

オンライン翻訳の一つのアプローチ

福島正俊 有田英一

(三菱電機(株) 中央研究所)

1. はじめに

計算機による自然言語処理の研究が進み、機械翻訳は実用化を検討する時期になった。実用化へのアプローチとしては2つの流れがある。

(1) 翻訳する対象を十分に小さくするアプローチ [長尾 他 1981]

(2) 原言語の文法を制限するアプローチ [長尾 1983] [吉田 他 1983]

これらは人間の介入なしに翻訳を行うことを想定しており、翻訳をする原言語の表現が入力されてから後の処理を中心に議論されている。我々は、翻訳における人間の役割も考慮した実用的な総合システムとして、機械翻訳システムを検討している。ある程度広い対象(一般文)が翻訳可能であり、システムの受理できる文について事前に教育をする必要のない、一般ユーザー向けの機械翻訳システムの開発を目的として、例文を利用した英文レター作成援助システムを作成している。

2. 例文を利用した機械翻訳のアプローチ

我々は機械翻訳を、人間と計算機がその能力に従って役割を分担することにより、高い機能を実現するマン・マシンシステムの1つと考えている。実用的な機械翻訳システムを作るためには人間の介入が不可欠であるからである。人間が積極的な役割をする機械翻訳システムの設計に当たって次の点を考慮した。

(1) 高度な意味処理は計算機には難しい。

概念変換という言葉に代表される高度な意味解析は、実験システムで研究が進められているが、現在のところ、実用システムで使える段階に達していない。特に、一般文を翻訳の対象として考える場合、システムが持つ文法モデルの意味の体系を矛盾なく作ることは難しい。

(2) 機械翻訳では目標言語での表現を意識して原言語を記述する必要がある。

現在、多くの機械翻訳システムで採用されているトランスファー方式においても目標言語での表現を意識して原言語で記述しなければ良い翻訳文は得られない。たとえば、

<1> 彼はテンポがのろいので一緒に話をすると調子が狂ってしまう。

という原言語での自然な表現から、

<2> He is so slow-paced that our conversations always seem disjointed.

([ドナルド 他 1982] より引用)

という目標言語での自然な表現をトランスファー方式で得ることは難しい。<2>の表現を得るためには、

<3> 彼はテンポがのろいので我々の会話はいつも調子はずれているようだ。

という目標言語の発想法での原言語による表現が必要である。

(3) 制限文法はユーザーに負担がかかる。

制限文法によるアプローチは入力文の構文上、意味上のあいまいさを減少させることを目的としており、機械翻訳の立場からすれば必要なアプローチである。翻訳することを前提とした大量の技術文書を書くことを仕事とする人の場合は、事前に制限文法を学習するのが効果的である。しかし、一般の人が手軽に翻訳文を作る場合には、制限文法を学習するのはめんどろである。事前の学習なしに翻訳システムの受理できる文の書き方がわかるようにシステムはサポートする必要がある。

(4) 母国語から外国語への翻訳は必ずしも表層文をそのまま翻訳する必要はない。

特に翻訳する内容を考える人と翻訳する人が同じである場合は、表層文を手がかりに翻訳する必要はなく、自分の表現したい意図から直接翻訳してもよい。

(5) 一般の日本人は英語が理解できる。

一般の日本人は自分で考えだせる英文よりも、読めば理解できる英文のほうがはるかに多い。そして例文のうまい表現を利用して、その一部分を置換えることにより自分の表現したい内容を表す英文を作ることができる。従って良い例文を見つけることができれば、英語らしい英文を書くことができる。

以上の点を考慮して、本報告では例文を利用した機械翻訳のアプローチ（*1）を提案する。図1にその概念図を示す。

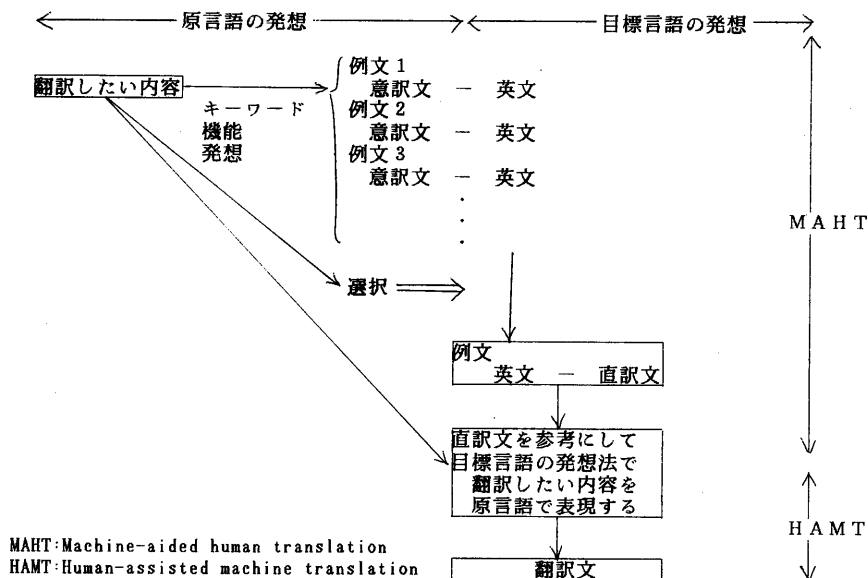


図1 例文を利用した機械翻訳のアプローチ

*1 [長尾 1982]では、例文とその訳文及びシソーラスを利用することにより新しい機械翻訳システムができる可能性を示唆している。我々の方法ではシソーラスに相当する部分を機械で処理せず人間が判断して行っている。

- (1) まず、翻訳したい内容を表現するのに適切な例文を検索する。表層文を意味解析するのではなく、表現したい内容のキーワード、機能、発想を手がかりに、利用できる可能性のある例文を捜す。
- (2) 検索された複数の例文の中から適切な例文を選択する。この時は、英文（目標言語の発想での表現）と共に表示される意訳文（原言語の発想での表現）を見ながらユーザーが判断する。
- (3) 選択された英文（目標言語の発想での表現）と直訳文（目標言語の発想での表現）をシステムは表示する。
- (4) ユーザーは直訳文を参考にして自分の表現したい内容を原言語で入力する。これは目標言語の発想法で、翻訳したい内容を原言語で表現していることになる。例文の直訳文を見ながら原言語で入力するので、ユーザーは、システムの受理できる文について事前の学習は必要ない。
- (5) システムは入力された文を翻訳する。この時、例文の構文パターンを参考にするので、パーザーの能力はあまり高くなくてもよい。

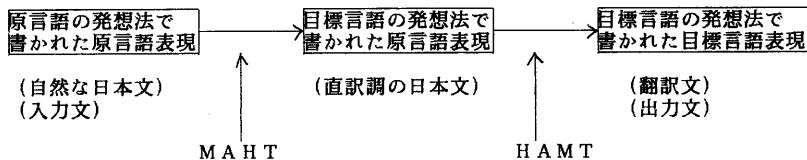


図2 2段階の翻訳

図2に示すように、我々は翻訳を

- I (自然な日本語) → (直訳調の日本語)
(原言語) (原言語)
- II (直訳調の日本語) → (翻訳文)
(原言語) (目標言語)

の2段階で行うことを考えている。

Iは、高度な意味解析が必要なので人間が行う。人間が主で機械が従のMAHT (Machine Aided Human Translation)である。IIは、計算機がトランスファー方式で行う。機械が主で人間が従のHAMT (Human Assisted Machine Translation)である。従来から発表されている機械翻訳システムのプリエディット処理はIに対応するが、そこでは、どのようにエディットすればよいかについては、ほとんど計算機側のサポートがなかった。本報告のアプローチでは、例文の直訳文を表示するという形でサポートしている。

ここで問題となるのは、自分の言いたいことを英語で表現する際に、利用できる英文をいかにして効率よく検索するか、ということである。従来の例文の検索の方法には、

- (1) キーワード (動詞、名詞などを中心に検索する)
- (2) パタン (SはIOにDOをVする ↔ S V IO DO)
- (3) 状況 (訪問の文例、紹介状の文例 etc.)
- (4) 機能 (依頼、提案、勧誘 etc.)

による方法がある。しかし、これらの方法では英語らしい表現を得るには不十分な場合がある。例えば、

<4> 塩を入れると味がよくなる。

の自然な英文は、

<5> Salt will improve the taste.

であるが、上記の(1)～(4)の方法では、「塩を入れること」を主語にして英文を作るという名詞構文の発想法は得られない。

我々は原言語の発想法で書かれた原言語表現と目標言語の発想法で書かれた原言語表現との対応をつけることにより、目標言語での自然な表現を得られるようにしている。現在のところ、原言語の表層文の付属語の系列に注目して対応づけを行っている。下にその例を示す。

<タイプ>

～の～は～

- 1 <日本文> 君の将来は明るい
<英文> You have a bright future.
<直訳文> 君は明るい将来を持っている
- 2 <日本文> 物語の幕切れはあっけないものだった
<英文> The story ended up abruptly.
<直訳文> 物語はあっけなく終わった

～は～が～

- 1 <日本文> 私は足が長い
<英文> I have long legs.
<直訳文> 私は長い足を持っている
- 2 <日本文> 彼女は仕事が丁寧だ
<英文> She works carefully.
<直訳文> 彼女は丁寧に仕事をする

この対応づけにより、ユーザーは適切な例文を検索することができる。

このアプローチは、手紙文のように(1)一度に大量の文書は扱わない、(2)個人が書くことが多い、(3)翻訳する対象を十分に小さくすることができない、(4)きまり文句を使う部分もある場合には有効である。

3. 例文を利用した英文レター作成援助システム

3.1 システムの概要

図3にシステムの処理の概要を示す。まず、どのような内容の手紙を書くのかを選択する。「訪問に関する文例」、「依頼に関する文例」、「招待に関する文例」等のメニューの中から選択する。それらは必要に応じて細分類されており、自分の表現したい内容に最も近い文例を捜す。その文例をもとに必要箇所の修正を行い、目的とする手紙を作成してゆく。修正処理には(1)置き換え処理、(2)例文検索、(3)例文を利用した日英翻訳がある。

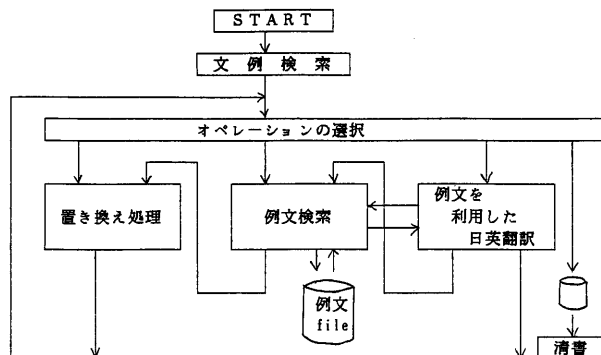


図3 例文を利用した英文レター作成援助システムの処理の概要

3.2 置き換え処理

置き換え処理は、例文の骨格パターンをそのまま利用し、その格要素をユーザーの表現したい内容に置き換える。訳文は英語の発想で表現してあるので日本語と英文の対応が明確であるから、句、節の単位の置き換え処理が容易になる。図4に置き換え処理の例を示す。2重下線の部分がユーザーの入力である。

- ① I have been engaged in the projection of city planning in consulting civil engineering firm of Osaka .
- ② 大阪の土木コンサルタント会社で、私は都市計画のプロジェクトに従事している。
- ③ 従事する 仕事 は 都市計画のプロジェクト です
- ④ 従事する 仕事 は : ((KIKAIHONYAKU) NIKANSURU (KENKYUU) DESU)
- ⑤ 機械翻訳にかんする研究です
|ok ?| Y
- ⑥ I have been engaged in a study on machine translation in consulting civil engineering firm of Osaka .
- ⑦ 大阪の土木コンサルタント会社で、私は機械翻訳にかんする研究に従事している。

図4 置き換え処理の例

①、②は例文及びその訳文である。これらは予め解析されて内部表現で表わされているものを合成して出力している。③はこの例文の中心動詞<従事する>の<仕事>格が「都市計画のプロジェクト」であることを示している。④はその<仕事>格の内容を換えるかどうかを、システムがユーザーに問い合わせてきたのに対して、ユーザーは自分の表現したい内容「機械翻訳に関する研究」に修正していることを示している。⑤は入力がローマ字であるので、まちがいがいなか漢字かな混じり文で表示して確認している。⑥、⑦は入力された修正情報をもとに例文の内部構造を変換し、その結果を合成したものである。

また、置き換え処理では、例文にない任意格の追加、不要な任意格の削除ができる。

3.3 例文の分類と検索

例文には「機能」を表すための例文と「発想の違い」を表すための例文がある。機能を表す例文は、たとえば「相手に何かを要求したい」という内容を表す時の例文であり、「発想の違い」を表す例文は、英

語の表現と日本語の表現の発想が違う時に使う例文である。

例文のもつ機能は予め人間が判断しておく。

We would appreciate it if you would ～ 。

I would find it extremely helpful if ～ 。

I will be grateful if ～ 。

のように直訳文に「依頼」という文字がなくても、これらの文は「依頼」の機能をもつ文として分類される。機能の分類項目の例を表1に示す。

表1 機能の分類項目の例

(1 1):書きだし	(5 13):祝賀	(5 31):紹介
(1 2):結言	(5 14):許可	(5 32):照会問い合わせ
(2 1):進捗状況	(5 15):後悔	(5 33):推薦
(3 1):時に關する表現	(5 16):おわび言いわけ	(5 34):勧誘
(3 2):數量に關する表現	(5 17):予想予測	
(4 1):話題特有表現	(5 18):通知	
(5 1):提案提示	(5 19):意志自決意向	
(5 2):困難不可能	(5 20):おくやみ	
(5 3):提供	(5 21):予定計画	
(5 4):助言忠告示唆	(5 22):確信判断	
(5 5):希望要望期待	(5 23):命令要求指示	
(5 6):依頼要求	(5 24):反論反对	
(5 7):願望欲求	(5 25):感覚感情	
(5 8):感謝お礼	(5 26):必要義務	
(5 9):拒否不都合断り不可能	(5 27):可能	
(5 10):受諾承諾同意	(5 28):能力	
(5 11):催促	(5 29):感嘆	
(5 12):喜び楽しみ	(5 30):興味関心	

日本語と英語との間の発想の違いは前節で述べたように多々ある。たとえば、「原因+結果」の英語表現が表2で示す単語を用いて表現可能である。表3に日本語の発想に基づく検索キーの一部を示す。

例文は、

- (1) キーワード
- (2) 機能
- (3) 発想

で検索が可能である。検索キーに対応して、システムが表示する例文の候補の中からユーザーが選択する。例文には、「この例文は公式の場では用いられない」のようなコメントや、その例文が使われている文例(コンテキスト)へのポインターをつけてあるので、例文の選択の時に参照できる。

表2 「原因+結果」を表現する単語の例

for	because of	on behalf of
at	with	by
due to	on account of	through
because	as	since
so	on the ground that	for reasons that
too ~ to		lead to
is the result of	result in	result from
cause	lose	end in
bring about	turn out	prevent
keep	stop	block
come	cost	concur

表3 日本語の発想に基づく検索キーの例

クラス	タイプ	クラス	タイプ	
叙述	～を～スル	原因+結果	～スルから～スル	
	～は～スル		～で～スル	
提題+説明	～が～スル	対比	～スルので～スル	
	～は～が～スル		～が原因で～スル	
	～は～スル		～シタ結果～スル	
	～では～スル		～ため(に)～スル	
	～には～スル		～により～スル	
	～については～スル		～タことから～	
	～は～タもので、		～なので～スル	
	～シテ～スル		～けれども～	
	～は～と違って、		～が～	
	～スル		～(運用中止), ～	
一般条件+文	～スルと～スル	状況+文	～シテ～スル	
逆接接続+文	～スレバ～スル	手段+文	～場合(は),	
	～が, ～スル		～スル	
	～けれども～スル		～において(は)～	
順接接続+文	～からといって	比較	～スル時(は)～	
	～ナイ		～ように, ～	
	～シナイで～スル		目的+文	～スルことによって
	～スルのに～スル			～により～
	～が, ～スル			～シテ～
～スルと, ～スル	～スルと～	並列	～より(も)～	
～シテ～スル	～スルには～			
～シテから～スル	～(運用中止), ～	その他	～スルため(に)～	
～(運用中止), ～	～で, ～			
逆接確定条件+文	～シテも～スル	その他	～(運用中止), ～	
逆接仮定条件+文	～シナイと～		～で, ～	
順接仮定条件+文	～タとしても～		～は～が～	
	～スルと～スル	～の～は～		
	～なら～スル			
	～スレバ～スル			
	もし～の場合には～			

3.4 例文を利用した日英翻訳

図5に処理のフローを示す。まず、3.3に示した方法で例文を検索する。システムは例文とその直訳文を表示するので、ユーザーは表示された日本語に従って(助詞、助動詞、キーワードの言い回しを真似て)自分の表現したい日本語を入力する。システムは、例文の直訳文と入力された日本語とのパターンマッチングにより日本語を分割してから、トランスファー方式で翻訳を行ってゆく。係受け関係や意味が解決できない時は、図6、図7のようにユーザーに表示し、選択してもらう。必要があれば、その後置き換え処理を行う。

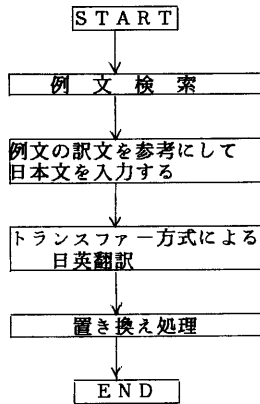


図5 例文を利用した日英翻訳

私はあなたに表紙の赤い本をあげる
 1: 私はあなたに表紙の赤い本をあげる
 2: 私はあなたに表紙の赤い本をあげる

番号を入れて下さい :

図6 係受け関係の問い合わせ

「機械翻訳の技術」の「の」は次のどの意味ですか

- 1: 目的 (外出の支度)
- 2: 対象 (命令の実行)
- 3: 同格 (息子の太郎)
- 4: 連体修飾

番号を入れて下さい :

図7 助詞の意味の問い合わせ

4. おわりに

実用的な機械翻訳システムを作るために、

- (1) 事前にシステムの受理できる文法を学習することなく目標言語を意識した原言語を入力できる
- (2) 例文を参考にすることにより英語らしい表現を生成する

アプローチについて述べた。

今後はマン・マシンシステムとして、計算機が何を知っていて、何がわからないのか、を人間にうまく問い合わせることができる機械翻訳システムを作りたいと考えている。

参考文献

- [長尾 他 1981] 長尾 真, 辻井潤一, 矢田光治, 柿元俊博:
 「技術論文表題の英和自動翻訳を利用した文献速報システム」
 第23回情報処理学会全国大会 (昭和56年後期) p829
- [長尾 1982] 長尾 真: 「機械翻訳」 電子通信学会誌 Vol 65 No. 4 (1982) p386-391
- [長尾 1983] 長尾 真: 「制限言語の試み」
 情報処理学会 自然言語処理シンポジウム論文集 (1983) p91-99
- [吉田 他 1983] 吉田 将, 田中武美, 松山晶子:
 「科学技術文書のための日本語文法の制限について
 —付属語的表現の語いおよび語義の制限を中心として—」
 情報処理学会 自然言語処理研究会資料 37-4
- [ドナルド 他 1982] ドナルド・キーン/羽島博愛 監修
 「会話作文英語表現辞典」 朝日出版社 (1982)