

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

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。2009年度から特許版産業日本語委員会委員。



1 はじめに

機械翻訳の方式には、大きく分けて、規則方式、用例方式、統計方式の3つの方式がある。

規則方式は、翻訳規則を主として人手によって構築する方式で、機械翻訳開発の当初から用いられてきた。そのため、長年の研究成果を利用できる利点がある。しかしながら、翻訳手順を単純な規則で記述することは難しく、規則数が多くなって人手での管理が困難な状態にある。

用例方式は、入力文に類似した翻訳用例を用い、それらの訳文を組み合わせることで翻訳する方式であり、1884年に長尾真によって提唱されて以来 [1]、精力的に研究開発が進められている。

統計方式は、大量の対訳ペアを集め、それらから翻訳規則を機械学習の方法で自動的に構築するもので、1990年に Peter F. Brown らによってシステムが示されて以来 [2]、既に20年の歴史を持ち、機械翻訳の主流の一つとなっている。統計方式は、統計的音声認識の考え方を機械翻訳にも適用したものであり、語順の類似した言語間では、相当の精度を達成している。しかしながら日本語と英語のように語順が大きく異なる言語間では、いまだに精度が不十分である。語順の相違が大きい言語間では構文情報の利用が必須と思われ、実際、統計方式に構文情報を利用する試みも成されている。

規則方式でも、構文情報を得るために構文解析を行っており、長い歴史を経て徐々に精度が向上してきている

ものの、特に特許文書のように構文が複雑な文書に対しての構文解析精度は不十分なものがある。

本文で述べる機械翻訳方式は規則方式の利点と統計方式の利点を組み合わせるものであり、規則方式機械翻訳の出力に対して統計的自動後編集を施し、翻訳精度を向上させようとするものである。

2 システムの構成

本システムの構成を図1に示す。入力日本語文は規則方式機械翻訳で英語に翻訳されたあと、統計的後編集によって後編集後の英語文に書き換えられる。規則方式機械翻訳では、主として人手によって構築された辞書と文法(構文規則や翻訳規則など)が用いられる。一方、統計的後編集では、文対応のついた大量の日英特許対訳データを訓練データとして機械学習を適用し、その結果得られた言語モデルと翻訳モデルを利用する。訓練デー

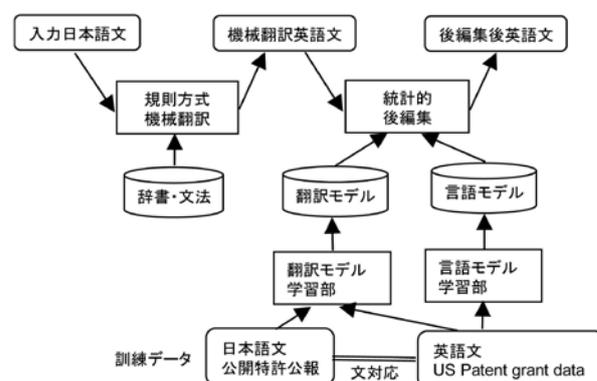


図1 システム構成

タとしては、NTCIR-7 の formal run で用いられた訓練データを用いた [3]。

ここで用いた統計的後編集の特徴は、翻訳対象文に適合した訓練文を用いる点である。たとえば、表 2 の例文 1『図5は回転羽根2を駆動するモータの構成例を示す図である。』に対する訓練文としては、付録に示すような訓練文を抽出して用いている。

規則方式のみの場合に比較して、後編集を加えた方式の精度が向上していることが分かる。

	規則方式	規則方式 + 後編集
BLEU	0.1961	0.2998
NIST	6.1913	7.3058

表 1 翻訳精度

3 翻訳実験の結果

NTCIR-7 で試験データとして提供された日本語文 1381 文を用いて翻訳実験を行った。実験の設定条件の詳細については文献 [4] を参照されたい。実験結果の翻訳精度を自動評価基準として標準的に用いられている BLEU と NIST で測定した。結果を表 1 に示す。規

4 翻訳例

表 2 に翻訳例を示す。表中、原文は公開特許公報の文であり、規準翻訳文は US Patent grant data の文である。例文 1 と例文 3 では、後編集によって訳語が改善されている。例文 2 では、規則方式の訳文で構文が崩れており、後編集でも回復できていない。なお、例文

1	原文	図5は回転羽根2を駆動するモータの構成例を示す図である。
	基準翻訳文	fig . 5 is a diagram showing a structural example of a motor for driving the rotating blade 2 .
	規則方式訳文	drawing 5 is a figure showing the example of composition of the motor which drives moving vane 2 .
	後編集後訳文	fig . 5 is a diagram showing an example of a motor for driving the rotator 2 .
2	原文	さらに、心線ワイヤ51の先端部がラミネートフィルム59により挟まれて保持され、その変形、ピッチの狂いを防止する。
	基準翻訳文	moreover , the front ends of the core wires 51 are sandwiched with laminated films 59 to prevent deformation of the core wires 51 for the purpose of maintaining their relative positions intact .
	規則方式訳文	it faces across the tip part of cable core wire 51 with laminate film 59 , it is held , and the deviation of the modification and a pitch is prevented .
	後編集後訳文	the distal end portion of the core wire 51 with the lamination film 59 is maintained , and the offset of the modification and its pitch is prevented .
3	原文	この絶縁ハウジング10の外面に取り付けられるシールドカバー30を図6に示している。
	基準翻訳文	fig . 6 shows the shield cover 30 , which is to be mounted on the insulative housing 10 .
	規則方式訳文	shield cover 30 attached to the external surface of this insulating housing 10 is shown in drawing 6 .
	後編集後訳文	the shield cover 30 attached onto the outer face of the insulating housing 10 is shown in fig . 6 .

表 2 翻訳例

2は、規準翻訳文の表現が原文の表現とかなり異なっている。

5 おわりに

規則方式機械翻訳に統計的後編集を組み合わせることで機械翻訳精度を向上できることが分かった。本文ではNTCIR-7のデータを用いたが、NTCIR-8のデータを用いた場合でも同様の結果が得られている [5] [6]。

翻訳例の例文2からも分かる通り、規則方式機械翻訳で訳文の構文が崩れている場合、統計的後編集ですそれを回復することは困難である。訳文が崩れる原因として日本文の構文解析に失敗していることが考えられる。特許文は長文であり、構文が複雑である。このような文に対して構文解析の精度を向上させることが今後の課題となる。そのためには、文パターンの利用や産業日本語プロジェクトの中で検討されている明晰な文章表現の利用が有効であろう。

[参考文献]

- [1] Makoto Nagao : Framework of a Mechanical Translation between Japanese and English by Analogy Principle, In Artificial and Human Intelligence, pp.173-180, 1984.
- [2] Peter F. Brown, John Cocke, Stephen Della Pietra, Vincent J. Della Pietra, Frederick Jelinek, John D. Lafferty, Robert L. Mercer, Paul S. Roossin : A Statistical Approach to Machine Translation, Computational Linguistics, Volume 16, Number 2, 79-85, June 1990.
- [3] Atsushi Fujii, Masao Utiyama, Mikio Yamamoto, Takehito Utsuro : Overview of the Patent Translation Task at the NTCIR-7 Workshop, Proceedings of NTCIR-7

Workshop Meeting, pp.389-400, December, 2008.

- [4] 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機械翻訳 (その2)、平成21年度 AAMT/Japio 特許翻訳研究会報告書、pp.56-60、2009.
- [5] Atsushi Fujii, Masao Utiyama, Mikio Yamamoto, Takehito Utsuro, Terumasa Ehara, Hiroshi Echizen-ya and Sayori Shimohata : Overview of the Patent Translation Task at the NTCIR-8 Workshop, Proceedings of NTCIR-8 Workshop Meeting, June, 2010.
- [6] Terumasa Ehara : Machine Translation for Patent Documents Combining Rule-based Translation and Statistical Post-editing, Proceedings of NTCIR-8 Workshop Meeting, June, 2010.

付録 表2の例文1に対して抽出された訓練データ例 (原文・規則方式訳文・基準翻訳文)

図7は、ワード線駆動回路の構成例を示す。
Drawing 7 shows the example of composition of a word line drive circuit.
FIG. 7 shows an example of a word line drive circuit.

図5は本実施例におけるモータ駆動回路の構成要素を示すブロック図である。
Drawing 5 is a block diagram showing the component of the motor drive circuit in this example..
FIG. 5 shows a block diagram of a configuration of a motor drive circuit in the present embodiment.

回転羽根60は、駆動源であるモータ81に連結されている。
Moving vane 60 is connected with motor 81 which is a source of a drive.
The rotator 60 is connected with a motor 81 (FIG. 1) as a drive source.

図5は、図4に示すフォーカルプレーンシャッタ羽根をフォーカルプレーンシャッタに組み込んだ例を示す。
Drawing 5 shows the example which built into the focal-plane shutter the focal-plane shutter shuttlecock shown in Drawing 4.
FIG. 5 shows an example in which the focal plane shutter blade of FIG. 4 is assembled on a focal plane shutter.

図6には、図1で示したシャッタ羽根7の詳細な構成が示され、図7には、駆動レバー8の詳細な構成が示されている。

The detailed composition of shutter shuttlecock 7 shown in Drawing 1 is shown in Drawing 6, and the detailed composition of driving lever 8 is shown in Drawing 7.

FIGS. 6 (A) and (B) show the detailed arrangement of the shutter blade 7 shown in FIG. 1, while FIGS. 7 (A) and (B) show the detailed arrangement of the drive lever 8.

すなわち、駆動モータ51の単体の回転数の変化を説明すると、図28に示すようになる。

That is, when change of the number of rotations of the simple substance of drive motor 51 is explained, it comes to be shown in Drawing 28.

It is FIG. 28 that explains the change in the number of revolutions of the drive motor 51 alone.

(モータ駆動制御回路／モータ駆動制御装置の構成) 図3は、モータ駆動制御回路4を示す回路図である。

(Composition of a motor drive control circuit / motor drive controller) Drawing 3 is a circuit diagram showing motor drive control circuit 4.

FIG. 3 is a circuit diagram of the motor driving control circuit 4.

図5にセンサ駆動パルスとモータ駆動パルスの発生部の構成の一例を示す。

An example of the composition of the generating part of a sensor drive pulse and a motor drive pulse is shown in Drawing 5.

FIG. 5 shows one example of the generators for the sensor driving pulse and the motor drive pulse.

図3にその構成例を示す。

The example of composition is shown in Drawing 3.

FIG. 7 shows the example of the construction.

図14は、パルスモータおよびその制御装置の構成例を示す図である。

Drawing 14 is a figure showing a pulse motor and the example of composition of the control device.

FIG. 14 is a view showing one example of the arrangement of a pulse motor and a pulse motor control apparatus.

図8にディスクを回転駆動する構成を示す。

The composition which rotates a disk is shown in Drawing 8.

FIG. 8 shows a configuration for rotating the disk.