

機械翻訳から見た日本語

小原 京子

(慶應義塾大学理工学部)

1. はじめに

昨今では、パソコン上で動く翻訳ソフトが多数市販され、またウェブ上で公開された「お試し版」翻訳ソフトを起動して文書やホームページを翻訳することも可能になった。しかし、翻訳ソフトの内部の仕組みはいわばブラックボックスで通常ユーザーには知る術もない。

自然言語処理や人工知能の分野では翻訳ソフトのことを機械翻訳システムあるいは単に機械翻訳という。筆者は、かつてコンピュータ・メーカーで日本語から英語への機械翻訳システム(日英翻訳システム)の研究に携わった経験がある。現在では言語学を専門とする者として、また理工系大学生に英語を教える者として、機械翻訳にとって翻訳困難な文は英語学習者にとっても習得困難なことが多いと感じている。本稿では、まず日英翻訳システムを例に取りながら、その内部処理の流れを紹介しつつ各局面で問題となる日本語の特徴について見る。次に、英語から日本語への翻訳システム(英日翻訳システム)のアウトプットと翻訳家による日本語訳とを比べてみて、機械翻訳にとっては「お手上げ」なものにもかかわらず翻訳家がこなれた訳に仕上げている例を分析し、日本語の特徴をいくつか挙げてみる。

2. 日英機械翻訳における日本語解析の問題点

機械翻訳システムは処理の各段階でルールと辞書を参照する。以下ではトランスファー方式と呼ばれる翻訳方式による日英機械翻訳の処理の流れを見ていく¹⁾。

2.1 形態素解析

ここではまず入力文を形態素(意味を持つ最小の言語単位)に分け、それら各々に品詞付けをする。日本語の形態素解析で厄介なのは、この「単語切り」である。

日本語は英語などのインド・ヨーロッパ言語と異なり語と語の間がスペースで区切られてはいない。従って、翻訳システムはシステム内部の形態素解析辞書を参照しながら単語切りをしていく。たとえば、(1)のようになすべて平仮名で書かれたフレーズが入力されたとする。

(1) くるまでまつわ

日本語話者がこのフレーズを聞いた場合には、韻律や文脈により「(あなたが)来るまで待つわ」と言っているのか「車の中で待つわ」と言っているのか推測がつく。ところが、現在の翻訳システムはまだ文脈を考慮に入れて解析するには至っておらず、二つの解析結果のうちどちらがその文脈にふさわしいか決定できない。

2.2 構文解析

ここでは形態素解析の結果を内部表現に変換し、文の係り受けを決定する。これは機械翻訳にとって最も困難なプロセスである。日本語話者は、いかに文が長くても、各々の語句の意味と文脈からたいていの場合瞬時に係り受け関係を決定して文の意味を理解することができる。しかしながら、機械翻訳では文が長くなれば急速に誤り率が增大する。文ではなく名詞句だけをとりまいても、日本語の名詞句では、形容詞や形容動詞さらに「名詞+『の』」の形の修飾語句が連なって一つの名詞に先行していることは多々ある。次の(2)では、いずれも「形容詞+名詞+『の』+名詞」の順に品詞が並んでいる。

(2) a. 鋭い目の子供
b. 優しい女の子

通常 (2a) では、形容詞「鋭い」が最初の名詞「目」にかかると思うのが最も自然であろう。これに対し、同じ品詞が同じ順に並んでいる (2b) では形容詞「優しい」は「名詞+『の』+名詞」である「女の子」全体にかかると思うのが妥当である。日本語話者はこのように同じ品詞列でもそれぞれの語句の意味や文脈から最も意味的にふさわしい係り受けを決定することができる。機械翻訳ではどうであろうか。筆者の所有するパソコン用の日英翻訳ソフトは (2a) を “the child of a sharp eye”、(2b) を “a girl kind to” と訳したので係り受けは正しく決定できたようである²⁾。しかし、文レベルになると日本語には重文や複文が多いこともあり、機械翻訳は係り受け決定に失敗することが多くなっていく。

2.3 変換

トランスファー方式では、ここで日本語入力文の内部構造を英語の内部構造に変換しさらに訳語を決定する。構造変換で問題になるのは、日本語と英語の構造が一对一に対応しないケースである。また、訳語決定で厄介なのは、日本語にはない区別が英語に存在する場合である。

まず、構造変換に関して、(3) を見てほしい。

- (3) a. 私は雨に降られた。
b. It rained on me.

日本語文 (3a) は下線部「られ」が用いられていることからわかるように、いわゆる受身構文の一例である。ところがこれに対する英語訳として最も適切な (3b) では受動態は用いられていない。(3a) が受身構文ではあっても意味的には特殊な「迷惑の受身」といわれる構文の一例であること、また、英語動詞 “rain” は自動詞であり受動態にできない動詞であることから、通常受動態ではなく “on me” を使って (3a) の意味を英語で表現する。機械翻訳システムで (3b) のような訳を生成するには (3a) の「に」格の前の名詞「雨」が通常受身文の「に」格の前に現れる名詞とは異なり「動作主」ではないこと、動詞「降る」が自動詞であること、従って (3a) は「迷惑の受身」であること、を解析しなくてはならない。

また、訳語決定に関しては、たとえば日本語動詞「焼く」を例にとると、対応する英語の動詞を決定する際には、対象物が何であるか、何を使ってどういう状態で焼くのか等によって “bake,” “toast,” “roast,” “broil,” “burn,” “grill,” “sizzle,” “fry” などと訳し分ける必要がある。機械翻訳では意味辞書を用いてこのような訳し分けを行う。筆者所有の日英翻訳ソフトは「パンを焼く」を “bread is baked” と訳し、次候補として “bread is toasted” を出力した。

2.4 英文生成

このように日英機械翻訳システムでは日本語の形態素解析、構文解析、そして変換のプロセスを経てはじめて英文生成の段階に到達するが、各々の段階で様々なハードルがあることがおわかりになったかと思う。英文生成では、冠詞などの決定詞を生成し、性・数の一致に考慮して最終的に英語文を出力する。

3. 英語らしさ、日本語らしさと英日機械翻訳システム

筆者は、英語で書かれた小説と翻訳家によるそれらの日本語訳との比較対照を行っている。主に登場人物の移動に関する表現を比較し、同じ状況を表現する際に、英語と日本語にとって最も自然な表現とはそれぞれどのようなものか、さらに両者はどのような点で異なるか、について調べている。そうすることにより、英語らしさ日本語らしさとは何か、ということを知る手がかりになると考えたからである。

さらに、同じ状況を描写する際の英語と日本語にとってそれぞれ最も自然な表現のパターンの対を抽出することにより、それらを機械翻訳システムの変換ルールとして使うことができるのではないかと、そうすれば目的言語における自然な表現（いわゆる「こなれた」訳）を機械翻訳が生成できることになり機械翻訳の翻訳の質の向上に役立つのではないかと、とも考えている。

小説とその翻訳における英日比較対照研究で明らかになった両言語の違いは、二つある。一つは文構造に関する違い、もう一つはレトリックに関する違いである。

そして、このような英日の違いを既存の翻訳ソフトがどの程度扱えることができるのかを見極めるために、英語小説から抜き出した登場人物の移動に関する文を翻訳

ソフトで翻訳し翻訳家の日本語訳と比べてみた。その結果の一部を以下に簡単に紹介する。

3.1 文構造

英語では、移動の経路を表わすのに “over,” “up,” “down,” “through” などの前置詞や動詞に後続する不変化詞(verb particle)を用いることが多い。これに対し、これらを日本語に翻訳する際には動詞を使わざるを得ないことがしばしばある。

(4) a. 英語の原文

Here was Matthew Cuthbert placidly driving over the hollow and up the hill.

(Lucy Maud Montgomery. *Anne of Green Gables*)

b. 翻訳家の日本語訳

マシュー・クスバートは悠々と窪地を抜けて丘を上って行くのである。

(村岡花子訳『赤毛のアン』)

c. 英日翻訳ソフト

ここに、マシュー・クスバートが、平穩に穴に関して、および丘を上へ運転していきました。

(4) では下線で示したように (4a) の “over” と “up” はそれぞれ翻訳家により (4b) で「抜けて」、「上って」と動詞を使って日本語に訳されている。その結果、英語の原文は単文だが、翻訳家による日本語訳の文構造は接続助詞「て」による重文となっている。このように、英語では前置詞その他の不変化詞を用いた単文構造で表現されるものが、対応する日本語訳では「て」接続による重文構造となることが非常に多い。ところが、(4c) の下線部からわかるように、既存の英日翻訳ソフトではこのような英日の違いはまだあまり考慮に入れられていないようである。

3.2 レトリック

さらに機械翻訳システムにとっては困難と思われるのが、英日のレトリックの違いである。英語では人物の移動を移動動詞を使って明示的に表現するのに対し、日本語では移動動詞をあまり用いず代わりに人物の移動経路

上に存在する建物やその一部を記述して移動が起こったことを暗示的に描写することが多い。その結果、英語では動的な表現となるが、日本語では動詞が使われないため静的な表現となる。

(5) a. 英語の原文

Finally he returned to the pawnbroker's, and, having thumped vigorously upon the pavement, he went up to the door.

(Sir Arthur Conan Doyle. *The adventures of Sherlock Holmes*.)

b. 翻訳家の日本語訳

そして最後に、質屋の前の敷石をとんとんと強く叩いてから、ドアを叩いた。

(延原謙訳『シャーロック・ホームズの冒険』)

c. 英日翻訳ソフト

最後に、彼は質屋に戻りました、また、舗道で活発に強打して、彼はドアまで行きました。

(5a) では、下線部で示したように “returned” と “went” の二つの移動動詞が使われている。ところが、翻訳家の日本語訳 (5b) では移動動詞は一つも出現しない。「質屋」と「ドア」の二つの場所表現がシャーロック・ホームズがこれらの場所に移動したことを暗示的に示唆しているだけである。一文ずつ入力文を直訳することが中心の既存の翻訳ソフトにはこのような翻訳家の離れ技は無理のようである。翻訳ソフトの出力 (5c) では、原文 (5a) の下線部の動詞を日本語動詞に「直訳」している。

4. おわりに

以上、機械翻訳から見た日本語の特徴を、日英機械翻訳の処理の流れから、また英日翻訳システムのアウトプットを翻訳家による日本語訳と比較しながら、いくつか挙げてみた。本稿で指摘した英日両言語の相違点を反映させることにより、機械翻訳の質がさらに向上することを願っている。

注

1) トランスファー方式の他に、中間言語方式と呼ばれる方

式もある。

2) 本稿の例では英日/日英翻訳ソフト「The 翻訳プロフェッショナル V6.0」(株式会社東芝デジタルメディアネットワーク社)を主に使用した。

参考文献

Ohara, K. H (1999). 'Cognitive and structural constraints on motion descriptions: Observations from Japanese and English.' Proceedings of the 2nd International Conference on Cognitive Science and the 16th Annual Meeting of the Japanese Cognitive Science Society Joint Conference 994-997. Tokyo, Japan.

池原悟(1998). 「機械翻訳」『岩波講座言語の科学』第9巻. 岩波書店.

著者紹介

小原京子：慶応義塾大学理工学部専任講師

「経歴」東京大学教養学部教養学科卒業、同大学新聞研究所修了、日本アイ・ビー・エム株式会社勤務を経て、カリフォルニア大学バークレー校言語学博士 (Ph.D.) 「専門」認知言語学