自 然 言 語 処 理 99−12
（1994. 1. 21）

# 類似性に基づいた日韓対訳テキストの文対応

黄 道三　　長尾 眞
京都大学工学部 電気工学第二教室

**要 旨**

最近、例文を用いた自然言語処理の研究が盛んに行なわれている。このような研究には大量の二言語間の対訳テキストが使われているが、原文と翻訳文とがかならず1文対1文に対応しているのではないので、実際の研究に使うためには原文と翻訳文とを対応させる必要がある。しかし、人手で文対応を行なうには非常に手間がかかる。ところで、人間がこのような文対応作業をするときには、まず、原文を翻訳し、その文と似ている文を翻訳文より探し、原文と対応させると考えられる。そこで、本稿では、日本語文と韓国語文の対訳テキストを対象にして、日本語原文を翻訳文と似ている文に変換し、その変換文と翻訳文との類似度に基づいて文対応をとる方法を示す。また、実際の例文に対して実験し、その評価を示す。

# Aligning of Japanese and Korean Texts by Analogy

Dosam Hwang　　Makoto Nagao
Department of Electrical Engineering, Kyoto University

**Abstract**

Text alignment has become an important problem for the contrastive study of languages. Several text alignment methods have already been proposed for western languages, but there is still no good method proposed Far Eastern languages. We propose here a new method for Japanese and Korean texts. Japanese and Korean languages are very similar in the word order and lexical properties. In such a language pair simple replacement of words in a sentence from one language to another may give a rough translation that may help to match a proper corresponding sentence in the text of the other language. We tried to realize this process by computer. We present in this paper a method to align Japanese sentences and their Korean translations automatically based on the similarity between the sentences. In addition, we present and evaluate the results of the alignment experiments for Japanese sentences and their Korean translations.

## 1 はじめに

1980 年中ごろ [1] によって類似性による機械翻訳が提案された後、最近になって、例文を用いた自然言語処理の研究が盛んに行なわれるようになった。このような研究には大量の二言語間の対訳テキストが使われているが、実際に大量の二言語間の対訳テキストを研究に利用するためには、原文と翻訳文とがそれぞれ対応している必要がある。

しかし、原文は翻訳されるとき、原文の 1 文が 1 文に訳される場合の外に、1 文が 2 文に、あるいは 2 文が 1 文に訳されたり、省略されたりする時も生ずるため、かならずしも原文と翻訳文とが 1 対 1 に対応しているとはいえない。したがって、対訳文を利用する前に、まず、原文と翻訳文との対応づけを正しく行なわなければならないが、人手で対応づけをするには非常に手間がかかる。

そこで、文字数、または単語数などの文対応の統計的な数値を用いて、文対応を自動的にとる研究が行なわれたことがある [2][3][4]。ところが、人間がこのような文対応作業をするときには、文対応に関する統計的な数値を利用するよりも、まず、原文を相手側の言語に訳し、文の長さを考慮しながらその文と似ている翻訳文を探して原文と対応させると考えられる。とくに、長い原文の場合には、完全に翻訳せず、文の一部分、またはいくつかの単語だけを翻訳し、それとマッチしている翻訳文を対応している文とすることができる。1 文中の単語数を用いて文対応をとるためには形態素解析が必要になるが、韓国語の形態素解析の成功率は低いので、できれば、韓国語の形態素解析は行なわず文対応をとる方法が望ましい。

そこで、我々は、もし原文を翻訳文と一致する文に変換できれば、その変換文と最もよく一致する翻訳文を原文に対応させることで、容易に、自動的に文対応がとれると考えた。しかも、日本語と韓国語とは、文の構造とともに文字においても非常に似ているので、翻訳文とかなり似ている文に容易に変換ができる。しかし、翻訳文と完全に一致する文に変換するのは不可能であるので、変換文と翻訳文の文字列一致度とともに、日本語と韓国語との文対応統計値を用いて対応させる方法を用いた。

したがって、本稿では、日本語文を韓国語文に変換する方法と、原文の変換文と翻訳文の類似度に基づいて対応をとる方法を述べ、実例文を対象にして実験し、その評価結果を示す。文対応の処理過程は図 1のようになっている。

- まず、日本語の原文を形態素解析し、形態素解析を通して出てきた日本語単語の対訳語を日韓対訳辞書 [5] より引いて韓国語文に変換する。
- このとき、変換に失敗した日本語文字列に対しては、日韓文字対訳隣接テーブルと日韓文字対訳テーブルとを用いて韓国語文に変換する。
- 最後に、変換文と翻訳文との類似度を求めて、日韓対訳テキストの文対応を行なう。

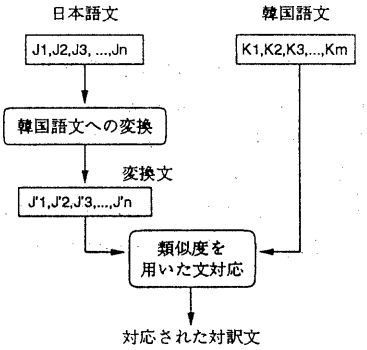



図 1: 日本語と韓国語文においての文対応の処理過程

## 2 日本語文の韓国語文への変換

### 2.1 日韓対訳辞書を用いた変換

日本語と韓国語とは文構造だけではなく、文字レベルにおいても非常に似ている。まず、文構造は SOV 型で、語順も似ている。そのため、図 2のように日本語単語をその韓国語対訳語で置き換えるだけで、翻訳文に似ている韓国語文になる場合も少なくなく、図 2の例では日本語から変換された文と韓国語翻訳文とが完全に一致している。

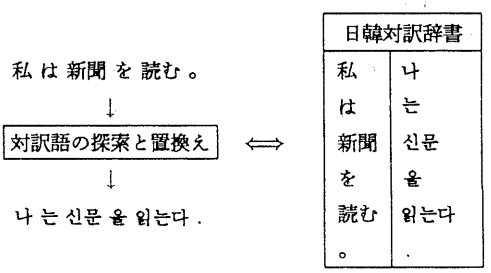



図 2: 日韓対訳辞書を用いた変換

そこで、本研究では、JUMAN[*1] を用いて日本語文の形態素解析を行ない、形態素解析されて出てきた日本語

[*1] 京都大学長尾研究室で開発した日本語形態素解析システム

単語の対訳語を日韓対訳辞書*2より引いて韓国語文に変換することにした[6]。

**日本語助詞の変換**

しかし、日本語助詞は同じ助詞であっても、前単語の韓国語対訳語に終声*3があるかないかによってその対訳語が違う。例えば、主格助詞「は」は前単語の韓国語対訳語が無終声の場合は、「는」に、有終声の場合は「은」に訳される。

そこで、表1のように終声の有無にしたがって日本語助詞の韓国語対訳語を書き込んだ日韓助詞対訳テーブルを用意する。

表 1: 日韓助詞対訳テーブル

| 助詞 | | が | は | ··· | を |
|---|---|---|---|---|---|
| 終声の区別 | | 01 | 02 | ··· | 41 |
| 無終声 | 00 | 가 | 는 | ··· | 를 |
| 有終声 | 01 | 이 | 은 | ··· | 을 |

また、日韓対訳辞書には対訳語が無終声か有終声かにしたがって、それぞれ対訳語の後ろに「J(00」、「J(01」の終声マークを書き加え、助詞については助詞対訳番号を対訳語として書いておく。次に、終声マークと助詞対訳番号とを組合せ、それに対応する対訳語を日韓助詞対訳テーブルから引いて、書き換えることによって日本語助詞を韓国語に変換する。

| | |
|---|---|
| 無終声の場合 | 学校 が → 학교J(00 01) → 학교가 |
| 有終声の場合 | 人間 が → 인간J(01 01) → 인간이 |

**日本語用言の変換**

また、日本語用言も韓国語に翻訳されるとき、韓国語語幹の活用規則と活用形によってその対訳語が違う。そのため、表2のような日韓語尾対訳テーブルを用意し、助詞と同じ変換方法を用いて韓国語に変換する。

**日本語名詞形動詞の変換**

最後に、「置き換え」のような日本語名詞形動詞はその次にくる単語によって名詞形、または動詞形に訳される。例えば、「置き換え」の次に「が」のような助詞がくる場合は、名詞に訳され、「る」か「た」などがくる場合は動詞に訳される。そこで、日韓対訳辞書に「置き換え」のような日本語名詞形動詞に対しては品詞属性とともに、名詞形対訳語「치환」と動詞形対訳語「치환하」とを与え、その次の単語の品詞を参照して名詞形、または動詞形に変換する。例えば、「置き換え」＋「は」の場合は、「치환」＋「은」、「置き換え」＋「た」の場合は、「치환하」＋「였다」に訳する。

表 2: 日韓語尾対訳テーブル

| 用言活用形 | | 終止形 | 冠形形 | ··· | 命令形 |
|---|---|---|---|---|---|
| 用言活用規則 | | 01 | 02 | ··· | 44 |
| 規則1(보) | 01 | 다 | 는 | ··· | 아라 |
| 規則2(먹) | 02 | 는다 | 는 | ··· | 어라 |
| ··· | ··· | ··· | ··· | ··· | ··· |
| 不規則15過去(낮) | 99 | 았다 | 았던 | ··· | ··· |

表 3: 用言の変換例

| | |
|---|---|
| 規則1の場合 | 見る → 보E(01 01) → 본다 |
| 規則2の場合 | 食べる → 먹E(02 01) → 먹는다 |

## 2.2 日本語文字の韓国語文字への変換

しかし、2.1の変換方法では日韓対訳辞書に登録されていない単語については変換ができない。ところで、漢字だけで構成されている日本語単語は、表4のように一意にその漢字を韓国語文字に変換しても韓国語対訳語になり、この場合、同じ長さの単語が多い。

表 4: 日韓文字の類似性

| | 考(고) | | |
|---|---|---|---|
| 思(사) | 思考(사고) | 考察(고찰) | 察(찰) |
| | 思想(사상) | 観察(관찰) | |
| 想(상) | 想起(상기) | 観光(관광) | 観(관) |

| | 考(고) | | |
|---|---|---|---|
| 熟(숙) | 熟考(숙고) | 考案(고안) | 案(안) |
| | 成熟(성숙) | 案内(안내) | |
| 成(성) | 完成(완성) | 内外(내외) | 内(내) |

そこで、2.1の方法で変換されなかった漢字文字列に対しては、日本語の漢字を一意に韓国語の文字に対訳させて、日本語文字列を韓国語文字列に訳する。

また、カタカナだけで構成されている日本語単語も、一意にそれぞれのカタカナを韓国語に変換しても、以下のように韓国語対訳語、あるいは対訳語に似ている単語になる場合が多いので、カタカナの日本語文字列もそれに対する一意の韓国語文字に対訳させて韓国語文字列に

*2 約57,000語が登録されている。

*3 韓国語では、母音と子音の組合せで一つの音節を成り立っており、この組合せの中で音節の中心音である母音の次に付く子音を終声とよぶ。

訳する。

- アーク(아크)、アイス(아이스)

この変換では日本語約 4,000 文字に対して韓国語文字の対訳が付与されている日韓文字対訳テーブルを用いた。

### 2.3 日韓文字対訳隣接テーブルを用いた変換

2.2に述べたように、日本語の漢字単語は一意にそれぞれの漢字を韓国語文字に変換してもかなり正しく訳されるが、次のように、後ろにくる文字にしたがって、異なる対訳文字をとる場合もある。例えば、「不」という文字は後ろに「安」、「潔」、「運」などの文字がくる場合には、「불」に訳され、「在」、「定」、「動」などの文字がくる場合には、「부」に訳される。その他、「如」が後ろにくると、「두」に訳される。

| 対訳の種類 | 対訳の例 |
|---|---|
| 불 | 不安(불안)、不潔(불결)、不運(불운) |
| 부 | 不在(부재)、不定(부정)、不動(부동) |
| その他 | 不如帰(두견새) |

また、同じ漢字文字であっても書かれている位置によって、対訳が異なる場合もある。例えば、「立」という文字は単語の最初にくるときには「입」に、最初以外のところにくるときには、「립」に翻訳される。

| 文字の位置 | 対訳の例 |
|---|---|
| 単語の最初 | 立案(입안)、立場(입장)、立法(입법) |
| 単語の最初以外 | 起立(기립)、国立(국립)、私立(사립) |

そこで、表5と表6のような日韓文字対訳隣接テーブル(NH-Table[*4])を用いて、日本語の漢字を韓国語文字に変換する。

#### 2.3.1 日韓文字対訳隣接テーブルの作成

日韓対訳辞書より日本語の見出し語が漢字だけになっており、その見出し語と韓国語対訳語との長さが同じであるレコードだけを取り出す。

その単語を対象にして、x という日本語文字の右側に y という日本語文字が隣接するとき、NH(x,y) に x の韓国語対訳文字を入れて表5のようなテーブルを作成する。このとき、対訳する韓国語文字が複数ある場合には、頻度数が一番高いのを選ぶ。

また、2.3に述べたように、漢字文字の位置によって対訳語が異なる場合があるので、x という日本語文字の左側に y という日本語文字が隣接するとき、NH(x,y) に x

[*4] NH-Table: Nihongo to Hangul conversion Table

表 5: 日韓文字対訳右隣接テーブルの例

| x＼y | 都 | 潔 | 法 | 安 | 運 | 如 | 場 | 在 | 定 | 学 |
|---|---|---|---|---|---|---|---|---|---|---|
| 京 | 경 | ∅ | ∅ | ∅ | ∅ | ∅ | ∅ | ∅ | ∅ | ∅ |
| 大 | ∅ | ∅ | ∅ | ∅ | ∅ | ∅ | ∅ | ∅ | ∅ | ∅ |
| 不 | ∅ | 불 | 불 | 불 | 불 | 두 | ∅ | 부 | 부 | ∅ |
| 立 | ∅ | ∅ | 입 | 입 | ∅ | ∅ | 입 | ∅ | ∅ | ∅ |

表 6: 日韓文字対訳左隣接テーブルの例

| x＼y | 京 | 国 | 私 | … | 不 |
|---|---|---|---|---|---|
| 案 | … | … | 안 | … | 안 |
| 都 | 도 | … | … | … | … |
| 立 | … | 립 | 립 | … | 립 |

の韓国語の対訳文字を入れて表6のようなテーブルを作成する。

そこで、2.1で変換されなかった文字に対して、日韓文字対訳隣接テーブルの中からその文字と隣接している文字にマッチしている韓国語対訳語を取り出し、韓国語文字に変換する。

#### 2.3.2 日韓文字対訳隣接テーブルの構造

テーブルはメモリを効率的に活用するために、図 3のように一方向リスト構造で構成する。x の文字は第一のノードに、y の文字は第二のノードに、NH(x,y) の対訳文字は第三のノードに置いて、各ノードに次のノードを指すポインターを置く。日韓文字対訳左隣接テーブルと日韓文字対訳右隣接テーブルの大きさはそれぞれ234Kbyteである。

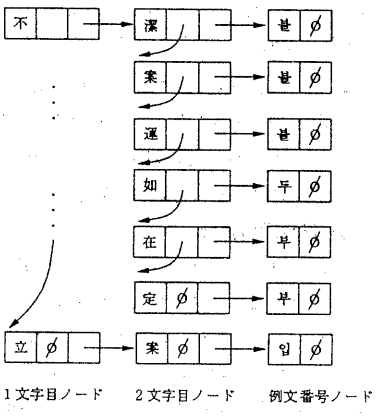
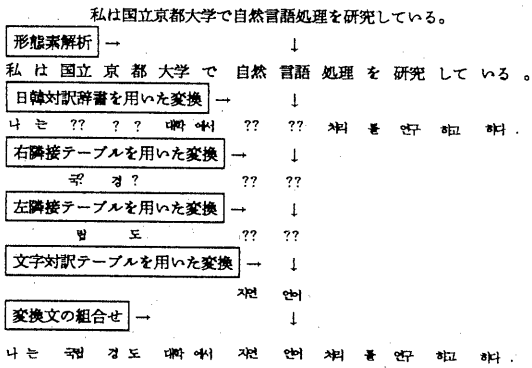



図 3: 日韓文字対訳隣接テーブルの構造

### 2.4 日本語文からの韓国語文への変換例

日本語文は表7のような変換段階を経て変換文に変換する。「私」、「は」、「大学」、「で」、「処理」、「を」、「研究」、「して」、「い」、「る」、「。」などは日韓対訳辞書に登録されていたので、それぞれ日韓対訳辞書よりその対訳語を引いて韓国語単語に変換し、日韓対訳辞書に登録されていない「国立」の「国」と「京」は日韓文字対訳左隣接テーブルによって、また「国立」の「立」と「都」は文字対訳右隣接テーブルによって韓国語文字に変換する。どちらも登録されていない「自」、「然」、「言」、「語」などは日韓文字対訳テーブルによって韓国語文字に変換する。最後に、各変換段階で変換した韓国語対訳語を組み合わせて韓国語変換文を生成する。この例では、「いる」の「い」が「하」に誤訳されたのを除くと、すべてが韓国語に正しく変換されている。

表 7: 日本語文からの韓国語文への変換過程




## 3 類似度による文対応

### 3.1 文対応の形態

日本語文は韓国語文に翻訳されるとき、1文が1文に訳される場合以外に、1文が2文、あるいは、2文が1文に訳されるときもある。日韓対訳文庫本5冊の5,818文を対象にして、日本語文と韓国語文との文対応の形態とその対応比率とを調査した結果を表8に示す。

また、日本語文と韓国語対訳文において文の文字数の比率は、約1,300文を対象にして調査した結果、約9:10であることが分かった。

表 8: 日韓文対応の形態と比率

| 文対応形態 | 頻度 | 比率 |
|---|---|---|
| 1-1 | 5734 | 98.6 |
| 1-0 or 0-1 | 2 | 0.0 |
| 2-1 or 1-2 | 78 | 1.3 |
| 3-1 or 1-3 | 4 | 0.1 |
| 合計 | 5818 | 100.0 |

### 3.2 文対応づけの概要

#### 3.2.1 文の類似度の計算

ところで、人間が文対応をするのを考えてみると、まず、原文をある程度まで訳し、その文に一番似ている翻訳文を原文に対応させると考えられる。そこで、本研究では、2章に述べた変換方法を通して生成された変換文に一番似ている文を翻訳文より探し、それを日本語文に対応させる方法を用いた。ここで、似ている翻訳文を探す方法としては、次のように定義した文字列一致度 $(MS)$ と文の長さの比率 $(CR)$ を用いた文の類似度 $(AD)$ を利用する。ここで日本語文を $Ji, i = 1, ..., n$、変換文を $J'i, i = 1, ..., n$、翻訳文を $Kj, j = 1, ..., m$ と表現することにする。

$$
\begin{array}{lll}
AD(J,K) & = MS(J',K) * CR(J,K) & \\
MS(J',K) & = \text{文字列一致度} & \\
CR(J,K) & = if\ (J < K * 0.9) & then\ J/(K*0.9) \\
 & & else\ (K*0.9)/J
\end{array}
$$

これは、文字列一致度が高いほど、また、文字数の比率が日本語文と韓国語文との平均文字比率に近いほど、対訳文である可能性が高いということを意味する。

また、文字列一致度は表9のように文字の順序を考慮した一致文字数をもって類似度を計算する。

表 9: 文字列一致度の計算式

$$S(A,B) = s(x,y)$$

$$s(i,j) = \begin{cases} 0 & \text{if } (i=0) \vee (j=0) \\ max \begin{pmatrix} s(i-1,j-1) + min(cm(i,j), W^*) \\ s(i-1,j), \\ s(i,j-1) \end{pmatrix} & if (1 \le i \le x) \wedge (1 \le j \le y) \end{cases}$$

$$cm(i,j) = \begin{cases} 0 & \text{if } (i=0) \vee (j=0) \\ cm(i-1,j-1) + m(i,j) & if (1 \le i \le x) \wedge (1 \le j \le y) \end{cases}$$

$$m(i,j) = \begin{cases} 1 & if a_i = y_j \\ 0 & if a_i \neq b_j \end{cases}$$

$*W$(連続値) = 4

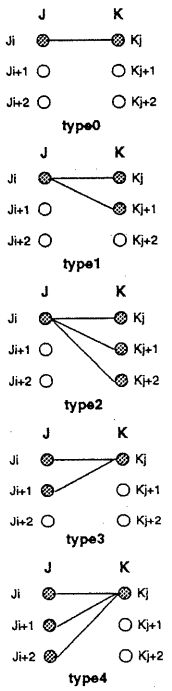
ここで、$W$ は、連続した一致に対するボーナスの最大値を定めるパラメータであり、我々は、$W = 4$ を採用した [5]。たとえば、図 4で「は」と「学」はそれぞれ 1 文字しかマッチしていないため、スコア 1 を与えるが、「生」は「学」に続いてマッチしているので、ボーナスを加えてスコア 2 を与える。これらのスコアを合わせた 4 がこの例文の文字一致度になる。

| | 彼 | は | 大 | 学 | 生 | で | あ | る |
|---|---|---|---|---|---|---|---|---|
| 私 | . | . | . | . | . | . | . | . |
| は | . | 1 | . | . | . | . | . | . |
| 学 | . | . | . | 1 | . | . | . | . |
| 生 | . | . | . | . | 2 | . | . | . |
| だ | . | . | . | . | . | . | . | . |

図 4: 文字列一致度の計算例

### 3.2.2 文対応

日本語文と韓国語文との対応には、表 8に示したように、5 つの対応形態がある。したがって、本稿では、文対応形態を決めるに、図 5のような五つの対応形態での類似度を求めて、その中で一番類似度が高い対応形態をとることにした。




$AD(type0)$
$= AD(J_i, K_j)$
$= MS(J'_i, Kj) * CR(J_i, K_j)$

$AD(type1)$
$= AD(J_i, K_j + K_{j+1})$
$= MS(J'_i, Kj + 1)$
$*CR(J_i, K_j + K_{j+1})$

$AD(type2)$
$= AD(J_i, K_j + K_{j+1} + K_{j+2})$
$= MS(J'_i, Kj + 2)$
$*CR(J_i, K_j + K_{j+1} + K_{j+2})$

$AD(type3)$
$= AD(J_i + J_{i+1}, K_j)$
$= MS(J'_{i+1}, Kj)$
$*CR(J_i + J_{i+1}, K_j)$

$AD(type4)$
$= AD(J_i + J_{i+1} + J_{i+2}, K_j)$
$= MS(J'_{i+2}, Kj)$
$*CR(J_i + J_{i+1}, +J_{i+2}, K_j)$

図 5: 日韓文対応の形態と類似度の計算

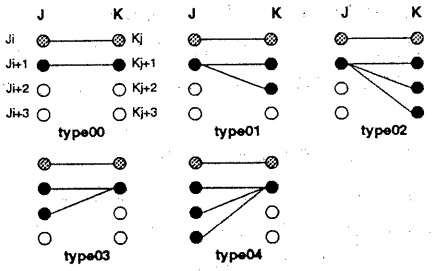


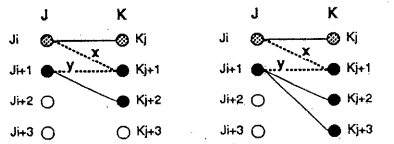

図 6: 日韓文対応の形態 2

ところが、1 文が複数文と対応している文対応形態 (type1,type2,type3,typ4) においての類似度を計算するときには、1 文と複数文での最後の文との文字一致度を用いる。例えば、type1 の類似度を計算するときに、文字列一致度は $MS(J'_i, Kj + 1)$ を用い、type3 の類似度には、$MS(J'_{i+1}, Kj)$ を用いる。これは、1 文と複数文の最後の文との類似度が高いというのが、その 1 文が複数文に対応していることを意味しているからである。

また、type0 の類似度が一番高いときには、文対応決定の信頼性を高めるために、図 6のようにその次の文対応形態の類似度も考慮して対応形態を決める。たとえば、日本語 1 文に対して韓国語 2 文が対応しても、2 つの文になっている韓国語文の 2 番目の文が短い場合には、type1 の類似度を求めるとき、$MS$ が小さくなりがちであるため、初めての文対応決定には type0 に決められる可能性が高い。そこで、その次の文同士の文対応も検討してから、現在の文対応を最終的に決める。

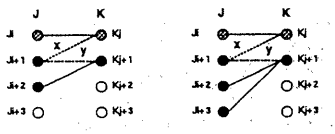
その次の文対応形態で類似度が一番高い対応形態が、

1. type00 であると、type0 に決まる。
2. type01、または type02 であると、
   x=$AD(J_i, K_{j+1})$,y=$AD(J_{i+1}, K_{j+1})$ を求め、x が y より大きければ、type1 に、そうでないと type0 に決める。




$if\ (x > y)\quad then\ type1\quad else\ type0$

3. type03、または type04 であると、
   x=$AD(J_{i+1}, K_j)$,y=$AD(J_{i+1}, K_{j+1})$ を求め、x が y より大きければ、type3 に、そうでないと type0 に決める。




$if\ (x > y)\quad then\ type3\quad else\ type0$

こうして、文対応を対訳テキストの先頭から順番に行ない、文対応がとられると、その次の文を順番に読み込んで文対応を行なう。

## 4 実験および考察

日韓学習文庫の4冊全部については入力その他の点から実験ができなかったので、特に対応が1対1になっていない58文とその周りの文を取り出し、それに対して実験を行なった。この実験文は文対応の形態が多様な文が続いて書かれており、また小説文であるため、変換文と翻訳文との文字一致度が低いので、文対応がうまくとられないと考えられるが、表10のように、100%という文対応の成功率を出した。図7の上半は文対応を行なう以前の日本語文と韓国語文であり、下半は文対応を行なった結果である。

表10: 実験1の評価

| 日韓 | 対応形態 | | | | | 合計 |
|---|---|---|---|---|---|---|
| 学習文庫 | 1-1 | 1-0 | 1-2 | 2-1 | 1-3 | |
| 文数 | 127文 | 1文 | 37文 | 19文 | 1文 | 185文 |
| 対応成功率 | 100% | 100% | 100% | 100% | 100% | 100% |

また、「人工知能と人間」という本の中の連続した約873文を取り出し、実験を行なった。この文は科学技術文で、type0の848文とその以外のtypeの25文になっている。これに対しても文対応が100%正しくとられた。ところが、翻訳誤謬によって翻訳されていない1つの日本語段落があったのでこれに対しては文対応が正しくとれなかった。評価の際、これは対象にしなかった。

表11: 実験2の評価

| 人工知能 | 対応形態 | | | | | | 合計 |
|---|---|---|---|---|---|---|---|
| と人間 | 1-1 | 1-0 | 1-2 | 2-1 | 1-3 | 3-1 | |
| 文数 | 847 | 1 | 9 | 13 | 1 | 2 | 873 |
| 対応成功率 | 100% | 100% | 100% | 100% | 100% | 100% | 100% |

## 5 おわりに

本研究では、人間が文対応を行なうであろう方法に基づいて、原文中の表現を3つの場合に分け、それぞれに対して妥当な変換の方法を考察することによって、原文を翻訳文に似ている文に変換し、その変換文と翻訳文との類似度を求めて一番類似度が高い文同士を文対応する方法を示した。また、実際例文1,058文に対して実験した結果、文対応が100%正しくとられた。これによって、日韓対訳テキストの文対応がほぼ完全に自動的に行なうことができるようになったので、これからの大量の対訳例文を用いる日韓自然言語処理研究に役立つと思われる。

本稿で提案したこの方法による文対応の成功率の信頼性を確かめるためには、今後、より大量の対訳文に対して実験を行なう必要があると思われる。文章には段落または章の区別のために普通その先頭に空白の記号をおくし、また段落の最後には、何文字かの空白があるので、文対応を段落別、または章ごとに区別して行なうのも当然のことと思われる。日韓対訳辞書の登録単語数を増やすことによって、さらに、変換率をあげることができ、文対応の成功率はさらに安定性をもつと思われる。

## 参考文献

[1] M. Nagao: A Framework of a Mechanical Translation between Japanese and English by Analogy Principle, in *ARTIFICIAL AND HUMAN INTELLIGENCE*, Elsevier Science Publishers, pp173-180 (1984).

[2] William A. Gale, and Kenneth W. Church: A Program for Aligning Sentences in Bilingual Corpora, *Computational Linguistics*, Vol.19, No.1, pp75-90 (1993).

[3] Peter F. Brown, Jennifer C. Lai, and Robert L. Mercer: Aligning Sentences in Parallel Corpora, *Proc. of ACL'91*, pp169-176 (1991).

[4] Martin Kay and Martin Röscheisen: Text-Translation Alignment, *Computational Linguistics*, Vol.19, No.1, pp121-142 (1993).

[5] 黄 道三、長尾 真、佐藤 理史: 日本語の分類語彙表からの韓国語の分類語彙表の作成、情報処理学会自然言語処理研究会報告、94-12 (1993).

[6] 長尾 真のほか: 日本語形態素解析システム JUMAN 使用説明書 version 0.6、京都大学工学部長尾研究室 (1992).

[7] 長尾 真: 人工知能と人間、岩波書店 (1983).

[8] 日韓対訳文庫、*Darakwon* (1991).

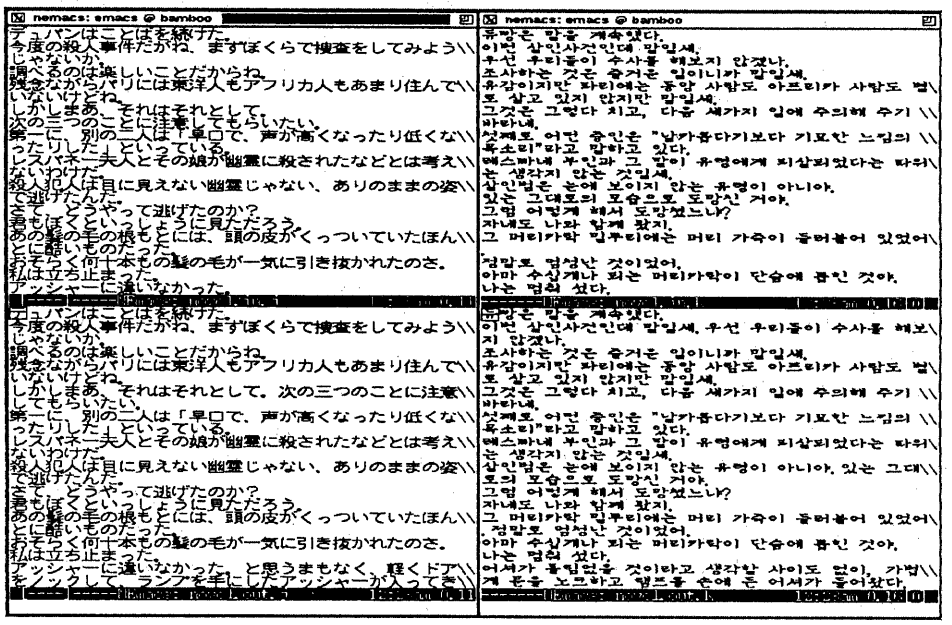

nemacs: emacs @ bamboo
デュパンはことばを続けた。
今度の殺人事件だがね、まずぼくらで捜査をしてみようじゃないか。
調べるのは楽しいことだからね。
残念ながらパリには東洋人もアフリカ人もあまり住んでいないけどね。
しかしまあ、それはそれとして。
次の三つのことに注意してもらいたい。
第一に、別の二人は「早口で、声が高くなったり低くなったりした」といっている。
レスパネー夫人とその娘が幽霊に殺されたなどとは考えないわけだ。
殺人犯人は見に見えない幽霊じゃない、ありのままの姿で逃げたんだ。
さてどうやって逃げたのか？
君もぼくといっしょうに見ただろう。
あの髪の毛の根もとには、頭の皮がくっついていたほんとに酷いものだった。
おそらく何十本もの髪の毛が一気に引き抜かれたのさ。
私は立ち止まった。
アッシャーに違いなかった。
と思うまもなく、駆くドアをノックしてランプを手にしたアッシャーが入ってきた。
Mark set
勇敢な男を演じ続けた。
今度の殺人事件だがね、まずぼくらで捜査をしてみようじゃないか。
調べるのは楽しいことだからね。
残念ながらパリには東洋人もアフリカ人もあまり住んでいないけどね。
しかしまあ、それはそれとして。次の三つのことに注意してもらいたい。
第一に、別の二人は「早口で、声が高くなったり低くなったりした」といっている。
レスパネー夫人とその娘が幽霊に殺されたなどとは考えないわけだ。
殺人犯人は見に見えない幽霊じゃない、ありのままの姿で逃げたんだ。
さてどうやって逃げたのか？
君もぼくといっしょうに見ただろう。
あの髪の毛の根もとには、頭の皮がくっついていたほんとに酷いものだった。
おそらく何十本もの髪の毛が一気に引き抜かれたのさ。
私は立ち止まった。
アッシャーに違いなかった。と思うまもなく、駆くドアをノックしてランプを手にしたアッシャーが入ってきた。
勇敢な男を演じ続けた。
今度の殺人事件だがね、まずぼくらで捜査をしてみようじゃないか。
調べるのは楽しいことだからね。
残念ながらパリには東洋人もアフリカ人もあまり住んでいないけどね。
しかしまあ、それはそれとして。次の三つのことに注意してもらいたい。
第一に、別の二人は「早口で、声が高くなったり低くなったりした」といっている。
レスパネー夫人とその娘が幽霊に殺されたなどとは考えないわけだ。
殺人犯人は見に見えない幽霊じゃない、ありのままの姿で逃げたんだ。
さてどうやって逃げたのか？
君もぼくといっしょうに見ただろう。
あの髪の毛の根もとには、頭の皮がくっついていたほんとに酷いものだった。
おそらく何十本もの髪の毛が一気に引き抜かれたのさ。
私は立ち止まった。
アッシャーに違いなかった。と思うまもなく、駆くドアをノックしてランプを手にしたアッシャーが入ってきた。
勇敢な男를 계속했다.
이번 살인사건인데 말일세.
우선 우리들이 수사를 해보지 않겠나.
조사하는 것은 즐거운 일이니까 말일세.
유감이지만 파리에는 동양 사람도 아프리카 사람도 별로 살고 있지 않지만 말일세.
그것은 그렇다 치고, 다음 세가지 일에 주의해 주기 바라네.
첫째로 어떤 증인은 "날카롭다기보다 기묘한 느낌의 목소리"라고 말하고 있다.
레스파네 부인과 그 딸이 유령에게 피살되었다는 따위는 생각지 않는 것일세.
살인범은 눈에 보이지 않는 유령이 아니야. 있는 그대로의 모습으로 도망친 거야.
그럼 어떻게 해서 도망했느냐?
자네도 나와 함께 봤지.
그 머리카락 밀푸이에는 머리 가죽이 들러붙어 있었어.
.점말로 엄청난 것이었어.
아마 수십개나 되는 머리카락이 단숨에 뽑힌 것야.
나는 멈춰 섰다.
어쩌가 틀림없을 것이라고 생각할 사이도 없이, 가볍게 문을 노크하고 램프를 손에 든 어쩌가 들어왔다.
Mark set

図 7: 文対応の実験結果 1

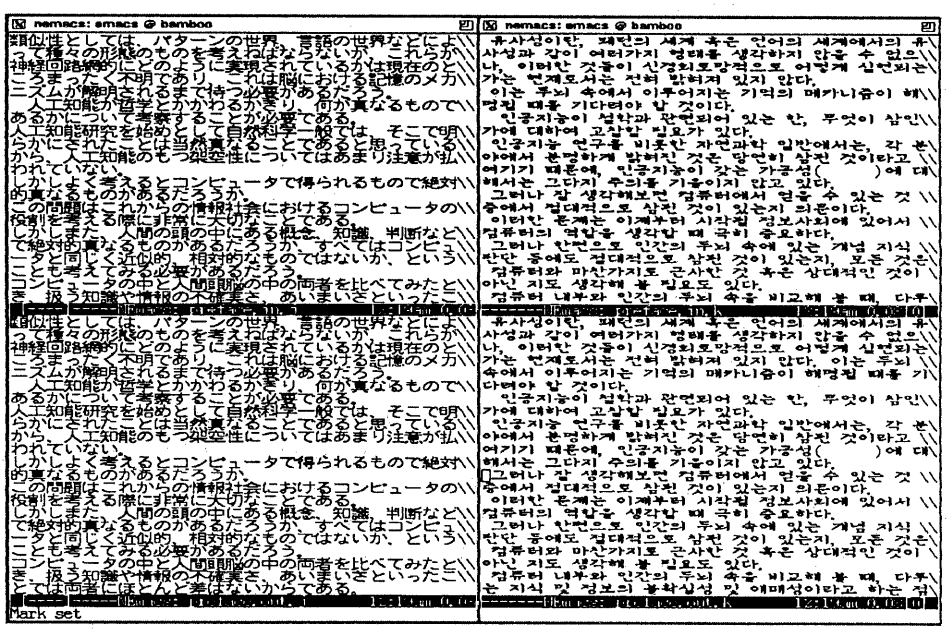

nemacs: emacs @ bamboo
類似性としては、パターンの世界、言語の世界などによって種々の形態のものを考えねばならないが、これらが神経回路網的にどのように表現されているかは現在のところまったく不明であり、これは脳における記憶のメカニズムが解明されるまで待つ必要があるだろう。
人工知能が哲学とかかわるかぎり、何が真なるものであるかについて考察することが必要である。
人工知能研究を始めとして自然科学一般では、そこで明らかにされたことは当然真なることであると思っている。
から、人工知能のもつ架空性についてはあまり注意が払われていない。
しかしよく考えるとコンピュータで得られるもので絶対的真なるものがあるだろうか。
この問題はこれからの情報社会におけるコンピュータの役割を考える際に非常に大切なことである。
しかしまた、人間の頭の中にある概念、知識、判断などは絶対的真なるものがあるだろうか。すべてはコンピュータと同じく近似的、相対的なものではないか、という二つとも考えてみる必要があるだろう。
コンピュータの中と人間脳の中の両者を比べてみたとき扱う知識や情報の不確実さ、あいまいさといったことでは両者にほとんど差はないからである。
Mark set
類似性としては、パターンの世界、言語の世界などによって種々の形態のものを考えねばならないが、これらが神経回路網的にどのように表現されているかは現在のところまったく不明であり、これは脳における記憶のメカニズムが解明されるまで待つ必要があるだろう。
人工知能が哲学とかかわるかぎり、何が真なるものであるかについて考察することが必要である。
人工知能研究を始めとして自然科学一般では、そこで明らかにされたことは当然真なることであると思っている。
から、人工知能のもつ架空性についてはあまり注意が払われていない。
しかしよく考えるとコンピュータで得られるもので絶対的真なるものがあるだろうか。
この問題はこれからの情報社会におけるコンピュータの役割を考える際に非常に大切なことである。
しかしまた、人間の頭の中にある概念、知識、判断などは絶対的真なるものがあるだろうか。すべてはコンピュータと同じく近似的、相対的なものではないか、という二つとも考えてみる必要があるだろう。
コンピュータの中と人間脳の中の両者を比べてみたとき扱う知識や情報の不確実さ、あいまいさといったことでは両者にほとんど差はないからである。
유사성이란, 패턴의 세계 혹은 언어의 세계에서의 유사성과 같이 여러가지 형태를 생각하지 않을 수 없으나, 이러한 것들이 신경회로망적으로 어떻게 실현되는가는 현재로서는 전혀 밝혀져 있지 않다.
이는 두뇌 속에서 이루어지는 기억의 메카니즘이 해명될 때를 기다려야 할 것이다.
인공지능이 철학과 관련되어 있는 한, 무엇이 참인가에 대하여 고찰할 필요가 있다.
인공지능 연구를 비롯한 자연과학 일반에서는, 각 분야에서 분명하게 밝혀진 것은 당연히 참된 것이라고 여기기 때문에, 인공지능이 갖는 가공성( )에 대해서는 그다지 주의를 기울이지 않고 있다.
그러나 잘 생각해보면 컴퓨터에서 얻을 수 있는 것 중에서 절대적으로 참된 것이 있는지 의문이다.
이러한 문제는 이제부터 시작될 정보사회에 있어서 컴퓨터의 역할을 생각할 때 극히 중요하다.
그러나 근본적으로 인간의 두뇌 속에 있는 개념 지식 등 한단 등에도 절대적으로 참된 것이 있는지, 모든 것은 컴퓨터와 마찬가지로 근사한 것 혹은 상대적인 것이 아닌 지도 생각해 볼 필요도 있다.
컴퓨터 내부와 인간의 두뇌 속을 비교해 볼 때, 다루는 지식 및 정보의 불확실성 및 애매성이라고 하는 점
Mark set
유사성이란, 패턴의 세계 혹은 언어의 세계에서의 유사성과 같이 여러가지 형태를 생각하지 않을 수 없으나, 이러한 것들이 신경회로망적으로 어떻게 실현되는가는 현재로서는 전혀 밝혀져 있지 않다. 이는 두뇌 속에서 이루어지는 기억의 메카니즘이 해명될 때를 기다려야 할 것이다.
인공지능이 철학과 관련되어 있는 한, 무엇이 참인가에 대하여 고찰할 필요가 있다.
인공지능 연구를 비롯한 자연과학 일반에서는, 각 분야에서 분명하게 밝혀진 것은 당연히 참된 것이라고 여기기 때문에, 인공지능이 갖는 가공성( )에 대해서는 그다지 주의를 기울이지 않고 있다.
그러나 잘 생각해보면 컴퓨터에서 얻을 수 있는 것 중에서 절대적으로 참된 것이 있는지 의문이다.
이러한 문제는 이제부터 시작될 정보사회에 있어서 컴퓨터의 역할을 생각할 때 극히 중요하다.
그러나 근본적으로 인간의 두뇌 속에 있는 개념 지식 등 한단 등에도 절대적으로 참된 것이 있는지, 모든 것은 컴퓨터와 마찬가지로 근사한 것 혹은 상대적인 것이 아닌 지도 생각해 볼 필요도 있다.
컴퓨터 내부와 인간의 두뇌 속을 비교해 볼 때, 다루는 지식 및 정보의 불확실성 및 애매성이라고 하는 점

図 8: 文対応の実験結果 2