解　説

実例に基づく翻訳†

佐藤　理史††

1. はじめに

計算機の出現とともに始まった機械翻訳の研究は，1980年代末，多数の商用機械翻訳システムを生み出すまでに至った．しかし，機械翻訳を実現する技術はまだまだ不完全であり，以下のような問題を抱えている．

●翻訳の品質　現在の機械翻訳システムは，翻訳作業を完全に任せられるような，高品質な翻訳ができるわけではない．通常は，対象とする文章の種類や分野を限定することによって，品質をあげようとするが，それだけでは不十分であり，機械が翻訳しやすいように前編集したり，機械が出力した結果を素訳として，後編集したりする必要がある．

●システムの構築と改良　現在の機械翻訳システムは，解析・変換・生成のための膨大な規則集合と辞書を必要とするが，その作成は，人手に頼っている．そのため，システムの構築に膨大な労力が必要である．また，そのような膨大な規則集合や辞書を改良していくのは，非常に困難な作業である．

つまり，機械翻訳は，完成した技術ではなく，まだまだ「難しさ」と「おもしろさ」を秘めた研究対象なのである．

本稿で紹介するのは，近年，機械翻訳の分野で現れてきたいくつかの新しい研究である．これらの研究は，ある一つの方向性をもっており，新しいパラダイムを形成しつつある．その中心的な考え方を一言で言ってしまえば，

抽象化された「規則」に頼るのでなく，むしろ，豊富な実例／用例を積極的に利用しようということになるだろう．おそらく，この考えは，自然言語という，個々の単語の個別性と膨大な多様性をもった対象に対しては，少数の抽象化された規則集合を用いるよりも，豊富な実例や用例を用いたほうが有効であろうという認識からきているのだろう．これをここでは「実例に基づく翻訳 (Example-Based Translation, EBT)」と呼ぶことにしよう*．もちろん，このパラダイムは，いま，まさに，黎明期であり，統一した用語や見解がないまま，個々の研究が進められている状態である．したがって，本稿は，解説というよりは，筆者の個人的見解による交通整理と考えていただきたい．

本稿の構成は，以下のとおりである．まず，**2.** で，EBT の基本コンセプトについて整理し，**3.** では，EBT を実現する要素技術について簡単に述べる．**4.** では，現在までに発表されている各種の EBT システムについて簡単に紹介する．**5.** では，自然言語処理，および，人工知能研究における EBT の位置付けを中心に，EBT を研究する「意義」について筆者なりに考えてみる．

2. 基本コンセプト

EBT は，長尾の論文 "A Framework of a Mechanical Translation between Japanese and English by Analogy Principle"[1] に端を発する．6頁のこの短い論文には，数多くのアイディアが詰まっている．

EBT の基本的なアイディアは，以下のように言い表すことができるであろう．

ある文を翻訳することを，それとよく似た文の翻訳例を見つけ，それを模倣することによって行う．

例を用いて，具体的に説明しよう．今，あなたは，

† Example-Based Translation by Satoshi SATO (School of Information Science, Japan Institute of Science and Technology, Hokuriku).

†† 北陸先端科学技術大学院大学情報科学研究科

* このほかに，Translation by Analogy, Memory-Based Translation, Example-Based Machine Translation, Analogy-Based Machine Translation などと呼ばれている．

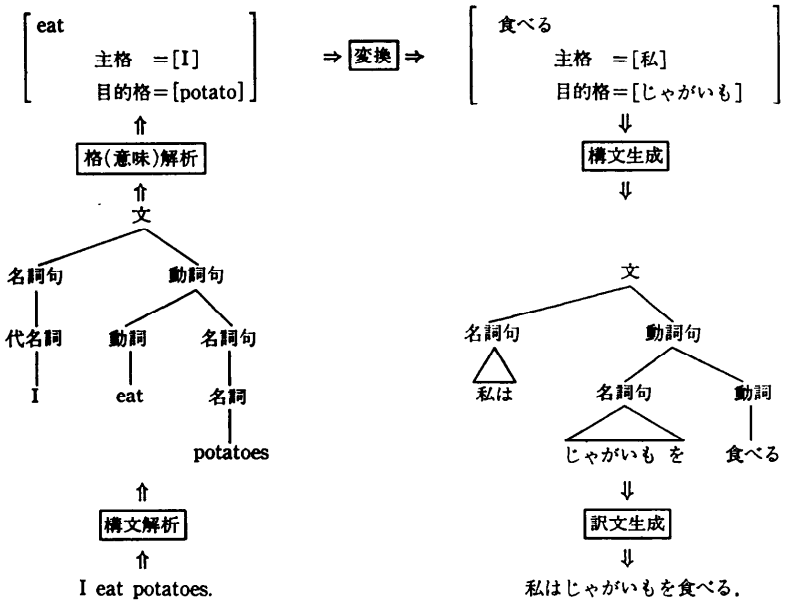

図-6　従来の機械翻訳の方式（トランスファ方式）

従来の機械翻訳の方式

ここでは，従来の機械翻訳の方式を非常に単純化して説明しよう．従来の機械翻訳の方式（トランスファ方式）は，大きく，解析，変換，生成の三つのステップによって行われる．**図-6**に示すように，解析では，まず，文の構造が調べられ（構文解析），次に，主動詞とその格が決定される（格解析）．その後，格解析された構造が相手側言語に移され（変換，あるいは，トランスファとも呼ばれる），構文生成を経て，訳文が生成される．これらの処理は，基本的には，すべて規則という形で記述されたプログラムによって行われる．たとえば，構文解析では，

文→名詞句 動詞句
動詞句→動詞 名詞句
名詞句→代名詞
名詞句→名詞
動詞→eat
代名詞→I
名詞→potatoes

といった規則が解析を司る．たとえば，最初の規則は，名詞句とそれに引き続く動詞句は文を構成することを，最後の規則は，potatoes という単語は名詞であることを，それぞれ表している．これらの規則によって，図に示すような構文解析木が作られるのである．

実際のトランスファ方式は，上記の説明を非常に複雑にしたものである．詳しくは，たとえば，文献 37) を参照されたい．なお，機械翻訳の一般向けの入門書としては，文献 38) が分かりやすい．

（1） 君は水泳が大変うまい．

という文を英語に翻訳しなければならないとしよう．英語があまり得意でないあなたは，すぐにはこの文を英訳できないかもしれない．ここで，もし，次のような翻訳例を思い出すか，あるいは，翻訳例文集や辞書で見つけたとしよう．

（2） 君はまったく芝居がうまい．
→You're a great actor.

この場合，この翻訳例をまねて，

（3） You're a great swimmer.

という訳を作ることは，比較的簡単にできるであろう．

もう一つの例を図-1に示そう．今，翻訳例(4)，(5)を知っていることを仮定しよう．そして，文(6)を翻訳することを考えよう．まず，(6)によく似た翻訳例を検索する．文法的には，(6)は，(4)や(5)と同じであるが，意味的には，'I' は 'acid' よりも 'he' に似ており，また，'potato' は 'metal' よりも 'vegetable' に似ているので，最終的には，(4)のほうがより(6)に似ていることになる．そこで，(4)を模倣して，(7)という翻訳が導かれるわけである．

EBT は，まさに，このようなプロセスを計算機上に実現しようというものである．

上記のストーリは，翻訳対象を文としたが，この考え方は，任意の大きさのテキストに対して適用可能である．文より小さな単位，たとえば，句や節に関して適用可能であることは，容易に想像できるだろう．逆に，文より大きなテキストを模倣利用する典型的な例は，(英語で) 手紙を書く場合である．アポイントメントをとる手紙を書くとき，お礼の手紙を書くとき，多くの人は，例文集のお世話になっているのではないだろうか．

さて，このような考え方に沿って機械翻訳システムを計算機上に実現する場合，翻訳例は当然のことながら，計算機上に格納しておくことになる．ということは，もし，よく似た翻訳例をまねて翻訳するという基本的な能力さえ実現できれば，あとは，どんどん新しい翻訳例を追加していくだけで，どんどん翻訳能力を向上させていくことができることになる．あたかも人間が数多くの翻訳例を覚えることによって，翻訳能力が向上していくようにである．

実は，EBT は，これ以外にも工学的に数多くの長所をもち得ると考えられている．今まで言われている長所を，ここでまとめて述べておこう．

1. **システムの構築・改良が容易である**．われわれがしなければならないことは，単に翻訳例を集め，それを計算機に入力することである．このことは，機械翻訳の非専門家でも，簡単に自分用の機械翻訳システムを構築できることを意味する．

2. **知識のポータビリティと安定性に優れている**．いったん作成された翻訳例データベースは，他のシステムに移植可能であるとともに，長期にわたって安定した知識源として利用でき，風化することが少ない．

3. **こなれた訳を出力できる可能性がある**．人間の訳した翻訳例は，翻訳に関する知識の宝庫である．こなれた訳をまねることによって，結果的にこなれた訳を出力できる可能性がある．

4. **システムの安定性に優れている**．大規模な翻訳例データベースを利用することによって，システムの挙動を安定させることができる．既存の規則に基づいた方法のように，一つ規則が間違っているだけで結果がめちゃくちゃになるといった不安定性を排除できる．

5. **その訳がどれくらい信頼できるかの指標を求めることができる**．翻訳例との類似性が，信頼性を測る指標となりうる．

一方，EBT の短所には，

1. **計算量が多い**．最もよく似たものを見つけるという処理は，本質的に全数探索であり，現在の逐次型計算機では，その計算コストはかなり大きなものになる．

がある．しかし，この問題は，並列計算機の利用などによって解決できると考えられている☆．

[翻訳例]
(4) He eats vegetables. → 彼は野菜を食べる．
(5) Acid eats metal. → 酸は金属を侵す．
[入力文]
(6) I eat potatoes.
[類似翻訳例の検索]
文法的に 入力文＝(4)，(5)
意味的に I～he, potato ～ vegetable
I / ～ acid, potato / ～ metal
全体として 入力文～(4)
[出力文]
(7) 私はじゃがいもを食べる．

図-1 EBT の基本的な流れ

3. 要素技術とその現状

次に，EBT はどのようなシステム構成となるかを示そう．残念ながら，翻訳処理全体をカバーする完全な EBT システムはまだ存在していない．しかし，前章のストーリに沿って考えれば，その

☆ 現段階では，このほかの EBT の欠点はあまりよく分かっていない．今後の研究の進展によって徐々に明らかになってくると思われる．なお，EBT に対する根強い懐疑として，
- EBT が必要とする，大規模なコーパスや対訳データベースを本当に作ることができるのか．

という意見があることを付記する．

おおよその姿は，図-2 のような構成となるだろう．つまり，

1. 翻訳の知識源となる**対訳データベース**
2. 入力から，それとよく似た対訳を検索する**最適照合検索機構**
3. 検索された対訳と入力との差異を調整し，最終的な出力を生成する**適用調節機構**

という三つが，EBT を実現する基本構成要素となる．以下では，これらの構成要素について述べる．

3.1 対訳データベース

EBT の主要な知識源となるのが対訳データベースである．言語Aのテキストとその翻訳である言語 B のテキストの対を一つのデータとするデータベースである．

データベースの作成には，いくつかの意思決定が必要である．

- テキストをどの程度解析処理しておくか．

候補 1　生のデータ．すなわち文字列．

候補 2　形態素処理したデータ．すなわち単語列．

候補 3　構文解析などを行ったデータ．典型的には木構造．

- 一つのデータ，あるいは，対応関係をとる単位をどのくらいの大きさにするか．

候補 1　単語．

候補 2　句，節．

候補 3　文．

候補 4　文章．

理想的には，できるだけ深く解析したテキストを，あらゆるレベルで対応関係を付け，データベースに格納することが望ましいことは言うまでもない．しかし，解析レベルが深くなればなるほど，マルチレベルで対応付けをすればするほど，データベースの作成コストは高くなり，また，標準化も難しくなる．

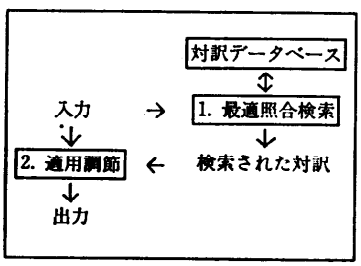

図-2　EBT のシステム構成

どのような対訳データベースを作るにせよ，そのコストはかなりなものであり，多くの人々が利用できるような形で作成しなければ，非常にもったいないことになる．そのためには，データの形式の標準化が不可欠であり，それを強く推進する必要がある*．これに関しては，まず，処理の浅いレベルから段階的に標準化を進めていく必要があるだろう．また，データベースを広く利用できるようにするためには，著作権の問題も解決しなければならない重要な問題である．

3.2 最適照合検索

最適照合検索は，与えられた入力文**によく似た文（とその訳文）を対訳データベースから検索することを実現するものである．その基本的な実現手法は，

1. 二つの文間に類似性を定義し，
2. 入力文と最も類似性が高い文を（高速に）探す検索アルゴリズムを与える

という方法である．この処理は，本質的には全数探索であるが，うまく類似性を定義すれば，真面目に全数探索する必要がないような検索アルゴリズムを作ることができる***．もちろん，そのようなトリックを使わずに，並列計算機のパワーに頼るという方法もある．

類似性の定義は，類似性をとる対象となるデータの大きさとデータ構造によっていろいろなものが提案されている．もちろん，そのベースとなるのは，あるデータ構造で表現された一般的なデータ間の類似度[5]であるが，それが言語テキストであるということ（さらには，それが日本語とか英語であるということ）を意識して，チューニングすることが重要である．それを行わないかぎり，意味のある有用な類似性の定義は得られない．

比較的よく研究されているのは，文字列間の類似性（距離），単語列間の類似性（距離）である．後者では，単語の類似度を与えるシソーラスを利用することもある．詳細については，**4.1**, **4.2**, および，個々の論文を参照されたい．

---

* アメリカでは，Text Encoding Initiative[3] がテキストデータベースの作成と標準化を推進している．日本でも，同様な動きがある．

** 必ずしも「文」である必要はないが，ここでは，文だとして話を進める．

*** 新日鉄の大規模文書データのあいまい検索システムの検索スピードは注目に値する[3],[4]．600 M バイトのテキストに対して最適照合検索を 5 秒以内で実現する．しかし，残念なことに，詳細は発表されていない．

3.3 適用調整

適用調整では，検索された類似対訳から望むべき訳文を作り出すことを行う．しかし，これをどのように実現すればよいかということは，ほとんど分かっていないに等しい．これは，EBT に限らず，一般の類推においても，同じような状況である．

解決しなければならない問題を非常に大雑把に言うと，入力文と検索された対訳の入力側の文との差異を抽出し，その差異を出力側の文に反映させるということである．これをルールでゴリゴリ書かずに，スマートに解くことはなかなか難しく，ほとんどの研究者がこの問題を避けようとしているとも見受けられる．ほとんど唯一のアイディアは，MBT2 で採用された，複数の対訳の部分を組み合わせるという方法である．これについては，**4.3** で述べる．

4. BET システム

本章では，今までに提案されている代表的な EBT システムを大きく三つのグループとその他に分けて概観する．

4.1 用例検索による翻訳支援システム

第一のグループは，対訳（用例）検索による翻訳支援システムである．このタイプのシステムは，翻訳したい文によく似た文とその訳文を検索し，それをユーザに提示することによってユーザによる翻訳を支援しようという考え方に基づいている．その実体は，事実上，対訳データベースと最適照合検索機構からなる，対訳データベース検索システムであるが，翻訳支援システムとしても，かなり有用である．

このようなタイプのシステムを最初に提案したのは，隅田らである．彼らの作成したシステム ETOC は，日本語文を入力として，最適照合した日英対訳文を出力する．データベース中の日本語文は，あらかじめ形態素解析が施されている．ETOC の最適照合検索機構は，入力文をまず，形態素解析し，次に，重要度の低い単語から順に変数化する一般化規則を適用して，その結果とデータベース中との文を照合する方法によって実現されている．この一般化規則の集合を変更することによって，いろいろな観点からの最適照合を実現することができる（文献[6),7)]）．

同じようなシステムが中村によっても提案された．ETOC との大きな違いは，付属語を無視し，自立語だけを利用して最適照合を行う点である．すなわち，入力文中の自立語集合と，比較するデータベース中の文の自立語集合の積集合の大きさ（要素の数）で，類似度を定義している（文献[8)]）．

オランダのソフトウェア会社 BSO では，BKB (Bilingual Knowledge Bank) という対訳データベースとその検索システムが作成されている．このデータベースは，テキストを係り受け解析した結果である，意味ラベル付き単語依存構造で表現している．中間言語としてエスペラント語を採用し，英語-エスペラント語，仏語-エスペラント語の二つの対訳データベースを利用して，英仏翻訳を実現しようとしている．Sadler による紙上シミュレーションはあるが，実際に翻訳を実現するシステムはまだ作成されてはおらず，データベースとその検索システムのみが作成されているに留まっている（文献[9)~12)]）．

最適照合検索については，まだまだ研究の余地はあるが，このようなタイプのシステムは，実用システムとしてすぐにでも動き出せる状態にある．そこでの問題は，対訳データベースを作成するコストである．上記の三つのシステムは，すべて，対訳データに対してなんらかの処理を施しているため，ユーザが簡単に対訳データベースを作成できるようにはなっていない．しかし，単に翻訳支援の目的だけならば，まったく処理していない文字列のままの対訳データを対象としても，有用な翻訳支援システムが作成できる可能性がある．筆者が現在作成中のシステム CTM[13)] は，日本語・英語とも生の文字列データで表現された対訳データを対象とした対訳データベース検索システムである．システムの検索例を**図-3** に示す．

4.2 訳語選択モジュール

第二のグループは，限定された範囲の訳語選択を行うシステムである．

筆者らの MBT 1 は，MBR (Memory-Based Reasoning) 的な考え方を機械翻訳に適用する最初の試みである．MBT 1 は，一つの動詞とその引数からなる動詞フレームの翻訳（訳語選択）を行う．**図-4** に実行例を示す．MBT 1 では，動詞とそれぞれの引数の名詞の訳語選択を行わなければ

CTM(Ab)>問題点を克服するために
Score=29, DB = Science 8710, ID = 266, File = 02.ej
先進的な計算機アーキテクチャの設計者たちは，計算を遅くしている問題を克服するために，次のような方式の開発に取り組んでいる．
Designers of advanced computer architectures are developing approaches that overcome both sources of slowness.
Score=18, DB = Science8710, ID = 1935, File = 09.ej
医師（または医学研究者）が，ある問題を解決するために必要なすべての情報を，読んだり，メモしたり，記憶したりすることが日増しに困難になってきたことから，「通信システム」に対する需要が発生してきた．
The need for communication systems arises in par because it is increasingly difficult for a physician (or a biomedical investigator) to read, memorize and remember all the information needed to solve a particular problem.
Score = 18, DB = ML1, ID = 290, File = 01.ej
彼は，領域知識，学習のヒューリスティックス，問題解決方法をコード化するために全面的にプロダクション・システムの枠組みに依存している．
He relies entirely upon a production system framework to encode domain knowledge, learning heuristics, and problem-solving strategies.
Time = 1499(266 + 1233(683))

図-3 CTM による検索例
ゴシック体の示された入力に対して，最適照合した上位三つの翻訳例が示されている．下線部が入力の一部と一致した文字列であり，照合の得点 (score) は，一致した文字の数を基本とし，連続して一致する文字にボーナスを付加する方法によって計算される．詳細は文献 13) を参照．

ならないため，最適照合検索以外に，適用調整の機構が必要になってくる．しかし，MBT 1 では，以下のような方法でこの問題を避けた．まず，入力から，出力の候補を求める．すなわち，各単語に訳語の候補を与えておき，その全ての組合せをもって出力の候補とする．次に，こうして求められた各候補に対して，それぞれ最適照合する翻訳例を求める．最適照合する翻訳例との距離が最も小さな出力の候補を，最終的な出力とする．これによって，動詞と引数の名詞，両者の訳語選択が可能になる．なお，MBT 1 の最適照合検索機構は，多少のトリックは使っているが，基本的には，単語間の距離の重み付き総和をもって動詞フレームの距離を定義する標準的な方法を採用している（文献[14),15)]）．

隅田らは，ほぼ同様の考え方で，日本語の「A の B」をどのような英語のパターン（たとえば，"B of A" や "A's B" など）に訳すかを決定するシステム EBMT を作成した．**図-5** に EBMT の入出力例を示す．EBMT は訳そのものではなく，訳のパターンという一つの値のみを出力するので，最適照合検索機構だけしかもっていない．距離の定義は，各スロット（A，付属語，B）の値間の距離の重み付き総和という標準的な方法である．なお，EBMT では，データ量と精度の関係，並列検索の効果なども検討されている（文献[16),17)]）．

MBT 1 や EBMT によって，規則ではなかなか書き切れなかった複雑な訳語選択問題に対して，EBT 的アプローチは有効であることが実証されつつある．このタイプのシステムは，今後，多くの機械翻訳システムのサブモジュールとして採り入れられていくだろう．その成否は，うまく問題を限定して切りとれるかどうかにかかっている．

4.3 変換モジュール

第三のグループは，従来の枠組では「変換」と呼ばれていた処理に相当する部分をカバーするシ

Source=(PLAY JAPANESE CARD) *Weight-List*=(.211 .789)

Rank	*Target*	*Distance*	*Most Similar Translation*
1	(する 日本人 トランプ)	1.25(4.34 .429)	(PLAY TARO TENNIS) -> (する 太郎 テニス)
2	(ひく 日本人 トランプ)	6.05(18.4 2.74)	(PLAY YOU VIOLIN) -> (ひく あなた バイオリン)
3	(ひく 日本人 カード)	6.72(18.4 3.58)	(PLAY YOU VIOLIN) -> (ひく あなた バイオリン)
4	(する 日本語 トランプ)	211. (999. 0.0)	(PLAY THEY CARD) -> (する 彼ら トランプ)
5	(する 日本語 カード)	212. (999. 1.45)	(PLAY THEY CARD) -> (する 彼ら トランプ)
5	(する 日本人 カード)	212. (999. 1.45)	(PLAY THEY CARD) -> (する 彼ら トランプ)
7	(ひく 日本語 トランプ)	213. (999. 2.74)	(PLAY I VIOLIN) -> (ひく 私 バイオリン)
8	(ひく 日本語 カード)	214. (999. 3.58)	(PLAY I VIOLIN) -> (ひく 私 バイオリン)
9	(演じる 日本人 トランプ)	792. (18.4 999.)	(PLAY YOU HAMLET) -> (演じる あなた ハムレット)
9	(演じる 日本人 カード)	792. (18.4 999.)	(PLAY YOU HAMLET) -> (演じる あなた ハムレット)
11	(演じる 日本語 トランプ)	999. (999. 999.)	(PLAY HE ROMEO) -> (演じる 彼 ロメオ)
11	(演じる 日本語 カード)	999. (999. 999.)	(PLAY HE ROMEO) -> (演じる 彼 ロメオ)

図-4 MBT1 の実行例
入力 (PLAY JAPANESE CARD) に対して，12 の翻訳の候補とその得点（最もよく似ている翻訳例への距離）が得られている．最も小さな距離をもつ候補が，最終的な翻訳出力となる．

入力 A 付属語 B	出力 (パターン)
八日 の 午後	B of A
会議 の 参加料	B for A
京都 での 会議	B in A
東京 での 滞在	?

図-5 EBMT の入出力

ステムである.

まず，現れたのは，Sadler の BKB を用いた紙上シミュレーションである．2言語間で対応関係のつくテキストフラグメントの対を翻訳ユニット (Translation Unit) と名付け，ジグソーパズル的に翻訳ユニットを組み合わせて入力文(依存構造)を覆い，変換を実現するというアイディアを示した．しかし，完全なアルゴリズム化，および，プログラム化はなされていない (文献[10),11)]).

実際に翻訳ユニットを組み合わせて利用し，変換を実現するプログラムは，筆者らの MBT 2 において初めて示された．その中心的アイディアは，照合表現と呼ばれる，翻訳ユニットの組合せを表す表現の導入である．入力（依存構造）は，まず，入力側の照合表現に変換される．次に，その照合表現を出力側の照合表現に変換する．最後に，出力側の照合表現を解いて，求めるべき出力(依存構造）を得る．一般に，入力は，複数の照合表現に変換することができ，これが複数の出力結果をもたらす．この中から最良の解を選択するために，類似度を利用した得点を導入している (文献[18)~20)]).

このようなタイプのシステムでは，いかにして適用調節を行うかが最大の問題となる．MBT 2 の解法は，言わば，再帰的に類似した翻訳例を利用することによって，適用調節の問題を回避しようとしていると言える．すなわち，まず，入力によく似た実例を探し，利用する．次に，その差異の部分に対して，同様に似ている実例を探し，利用する．このような再帰を差異がなくなるまで繰り返すわけである.

4.4 その他のシステム

これら以外にも，EBT 的なシステムは，いくつか現れてきている.

古瀬らの変換主導型翻訳[21)]は，翻訳処理全体をカバーする完全な EBT システムを実現する一つの試みである．国際会議の参加登録に関するドメインにおいて，典型的な文を訳すことができる EBT を作成しつつある．また，筆者の MPEBT[22)] は，MBT 2 を発展させ，解析・生成を含む形で並列計算機上で翻訳システムを作成しようという試みである．さらに，統計的アプローチによる翻訳[23)]は，ある種の EBT とも考えることができる.

解析における曖昧性の解消を実例（あるいは，実例から抽出した情報）を用いて行うことは，かなり多くの研究がある*．また，規則を用いずに，実例（解析例）だけに基づいて解析を行うことも試みられている[26),27)]．さらに，超並列計算機を利用した Memory-Based Parsing の研究[28)]もある.

そのほかに，渡辺の RCT[29)]，ArchTran[30)] なども，EBT とかなり関連がある.

5. 自然言語処理，および，人工知能研究における EBT 研究の意義

EBT の自然言語処理研究，および，人工知能研究における位置付けは，かなり微妙である．筆者自身，明確な答をもっているわけではない．ここでは，現時点における筆者の考えをまとめてみる.

5.1 自然言語処理と EBT

「言語」というのは，少数の抽象化された規則で記述できるものなのだろうか．おそらく，言語理論というものは，暗黙的にそれができると仮定しているのだろう．あるいは，翻訳というものはどうか．翻訳理論というものは作れるのだろうか.

EBT は，ある意味では，言語（翻訳）の「理論化」ということを放棄した「あきらめの境地」に達しているということができるかもしれない．その意味においては，EBT は科学ではなく，純粋工学である．つまり，翻訳の再利用である．すなわち，EBT は，自然言語処理あるいは翻訳の理論を作ろうという研究ではない.

しかし，一方では，EBT は，「量」というものにまともに向き合うことを強いるものである．「量」が「質」に転じることは起こりうるものなのだろうか．現時点では「量」に対する考察は十分にはなされていないが，大量なデータの蓄積にともなって，量に対する科学といった研究が現れてくることを期待したい.

* たとえば，長尾のシステム[24),25)]など.

5.2 人工知能と EBT

人工知能の中の，学習研究，とりわけ，類推[31]，Case-Based Reasoning (CBR)[32]，Memory-Based Reasoning (MBR)[33] とは，かなり直接的な関連を EBT はもっている[34),35]．似たものを利用して推論を行うという意味においては，EBT は，明らかに類推の一種と考えることはできるであろう．翻訳例を Case（事例）と考えるならば，EBT は，翻訳というドメインにおける CBR と言えるだろう．MBR は，MBT 1 に強い影響を与えている．

そのような表面的な関連を縦糸と考えた場合，横糸は，おそらく以下のようなものになるのではないかと思う．すなわち，

問い あらかじめ全ての場合を想定しておくことができないような「開かれた問題」に対して，有効な推論機構，あるいは，学習機構は何か．

推論機構に対する一つの答え 与えられた問題に完全には照合しないが，その問題を解くのに有効な規則，あるいは，事例を類推的に（あるいはルーズに）適用すること．

学習機構に対する一つの答え 経験した事例を組織化して記憶しておくこと．

「開かれた問題」を人工知能は扱おうとしてきたが，従来の方法は，分野を限るとか用途を限ることによって，できるだけその問題を「閉じた問題」に押し込み，限られた範囲における全ての場合を数えあげる（つまり，それに対する規則を書き切る）ことによって対処してきた．この方法論は，確かに，ある意味では有効であった．

しかし，現実の問題においては，たとえば，「状況」というものは，明らかに数えあげることはできない．人間はそのような問題に対して，どう対処しているのか，あるいは，機械に対処させるにはどうしたらよいのか．翻訳というとてつもない「開かれた問題」を前にして，われわれは，これらの問題について真面目に考えていく時期にきたのかもしれない．

5.3 説明原理と動作原理

EBT の研究を進めていく上で筆者が至った一つの仮説は，

説明原理と動作原理はもしかしたら異なるものなのかもしれない．EBT は説明原理を与えるものではなく，動作原理を与えるものである．

ということである[36]．ある問題を解くシステムが実際にどう動いているのかということと，それを簡単に（人間が理解できるように）説明することとは，別のものである．あるいは，機械翻訳システムを作ること（動作原理）と翻訳の理論を作ること（説明原理）は分けて考えてもよいということである．これは，かなり過激な考え方かもしれない．しかし，少なくとも筆者にとっては，EBT を実現することは，この仮説を検証してみるという意味をもっている．

6. おわりにかえて

本稿では，「実例に基づく翻訳」のアプローチを紹介した．

現段階では，実例に基づく翻訳はスローガンの域を出ていないのかも知れない．しかし，少なくとも近いうちに，その副産物として，対訳テキストデータベースと，それを利用した翻訳支援システムは，実用に供することになるだろう．また，従来の機械翻訳システムの一部として，実例に基づく翻訳の手法が部分的に採り入れられていくことになるだろう．

最後に，今後の研究課題のリストを示すことによって，本稿を締めくくろう．

- **大規模対訳データベースの構築** どのようなデータベースシステムが，対訳テキストデータベースのプラットホームとして適しているか（ハードウェア，ソフトウェアを含めて）．データの加工法の確立．特に，ブートストラッピング（作成したデータベースを利用して，新たなデータを加工する技術）など．
- **定量的な評価**．EBT を実現するためには，どのくらいのデータ（翻訳例）が必要か．
- **適用調節の手法の確立**．
- **高速化の手法**．
- **EBT の利点の実証**．

道は，まだ遠い．

謝辞 本稿の予稿に対してコメントしてくださった，隅田英一郎氏（ATR 自動翻訳電話研究所）と査読者の方々に感謝します．

参 考 文 献

1) Nagao, M.: A Framework of a Mechanical Translation between Japanese and English by Analogy Principle, in ARTIFICIAL AND HUMAN INTELLIGENCE (Elithorn & Banerji,

Eds.), Elsevier Science Publishers, pp. 173-180 (1984).
2) Association for Computers and the Humanities, Association for Computational Linguistics, and Association for Literary and Linguistic Computing: Text Encoding Initiative: Guidelines for the Encoding and Interchange of Machine-Readable Texts, draft (1990).
3) 日経 AI, 1991.7.1, pp. 2 (1991).
4) 日経コンピュータ, 1991.7.29, pp. 57 (1991).
5) 田中栄一: 構造をもつものの距離と類似度, 情報処理, Vol. 31, No. 9, pp. 1270-1279 (1990).
6) Sumita, E. and Tsutsumi, Y.: A Translation Aid System Using Flexible Text Retrieval Based on Syntax-Matching, TRL Research Report, TR 87-1019, IBM (1988).
7) 隅田栄一郎, 堤 豊: 翻訳支援のための類似用例の実用的検索法, 電子通信情報学会論文誌, D-II, Vol. J 74-D-II, No. 10, pp. 1437-1447 (1991).
8) 中村直人: 用例検索翻訳支援システム, 情報処理学会第 38 回全国大会, pp. 357-358 (1989).
9) Sadler, V.: The Bilingual Knowledge Bank (BKB), BSO/Research (1989).
10) Sadler, V.: Translating with a simulated Bilingual Knowledge Bank (BKB), BSO/Research (1989).
11) Sadler, V.: Working with Analogical Semantics, Foris Publications (1989).
12) Sadler, V. and Vendelmans, R.: Pilot Implementation of a Bilingual Knowledge Bank, Proc. of COLING-90, Vol. 3, pp. 449-451 (1990).
13) Sato, S.: Example-Based Translation Approach, Proc. of International Workshop on Fundamental Research for the Future Generation of Natural Language Processing, ATR Interpreting Telephony Research Laboratories, pp. 1-16 (1991).
14) 佐藤理史, 長尾 真: 実例に基づいた翻訳, 情報処理学会研究報告, NL-70-9 (1989).
15) 佐藤理史: MBT 1: 実例に基づく訳語選択, 人工知能学会誌, Vol. 6, No. 4, pp. 592-600 (1991).
16) Sumita, E., Iida, H. and Kohyama, H.: Example-based Approach in Machine Translation, IPSJ, Proc. of InfoJapan '90, Part: 2, pp. 65-72 (1990).
17) 隅田英一郎, 飯田 仁: 用例主導型機械翻訳, 情報処理学会研究報告, NL-82-5 (1991).
18) 佐藤理史: 実例に基づく翻訳 II, 情報処理学会研究報告, AI-70-3 (1990).
19) Sato, S. and Nagao, M.: Toward Memory-based Translation, Proc. of COLING-90, Vol. 3, pp. 247-252 (1990).
20) 佐藤理史: MBT 2: 実例に基づく翻訳における複数翻訳例の組合せ利用, 人工知能学会誌, Vol. 6, pp. 861-871 (1991).
21) 古瀬 蔵, 隅田英一郎, 飯田 仁: 変換主導型機械翻訳の実現手法, 情報処理学会研究報告, NL-80-8 (1990).
22) 佐藤理史: 実例に基づく翻訳の超並列化に向けて, 情報処理学会研究報告, AI-77-16 (1991).
23) Brown, P. et al.: A Statistical Approach to Language Translation, Proc. of COLING-88, pp. 71-76 (1989).
24) 長尾 確: 制約と選好による構造的多義性の解消, 情報処理学会研究報告, NL-78-1 (1990).
25) Nagao, K.: Dependency Analyzer: A Knowledge-Based Approach to Structural Disambiguation, Proc. of COLING-90, Vol. 2, pp. 282-287 (1990).
26) van Zuijlen, J.: Probabilistic Methods in Dependency Grammar Parsing, Proc. of International Parsing Workshop '89, pp. 142-151 (1989).
27) van Zuijlen, J.: Notes on a Probabilistic Parsing Experiment, BSO/Language Systems (1990).
28) Kitano, H. and Higuchi, T.: Massively Parallel Memory-Based Parsing, Proc. of IJCAI-91, pp. 918-924 (1991).
29) Watanabe, H.: A Method for a Transfer Process Using Combinations of Translation Rules, Proc. of PRICAI '90, pp. 215-220 (1990).
30) Chen, S. et al.: ArchTran: A Corpus-based Statistics-oriented English-Chinese Machine Translation System, Proc. of MT SUMMIT III, pp. 33-40 (1991).
31) Hall, R. P.: Computational Approaches to Analogical Reasoning: A Comparative Analysis, Artificial Intelligence, Vol. 39, pp. 39-120 (1989).
32) Case-based Reasoning from DARPA: Machine Learning Program Plan, Proc. of Case-based Reasoning Workshop 89, Morgan Kaufmann Publisher, pp. 1-13 (1989).
33) Stanfill, C. and Waltz, D.: Toward Memory-based Reasoning, Comm. of ACM, Vol. 29, No. 12, pp. 1213-1228 (1986).
34) 佐藤理史: 「実例に基づく翻訳」における類推, ICOT TM-1049 (1991).
35) 佐藤理史: Memory-based Reasoning の挑戦—もう, ルールなんていらない?—, 1990 年度日本認知科学会シンポジウム資料集, pp. 1-10 (1990).
36) 佐藤理史: Memory-based アプローチと規則学習, 「学習のパラダイムとその応用」シンポジウム論文集, pp. 69-77, 情報処理学会 (1989).
37) 野村浩郷, 田中穂積(編): 機械翻訳, bit 別冊, 共立出版 (1988).
38) 長尾 真, 機械翻訳はどこまで可能か, 岩波書店 (1986).

(平成 3 年 10 月 1 日受付)

佐藤 理史 (正会員)

1983 年京都大学工学部電気工学第二学科卒業. 1988 年同大学院工学研究科博士課程研究指導認定退学. 同年より京都大学工学部電気工学第二教室助手, 1992 年より北陸先端科学技術大学院大学情報科学研究科助教授. 工学博士. 人工知能, 特に, 機械学習, 自然言語処理に興味を持つ. 人工知能学会, 認知科学会, ソフトウェア科学会各会員.