

機械翻訳を利用したぎこちない翻訳文の検出への試み

阿辺川 武

影浦 峽

東京大学大学院 教育学研究科

{abekawa,kyo}@p.u-tokyo.ac.jp

1 はじめに

経験のある翻訳者による翻訳と初心者による翻訳では、たとえ伝わる情報が同一であるとしても、文の読み易さに関しては格段に前者が勝る。我々はそのひとつの原因として、翻訳初心者は原文の意味を再現しようとするあまり、原文を直訳してしまう傾向があると考えている。英日のように構文構造が大きく異なる言語間の翻訳では、原言語の直訳は対象言語においてぎこちない文となることが多い。他方で、機械翻訳の世界でも、現状は読み易さよりも原文を正しく翻訳することを重視しており、翻訳初心者と同様の傾向が診てとれる。

そこで我々は、翻訳初心者による翻訳文の中から、ぎこちない文を検出するために機械翻訳の翻訳結果を指標のひとつとして利用することを考えた。本稿では、ベテラン翻訳者、初心者翻訳者、機械翻訳それぞれによる翻訳文を比較分析した結果と、ぎこちない翻訳文の検出への試みについて報告する。

2 データ

まず最初に本研究で用いているデータについて説明する。翻訳データは、石油の枯渇に関する英語で執筆された一般向け書籍¹と、それに対する日本語訳²である。現在手元にある翻訳データは、翻訳経験の浅い3人(2~5年)が分担して下訳を担当し、それに対して経験12年の翻訳者が1人で修正を施したものである。データの基本的な統計データを表1に掲載する。なお、簡単なDPマッチングによる下訳と修正訳の形態素単位の一致率は約78.9%であった³。文数がそれぞれ異なることから、文の対応は、1対2や2対3といったものがいくつも存在することがわかる。

3 ぎこちない翻訳と機械翻訳

ぎこちなさのひとつとして元の原文の構造が垣間みられる文がある。たとえば原文では受動態で表現されて

¹Jeremy Leggett, "Half Gone: oil, gas, hot air, and the global energy crisis", Portobello Books Ltd, 2005

²ジェレミー・レゲット (原著). 益岡 賢, 楠田泰子, 植田 那美, リックタ ナカ (訳). 「ピーク・オイル・パニック-迫る石油危機と代替エネルギーの可能性」. 作品社. 2006.

³本研究では、形態素解析器に Mecab を使用した。
<http://mecab.sourceforge.net/>

表 1: データの統計

	英文	下訳	修正訳
文数	4,414	4,628	4,648
単語 or 形態素数	81,207	127,721	132,823
(一文あたり平均)	18.4	27.6	28.6

いるが、そのまま下訳でも受動態で表現されている場合である。一方で、それに対する修正訳では能動態に修正されており、原文とは構造も表現も大きく変わっていることがある。ただ、必ずしもすべての原文の受動態が翻訳で能動態へと変換されるわけではなく、その動詞の性質、文中の他の構成語、前後の文など総合的に考慮した上で訳文が作成される。

また、別のぎこちなさとして、適切な訳語が正しく選択されていないことがある。たしかにその訳語でも意味は通るが、文脈を考慮した場合は、たとえ辞書に掲載されていない表現でも、その表現の方が文章の流れとしてよい場合がある。また文章全体の雰囲気として堅い表現や柔らかい表現を使いわけるといったことも必要とされる。

機械翻訳の世界では、近年、SMTの手法が盛んに研究されているが、市販の機械翻訳ソフトでは、精度の観点や構築された規則の豊富さからルールベースの翻訳手法を採用している会社が多い。ルールベースの手法では、決まりきった訳出が多く、臨機応変な訳語の選択をすることは難しい。したがって個々の訳語は意味として正しくても、単語の接続として考えると不適切なことが多々存在する。

以下では、人間訳と機械翻訳機による訳文について比較を行なう。機械翻訳の出力は、市販の機械翻訳ソフトを利用した。翻訳出力の例を表2に掲載する。

3.1 言語モデル

まず、語の接続の観点から訳文の質を眺める。表3は、2節で説明したデータに対して、n-gramモデル⁴を用いて計算した各訳文のパープレキシティであり、機械翻訳機により生成された訳文(以降、MT訳)は、人間訳に比べて格段に質が悪いのが現状である。n-gram

⁴コーパスは Web から抽出した日本語文約 4,000 万文を使用し、n グラムの接続数は n=4 とした。

表 2: 翻訳の例

原文:	We can think of these as the “existing-reserves” question, the “reserves-additions” question, and the “speed-to-market” question.
MT 訳:	「既存の蓄え」質問、「蓄え追加」質問、及び「売り出す速度」が質問されるとき、これらを考えることができる。
下訳:	それは「残存埋蔵量」問題および「累積埋蔵量」問題、「スピード・トゥー・マーケット」問題として捉えることができる。
修正訳:	それは「現存する埋蔵量」の問題、「埋蔵量の増加」の問題、「市場に供給されるスピード」の問題である。

モデルでは正しく原文の意味を翻訳しているかについては知る由もないが、MT 訳では、語の接続関係が正しく機能しておらず、訳語の選択がうまくいっていないことがわかる。

表 3: 各訳文のパーブレキシティ

	全文	翻訳者 A	翻訳者 B	翻訳者 C
担当文数	3,948	1,676	768	1,504
下訳	65.8	66.8	57.3	70.5
修正訳	58.7	58.5	52.7	62.9
MT 訳	119.2	115.2	116.8	124.9

3.2 品詞の増減

日本語と英語のような言語構造が大きく異なる言語間の翻訳では、対応する単語が同一の意味を表わしていても、品詞が異なる場合がある。例えば原文では名詞を修飾する形容詞であったものが、対象文では動詞を修飾する副詞として翻訳される例はよくある。そこで各翻訳間で品詞の増減が 1 文あたりどのくらいあるかを統計的に調査した結果を結果を表 4 に示す。

表 4: 品詞の増減 (値は絶対値)

	原文 vs 下訳		原文 vs MT 訳		下訳 vs MT 訳	
	平均	分散	平均	分散	平均	分散
形容詞	1.14	1.48	1.04	1.32	0.52	0.47
動詞	1.03	0.98	0.89	0.76	0.83	0.84
副詞	-	-	-	-	0.59	0.55
名詞	1.59	1.79	1.47	1.35	1.27	1.72
総計	3.76	5.36	3.40	3.83	3.21	5.19

「原文 vs 下訳」と「原文 vs MT 訳」を比較すると品詞の増減の平均は MT 訳よりも下訳の方が高く、分散をみても下訳の方が大きく、文によって増減の差がはげしいことがわかる。このことから、MT 訳の方が原文を忠実に訳していることがいえる。また、どの翻訳対でも名詞の増減が大きいことから、名詞は翻訳の仕方により多様な表現方法があることがわかる。

4 ぎこちない翻訳文の検出

下訳のぎこちなさを検出するため、機械翻訳の出力結果をひとつの指標として利用することを試みる。

4.1 実験データ

今回の実験で使用するデータは、原文と下訳を人手で文単位の対応を取り、その中で原文と下訳と修正訳の 3 文対がすべて 1 対 1 で対応している文に限定して、文の結合や分割は考えないこととした。さらに章見出しなど文末に句点、疑問符がない文を除外した。その結果、下訳総数 4,628 文の中から 3,826 文が実験の対象となった。

ぎこちなさの評価は、下訳と修正訳を比較してどのくらいの割合が修正がなされているかで行う。参考文献 [1] で分析されているように修正されている要因は様々で、すべてが訳文のぎこちなさからくるものではない(誤訳や編集の方針など)が、本実験では、大きく修正されているほど下訳がぎこちないとみなす。

下訳と修正訳の比較は n-gram ベースの BLEU や NIST といった評価手法ではなく、TER (Translation Error Rate) [3] を用いる。TER は、通常の DP マッチングの比較で対象とする操作(挿入・削除・置換)に加え、フレーズのシフトも操作のひとつとしてカウントし、最も操作数の少ない対応を求め、総操作数を対象文長で正規化した値として算出する。図 1 に、我々のデータにおける TER の分布を示す。TER は、対象文が元の原文よりも長い場合、100(1.0)を越えることがあるので、そのときは 100 として扱った。

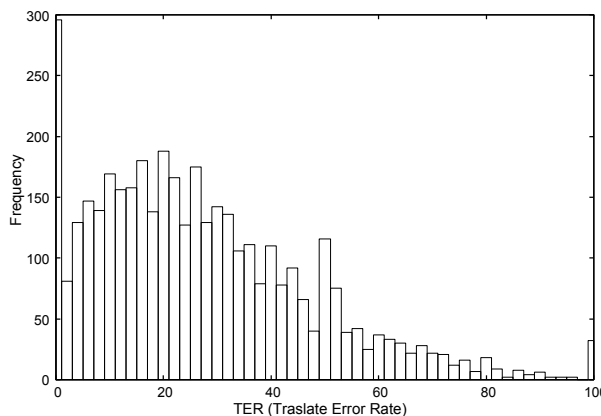


図 1: 下訳と修正分間の TER の分布
修正がまったくない TER 値=0 が最頻値であるが、

その後は単純な右下りではなく、TER 値=20 付近でピークとなるような分布となっている。これは1箇所修正されるとその周辺も同時に修正されることもひとつの理由である。

4.2 手法

今回は、データセットを訓練データ、テストデータに分割し、機械学習アルゴリズムを利用してぎこちなさをモデル化する手法をとる。下訳1文と対応する原文、MT 訳を入力とし、TER を算出し、実際の TER 値との相関を求めることで評価を行なう。そのため学習アルゴリズムには出力を連続値として扱うことのできる Support Vector Regression (SVR) を用いる。

素性については、binary 素性と数値素性を用い、数値素性については z-score⁵ を利用して正規化を行なった。カーネルには線形関数を使用した。評価尺度は、テストセットの出力値が TER と比べてどのくらい相関しているかをピアソンの積率相関係数で測る。

4.3 素性

機械学習で使用する素性は次のものを設定した。

(a) 下訳と MT 訳の対応

下訳中の語で、MT 訳で使用されている単語と近い語を直訳に近いと仮定する。最初に下訳と MT 訳との間で DP マッチングをとり対応する語のペア、下訳の語、MT 訳の語をそれぞれ素性として利用する。このとき対応しているとみなす語は、表記が完全一致した場合だけでなく、品詞が同じ場合 (ex. 「単純」と「簡単」) や1文字目が同一の文字の場合 (ex. 「濃厚」と「濃かつ」も含める。

(b) 原文と下訳の対応

最初に GIZA++[4] により原文と下訳の語対応を求める。本データのみでアライメントを求めるには文の量が足りないため、同時に内山らの対訳コーパス [5] を使用した。今回は、原文 → 下訳、下訳 → 原文の両方向で別々に求めてたものを使用する。1対多対応があるときは、そのすべてを利用すると誤りが含まれる可能性が高いため、GIZA++ が出力するそれぞれ対の翻訳確率を利用し、確率が 0.2 以上の対応を利用した。また対応がない語についても、重要な指標のひとつと考え、対応先の単語を NULL として素性として用いる。

(c) 下訳

下訳の語の unigram と bigram を素性として利用する。データを観察したところ、特定の品詞が多く含まれると修正されやすい傾向にあることから、

⁵z スコアは、観測値と平均値との間が標準偏差何個分があるかを示す指標 $zscore(x) = \frac{x - avg(X)}{\sqrt{var(X)}}$

表 5: 各数値素性の TER 値との相関係数

助詞, 格助詞, 連語	0.123	名詞, 形容動詞語幹	0.090
名詞, 一般	-0.103	助動詞	0.090
記号, 読点	-0.099	名詞, 代名詞, 一般	0.087
記号, ?	0.094	名詞, 非自立, 一般	0.078
記号, 一	0.091	名詞, 接尾, 助数詞	-0.084

文中に含まれる品詞をカウントし、文長で除した割合を素性とする。表 5 に、素性として使用した品詞の1文中の出現割合と TER 値の相関係数を載せる。「によって」「に関する」といった「助詞, 格助詞, 連語」が相関係数が高い。また「記号, 読点」と TER 値との相関係数が負の値になっているのは、1文中の読点の数が少ないと、TER 値が上がるということを示している。

使用しなかった素性

次の素性は検討したが予備実験で精度が向上しないため使用しなかった。

- 品詞の増減

それぞれの翻訳対での増減を素性として使用した。向上しなかった理由として、英語では Penn Treebank の品詞体系を、日本語では IPA の品詞体系を用いたが、両体系間での単純な対応では、期待する品詞の増減が計算できないことが挙げられる。

- 下訳文のパープレキシティ

下訳1文の言語モデルの値を利用したが、1文だけでは大小様々な値をとり、ぎこちなさの指標としては、意味を持たなかった。

4.4 結果

最初に、学習器の予測能力を調べるために10分割交差検定を行なった。全部の素性を使用して計算した場合の相関係数、およびそれぞれの素性群の影響度を見るために、素性群を除いた場合の相関係数を表 6 に示す。

表 6: 実験結果

TER との相関係数	
全部の素性	0.355
以下の素性を除いたとき	
(a)MT 訳対応	0.331
(b) 原文と下訳対応	0.349
(c) 下訳	0.344

次に実際の運用を想定をした場合の評価を行う。表 6 の結果は、ランダムに並べ換えたデータを10分割したものだが、実際には最初からすべての修正訳が入るわけではない。本データは3人の翻訳者から下訳が作成されているので、おのおの2人の下訳が先に完

成し、それに対する修正されたデータがあるという状況のもと、残りの1人の下訳の検出精度がどうなるかという実験である。つまり訓練データが2人分のデータ、テストデータが1人分のデータである。表7が3通りパターンで評価した結果である。

表7: 下訳翻訳者別の評価

訓練データ	テストデータ	TER との相関係数
翻訳者 B,C	翻訳者 A	0.220
翻訳者 A,C	翻訳者 B	0.283
翻訳者 A,B	翻訳者 C	0.180

4.5 考察

機械学習による出力結果は、実際の TER との相関係数では 0.355 と弱い相関が見られたが、実際の使用をめざすには、まだまだ低い値である。さらに実際の運用を模して行なった結果では、3分割により訓練事例数が少なくなった影響を汲みしても⁶、翻訳者が異なることの要因が大幅な精度低下を招いている。ここでは、低精度となった原因として考えられる3つの理由を挙げる。

MT 訳の利用方法

表6の結果より、MT 訳と関連する素性を使用しないときでも、それほど相関係数が低下していない。MT 訳の直訳性を利用することが目的であったが、この素性の影響は限定的であるといえる。翻訳の直訳性には大きく2つの指標があり、1つは今村ら [2] の提唱するような単語の直訳性であり、もう1つは構文の直訳性である。本手法では、広い意味で MT 訳の単語の直訳性を利用したが、構文の直訳性も有効に利用していければ、MT 訳の利用が精度の向上に寄与できると考えている。

誤訳

下訳において誤訳とぎこちなさの区別がつかないことが原因のひとつである。ぎこちなさは対象言語の言語モデルとの差から識別することができるが、誤訳は原文との意味の乖離であって、たとえ対象言語の言語モデルの数値が高くても修正は大規模となる。今回は原文と下訳の意味の対応までは考慮していないので、誤訳かどうかは判別できない。

ただし、誤訳は、完全な勘違いからくることもあるが、翻訳者自身が原文の意味を把握できていないことから発生する。そのような場合、意識的にしろ無意識的にしろ、原文からの意味の乖離が少なくなるように直訳調になることが多いというのが翻訳データを観察した実感である。

⁶通常の3分割交差検定では 0.334 であった

検出の単位

本実験では、どの程度修正されているかという TER 値を評価の指標として文単位で用いた。しかし、ぎこちない翻訳というのは多くの場合、文全体がぎこちないのではなく、ある範囲(最小は単語1語)がぎこちないのであり、TER 値でまとめてしまうのには多少無理がある。したがって、別の手法として、文中のある範囲に限定し、その範囲が修正されるのか否かを判定していく手法が考えられる。そのためには、原文と下訳で単語単位の正確なアライメントが必要であるが、上手な翻訳になればなるほど原文とは離れた構造・表現になり、正しいアライメントが取得できないという問題が発生する。

5 おわりに

本稿では、英日翻訳支援システムの開発の一環として、ぎこちない翻訳を検出する試みを紹介した。ぎこちない翻訳では原文をそのまま翻訳する「直訳」を施す傾向が高いこと、同様に機械翻訳においても「直訳」に近い傾向があることから、ぎこちない翻訳の検出に際して、機械翻訳機の翻訳結果と一致するかどうかという指標を提案した。実際にこの指標を使用することで、ぎこちない翻訳の検出精度が向上することを示した。

今後の課題として、ぎこちない翻訳と MT 訳との関連についてさらなる調査を行った上で構文の直訳性を利用することや、誤訳と上手な翻訳との違いを検出することなどを考えていきたい。

付記

本研究の一部は、日本学術振興会科学研究費補助金基盤(A)「翻訳者を支援するオンライン多言語レファレンス・ツールの構築」(課題番号 17200018)、および財団法人博報児童教育振興会の支援を得て行われた。

参考文献

- [1] 阿辺川武, 影浦峽. 下訳から修正訳への訳文修正要因の分析. 言語処理学会 第 14 回年次大会 発表論文集, 2008. (発表予定).
- [2] 今村賢治, 隅田英一郎, 松本祐治. 直訳性を利用した機械翻訳知識の自動構築. 自然言語処理, Vol. 11, No. 2, pp. 85-99, 2004.
- [3] Matthew Snover, Bonnie Dorr, Richard Schwartz, Linnea Micciulla, and John Makhoul. A study of translation edit rate with targeted human annotation. In *Proceedings of Association for Machine Translation in the Americas*, pp. 223-231, 2006.
- [4] Franz Josef Och and Hermann Ney. A systematic comparison of various statistical alignment models. *Computational Linguistics*, Vol. 29, No. 1, pp. 19-51, 2003.
- [5] 内山将夫, 高橋真弓. 日英対訳文対応付けデータ, 2007. <http://www2.nict.go.jp/x/x161/members/mutiyama/align/index.html>.