

翻訳ツールが苦手とする箇所の英日翻訳

山岡幸高

九州大学

y-yamaoka@kyudai.jp

1 はじめに

ニューラル機械翻訳の登場以来, 英日翻訳の精度は飛躍的に向上したが, 致命的な誤訳も僅かにある。その誤訳は学校文法における仮定法, 複合関係詞, 比較, 否定などの特定の箇所に特に多いように思われる。

そこで, 以下のようなツールを考えた。

- 1) 学校文法における主な文法項目を検出・表示し,
- 2) 特定の項目の誤訳の多い箇所だけを翻訳する。

2 実装

筆者は以前から, 既存の英文タグ付けツールの誤りを訂正し, 品詞・文型・含まれている文法項目を表示するというシステムを開発している。

今回そのシステムを用いて文法項目を検出し, 翻訳ツールが苦手とする項目が検出された文に対し, 誤訳の多い箇所のみ並び替えて部分的に翻訳するツールを開発した。

文法項目は学校文法の主なものを網羅しているが, 具体的には下記の Web サイトに記す。部分訳に選んだ項目も記してある。

3 実行結果

ここでは実行結果のごく一部を紹介する。

原文

↓

既存の翻訳サイトでの実行結果

↓

(コメント)

↓

筆者のツールでの実行結果

↓

コメント

の順で記す。

原文

If I had left ten minutes earlier, I would not have missed the train.

↓

翻訳サイト

私が 10 分早く出発した場合、私は列車を逃したことはありませんでした。

↓

筆者のツール

品詞

if(接) i(N) had(v-d) left(V-n) ten(X) minutes(副) earlier(副比), i(N) would(v-d) not(副) have(v) missed(V-n) the(冠) train(N) .

文型 {if {S} {V} {副比}, {S} {V} {O}} .

文法

If S had V-n ..., S would[could] have V-n 「もし S が...したなら、S は...した[できた]だろうに」

和訳

If(接) i(N) left(V)=たなら(v-d) ten(X) minutes(副) earlier(副比), i(N) missed(V)=なかっただろうに(v-d) the(冠)=train(N) .

↓

筆者のツールでは, 構文解析し仮定法過去完了を検出し, 動詞句の語順を替え, 動詞の語尾のみ訳した。

原文

However tired she is, she always smiles.

↓

翻訳サイト

しかし、彼女は疲れている、彼女はいつも笑顔。

↓

この **However** は複合関係副詞だが、副詞であるかのように訳してしまっている。

↓

筆者のツール

品詞

however(関副) tired(形) she(N) be(V) , she(N) always(副) smile(V) .

文型

{however {C} {S} {Vi}} , {S} {V} .

文法

however 形容詞[副詞] S V {副詞節を導く複合関係副詞}

和訳

どんなに(関副) tired(形) she(N) be(V)=でも , she(N) always(副) smile(V) .

↓

筆者のツールでは構文解析した上で **However** のみ訳してある。

原文

She doesn't play tennis as well as her sister.

↓

翻訳サイト

彼女はテニスだけでなく、妹も演じていません。

↓

not as ... as を「ほど…ではない」と訳してほしいところだが、熟語の **as well as** の訳のようになっている。

↓

筆者のツール

品詞

she(N) do(v) not(副) play(V) tennis(N) as(副) well(副) as(接) her sister(N) .

文型

{S} {V} {O} {副} {as/ {N}} .

文法

A ... not as[so] 原級 as B 「ほど形ではない」

和訳

she(N) play(V)=do(v)=not(副) tennis(N) well(副) ほど(接) her(所) sister(N) .

↓

筆者のツールでは、構文解析し原級比較を検出し、前の **as** を消し後ろの **as** を「ほど」と訳し、動詞句の語順を替えた。

原文

His story was anything but boring.

↓

翻訳サイト

彼の話は退屈だった。

↓

anything but の訳抜け？

意味が逆になってしまっている。

↓

筆者のツール

品詞

his story(N) be-d(V-d) anything(N) but(前)

boring(形) .

文型

{S} {V} {C} {前} {形} .

文法

anything but

文法

be 名詞[形容詞] {V C}

和訳

his(所) story(N) be-d(V-d) 決して boring(形)でない(前) .

↓

筆者のツールでは、構文解析し anything but を検出し、その周辺のみ並び替え「決して…でない」と訳した。

4 精度

訓練データに

『Vision Quest 総合英語』

『デュアルスコープ総合英語』

『ブレイクスルー総合英語』

『Next Stage 英文法・語法問題』

『英文法・語法 Vintage』

『スクランブル英文法・語法』

等の参考書・問題集を用いた後、高校教科書

『CROWN English Communication I』

『CROWN English Communication II』

の英文の難しい部分(教科書ガイドで解説されている文)に対してテストを行った。

文法項目の検出精度は、全体では 8 割以上、部分訳に選んだ項目に限っては 9 割以上という結果であった。

部分訳は検出に基づいて行われるため、両者の精度は同じである。

5 公開

今回開発したツールは以下の Web サイトで公開しているため、是非使っていただきたい。

<http://yamaoka.rosx.net/yamaoka-nltk.html>

参考文献

- [1] Natural Language Toolkit
<https://pypi.python.org/pypi/nltk/3.1>
- [2] Numerical Python
<https://sourceforge.net/projects/numpy/files/NumPy/>
- [3] <http://yamaoka.rosx.net/yamaoka-nltk.html>
- [4] 瓦 祐希, Chenhui Chu, 荒瀬 由紀: Recursive Neural Network を用いた事前並び替えによる 英日機械翻訳, 言語処理学会, 第 24 回年次大会, 発表論文集 (2018)
- [5] 鶴川 新, 田村 晃裕, 二宮 崇, 高村 大也, 奥村 学: 固有表現情報を用いたニューラル機械翻訳, 言語処理学会, 第 24 回年次大会, 発表論文集 (2018)
- [6] 西村 友樹, 秋葉 友良, 塚田 元: 大規模単言語資源を用いた大語彙ニューラル機械翻訳の検討, 言語処理学会, 第 24 回年次大会, 発表論文集 (2018)
- [7] 山岸駿秀, 叶内晨, 佐藤貴之, 小町守: ニューラル日英翻訳における出力文の態制御, 言語処理学会,

第 23 回年次大会, 発表論文集 (2017)

- [8] 藤原菜々美, 今出昌宏, 山内真樹, 内山将夫, 隅田英一郎: 自動コーパス生成とユーザフィードバックによる機械翻訳, 言語処理学会, 第 23 回年次大会, 発表論文集 (2017)
- [9] 竹野峻輔, 永田昌明, 山本和英: ニューラル日英機械翻訳における欠落語に関する分析, 言語処理学会, 第 23 回年次大会, 発表論文集 (2017)
- [10] 高橋寛治, 竹野峻輔, 山本和英: 機械翻訳システムの安定性評価, 言語処理学会, 第 23 回年次大会, 発表論文集 (2017)
- [11] 田中省作, 小林雄一郎, 徳見道夫, 後藤一章, 富浦洋一, 柴田雅博: 学校英文法の学参例文データベースとその応用, 情報処理学会研究報告, 第 2012-CH-93 巻第 5 号:1-8 (2012).
- [12] 石黒昭博: 総合英語 Forest 7th Edition, 桐原書店 (2013)
- [13] クラウン コミュニケーション英語 I 予習と演習, 三省堂
- [14] クラウン コミュニケーション英語 II 予習と演習, 三省堂
- [15] クラウン コミュニケーション英語 III 予習と演習, 三省堂
- [16] Sebastian Raschka: [第 2 版]Python 機械学習プログラミング, インプレス (2018)
- [17] 奥村学: 構文解析, コロナ社 (2017)
- [18] 奥村学: 言語処理のための機械学習入門, コロナ社 (2010)
- [19] 野村恵造: Vision Quest 総合英語, 新興出版社啓林館 (2013)
- [20] 小寺茂明: デュアルスコープ総合英語 四訂版, 数研出版 (2011)
- [21] 吉波和彦, 北村博一: ブレイクスルー総合英語 改訂二版, 美誠社 (2011)