

C22

韓国特許調査における機械翻訳クレーム検索:

機械翻訳されたクレームを
検索する際の網羅性向上の検討

○ 富士フイルム(株)

田畑文也

発表内容目次

1. はじめに
2. 韓国特許出願動向と分析
3. 韓国特許クレーム検索DB(データベース)について
4. 機械翻訳クレーム検索の精度検討方法
5. 機械翻訳クレーム検索の精度検討結果
6. まとめ

(注)本発表は、2016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 韓国特許データの流れ

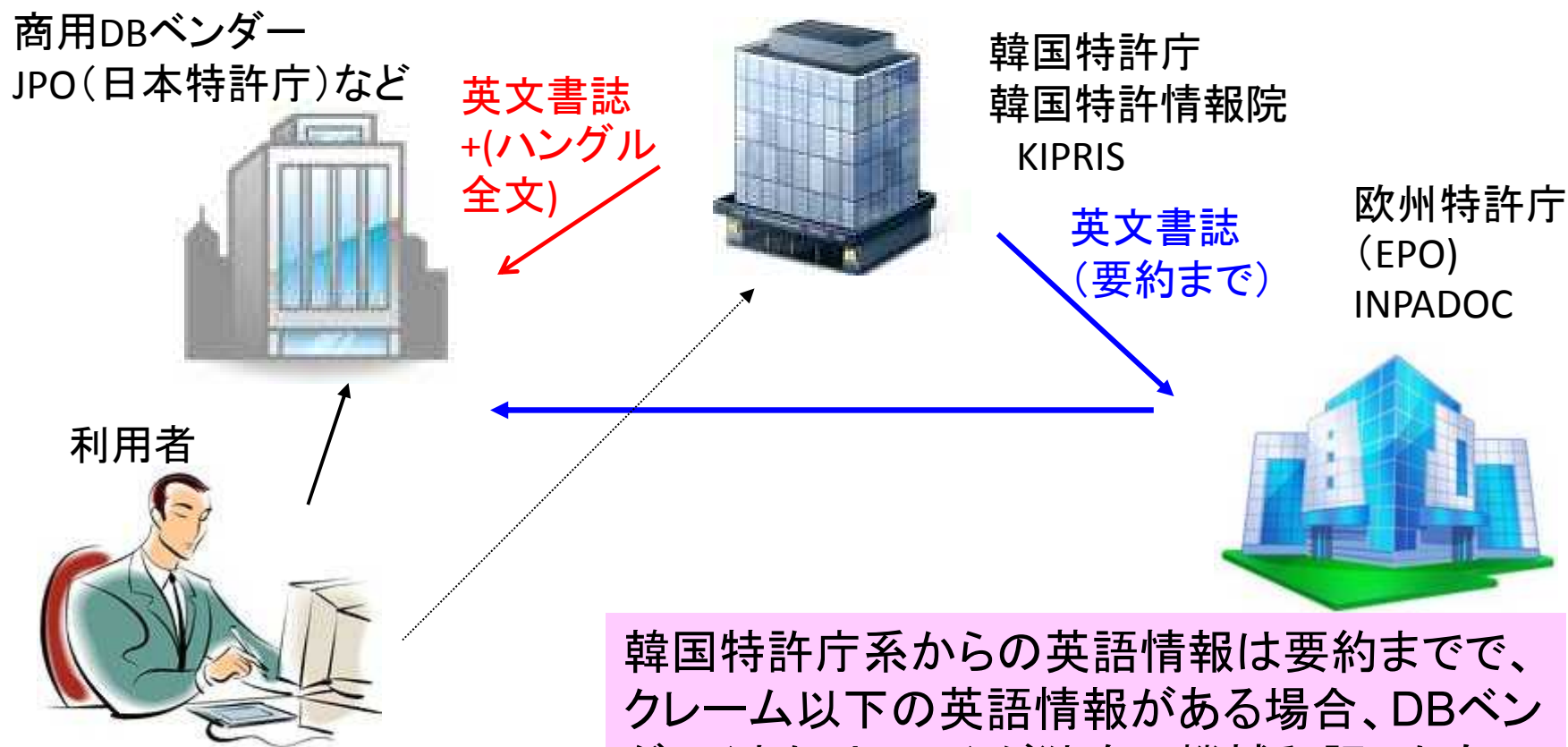


図1.
韓国特許データの流れ

1.2 韓国特許クレーム検索について

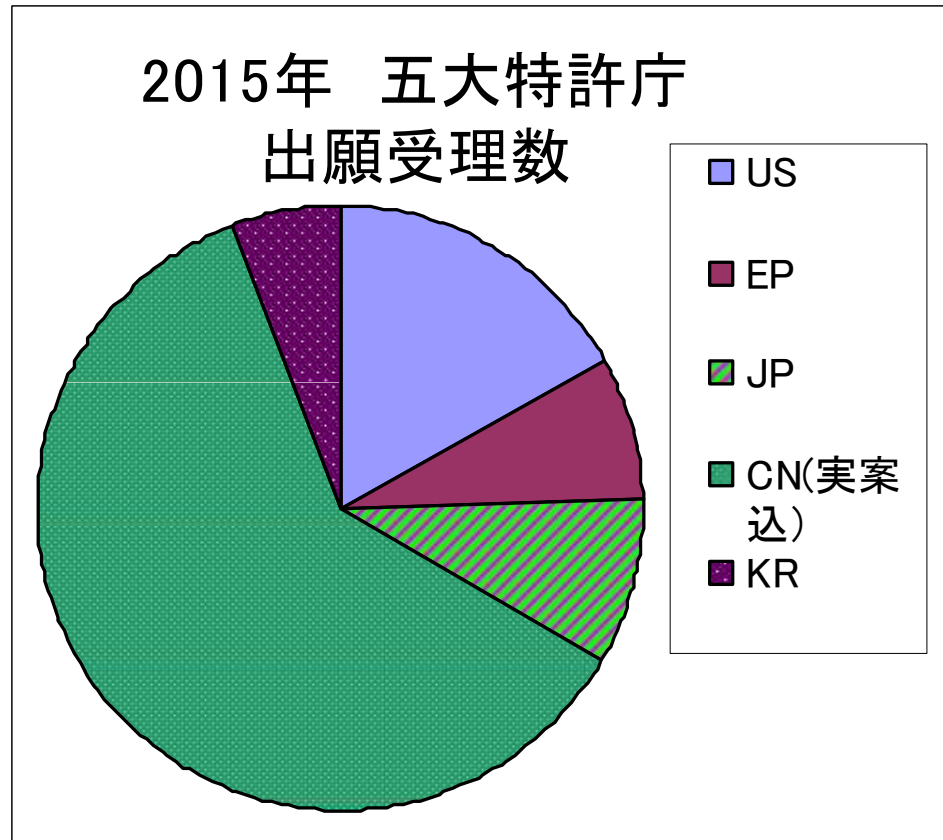
・韓国特許情報のキーワード検索環境（表1）

情報項目 /DB情報	KIPRIS(韓国特許庁系)		その他DB(英語、日本語検索)	
	ハングル	英語	EPO Espacenet	商用DBなど
要約までの書誌	○	○	○	○
クレーム	○	△(機械翻訳)	×	× または△(機械翻訳)
実施例	○	△(機械翻訳)	×	× または△(機械翻訳)

韓国特許クレームのキーワード検索について、ハングル以外は、検索できないか、できても機械翻訳(英語、または日本語)されたものしかなく、今回は機械翻訳されたクレームのキーワード検索について検討した。

2. 韓国特許出願動向と分析

2.1 2015年 五大特許庁への特許出願状況



数的には、米国、欧州以外の日中韓の非英語圏からの特許出願が3/4近くを占めている。

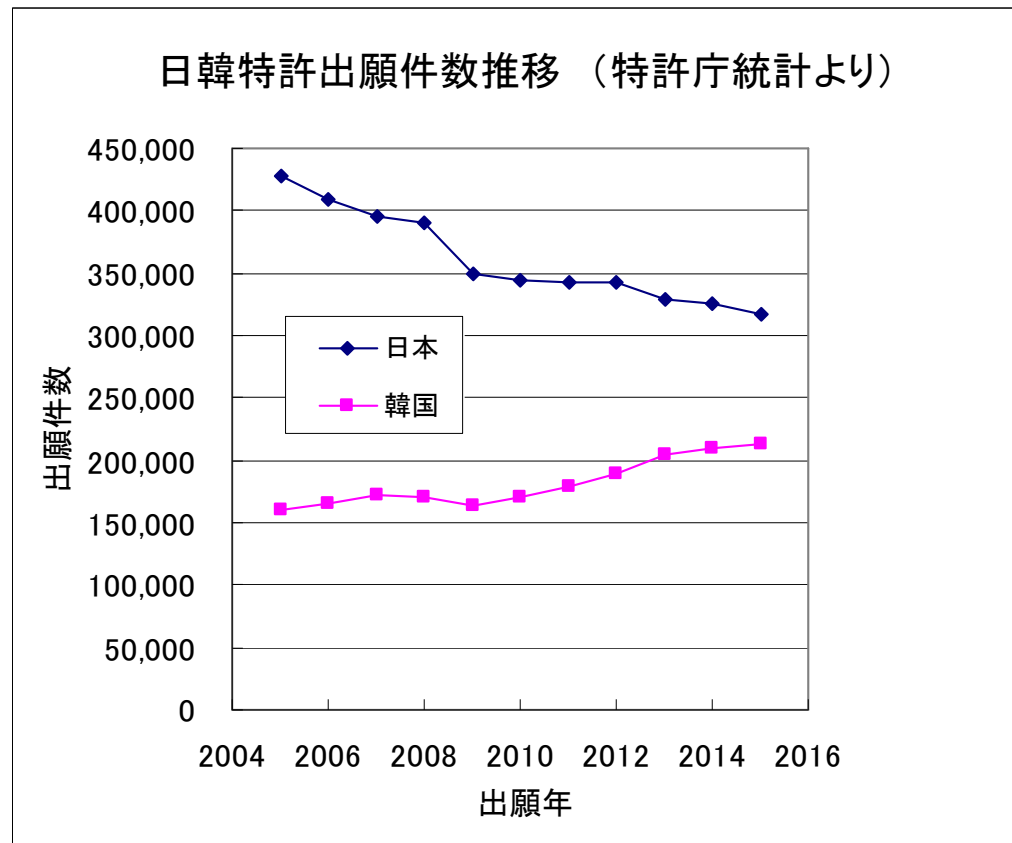


本発表では、このうちの韓国特許について検討した。

図2. 2015年 五大特許庁への特許出願受理数
(各特許庁統計を集計)

2.2 韓国特許出願推移

・韓国と日本の特許出願件数推移



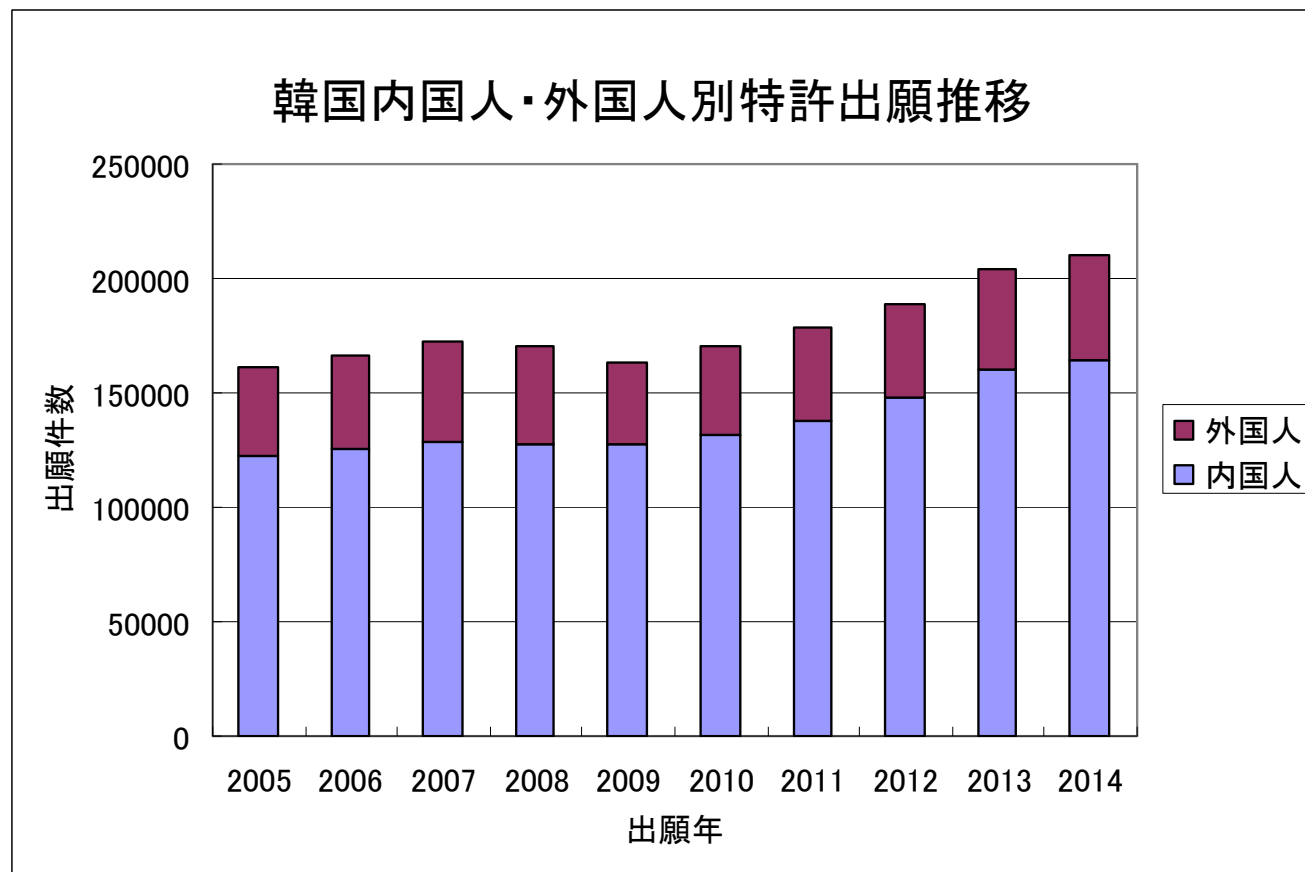
・日本特許が減少傾向に対し、韓国特許は継続的に増加している。



韓国の技術力向上と合わせ、特許的な脅威は高まっていると考える。

図3. 日韓特許出願件数推移（特許庁統計を集計）

2.3 韓国特許 内国人・外国人別出願推移



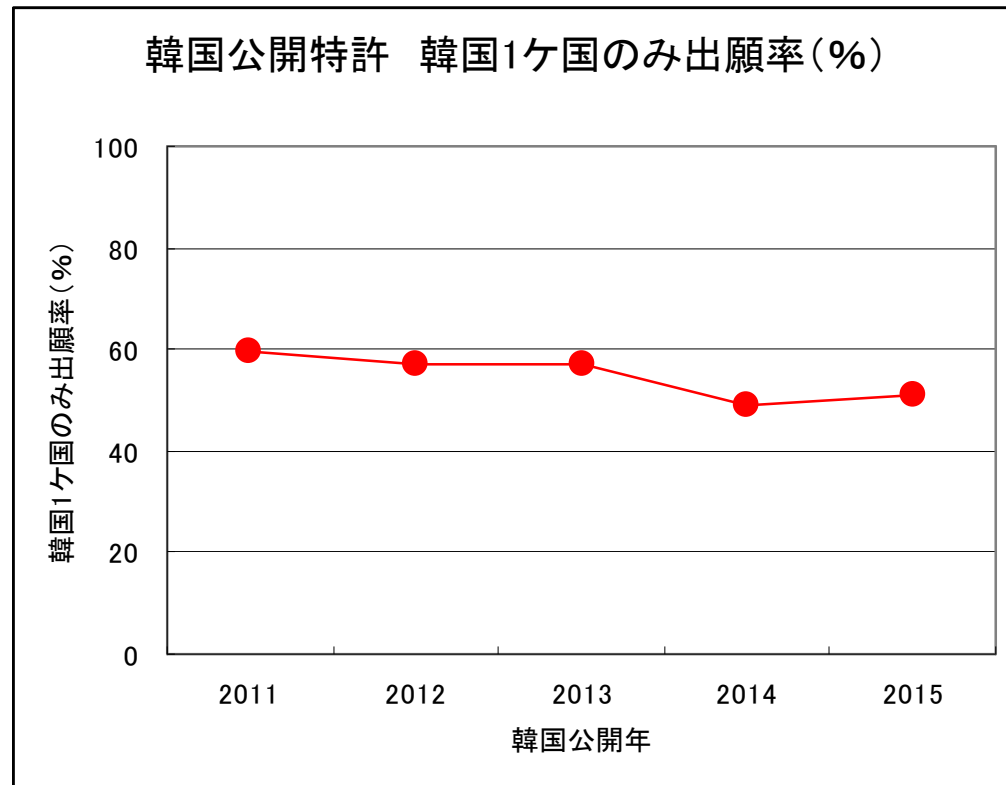
・韓国特許について、内国人出願率が高く(約8割)、かつ、数も増大。

↓
韓国特許調査においては、韓国からの現地出願人に注意する必要がある。

図4. 韓国特許内国人・外国人別出願推移 (特許庁統計を集計)

2.4 韓国特許 対応海外ファミリーの有無

- 韓国特許 韓国1ヶ国のみ出願率 (対応海外ファミリー無し)



・韓国特許については、5～6割の特許が韓国のみへの出願。



対応海外ファミリーの情報で検索補完できないため、クレーム検索については、ハングル検索または、機械翻訳されたクレームをキーワード検索するか手がない。

図5. 韓国公開特許 韓国1ヶ国のみ出願率
(2016年7月STN WPIにて確認)

3. 韓国特許クレーム検索DBについて

3.1 ハングル検索か、機械翻訳検索か？

韓国特許クレーム検索 メリット・デメリット対比(表2)

	ハングル検索	機械翻訳検索
メリット	網羅性の向上。 (機械翻訳における誤訳問題が回避可能)	工数負荷が少ない。 (日本語または英語でそのまま検索可能)
デメリット	言語的に日本人サーチャーの負荷が激増。 工数増大(検索結果マージ工数を含む)。	機械翻訳における誤訳問題を回避できない。 網羅性がどこまで担保できているか不明。

↓
できる限り回避したい。
(本音)

↓
今回の検討内容

3.2 韓国特許 ハングル検索DB例

韓国特許 ハングル(キーワード)検索可能なDB例(表3)

DB名	提供元	備考
KIPRIS	韓国特許庁系	無料
PatBridge	韓国WIPS	
PatBase	RWSグループ	
Orbit.com	Questel	
TotalPatent	LexisNexis	
KRFULL	STN(Questel)	

など他多数

今回の検討では、ハングル検索のレファレンスとして、韓国WIPS社のPatBridgeを用いた。
(注:日本語表示、日本語検索も可能)

今回、ハングル検索に用いたDB



<http://www.patbridge.com/>

KR 1416252 B1

(54) ポリエチレンイミン系顔料分散剤
Polyethylene imine based pigment dispersants
폴리에틸렌이민계 안료 분산제

多段表示 検索式ハイライト エチレンイミン(8) 에틸렌이민(8)

請求項1項

化学式 Iのポリエチレンイミン(PEI)系顔料分散剤(pigment dispersant).

化学式 I

$$X-(T)_m-P-(T)_n-H$$

上の化学式 Iで、Pはポリエチレンイミン(PEI)主鎖(backbone)であり、Tは残基 $-CO-A-O-$ でありここで、Aは置換されないか

청구항 1항

화학식 I의 폴리에틸렌이민(P dispersant).

화학식 I

$$X-(T)_m-P-(T)_n-H$$

위의 화학식 I에서, P는 폴리에틸렌이민(PEI) 주

図6. PatBridge検索結果表示例

3.3 機械翻訳クレーム検索可能DB例

韓国特許 機械翻訳クレーム検索可能DB例(表4)

日本語でクレーム検索可能	提供
中韓文献翻訳・検索システム	JPO
PatBridge	WIPS
Japio-GPG/FX	Japio

など多数

英語でクレーム検索可能	提供
Thomson Innovation	Clarivate Analytics
Shareresearch	日立総合特許情報システム
PatentSQUARE	パナソニック ソリューションテクノロジー
PatBase	RWSグループ
Orbit.com	Questel
TotalPatent	LexisNexis
KIPRIS	韓国特許庁系

など多数

今回、青字のDBを用いて検証した。

基本的に機械翻訳されたクレームを検索するが、ファミリー情報で補完される場合有り。

(注)収録範囲に要注意。要オプション契約等もあり。

3.4 JPO 中韓文献翻訳・検索システム

http://www.ckgs.jpo.go.jp/full_text

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 [] ~ []

フリー検索条件 ? NOT検索

項番	項目間接続 ?	検索項目 ?	検索キーワード ?	項目内接続 ?	項番	項目間接続
<input checked="" type="checkbox"/>	1	要約+請求の範囲		OR	<input checked="" type="checkbox"/>	1
<input checked="" type="checkbox"/>	2	AND	公報全文(書誌を除く)	OR	<input checked="" type="checkbox"/>	2

スクリーニング

< 前の文献 KRB1-101654845 次の文献 > [] n件目の文献

キーワード反転 ジャンプ

【請求項1】

【請求の状態】 Currently amended

ポリエチレン40～60重量%；エラストマ8～15重量%；膨張黒鉛23～30重量%；酸化亜鉛4～7重量%；二酸化チタン4～7重量%；及び引き継ぎ難燃剤、水酸化アルミニウム及び臭素系難燃剤の中で選択されたいずれか一つの難燃剤1～2重量%に架橋されていることであり、前記膨張ポリエチレンイミン樹脂はポリエチレンイミン樹脂100重量部に対して硬化触媒10～30重量部を使用したことであり、前記膨張黒鉛は密度が1.5～2.3g

- ・JPO(日本特許庁)提供の無料で使えるツールである。
- ・日本語で検索及び全文表示可。
- ・収録範囲が2003年以降発行分のみであること、収録漏れの発生が確認されていることに注意が必要。

図7. JPO 中韓文献翻訳・検索システム

4. 機械翻訳クレーン検索の精度検証方法

4.1 検証に用いたキーワード

検証に用いたキーワード(表5)

キーワード概念 (日本語)	対応キーワード例(ハングル)
層間絶縁(膜)	층간절연(分かち書き無し) 층간 절연(分かち書き有)
ポリエチレンイミン	폴리에틸렌이민 폴리에칠렌이민(4文字目が異なる)
赤外吸収	적외흡수(赤外吸収) 적외선흡수(赤外線吸収) 적외선광흡수(赤外線光吸収)
4級アンモニウム	4급암모늄(4級アンモニウム) 사급암모늄(四級アンモニウム) 4차암모늄(4次アンモニウム)

表5の4系統のキーワードで、出現頻度高い表現についてのみ、検討を実施した。
(対象は2013年～2015年の韓国登録特許で実施した。)

4.2 ハングル検索例 ハングルクレームの検索式例

使用DB: PatBridge

ハングルクレーム検索式例(層間絶縁(膜)の例)

(注:赤字は日本語解説)

(층간절연막* (層間絶縁膜) or

(층(層) adj2 간(間) adj2 절연(絶縁) adj2 막*(膜))

or (층간(層間) adj2 절연막*(絶縁膜))

or (층간(層間) adj2 절연(絶縁) adj2 막*(膜))).CLA. AND
(@FD>=20130101<=20151231)

4.3 正解母集団作成スキーム

ファミリーないものに限定 & 目視でノイズ除去

PatBridge ハングルキーワード検索
(対象:
2013年~2015年登録特許クレーム)

↓ 特許番号移行

Thomson Innovationで
海外ファミリーが無いものを抽出

↓ 特許番号移行

PatBridge(ハングル+日本語表示)にて、
目視でノイズ除去

→これを正解母集団とした。

各DBを公平に評価するため、ハングルを用いて検索後、

- ・海外ファミリーのないもののみ抽出
- ・ノイズを目視で除去

することにより、正解母集団を作成し、再現率(得られるべき正解母集団に対する網羅性)を算出した。

正解母集団公報数(表6)

キーワード概念	検証正解公報数
層間絶縁(膜)	100
ポリエチレンイミン	156
赤外吸収	69
4級アンモニウム	125

図8. 正解母集団作成スキーム

4.4 再現率検証DBと検索式

機械翻訳精度検証に用いたDB(表7)

DB	提供元	検索に用いた用語
中韓文献翻訳 ・検索システム	JPO	日本語
PatBridge	WIPS	日本語
Thomson Innovation	Clarivate Analytics	英語
Shareresearch	日立総合特許情報システム	英語

表7に示す4種のDBを用いて、日本語または、英語に機械翻訳されたクレームを、表8に示す検索式例で検索し、再現率を算出した。

機械翻訳クレーム
検索式例(表8)

機械翻訳を想定して、想定できる範囲の広めのキーワードで検索を実施した。

キーワード概念 (日本語)	日本語検索式例	英語検索式例
層間絶縁(膜)	層間絶縁 層間誘電	((interlayer*)+(interlevel*+(intermedia*)) adj3 ((isolat*)+(insulat*+(dielect*))) (inter adj3 (layer+level+media)))*(inter adj5 ((isolat*)+(insulat*+(dielect*))))
ポリエチレンイミン	ポリエチレンイミン ポリ adj3 エチレンイミン ポリエチレン adj3 イミン (ポリ adj3 エチレン adj3 イミン)	(Polyethyleneimine*)+PEI Poly adj3 ethyleneimine* Polyethylene adj3 imine* Poly adj3 ethylene adj3 imine*
赤外吸収	(赤外 or IR or IR) near15 (吸収 or 吸光)	(IR+infrared) near3 ((abso*)+(filter*))
4級アンモニウム	((4級 or 4級 or 四級) adj10 アンモニウム	Quaternary* near3 ammonium*

4.5 検索でヒットしなかった原因カテゴリ

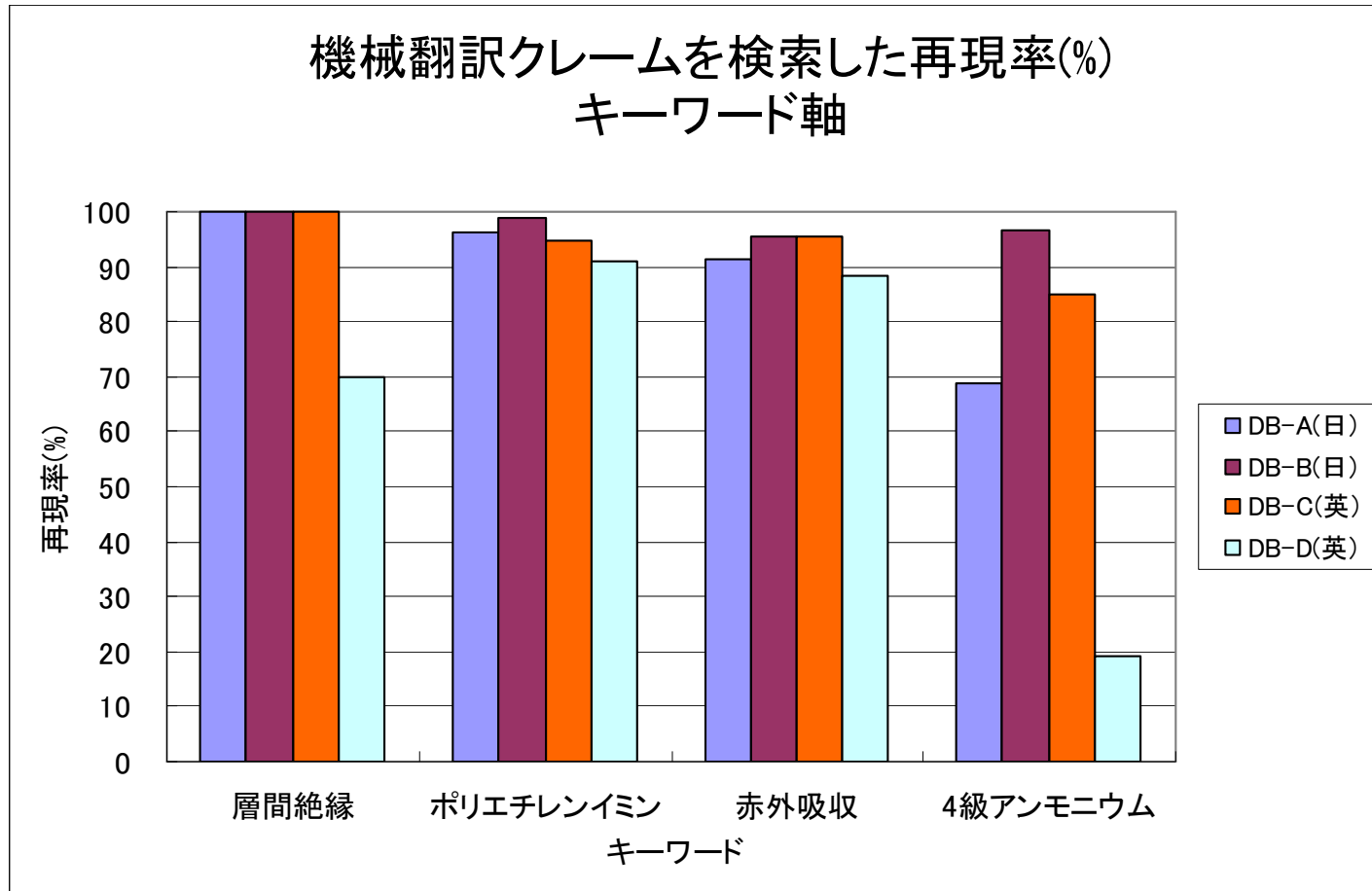
ヒットできない原因カテゴリ定義(表9)

カテゴリ	定義	例
対処可能訳	予備検索を実施し、ハングル表記から、機械翻訳のくせを考慮すれば、なんとか対処可能なレベルのもの。(通常 of 検索を超えた対応レベル)	4級アンモニウム →4次アンモニウム
対処不能訳	機械翻訳されたものが、想像し得ない訳になっている。または、機械翻訳されずに訳が消えるなど。	層間絶縁 → getozoruyonmakku (英語にない発音表記)
DB収録など	DBにデータが収録されていない	
異義語	対応するハングル表記が、異なる意味を持ち、全く別の単語として約される。	吸収→吸水 (ハングルでは同表記)
その他	上記以外のもの	

本検討では、表9に示すように、ヒットしなかったものをカテゴリ分けして、原因分類した。

5.機械翻訳クレーム検索の精度検討結果

5.1 キーワード軸の再現率結果

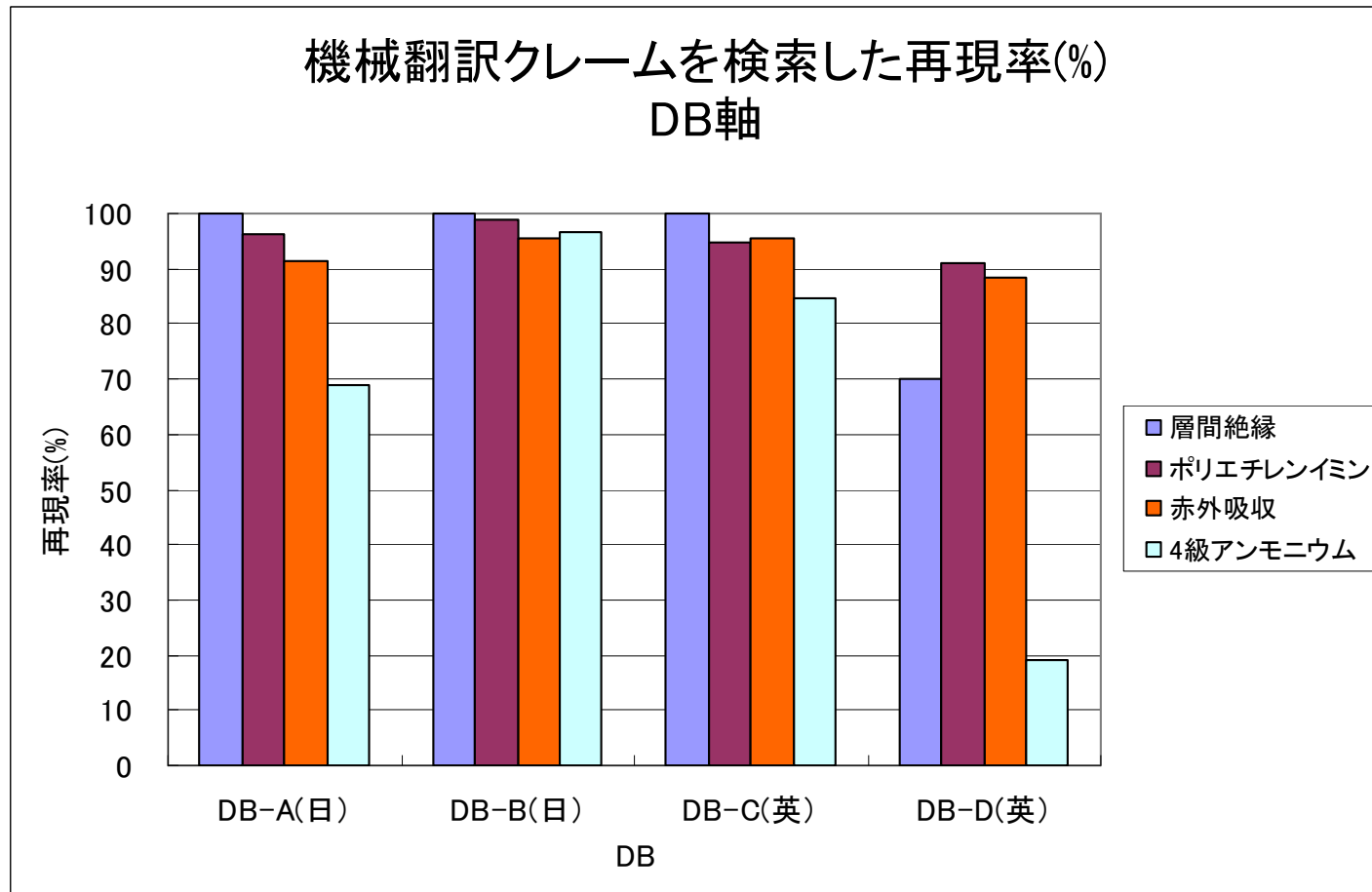


キーワード、DB毎に再現率が異なり、9割以上のものもあるが、2割未満の場合も有る。

(注: DB名は明示せず)

図9. 機械翻訳クレーム検索再現率 (キーワード軸)

5.2 DB軸のクレーム検索再現率結果



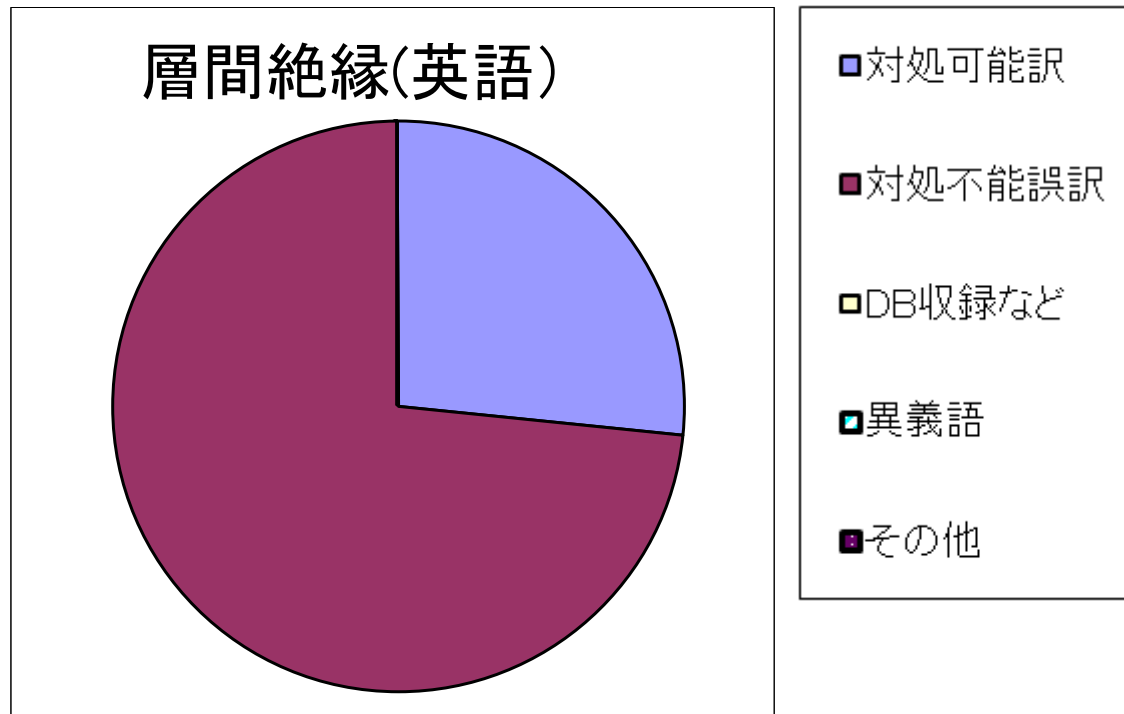
・日本語への機械翻訳DBの方が、英語へのものよりやや良好な傾向ある。
・DB-Bが比較的良好な結果を示しているが、全体として、侵害予防調査時に要求されるレベルの網羅性は有していないと考える。

(注: DB名は明示せず)

図10. 機械翻訳クレーム検索再現率 (DB軸)

5.3 ヒットしない原因解析1

層間絶縁



予備検索で機械翻訳の癖を掴めば対応可能なものとして、“between a floor”などがあるが、“getozoruyonmakku”など、ハングル表記の発音からの当て字に訳されるような、対処不能誤訳が約7割あった。(英語)

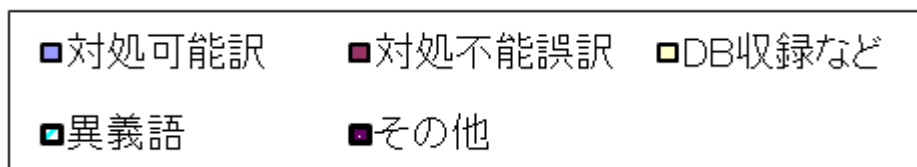
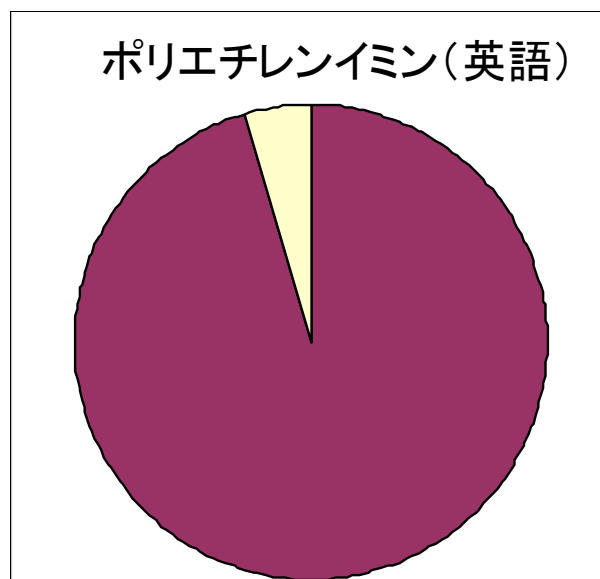
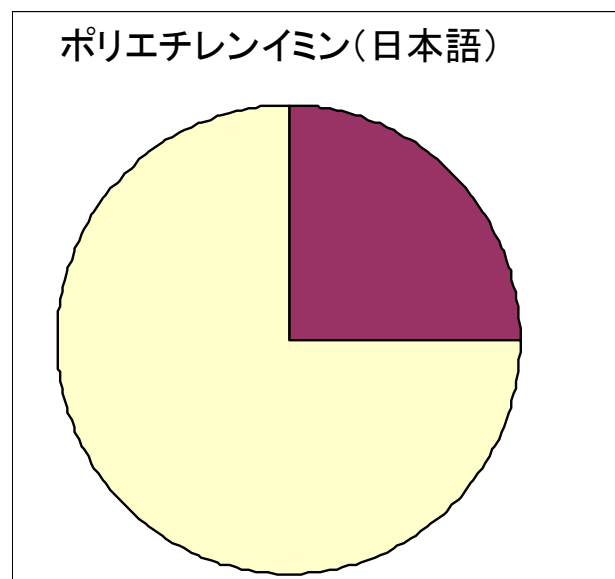
→機械翻訳特有の異表記を考慮することが必要だが限界もある。

図11. 層間絶縁 ヒットしない原因

(注)

日本語検索の再現性は良好な為、解析を割愛した。

5.4 ヒットしない原因解析2 ポリエチレンイミン



・日本語
DB収録漏れが原因の7割以上。”ポ
ルリエチルレンイ
ミン“と対処不能な表
記も有り。

・英語
”polyethylenemine“
(イミンの”i“が抜け
ている)のような対
処不能なミススペル
がほとんどであった。
→発音を英語的に
表記化？

図12. ポリエチレンイミン ヒットしない原因

(注)各言語への機械翻訳
DB2つずつの結果をマー
ジして表示した。

5.5 ヒットしない原因解析3 赤外吸収

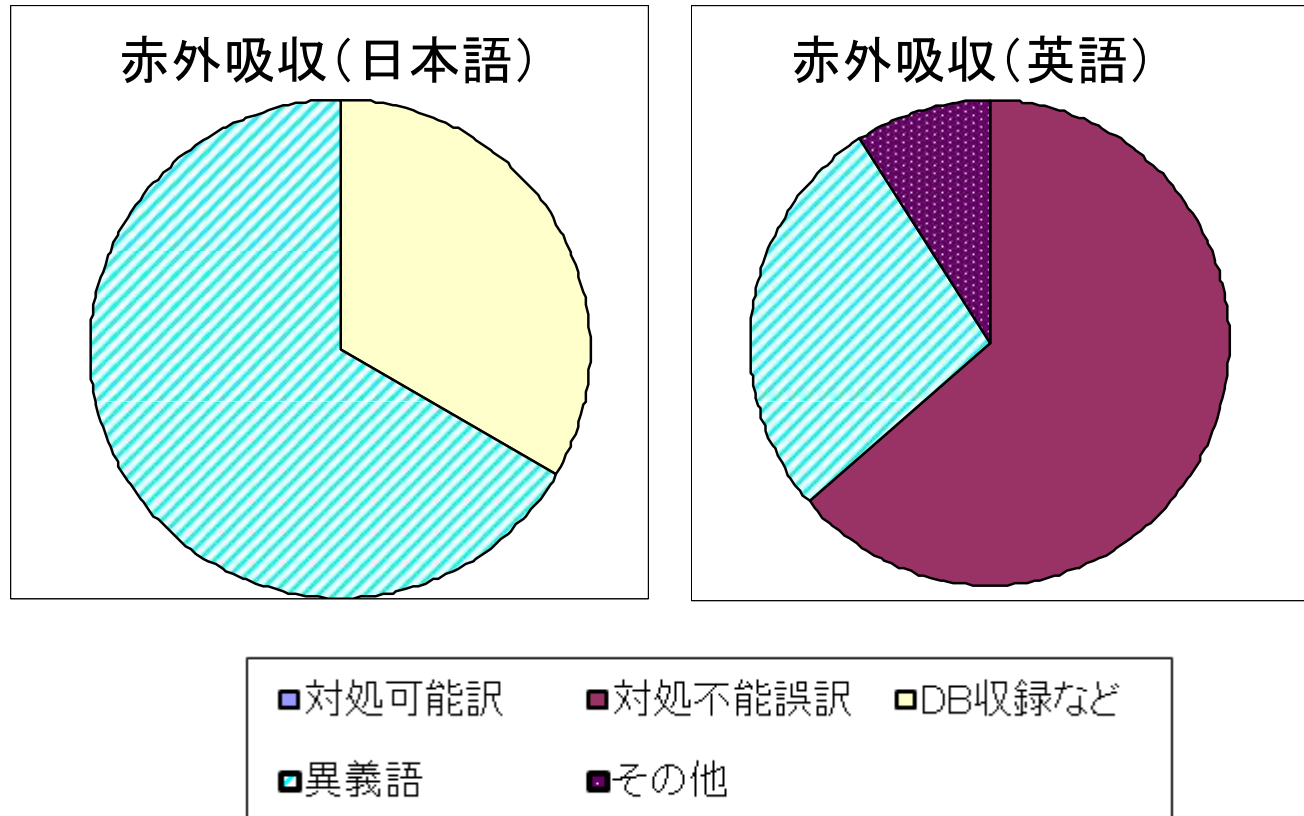


図13. 赤外吸収 ヒットしない原因

・日本語
ハンゲルの吸収にあたる”흡수“が、吸水と言う(同表記)異義語を持つため、吸水に訳されるものが、漏れのうち約2/3。残り1/3は収録漏れの問題。

・英語
(うまく訳せず)訳が消えたり、赤外線が”far-red light”などと、対処不能な誤訳をされるものが、漏れの6割以上であった。

5.6 ハングル同表記異義語の例

ハングル同表記異義語例(表10)

ハングル	日本語
산	酸 山
충전	充電 充填
발전	発電 発展
진통	陣痛 鎮痛

吸収にあたるハングル”
흡수“が、吸水という(同
表記)異義語を持つ。他
にも、表10のように、同じ
ハングル表記が複数の
意味を持つ場合あり、誤
訳が生じる原因になる。



漢字を持たないハングル
は、機械翻訳において、
同表記異義語が多く、誤
訳が生じやすい。対応の
ハングル表記とその意味
を調べることも重要と考
える。

5.7 ヒットしない原因解析4 4級アンモニウム

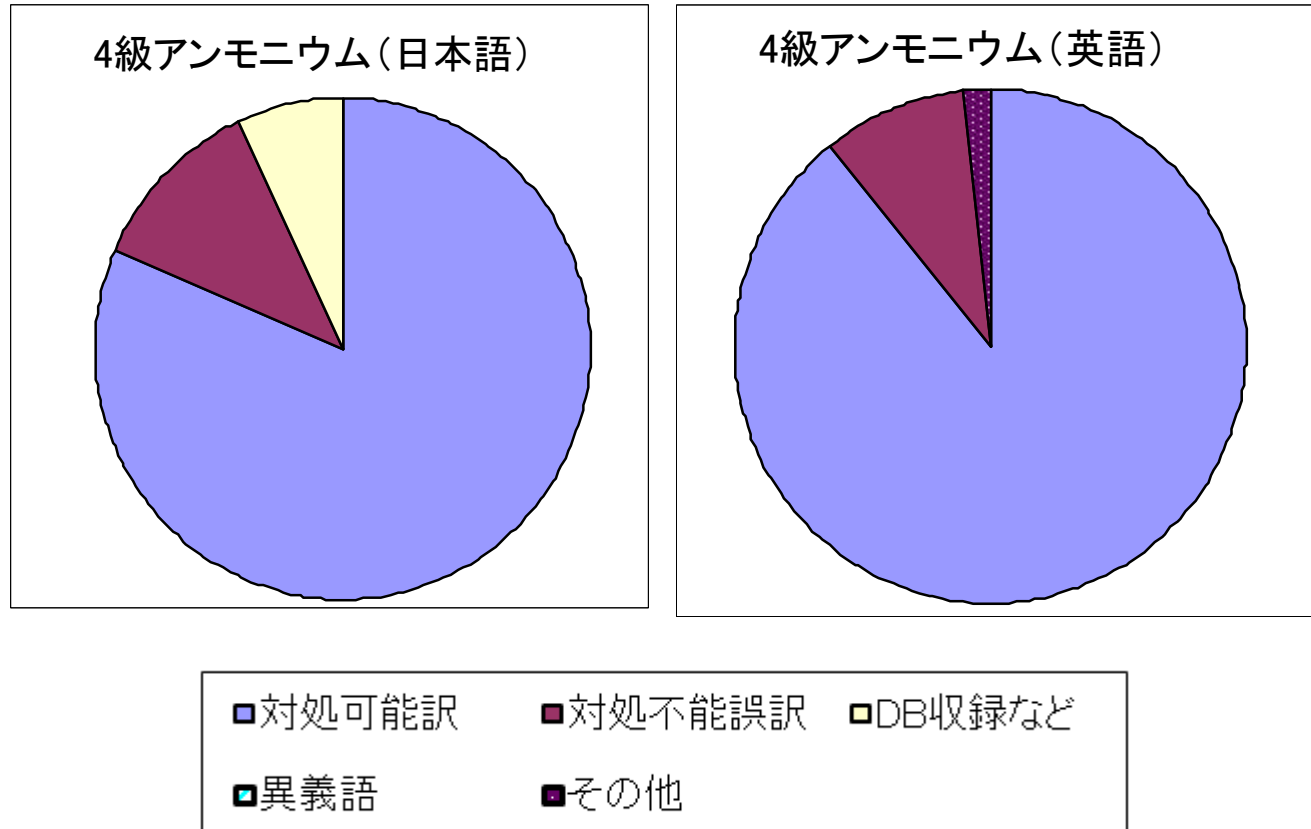


図14. 4級アンモニウム ヒットしない原因

・ 4級アンモニウム (quaternary ammonium) について、ハンゲルでは、大きく分けて “4급 암모늄” (4級アンモニウム) と、”4차 암모늄” (4次アンモニウム)、つまり “級” または “次” の2系統で表記される。この為、“次” でハンゲル表記されたものは、日本語では、4次アンモニウム、英語では ”forth ammonium” のように機械翻訳されるものが多かった。(対処可能訳に分類)。

5.8 4級アンモニウムの誤訳原因

4級アンモニウムの4次アンモニウムへの
誤訳原因 模式表(表11)

キーワード	対応ハングル例	対応ハングルからの 機械翻訳例
4級アンモニウム quaternary ammonium	4급 암모늄 (4級アンモニウム)	4級アンモニウム quaternary ammonium
	4차 암모늄 (4次アンモニウム)	4次アンモニウム forth ammonium

言語が単純な一対一対応でないため、4級アンモニウムのハングル表記を調べ、ハングル検索して、どのように機械翻訳されるか等で調べないと、漏れたことすら気づかない可能性有る。

(注)4級アンモニウムのハングル表記には分かち書きを含め、表10以外の表記も有る。

5.9 “級”のつくキーワードの誤訳頻度確認

2級アルコール、3級アミンについて、
JPO中韓文献翻訳・検索システムにて韓国特許クレーム検索

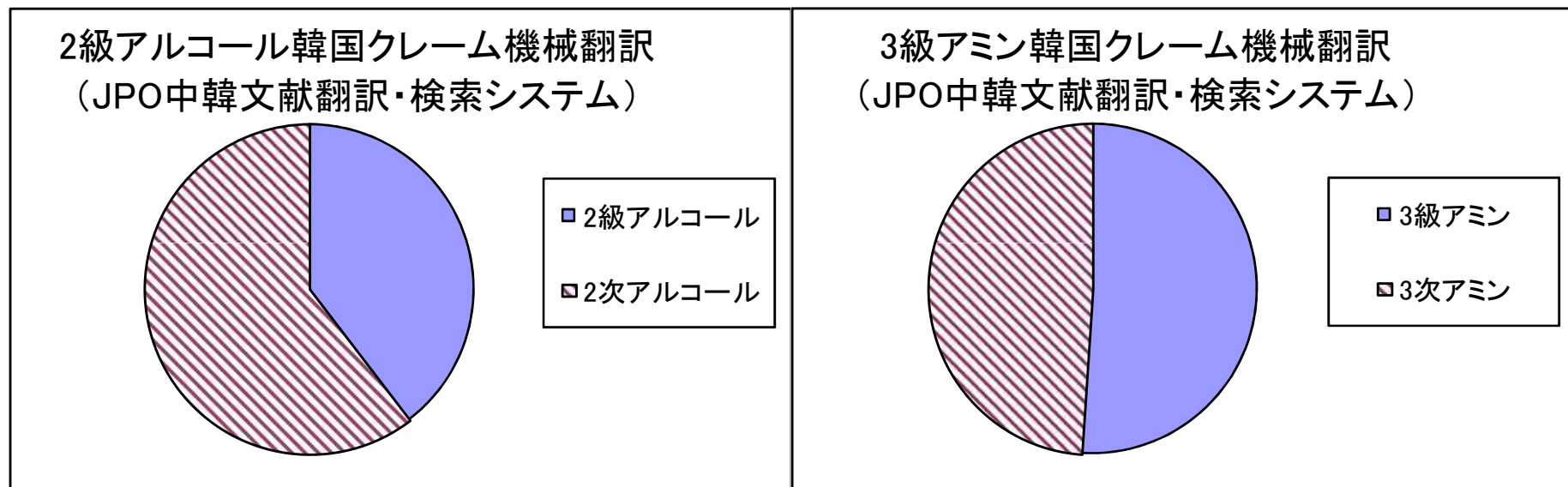


図15. 2級アルコール、3級アミン クレーム機械翻訳 (2016.11.21確認)
(JPO 中韓文献翻訳・検索システム 韓国特許)

JPO 中韓文献翻訳・検索システムに於いて、4級アンモニウムの例だけでなく、2級アルコールや、3級アミンについても、約半数以上が、“級”ではなく、“次”と訳されることがわかった。

5.10 機械翻訳検索で網羅性を上げるには？

キーワードの**ハングル表記**を調べる。



ハングル検索で、**どのように機械翻訳されるか**を調べる。



機械翻訳された結果を、検索式にフィードバックし、再度検索し、網羅性を向上させる。

→機械翻訳の表記ゆれを完全に網羅することは、現実的には不可能と考える。侵害予防調査などで、現地出願人の明確な脅威がある場合は、右のスキームも併用する。

図16. 網羅性向上スキーム

研究部門や、営業からの情報、予備検索などより、**脅威のある競合現地出願人**を把握する。



脅威のある競合現地出願人に対しては、**広めの特許分類(IPC,CPC)及び、広めのキーワードで検索補完する。**

→脅威のある競合現地出願人に対するリスクを最小限化することで、実質的な網羅性を向上させることができると考える。さらに、網羅性が必要な場合は、全体の検索において、ハングル検索も併用する。

図17. 競合現地出願人
検索補完スキーム

6.まとめ

・機械翻訳された韓国特許クレームについて、キーワード検索の網羅性を上げるには、

(A)機械翻訳特有の異表記を考慮する だけでなく、

(B)用いるキーワードのハングル表記を調べ、それがどのよう

に機械翻訳されるかを調べ、検索式にフィードバックする

ことにより、網羅性を向上できるが、限界あると考える。

・侵害予防調査などで、現地競合出願人の脅威がある場合、

(C) 競合現地出願人について、広めの特許分類、キーワードで

検索補完する

ことで、実質的な網羅性を向上させることができると考える。

ただし、それ以上の網羅性が必要な場合は、

(D)ハングルによる検索も併用する ことも必要と考える。

ご清聴、ありがとうございました。

(끝까지 경청해 주셔서 감사합니다.)

本発表は、2016年度 アジア特許情報研究会のWG活動の一環で、メンバー各位にも、アドバイス頂きました。(感謝!)



補充資料1 辞書ツール例

Naver辞書 <http://dic.naver.com/>

The screenshot shows the Naver dictionary interface. At the top, there is a green navigation bar with links for 'NAVER 사전', '번역기', '참여번역', and '지식백과'. Below this, a horizontal menu lists various languages: '사전 홈', '영어', '국어', '한자', '일본어', '중국어 U', '프랑스어', '스페인어', and '독일어'. The '영어' (English) option is highlighted. In the search section, the text '어학사전' (Language Dictionary) is followed by a search input field containing 'Polyethyleneimine' and a '검색' (Search) button. Below the search results, there are two sections: '영어사전 단어·속어 1-1 / 1건' (English Dictionary Word·Slang 1-1 / 1 item) and '일본어사전 단어 1-1 / 1건' (Japanese Dictionary Word 1-1 / 1 item). The English section shows the word 'polyethyleneimine' with its Korean translation '폴리에틸렌이민'. The Japanese section shows 'ポリエチレンミン (polyethyleneimine)' with its Korean translation '[명사] [생명공학] 폴리에틸렌이민'.

補 図1. Naver辞書

補充資料2

KIPRISを用いたハンゲル表記確認

http://eng.kipris.or.kr/enghome/main.jsp

The screenshot shows the KIPRIS patent search interface. At the top, there are navigation tabs for 'SEARCH', 'TODAY KIPRIS', 'PR', 'GUIDE', and 'KIPRIS'. Below these are sub-tabs for 'Patent', 'Design', 'Trademark', and 'KPA'. The search bar contains the query 'TL=[Polyethyleneimine]' and a 'Spread' button. The search history shows 'TL=[Polyethylen... RN=[KR] AP=[korea] IOPD=[19850101~... PD=[20160101~20...'. The main content area displays search results for 'Patent' with 16 articles found. The first three results are highlighted:

- Registered [1] Glycol Chitosan-Methyl Acrylate-Polyethyleneimine nanoparticle for Gene Delivery(유전자 전달을 위한 GMP 나노입자)**
IPC: A61K 9/16 A61K 47/48 Applicant: The Industry & Academ...
Application No.: 1020140009387 Application Date: 2014.01.23
Registration No.: 1014844410000 Registration Date: 2015.01.13
Unex. Pub. No.: Unex. Pub. Date:
Agent: WIE Byoung Gap Inventor: CHOI, Joon-Sig | LEE, Y...
- Registered [2] Gene delivery system comprising branched polyethyleneimine and uses thereof(가지형 폴리에틸렌이민 및 산화그래핀을 포함하는 유전자 전달체, 및 이의 이용)**
IPC: A61K 47/48 A61K 9/16 Applicant: POSTECH ACADEMY-IN...
Application No.: 1020120019928 Application Date: 2012.02.27
Registration No.: 1013681700000 Registration Date: 2014.02.20
Unex. Pub. No.: 1020130098090 Unex. Pub. Date: 2013.03.04
Agent: ERUUM PATENT&LAW FIRM Inventor: KIM, Won Jong | KIM, Hy...
- Registered [3] NOVEL FABRICATION OF POLYEHTYLENEIMINE-CAPPED Au-Ag ALLOY NANOPARTICLES(폴리에틸렌이민을 이용한 고농도 금-은 합금 나노입자 제조방법)**
IPC: B82B 3/00 B82B 1/00 Applicant: Foundation of Soongsil...

補 図2. KIPRIS検索