |
| | | NLNK | + | FNOUN | | | | NOUN | + | FRONT |
| | | DET | + | NOUN | | VBASE | → | BACK5 | + | PML |
| | | NOUN | | | | | | NOUN | + | BASE |
| | | QUANT | | | | SAUX | → | SAUX | + | RUAX |
| | | FNOUN | | | | | | ADJ34 | + | ADJ34 |
| VP34 | → | VHEAD | + | BACK3 | | AP | → | PP | + | ADJ34 |
| | | VHEAD | + | BACK12 | | | | ADJ34 | | |
| | | VHEAD | + | BACK133 | | PP | → | NP | + | P |
| | | VHEAD | + | BACK233 | | | | NP | | |
| | | VHEAD | + | BACK333 | | | | | | |

Fig. 4 文法規則の例

| | | | |
|---------------|---------|-------------|-----------------------|
| (RU | BACK3 | ((RU) 0) | NIL) |
| (NAKATTA | BACK12 | ((TE-TA) 0) | NIL) |
| (E | BACK5 | ((E) 0) | NIL) |
| (NARA | COND34 | ((NARA) 0) | NIL) |
| (U | FRONT | ((U) 0) | NIL) |
| (MANA | VHEAD | (NIL 0) | LERNI) |
| (YUUMEI | NOUN | (NIL 0) | FAMO) |
| (BANYUUNRYOKU | NOUN | (NIL 0) | (UNIVERSALA GRAVITO)) |
| (GA | P | (NIL 0) | NIL) |
| (WA | P | (NIL 0) | NIL) |
| (WO | P | (NIL 0) | NIL) |
| (SASE | SAUX | (NIL 0) | NIL) |
| (RARE | RAUX | (NIL 0) | NIL) |
| (RE | RAUX | (NIL 0) | NIL) |
| (KIREI | ASP | (NIL 0) | BELA) |
| (AKA | ADJ | (NIL 0) | RUĜA) |
| (MIGI | FNOUN | (NIL 0) | DEKSTRO) |
| (TO | NPARA | (NIL 0) | KAJ) |
| (CM | UNIT | (NIL 0) | (CM (DE)) |
| (KOUDO | PREUNIT | (NIL 0) | ALTO) |

Fig. 5 辞書の記述例

| | | | |
|------|---|---------|--|
| S | <p>名詞句、動詞句のそれぞれの処理を行った生成結果から文を構成する。</p> <p>文の構成要素が名詞句、動詞句の一方だけの場合は、全体を動詞句として処理する。また、"それは赤い"のような属辞（形容詞、名詞句あるいは前置詞句）とともに述語を構成する場合は、"esti"によって名詞句と動詞句を連結する。</p> | VP 3 4 | <p>VP 3 4の生成処理は次の場合に分けられる。</p> <ul style="list-style-type: none"> * VP 3 4→VHEAD+BACK 3 * VP 3 4→VHEAD+BACK 1 2 * VP 3 4→VHEAD+BACK 1 3 * VP 3 4→VHEAD+BACK 2 3 * VP 3 4→VHEAD+BACK 3 3 <p>BACK 3, BACK 1 2, BACK 1 3, BACK 2 3, BACK 3 3は、それぞれ現在形、過去形、否定の過去形、否定の現在形、未来形の語尾である。</p> <p>VHEADの生成処理で得られた動詞の原形をこれらの語尾に応じて変化させVP 3 4の生成結果とする。</p> |
| NPD | <p>文のトップレベルの構成要素である名詞句は、主格を表わす助詞を含んだCASE部分である。従って、CASEの生成処理を行うことによってNPDの生成をする。</p> | VHEAD | <p>VHEADの生成処理は次の場合に分けられる。</p> <ul style="list-style-type: none"> * VHEAD→VHEAD+SAUX <p>SAUXは使役を表わす語尾である。VHEADの生成処理で得られた動詞に使役を表わす接尾辞を付加して生成する。</p> <ul style="list-style-type: none"> * VHEAD→VHEAD+RAUX <p>RAUXは可能を表わす語尾である。VHEADの生成処理で得られた動詞に可能を表わす接尾辞を付加して生成する。</p> <ul style="list-style-type: none"> * VHEAD→VHEAD+WA <p>VHEAD+A VHEAD+I VHEAD+T VHEAD+N</p> <p>VHEAD部を辞書引きによって生成する。</p> <ul style="list-style-type: none"> * VHEAD→NOUN+SHI <p>NOUN+SU</p> <p>NOUNの生成処理で得られた名詞を動詞に変形して生成する。</p> |
| VPD | <p>エスペラントは通常目的語が動詞の後に現われるSVO言語なので、動詞句を構成する各部の生成結果の語順を入れ換えて生成する。</p> | ADJ 3 4 | <p>ADJ 3 4の生成処理は次の場合に分けられる。</p> <ul style="list-style-type: none"> * ADJ 3 4→ADJ+I <p>ADJ+NA ASP+NA</p> <p>辞書引きによってADJ 3 4を生成する。</p> <ul style="list-style-type: none"> * ADJ 3 4→NP+NA <p>NPの生成処理で得られた名詞を形容詞に変形して生成する。</p> |
| CASE | <p>CASEの生成処理は次の場合に分けられる。</p> <ul style="list-style-type: none"> * CASE→CASE+NP <p>前置詞句の生成を行う。</p> <ul style="list-style-type: none"> * CASE→PP <p>PPの生成結果をCASEの生成結果とする。</p> | | |
| NP | <p>NPの生成処理は次の場合に分けられる。</p> <ul style="list-style-type: none"> * NP→NLNK+NP <p>CASE→CASE+PPの生成処理と同じように生成する。</p> <ul style="list-style-type: none"> * NP→AP+NP <p>NPの中に助詞があれば、APをその助詞の後に、助詞がなければ、NPの先頭に付け加える。</p> <ul style="list-style-type: none"> * NP→NLNK 1+NP <p>NLNK 1の生成結果とNPの生成結果を結合する。</p> <ul style="list-style-type: none"> * NP→NLNK+F NOUN <p>F NOUNをNPに変換してNP→NLNK+NPの処理で生成を行う。</p> <ul style="list-style-type: none"> * NP→DET+NOUN <p>DET部は辞書引きによって得、NOUN部は生成処理を行い結合する。</p> <ul style="list-style-type: none"> * NP→NOUN <p>NOUNの生成結果をNPの生成結果とする。</p> <ul style="list-style-type: none"> * NP→QUANT <p>QUANTの生成結果をNPの生成結果とする。</p> <ul style="list-style-type: none"> * NP→F NOUN <p>辞書引きによってF NOUNを生成する。</p> | | |

Fig. 6 生成処理関数の例

3. 訳文例

訳文例を Fig. 7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。いくつかの例では構文解析結果も出力している。エスペラント訳文例としては、阪氏のテキストの水準を参考にした。⁷⁾

拡張LINGOLの日本語入力ではローマ字を用いるが、見やすくするため、上に漢字かな混じり文を付加した。

Fig. 7 訳文例

エスペラントは国際補助語である。
入力文 => ESPERANTOWAKOKUSAIHOJOGODEARU.
出力文 => (ESPERANTO ESTAS INTERNACIA HELPA LINGVO .)

エスペラントは機械翻訳の仲介言語である。
入力文 => ESPERANTOWAKIKAIHONYAKUNOCHUUKAIGENGODEARU.
出力文 => (ESPERANTO ESTAS PONTOLINGVO DE MASIN-TRADUKO .)

私は彼を訪問しなかった。
入力文 => WATASHIWAKAREWOHOUOMONSHINAKATTA.
出力文 => (MI NE VIZITIS LIN .)

ニュートンは有名な万有引力を発見した。
入力文 => NEWTONWAYUUMEINABANYUUIINRYOKUWOHAKKENSHTA.
(BUN ((SENTENCE
((NPP (CASE ((PP ((NP ((NOUN NEWTON))) (P WA))))))
(VPP ((CASE ((PP ((NP
((AP ((ADJ34 ((NP ((NOUN YUUMEI))) (NA NA))))))
(NP ((NOUN BANYUUIINRYOKU))))))
(P WO))))))
(VP34 ((VHEAD ((NOUN HAKKEN) (SHI SHI))) (BACK23 TA))))))
(END .)))
出力文 => (NEWTON ELTROVIS FAMAN UNIVERSALAN GRAVITON .)

私は彼にその問題に対して腹を立てた。
入力文 => WATASHIWAKARENISONOMONDAINITAISHITEHARAWOTATETA.
出力文 => (MI EKKOLERIS AL LI PRI LA PROBLEMO .)

彼の弟はニュートンの法則を学んだ。
入力文 => KARENOOTOUTOWANEWTONNOHOUSOKUWOMANANDA.
出力文 => (LIA FRATO LERNIS LEGON DE NEWTON .)

彼の弟は野球を公園の右の空地でする。
入力文 => KARENOOTOUTOWAYAKYUWOKOUEENNOMIGINOAKICHIDESURU.
(BUN ((SENTENCE
((NPP (CASE ((PP ((NP
((NLNK ((NOUN KARE) (NO NO)))
(NP ((NOUN OTOUTO))))))
(P WA))))))
(VPP ((CASE ((CASE ((PP ((NP ((NOUN YAKYUU))) (P WO))))))
(PP ((NP
((NLNK ((NOUN KOUEEN) (NO NO)))
(NP
((NLNK ((FNOUN MIGI) (NO NO)))
(NP ((NOUN AKICHI))))))
(P DE))))))
(VP34 ((VHEAD ((VBASE S) (FRONT U)) (BACK3 RU))))))
(END .)))
出力文 => (LIA FRATO LUDAS BASBALON EN DEKSTRA LIBERA PLACO DE PARKO.)

4. おわりに

以上、拡張LINGOLによる解析結果からエスペラント文を生成する翻訳システムについて、エスペラントの特長を生かすことによって、簡単で効率がよく比較的よい訳文を生成するシステムが作成できることを説明した。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入力する自動翻訳システムと組み合わせれば、たとえば日本語からハンガリー語とかポーランド語への翻訳が容易に行えるであろう。

エスペラントは、文法情報を文の表層に含み、また意味的要素もマーカとして表層に持つので、より深いレベルの解析結果をエスペラント文に反映させることができると考えられる。この立場からエスペラント文の生成システムを改良する作業が現在行われていて、近く発表する予定である。

この研究で日本文の解析に当って、工業技術院電子技術総合研究所・井佐原 均氏から、拡張LINGOLに関する多くの資料を提供して頂き、有益なご助言を頂いたことを厚く御礼申し上げます。

参考文献

- 1) 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.
- 2) 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 (1985) p. 395.
- 3) 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). 勝守 寛, 牧野 悟, 平田 満: 中部大学工学部紀要 22 (1986), p. 57
- 4) 井佐原 均, 本吉 文雄, 田中 穂積: N進木拡張LINGOLのユーティリティ関数について, 電子技術総合研究所彙報 第46巻 第12号 (1982).
- 5) 田中 穂積: 計算機による自然言語の意味処理に関する研究, 電子技術総合研究所研究報告 第797号.
- 6) たとえば, 三宅 史平編: エスペラント小辞典, 大学書林 (1965) pp. 435~448 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).

付録 エスペラント文法の要点

1. 不定冠詞はなく、定冠詞 la はあらゆる性・数・格に対して共通である。
2. 名詞の語尾は o であり、複数には j を付ける。格は主格・目的格だけで、目的格には n を付ける。
3. 形容詞の語尾は a であり、数・格は名詞にそろえる。

4. 基本数詞

1 unu 2 du 3 tri 4 kvar 5 kvin 100 cent
6 ses 7 sep 8 ok 9 naŭ 10 dek 1000 mil
十位、百位…の数は日本語と同じ。例 234 ducent tridek kvar

5. 人称代名詞

私 mi われわれ ni あなた vi あなたがた vi
彼 li 彼女 ŝi それ ĝi 自身 si 人々 oni
彼ら、彼女ら、それら ili
所有格は a を付ける mia 私の 目的格は n を付ける min 私を
他の格は前置詞で al mi 私に

6. 動詞は人称や数によって変化しない。動詞の語尾形式を次に示す。

現在時 as 過去時 is 未来時 os
仮定法 us 命令法 u 不定法 i

分詞（形容詞的または副詞的意義を持つ）は次のようである。

継続 完了 将然
能動 ant int ont
受動 at it ot

受動態は esti と受動分詞で作る。前置詞は de が用いられる。

7. 副詞には品詞語尾 e のある派生副詞と品詞語尾のないものがある。
8. すべての前置詞は主格をとる。移動の方向を示すときは、目的格をとる。
9. 合成語は単語の簡単な結合によって作る。