# 第 6 回　統計的機械翻訳は役に立つか

従来の機械翻訳は文法規則や辞書を作り込んで言語学の手法で訳文を出力するものですが、極めて複雑な「言語」を規則だけでコントロールしようとしても、例外が数多く発生し、なかなかうまくいきません。機械翻訳が実用化されてすでに 20 年以上たっていますが、まだ完成の域に達してないのは皆さんもご存じの通りです。
その問題点を解決しようと、1990 年前後に統計的機械翻訳という新しい手法が提案されました。大量の対訳コーパスから統計的手法を使って自動的に翻訳規則や対訳辞書を「学習」して利用するものですが、これは言語学の知識ではなく共起頻度などを手掛かりにしています。
この統計的機械翻訳をベースとしているのが Google 翻訳ですが、はたしてその訳文は従来の機械翻訳に比べて大きく改善されているのでしょうか。

## ● Google 翻訳の訳文はどうなのか

とりあえず VOA のニュースを訳してみましょう。
【T】は PC-Transer 翻訳スタジオ 2009
【G】は Google 翻訳（http://translate.google.co.jp/translate_t#）
PC-Transer は、スタイルを「一般」にした他はデフォルトの出力です。

### ■【原文】

In an unprecedented response to a domestic terrorist attack, Indian military forces have taken over the battle against gunmen who plunged Mumbai, a city of 18 million, into confusion.

### ■【T】

国内のテロ攻撃への先例のない反応において、インド軍隊は、ムンバイ（1800 万の都市）を混乱に陥れた銃撃者との戦いを引き継いだ。

### ■【G】

国内のテロ攻撃に前例のないレスポンスでは、インド軍は混乱にムンバイ、18 万人の都市で、急落した武装集団との戦いをしている。

読みやすい訳文とは言えませんが、【T】はほぼ正しく訳しています。訳語を修正する必要がありますが構文はそのまま使えそうです。【G】のほうは、「混乱にムンバイ」「急落した武装集団」と、構文を読み違えています。さらに、18 million を「18 万人」と誤訳しています。ただ、「18 万人の都市」と「人」を補ったり、gunmen を「武装集団」と訳しているのがちょっと興味を引きます。

今度は、もっと構文の簡単なものを見てみます。

### ■【原文】

Air force helicopters flew overhead along the shoreline.

### ■【T】

空軍ヘリコプターは、海岸線に沿って頭上を飛んだ。

### ■【G】

海岸線に沿っ空軍のヘリコプターが頭上を飛んでいた。

Google 翻訳の特徴をよく表しているのが「海岸線に沿っ空軍のヘリコプター」の「沿っ」です。フレーズの接続が統語的に不完全です。ただ、【T】が「Air force ＝名詞」「helicopters ＝名詞」と解析して、「空軍ヘリコプター」と訳したところを、【G】では「空軍のヘリコプター」と「の」を入れているのは、おそらくこのフレーズが辞書に「学習」されていたからでしょう。

## ●構文解析かフレーズ訳の自然さか

さて、このまま見て行ってもきりがないのでやめますが、統計的機械翻訳をベースとした Google 翻訳の特徴が少しはお分かりいただけたでしょう。
部分的に大変こなれた良い訳があるかと思えば、文全体ではどうしてこんな訳が出力されたのか説明がつかないようなところが多いという印象です。
統計的機械翻訳は文構造が同じような言語間では有効ですが、英語と日本語のように文構造が大きく違う場合は、当然このような結果になります。そのため、現在では構文情報の利用ができるシステムが研究されています。

今回の少ない例を見ただけでも、PC-Transer 翻訳スタジオ 2009 の構文解析力と訳文生成力がかなり優れていることが分かると思います。ただ、訳語がこなれていないので全体的に「ぎこちない訳文」という印象を与えてしまうのです。このぎごちなさをなくす最も効果的な方法はフレーズ辞書の登録であることはこの連載で何度も書いています。翻訳者が適切に登録したフレーズ辞書は大変効果的ですが、大量に蓄積するには時間がかかります。

統計的機械翻訳では対訳データから自動的にフレーズ辞書を作成します。ですから対訳データが大量にあれば人手で登録するよりも格段に速く辞書の蓄積ができます。ただし、統語的に正しい訳文が出力されることはすぐには期待できません。

結局、翻訳者が利用するには、統語的に適切な訳文を出力することのできる PC-Transer 翻訳スタジオをメインの作業ツールとし、必要に応じて Google 翻訳のフレーズ訳を参考にするのが現在のところ最も効果的だと言えるでしょう。

## ●「Web 検索」機能で効率アップ

作業しやすくするために、翻訳スタジオ 2009 で新たに追加された「Web 検索」に Google 翻訳を設定して、「Google 翻訳」の機能を組み込んでしまいましょう。

＜ Web 検索に Google 翻訳を設定する方法＞

メニューバーの「ツール」→「環境設定」をクリックして「環境設定」画面を表示します。
「Web 検索」タブをクリックし、検索エンジンの「追加」をクリックします。
「検索エンジン追加」画面が開いたら以下のように入力します（英日翻訳）。

タイトル (T):「Google 翻訳 (EJ)」
サイト (S):「http://translate.google.co.jp/translate_t#」
キーワード指定 (K):「http://translate.google.co.jp/translate_t#en|ja|」
キーワードエンコード (N):「UTF-8」

＊日英の場合は、タイトルを「Google 翻訳 (JE)」とし、キーワード指定を「http://translate.google.co.jp/translate_t#ja|en|」とするだけです。

入力したら「OK」ボタンをクリックして設定完了です。

翻訳エディタで作業しているときに Google の訳が見たい場合は、原文を選択して「Web 検索」ボタンをクリックしてください。翻訳エディタ内に Google 翻訳の Web ページが表示されると共に、選択した文章が自動的に「原文」欄に入力され、一息おいて「翻訳」欄に訳文が表示されます。作業中の画面とはタブで切り変えられるので交互に見比べることができます。気になるフレーズだけ Goole 翻訳で訳して参考にするのも良いでしょう。

---
今回は、PC-Transer 翻訳スタジオを基本的な翻訳支援ツールとして中心に据え、足りない機能をその他のソフトと組み合わせて使うという考え方を具体的に説明する形になりました。PC-Transer 翻訳スタジオはこのように柔軟なカスタマイズができるのが大きな特徴です。今後、Web 上のサービスを利用する機会が増えてくると思われますが、「Web 検索」機能は強力なツールとなるはずです。

【eTrans Technology】

---