

## 言語学的視点からの和英翻訳エンジンの評価と改良のための提案

加藤 鉦三 (信州大学)

kinoene@shinshu-u.ac.jp

### 【1】動機

本研究の動機は次のものである。

- (1) 社会的に意味のある研究がしたい
- (2) 機械翻訳, 特に和英翻訳には改善の余地が大きい

自分の研究に関する限り, これまでは, 「言語研究を通して学生の思考力を高める」授業をする基盤となる, という間接的なものしか社会的意味はなかったと言いきってよいかもしれない。しかし, もし自分の言語研究が機械翻訳の精度向上に少しでも寄与できるとしたら, それはそれなりの社会的意味を持つことになる。

### 【2】素材

本研究では, 「ところ」を含む文の, 次の3つの機械翻訳エンジンの出力を素材とする。

- (3) Yahoo! 翻訳 (<http://honyaku.yahoo.co.jp>)
- (4) Livedoor 翻訳 (<http://translate.livedoor.com>)
- (5) Google 翻訳 ([http://translate.google.com/translate\\_t?hl=ja](http://translate.google.com/translate_t?hl=ja))

Google 翻訳の特徴については, 同サイト「よくある質問」で次のように解説されている。

#### 統計的機械翻訳とは何ですか？

今日市場に出回っている自動翻訳システムのほとんどは, 規則ベースで開発されており, 言語研究者による語彙や文法の定義など多くの作業を必要とします。

Google の翻訳システムの手法は異なります。ターゲットとなる言語で記述された単一言語のテキストと, 人間が翻訳した他言語のサンプル翻訳テキストを対にしたものを大量にコンピュータに入れます。そしてこれらのテキストに統計的学習手法を適用して, 翻訳モデルを構築しています。Google のリサーチ評価では, この手法が優れた結果をもたらすことが判明しています。

### 【3】「ところ」の諸用法

本研究では, 「ところ」の用法を暫定的に次のように分類する。

#### (6) 「ところ」の用法

- |         |                          |
|---------|--------------------------|
| <場所>    | 私が住んでいるところ               |
| <代名詞>   | 私が予約したところ (= 「ホテル」等)     |
| <点>     | 彼のそういうところが嫌いだ            |
| <アスペクト> | 投げるところ, 投げたところ, 投げているところ |
| <事態認知>  | 私が見るところ (では), それは間違っている  |

<場所>は, 関係節内の動詞が「いる, ある, 住む, 立つ」のような静止位置や, 「行く, 帰る, 落ちる」のような移動後の位置を要求するものである時に現れる。「かゆいところない? / かゆいところに手が届かない」のような事例も<場所>である。<場所>は名詞としての用法である。

<代名詞>は, 場所そのものではなく, 「私が勤めているところ / いつも昼飯を食べるところ」のように, 店や機関等を指すのに本来は<場所>である「ところ」を代名詞的に使う用法である。<代名詞>は名詞としての用法である。

<点>は, 「彼の偉さは動じないところにある」のように, 「点」で言い換えられる事例である。<点>は名詞としての用法である。

<アスペクト>は, 助動詞用法である。前接する動詞が終止形ならば直近の未来を, タ形ならば直後を, テイル形ならば動作の途中を, というように, 「ところ」自体がアスペクト的意味を持つ

のではなく、動詞のアスペクトをそのまま強調するという機能を持つ。

<事態認知>は、「それで当局が知るところとなった」「それは私の関知するところではない」のように、関係節内の動詞が「知る、聞く、理解する」のように、判断を伴わない認知活動を表す場合に限られるようである。この用法の「ところ」自体の意味はよく分らない。<事態認知>は名詞としての用法である。

以下、4節ではこれらの用法を翻訳エンジンがどう訳すかを概観し、5節で評価基準を提案した後、6節で改善のための提案を行う。

#### 【4】出力の実際

##### (7) <場所>

[原文 A] 昨日行ったところは寒かった。

Y : The place that I went was cold yesterday.

L : It was cold in the place yesterday.

G : Just yesterday was cold.

[原文 B] 私は住むところを探しています。

Y : I look for a place to live in.

L : I am looking for the place in which it lives.

G : I am looking for homeless.

<場所>の「ところ」は place で訳してよい。実際、place で訳してよいのは、この<場所>だけである。Y と L は、非常に多くの場合、「ところ」を place で訳出する。Y は、主語がない場合、I を補う傾向があるが、それはおそらく統計的に見て正解だろう。しかし、that ではなく where であるべきである。Y と違って、L は表面上主語がない時 it を補う傾向があり、それが[原文 B]で大きな傷となってしまっている。G はいずれも訳になっていない。

##### (8) <代名詞>

[原文 C] 彼が勤めているところは倒産した。

Y : The place where he worked at went bankrupt.

L : The place where he is working went bankrupt.

G : Where he was working for went bankrupt.

[原文 D] 予約したところはカードが使えないようだ。

Y : In the place that made reservations, a card does not seem to be usable.

L : It seems that the reserved place cannot use a card.

G : There is a reservation card is not available.

[原文 C]では、いずれも正解から遠くはないが、place は誤りであり、company 等を「勤めている」から補う仕組みが必要である。[原文 D]は、L と G は正解からとても遠い。ここでは、本当の正解を出すためには、(i)「カードが使える」からレストランやホテルを候補にし、(ii)この文が含まれる段落の文脈からそれを絞り込む、というかなり複雑な仕組みが必要となる。

##### (9) <点>

[原文 E] 彼のそういうところが嫌いです。

Y : I dislike his such place.

L : He dislikes such a place of his.

G : That's just not like him.

[原文 F] 子供のいいところを見つけてあげましょう。

Y : I will find the good place of the child.

L : Let's find a place with a sufficient child.

G : Children good place to find them.

G はいずれも英語になっていない。[原文 E]では、Y と L は place ではなく character を引き出す仕組みを組み入れれば正解になる。[原文 F]では、Y は place ではなく points を引き出す仕組みを組み入れなければならないが、更に定・不定と数の問題もある。これについては最後に触れる。L は不可解で、G は問題外である。

#### (10) <アスペクト>

[原文 G] 私はこれから出かけるところです。 [原文 H] これから出かけるところです。

Y : I am going to go out from now on.

Y : I am going to go out from now on.

L : I am just going to go out after this.

L : From now on, he will just be going to go out.

G : I am now going.

G : Go now.

[原文 I] 私はさっき食事したところです。

Y : I ate a while ago.

L : I just had a meal a while ago.

G : I was eating a little while ago.

[原文 J] 私は今それを読んでいるところです。

[原文 K] 私は今それを読むところです。

Y : I am reading it now.

Y : I am going to read it now.

L : I am just going to read it now.

L : I am just going to read it now.

G : I have read it now.

G : So I read it.

<アスペクト>では、いずれも(さすがに) place は用いていない。[原文 G]を除いて、G はその他全てを誤訳している。Y はアスペクトの違いを正確に訳しており、とても優秀である。L は、「タ形かそれ以外」という区別の仕方をしており、終止形[原文 J]とテイル形[原文 K]の差別化ができていない。

#### (11) <事態認知>

[原文 L] 私が見るところではそれは間違っている。

Y : I am going to look, and it is wrong.

L : It is wrong in the place which I look at.

G : I can see where it is wrong.

[原文 M] 機械翻訳について知るところを述べよ。

Y : Speak a place knowing machine translation.

L : Describe the place to get to know about machine translation.

G : 述べよ to know about machine translation.

[原文 L]では、Y は<アスペクト>の「ところ」と誤認しており、G では where の位置が主従逆になっている。L は誤った直訳になっている。[原文 M]はいずれも問題外である。

### 【5】翻訳エンジンの評価

前節で見たような出力から、翻訳エンジンの実力がある程度判定できる。<アスペクト>を評価基準とすれば、Y が最も正確に訳している。それでも、その<アスペクト>に関しても、[原文 L]のように過剰般化が見られる。<アスペクト>以外では、「ところ」の訳出だけに注目すれば、Y と L には大して差はない。それでも全体評価としては、Y>L となるだろう。そのような印象は、次のように分析できるのではないだろうか。

(12) 可能な判定基準

- ・ 統語を（大きくは）外さない
- ・ （表面に表れていない）主語の認定がある程度確かである [原文 A, D, E, H]

2節で、Google 翻訳は自分の統計的翻訳手法が優れていると述べているのを見たが、4節のデータを見る限り、Google 翻訳は実際には Y や L に比べてかなり劣っていると判定される。

【6】提案

本節では、4節で見た問題点を解決するための提案をする。

まず、①<場所>以外では place を使わないようにする必要がある。

次に、<代名詞>では、(8)の解説ですでに提案したが、それは現段階ではおそらく要求しすぎであろう。しかし他に有効な手があるようには思えない。<代名詞>はいきなり用いられることはなく、文脈依存性が高いため、当面は、②代名詞的な「ところ」があることを認識し、文脈からそれが指すものを探してくる仕組みを志向すべきだろう。

<点>では、③述部が「好き、嫌い、すばらしい」のような<評価>系である場合には、人ならば character で、モノならば aspect で、④「ところ」の前接修飾部が「いい、悪い、すばらしい、偉い」のような<評価>系である場合には point を訳語にするような仕組みである程度対処できるだろう。

<アスペクト>では、3つのアスペクト以外に、「住むところを探す」のような事例も対処できるように、次の規則が有効であろう。

⑤V=終止形で、トコロ句が主語・目的語・補語ならば ⇒ a place to V

⑥V=終止形で、トコロ句が主語・目的語・補語ではなく、かつ副詞成分でもないならば（つまり助動詞用法ならば）⇒ S is going to V

⑦V=テイル形で、トコロ句が主語・目的語・補語ではなく、かつ副詞成分でもないならば（つまり助動詞用法ならば）⇒ S is Ving

⑧V=タ形で、トコロ句が主語・目的語・補語ではなく、かつ副詞成分でもないならば（つまり助動詞用法ならば）⇒ S has just Ven

<事態認知>では、⑨「私が見るところ（では）」のような副詞用法の場合には、副詞節構造ではなく、I see ...のように複文構造にし、⑩「知るところを述べよ」のように副詞用法でない場合には、what you know のように what を使った関係節にするような仕組みで対処可能であろう。

最後に、(7)の[原文 B]のような事例が示唆するところを見ておきたい。Y は a place で、L は the place で訳出している。Google 検索では次の結果が出て、決め手にはならない。

“I am looking for the place” に一致する英語のページ 約 244,000 件

“I am looking for a place” に一致する英語のページ 約 307,000 件

しかし to live を追加すると、a place でなければならないことが分かる。

“I am looking for a place to live” に一致する英語のページ 約 58,700 件

“I am looking for the place to live” に一致する英語のページ 1 件

機械翻訳の精度向上のためには、意味論的アプローチよりは統計処理の方が容易であろう。この事例は、統計を訳文のフィルターとして使うことが有効であることを示しているように思われる。少なくとも、名詞の選択と定・不定、それに数については、訳の候補から統計処理で出力を選ぶという方向が考えられるのではないだろうか。

《謝辞》

機械翻訳に関する基礎的な知識と研究対象としての可能性の示唆を黒田航氏（けいはんな情報通信融合研究センター）より受けた。この場を借りてお礼申し上げたい。