

基本的な事例構文と翻訳例を利用した 英日機械翻訳

武田 紀子

成蹊大学工学部経営工学科

E-mail:takeda@is.seikei.ac.jp

簡単な構文例と、その構文により生成される英文と訳文の組からなる翻訳例を事例として学習して、英語の入力文に対し、学習された事例のいくつかを組み合わせて訳文を生成する英日機械翻訳システムについて述べる。ここで、事例としては、中学英語の学習課程に沿って、重要とされている構文、典型的な構文や例文が与えられる。このため、事例データベースは、より少ない、有効な事例から構成される。そして、入力文の処理に対する事例の検索は、構文例と、翻訳例の2段階の検索で行われることにより、検索範囲が制限され、効率を上げることができた。

また、システムは、より高度な事例の追加によって進化することもできるし、いくつかの事例の組み合わせにより生成される新しい事例を加えることによっても進化することもできる。

English-Japanese Machine Translation Using a Collection of Basic Syntactic Examples and Translation Examples

Noriko Takeda

Department of Information Science,
Faculty of Engineering,
Seikei University

This paper describes the English-Japanese machine translation system that generates Japanese text by a study and a combination of basic syntactic examples and translation examples. The examples registered to this system are those important and basic ones extracted from English textbook of Japanese junior high school standard. Since the example database contains only effective examples so its size is fairly small. Retrieval of examples that is similar to input sentence is done through two steps, retrieve syntactic examples, and then translation examples. Consequently, the retrieval efficiency has been improved by limiting the range of retrieval.

The system can be improved by adding more complicated examples and new examples generated through a combination of several examples.

- 119 -

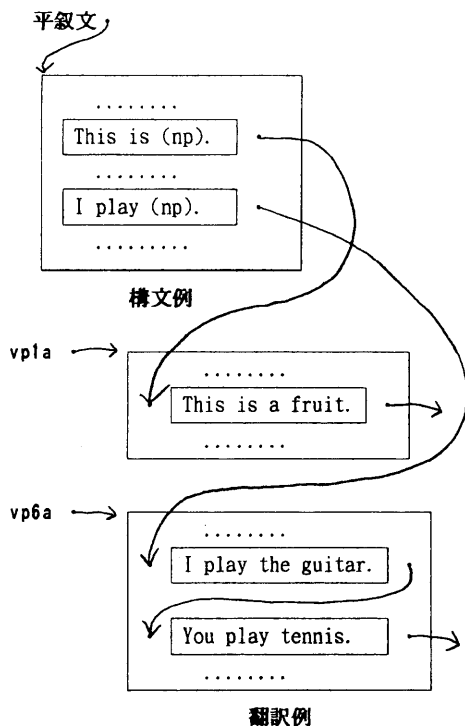


図3. データベースの関係

2. 入力文と似た事例が見つかり、簡単な語の入れ替えで、選択された事例を入力文に合うものに変更できる場合は、語の入れ替えにより、解析は終わる。

また、入力文に対する適当な事例が見つからない場合は、文全体に対する構文解析を行う。これには、まず、各語の形態素解析を行い、主動詞の候補を決定する。次のような場合は、主動詞は、一意的に決まる。

1. 対応する構文例が見つかり、それにより動詞が決定される。
2. 品詞として、動詞を持つ語が、1つしかない。
3. 品詞として、動詞しか持たない語がある。

動詞が、一意的に決まらない場合は、以降の解析で、構文的に矛盾のあるものを除きながら動詞の候補をしぼっていく。

動詞が決まると、その動詞の動詞パターンから、文の各部分の文法的範疇を予測して、解析を行う。この解析においても、各句（名詞句、前置詞句等）に対する事例を参照し、語の比較による解析を行い、それで解決出来ない場合には、文法規則による解析

を行う。

3. 2. 意味距離計算

入力文と、各翻訳例との類似性の判断は、簡単な意味距離計算によってなされる。各翻訳例には、[2. 2.]で述べたように文のキーワードとなる語が、明示的に示されている。入力文の解析結果から、キーワードを取り出し、これと、事例のキーワードから次のように類似度の計算をする。

$$\text{類似度} = \sum_{i=1}^n W_i S_i$$

W_i は、各キーワードの重み

S_i は、キーワード間の意味的距離

値が、大きいほど意味的に近い。

n は、キーワードの個数

ここで、動詞は重み(W_i)を大きくしているのので、入力文と同じ動詞からなる翻訳例が選ばれる。意味的距離は、最大1(同じ語)として、意味シソーラスにおける互いの距離によってその値は決定される。そして、計算された類似度の値が最大となる事例が、最も近い事例とみなされる。

【例4】You play the piano.

に対する類似度の計算

翻訳例1) I play the guitar.

$$0.8(\text{you}, \text{I}) + 0.8(\text{piano}, \text{guitar}) = 1.6$$

翻訳例2) You play tennis.

$$1(\text{you}, \text{you}) + 0.3(\text{piano}, \text{tennis}) = 1.3$$

翻訳例1の方が、意味的に近いとみなされる。

3. 3. 翻訳のアルゴリズム

入力された英文の翻訳は、以下のようなアルゴリズムに沿ってなされる。

[1] 英文の入力

[2] 一致する構文例さがし、見つかった場合は、

[3]へ、見つからなかった場合は、[8]へ行く。

[3] 構文例と対応する翻訳例の中に、入力文と同じ英文が見つかる、その訳文をそのまま出し、翻訳を終える。

[4] 同じ動詞を持つ翻訳例の中から、キーワードの意味距離計算により、意味的に最も近い事例を見つける。(翻訳例1とする)

[5] 入力文と同じ構文をもつ翻訳例を1つ選ぶ(翻訳例2とする)

[6] 翻訳例1のキーワードが、入力文と等しい場合、翻訳例2を入力文と翻訳例1を参照しながら、入力文に合うように書き換える。そして、この操作によって新しく作成された翻訳例から、訳文を生成し、翻訳を終える。

[7] 翻訳例1のキーワードと、入力文が等しく

平叙文	
.....	
This is (np).	(1)
That is (np).	(2)
.....	
否定文	
.....	
Taht is not (np).	(3)
.....	

図 4. 構文例データベース

- ない場合は、異なっている部分を入れ替え、入力文と同じキーワードを持つ翻訳例（これを新翻訳例 1 とする）に変更し [5] に行く。
- [8] 一致する構文例が見つからない場合は、英文の形態素解析を行い、動詞を見つける。
- [9] この動詞の動詞パターンをを参照し、最も近い文法的構造をしている構文例を見つける。そして、この構文からなる翻訳例を 1 つ選ぶ。（翻訳例 1 とする）
- [10] 同じ動詞を持つ翻訳例の中からキーワードの意味距離計算により、意味的に最も近い翻訳例を見つける。（翻訳例 2 とする）
- [11] 翻訳例 1 と 2 の対応している部分と比較し、異なっている部分の訳を入れ替え、新しい翻訳例を作成する。（翻訳例 3 とする）
- [12] 翻訳例 3 と入力文を比べ、翻訳例 3 の入力文と異なった部分を変更したのち、訳文を生成し、翻訳を終える。

3. 4. 翻訳の流れ

次に上で述べたアルゴリズムに沿って実際にどのように英日の翻訳がなされるかについて述べる。

図 4、図 5 のように事例データベースに構文例、翻訳例が、登録されているとする。

[例 5] That is not Tom. の翻訳の流れ

1. alg. 2 (アルゴリズムの [2]) より、同じ構文を持つ構文例 (3) が見つかる。
2. 入力文と同じ英文は、事例として登録されていないので、alg. 4 より、意味的に最も近い翻訳例 (6) を見つける。
つまり、入力文は、構文的には、
That is not a table.
に近いが、意味的には、
That is Tom.
に近いとみなされる。
3. 翻訳例 (6) のキーワードと、入力文のキーワードは等しいので、alg. 6 より、まず、構文例 (3) と同じ構造を持つ翻訳例 (7) を取り

vpla	
.....	
This is Tom.	(4)
((“こちら” this)(“は、” subj_jyoshi)	
(“トム” Tom)(“です” is)(“。” nil))	
(is) (This) (Tom)	
(this Tom) ((hum) (hum))	
That is an apple.	(5)
((“あれ” that)(“は、” subj_jyoshi)	
(“” an) (“りんご” apple)	
(“です” is) (“。” nil))	
(is) (That) (an apple)	
(that apple) ((conc) (conc plant))	
That is Tom.	(6)
((“あちら” that)(“は、” subj_jyoshi)	
(“トム” Tom)(“です” is)(“。” nil))	
(is) (That) (Tom)	
(that Tom) ((hum) (hum))	
.....	
That is not a table.	(7)
((“あれ” that)(“は、” subj_jyoshi)	
(“” a) (“テーブル” table)	
(“ではありません” is not) (“。” nil))	
(is not) (That) (a table)	
(that table) ((conc) (conc))	
.....	

図 5. 翻訳例データベース

出し、これと、翻訳例 (6) とを比較して、異なっている部分を入れ替える。

That の訳 あれ → あちら
a table → Tom

4. 訳をだす。

“あちらは、トムではありません。”

[例 6] This is not my mother. の翻訳の流れ

1. 一致する構文例見つからなかったので、alg. 8, 9 より、近い構文例 (3) を見つける。
2. alg. 10 より、意味的に近い翻訳例 (4) を見つける。
3. alg. 11 より、(3) と同じ構造をもつ翻訳例 (7) と (4) の下に示した対応する部分を入れ替えて、図 6. のような新しい翻訳例 (7') を作る。
That (“あれ”) → This (“こちら”)
a table (“テーブル”) → Tom (“トム”)
4. これと、入力文比較して、異なっている部分を入れ替え、訳をだす。 (alg. 12)
Tom → my mother
“こちらは、私の母ではありません。”

This is not Tom.
 ("こちら" this)("は," subj_jyoshi)
 ("トム" Tom)("ではありません" is not)
 ("。" nil))
 (is not) (This) (Tom)
 (this Tom) ((hum) (hum))

図 6. 翻訳例 7'

4. 結果と考察

英語の学習においては、構文的には、より単純なものから複雑な構文へ、順次、いろいろな文法規則を学び、それらに応用し、組み合わせてできる文を学び、それと同時に、語彙を増やすことにより、表現力、理解力を増している。この過程において、学校の学習課程に沿った方法は、ある程度有効であると思われる。本システムでは、この課程に沿って、構文例を登録し、システムを進化させている。

現在、中学校 1 年程度の構文例（文の構文例は 58 例、その他の句の例は 7 例）、翻訳例（文の翻訳例は、89 例、その他の句の例は、33 例）を事例データベースに登録し、教科書、参考書等にある英文を入力して翻訳実験を行っている。

ここで、入力文に対する事例の適用が、構文例と翻訳例の 2 段階でなされているので、構文例により、検索されるべき翻訳例の範囲を狭くすることができ、検索効率を上げることができた。

また、

- ・構文例を利用した構文解析では、語の比較でなされる部分が多いので、文法規則のみによる処理よりも効率良くなされる。

- ・翻訳例が、構文例と対応しており、キーワードが、明示的に示されているため、入力文と、翻訳例との対応がとり易い。このため、複数の事例の組み合わせによる処理が容易にできる。

- ・名詞句、前置詞句等の句の構文例、翻訳例も登録されているので、各句の段階で事例を適用したり、事例を狭い範囲から適用して、それらをより広い範囲で組み合わせたりすることが可能である。

- ・入力文に対しては、文の翻訳例が、与えられるので読みやすい訳文を生成することができる。

しかし、構文解析において参照される構文例が、語で与えられているということは、反面、一般性に欠けるということの意味する。中学生レベルの英文の翻訳では、与えられている単語も限られており、構文も比較的単純であるので、よく使われる構文が、容易に特定できるため、この方法は、有効であるが、語彙が増えてきた場合や、より複雑な文を処理する場合には、事例を増やすことによる対応のみでは、構文例のデータベースが大きくなりすぎて、検索効

率が著しく低下することになる。今は、入力文と対応する構文例が見つからないと、文法規則による解析を行った後、より近い構文例を求めているが、

- ・構文例を、段階的により一般的なものになるように定義して、それらを登録する。

ことも考えられる。

また、本システムでは、翻訳過程において、入力文は、登録されている翻訳例と同じ形に変換されてから、訳文が生成される。従って、これらを新しい事例として登録することにより、語彙の知識を増やすことができるし、構文的にも、拡張した用法に対処することができるようになる。

これからは、中学の学習課程の進度に沿って、より多くの構文例、翻訳例を登録して、実験を続けていくが、そのときの課題として、

- ・より複雑になる、各事例の組み合わせをいかに効率よく行うか

- ・事例の組み合わせの優先順位をどうするか等が挙げられる。

そして、構文例では、十分対応できない

- ・文のどこにでも挿入することのできる副詞句、挿入句の処理

- ・複雑な並列関係の処理

についても、考慮していかなければならない。

【参考文献】

- 1) 古瀬 蔵, 隅田 英一朗, 飯田 仁: 経験的知識を活用する変換主導型機械翻訳, 情報処理学会論文誌, Vol. 35, No. 3, pp. 414-425(1994).
- 2) 宇津呂 武仁: 類似度テンプレートを利用した高速類似用例検索, 情報処理学会研究会報告, Vol. 94, No. 77, pp. 33-40(1994).
- 3) 好田 実 監修: チャート式シリーズ, 中 1 英語, 中 2 英語, 数研出版(1995).
- 4) 斎藤 次郎 監修: ハイトップ中 1 英語, 中 2 英語, 旺文社(1996).
- 5) 藤井 篤, 乾 健太郎, 徳永 健伸, 田中 穂積: 多義性解消に用いる事例の獲得, 言語処理学会第 2 回年次大会, pp. 261-264(1996).
- 6) 宇津呂 武仁, 松本 祐治: 対訳辞書および統計情報を用いた二言語対訳テキスト照合, コンピュータソフトウェア, Vol. 12, No. 5, pp. 12-21(1995).
- 7) 側嶋 康博: 多元的類似度計算に基づく文脈を考慮したボトムアップ構文解析法, 情報処理学会研究会報告, Vol. 95, No. 27, pp. 61-66(1995).
- 8) 野美山 浩: 事例の一般化による機械翻訳, 情報処理学会論文誌, Vol. 34, No. 5, pp. 905-912(1993).
- 9) A.S. Hornby: Guide to Patterns and Usage in English, second edition, Oxford University Press(1977).