

## A32

ニューラル翻訳を用いた中国特許  
機械翻訳精度の検証:

中国特許の日本語及び英語への  
機械翻訳精度の検証

○ 富士フイルム(株)

田畑文也

# 発表内容目次

1. はじめに
2. 機械翻訳精度評価に用いたシステム
3. 機械翻訳精度評価方法
4. 機械翻訳精度評価結果と考察
5. まとめ

(注)本発表は、2017年度 アジア特許情報研究会の  
WG活動の一環です。

# 1.はじめに

## 1.1 AIとニューラル翻訳

- AI(人工知能)の急激な進化で、人間の脳を模倣したニューラルネットワークを用いた手法で、自然言語に関する処理技術も進み、この手法を用いたニューラル(機械)翻訳(NMT: Neural Machine Translation)が登場！

## 1.2 機械翻訳のトレンド

- ・ルールベース(機械)翻訳

(**RBMT**: Rule-Based Machine Translation)



- ・統計(的機械)翻訳

(**SMT**: Statistical Machine Translation)



- ・ニューラル(機械)翻訳

(**NMT**: Neural Machine Translation)

(注: **トレンドと翻訳精度は一致する訳ではない**)

# 1.3 機械翻訳のアルゴリズム

翻訳方法名	翻訳のアルゴリズム
ルールベース翻訳 (RBMT)	ルールに基づいて翻訳する方法
統計翻訳 (SMT)	大量の対訳データを解析し、その統計結果から適した訳し方を割り出す翻訳方法
ニューラル翻訳 (NMT)	機械学習を用いたディープニューラルネットを使用し、単語の意味だけでなく接頭辞や語幹、単語の位置なども考慮し、自然な文の流れを分析して翻訳する方法

表1. 機械翻訳のアルゴリズム

## 1.4 NMT(ニューラル翻訳)対応例

- (例)Google翻訳 SMT→NMTへ  
英語←→日本語 2016年11月頃  
中国語←→日本語 2017年8月下旬頃



ただし、2017年8月時点で、NMTに対応した中→日翻訳システムは非常に少ないのが現状である。

図1.Google翻訳システム例

(注: Google翻訳は上記図1と、Webページ翻訳等では翻訳アルゴリズムが異なること有り。本報告では上記図1に表示されるものを用いた。)

## 1.5 今回の評価の目的

実務で困っている中国特許の機械翻訳レベルの把握のため

- ・NMTの翻訳精度を従来のアルゴリズムと比較し、レベル確認する。
- ・NMT特有の問題がないか、またその問題レベルを確認する。

を実施した。なお、中→日のNMTに対応したシステムが非常に少ないので、中→英も合わせて評価した。

## 2.機械翻訳精度評価に用いたシステム

### 2.1 評価に用いたシステム(中→日)

- 中国語→日本語 (無料で一般公開されているものから選択)

システム	URL	中国語→日本語 翻訳アルゴリズム
Google翻訳	<a href="https://translate.google.co.jp/">https://translate.google.co.jp/</a>	NMT(推定)
MS翻訳	<a href="https://translator.microsoft.com/neural">https://translator.microsoft.com/neural</a>	NMTおよびSMT
JPO中韓文献	<a href="http://www.ckgs.jpo.go.jp/">http://www.ckgs.jpo.go.jp/</a>	RBMT

表2. 中国語から日本語への機械翻訳システム

注1: Google翻訳については、2017年8月以前のSMTのデータも一部残っており、これも使用した



## 2.2 Google翻訳

### 中→日、英のアルゴリズムはNMT(推定)



図.2 Google翻訳

URL: <https://translate.google.co.jp/>

(注) Google翻訳については、マウスオーバーした時、単語でなく文がハイライトされると翻訳アルゴリズムがNMTとのこと。

## 2.3 MS翻訳

# NMTとSMTの両方に対応

The screenshot shows the Microsoft Translator interface. At the top, there is a dark grey header with the Microsoft Translator logo and name. Below this, a green icon of a flask is followed by the text "Try & Compare" and "Artificial Intelligence, powered by neural networks". The main input area is a white box with a dropdown menu set to "Chinese Simplified" and a character count of "28/1000". The input text is "含有聚对苯二甲酸乙二酯膜作为保护膜の偏光板及其制造方法". Below the input box is a green button labeled "Translate & Compare!". Below this, there are two output boxes. The left box is labeled "Japanese" and contains the text "ポリエチレンテレフタレート膜を保護膜として含む偏光子と2つのエステル製造方法". The right box is labeled "English" and contains the text "ポリエチレンテレフタレートフィルム偏光板保護フィルムを含むとその製造方法". At the bottom, there are two buttons: "Neural" (dark teal) and "Statistical" (brown).

URL:  
<https://translator.microsoft.com/neural>

一つのシステム上で、NMTとSMTに対応しており、同じシステム上で比較できる。

図.3 MS翻訳

## 2.4 JPO中韓文献翻訳・検索システム RBMTで機械翻訳された文を提供するシステム



### スクリーニング

< 前の文献 CNA-106929832 [次の文献](#) >  n件目の文献を表示 表示：

キーワード反転

(54) 【発明の名称】 ポリアセトアミドあるいはポリエチレンテレフタレート基材基材のウェット金属化プロセスモニタ制御システム

(71) 【出願人】

[図 1](#)

図.4 JPO中韓文献翻訳・  
検索システム

URL: <http://www.ckgs.jpo.go.jp/>

## 2.5 評価に用いたシステム(中→英)

- 中国語→英語 (無料で一般公開されているものから選択)

システム	URL	中国語→英語 翻訳アルゴリズム
Google翻訳	<a href="https://translate.google.co.jp/">https://translate.google.co.jp/</a>	NMT
MS翻訳	<a href="https://translator.microsoft.com/neural">https://translator.microsoft.com/neural</a>	NMTおよびSMT
WIPO翻訳	<a href="https://patentscope.wipo.int/translate/translate.jsf?interfaceLanguage=en">https://patentscope.wipo.int/translate/translate.jsf?interfaceLanguage=en</a>	NMTおよび 従来型(非公表)
百度翻訳	<a href="https://fanyi.baidu.com/">https://fanyi.baidu.com/</a>	NMT(推定)

表3. 中国語から英語への機械翻訳システム

# 2.6 WIPO翻訳

## 中→英 NMTと従来型（非公表）で翻訳

The screenshot shows the WIPO Translate NMT interface. At the top, there is the WIPO logo and the word "TRANSLATE" in blue. Below it, it says "Instant patent translation". There are navigation links for "Home", "IP Services", "PATENTSCOPE", "Database Search", and "WIPO translate". A language selection bar at the top right includes "العربية | English | Español | Français | Русский | 中文".

The main content area has a blue header "Translate" and a "[help/user guide]" link. The text below reads: "WIPO Translate NMT is a powerful instant translation tool, designed specifically to translate patent texts (now almost all languages are available using Neural Machine Translation technology). Simply cut and paste text from a patent document into the box below and select from the available language pairs, then click on 'Translate'."

The "Text to be translated:" field contains the Chinese text: "含有聚对苯二甲酸乙二醇膜作为保护膜 的偏光板及其制造方法".

The "Language pair:" dropdown is set to "Chinese->English (Neural MT)".

The "Domain:" dropdown is set to "ADMN-Admin, Business, Management & Soc Sci".

A "Translate" button is located below the form.

Below the form, a warning box states: "This automatic translation is provided for information only, it may contain discrepancies or mistakes and does not have any juridical value." It includes three bullet points: "Please hover your mouse over parallel segments of text", "Click to view other proposals", and "Select words or phrases on the left to access other translation proposals".

The translation result is shown in a table with two columns:

含有聚对苯二甲酸乙二醇膜作为保护膜 的偏光板及其制造方法	polarizing plate containing polyethylene terephthalate film as protective film and manufacturing method thereof
------------------------------	-----------------------------------------------------------------------------------------------------------------

An "Edit translation" button is located at the bottom left of the result area.

URL:  
<https://patentscope.wipo.int/translate/translate.jsf?interfaceLanguage=en>

(注: 中国語→日本語への翻訳には対応していない)

図5. WIPO翻訳

## 2.7 百度翻訳

### 中→英の翻訳はNMT（推定）を使用



図6. 百度翻訳

URL: <https://fanyi.baidu.com/>

百度翻訳は、中国人の方がよく使われている機械翻訳システムです。

# 3.機械翻訳精度評価方法

## 3.1 評価方法、基準

特許庁 情報技術統括室, 特許文献機械翻訳の  
品質評価手順Ver1.0 (2014年6月)を参考に

[http://www.jpo.go.jp/shiryou/toushin/chousa/pdf/tokkyohonyaku\\_hyouka/01.pdf](http://www.jpo.go.jp/shiryou/toushin/chousa/pdf/tokkyohonyaku_hyouka/01.pdf)

(1) 単語としての評価

(2) 文としての評価

を、中→日(4~5システム)、中→英(単語評価のみ 6システム)について、中国語を理解できる日本人評価者1名で実施した。

## 3.2 単語としての評価基準

ランク	レベル	点数
A(適訳語)	人手翻訳に照らし、技術的に同義かつ一般的に用いられる訳語である。	2
B(可訳語)	技術用語として一般的に用いられる訳語ではないが、意味はおおむね正しい。	1
C(誤訳語)	誤訳である。	0
D(不訳語)	未知語、訳漏れである。	0

表4. 技術単語用語の評価基準



# 3.3 文としての評価基準1

## 内容伝達レベルの評価基準

レベル	点数
5:すべての重要情報が正確に伝達されている。(100%)	5
4:ほとんどの重要情報は正確に伝達されている。(80%~)	4
3:半分以上の重要情報は正確に伝達されている。(50%~)	3
2:いくつかの重要情報は正確に伝達されている。(20%~)	2
1:文意がわからない、もしくは正確に伝達されている重要情報はほとんどない。(~20%)	1

表5. 内容伝達レベルの評価基準

## 3.4 文としての評価基準2 流暢さの評価基準

レベル	点数
5: 文意が明解で、人間が書いた日本語文に近い。	5
4: 日本語文として不自然な箇所を若干含むが、文意は明解である。	4
3: 日本語文として不自然な箇所があり、文意がわかりにくい。	3
2: 日本語文法規則に反する表現をかなり含む。文意がわからない。	2
1: 日本語文として成立していない。	1

表6. 流暢さの評価基準

## 3.5 単語として評価した技術用語

日本語	中国語	意図
ポリエチレンテレフタレート (PET)	聚对苯二甲酸乙二酯	確立された汎用用語
シェールガス	页岩气	比較的新しい用語
クラッシュボックス (衝撃吸収ボックス)	碰撞吸能盒	業界専門用語的で、あまり一般用語ではないもの

表7. 単語として評価した技術用語

## 3.6 評価内容一覧

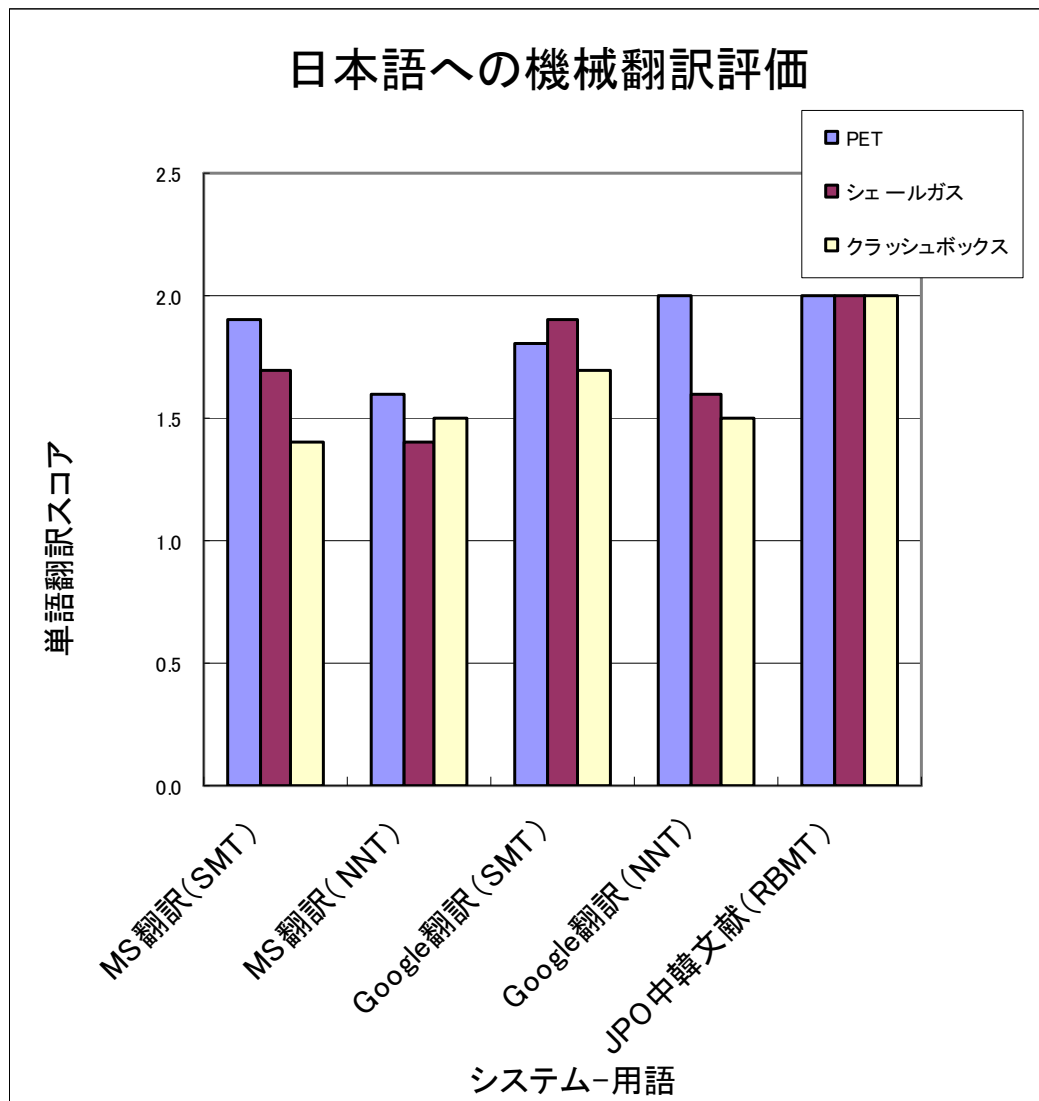
評価単位	技術用語	評価対象部分	対象システム
単語	ポリエチレンテレフタレート	発明の名称 (3用語各N=10 計30種)	中→日 5システム 中→英 6システム
文	シェールガス クラッシュボックス	要約部分 (3系統 各N=3 計9種)	中→日 4システム

表8. 評価内容一覧

(注)文の評価としては、要約のうち、着目した技術用語を含む部分の文を評価した。

# 4.機械翻訳精度評価結果と考察

## 4.1 中→日の単語としての翻訳精度評価

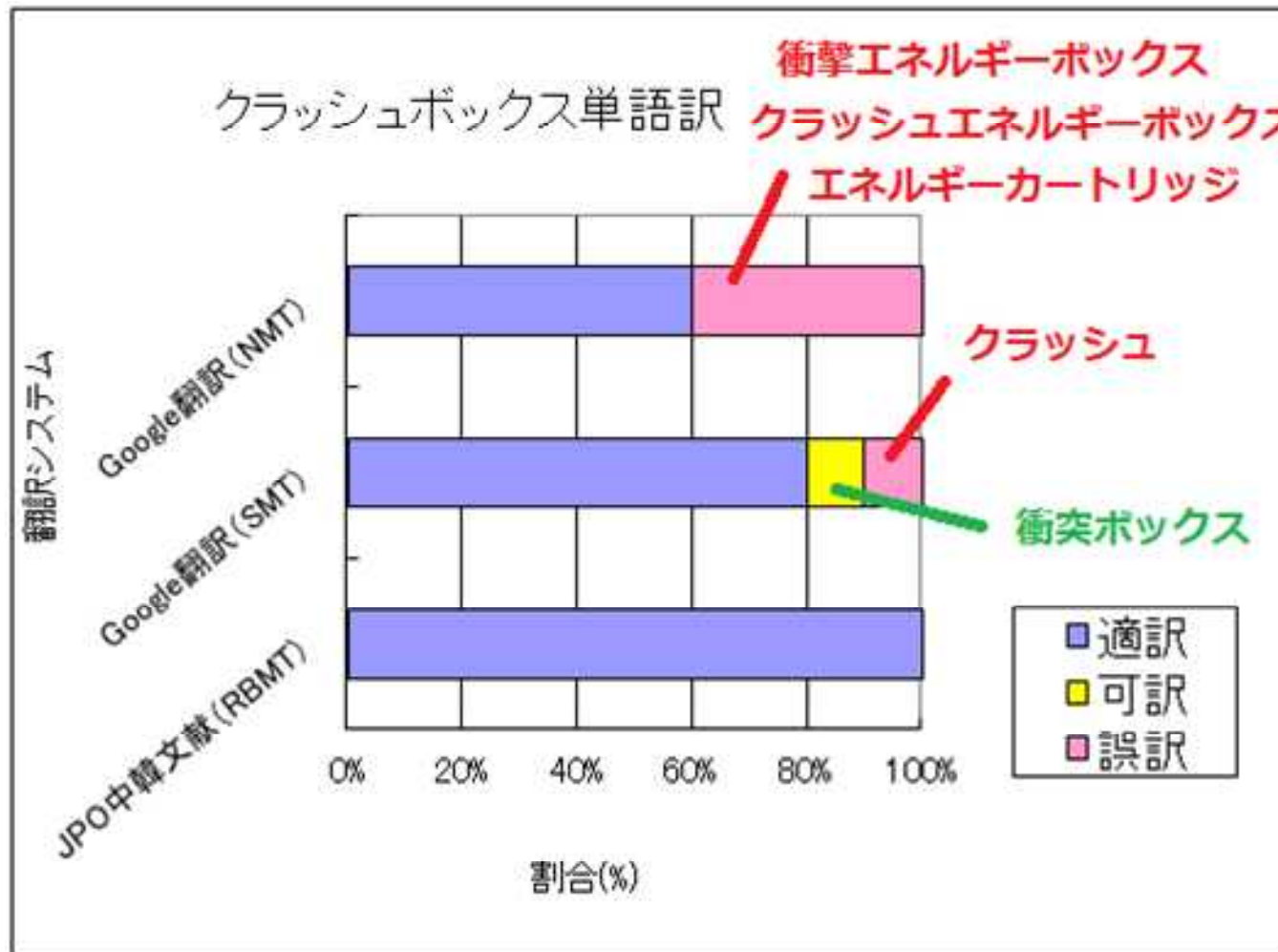


翻訳アルゴリズムがRBMTのJPO中韓文献が、最も単語として翻訳精度良く、従来の翻訳アルゴリズムのものより、むしろNMT方が翻訳精度悪い結果である。

図7. 中国語から日本語への単語翻訳精度評価

# 4.2 中→日単語訳の表記例

## 碰撞吸能盒(クラッシュボックス)の翻訳例

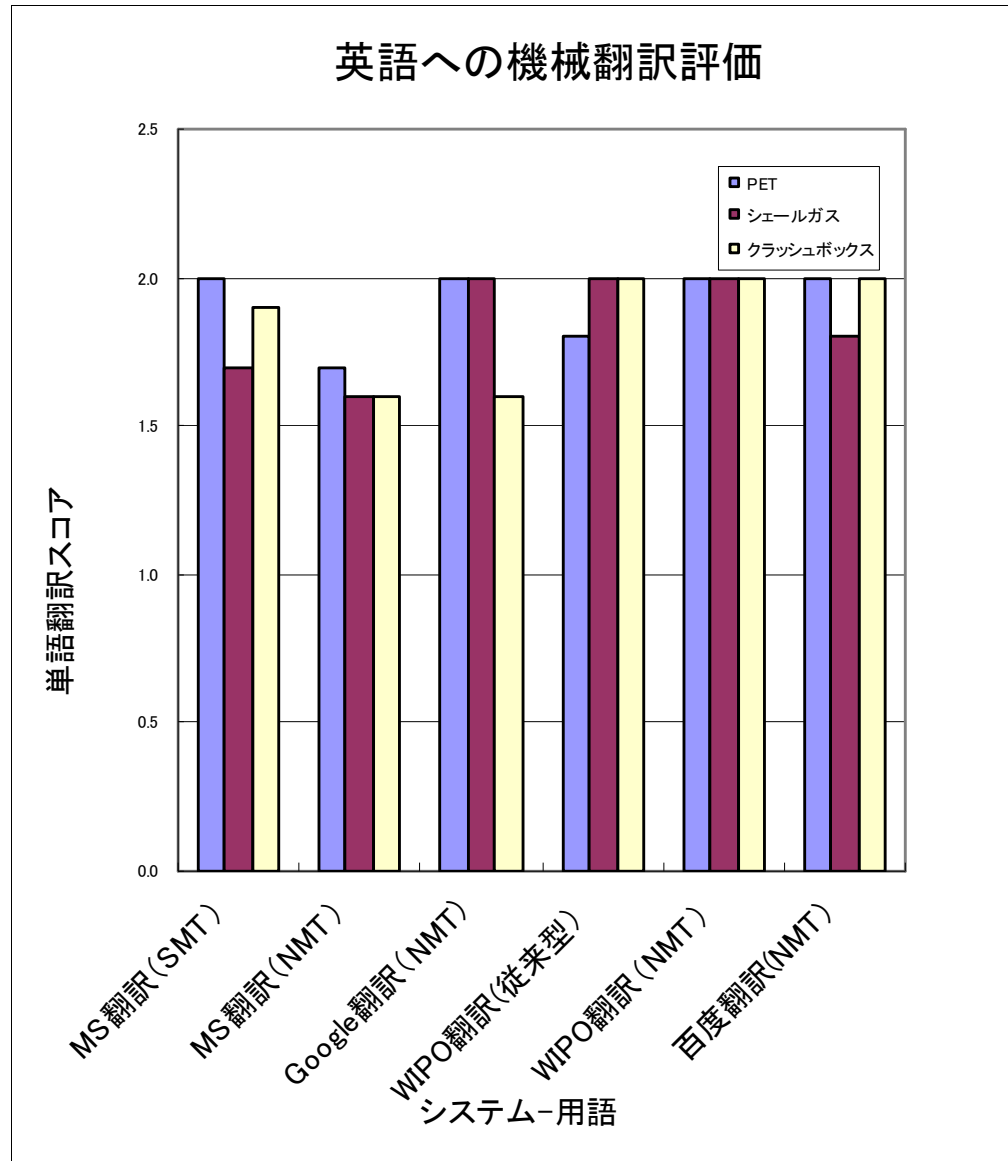


本例では、Google翻訳 (NMT)の誤訳が顕著であることがわかる。

“能”には“できる”と“エネルギー”の二つの意味があり、誤訳が生じたと考えられる。

図8. クラッシュボックスの単語翻訳例

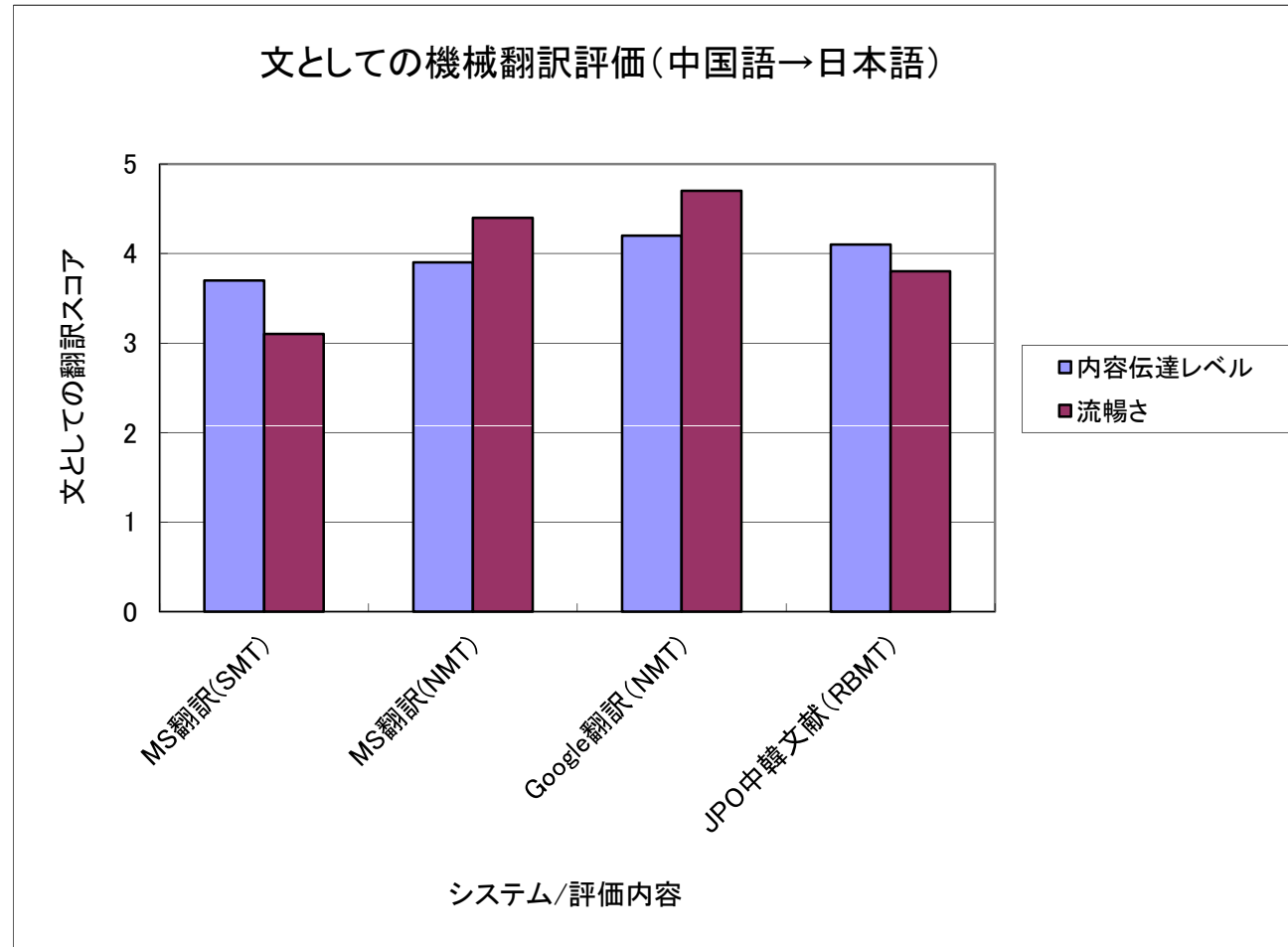
## 4.3 中→英の単語としての翻訳精度評価



中国語から英語への、単語としての機械翻訳精度は、全体的に日本語への機械翻訳よりレベルが高く、翻訳アルゴリズムの差より、システムの差の方が顕著である。

図9. 中国語から英語への単語翻訳精度評価

## 4.4 中→日の文としての翻訳精度評価



単語での翻訳精度と異なり、全体的にNMTの方が精度良く、特に流暢さについてその傾向が強い。

図10. 中国語から日本語への文としての翻訳精度評価



## 4.4 文としての翻訳例

原文 & 人間翻訳	従来型	NMT
CN105579663A [原文] 通过本发明的采收方法，可以有效率地采收页岩气、页岩油等。	[JPO中韓文献] RBMT 本発明の回収法によって、 <b>有効率の地獲得</b> シェールガス、 <b>頁岩油</b> など。 <b>(動詞の訳抜け)</b>	[Google翻訳 NMT] 本発明の回収方法によれば、シェールガス、シェールオイル等を効率良く回収することができる。
[人間翻訳] 本発明の採取方法によって、効率的にシェールガスやシェールオイル等を採取できる。	[MS翻訳 SMT] <b>(訳抜け: 本発明の) 収穫</b> によるは、シェールガス・シェールオイルが効率的に <b>収穫</b> することができます。	[MS翻訳 NMT] <b>(訳抜け: 本発明の) 収穫</b> の方法を通じて、シェールガスやシェールオイルを効率的に <b>収穫</b> できる。

表9. 文としての翻訳例

NMTの文としての流暢さが顕著にわかる。

## 4.5 NMTの訳抜け例 跡形もなく、訳抜けが生じる

原文&人間翻訳	[JPO中韓文献] RBMT	[Google翻訳] NMT
CN102414261A [原文] 本发明提供太阳能电池用双轴拉伸聚酯膜， [人間翻訳] 本発明は太陽電池用二軸延伸ポリエステルフィルムを提供し	本発明は <b>太陽電池用二軸延伸ポリエステル膜</b> を提供し、それは理由としてアンチモン化合物及び／又はチタン化合物を重縮合触媒のポリエチレンテレフタレート組成物組が <b>なった二軸延伸ポリエステル膜</b> とし、	本発明は、重縮合触媒としてアンチモン化合物及び／又はチタン化合物を含有するポリエチレンテレフタレート組成物からなる <b>二軸延伸ポリエステルフィルム</b> であって、 <b>(訳抜け： 太陽電池用)</b>

表10. NMT訳抜け例

NMTは流来型機械翻訳に比べ、流暢に訳されており、一見、誤訳が無いように感じやすいが、訳抜けが散見され、注意が必要と考える。

↑ 文は滑らかだが、“**太陽電池用**”の部分に訳抜けがあり、対象に関する判断ミスを起こしかねない。

# 4.6 NMTの訳語の非統一例

## 碰撞吸能盒(クラッシュボックス)の訳の統一性について

原文	[JPO中韓文献] RBMT	[Google翻訳] NMT
CN100500495C [原文] 本发明涉及一种汽车车身(1)，其有纵梁(2)和插入在纵梁(2)中的碰撞吸能盒(4)，在与纵梁(2)相分离的状态下碰撞吸能盒(4)比纵梁(2)的两个相对壁(7, 7')之间的间距窄，由此可以将碰撞吸能盒(4)	本発明は自動車車体(1)に関し、それはメンバー(2)とメンバー(2)中に挿入する衝突クラッシュボックス(4)があり、メンバー(2)と離間して衝突クラッシュボックス(4)が比べたメンバー(2)の2個の対向壁(7, 7')の間のピッチが狭い状態で、これによって衝突してクラッシュボックス(4)簡便のようにメンバー(2)中に挿入することができる。取り付けした後、衝突クラッシュボックス(4)	本発明は、長手方向のビームと、長手方向のビームに挿入された衝撃エネルギー吸収ボックスとを有する自動車車体に関し、衝撃吸収ボックスは、エネルギーボックス4はサイド部材2の2つの対向する壁7,7'の間の空間よりも狭く、クラッシュボックス4を

→訳の重複も有り  
(同じ訳語が二度出現)

表11. NMTの訳語の非統一例

NMTでは同じ語についての訳語の非統一が散見される。

## 4.7 実務としてのNMT利用 訳抜け等対策方法例

- NMTを用いた機械翻訳が、現時点で、読みやすいと考えるが、**文が滑らかで、訳抜け等の誤訳に気づきにくい**ことも考えられる。当面、査読時には、**単語が決まった訳に設定されやすいRBMTなども併用する**などの対策も有効と考える。

原文	NMT (内容理解用)	RBMT,SMT (誤訳確認用)
----	----------------	---------------------

図11. 査読時のイメージ

# 5.まとめ

## 5.1 まとめ1

中国特許について、日本語及び英語への機械翻訳精度について調べた結果、

(1) 日本語への翻訳精度は、

単語としては、NMTより従来型翻訳の方が精度良い傾向である。

ただし、文としてはNMTの方が精度良く、特に流暢に訳され、読みやすい傾向にある。

(2) 英語への単語の翻訳精度は、全体的に日本語への翻訳より精度良く、翻訳アルゴリズムより、むしろシステム間の差が顕著な傾向であった。

## 5.2 まとめ2

NMTの翻訳について、

従来型機械翻訳に比べ、流暢で読みやすい傾向にあるが、

(1) 訳抜け

(2) 訳語の非統一

が散見される。ただし、**文が滑らかで誤訳に気づきにくい**。

翻訳文を読む際には、従来の観点とは異なる注意

(例: RBMTと並べて確認するなど)が必要と考える。

今後のNMTの精度向上に期待したい。

ご清聴ありがとうございました！

谢谢大家！

