

実例に基づく翻訳 II

佐藤理史

京都大学工学部電気工学第二教室

606 京都市左京区吉田本町

(sato@kuee.kyoto-u.ac.jp)

要 旨

本稿では、実例に基づく翻訳モデル MBT2 について述べる。MBT2 で解決しようとした問題は、一つの文を翻訳する際に、必要に応じて、複数の翻訳例の有用な部分を組み合わせ、それに基づいて翻訳するにはどうすればよいかという問題である。この問題を解決するために、複数の翻訳例の部分の組み合わせ方を表現する「照合表現」と呼ぶ表現を導入する。翻訳は、この照合表現を介して行なわれる。すなわち、まず、入力から入力側の照合表現を作り、次に、その照合表現を出力側の照合表現に変換し、最後に、出力側の照合表現を解いて出力を得る。この機構は、一般に複数の翻訳候補を出力する。翻訳の得点を定義し、最終的に最もよさそうな翻訳を選択する。

Memory-based Translation II

Satoshi Sato

Dept. of Electrical Engineering, Kyoto University

Yoshida-honmachi, Sakyo, Kyoto, 606, Japan

Abstract

An essential problem of memory-based translation is how to utilize more than one translation example for translating one source sentence.

This paper proposes a method to solve this problem. We introduce the representation, called *matching expression*, which represents the combination of fragments of translation examples. The translation process consists of three steps: (1) Make the source matching expression from the source sentence. (2) Transfer the source matching expression into the target matching expression. (3) Construct the target sentence from the target matching expression.

This mechanism generates some candidates of translation. To select the best translation out of them, we define the score of a translation.

1 はじめに

人間は、過去の実験から色々なことを学び、それを現在や将来の問題解決に役立てている。過去の事例なり実例を記憶しておき、それらを模倣して、現在の問題を解こうとすることは、こうした能力の典型的な一形態であり、多くの場面で、日常茶飯事に行なっていると考えられる。

事例に基づく推論 (Case-based Reasoning)[1] に代表されるアプローチは、このような能力をモデル化し、計算機上に実現しようというものである。従来の規則に基づくアプローチが、十分に抽象化され、明確に適用条件が規定された「規則」を駆動して推論を行なうというものだったのに対して、本アプローチは、比較的生のデータに近い事例・事例・用例を直接駆動 (利用) して推論しようという点に特徴がある。これは、類推の一形式として考えることもできる。

これとは独立に、自然言語処理では、用例の利用が以前から考えられてきた。特に、アナロジーによる翻訳 [2] の提案以来、自然言語処理における用例・事例・コーパスの利用が盛んに研究されてきている [3,4,5,6,7]。これらは、事例に基づくアプローチの、自然言語処理への応用と考えることができる。

事例に基づくアプローチでは、解こうとしている問題に記憶中の事例をどのように適応させるかということが重要な問題になる。もし、全く同じ問題に対してのみしかその事例を利用できないならば、その能力は、単なる暗記学習と等しい。記憶中の事例を変形して、似たような問題に適応したり、いくつかの事例を組み合わせることで利用することができるようにする必要がある。

筆者は、先に、格フレームの翻訳において、名詞と動詞の訳語選択を実例に基づいて行なうモデル MBT1 を提案した [4]。このモデルでは、訳語選択を行なうために、1つの実例だけに基づいた類推を使っており、複数の実例を利用することは実現できていなかった。そこで、本稿では、この未解決の問題、すなわち、複数の実例の利用を、翻訳というタスクにおいて実現する方法について検討する。

2 複数の実例の利用

2.1 何故、複数の実例の利用が必要か

実例に基づく翻訳のスローガンは、以下の通りである。

ある文を翻訳する際に、その文とよく似ている既知の文の翻訳例を模倣することによって翻訳を行なう。

しかし、多くの場合、1つの翻訳例を模倣するだけでは不十分であり、複数の翻訳例を利用する必要がある。

例えば、以下の文を翻訳することを考えよう。

- (1) 彼は国際政治について書かれた本を買った。

この文を翻訳する際に、次の翻訳例が利用できるであろう。

- (2) 彼はノートを買った。

He bought a notebook.

しかし、この翻訳例から引き出せる情報は、最大限

- (3) 彼は X を買った。

He bought a X.

という対応関係に関する情報であり、

- (4) 国際政治について書かれた本

をどのように訳せばよいかという情報は決して得られない。もし、訳したい文が

- (5) 彼は本を買った。

ならば、その差分の「本」を、辞書を使って訳すという方法も考えられるが、それでは能力が限定されてしまう。根本的な解決策は、(4)を含む翻訳例、例えば、

- (6) 私は国際政治について書かれた本を読んだ。

I read a book on international politics.

を利用するということである。これは、すなわち、(1)を訳すために、(2)の一部と(6)の一部を組み合わせる利用し、

- (7) 彼は国際政治について書かれた本を買った。

なる訳を作り出すということに他ならない。

このように、複数の翻訳例の部分を組み合わせて翻訳を構成することは、実例に基づく翻訳にとって不可欠な能力であり、ぜひともこれを実現する必要がある。

2.2 実現への道筋

では、どのようにしたら実現できるであろうか。

これを実現する単純な方法は、複数の翻訳例の部分を組み合わせて翻訳する際の、組合せの対象となる単位を決めることである。

自然言語の文の翻訳例、例えば、日本語の文と英語の文の対訳例を考える場合、文全体の対応関係以外

に、多くの場合、部分的な対応関係が存在する。この対応関係が付けられた部分は、少なくともその例の翻訳においては、部分的に翻訳可能な1単位となる。また、全体から、部分的な対応関係が付けられた部分を差し引いて残った部分も同様に、部分的に翻訳可能な1単位となる。これを翻訳ユニットと呼ぼう [6]。

この翻訳ユニットを先の組合せの対象となる単位とする。これによって、複数の翻訳例の部分を組み合わせて翻訳することは、以下のように実現できるであろう。

ある入力を翻訳する場合、まず、その入力を翻訳ユニット (部分的に翻訳可能な単位) に分解する。次に、各々の翻訳ユニットを翻訳し、それを合成することによって出力を得る。

しかし、これだけでは不十分である。例えば、(1) を翻訳する際に、(2)、(6) の他に、次の (8) の翻訳例が記憶されている場合を考えよう。

(8) 彼は前売券を買った。

He got the advance ticket.

この場合は、(2) と (6) を組み合わせて翻訳する方法と、(8) と (6) を組み合わせて翻訳する方法の、2通りの方法が考えられる。このように、入力をどのような翻訳ユニットに分解するかということは、一般に複数の分解結果があり、これが複数の翻訳結果を生じさせる。そのため、得られた複数の結果から最もよさそうなものを選ぶことが必要になる。

MBT1 では、最もよさそうな翻訳結果を選ぶ指標として、翻訳例との距離を用いていた。しかし、今回は、入力や出力が複数の翻訳例の部分の組合せとして表現されるため、ある1つの翻訳例とどの程度似ているかということは、翻訳の良さを測る指標として、もはや適切ではない。複数解を生じさせる直接の原因は、入力をどのような翻訳ユニットに分解するかということにある。それゆえ、入力をどのような翻訳ユニットの組み合わせで覆うかということに関して、良さの指標を割り当てる必要がある。

次節以降では、このような考え方に沿って、複数の翻訳例の部分の組み合わせで翻訳を行なうモデル MBT2 を組み立てていく。

3 照合表現

本節より MBT2 の具体的な説明にはいる。MBT2 は、日英の単語係り受け構造間の翻訳を対象とする。本節では、翻訳例の形式と翻訳ユニットを定義し、単

語係り受け構造を翻訳ユニットの組み合わせとして表現する表現形式 (照合表現) を導入する。

なお、MBT2 は Sicstus Prolog で実現されており、以下の説明では、Prolog の表記法を用いる。

3.1 翻訳例の形式

翻訳例は、日本語、英語それぞれの単語係り受け構造とその間の対応関係から成る。単語係り受け構造は、単語 (原形) と品詞のペアを節点とする木構造であり、対応関係は、部分木間に設定できるものとする。すべての部分木間に対応関係を付ける必要はなく、対応関係が付くものだけに付けられればよい。この対応関係を表現するために、単語係り受け構造のすべての部分木に ID を付ける。対応関係は、ID のペアとして表現する。以下に例を示す。

[翻訳例 1]

```
jwd_e([j1,[買う,動詞],
        [j2,[は,係助詞],
        [j3,[彼,代名詞]],
        [j4,[を,格助詞],
        [j5,[ノート,名詞]]]]),
ewd_e([e1,[buy,v],
        [e2,[he,pron]],
        [e3,[notebook,n],
        [e4,[a,the]]]]),
jwd_ewd_id(j1,e1).
jwd_ewd_id(j3,e2).
jwd_ewd_id(j5,e3).
```

3.2 翻訳ユニット

このような形式で与えられる翻訳例に対して、翻訳可能な部分木と翻訳ユニットを定義する。

翻訳可能な部分木 対応付けが存在する部分木を翻訳可能な部分木と呼ぶ。

[翻訳例 1] における翻訳可能な部分木は以下の通りである。

日本語側	英語側
j1	e1
j3	e2
j5	e3

翻訳ユニット 以下の2つのどちらかに当てはまるものを翻訳ユニットと呼ぶ。

- 翻訳可能な部分木

- 翻訳可能な部分木から、その部分木に含まれる任意の翻訳可能な部分木を削除したもの

[翻訳例 1] における翻訳ユニットは、以下の通りである。

日本語側	英語側
j1	e1
j3	e2
j5	e3
j1-j3	e1-e2
j1-j5	e1-e3
j1-j3-j5	e1-e2-e3

3.3 照合表現

以上の準備に基づいて、照合表現 (Matching Expression, ME) を導入する。照合表現とは、単語係り受け構造を翻訳ユニットの組み合わせとして表現したものである。これを以下のように定義する。

[照合表現の定義]

```

<ME> ::= [<ID>|<DifList>]
<DifList> ::= [] or [<Dif>|<DifList>]
<Dif> ::= [d,<ID>]
          or [r,<ID>,<ME>]
          or [a,<ID>,<ME>]

```

照合表現は、翻訳可能な部分木 (<ID>) とそれに対する操作リスト (<DifList>) から成る。その操作 (command) として、以下の 3 種類がある。

1. [d,<ID>] : ある翻訳可能な部分木 (<ID>) を削除する。
2. [r,<ID>,<ME>] : ある翻訳可能な部分木 (<ID>) を別の照合表現 (<ME>) で表される木構造で置き換える。
3. [a,<ID>,<ME>] : ある翻訳可能な部分木 (<ID>) の根節点の直下にぶら下がる子木として、ある照合表現 (<ME>) で表される木構造を追加する。

の 3 種類がある。

例えば、[翻訳例 1] の他に、次の翻訳例が記憶されているとしよう。

[翻訳例 2]

```

jwd_e([j11,[読む,動詞],
      [j12,[は,係助詞],
      [j13,[私,代名詞]],
      [j14,[を,格助詞],
      [j15,[本,名詞],
      [j16,[た,助動詞],
      [j17,[れる,助動詞],
      [j18,[書く,動詞],
      [j19,[について,格助詞],
      [j20,[国際政治,名詞]]
      ]]]]]]).
ewd_e([e11,[read,v],
      [e12,['I',pron]],
      [e13,[book,n],
      [e14,[a,det]],
      [e15,[on,p],
      [e16,[politics,n],
      [e17,[international,adj]]
      ]]]]).
jwd_ewd_id(j11,e11).
jwd_ewd_id(j13,e12).
jwd_ewd_id(j15,e13).
jwd_ewd_id(j20,e16).

```

このとき、[入力 1] は、[照合表現 1] で表わされる。

[入力 1]

```

[[買う,動詞],
 [[は,係助詞],
 [[彼,代名詞]],
 [[を,格助詞],
 [[本,名詞],
 [[た,助動詞],
 [[れる,助動詞],
 [[書く,動詞],
 [[について,格助詞],
 [[国際政治,名詞]]]]]]]]

```

[照合表現 1]

```
[j1,[r,j5,[j15]]]
```

この照合表現は、j1-j5, j15 の 2 つの翻訳ユニットから構成され、j5 のところを j15 で置き換えて組み合わせることを表現している。

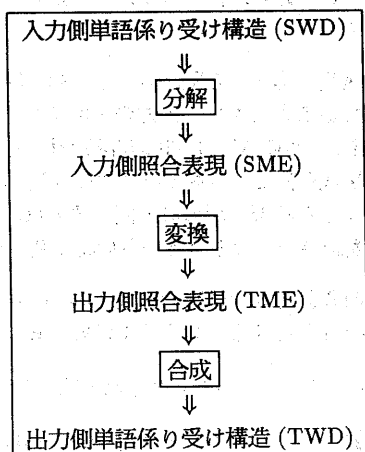


図 1: 翻訳処理の流れ

4 照合表現を介した翻訳

翻訳は、前節で述べた照合表現を介して行われる。翻訳処理は、分解、変換、合成の3つの処理から構成される(図1参照)。なお、これらの処理では、Prologの後戻り機構によって、全ての解が出力される。

4.1 分解

分解処理では、入力単語係り受け構造(SWD)を翻訳ユニットに分解し、入力側の照合表現(SME)を作り出すことを行う。

```

SWD = [[買う, 動詞],
        [[は, 係助詞],
         [[彼, 代名詞]]],
        [[を, 格助詞],
         [[本, 名詞],
          [[た, 助動詞],
           [[れる, 助動詞],
            [[書く, 動詞],
             [[について, 格助詞],
              [[国際政治, 名詞]]]]]]]]]]]]]]
SME = [j1, [r, j5, [j15]]]
  
```

分解処理における中心的な処理は、データベースから翻訳ユニットを検索し、入力単語係り受け構造(以下、照合対象と呼ぶ)と照合することである。ここで決めなければならない問題は、

- どのようにして翻訳ユニットを検索するか(検索の高速化をどのようにして実現するか)

- 照合対象と翻訳ユニットの照合において、どのような合わせ方を許すか

- 一致しない部分に対して、どのような差異(照合表現の command)を作り出すか¹

ということであり、これを

- 無意味は合わせ方を減らすために、どのようなヒューリスティクスを用いるか

を考慮して決める必要がある。

検索については、翻訳ユニットの検索を翻訳ユニットの根節点の内容でハッシュしたテーブルを使って行なう方法をとる。このため、翻訳例は、先に示した形式から、より検索に適した形式に前もって形式変換しておく。

合わせ方については、日本語用と英語用の2つの照合プログラムを用意する。日本語単語係り受け構造の照合では、ある節点にぶらさがる子木の順序は交換可能として照合する。これに対して、英語では、順序を考慮して照合する。

差異の作成の決定と効率化のために、各単語に付けられている品詞をガイド情報として使う。品詞間に置換可能であるかどうかを定義する。そして、置換可能なものに対しては、置換操作を生成し、追加と削除の組み合わせによる差異を作らないことにする。

なお、入力のある部分に対して作成された照合表現は、中間結果として保存し、無意味な探索を繰り返さないようにする。

4.2 変換

変換処理では、分解処理で得られた照合表現に含まれるIDを、対応関係に基づいて相手側のIDに置き換え、出力側の照合表現(TME)を作る。

```

SME = [j1, [r, j5, [j15]]]
TME = [e1, [r, e3, [e13]]]
  
```

4.3 合成

合成処理では、変換処理で得られた出力側の照合表現から、出力側の単語係り受け構造(TWD)を合成する。

```

TME = [e1, [r, e3, [e13]]]
TWD = [[buy, v],
        [[he, pron]],
        [[book, n],
  
```

¹置換は、削除と追加によって表すことができるので、どのような場合に、置換を作ることにするのかを決める必要がある。

```
[[a,det]],
[[on,p],
[[politics,n],
[[international,adj]]]]]]
```

この合成処理は、大きく、照合表現を解いて出力を合成する処理と、合成された出力が適格であるかどうか調べる処理の、2つに分けられる。

4.3.1 出力の合成

出力の合成は、具体的には、照合表現に含まれる操作リスト (DifList) を解くことによって実現される。置換と削除に関しては、何の問題もないが、追加に関しては、親となる節点が指定されるだけなので、英語においては、左から何番目の子木として追加するかという複数の選択肢が存在する。ここでは、すべての可能な構造を作り、次の適格かどうかを調べる処理で、不適格なものを削除する。

4.3.2 適格性の検査

上で述べたように、照合表現に含まれる追加操作を解く場合に、不適格な構造を合成してしまう可能性がある。これを削除するために、ある構造が適格であるかどうかを以下のように検査する。

適格性の検査は、1つの節点と、そのすぐ直下における下がつている子節点のリストからなる単位を1単位として行う。その単位において、それぞれ対応する節点の品詞が一致する単位が実例に存在する場合、その単位は適格とする。ある構造に含まれるこの単位がすべて適格であるとき、その構造を適格とする。

5 翻訳の得点付け

前節で見てきたように、本方式において、翻訳は照合表現を介して行われる。照合表現は、翻訳ユニットとその組み合わせ方を表わす表現であり、翻訳の得点は、この照合表現の良さを反映するものに必要がある。

5.1 翻訳ユニットの得点

照合表現の得点を定義するために、まず、照合表現の要素となる翻訳ユニットの得点を定義する。

どのような翻訳ユニットが良い翻訳ユニットであろうか。これに対する答えは、

- できるだけ大きなサイズの翻訳ユニットがよいだろう。これは、できるだけ大きな単位で翻訳し

た方がよい翻訳が得られるだろうという仮説に基づく。

- 照合表現に含まれる翻訳ユニットは、翻訳例の一部であるとともに、入力の一部でもある。すなわち、2つの環境 (周囲の状況) を持つ。この2つの環境が似ているものがよいだろう。これは、似たような状況で現われる同一要素は、同じように翻訳されるだろうという仮説に基づく。

まず、前者を具体化するために、翻訳ユニット TU のサイズ $size(TU)$ を以下のように定義する。

$$size(TU) = "TU \text{ に含まれる節点数}"$$

次に、翻訳ユニットの2つの環境がどの程度似ているかということを実体化するために、まず、考慮の対象とする環境 (制限された環境) を以下のように限定する。

制限された環境 翻訳ユニットを含む最大一致部分とその1節点外側までの範囲から、翻訳ユニットを除いた部分。

すなわち、制限された環境は、通常は、翻訳ユニットの1段外側にある節点の集合であり、対応する2つの節点在同一の場合は、もう1段拡張することを繰り返すによって得られる節点集合である。図2に制限された環境の例を示す。

翻訳ユニットの環境の類似性は、2つの制限された環境の最適照合 (best match) 時の一致の割合によって見積る。これを計算するために、まず、制限された環境に含まれる節点間の対応付けを行なう。節点の対応付けでは、複数の候補がある場合がある。例えば、図2において、 n_7 は、 m_6 または m_7 に対応する。このような場合は、最も似ている節点を選ぶ。これを可能にするために、節点 (単語) 間の類似度が必要になるが、これは、あらかじめシソーラスとして定義されていると仮定する。なお、単語間の類似度の値域は $[0,1]$ とし、一致するものを1とする。

最適照合が見つかった時点で、以下の値が計算できる。²

$$mpoint(TU, WD) =$$

" TU の翻訳例における制限された環境と WD における制限された環境間の、最適照合時における対応する節点間の類似度の総和"

この値を、2つの環境の類似度を測る指標として用いる。

² この $mpoint(TU, WD)$ という記法は、 TU が WD のどの部分に合わせられているかの情報が記述されていないため、厳密ではない。しかし、簡略化のためこのまま用いる。

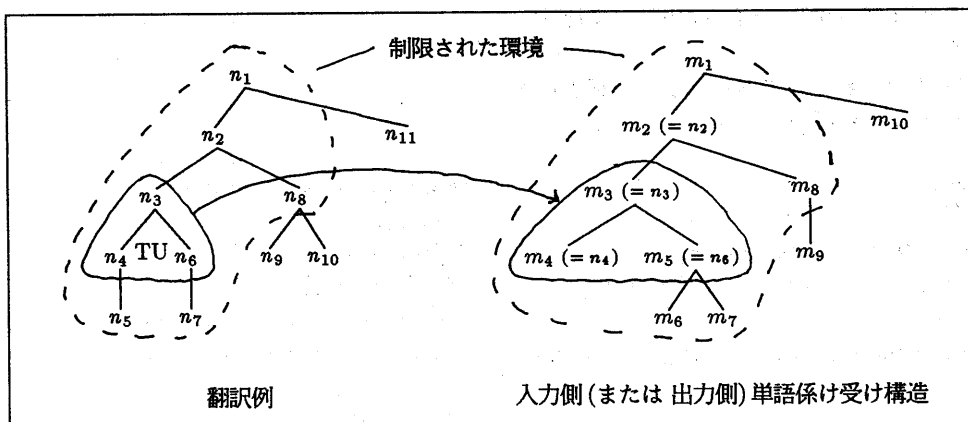


図 2: 翻訳ユニットの制限された環境

以上の準備を経て、翻訳ユニットの得点を以下のように定義する。

$$\text{score}(TU, WD) = \text{size}(TU) \times (\text{size}(TU) + \text{mpoint}(TU, WD))$$

例えば、シソーラスに以下の単語の類似度が定義されているとしよう。

$\text{sim}([\text{本}, n], [\text{ノート}, n], 0.8).$
 $\text{sim}([\text{買う}, v], [\text{読む}, v], 0.5).$
 $\text{sim}([\text{book}, n], [\text{notebook}, n], 0.8).$
 $\text{sim}([\text{buy}, v], [\text{read}, v], 0.5).$

このとき、前節の翻訳ユニットの得点は、以下のようになる。

TU	size	mpoint	score
j1-j5	4	0.8	19.2
j15	6	1.5	45
e1-e3	2	0.8	5.6
e13	5	0.5	27.5

5.2 照合表現の得点

照合表現 ME の得点は、照合表現を構成する翻訳ユニットの得点の総和を照合対象の単語係受け構造 WD の大きさと正規化したものとして定義する。すなわち、

$$\text{score}(ME, WD) = \frac{\sum_{TU \in ME} \text{score}(TU, WD)}{[\text{size}(WD)]^2}$$

単語係受け構造の大きさは、翻訳ユニットの大きさと同じように、それに含まれる節点の数とする。以上

の定義から明かなように、照合対象が1つの翻訳ユニットと全く同一である場合、得点は1となる。

前節の照合表現の得点は、

ME	score
[j1, [r, j5, [j15]]]	0.642
[e1, [r, e3, [e13]]]	0.676

となる。

5.3 翻訳の得点

翻訳においては、入力側と出力側の、2つの照合表現を介する。このため、2つの照合表現の得点が得られる。この両者の低い方の値をとって翻訳全体の得点とする。すなわち、

$$\text{score}(SWD, SME, TME, TWD) = \min(\text{score}(SME, SWD), \text{score}(TME, TWD))$$

前節で示した翻訳の得点は、0.642 となる。

6 実行例

英語の動詞 eat の訳し分けの例を以下に示す。eat に対して、以下の2つの翻訳例を用意した。

(9) A man eats vegetables.

人は野菜を食べる。

(10) Acid eats metal.

酸は金属を侵す。

このような翻訳例を用意したときの MBT2 の実行例を図 3 に示す³。

前者の場合は「食べる」を後者の場合は「侵す」を選択していることがわかる。

7 考察

本節では、MBT2 と規則学習の関連と、類推という立場から見たときの MBT2 について考察する。

7.1 規則学習との関連 — 実例からの広がり —

翻訳例は、それ自身、実例規則と考えることができる。実例規則を記憶しておき、完全照合 (exact match) で照合をとる従来のルール・システムで動かすならば、与えられた実例しか翻訳できない。これに対して、実例に基づく翻訳では、与えられた実例以外の翻訳をすることができる⁴。規則学習の見地から見ると、この能力の向上は、実例規則の適用範囲の暗黙的な拡大として解釈することができるということは、すでに [8] に述べた。ここでは、MBT2 において、どのようなメカニズムで適用範囲の拡大がおこなわれるかを考えてみよう。

再び、3.1 で示した [翻訳例 1] を考えよう。この翻訳例は、[入力 1] の翻訳に利用されるうるわけであるが、これが可能なのは、部分への分解と変数化というメカニズムのおかげである。すなわち、この場合、上記の翻訳例は、以下のような翻訳規則として働く。

```
trans([[買う, 動詞],
      [[は, 係助詞],
       [[彼, 代名詞]]],
      [[を, 格助詞],
       x]],
      [[buy, v],
       [[he, pron]],
       y])
:- trans(X, Y).
```

このことから明かなように、翻訳ユニットは、1 つの翻訳規則として考えることができる。

しかし、翻訳ユニットが通常の翻訳規則と異なっている点は、変数化された部分が元々は何だったかという情報が保持される点である。この例の場合、

³このほかに、訳したい文に出て来る単語を含んだ翻訳例やシソーラスを用意した。

⁴この意味において、実例に基づく翻訳は、従来の「学習」の一部を取り込んでいることになる。

*** Translation Source ***

```
[[eat, v],
 [[he, pron]],
 [[potato, n]]]
```

*** Translation Results ***

No. 1 (Score = 0.5889)

```
[[食べる, 動詞],
 [[は, 係助詞],
 [[彼, 代名詞]]],
 [[を, 格助詞],
 [[じゃがいも, 名詞]]]]
```

No. 2 (Score = 0.4556)

```
[[侵す, 動詞],
 [[は, 係助詞],
 [[彼, 代名詞]]],
 [[を, 格助詞],
 [[じゃがいも, 名詞]]]]
```

*** Translation Source ***

```
[[eat, v],
 [[acid, n],
 [[sulphuric, adj]]],
 [[iron, n]]]
```

*** Translation Results ***

No. 1 (Score = 0.5500)

```
[[侵す, 動詞],
 [[は, 係助詞],
 [[硫酸, 名詞]]],
 [[を, 格助詞],
 [[鉄, 名詞]]]]
```

No. 2 (Score = 0.4688)

```
[[食べる, 動詞],
 [[は, 係助詞],
 [[硫酸, 名詞]]],
 [[を, 格助詞],
 [[鉄, 名詞]]]]
```

図 3: MBT2 の実行例

	Input	Output
Analogy Target	x	y
	$\downarrow$	$\uparrow$
Analogy Source	x'	y'
	$\rightarrow$	

図 4: 単一類推源に基づく類推の基本的な図式

X : [[ノート, 名詞]]

Y : [[notebook, n], [[a, det]]]

という情報が保持されているということである。この情報が保持されるため、その規則の実際のデータ (入力) への適用において、この適用がどの程度好ましいかという指標を求めることができる。これは、変数化された部分の元々の内容と実際のデータとの類似性に基づいて計算されるので、別の見方をすれば、実例規則の類似語への広がりと考えることもできる。

以上、まとめると、実例規則の拡大は、主に以下の2つのメカニズムによって行われる⁵。

- 部分への分解と変数化
- 類似語への広がり

前者は、規則学習において一般化 (generalization) と呼ばれるものに相当し、後者は、MBT1 において起こっていた規則適用の拡大に相当する。MBT2 では、この2つがこん然一体となって実例規則適用の拡大を可能にしている。

7.2 類推としての MBT2

事例に基づくアプローチは、一種の類推と考えることができる。ここでは、類推という立場から、MBT2 を眺めてみよう。

今までの類推研究の多くは、1つの類推源に基づく類推に限定されていた。この類推の基本的な図式を、図 4 に示そう。

すなわち、入力 x が与えられると、それとよく似た入力 x' を事例データベースから探しだし、その出力 y' から最終的な出力 y を得る。ここで、入出力ペア $\langle x, y \rangle$ を類推目標 (analogy target)、推論の根拠となった事例 $\langle x', y' \rangle$ を類推源 (analogy source) と呼ぶ。この類推の確からしさは、通常、入力間の距離 $d(x, x')$ 、あるいは、事例間の距離 $d(\langle x, y \rangle, \langle x', y' \rangle)$ を定義し、この距離によって測るという方法が採られる。MBT1 は完全にこの図式にのることから、単一類推源を用いた類推と考えることができる。

⁵ この2つ以外に、順序任意の照合法と、挿入、削除の2つの command が実例規則の拡大に関係している。

	Input	Output
Analogy Target	x	y
	$\downarrow$	$\uparrow$
Matching Expression (Analogy Sources)	ME_x	ME_y
	$\rightarrow$	

図 5: 類推としての MBT2 の図式

これに対して、MBT2 では、この図式にはのらない。MBT2 の図式は、図 5 のようになる。これは、複数の類推源に基づく類推の一種と考えることができる。

複数類推源を用いた類推に対して、MBT2 から得られた知見は以下の通りである。

- 複数の類推源を用いた類推では、複数の類推源 (の部分) をどのように組み合わせて使うかということを決める必要がある。これを簡単に行うために、その組み合わせ方を表現する表現 (MBT2 では、照合表現) を持ち込むのがよいであろう。
- 複数の類推源を用いた類推では、類推の確からしさの指標として、単一の類推源 (事例) との距離を用いることはもはやできない。複数の類推源の組み合わせ方に対する得点付けを行う必要がある。

8 おわりに

本稿では、MBT2 について述べた。MBT2 で解決しようとした問題は、一つの文を翻訳する際に、必要に応じて、複数の事例の有用な部分を組み合わせ、それに基づいて翻訳するにはどうすればよいかという問題である。MBT2 では、複数の事例の部分の組み合わせ方を表現するために、照合表現と呼ぶ表現を導入し、それに対する得点付けを定義するという方法で、この問題の解決を試みた。

MBT2 で扱った問題は、より一般的には、事例に基づく推論において、複数の事例をいかにして利用するか、あるいは、複数の類推源に利用した類推をいかにして実現するかという問題である。MBT2 の解法は、これらの問題に対する指針を与える。

しかし、残念ながら、機械翻訳という見地から見れば、現在の MBT2 は、甚だ限られた翻訳能力しかなく、実用レベルには程遠いレベルに留まっている。

事例に基づく翻訳は、まだまだ試行錯誤の段階にあり、未知の問題、未解決の問題が山積している。今後

考えていきたい興味深い問題としては、以下のような問題がある。

- **いわゆる文法的変形の問題** 例えば、能動態の文を知っている場合、受動態の文を訳すことができるようにする必要はあるか。すなわち、文法的変形を受けていない文を実例として持つ場合、文法的変形を受けている文も訳せるようにする必要はあるか。もし、あるとすれば、どのような方法で取り扱うべきか。これは、一つの実例の適応範囲をどこまで広げるべきかという問題として考えるべきである。
- **解析の問題** どうやって入力単語列から単語依存構造を構成するか。実例に基づく解析は可能か。意味があるか。

これらの問題は、今まで「文法」として比較的容易に取り扱えていた問題を実例に基づく翻訳ではどう取り扱うべきかという問題である。従来の方法と組み合わせるという道がよいのか、それとも、違う道があるのかは、現段階ではなんとも言えない。今後、この問題について考えていきたい。

参考文献

- [1] Kolodner, J. (ed.), Proceedings: Case-based Reasoning Workshop, Morgan Kaufmann Publishers, 1988.
- [2] 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, pp173-180, 1984.
- [3] Sumita, E. and Tsutsumi, Y., A Translation Aid System Using Flexible Text Retrieval Based on Syntax-Matching, TRL Research Report, TR87-1019, IBM, 1988.
- [4] 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.
- [5] 中村直人, 用例検索翻訳支援システム KSA フォーラム/自然言語処理分科会, 1989.
- [6] Sadler, V., Translating with a simulated Bilingual Knowledge Bank(BKB), BSO/Research, 1989.
- [7] Sumita, E., Iida, H., and Kohyama, H., Translating with Examples: A New Approach to Machine Translation, The Third International Conference on Theoretical and Methodological Issues in Machine Translation of Natural Languages, to appear.
- [8] 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, pp69-77, 情報処理学会, 1989.