

## 中韓文献翻訳・検索システム: 検索漏れの低減案: 収録率及び翻訳不良の文法的な観点からの検証

○西尾潤<sup>1)</sup>, 雙田飛鳥<sup>2)</sup>, 石田政司<sup>3)</sup>, 伊藤徹男<sup>4)</sup>  
株式会社ユポ・コーポレーション<sup>1)</sup>, 北京銀龍知識産権代理有限公司<sup>2)</sup>, 株式会社神戸製鋼所<sup>3)</sup>, アジア特許情報研究会<sup>4)</sup>  
〒101-0062 東京都千代田区神田駿河台四丁目3番 新お茶の水ビルディング  
Tel: 03-5281-6675 FAX: 03-5281-0822  
E-mail: Nishio.Jun@mk.yupo.co.jp

## Translation and Search System Operated by JPO Search Leakage Reduction Plan: Coverage and Grammatical Verification of the Translation Failure

NISHIO Jun<sup>1)</sup>, SODA Takatori<sup>2)</sup>, ISHIDA Seiji<sup>3)</sup>, ITO Tetsuo<sup>4)</sup>  
YUPO Corp.<sup>1)</sup>, Dragon IP Law Firm<sup>2)</sup>, Kobe Steel Ltd.<sup>3)</sup>, Asia Patent Information Society<sup>4)</sup>  
Shin-Ochanomizu Building, 4-3, Kanda Surugadai, Chiyoda-ku, Tokyo 101-0062 Japan  
Tel: +81-5281-6675 FAX: +81-3-5281-0822  
E-mail: Nishio.Jun@mk.yupo.co.jp

### 【発表概要】

日本特許庁から提供されている「中韓文献翻訳・検索システム」には、中国・韓国特許文献が機械翻訳で日本語に変換されて収録されている。そのため、公報番号を入力するだけで、誰でも中国・韓国特許文献を日本語で査読できるようになった。このことは先行技術文献調査やSDIの場面における進歩である。

ところが、同システムにおいて未収録があったり、誤訳を含んだまま収録されていたりするならば、査読に支障するだけでなく公報テキスト検索において検索漏れの原因となりうるため、収録率や機械翻訳精度を向上させることは同システム構築上の重要なポイントである。

同システムの誤訳事例の解析から、原語の文法や表記法が誤訳に大きく影響することが分かった。そこで、原語別に同システムの誤訳傾向と翻訳精度向上に必要な事項を紹介する。また、ユーザーの立場から同システムの翻訳精度向上が可能か試行して報告する。

### 【キーワード】

中国特許, 韓国特許, 特許検索, 中韓文献翻訳・検索システム, キーワード検索, 文法, 誤訳, 査読

## 1. はじめに

特許文献は世界最大の技術情報データベースであるといつてよい。特許庁によれば、2013年の推定値では世界の特許文献(実用新案を含む)中、中韓文献が65%を占めるまでになった(図1)。

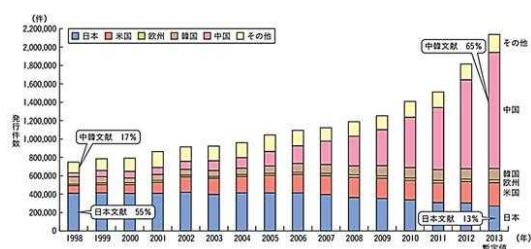


図1. 外国語文献数の増加[1]

一方、中国では外国企業が中国国内の特許権の侵害で提訴され、敗訴するケースが出ている。また、韓国では、優先審査や超高速審査出願によって短期間で設定登録されるケースが年々増加し、クリアランスの重要性が増している。

加えて、グローバルドシエ構想の進展に伴い、拒絶引例で中韓文献を引かれる事例が今後増加すると予想される。

これらの背景から、知財実務者のみならず発明者も中国及び韓国の特許情報検索に慣れるとともに、得られた情報のスクリーニングを短時間に精度よく実行することが求められている。

従来、中国及び韓国の特許情報を把握するには、無料データベース、商用データベースの如何を問わず主として英語に翻訳されて提供された情報を英語で読むことが主流であった。近年では無料翻訳サイトを併用して日本語で読むことも行われている。

しかし、英語情報は原語から英語への翻訳精度等の問題があり、侵害予防調査等の場面では原語情報による補完が必要になる[2]。ところが、先行技術文献調査やSDI[3]のスクリーニングの場面では、発明の内容を理解することが主目的であり、多少精度が悪くても機械翻

訳によって言語のハードルを下げるニーズが存在する。

日本特許庁では、中国・韓国特許文献を日本語で検索し、日本語で全文が表示できる「中韓文献翻訳・検索システム」(以下、「JPO 中韓システム」と称する)を2015年1月15日から正式に運用開始した。

JPO 中韓システムは、中国及び韓国の特許公報を機械翻訳で日本語に変換したうえでデータベース内部に収録している[4][5]。

しかし、JPO 中韓システムにおいて未収録があったり、誤訳を含んだまま収録されていたりするならば、特定公報の査読において支障するだけでなく、公報テキスト検索において検索漏れの原因となりうる。そこで、JPO 中韓システムの収録率や機械翻訳精度を向上させることは同システム構築上の重要なポイントである。

## 2. 目的

本報告は、JPO 中韓システムに収録された状態における精度をユーザーの立場から向上させる方法を見出すことを目的とする。報告に当たっては、以下の観点から検討した。

(1)原語別に、JPO 中韓システムの収録状況を公報発行時期の時系列で調査する。

(2)原語別にJPO 中韓システムの誤訳傾向を明らかにし、機械翻訳精度向上に必要な事項を検討する。

(3)ユーザーの立場から、JPO 中韓システムの翻訳精度向上が可能か試行する。

## 3. JPO 中韓システムの収録状況

### 3-1. 検討方法

2010~2014年の公開公報番号を中国特許はCNIPR日本語版[6]、韓国特許はKIPRIS[7]で全件取得し、これを

収録率 100%とした。また、得られた公開番号を Sharerresearch ((株)日立製作所)で公開番号検索し、結果を Excel データとしてダウンロードし、Excel 上で日本ファミリーを持つ公報を抽出した。

一方、得られた公開番号を JPO 中韓システムの公開番号索引照会に入力して、収録の有無を取得した。

### 3-2. 結果

中国特許公開公報の収録率に特徴があったので、中国の結果を示す。

表 1. 中国公報年次別収録率(2015/9/12 時点)

発行年	CNIPR全件	収録率	JP Family	収録率
2010	315836	99.5%	63746	98.9%
2011	368434	98.3%	72441	96.1%
2012	543296	98.5%	87462	96.5%
2013	632585	99.3%	80665	98.2%
2014	777336	95.8%	83315	85.2%

表 1 から、2010～2013 年の中国公開全件を対象とした収録率は 98～99%で推移しているが、JP ファミリーを有する中国公開に限定すると収録率は各年とも 1～2%低下する。また 2014 年公開分の収録率が低い。

### 3-3. 考察

そこで、JP ファミリーを有する中国公開について、月別収録数と収録率を整理して図 2 に示す(グラフの濃いバーが未収録を示す)。

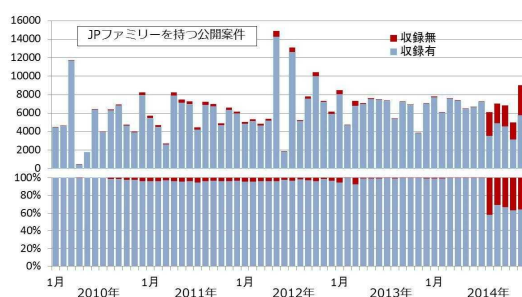


図 2. JP ファミリーを持つ中国公報の公開月別収録傾向(2015/9/12 時点)

図 2 から、2013 年 4 月～2014 年 7

月の収録率はほぼ 100%に近いが、2013 年 3 月以前は毎月約 5%程度の未収録があったことがわかる。このような未収録が生じる理由は推測の域を出ない。

### 3-4. 結論

各年とも未収録が 1%程度存在する。このことは、JPO 中韓システムのテキスト検索を用いてスクリーニングする際、未収録公報が漏れてしまうことを意味する。また、検索日の直近約 1 年の収録率が低いことは、SDI の場面での活用機会が減殺されることを意味する。システム改修等で収録率が改善されることを期待する。

## 4. JPO 中韓システムの誤訳傾向

### 4-1. 検討方法

2013 年 1 月出願の、JP ファミリーを持つ中国・韓国公報番号を JPO 中韓システムに入力して、発明の名称(日本語)を取得した(中国 413 件、韓国 418 件)。一方、当該日本ファミリーの発明の名称を取得して、両者を突き合わせて比較した。

### 4-2. 結果

JPO 中韓システムによる、中日翻訳・韓日翻訳における典型的誤訳例を表 2 及び表 3 に示す。

表 2. JPO 中韓システムによる典型的誤訳例(1)

日本語公報: WO2013111709A <b>金型清掃用樹脂組成物</b>
中国公報: CN104066563A (原文) <b>模具清扫用树脂组合物</b> (中日翻訳) <b>モールドが樹脂組成物を用いることを掃除すること</b>
韓国公報: KR1020140117405A (原文) <b>금형 청소용 수지 조성물</b> (韓日翻訳) <b>金型清掃用樹脂組成物</b>

注: 表中の中日翻訳・韓日翻訳は、JPO 中韓システムによる。

表 3. JPO 中韓システムによる典型的誤訳例(2)

日本語公報: WO2013108810A 溶銑の予備処理方法
中国公報: CN104053794A (原文) 鉄水 预处理方法 (中日翻訳) (JPO 中韓翻訳システム未収録)
韓国公報: KR1020140102742A (原文) 용선의 예비 처리 방법 (韓日翻訳) 傭船の予備処理方法

表 2 に記載の例では、中日翻訳において「清掃用」(清掃用)と副詞的に訳出すべきところ、「清掃」「用」をそれぞれ「掃除する」「用いる」と、動詞として認識している。一方、韓日翻訳において正確に訳出している。

また、表 3 に記載の例では、韓日翻訳において「溶銑」と訳出すべきところ、「傭船」と同音異義語が当てられたため意味が不明になった。一方、中国公報は JPO 翻訳システムに未収録であった。

#### 4-3. 考察

本項では JPO 中韓システムが中国・韓国特許文献を機械翻訳する際に文法解釈を誤ることに由来する誤訳を取り扱い、テクニカルタームの誤収録があったとしてもこれを無視する。

ハングル表記は表音文字であること、語順が日本語とほぼ一致すること、文節間にスペースを挿入すること(分かち書き)等の特徴を有する。そのため、韓日機械翻訳において、品詞の解釈を誤る確率が低い。しかしながら、表記が表音文字であるため、同音異義語や語感が近い用語の取り違えが多発している。

図 3 に韓日翻訳の検証結果を示す。単語の異表記(例: 金型とモールド)も少なくはないが、同音異義語等の誤訳が 14%あった。

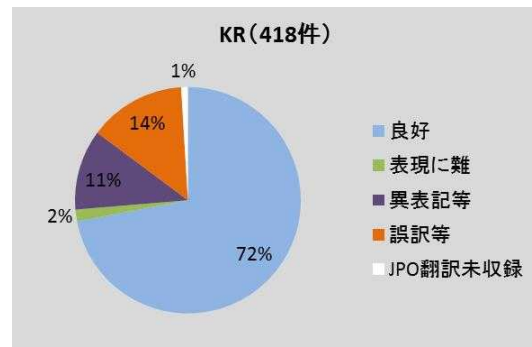


図 3. 韓日翻訳傾向まとめ(2015/4/7 時点)

一方、簡体字中国語表記は表意文字であること、語順が主語+動詞+目的語を基本とすること、分かち書きがないこと等の特徴を有する。分かち書きがないため、複合名詞を主語+動詞として解釈する事例が多発した。また、一文中に「的」が複数存在する場合などの複雑な係り受けの誤解釈が目立った。また、「用于~的」等の介詞を認識できず、この場合「用」を動詞として訳す例が目立った。さらに一語が複数の意味を有する(多義語)場合、「可」を「しかし」、「多」を「あまり」など、一律に訳すため意味不明になる事例が散見された。

図 4 に中日翻訳の検証結果を示す。

誤訳は 8%と、割合はあまり多くはないが、文意を損なう事が多く、場合によっては意味を取り違ふことになり、信頼性を低下させかねない。

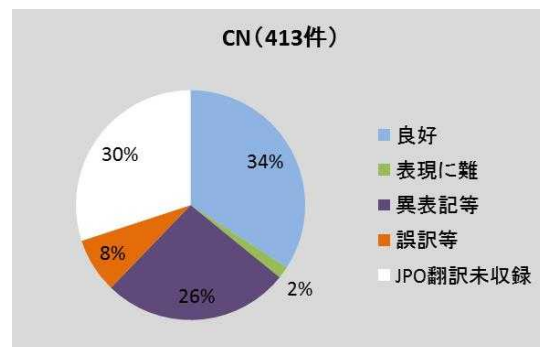


図 4. 中日翻訳傾向まとめ(2015/4/7 時点)

#### 4-4. 結論

上記検証結果から、中日翻訳と韓日

翻訳とで誤訳傾向が異なることが分かったことから、翻訳精度向上の考えも両者で異なる。

中日翻訳の場合、係り受けと介詞を正確に認識することが重要で、形態素分析を強化した翻訳エンジン改良が最も効果的と考える。なお、ユーザー側の立場からは、介詞構造を含んだ文の誤訳報告をして対訳コーパスに登録することが翻訳精度向上の一助になると考えられる。

韓日翻訳の場合、同音異義語を正確に訳出することが最も効果的であるが、ハングル表記は表音文字であることから、単語単位の辞書登録は意味をなさず、前後の文脈から類推できるように対訳コーパスに登録する必要がある。ユーザー側の立場からは、複文節に対する誤訳訂正をこまめに行って対訳コーパスを構築していくことが翻訳精度向上の一助になると考えられる。

## 5. 中韓文献翻訳・検索システムの翻訳精度向上テスト

### 5-1. 検討方法

2013年1月出願の、JPファミリーの有無にかかわらず中国・韓国公報番号をJPO中韓システムに入力して、発明の名称(日本語)を取得した。この中から表4に記載の2例を中日翻訳で誤訳報告し、どのように訂正されるか検証する。

### 5-2. 結果

結果はまだ出ておらず、当日報告することとしたい。

表 4. JPO 中韓システムの誤訳報告候補

<p>例 1. 「用于<u>~</u>的」(「<u>~</u>用」を「<u>~</u>に用いた」などと誤訳する)</p> <p>中国公報:CN103014564A            (原文) 用于制造渣浆泵过流部件的金属材料及其制作方法            (誤訳) <u>スラリーポンプストリーム部材を過ぎる</u>金属材を製造に用い及びその製造方法</p> <p>中国公報:CN103018189A            (原文) 一种用于测定多糖的苯酚试剂的制备方法            (誤訳) <u>糖分の多い</u>フェノール試剤を<u>測定</u>に用いた製造方法</p> <p>中国公報:CN102973744A            (原文) 一种用于混合型痔疮护理的外用药膏            (誤訳) <u>混合型痔介護</u>の外用ぬり薬に<u>用いること</u></p>
<p>例 2. 「可」(「可能」の意を「しかし」と誤訳する)</p> <p>中国公報:CN103071502A            (原文) 一种可磁分离的复合光催化剂及其制备方法和应用            (誤訳) <u>しかし</u>磁気分離の複合光触媒及びその調製方法とアプリケーション</p> <p>中国公報:CN103112332A            (原文) 用于混合动力和纯电动客车的可插拔式燃气动力空调系统            (誤訳) <u>しかし</u>ハイブリッドと純電動車に用いる着脱式可燃ガス動力空調システム</p> <p>中国公報:CN103046785A            (原文) 一种可快速存取立体车库            (誤訳) <u>しかし</u>高速アクセスの立体的のガレージ</p>

## 6. おわりに

JPO中韓システムの翻訳精度について検証した。現状では本システムによる翻訳不良を原因として意味を取り違えることがあり、クリアランスの場面では原語を参照して正確を期さねばならないことは言うまでもない[2]。

しかし、SDIや公報ウォッチングの場

面では、発明を理解することが最も重要であり、中国・韓国公報を誰もが日本語で読むことができるようになったことは格段の進歩といえよう。

なお、JPO 中韓システムを SDI ルーチンに組み込んで運用するためには、タイムラグを極力短くしなければならない。また、2013 年 3 月以前の収録率が改善されない等、収録の問題があり、JPO 中韓システムの改造事業[8]で収録問題が改善されることを期待したい。

中韓特許文献の日本語への機械翻訳に関して、2009 年の日立製作所の調査報告書[9]を嚆矢として、2011 年～2014 年 JAPIO において精力的に研究された一連の報告書[10]が日本特許庁から発行されており、これらの中でも文法的な観点からの検証がなされている。JPO 中韓システムの検証にあたっては、これらの報告が参考文献として有用であった。

謝辞:

本報告は2015年度の「アジア特許情報研究会」のワーキングの一環として報告するものであり、研究会の皆様には情報の提供及び数々のアドバイスをいただきました。ここに改めてお礼申し上げます。

## 7. 参考文献等

- [1] 特許行政年次報告書 2015 年版
- [2] 網羅的な中国特許調査のための「中国語の拾い方」(伊藤 JAPIO YEARBOOK 2014)
- [3] SDI (Selective Dissemination of Information)  
特定条件でヒットする各種公報を定期的に配信する機能
- [4] 中韓文献翻訳・検索システム(榎本 Japio JAPIO YEARBOOK 2014)

- [5] 中国・韓国語の特許文献を日本語で検索可能なシステムのご紹介 (櫻井 他 特技懇 No.276, pp.67-70)
- [6] CNIPR 日本語版  
[http:// www.cnipr.jp/](http://www.cnipr.jp/)
- [7] 韓国特許技術情報センター  
<http://www.kipris.or.kr/enghome/main.jsp>
- [8] 中韓文献翻訳・検索システムの改造(入札結果)  
[https://www.jpo.go.jp/koubo/choutatu/choutatu2/h27system/pdf/h27menu/42\\_270728.pdf](https://www.jpo.go.jp/koubo/choutatu/choutatu2/h27system/pdf/h27menu/42_270728.pdf)
- [9] 「審査関連情報を活用した次世代検索システム開発に向けた調査調査報告書」(株式会社 日立製作所 2009)
- [10] 機械翻訳に関する調査報告書 (JPO 2011～2014)  
のうちの特に以下の 2 報告書
  - ・中国特許文献の機械翻訳のための辞書整備及び機械翻訳の品質評価に関する調査(JAPIO 2014)
  - ・日本語特許出願書類の中国語への機械翻訳に関する調査(JAPIO 2011)