

再翻訳を用いた複数の機械翻訳候補からの出力選択

山梨英和大学人間文化学部人間文化学科教授 **江原 暉将**

PROFILE

1967年早稲田大学理工学部卒。同年NHK入局。2003年、諏訪東京理科大学教授。2009年より現職。アジア太平洋機械翻訳協会(AAMT) / Japio 特許翻訳研究会副委員長。

1 はじめに

筆者らは規則方式機械翻訳(RBMT: Rule Based Machine Translation)に統計的後編集(SPE: Statistical Post Editing)を組み合わせたハイブリッド翻訳方式を研究している[1, 2]。この方式は、人手によるきめ細かな翻訳規則を備えた規則方式の利点と大量の対訳コーパスから翻訳規則を自動獲得する統計方式の利点を兼ね備えており、RBMT単独のシステムを上回る精度が期待できる。

この方式を用いて、ワークショップ型共同研究であるNTCIR-9の特許翻訳タスク(PatentMT)[3]に参加した。このタスクは特許文書を対象にして、機械翻訳の精度を比較するものである。翻訳方向としては、日英、英日、中英の3方向のサブタスクが設定されているが、ここでは日英サブタスクについて記述する。日英サブタスクには12チームが参加している。各チームの翻訳結果に加えて、主催者が3種類のRBMTシステムの結果を参考として提示している。

NTCIR-9での精度評価の規準としては、自動評価基準と人手評価基準が用いられている。筆者らのシステム(EIWA)[4]は、日英方向において、自動評価基準では、参加チーム中最高レベルの評価を得たが¹、人手評価基準では主催者が提示した規則方式単独の機械翻訳システ

ムの一つであるRBMT1よりも低いレベルであった²。せっかくのSPE機能が有効に働いていないことがわかる³。しかし両者の翻訳結果を詳細に比較すると、SPEによって訳質が改善されている例も見られる。そこで、RBMT出力とSPE出力の良い方を選択して出力することが出来れば、全体としての翻訳品質を向上させることができる。そのためには、RBMT部の出力とSPE部の出力を比較して良い方を自動的に選択する手段が必要となる。

2 機械翻訳出力選択方法

複数の機械翻訳システムの出力を比較して良い訳文を選択する手法としてはいくつかの方法が提案されている[6, 7, 8]。ここでは、機械翻訳結果を原言語側に再翻訳して原文と比較する方法[9]を利用する。

この方法の処理手順を図1に示す。原文をn種類の機械翻訳システムで翻訳し、その結果の訳文をn個得る。今回の場合は、RBMTとSPEの2種類であるからn=2となる。次に訳文を原言語側に再翻訳する。再翻訳には、やはり機械翻訳システムを用いる。再翻訳で得られたn個の訳文(再訳文)を原文と比較し、原文に近い再訳文を選び、その再訳文の元となった訳文を最適訳文として選択して出力する。

1 自動評価規準のBLEUとRIBESでは1位、NISTでは4位であった。

2 RBMT1の翻訳結果は筆者らの規則方式翻訳部(RBMT)の出力と酷似しており、類似のシステムと推定できる。

3 同様の結果は[5]でも示されている。ちなみに人手評価の1位はJAPIOチームであった。

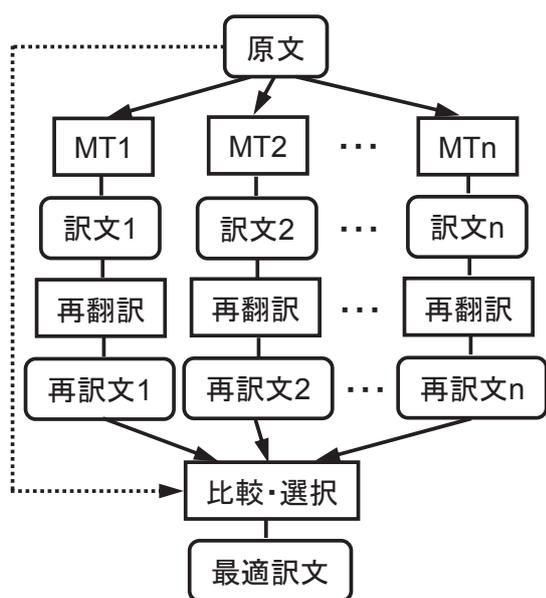


図1 再翻訳を利用した最適訳文の選択

再訳文と原文の比較には機械翻訳の自動評価規準の一つである IMPACT を用いた [10]。IMPACT は、機械翻訳の翻訳精度を測る指標であり、1 つ以上の規準訳文と機械翻訳文の近さを測定するものである。ここでは IMPACT を再訳文と原文の間の近さを測定するために用いる。IMPACT は数値が大きいほど近いと判断される。そこで、原文と各再訳文の間の IMPACT 値を計算し、最大の値を取る再訳文を選択し、その元となった訳文を最適訳文として出力する。

ここで、訳文に訳し漏れが存在する場合は、不適格な訳文として除外する。極端な場合、原文をそのまま訳文として出力するシステムは再翻訳の結果、原文そのものに戻るので IMPACT 値は最大となるが、そのようなシステムは翻訳とはいえず不適格なものである。今回の場合は、RBMT 出力の英語訳文に日本語の単語（文字列）がそのまま含まれている場合があり、そのような場合は強制的に SPE 出力を選択するようにした。

3 実験結果

NTCIR-9 の日英特許翻訳タスクの試験文の中で人手

評価が行われた 300 文を対象に実験した。人手評価基準としては Adequacy と Acceptability が用いられている [3]。Adequacy は訳の正確性を評価する指標で 5 段階評価となっている。5 が最高、1 が最低レベルである。一方 Acceptability は NTCIR-9 で新たに導入された人手評価基準であり、原文の文意が訳文に表現されているかどうかを示す指標である。AA, A, B, C, F の 5 段階で評価され、AA が最高、F が最低である。AA から C までは、原文の文意が訳文に表現されており、F は表現されていないと判断される。

表 1 には試験文 300 文のうち、人手評価の各ランクにカウントされた文数を示す。表中、理想的選択とは、RBMT と SPE のうち上位にランクされた方を確実に選択できたという理想的な場合を想定したときのカウント数である。図 2 は、表 1 を割合で示したものである。

[3] では、各システムの手評価結果に対して符号検定を実施している。その結果、Adequacy においては、RBMT1 が EIWA より有意水準 5% で有意に評価が高かった。Acceptability においては、両者の有意差が認められなかった。

表 2 に本手法と RBMT および SPE の結果を比較して、改善数と改悪数、符号検定の結果を示す。RBMT と本手法を比較すると Adequacy、Acceptability とともに改善数が改悪数を上回っているものの、有意性が認められるまでには至らなかった（表中 NS で示す）。

表 1 人手評価結果

Adequacy	RBMT	SPE	本手法	理想的選択
5	82	78	91	111
4	77	64	72	83
3	70	76	63	59
2	60	73	63	45
1	11	9	11	2

Acceptability	RBMT	SPE	本手法	理想的選択
AA	23	27	30	33
A	35	27	36	55
B	83	60	80	83
C	30	34	30	36
F	129	152	124	93

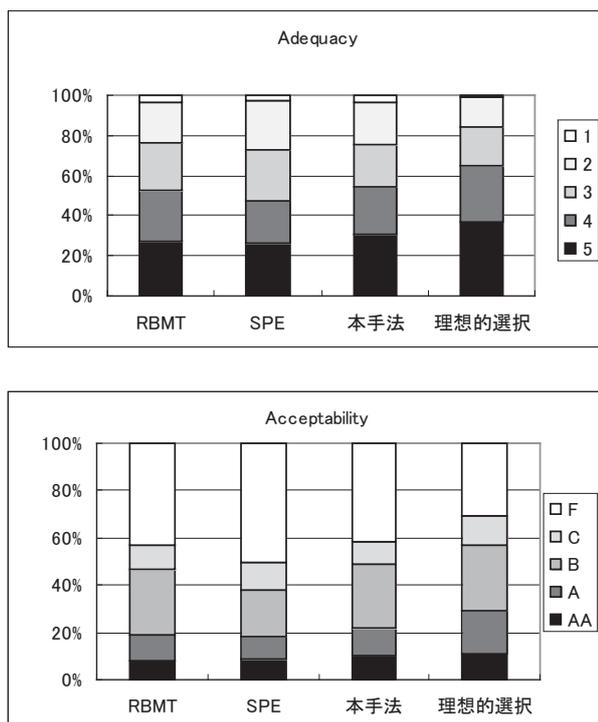


図2 人手評価結果

SPE と本手法を比較した場合は、有意水準 1% で有意に改善が認められた (表中 ** で示す)。

表 3 には、本方式での翻訳例を示す。(a) は RBMT が選択された例であり、(b) は SPE が選択された例である。

表 2 符号検定の結果
(a) Adequacy

改善・改悪\比較対象	RBMT	SPE
改悪	34	31
同値	228	206
改善	38	63
有意性	NS	**

(b) Acceptability

改善・改悪\比較対象	RBMT	SPE
改悪	26	28
同値	235	212
改善	39	60
有意性	NS	**

表 3 翻訳例
(a) RBMT が選択された例

原文	突起を切削機構に深く挿入することにより、鉛筆は切削後すぐに突起に当接するので、芯の太い鉛筆を削ることができる。
規準訳	Moving the projection deep into the cutting mechanism enables a pencil to strike on the projection immediately after shaving, thereby enabling sharpening of a pencil having a thick core.
RBMT	Since a pencil contacts a projection immediately after cutting by inserting a projection in a machining device deeply, a pencil with a thick core can be shaved.
RBMT の Adequacy / Acceptability	5 / A A
RBMT の再訳文	切削機構に突内部を深く挿入することにより切断した直後に、鉛筆が突内部と接触するので、厚い芯を備えた鉛筆は削ることができる。
SPE	Since the pencil comes into contact with the projection immediately after the cutting by inserting a projection in a machining apparatus and a light beam having a thick core can be scraped off.
SPE の Adequacy / Acceptability	2 / F
SPE の再訳文	厚い芯がある、機械加工できる装置およびビームに突内部を挿入することによる切断をこすり取ったことができる直後に、鉛筆が突内部と接触するので。

(b) SPE が選択された例

原文	そのため、車両用サンバイザーでは、遮光板がシャフトから外れないように抜け止め構造を採用している。
規準訳	Therefore, such a vehicular sun visor employs a retaining structure for preventing the shielding plate from being disengaged from the shaft.
RBMT	Therefore, in the sun visor for vehicles, structure is adopted stop escaping so that a gobo does not separate from a shaft.
RBMT の Adequacy / Acceptability	2 / F
RBMT の再訳文	したがって、乗り物のためのサンバイザーでは、ゴーボーが軸体から分離しないように、構造は採用された停止回避である。
SPE	Therefore, in the sun visor, structure is adopted Configuration so that the light shielding plate does not separate from the shaft.
SPE の Adequacy / Acceptability	4 / B
SPE の再訳文	したがって、サンバイザーでは、光シールド板が軸体から分離しないように、構造は採用された構成である。

4 おわりに

規則方式機械翻訳 (RBMT) と統計的後編集 (SPE) を組み合わせたハイブリッド翻訳システムにおいて、RBMT 出力と SPE 出力を比較してシステムが良いと判断する方を選択して出力する方法を実験した。本システムの有効性を NTCIR-9 の結果を用いて評価した結果、改善数が改悪数を上回ったものの、統計的有意性までは示せなかった。今後の課題として、複数の指標を組み合わせるなどにより、改悪数を少なくすることがあげられる。

参考文献

- [1] 江原暉将：句レベルを用いた統計的後編集の精度向上、Japio 2008 YEAR BOOK、pp.262-265、2008。
- [2] 江原暉将：規則方式機械翻訳と統計的後編集を組み合わせた特許文の日英機械翻訳、Japio 2010 YEAR BOOK、pp.280-283、2010。
- [3] Isao Goto, Bin Lu, Ka Po Chow, Eiichiro Sumita and Benjamin K. Tsou: Overview of the Patent Machine Translation Task at the NTCIR-9 Workshop, Proceedings of NTCIR-9 Workshop Meeting, pp.559-578, 2011.
- [4] Terumasa Ehara: Machine translation system for patent documents combining rule-based translation and statistical post-editing applied to the PatentMT Task, Proceedings of NTCIR-9 Workshop Meeting, pp.623-628, 2011.
- [5] 江原暉将、後藤功雄：機械翻訳結果の自動評価と人手評価の比較、Japio 2011 YEAR BOOK、pp.238-241、2011。
- [6] 秋葉泰弘、渡辺太郎、隅田英一郎：SMT 指標を用いた複数翻訳エンジンからの最適訳選択、情報処理学会研究会資料、自然言語処理 150-10, 2002。
- [7] 江原暉将：規則方式機械翻訳と統計的後編集を組み合わせた特許文の日英機械翻訳 (その 3)、平成 22 年度 AAMT/Japio 特許翻訳研究会報告書、pp.43-46、2011。
- [8] 鈴木博和：統計的後編集の効果的利用に関する検証、言語処理学会第 18 回年次大会発表論文集、pp.495-498, 2012。
- [9] Shoichi Yokoyama, Akira Kumano, Masaki Matsudaira, Yoshiko Shirokizawa, Mutsumi Kawagoe, Shuji Kodama, Hideki Kashioka, Terumasa Ehara. Shinichiro Miyazawa and Yasuo Nakajima: Quantitative evaluation of machine translation using two-way MT, Proceedings of the Machine Translation Summit VII, pp.568-573, 1999.
- [10] Hiroshi Echizen-ya, Kenji Araki: Automatic Evaluation Method for Machine Translation using Noun-Phrase Chunking, Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics, pp.108-117, 2010.