

中韓文献翻訳・検索システムの検証

○佐野浩太郎¹⁾, 前田佳治²⁾, 石田政司³⁾, 伊藤徹男⁴⁾
株式会社 IHI¹⁾, 東亜合成株式会社²⁾, 株式会社神戸製鋼所³⁾,
アジア特許情報研究会⁴⁾

〒135-8710 東京都江東区豊洲三丁目1番1号 豊洲 IHIビル
Tel: 03-6204-7093 FAX: 03-6204-8651
E-mail: koutarou_sano@ihi.co.jp

Study of the Chinese and Korean Gazette Translation and Search System

SANO Kotaro¹⁾, MAEDA Keiji²⁾, ISHIDA Seiji³⁾, ITO Tetsuo⁴⁾
IHI Corp.¹⁾, Toagosei CO., LTD²⁾, Kobe Steel Ltd.³⁾, Asia Patent Information Society⁴⁾
Toyosu IHI Building, 1-1, Toyosu 3-chome, Koto-ku, Tokyo 135-8710 Japan
Tel: +81-3-6204-7093 FAX: +81-3-6204-8651
E-Mail: koutarou_sano@ihi.co.jp

【発表概要】

従来、中国・韓国の特許文献の調査で行われていた機械翻訳を含む英語情報を用いた検索(及び査読)では漏れが発生する。そのため、中国語やハングルによる検索で補完する作業が必要となっている。

そのような状況のもと、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」と略)の提供を開始した。これにより、調査担当者だけでなく開発者などが容易に中国・韓国特許文献を検索し、内容を査読できる。

本報告では、開発者などが「JPO 中韓システム」を利用して中国・韓国特許文献を査読する際の活用法を提案する。具体的には、「JPO 中韓システム」の機能の概略を紹介し現段階で考えられる有効な利用方法を提案する。加えて、機械翻訳された収録内容の検索精度(翻訳異表記・誤表記)を複数の技術分野で検証し、特許文献を検索し翻訳文を査読する際の留意点を明らかにする。

【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, 異表記, 誤訳, 査読

1. はじめに

ここ数年、中国では特許出願の件数が、82.5 万件(2013 年)→92.8 万件(2014 年)と推移しており、大量の特許情報が公開されることが予想される。また、韓国では、出願から登録までの審査が促進されている。例えば、出願人が特許を出願してから最初の通知を受けるまでの審査期間が、14.8 ヶ月(2012 年)→13.2 ヶ月(2013 年)と推移している。

これらの背景から、中国及び韓国の特許情報においては、収集した情報について、効率的に、素早く内容を把握することが求められている。従来、主として英語データベースから得られた情報を把握してきたが、英語情報のみでは情報の漏れが発生するため、近年では原語である中国語やハングルによる検索で補完する作業も必要となっている。

そのような状況のもと、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」(機械翻訳日本語システム、以下、「JPO 中韓システム」と略)の提供を開始した。これにより、調査担当者だけでなく開発者などが容易に中国・韓国特許文献を検索し、内容を査読できるようになった。

2. 目的

発表では、JPO 中韓システムの利用方法の概略を紹介すると共に、開発者などが上記 JPO 中韓システムを利用して中国・韓国特許文献を査読する際の活用法を提案する。上記活用法については、以下の観点より検討した。

(1) 中国語やハングルの翻訳を多分野にわたって検証し、異表記・誤表記の分野別特異性などを検証する。

(2) 中国及び韓国特許文献を日本語で検索及び査読する際の現状を紹介すると共に、留意点を抽出する。

(3) 現状の機械翻訳精度におけるベストな日本語検索式を検討する。

3. JPO 中韓システムの機能

JPO 中韓システムで提供されている、公報テキスト検索、公報番号索引について、留意点を含めて簡単に紹介する。

3-1 データの収録

中国特許は CNIPR、韓国特許は KIPRIS の収録を 100%とした場合の収録率を確認した。

中国特許・実用新案とも 2003 年以降、2013 年までの各年の収録状況は 98%～99%の収録率を示すが、IPC での分野別収録を見てみると A61K、C08F、C12P などの 2011 年、2012 年において 90%前後の低い値を示す分野もあり、韓国特許においては収録率 101～103%を示すものがあるなどばらつきもある。

3-2 公報テキスト検索

公報テキスト検索では、検索項目を選択し、それらを組み合わせた検索が可能である。公報の文章(発明の名称、請求の範囲等)では、日本語キーワードによる検索が可能である。キーワードは、1 項目 100 文字以内、1 用語の長さは 20 文字以内の入力制限がある。国際特許分類では、サブクラス以下まで入力しないと検索できない。出願人では、原語(中国語又はハングル)による検索しかできない。また、近傍検索などの機能も有する。

特許庁 中韓文献 翻訳・検索システム サービスヘルプデスク 受付時間：9:00～18:00 TEL：0120-008525

公報テキスト検索

発行国 中国 韓国 発行種別 特許 実用 文献種別 公開 登録 公知日 ~ 出願日

フリー検索条件			検索キーワード			項目内検索			NOT検索条件		
項目番号	項目名称	検索項目	検索キーワード	項目内検索	項目番号	項目名称	項目名称	項目名称			
1		発明の名称	ディーゼルエンジン ディーゼル機関	OR	1		要約+請求				
2	AND	公開日	20030101:20131231	OR	2	AND	公報全文書				
3	AND	国際特許分類		OR	3	AND	国際特許分類				
4	AND	出願人/権利者		OR	4	AND	出願人/権利者				

検索条件クリア 行追加 最終行削除

図 1. 公報テキスト検索画面(抜粋)

3-3 公報番号索引

所定の形式で公報番号を入力することにより、番号検索が可能である。一度に 1000 件の番号を入力し、検索することが可能である。入力された番号が完全一致することで検索可能となるので、余分なスペースを削除しておく必要がある。



図 2. 公報番号索引画面(抜粋)

4. 検証結果

本報告では、複数の技術分野の事例により、異表記・誤訳の実態等について説明すると共に、中国及び韓国特許文献を日本語として査読する際の留意点を紹介する。

検証は、確認が容易であることから「発明の名称」中の用語の検索件数とし、検証期間は JPO 中韓システムが 2003 年発行以降の公報を対象にしていることから「2003～2013 年公開特許」とした。

なお、中国語やハングルにも複数の異表記が存在し、それぞれについて紹介することは紙数の関係から割愛した。

4-1 カーボンナノチューブの事例：中国

カーボンナノチューブの中国語「碳納米管」で検索した 2918 件について JPO 中韓システムではどのような翻訳日本語が充てられているかを確認した結果を表 1 に示した。

表 1 のように同一の中国語が 3 種の日本語に翻訳されている。「炭ナノチュー

ブ」もカーボンナノチューブであろう、と予想できるが、カーボンナノチューブの検索用語としては予想しがたい。

表 1. 中国語翻訳の日本語表記ゆれ

CNIPR	TI	JPO 中韓システム	TI
碳納米管	2918	カーボンナノチューブ	2694
		炭素ナノチューブ	118
		炭ナノチューブ	57

同一の中国語公報の要約中に「碳納米管」として記述されているのに(図 3) JPO 中韓システムでは、「カーボンナノチューブ」と「炭ナノチューブ」の 2 つの訳があるものもある(図 4)。

CNA-101462713

要約：本発明披露一種処理碳納米管的方法，包括如下的步骤：将碳納米管分散在分散介质中，制备分散体；将所述分散体与吸附剂混合，以使所述分散体中特定类型或特性的碳納米管吸附在所述吸附剂上，其中所述吸附剂经化学/生物改性剂改性而对不同类型或特性的碳納米管具有不同的选择吸附性；以及将吸附剂与分散体分离，由此使得吸附在所述吸附剂上的特定类型或特性的碳納米管富集在分散体中其它类型或特性的碳納米管分离。本发还披露了经该方法处理的碳納米管以及碳納米管器件。

図 3. 中国語原文

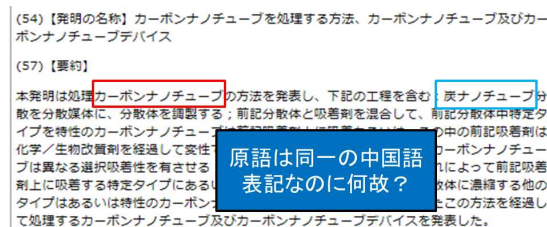


図 4. JPO 中韓システム

4-2 クラッシュボックスの事例：中国

同様に、自動車部品用語であるクラッシュボックスの中国語「吸能盒」について検証した結果を表 2 に示した。

表 2. 中国語翻訳の日本語表記ゆれ

CNIPR	TI	JPO 中韓システム	TI
吸能盒	19	クラッシュボックス	1
		含ませは容器	7
		含ませ・・・ 4 種	4
		ことの容器を吸入すること	4
		ことの容器・・・吸入すること	3

19 件中、正しい日本語に翻訳されたのはわずかに 1 件であり、他は検索用語としてはもちろん、査読時にもクラッシュボックスを予想させるような翻訳とは言えない。

4-3 シェールガスの事例:中国

表 3 に示すシェールガスの例では、中国語の「页岩气」が部分的に訳されているもの、全く訳されず該当する日本語漢字が充てられているものなどがあつた。さらには「炭質けつ炭気」のように予想もできないような翻訳となっている場合があつた。

表 3. 中国語翻訳の日本語表記ゆれ

KNIPR	TI	JPO 中韓システム	TI
页岩气 (shale gas)	64	シェールガス	59
		シェール気	1
		頁岩ガス	1
		頁岩気	1
		炭質けつ炭気	2

4-4 ターボチャージャーの事例:韓国

ターボチャージャーのハングル表記「터보차저」などの異表記について検証した結果を表 4 に示した。

異表記のいずれもほぼ「ターボチャージャー」と翻訳されているが、一部の表記「터보 차저」は「ターボチャヅ」と翻訳され、検索用語としてはもちろん、査読時にも「ターボチャージャー」とは判別しがたい。

表 4. ハングル翻訳の日本語表記ゆれ

KIPRIS	TI	JPO 中韓システム	TI
터보차저	283	ターボチャージャー	280
		翻訳不良	3
터보 차저	61	ターボチャージャー	61
터보차저	93	ターボチャージャー	93
터보 차저	28	ターボチャージャー	28
터보차저	21	ターボチャージャー	21
터보 차저	12	ターボチャージャー	4
		ターボチャヅ	8
터보차저	7	ターボチャージャー	7

4-5 カーボンナノチューブの事例:韓国

カーボンナノチューブのハングル表記「탄소나노튜브」などの異表記について検証した結果を表 5 に示した。

「탄소나노튜브」やスペースを含む異表記に対しては「炭素ナノチューブ」が充てられ、「카본나노튜브」やスペースを含む異表記に対しては「カーボンナノチューブ」が充てられており、検索用語としても、査読に際してもほぼ問題はない。

表 5. ハングル翻訳の日本語表記ゆれ

KIPRIS	TI	JPO 中韓システム	TI
탄소나노튜브	1471	炭素ナノチューブ	1471
"탄소 나노튜브"	324	炭素ナノチューブ	324
"탄소나노 튜브"	3	炭素ナノチューブ	3
"탄소 나노 튜브"	62	炭素ナノチューブ	61
		タンソナノチューブ	1
"탄소 나노관"	1	炭素ナノ管	1
카본나노튜브	187	カーボンナノチューブ	187
"카본 나노튜브"	79	カーボンナノチューブ	79
"카본나노 튜브"	0		
"카본 나노 튜브"	32	カーボンナノチューブ	32

ターボチャージャーの例でも示したように韓国文献では中国文献に比べ翻訳精度は比較的良好のようである。

4-6 その他の翻訳不良事例

化学分野では、tetracarboxylic acid diimide(tetra=4 つの)や azabicyclo(3.1.0) hexane のように数字を含む化合物や、poly(ethylene terephthalate) (PET) のように括弧で表される化合物なども多い。

特に、中国特許文献においては、数字や括弧が入ると、四羧二酰亚胺が「4 のカルボキシルジイミド」や氮杂二环(3.1.0)己烷が「アザビシクロ(3. 1. 0)もうジン」と訳され、「聚(对苯二甲酸乙二醇酯)」も「ポリ(正しいポリエチレン-テレフタレート)」と訳されてしまう傾向があるよう

である。

4-7 網羅的検索のための検索式例

同一の意味を有する用語でも日本語はもちろん英語や中国語、ハングルにも異表記が存在する。したがって、検索する際にはこれらの異表記を考慮して検索することになる。そこで原語表記に対応する JPO 中韓システムの日本語表記を網羅的に検索し、抽出した。この検索結果と原語検索の検索結果とを比較し、同等の結果が得られるか否を検証した。

表 6. 網羅的検索式例

CNIPR	碳纳米管 or 纳米碳管 or 炭纳米管 or 纳米炭管 or 碳奈米管 or 奈米碳管 or 碳毫微管 or 碳纳米细管 or 碳基纳米管 or 纳米级碳管	3084件
中韓	カーボンナノチューブ / 炭素ナノチューブ / 炭ナノチューブ / ナノカーボンチューブ / カーボンナノ管 / ナノカーボン管 / 炭素ナノ管 / 炭ナノ管 / ナノ炭管 / 炭素基ナノチューブ / ナノ木炭管	3034件
KIPRIS	탄소나노튜브 + “탄소 나노튜브” + “탄소나노 튜브” + “탄소 나노 튜브” + “탄소 나노관” + 카본나노튜브 + “카본 나노튜브” + “카본나노 튜브” + “카본 나노 튜브” + “carbon nanotube”	1733件
中韓	カーボンナノチューブ カーボンナノ管 ナノカーボンチューブ ナノカーボン管 炭素ナノチューブ 炭素ナノ管 炭ナノチューブ 炭ナノ管 ナノ炭素チューブ ナノ炭素管 ナノ炭チューブ ナノ炭管	1677件

表 6 の日本語異表記は、中国語の 3084 件、ハングルの 1733 件に対応する公報番号を元に JPO 中韓システムで調べて抽出したものである。この例のように検索時に各用語についてどのような日本語が充てられているかどうか(誤訳を含めて)逐一確認することは不可能ではない(原語検索の検索結果と同等の結果が得られると考えられる)。しかし、現段階ではキーワードの抽出に相当な手間とスキルを要するため実務的でない。

5. まとめ

以上、主に JPO 中韓システムの翻訳不良部分について紹介してきたが、本システムは翻訳不良が多いため利用できない、ということを示すためではなく、良好に翻訳されている用語も多い。

従って、現段階での JPO 中韓システムの有効な利用法として以下の内容が考えられる。

1) JPO 中韓システムを利用した査読

技術内容は充分理解しているが、中国語やハングルはもちろん、英語も苦手な研究者、技術者あるいは調査担当者が日本語で査読できる、という点では素晴らしいツールであり、査読作業の効率が格段に向上するものと思われる。

JPO 中韓システムの翻訳不良部分は Google 翻訳ツールを利用、あるいは該当部分の英語翻訳を参照しながら査読することでかなりの部分をカバーできる。

それでもなお、理解しがたい請求項などについてのみ、適宜、人の手による翻訳を依頼するとよい。

2) JPO 中韓システムを利用した検索

これまで述べてきたように翻訳不良部分が少なからず存在するために JPO 中韓システムのみを利用して網羅的な検索をすることには難がある。

一方で、上記のように JPO 中韓システムは査読がしやすいというメリットがあるため、検索については CNIPR や KIPRIS などの原語データベースで網羅的検索を実施して、その公報番号を本システムの番号検索(最大 1000 件まで検索できる)で日本語表示させ、内容を査読する、という利用法が好ましいと考える。

3) システム改善に対する要望

商用の英語データベースを含めた英語キーワードだけでは網羅的な検索はできない(特に、中国特許文献の英語情

報には誤翻訳やスペルミスが多い)点については、既にいくつかの文献で紹介されている[1]。

現在、本システムの改良の作業が予定されているが[2]、翻訳不良改善のための翻訳辞書登録の充実はもちろん、可能であれば中国語やハングルの原文も収録して日本語と原語で同時検索できる(ハイブリッド検索)システム機能を加えていただければ、本システムのみで網羅的検索が可能となり、ユーザの利便性がより向上すると思われる。

また、検証部分でも触れたように、本システムの翻訳不良を解消するには単に中国語やハングルの相当する日本語翻訳辞書を充実させるだけでは難しい。

このことは、特に中国語の翻訳において同一の中国語が複数の日本語に翻訳されている例を見ても明らかであると考えられる。

通常、機械翻訳文は辞書登録された用語に従って翻訳され、同一用語の異表記(表記のゆれ)は生じないはずである。しかし、本システムでは検証例で示したように、同一の用語がいくつもの日本語に表記割れしている。これは単に辞書登録の問題ではなく、特に中国語の場合には同一の用語も文章の前後の位置、言い換えれば文法的な観点からの翻訳(品詞の種類も加味した翻訳)も考慮しないと適切な訳語が与えられず、満足な翻訳結果にならないものと考えられる。

この点については、本発表とは別に、「翻訳不良の文法的な観点からの検証」として翻訳不良の原因について考察した発表を予定しているので参照していただきたい。

また、2011年～2014年 JAPIO において中国特許文献の日本語への機械翻訳に関して精力的に研究された一連の報告書が日本特許庁から発行されてお

り、この中でも文法的な観点からの検証がなされている[3]。参考文献として本検証に有用であった。今後予定されているシステム改善にも研究結果を反映させていただきたい。

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。

6. 参考文献

[1] 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)

http://www.japio.or.jp/00yearbook/files/2014book/14_2_03.pdf

[2] 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

[3] 機械翻訳に関する調査報告書(JPO 2011～2014)

https://www.jpo.go.jp/shiryoutoushin/chousa/kikai_honyaku.htm

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)

https://www.jpo.go.jp/shiryoutoushin/chousa/pdf/kikai_honyaku/h26_01.pdf

・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)

http://www.jpo.go.jp/shiryoutoushin/chousa/pdf/kikai_honyaku/h22_01.pdf