

第 6 回特許情報シンポジウム

論文資料集

2021年2月26日

オンライン開催

一般社団法人アジア太平洋機械翻訳協会

一般財団法人日本特許情報機構

特許庁における機械翻訳の活用

令和3年2月26日

特許庁 総務部 総務課

情報技術統括室 室長補佐（特許情報利用推進班長）

中西 聡



1 はじめに

1-1 翻訳した特許情報の活用

1-2 基礎資料と関連施策

2 特許庁の機械翻訳に関する取組み

2-1 機械翻訳文の対外提供、機械翻訳事業

2-2 関連事業

3 海外の動向

WIPO/EPO

4 おわりに

1. はじめに

1-1 翻訳した特許情報の活用

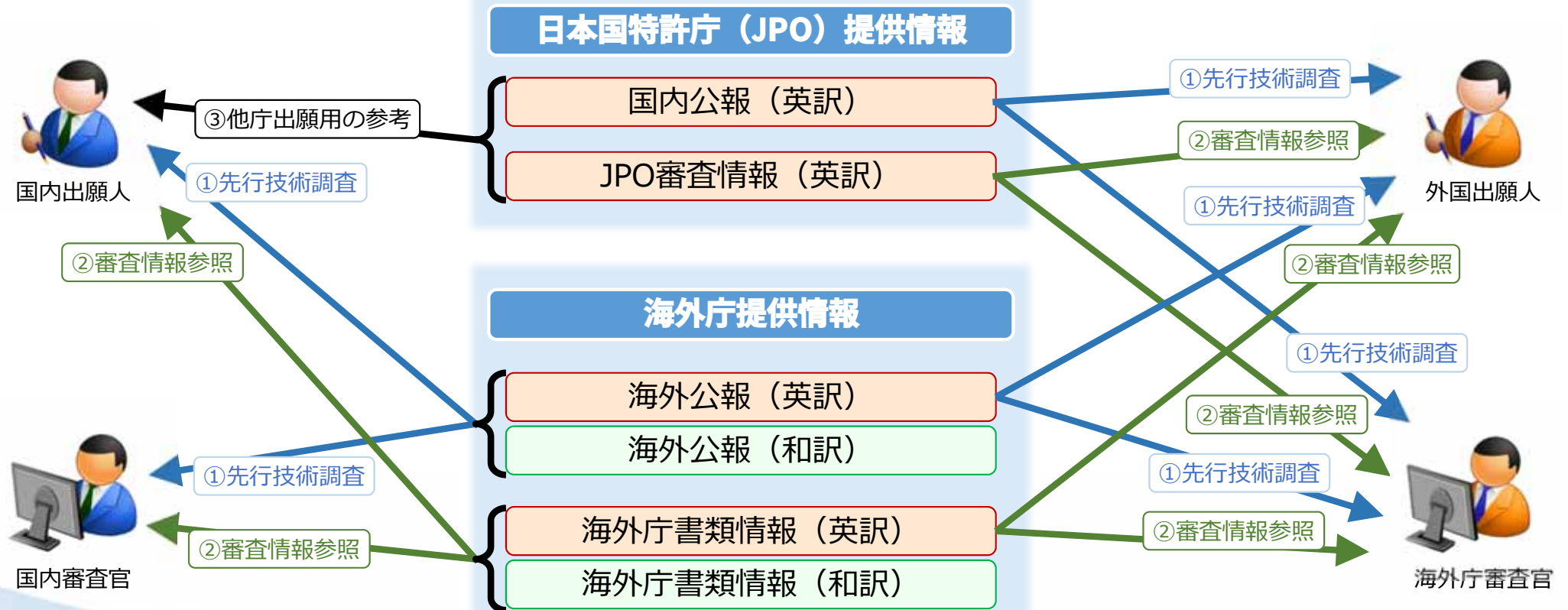
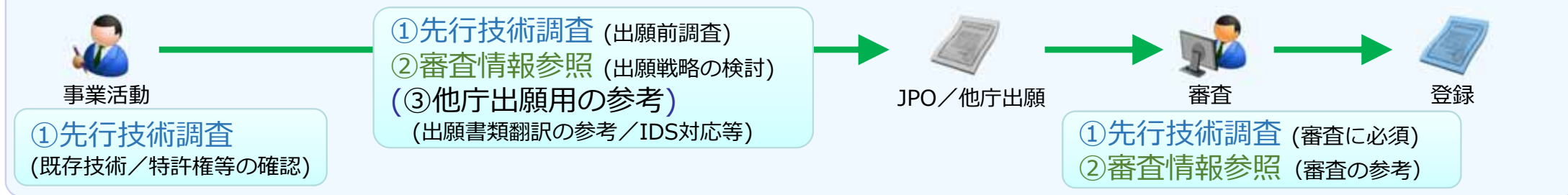
1-2 基礎資料と関連施策

1-1 翻訳した特許情報の活用

翻訳した特許情報の活用

・全体図

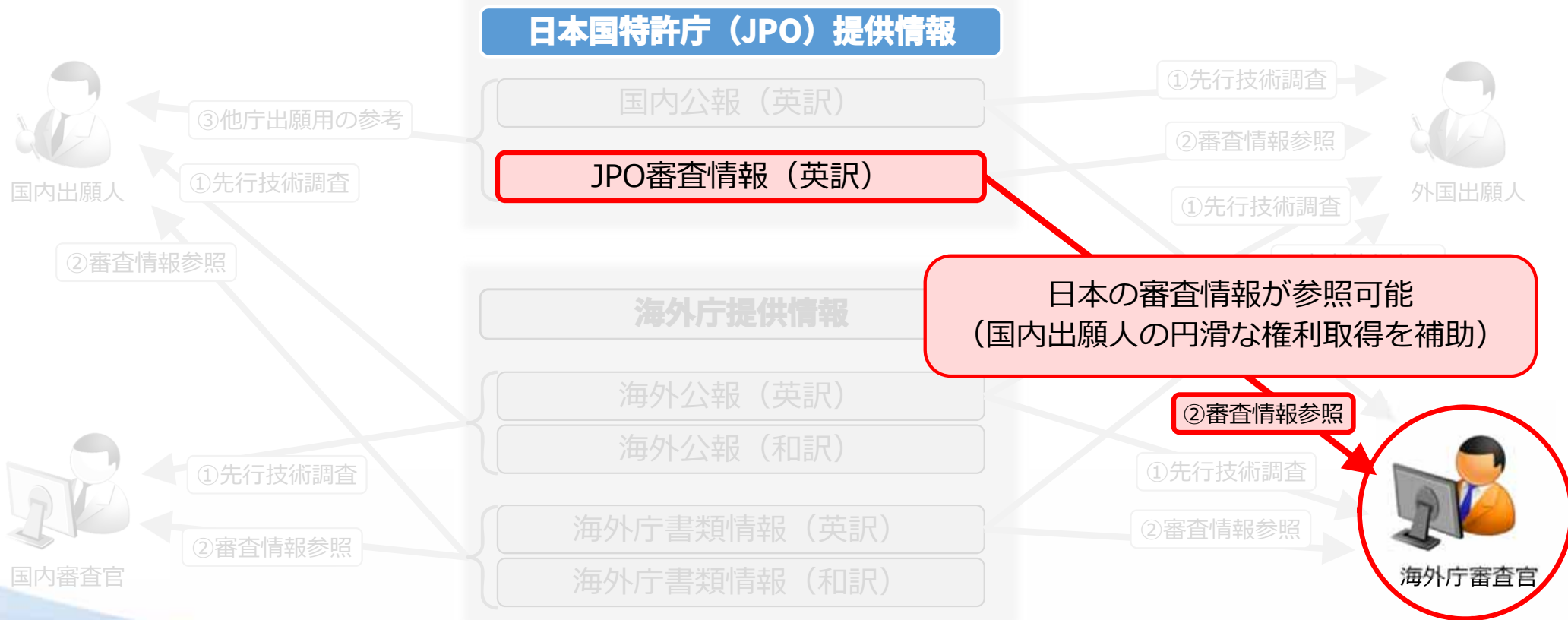
参照する特許情報の種類と時期



翻訳した特許情報の活用

・具体例①（日本の審査情報の海外発信）

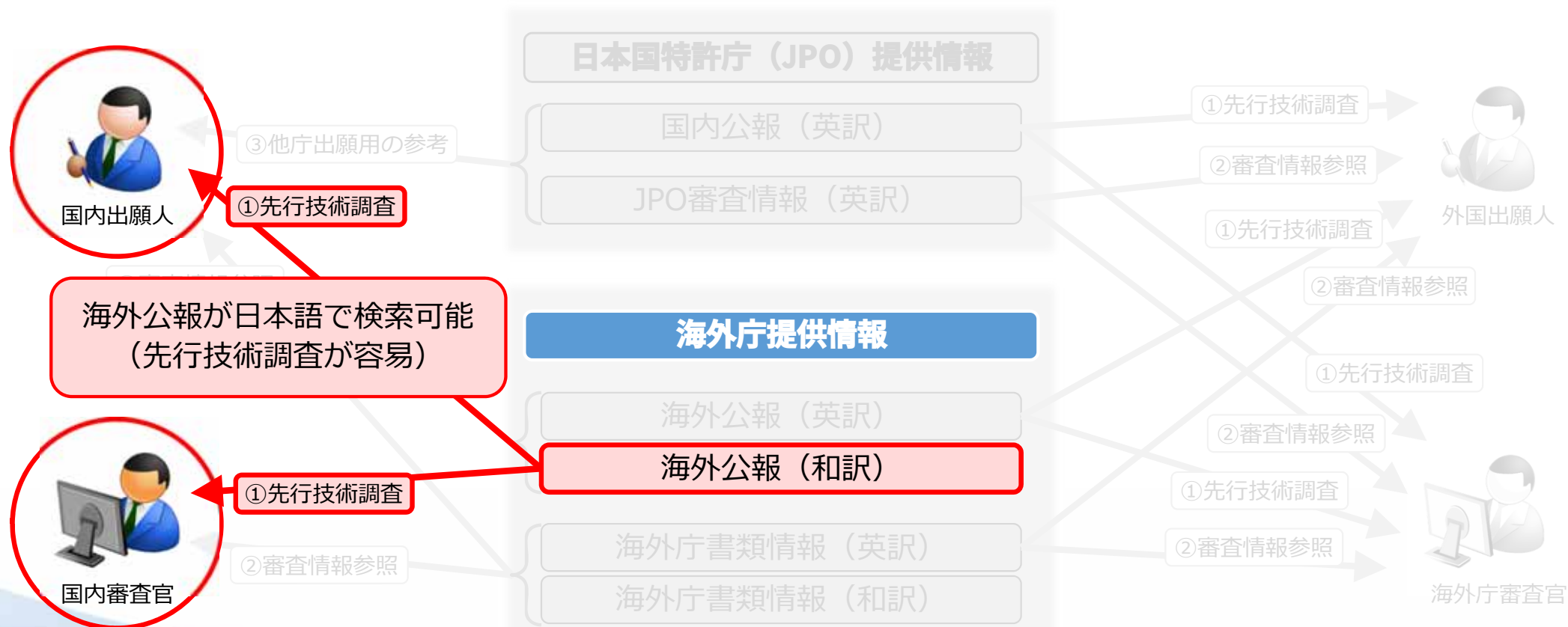
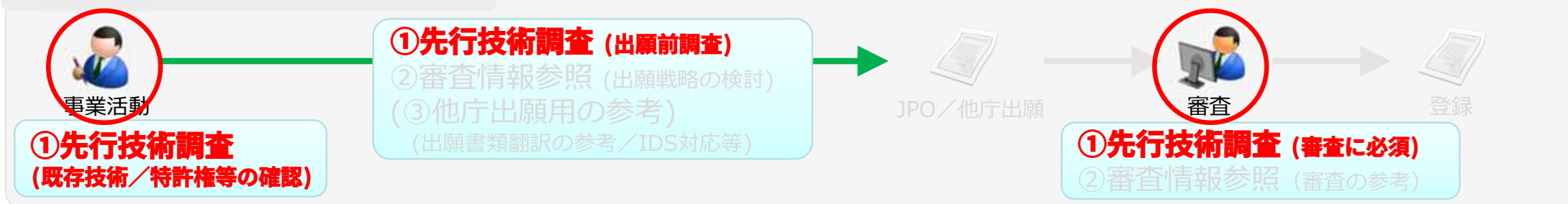
参照する特許情報の種類と時期



翻訳した特許情報の活用

・具体例②（海外出願の検索、照会環境整備）

参照する特許情報の種類と時期

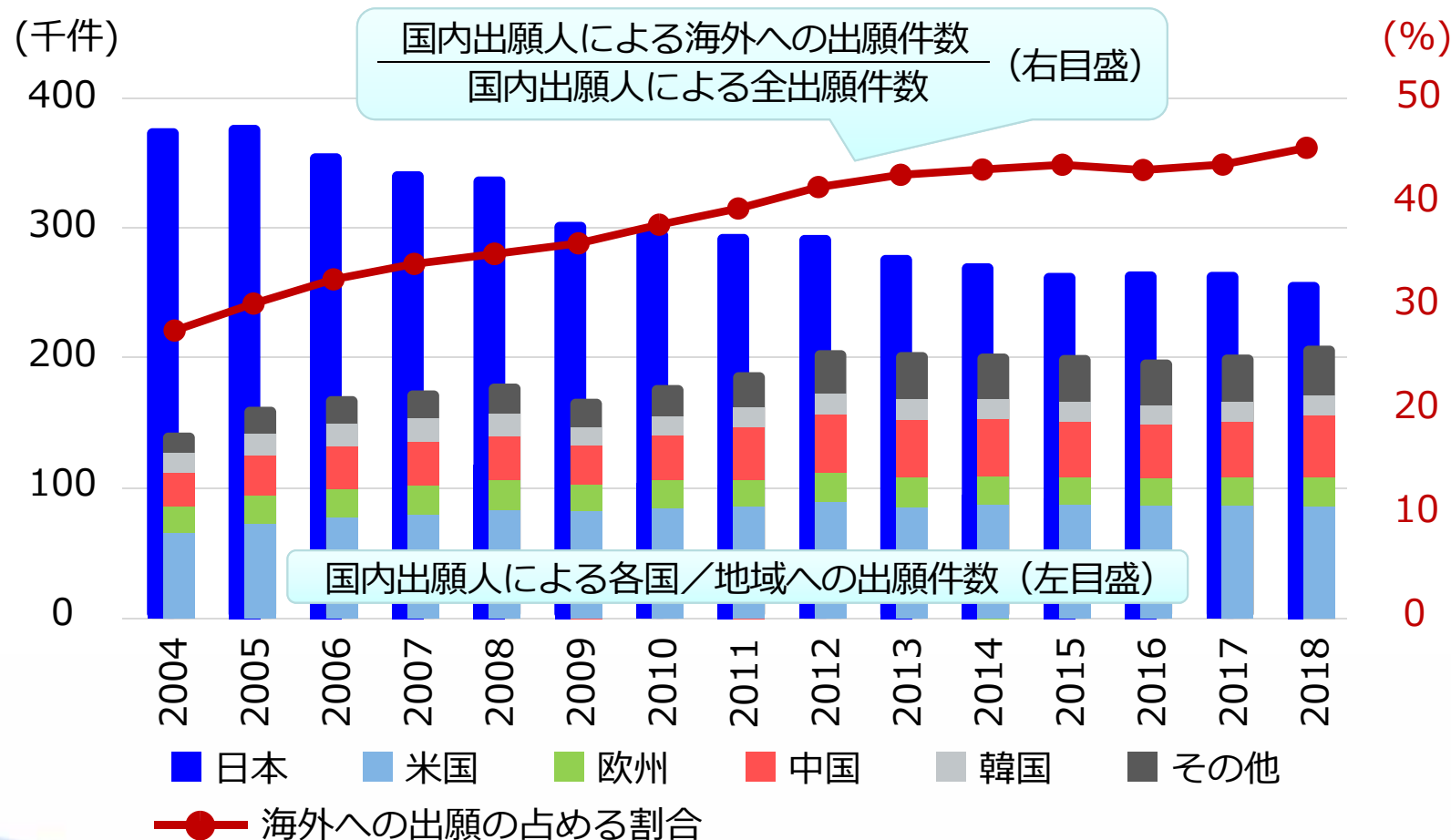


1-2 基礎資料と関連施策

基礎資料①

・国内出願人の出願動向（国内／海外）

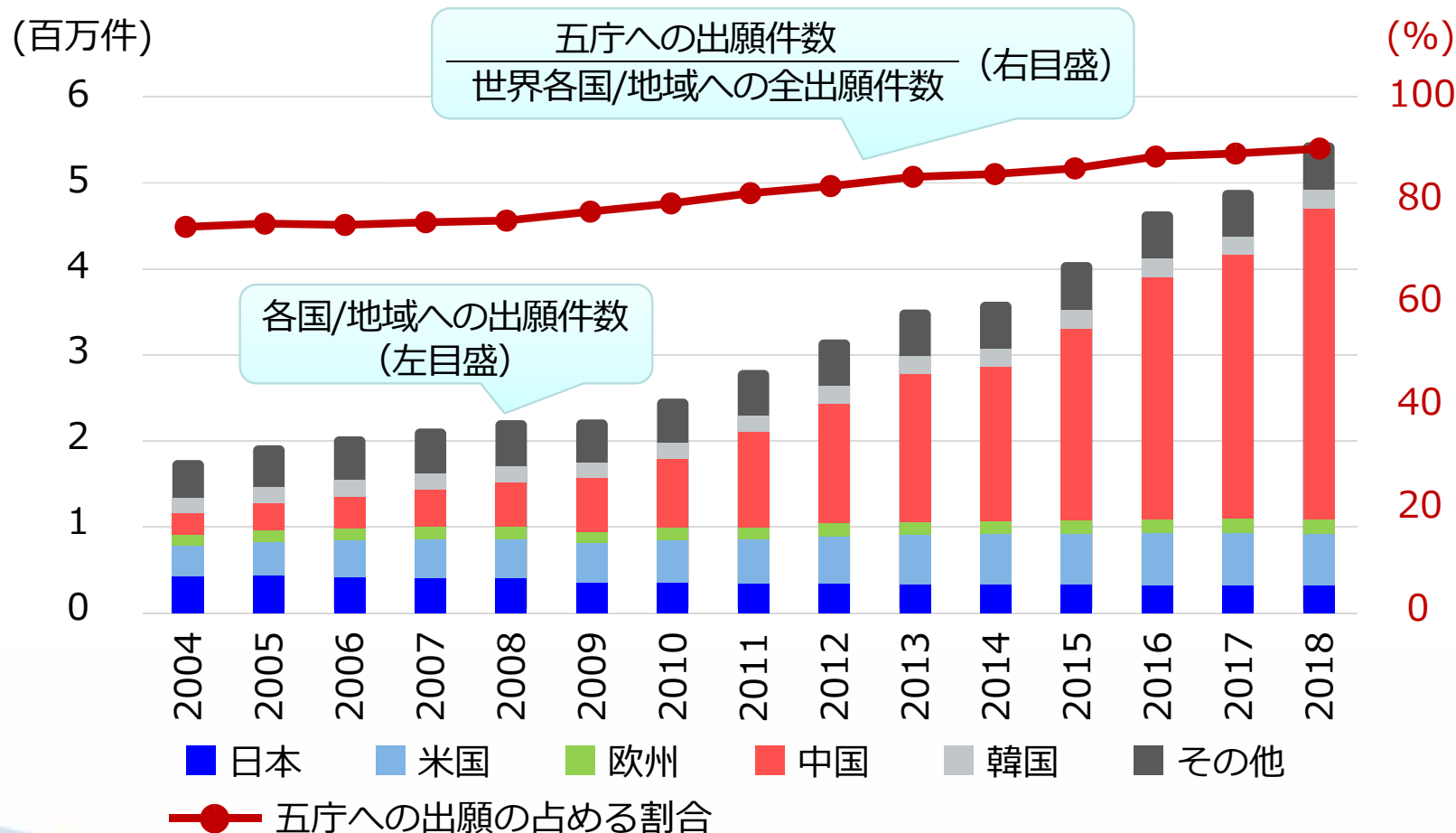
- 国内出願人による特許及び実用新案の出願先は、国内が減少傾向にある中、海外は、増加ないし横ばい。
- 国内出願人の、海外出願に対する関心や重要度の高さが窺える。



基礎資料②

・国内外の特許・実用新案出願件数

- 全世界の特許及び実用新案の出願は、年々増加（2018年には約550万件）しており、特に五庁（日米欧中韓）で全体の約9割を占める。
- 特に中国の増加が著しく、中国特許及び実用新案の検索環境は重要である。



基礎資料①②に関する施策

・日本の審査情報の海外発信／海外出願の検索、照会環境整備

- 国内出願人の関心／重要度が高い海外出願において、国内出願人が円滑に権利取得する一助となるように、日本の審査情報を翻訳して海外発信している。
- 海外出願、特に急増中の中国特許・実用新案の検索、照会環境を整備している。

日本の審査情報の海外発信（スライド5：具体例①）

迅速な審査

2018年平均（特許）

	一時審査期間（月）	最終判断までの期間（月）
JPO（日本）	9.3	14.1
USPTO（米国）	15.6	23.8
EPO（欧州）	4.4※	25.1
CNIPA（中国）	15.4	22.5
KIPO（韓国）	10.3	15.8

※ 出願日から拡張欧州調査報告の発表までの中央値
（出典）特許行政年次報告書2020年版

高品質な審査

特許審査全般の質（全体評価）2020年度調査

満足
9.1%

比較的満足
55.2%

普通
33.0%

比較的不満
2.4%

不満
0.3%

（出典）令和2年度 特許審査の質についてのユーザー評価調査報告書

翻訳・海外発信

日本の審査結果の
活用促進



海外の審査官

海外出願（中国特許・実用新案）の検索、照会環境整備（スライド6：具体例②）

中国公報を
日本語で検索

特許庁



特許庁が
中国公報データを受領



日本語に
機械翻訳



検索データ
ベースに蓄積

海外特許情報への
アクセス性向上



一般利用者/審査官

2. 特許庁の機械翻訳に関する取組み

2-1 機械翻訳文の対外提供、機械翻訳事業

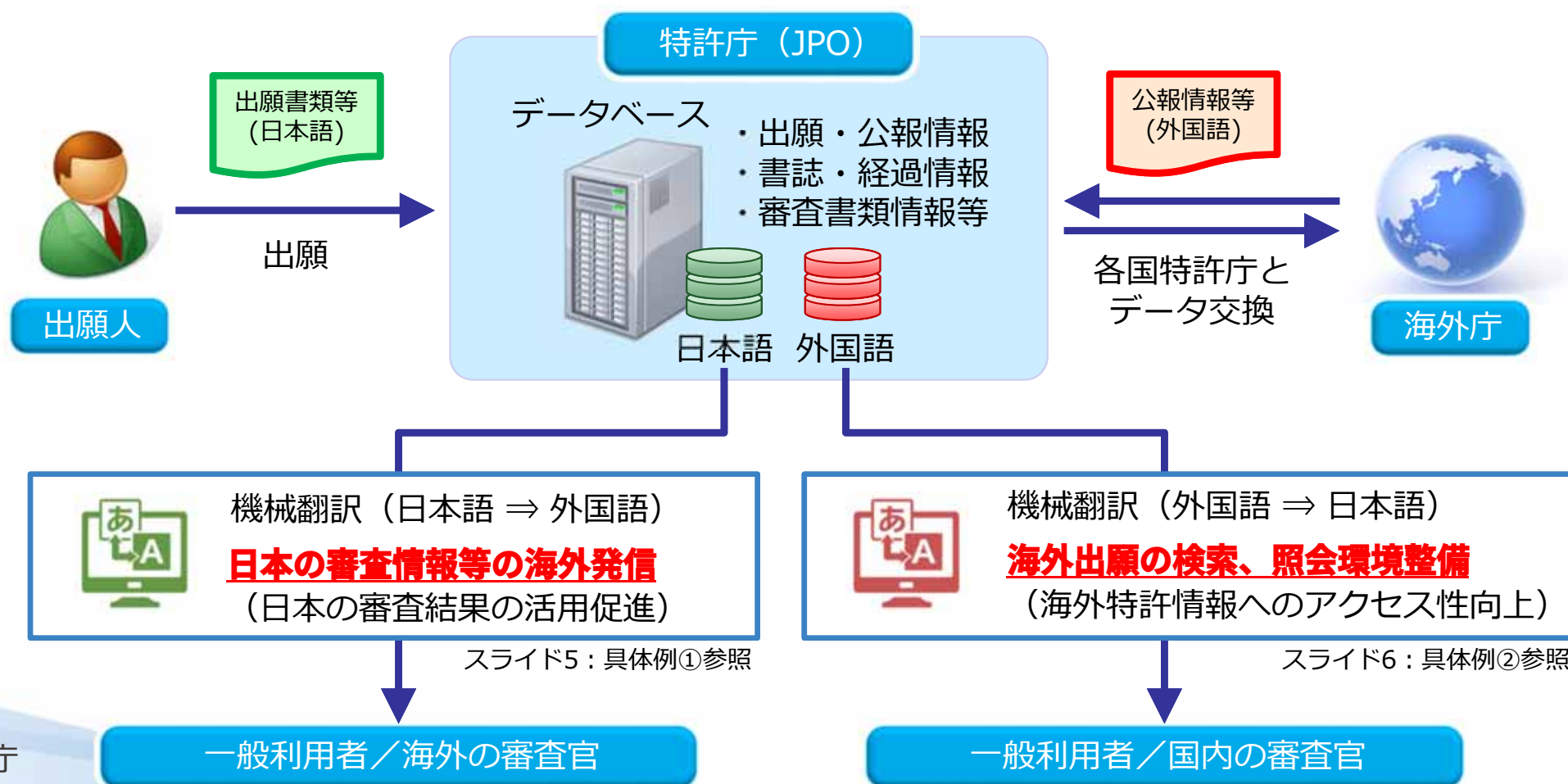
2-2 関連事業

2-1 機械翻訳文の対外提供、機械翻訳事業

機械翻訳の活用、及び、翻訳文の対外提供

- 特許庁では、日本の審査情報等の海外発信／海外出願の検索、照会環境整備の施策※における翻訳として、機械翻訳を活用している。
- 機械翻訳した、出願人や海外庁から得られた各種データは、国内外の審査官や一般利用者に提供している。

※ スライド10参照



「特許情報」の翻訳文を提供している主なインターフェース

- 「特許情報」とは、特許・実用新案・意匠・商標の出願や権利化に伴って生み出される情報全般を指す。日本国特許庁が保有する特許情報は、J-PlatPat、OPD、FOPISE等を通じて、無償で対外提供されている。

特許情報プラットフォーム J-PlatPat

<https://www.j-platpat.inpit.go.jp>

- 国内外の特許、実用新案、意匠、商標、審決に関する公報、手続、審査経過等を収録
- 「ぷらっと」寄って、情報を「ぱっと」見つけれられるような、ユーザーフレンドリーなサービスが提供できるようにとの意味
- 2015.3より運用開始（「特許電子図書館(IPDL ; 1999.3運用開始)」の後継）

ワンポータルドシエ OPD

<https://www.j-platpat.inpit.go.jp/p0000>
(J-PlatPatインターフェース（一般用）)

- 各庁の特許審査に関連する、手続や審査の状況、各種書類データ等（ドシエ情報）を、原語及び英語で閲覧できるプラットフォーム
 - AIPNインターフェースは、JPOに接続申請した海外庁が使用可能
- <経緯>
- 2003：アジア産業財産ネットワーク (AIPN)運用開始（国内の審査経過・結果情報のみ提供開始）
 - 2004：ドシエ情報全般も提供開始；高度産業財産ネットワーク (AIPN) に名称変更
 - 2006：日米欧の三極でドシエ情報が相互参照可能なシステム構築
 - 2013：日米欧中韓の五庁でドシエ情報が相互参照可能なOPD運用開始（2008よりJPO主導で開発）
 - 2016：J-PlatPat上から一般ユーザーにも提供開始

dossier (仏語で関係書類)

外国特許情報サービス FOPISE

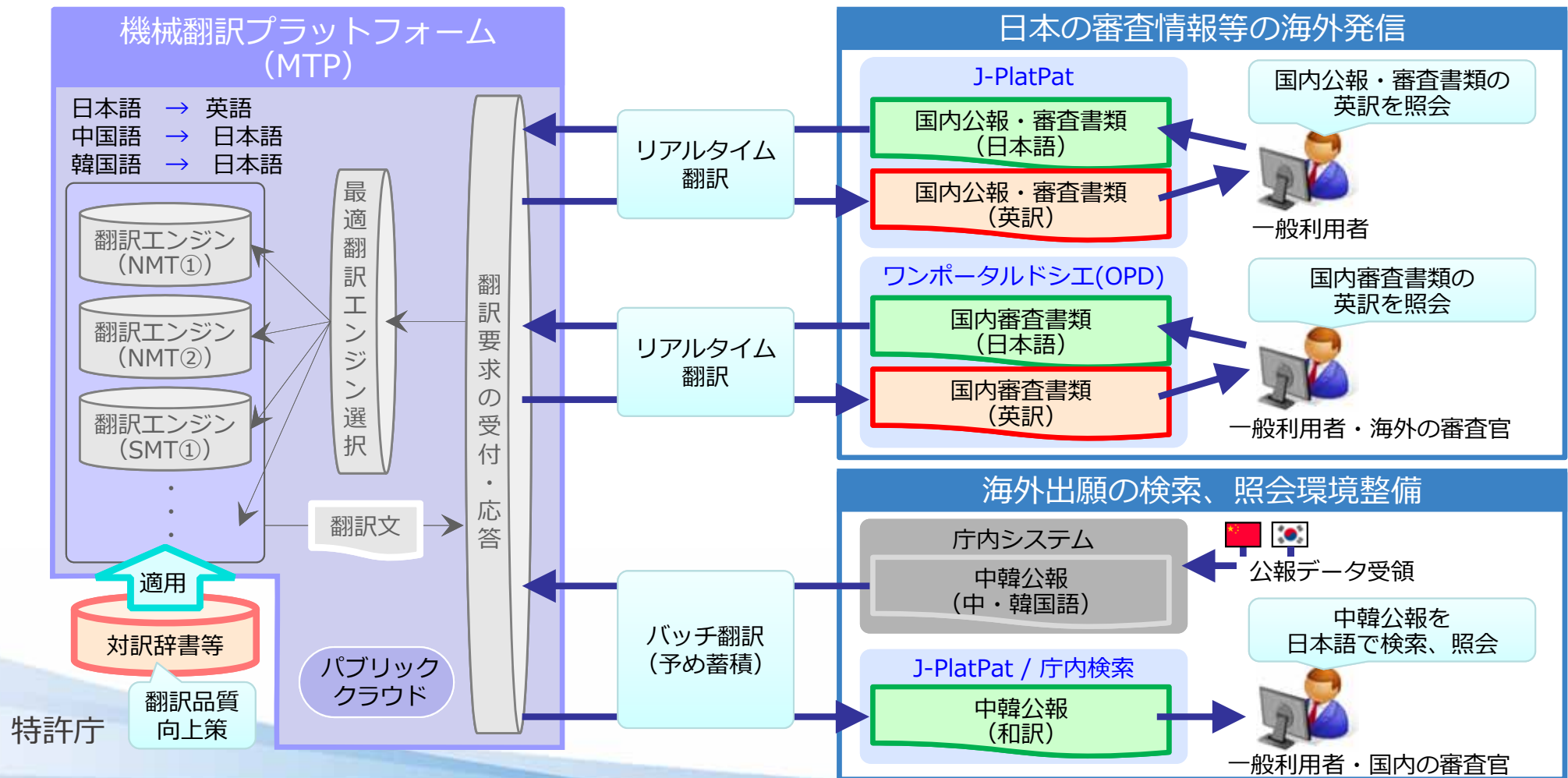
<https://www.foreignsearch2.jpo.go.jp>

- 露・台・オーストラリア・シンガポール・ベトナム・タイ・EUIPOの特許、実用新案、意匠又は商標の公報情報等、中国の特許及び実用新案に関する審決の和訳文を収録（J-PlatPat未収録）
- Foreign Patent Information Serviceの略称
- 提供時間は、開庁日の9時から20時に限られる
- 2015.8より運用開始

機械翻訳プラットフォーム（MTP）による機械翻訳

・全体図

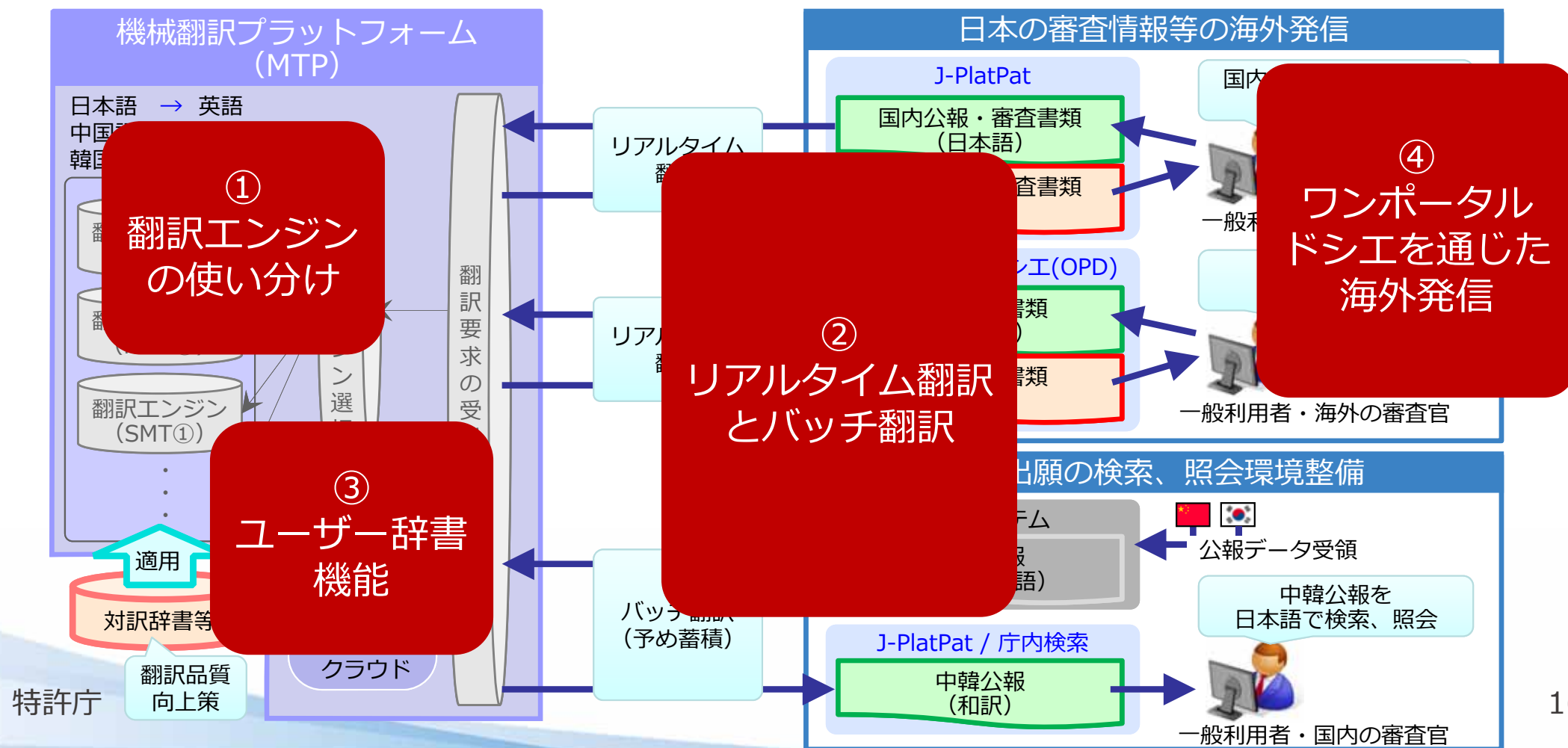
- 特許庁では、機械翻訳プラットフォーム（MTP）によって、日本の審査情報や公報の英訳、中韓の特許/実用新案の和訳を行っている。
- MTPは、複数の翻訳エンジンを搭載しており、令和元年5月に日英翻訳機能、翌令和2年4月に中日・韓日翻訳機能をリリースし、現在運用中である。



機械翻訳プラットフォーム（MTP）による機械翻訳

・全体図

- 特許庁では、機械翻訳プラットフォーム（MTP）によって、日本の審査情報や公報の英訳、中韓の特許/実用新案の和訳を行っている。
- MTPは、複数の翻訳エンジンを搭載しており、令和元年5月に日英翻訳機能、翌令和2年4月に中日・韓日翻訳機能をリリースし、現在運用中である。



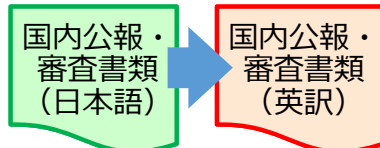
機械翻訳プラットフォーム（MTP）による機械翻訳

① 翻訳エンジンの使い分け

- MTPには、複数の翻訳エンジンが搭載されており、NMTをメインに利用しつつ、SMT及びRBMTの併用や、翻訳前後の文章に対する各種処理等を含む独自システムを通じて、高い翻訳品質を達成している。

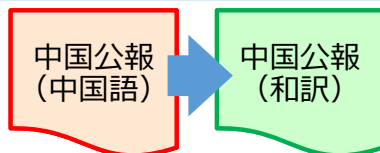
MTPにおけるメインの翻訳エンジンの使い分け

日英翻訳



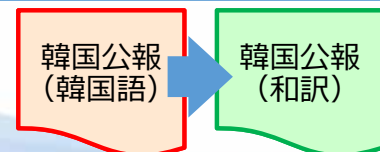
翻訳対象	メイン翻訳エンジン
特許公報（明細書等）	NMT（公報用カスタマイズ）
特許公報（発明の名称）	SMT
審査書類等	NMT（審査書類用カスタマイズ）
書誌情報・特許以外の公報 等	RBMT

中日翻訳



翻訳対象	メイン翻訳エンジン
特許公報（明細書等）	NMT
特許公報（発明の名称）	SMT

韓日翻訳



翻訳対象	メイン翻訳エンジン
特許公報	SMT

機械翻訳プラットフォーム（MTP）による機械翻訳

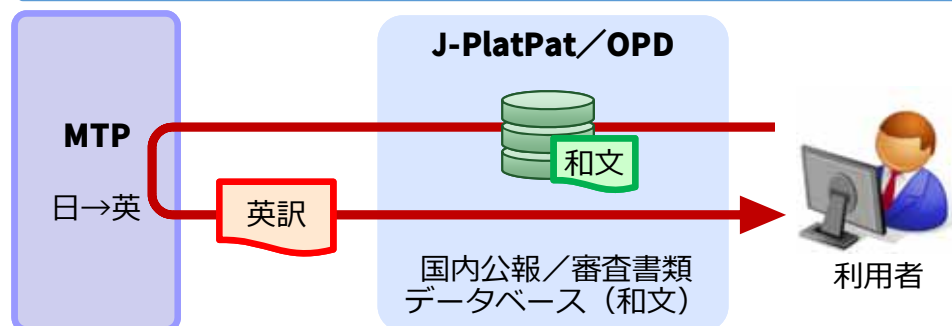
② リアルタイム翻訳とバッチ翻訳

- 日英翻訳は、リアルタイム翻訳（照会の度に都度翻訳）、中日・韓日翻訳はバッチ翻訳（予め検索システムに翻訳文を蓄積）を実施している。

（公的サービスが提供する機械翻訳は、当面、翻訳精度の向上が期待されることから、リアルタイム翻訳が原則ただし、中韓は過去事業の継続として例外。） 情報普及活用小委員会報告書「特許情報のさらなる活用に向けて」（p.31脚柱16参照）

リアルタイム翻訳とバッチ翻訳の違い

リアルタイム翻訳（都度翻訳・蓄積無し）



メリット

最新品質の翻訳文が参照できる

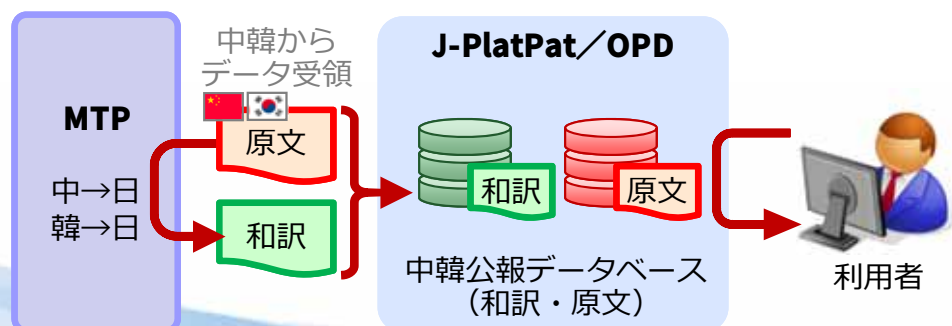
デメリット

翻訳文を対象としたテキスト検索ができない

国内公報・審査書類（日本語）

国内公報・審査書類（英訳）

バッチ翻訳（蓄積あり）



メリット

翻訳文を対象としたテキスト検索ができる

デメリット

蓄積時点の翻訳品質（最新の品質でない）

中韓公報（原文）

中韓公報（和訳）

特許庁

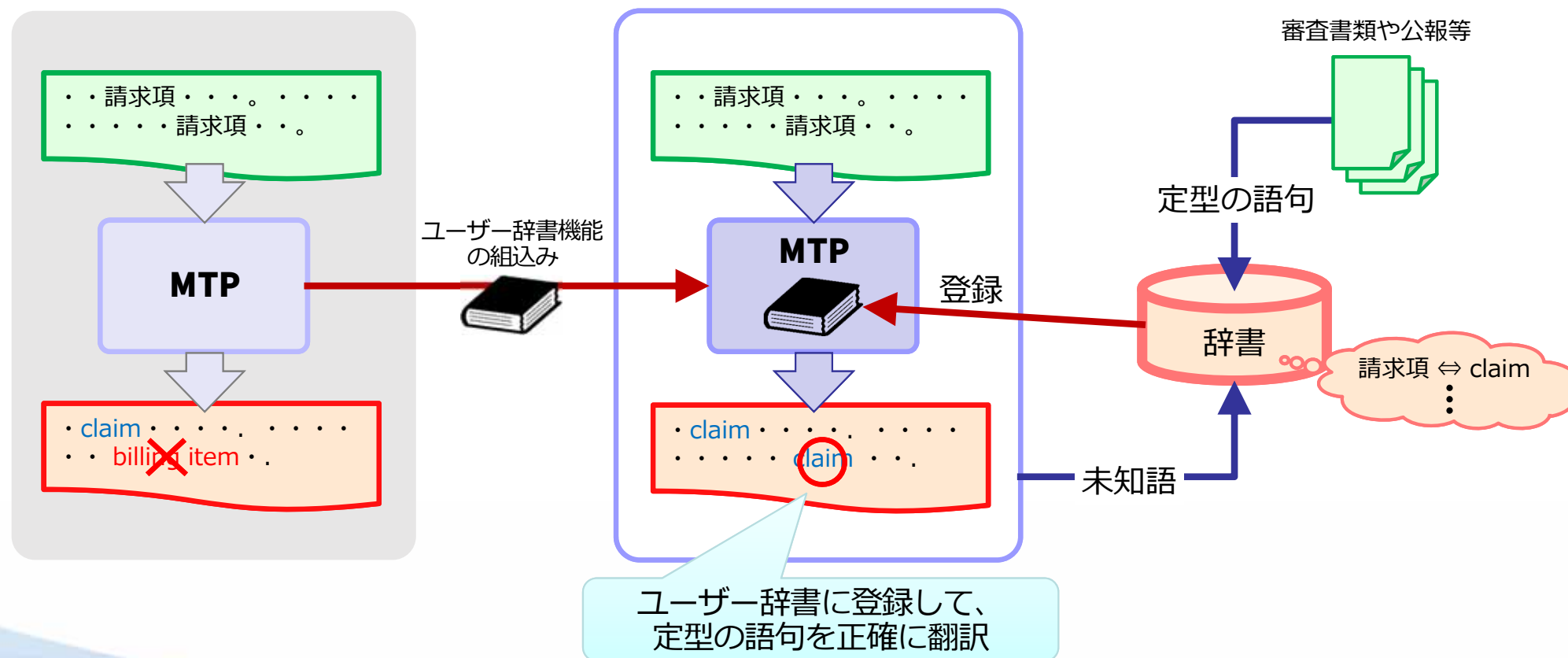
MTPによる高品質の和訳は、主に2020.03以降の発行分（過去分は、(旧)中韓文献翻訳・検索システムによる和訳）

機械翻訳プラットフォーム（MTP）による機械翻訳

③ ユーザー辞書機能

- MTPには、ユーザー辞書機能が組み込まれており、定型の語句の正確な翻訳及び翻訳揺れ防止を実施している。
- 未知語（機械翻訳できなかった用語）も、発見し次第積極的に登録している。

ユーザー辞書機能の組み込み

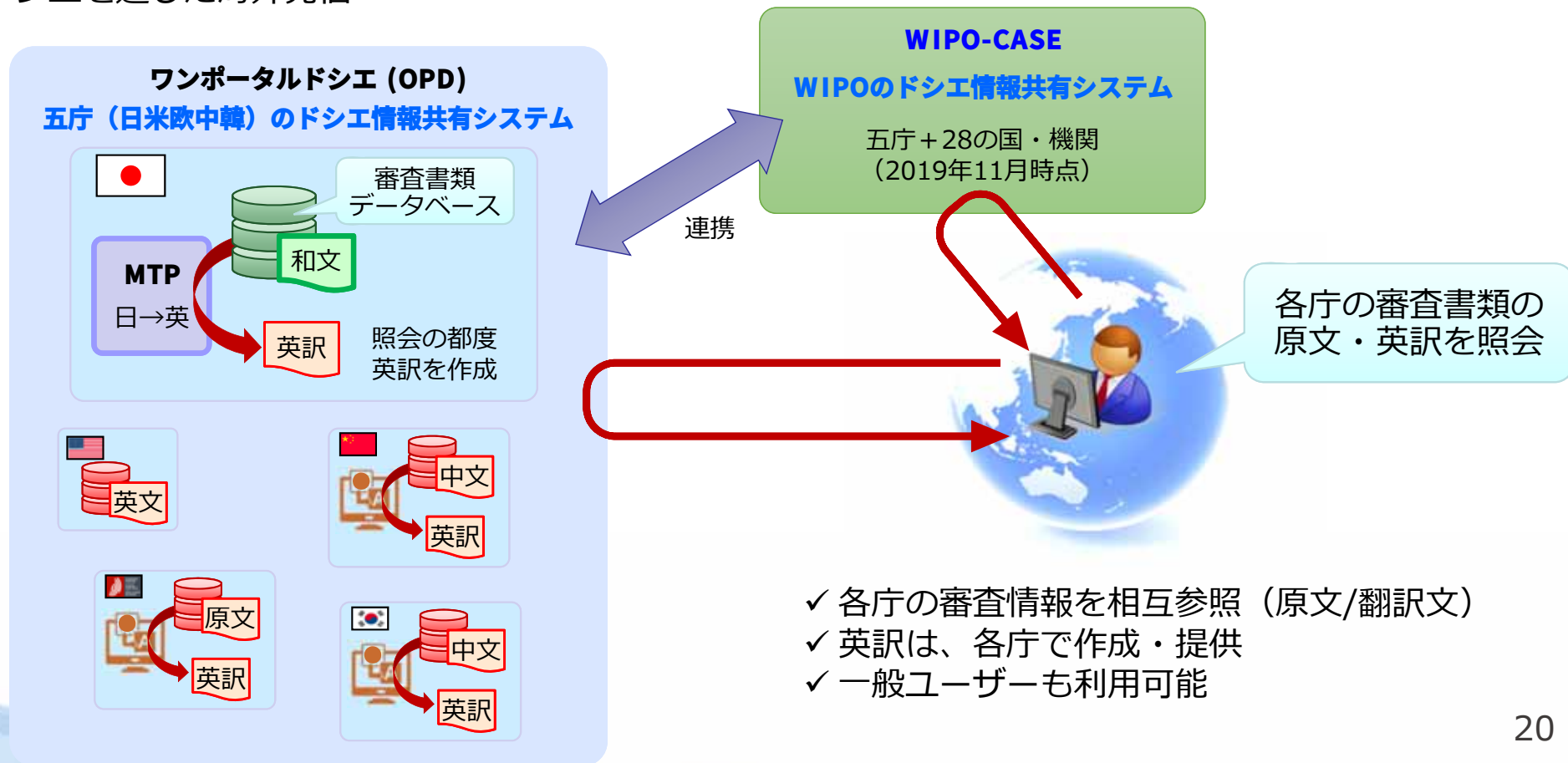


機械翻訳プラットフォーム（MTP）による機械翻訳

④ ワンポータルドシエを通じた海外発信

- 各国の審査書類（出願人や審査官が作成した各種書類；ドシエ情報）は、ワンポータルドシエ（OPD）／WIPO-CASEを通じて世界中で共有されている。
- 日本の国内審査書類については、原文（日本語）及びMTPによる英訳を、OPD／WIPO-CASEを通じて参照することができる。

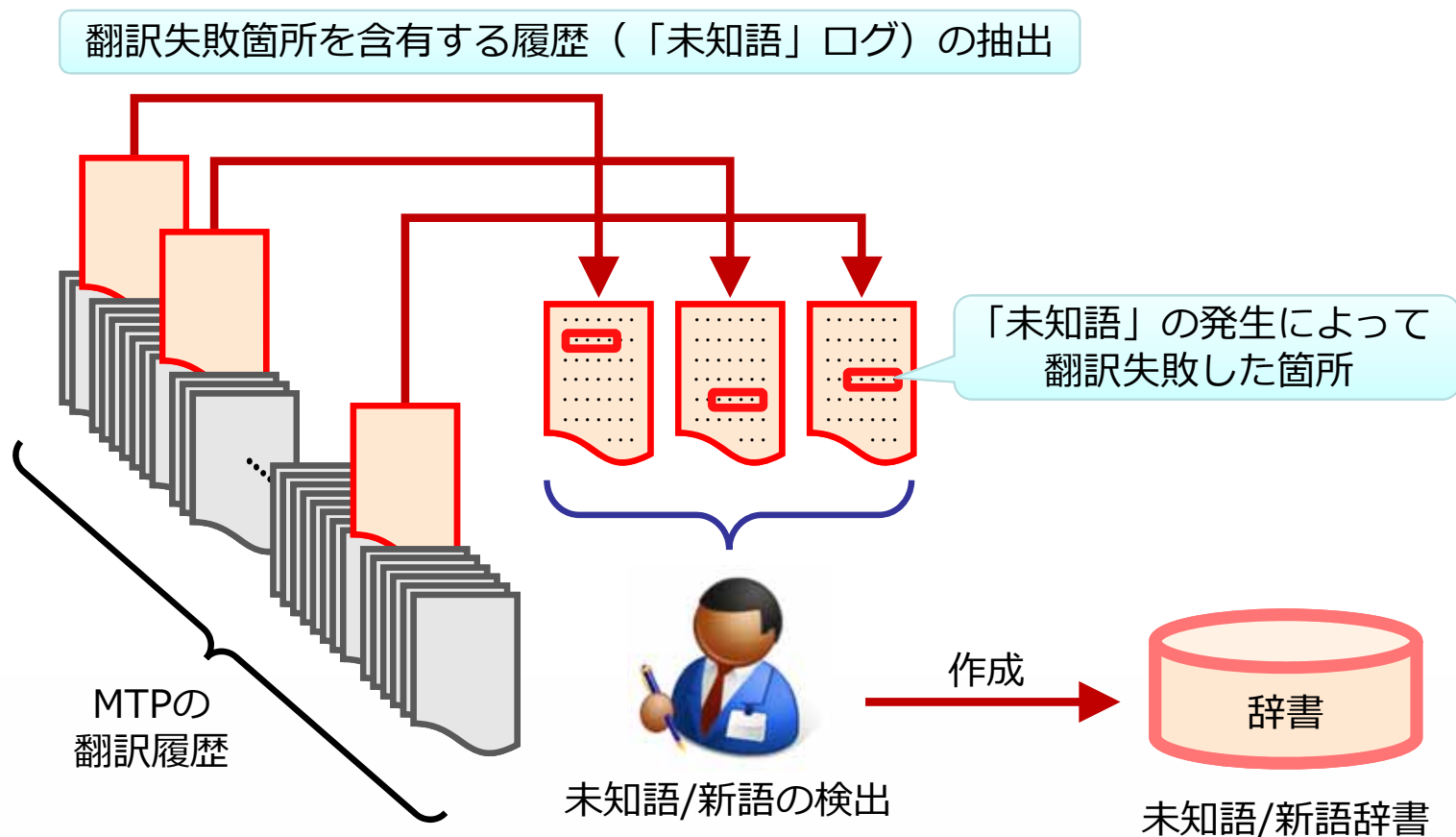
ワンポータルドシエを通じた海外発信



2-2 関連事業

新語対応による機械翻訳精度向上に関する調査

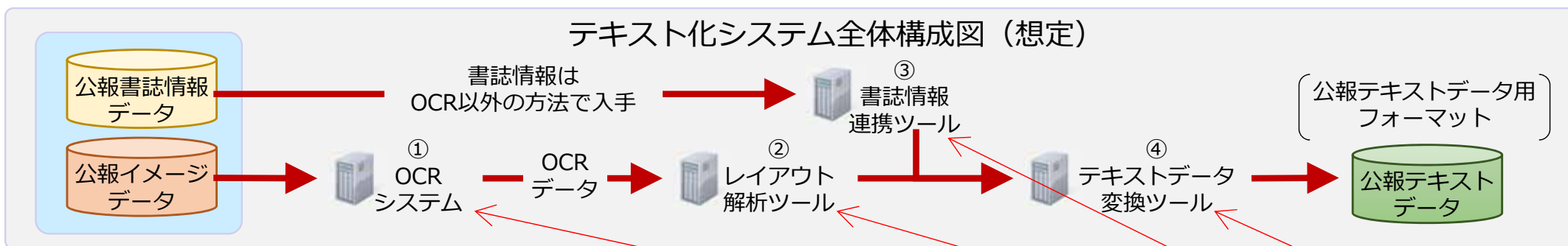
- MTPの日英翻訳によって翻訳できなかった未知語／新語の辞書を作成すると共に、機械翻訳の品質維持向上のため、未知語／新語の出現傾向や、辞書登録すべき未知語の効率的な選定方法について調査した。



最新のOCR技術による公開特許公報等のテキストデータ作成等を通じた機械翻訳サービスの充実化に向けた調査

- 国内の古い公報のイメージデータを、既存のOCR及び人工知能を活用したAI-OCRによりテキスト化して、機械翻訳する際の課題等を整理することが目的。
- 公報サンプルデータをフォーマット毎に分類し、OCRの手法及びテキスト費用を調査した。

公報の年代別フォーマット区分毎のOCRコスト



公報サンプルデータ

公開特許	1971-1992	1,001件
公表特許	1979-1995	701件
再公表特許	1979-1995	700件
公告特許	1922-1993	1,754件
公開実用新案	1971-1992	1,001件
公表実用新案	1979-1995	166件
再公表実用新案	1981-1992	19件
公告実用新案	1922-1993	1,754件
公開実用新案全文	1971-1992	1,002件

特許庁

公報レイアウトの特徴別分類	想定件数	①OCRシステム		②レイアウト解析ツール	③書誌情報連携ツール	④テキストデータ変換ツール	合計
		単価	最適ツール使用				
[I] 旧字旧仮名	51万件	40円/件	約2千万円 精度：約93%	約2千万円 精度：約50%	約2千万円	約2千万円	約8千万円
[II] 画質が悪い (新字新仮名)	507万件	40円/件	約2億円 精度：約87%	約2千万円 精度：約80%			約2.6億円
[III] 画質が良い (新字新仮名)	1,072万件	6.7円/件	約7千万円 精度：約95%	約2千万円 精度：約90%			約1.3億円

最新のOCR技術による公開特許公報等のテキストデータ作成等を通じた機械翻訳サービスの充実化に向けた調査

(参考) 特許公報のフォーマット変遷

公開特許

開始日	終了日	改訂箇所
1971/7/16	1975/3/17	発行開始
1975/3/18	1977/6/30	Int.Cl2.追加
1977/7/1	1979/12/27	書誌事項全体のレイアウトが変わる
1980/1/5	1985/2/13	日本分類の記載が無くなる
1985/2/14	1992/7/30頃	発明の数、審査請求の位置が移動 書誌事項が1段表記になる
1992/7/30頃		現行公開公報に近い形になる

公表特許

開始日	終了日	改訂箇所
1979/7/26	1979/12/27	発行開始
1980/1/10	1983/7/7	日本分類の記載が無くなる
1983/7/14	1984/12/27	「予備審査請求」の欄が登場
1985/1/10	1993/12/22	「審査請求」、「予備審査請求」の欄移動 書誌事項が1段表記になる
1994/1/6		現行実用新案公報に近い形になる

再公表特許

開始日	終了日	改訂箇所
1979/8/9	1979/12/6	発行開始
1980/1/10	1983/7/7	日本分類の記載が無くなる
1983/8/4		「審査請求」、「予備審査請求」の欄登場

特許庁

公告特許

開始日	終了日	改訂箇所
1922/6/9	1922/11/17	発行開始
1922/11/22	1923/3/7	「出願ノ要旨」が「発明ノ性質及ヒ目的ノ要領」と「特許請求ノ範囲」とに分かれる
1923/3/9	1923/8/24	「願書番号」が「特許願番号」に変わる
1923/8/25	1924/5/12	(関東大震災の影響)
1924/5/13	1924/7/16	公告番号を初期化して公報発行再開 「願書番号」に戻る
1924/7/18	1933/1/30	「特許出願公告」の位置変更 ※1927年～公告番号が年ごとになる
1933/2/1	1934/2/9	ページ番号の位置が変わる
1934/2/12	1936/3/30	「(特許局発行)」との記載が登場
1936/4/1	1938/7/29	冊子全体の通番が登場
1938/8/1	1940/8/23	1段表記から2段表記に変わる
1940/8/26	1940/11/30	「発明ノ性質及目的ノ要領」表記が無くなる
1940/12/5	1943/11/11	紙面のサイズが縮小する
1943/11/12	1947/2/14	(戦時特例により出願公告制度が停止)
1947/2/15	1947/12/26	「(特許標準局発行)」との記載となる
1948/2/10	1949/12/23	1紙面に対し、複数文献が記載される
1950/1/9	1951/10/30	文献ごとの記載に戻る
1951/11/4	1952/12/23	公報下部の書誌事項が無くなる 「日本國政府」→「特許庁」に表記が変更 「(全 ○ 頁)」との記載が登場
1953/1/6	1963/12/28	「発明の性質及目的の要領」表記が無くなる
1964/1/9	1969/3/31	書誌事項が公報上部から本文に移動
1969/4/2	1970/9/30	書誌事項に番号が導入される 欄の番号、発明の数の欄が登場
1970/10/1	1972/12/27	Int.Cl.導入
1973/1/5	1975/3/31	クレームに丸囲み文字が記載される
1975/4/1	1977/3/31	庁内整理番号導入
1977/4/1	1979/3/31	書誌事項が「特許公報」の下に移動する
1979/4/2	1979/12/27	JP、Bの表示
1980/1/5	1984/12/27	日本分類の記載が無くなる
1985/1/5	1993/12/24	書誌事項が1段表記になる

最新のOCR技術による公開特許公報等のテキストデータ作成等を通じた機械翻訳サービスの充実化に向けた調査

(参考) 実用新案公報のフォーマット変遷

公告実用新案

開始日	終了日	改訂箇所
1922/6/1	1923/8/31	発行開始
1923/9/1	1924/5/18	(関東大震災の影響)
1924/5/19	1924/7/24	公告番号を初期化して公報発行再開
1924/7/31	1933/1/31	「実用新案出願公告」の位置が変更 ※1927年～公告番号が年ごとになる
1933/2/2	1934/2/10	ページ番号の位置が変わる
1934/2/13	1936/3/31	「(特許局発行)」との記載が登場
1936/4/2	1938/1/29	冊子全体の通番が登場
1938/2/1	1938/7/30	冊子全体の通番の位置が変わる
1938/8/2	1940/11/30	1段表記から2段表記になる
1940/12/3	1943/11/13	紙面のサイズが縮小する
1943/11/14	1947/2/27	(戦時特例により出願公告制度が停止)
1947/2/28	1947/12/27	公報発行再開 「(特許標準局発行)」との記載となる
1948/2/10	1949/12/23	1紙面に対し、複数文献が記載される
1950/1/9	1951/10/30	文献ごとの記載に戻る
1951/11/4	1963/12/28	公報下部の書誌事項がなくなる 「日本國政府」→「特許庁」に表記が変更 「(全 ○ 頁)」との記載が登場
1964/1/9	1969/3/31	書誌事項が公報上部から本文に移動
1969/4/2	1970/9/30	書誌事項に番号が導入される 欄の番号、発明の数の欄が登場
1970/10/1	1972/12/27	Int.Cl.導入
1973/1/5	1975/3/31	クレームに丸囲み文字が記載される
1975/4/1	1977/3/31	庁内整理番号導入
1977/4/1	1979/3/31	書誌事項が「実用新案公報」の下に移動
1979/4/2	1979/12/27	JP、Yの表示
1980/1/5	1984/12/27	日本分類の記載が無くなる
1985/1/5	1993/12/24	書誌事項が1段表記になる

特許庁

公開実用新案

開始日	終了日	改訂箇所
1971/9/13	1972/12/27	発行開始
1973/1/5	1975/3/24頃	クレームに丸囲み文字が記載される
1975/3/24頃	1977/3/31	Int.Cl2.追加
1977/4/1	1977/6/30	書誌事項が「公開実用新案公報」の下に移動
1977/7/1	1979/3/31	「(全 ○ 頁)」との記載が登場 書誌事項が2段表記になる
1979/4/2	1985/2/13	JP、Uの表示
1985/2/14	1989/7/14頃	審査請求の位置が移動 書誌事項が1段表記になる
1989/7/14頃	1992/7/28頃	請求項の数の欄が登場
1992/7/28頃		現行実用新案公報に近い形になる

公開実用新案全文

開始日	終了日	改訂箇所
1971/9/13	1983/3/31	発行開始
1983/4/1	1985/2/13	書誌事項表記が公開公報と近い形式に変更
1985/2/14	1989/7/14頃	審査請求の位置が移動 書誌事項が1段表記になる
1989/7/14頃	1992/7/28頃	請求項の数の欄が登場
1992/7/28頃		現行実用新案公報に近い形になる

公表実用新案

開始日	終了日	改訂箇所
1979/9/6	1979/10/25	発行開始
1980/9/6	1983/6/23	日本分類の記載が無くなる
1983/8/4	1983/12/20	「予備審査請求」の欄が登場
1985/1/17	1993/9/2	「審査請求」、「予備審査請求」の欄が移動 書誌事項が1段表記になる
1994/1/6		現行実用新案公報に近い形になる

再公表実用新案

開始日	終了日	改訂箇所
1982/3/11	1982/3/11	発行開始
1984/7/19		「審査請求」、「予備審査請求」の欄が記載

3. 海外の動向

WIPO/EPO

WIPO Translate

- WIPO(世界知的所有権機関)が開発した機械翻訳ツール。PCTの国際公開言語(日本語を含む10言語)について、自由入力文の翻訳が可能。
- WIPOの特許検索(PATENTSCOPE)において、公報／書誌情報等をリアルタイム翻訳する翻訳エンジンとして、Google翻訳と併用している。
- 2016年9月よりNMTを採用。

WIPO Translate



<https://patentscope.wipo.int/translate/translate.jsf>

PATENTSCOPE



<https://patentscope2.wipo.int/search/ja/search.jsf>

Patent Translate (EPO)

- EPO（欧州特許庁）がGoogleと協力して開発した機械翻訳ツール。日本語を含む32言語の翻訳が可能。
- EPO提供の特許検索サービス（Espacenet）を通じて提供。（Espacenetは2019年11月に正式リニューアル。）
- 2017年5月よりNMTを採用。

The screenshot displays the Espacenet Patent search interface. At the top, there is a search bar with the text "Enter your search terms" and a magnifying glass icon. To the right of the search bar is a dropdown menu labeled "Office/Language". Below the search bar is a navigation bar with links: "My Espacenet", "Help", "Classification search", "Results", "Advanced search" (with a toggle switch), "Filters" (with a toggle switch), "Popup tips", and "Feedback". The main content area shows the search form with a "Query language: en de fr" dropdown. Below this is a section for building search criteria, starting with "AND" and a "+Field" button. A dropdown menu is open, showing options for "Title" (with a dropdown set to "all") and "Title or abstract" (with a dropdown set to "Title or abstract"). There is also a "Group" button and a "Search" button at the bottom left of the form.

4. おわりに



特許庁は今後も、特許情報の提供に機械翻訳を引き続き活用し、
更なる特許情報の普及活用に努めて参ります。

ありがとうございました



機械翻訳コンペティション参加報告

清野 舜

国立研究開発法人理化学研究所 革新知能統合研究センター

目的指向基盤技術研究グループ 自然言語理解チーム

※スライドは後日公開予定です



謝辞

- 講演のアイデア出し・スライドの推敲に協力してくださった東北大学の鈴木潤先生、小林颯介さん、伊藤拓海さんに感謝します

自己紹介



- 経歴

- 2013 – 2019: 東北大学(学士 – 修士)
- 2019 – : 理化学研究所 革新知能統合研究センター
- 2020 – : 東北大学(博士後期課程)

- これまでの研究

- 生成型要約 [BlackboxNLP 2018], [PACLIC 2018]
- 高速 & 大規模な半教師あり学習 [AAAI 2019]
- 文法誤り訂正 [EMNLP 2019], [TASLP 2020]



機械翻訳の基盤技術は
タスクをまたいで使われている
⇒他タスクの経験が役立つ

機械翻訳の研究をしたことがない

つまり、情熱さえあれば大丈夫！

今日話すこと

① 機械翻訳コンペティション(WMT)の紹介 (10%)

② 東北大 理研AIP NTT CS研 - - NTT で参加してみて分かったこと3つ (90%)

今回参加したコンペティション：WMT

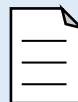
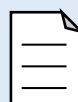
- WMT: もともと機械翻訳のワークショップ
 - 数年前に国際会議になりました
- さまざまなコンペティションが併設
 - 新聞記事翻訳 **我々が参加したタスク**
 - 未知のドメインの文に対する翻訳
 - 教師なし機械翻訳
 - 生物学・医療系文書の機械翻訳
 - チャット文の機械翻訳

最も歴史が古い
かつ
競争の激しいタスク

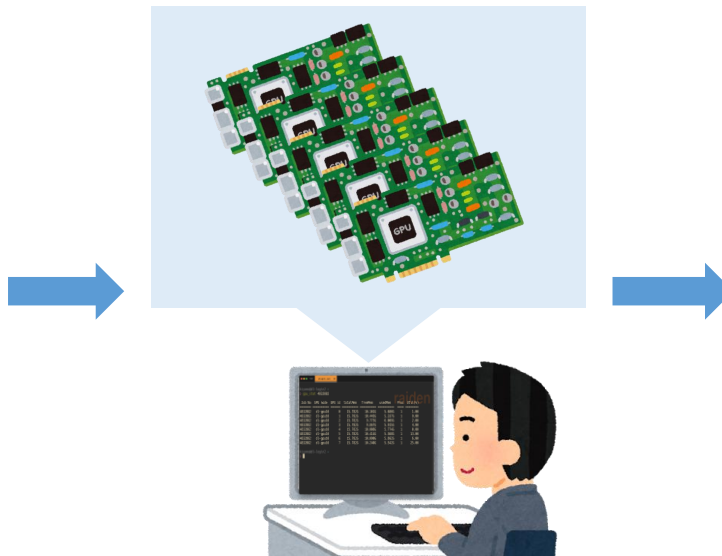
コンペティションの流れ（日英の場合）

① システム構築

データセットの準備・前処理

-  日英の対訳コーパス
-  単言語コーパス(日)
-  単言語コーパス(英)
-  前年度までのテストデータ

超大量(100枚~)のGPUで
試行錯誤



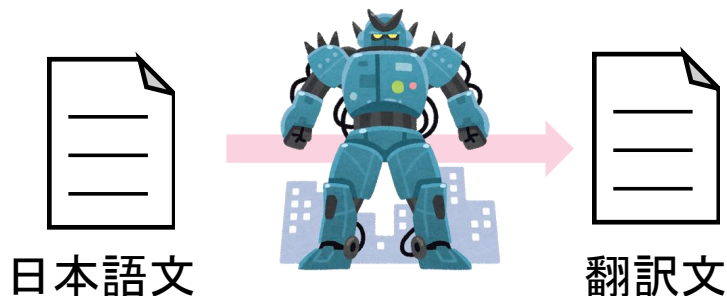
チームメンバー

システム完成

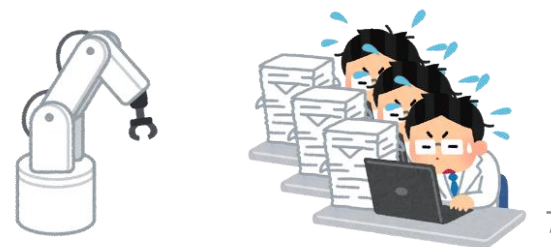


② システム評価

テストデータの翻訳



自動評価 (BLEU)・人手評価



参加チームの紹介

- 我々：東北大-理研AIP-NTTチーム

コンペティション上位の常連
WAT2017・WMT2018・WAT2019
で1位



Shun Kiyono^{1,2}

 ¹RIKEN AIP



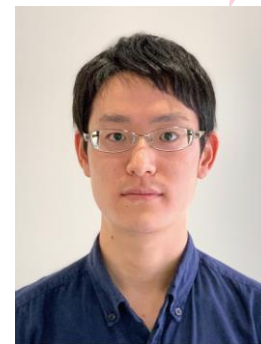
Takumi Ito^{2,1}



²Tohoku University



Ryuto Konno^{2,1}



Makoto Morishita^{3,2}



³NTT CS Laboratory



Jun Suzuki^{2,1}

- その他に 京大・NICT・DeepMind・Facebook・エジンバラ大・NAVER・OPPO・Tencent・WeChat・DiDi などが参加

結果：自動評価尺度（**BLEU**）で上位

独→英: 1位

Team	BLEU
Tohoku-AIP-NTT	43.8
Huoshan_Translate	43.5
OPPO	43.2
UEDIN	42.3
Online-B	41.9

英→独: 1位

Team	BLEU
Tohoku-AIP-NTT	38.8
Tencent_Translation	38.6
OPPO	38.6
Huoshan_Translate	38.2
eTranslation	37.9

日→英: 2位

Team	BLEU
NiuTrans	26.7
Tohoku-AIP-NTT	25.5
OPPO	24.8
NICT_Kyoto	22.8
eTranslation	22.2

英→日: 4位

Team	BLEU
NiuTrans	28.4
OPPO	27.3
ENMT	25.9
Tohoku-AIP-NTT	25.8
NICT_Kyoto	23.9

人手評価：全言語対で1位を達成

(人手評価に統計的有意差がない場合、同率一位という扱い)

German→English

Ave.	Ave. z	System
82.6	0.228	VolcTrans
84.6	0.220	OPPO
82.2	0.186	HUMAN
81.5	0.179	Tohoku-AIP-NTT
81.3	0.179	Online-A
81.5	0.172	Online-G
79.8	0.171	PROMT-NMT
82.1	0.167	Online-B
79.5	0.121	UEDIN

Japanese→English

Ave.	Ave. z	System
75.1	0.184	Tohoku-AIP-NTT
76.4	0.147	NiuTrans
74.1	0.088	OPPO
75.2	0.084	NICT-Kyoto
73.3	0.068	Online-B
70.9	0.026	Online-A
71.1	0.019	eTranslation
64.1	-0.208	zlabs-nlp
66.0	-0.220	Online-G
61.7	-0.240	Online-Z

English→German

Ave.	Ave. z	System
90.5	0.569	HUMAN-B
87.4	0.495	OPPO
88.6	0.468	Tohoku-AIP-NTT
85.7	0.446	HUMAN-A
84.5	0.416	Online-B
84.3	0.385	Tencent-Translation
84.6	0.326	VolcTrans
85.3	0.322	Online-A
82.5	0.312	eTranslation

English→Japanese

Ave.	Ave. z	System
79.7	0.576	HUMAN
77.7	0.502	NiuTrans
76.1	0.496	Tohoku-AIP-NTT
75.8	0.496	OPPO
75.9	0.492	ENMT
71.8	0.375	NICT-Kyoto
71.3	0.349	Online-A
70.2	0.335	Online-B
63.9	0.159	zlabs-nlp
59.8	0.032	Online-Z

その他：言語現象テストでも好成績

[Avramidis+2020]

category
Ambiguity
Composition
Coordination & ellipsis
False friends
Function word
LDD & interrogatives
MWE
Named entity & terminology
Negation
Non-verbal agreement
Punctuation
Subordination
Verb tense/aspect/mood
Verb valency



4.1 Comparison between systems

Two systems are standing out for their overall performance. **Tohoku-AIP-NTT** (Kiyono et al., 2020) achieves the best category macro-averaged accuracy of 88.1%, whereas together with **Huoshan** they share the first position based on their micro-averaged accuracy (85.3-85.4%). The systems UEdin (Germann, 2020), Online-B, Online-G and Online-A are next.

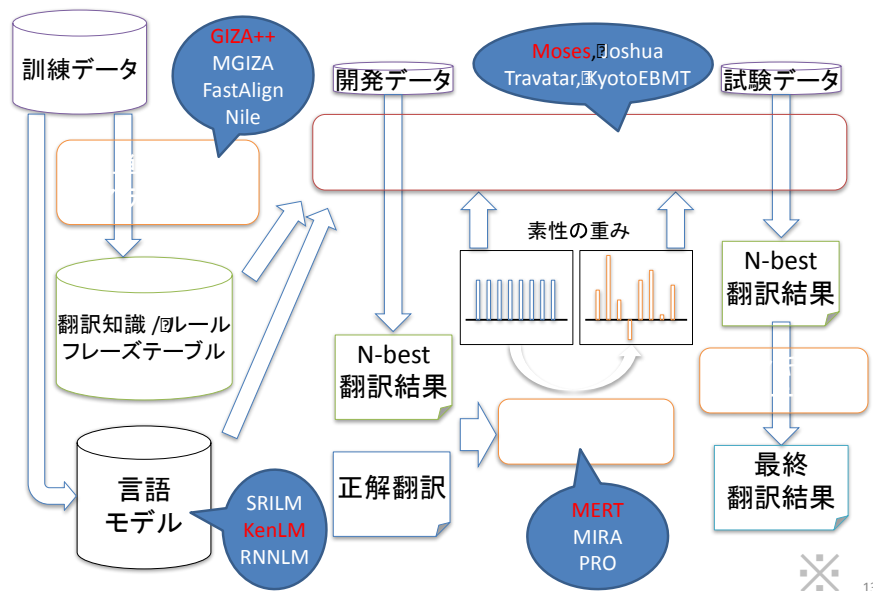
複単語表現・固有名詞
機能語・動詞の時制といった
言語現象の取り扱いに関する
テスト

我々のシステムが
マクロ平均値で最も良い成績

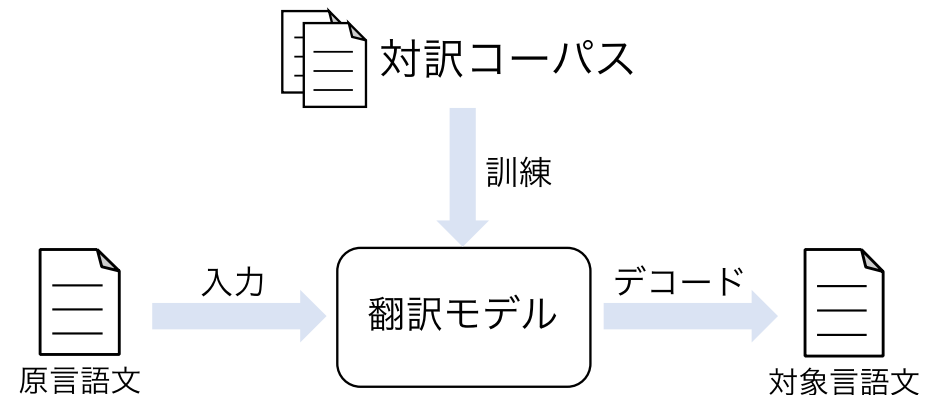
分かったこと①
最先端の**NMT**システムは泥臭い

よくある SMT vs NMT 観

SMT (統計的機械翻訳)



NMT (ニューラル機械翻訳)



- ☹️ 複数モジュールによる煩雑なシステム
- ☹️ エラー伝搬が翻訳精度に悪影響

- 😊 単一のモデルで一気通貫学習が可能
- 😊 エラー伝搬を解消→高い翻訳精度

よくある SMT vs NMT 観

SMT（統計的機械翻訳）

NMT（ニューラル機械翻訳）

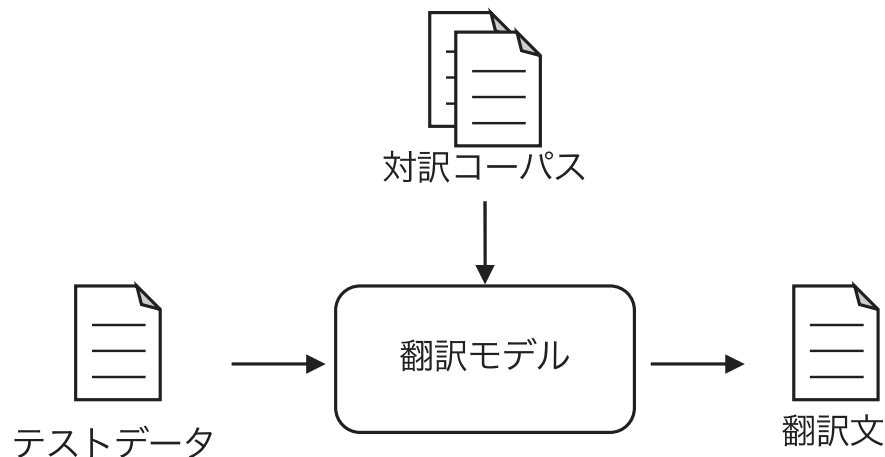


多くの場面でこれは正しいが...
「最先端のNMT」では状況が違う

- ☹️ 複数モジュールによる煩雑なシステム
- ☹️ エラー伝搬が翻訳精度に悪影響

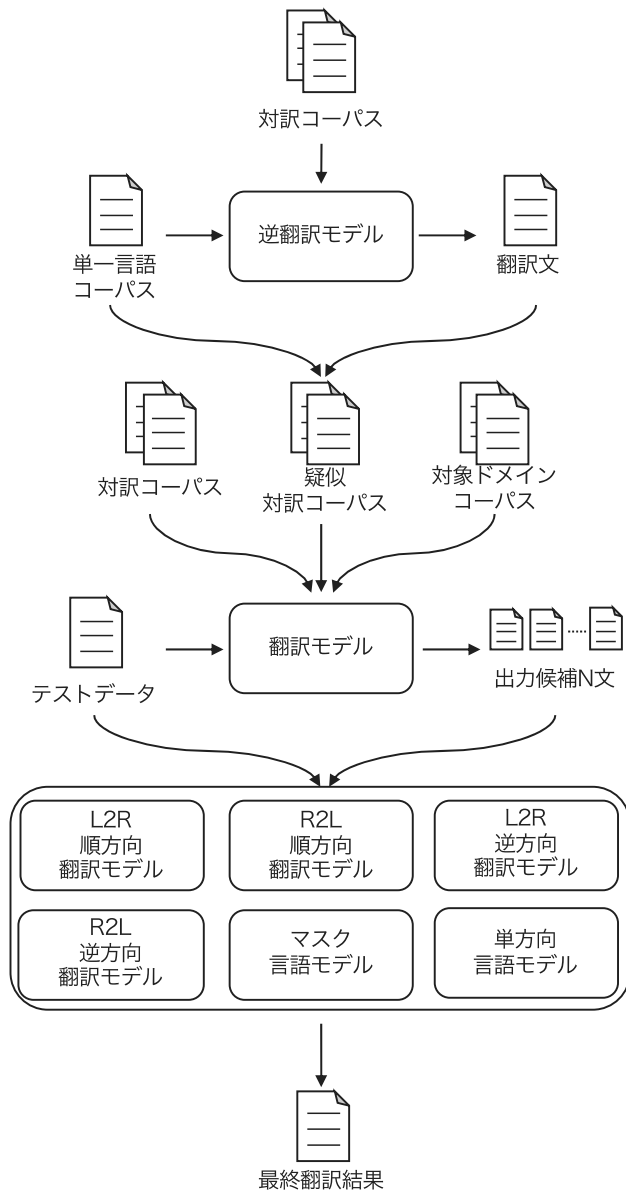
- 😊 単一のモデルで一気通貫学習が可能
- 😊 エラー伝搬を解消→高い翻訳精度

いつものNMTから最先端NMTへ



普段我々が慣れ親しんだNMTの枠組み

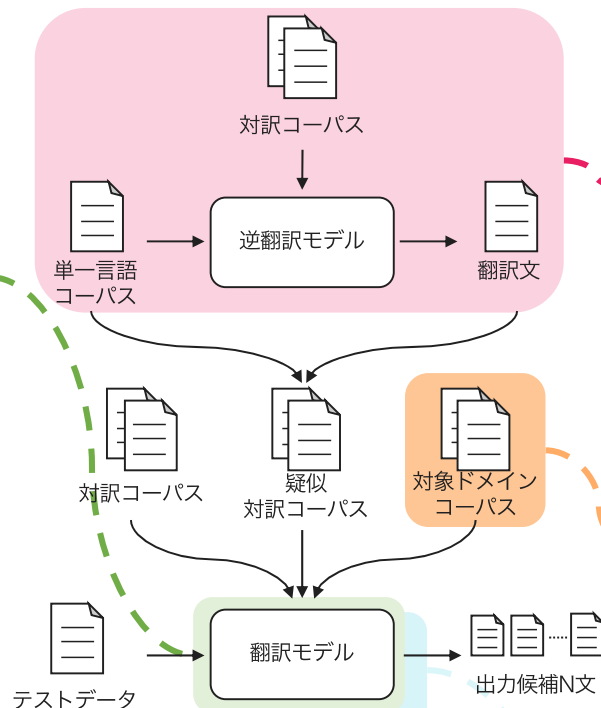
いつものNMTから最先端NMTへ



最先端NMTは“泥臭い”技術の集合体



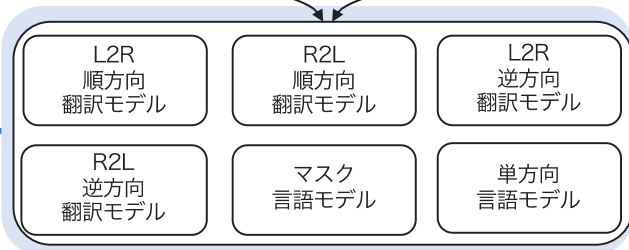
翻訳モデル(Transformer)はハイパーパラメータに
対して非常にセンシティブ
ハイパラの知見も日進月歩
⇒ハイパーパラメータの調整



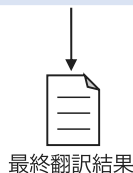
対訳コーパスだけでは
データが足りない
単一言語コーパスを使ってデ
ータを増やしたい
⇒逆翻訳によるデータ拡張

テストデータは新聞記事ドメ
インなので、
新聞記事のデータに適応さ
せたい
⇒ファインチューニング

三人寄れば文殊の知恵！
他のモデルの意見も取り入
れて出力を決めたい
⇒リランキング



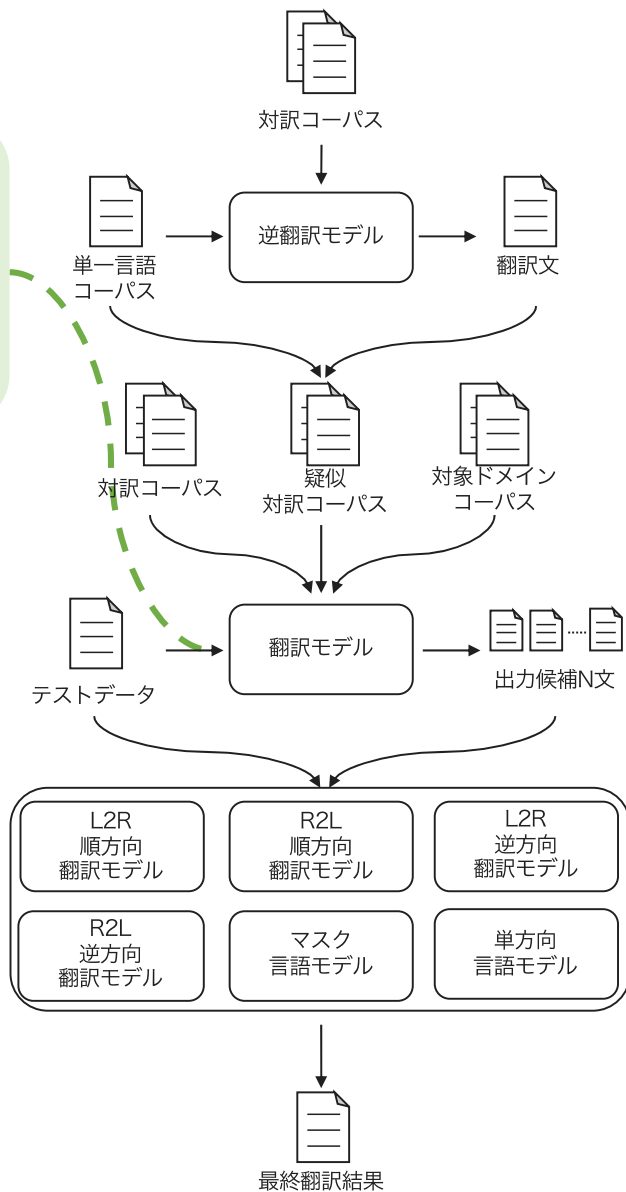
モデル学習結果にはムラが
あるので
複数のモデルを独立に訓練
して多数決しよう
⇒アンサンブル



最先端NMTは“泥臭い”技術の集合体

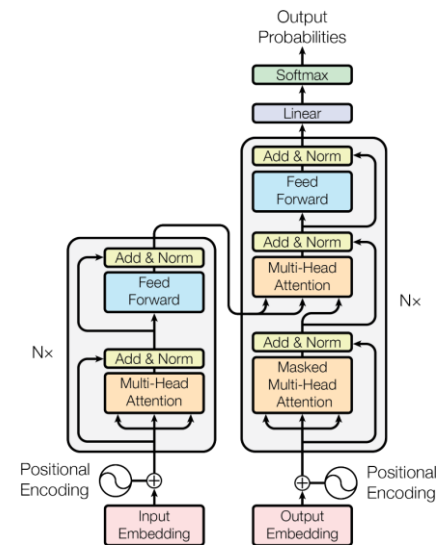


翻訳モデル(Transformer)はハイパーパラメータに
対して非常にセンシティブ
ハイパラの知見も日進月歩
⇒ハイパーパラメータの調整

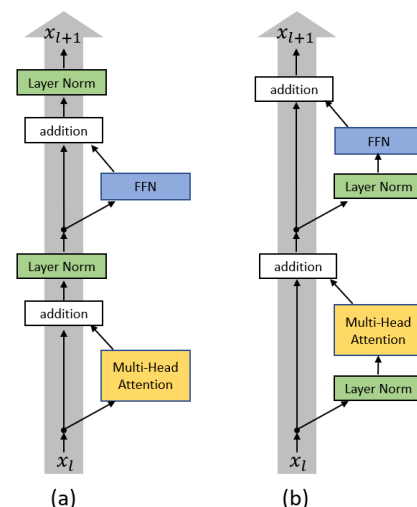


ハイパーパラメータの調整

- モデル: Transformer [Vaswani+2017]
 - 近年のデファクツ的なモデルであり、使わないという選択肢はない
 - フィードフォワード層の次元数を倍・層の数を $6 \rightarrow 9$ に設定し、より多くのデータを学習することをねらう
- 超巨大バッチサイズ [Ott+2018]
 - 通常 4,000 トークン $\rightarrow$ 512,000 トークンへ
 - 収束速度UP・汎化性能向上
 - 経験的に学習も安定する
 - Update delay (別名 ゴーストバッチ) を活用することで実現
- 巨大学習率 [Ott+2018]
 - Adamのステップサイズ $0.0005 \rightarrow 0.001$
 - 収束速度UP
 - 巨大バッチサイズとの組み合わせが非常に重要
- チェックポイント平均法
 - 適当な単位(例: 毎Epoch, 2k Updates)でモデルを保存しておく
 - 学習後、保存したモデルの平均値を計算し、推論に用いる
 - BLEUスコアが0.1~0.2ほど改善する [Popel+2018]
- Pre-layer-normalization
 - フィードフォワード層とアテンション層の前でLayerNormを計算
 - 多層Transformerの学習が安定するとの報告 [Xiong+2020]

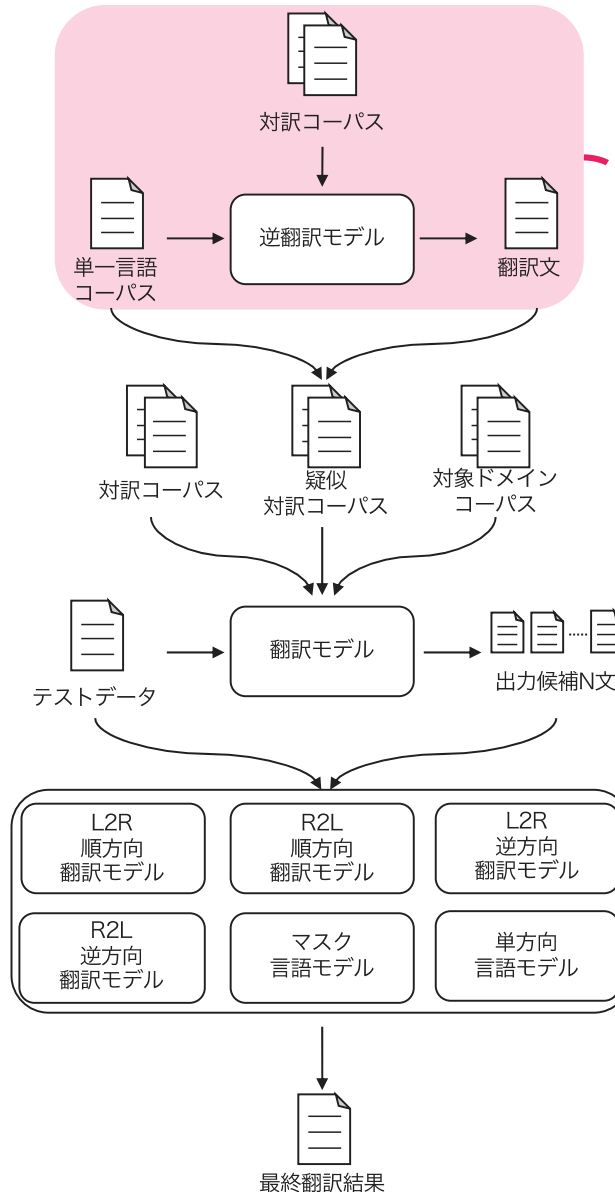


[Vaswani+2017]より引用



[Xiong+2020]より引用

最先端NMTは“泥臭い”技術の集合体



対訳コーパスだけでは
データが足りない
単一言語コーパスを使ってデ
ータを増やしたい
⇒逆翻訳によるデータ拡張

逆翻訳によるデータ拡張

- More data, better performance
 - 自然言語処理における標語
- もっと多くの対訳データがあれば性能向上するはず
- とはいえ、どうやってデータを増やすのか？

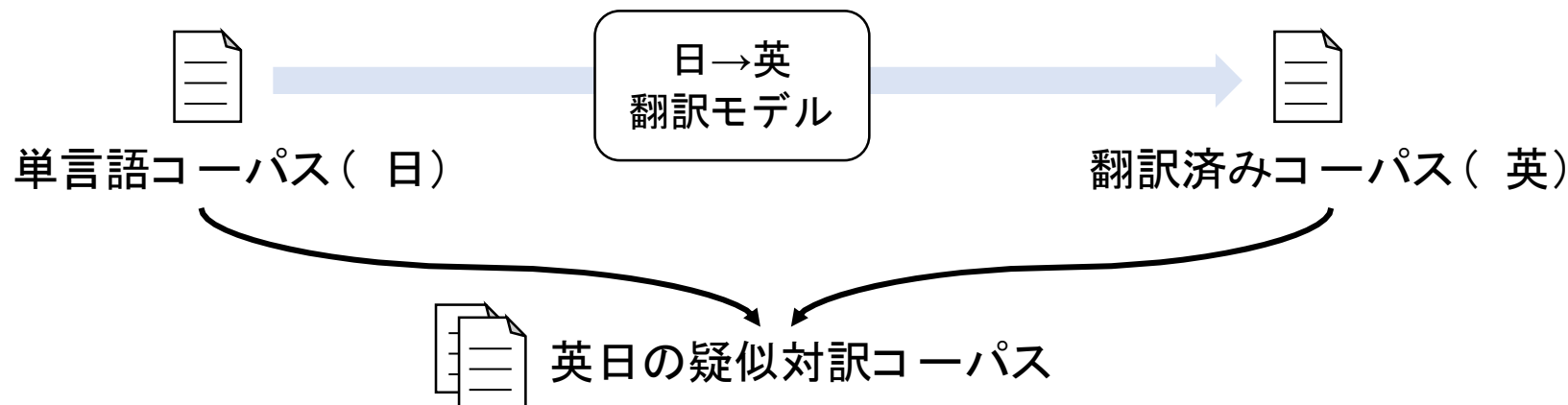


逆翻訳

逆翻訳とは何か？

- 逆翻訳 (Back-translation; BT) [Sennrich+2016]
 - 単一言語コーパスから疑似対訳コーパスを生成するための方法論
 - NMT用データ拡張のデファクト的な存在
- 逆翻訳モデルを用いて目的言語の文を原言語に「逆」翻訳

英日モデルを作る場合...

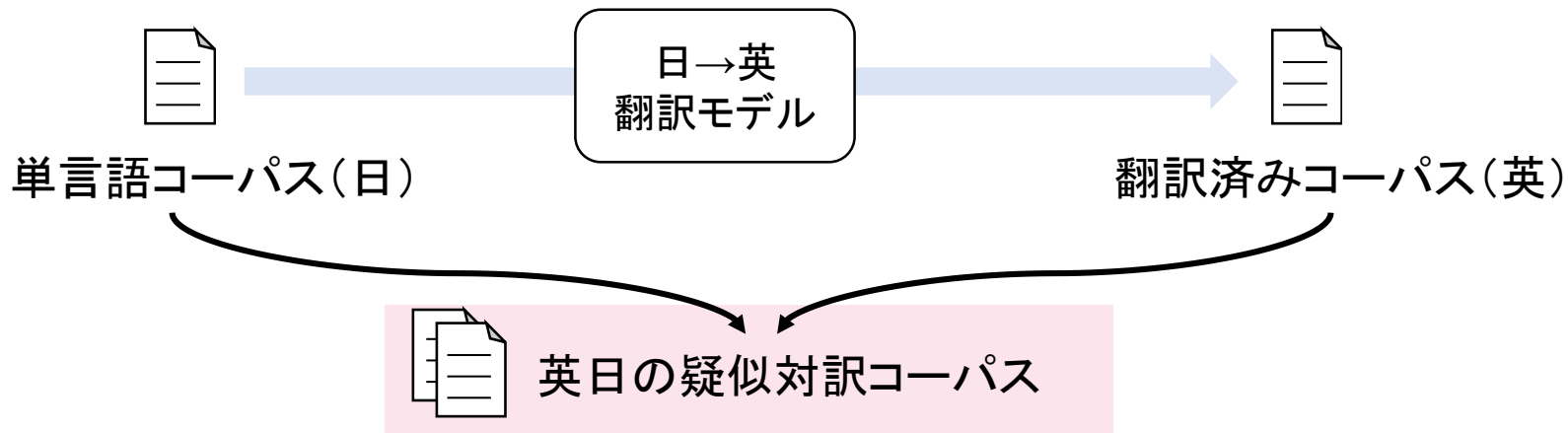


逆翻訳のプロセス

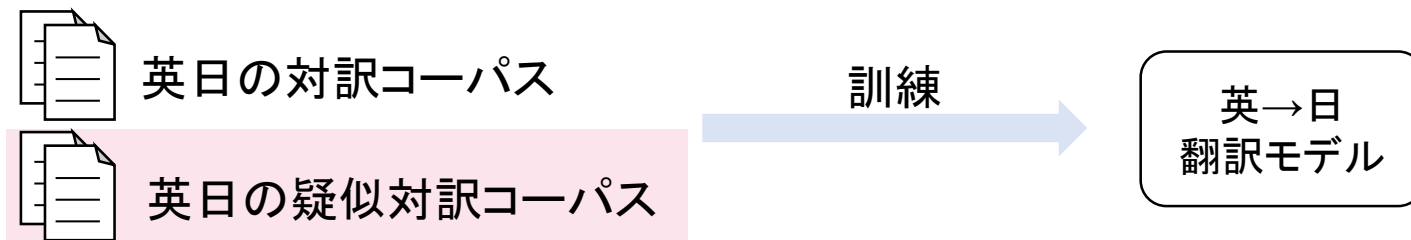
① 逆翻訳モデルの訓練



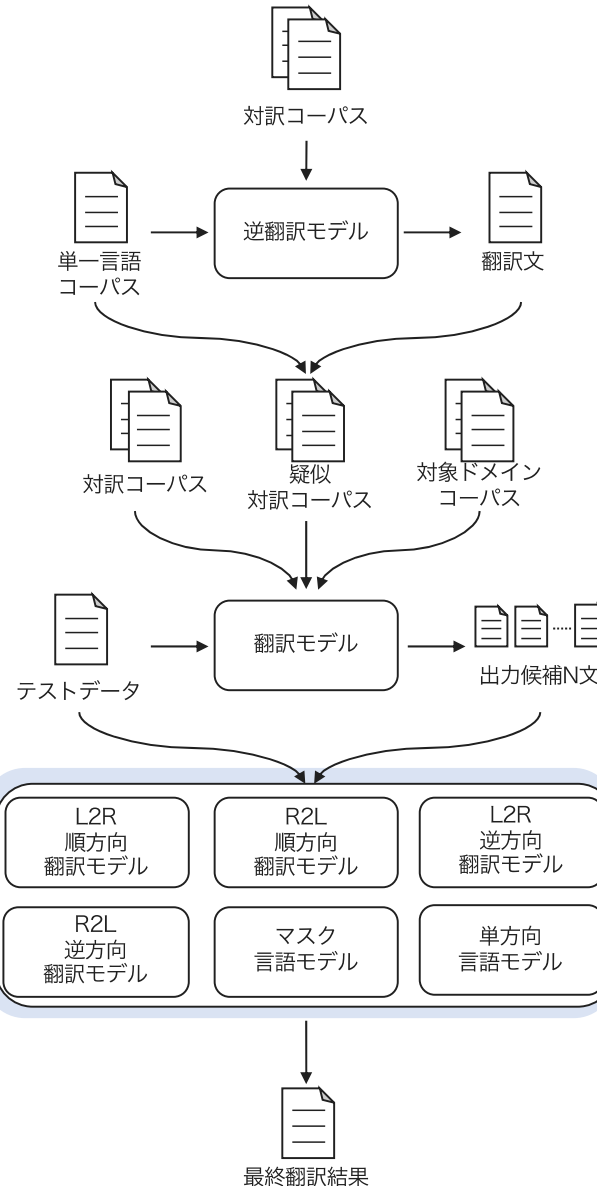
② 日本語単言語コーパスを翻訳し、疑似データを生成



③ 疑似データを用いて学習

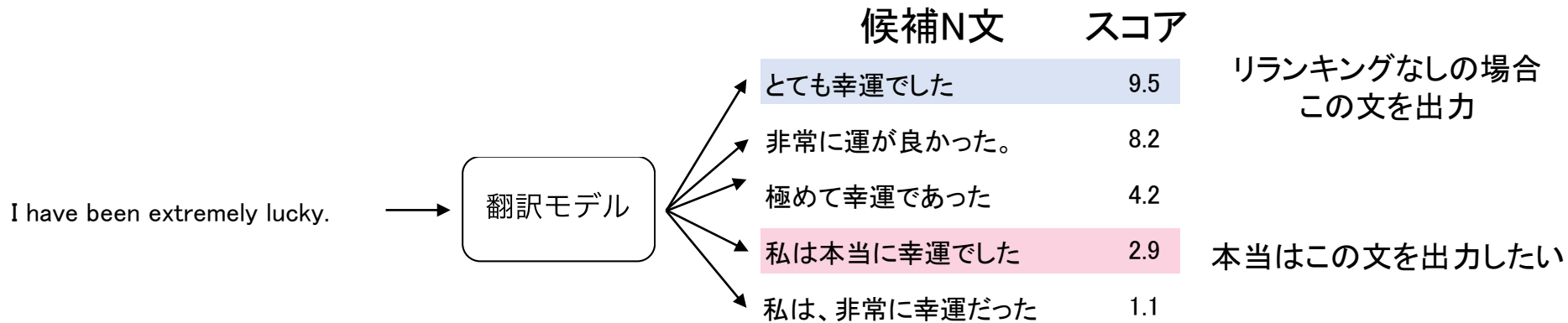


最先端NMTは“泥臭い”技術の集合体



三人寄れば文殊の知恵！
他のモデルの意見も取り入れて出力を決めたい
⇒リランキング

リランキングで候補文から良い翻訳を選ぶ



- リランキング無しの場合

1. ビームサーチによって候補文をN文生成
2. スコアの最も高い1文を出力

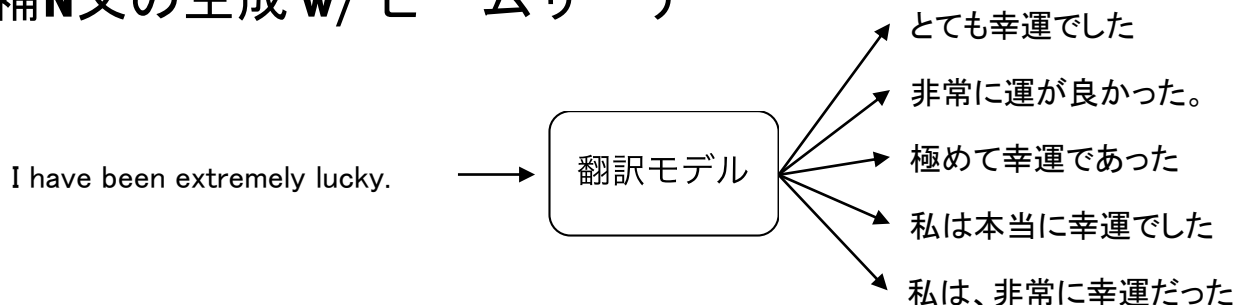
- 高スコア = 最も良い翻訳 ではない

- 他の候補文のほうが良い翻訳になっている可能性

- リランキング: 良い翻訳を見つけ出すための後処理

モデルの集合知で良い翻訳を目指す

① 候補N文の生成 w/ ビームサーチ



② N文を各モジュールでスコア付け + スコアの合計でソート



単方向言語モデル

逆翻訳モデル

双方向言語モデル 

逆方向翻訳モデル

などなど...

実験結果

ID	Setting	En→De	De→En	En→Ja	Ja→En
(a)	ベースライン	42.4	42.0	19.7	21.6
(b)	ベースライン+逆翻訳	42.7	42.5	22.0	23.9
(c)	(b)+ファインチューニング	44.9	42.3	23.1	24.4
(d)	(c) x 4 (アンサンブル)	45.5	42.8	23.9	25.4
(e)	(d)+リランキング	45.7	43.8	24.9	26.2
-	前年度の優勝システム	44.9	42.8	-	-

- 各技術の組み合わせで性能向上を達成
- 最先端の性能を出すためには複雑なシステムが必要

分かったこと②
膨大なリソースが必要

構築したシステム：色々な物が増えている

- 訓練データの量が増えている
 - 通常：最大で **5M** 程度
 - 今回：英独では **372M**
- モデルのパラメータが増えている
 - 通常：エンコーダとデコーダでそれぞれ**6層**ずつ
 - 今回：それぞれ**8層**
- モデルの数が増えている
 - アンサンブル・リランキング用に複数のモデルが必要
 - 各言語で8モデル必要 → 合計**32モデル**

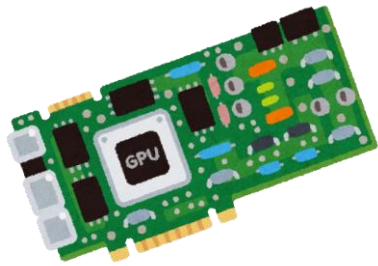


システム構築に必要なリソースの増加

必要なものは、すごいGPUと時間

1個のモデルを作るために必要なもの

GPU



V100 32GB (一枚100万円)

時間

16日間

16日間で終わるなら
待てなくもないけど…

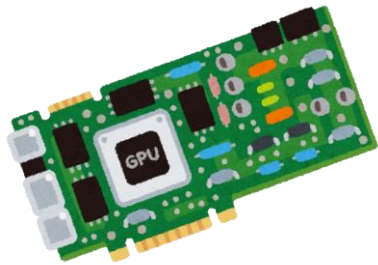
実は32モデル
必要なので…



必要なものは、すごいGPUと長い時間

32個のモデルを作るために必要なもの

GPU



V100 32GB (一枚100万円)

時間

512日間

一年以上必要なんて
コンペが終わるでしょ

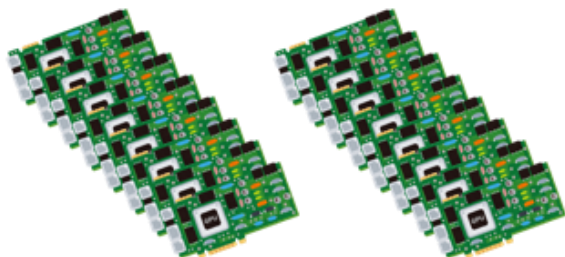


並列化して速くします…

必要なものは、**かなりすごいGPU**と**そこそこ**の時間

32個のモデルを作るために必要なもの

GPU



DGX-2

V100 16枚刺しマシン (????万円)

時間

32日間

つまりDGX-2が10台あれば
3日で終わるんだね

その通りですが今度は…



必要なものは、結局お金

- DGX-2 相当のマシンは AWS で 60ドル/hour
 - つまり、1モデル作るのに1440ドル必要
 - つまり、32モデル作るのに46080ドル必要
 - これは日本円にして500万円弱
- 途中の試行錯誤にかかった金額を合わせると
?千万円ほどかかった計算
 - あくまでもAWSでGPUを借りた場合の概算
 - 所属組織のマシンを使ったため、実際の金額とは異なります

結局いくら使ったの？

秘匿事項です

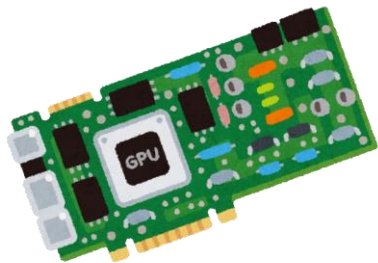


(余談) いつだって上には上がっている

- GPT-3を作るために必要なもの

- データは <https://lambdalabs.com/blog/demystifying-gpt-3/> より引用

GPU



V100 32GB (一枚100万円)

時間

355年

ご先祖様の作り始めた
GPT-3が来週完成です



この前出たGPT-371の
間違いとかではなく？

分かったこと③
“泥臭い”**NMT**の先に
新しい世界が見えてきている

ふたたび、実験結果

ID	Setting	En→De	De→En	En→Ja	Ja→En
(a)	ベースライン	42.4	42.0	19.7	21.6
(b)	ベースライン+逆翻訳	42.7	42.5	22.0	23.9
(c)	(b)+ファインチューニング	44.9	42.3	23.1	24.4
(d)	(c) x 4 (アンサンブル)	45.5	42.8	23.9	25.4
(e)	(d)+リランキング	45.7	43.8	24.9	26.2
–	前年度の優勝システム	44.9	42.8	–	–

観察1：逆翻訳は効果なし？

ID	Setting	En→De	De→En	En→Ja	Ja→En
(a)	ベースライン	42.4	42.0	19.7	21.6
(b)	ベースライン+逆翻訳	42.7	42.5	22.0	23.9
(c)	(b)+ファインチューニング	44.9	42.3	23.1	24.4
(d)	(c) x 4 (アンサンブル)	45.5	42.8	23.9	25.4
(e)	(d)+リランキング	45.7	43.8	24.9	26.2
–	前年度の優勝システム	44.9	42.8	–	–

- 逆翻訳で訓練データは約10倍に
- 英独ではBLEUスコアはほとんど向上せず
 - 例えば 42.4 → 42.7
 - 性能ゲインが労力に見合っていない…
- More data, ~~better~~ same performance なのか？

逆翻訳の効果はBLEUで測れない

- 最先端NMTにおける逆翻訳の効果について
 - [Edunov+2020] [Bogoychev+2019]

BLEU

逆翻訳あり $\doteq$ 逆翻訳なし

人手評価

逆翻訳あり $>$ 逆翻訳なし

疑似対訳コーパスでの学習により
出力が流暢になっている
[Edunov+2020]

😄 逆翻訳は無駄ではなかった

😞 人手評価とBLEUが相関しない世界になっている…

観察2：リランキングの効果が薄い

ID	Setting	En→De	De→En	En→Ja	Ja→En
(a)	ベースライン	42.4	42.0	19.7	21.6
(b)	ベースライン+逆翻訳	42.7	42.5	22.0	23.9
(c)	(b)+ファインチューニング	44.9	42.3	23.1	24.4
(d)	(c) x 4 (アンサンブル)	45.5	42.8	23.9	25.4
(e)	(d)+リランキング	45.7	43.8	24.9	26.2
–	前年度の優勝システム	44.9	42.8	–	–

- リランキングには、倍以上のモデルが必要
 - つまり、倍以上のお金がかかっている
- しかし、BLEUスコアが思ったように向上しない
 - 例えば 45.5 → 45.7
 - またしても性能ゲインが労力に見合っていない…

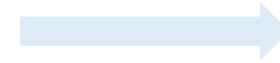
解けない問題を解こうとしている？

- 原文と候補文からだけでは良い翻訳文を判断できないのでは？
- 人間にも「良い翻訳文」はわからない
 - 判断するだけの情報が足りていない
 - どんな情報があれば良いか ⇒ 文脈？



翻訳システム

和平プロセスに影響を及ぼしたくはない
和平プロセスに影響を与えたくありません。
和平プロセスに影響を及ぼして欲しくない
和平プロセスに影響を与えたくないのです。
和平プロセスに影響が出ないようにしたい。
和平プロセスに影響を及ぼしたくありません
和平プロセスに影響を与えることは望まない。
和平プロセスに影響を与えることを望まない。



人間

まとめ

- 最先端のNMTシステム構築からわかったこと3つを紹介
 - ① 最先端のNMTシステムは泥臭い
 - ② 膨大なリソースが必要
 - ③ “泥臭い”NMTの先に新しい世界が見えてきている
- 日本の組織でも、リソースがあれば世界は取れる！

参考文献

- [Avramidis+2020]: Avramidis, E., Macketanz, V., Lommel, A., & Uszkoreit, H. (2018). Fine-grained evaluation of Quality Estimation for Machine translation based on a linguistically motivated Test Suite. In Proceedings of the AMTA 2018 Workshop on Translation Quality Estimation and Automatic Post-Editing (pp. 243–248). Association for Machine Translation in the Americas.
- [Vaswani+2017]: Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A., Kaiser, ., & Polosukhin, I. (2017). Attention Is All You Need. In *Advances in Neural Information Processing Systems 31 (NIPS 2017)* (pp. 5998–6008).
- [Ott+2018]: Ott, M., Edunov, S., Grangier, D., & Auli, M. (2018). Scaling Neural Machine Translation. In Proceedings of the Third Conference on Machine Translation: Research Papers (pp. 1–9). Association for Computational Linguistics.
- [Xiong+2020]: Ruibin Xiong, Yunchang Yang, Di He, Kai Zheng, Shuxin Zheng, Chen Xing, Huishuai Zhang, Yanyan Lan, Liwei Wang, & Tie-Yan Liu (2020). On Layer Normalization in the Transformer Architecture. In Proceedings of the 37th International Conference on Machine Learning, ICML 2020, 13–18 July 2020, Virtual Event (pp. 10524–10533). PMLR.
- [Sennrich+2016]: Sennrich, R., Haddow, B., & Birch, A. (2016). Improving Neural Machine Translation Models with Monolingual Data. In Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers) (pp. 86–96). Association for Computational Linguistics.
- [Edunov+2020]: Edunov, S., Ott, M., Ranzato, M., & Auli, M. (2020). On The Evaluation of Machine Translation Systems Trained With Back-Translation. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (pp. 2836–2846). Association for Computational Linguistics.
- [Bogoychev+2019]: Nikolay Bogoychev, & Rico Sennrich (2019). Domain, Translationese and Noise in Synthetic Data for Neural Machine Translation CoRR, abs/1911.03362.
- [Freitag+2020]: Freitag, M., Grangier, D., & Caswell, I. (2020). BLEU might be Guilty but References are not Innocent. In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP) (pp. 61–71). Association for Computational Linguistics.

産業日本語特許文書分科会活動報告

特許明細書の標準化に向けた取り組み

～特許文書品質特性モデル～

IRD国際特許事務所

谷川英和

2021/2/26

目次

1. 背景
2. 関連研究
3. 特許文書品質特性の概要
4. 特許文書品質特性を用いた評価
5. 評価目的・評価シチュエーション別の重要度
6. 各特性の事例
7. 特許文書品質の教育用テキスト
8. 今後

産業日本語研究会 特許文書分科会 メンバー

		(敬称略)
	氏 名	所 属
委員／ 主 査	谷川 英和	IRD国際特許事務所 所長・弁理士
委 員	安彦 元	ミノル国際特許事務所 所長 弁理士
委 員	黒川 恵	阿部・井窪・片山法律事務所 弁理士
委 員	久保田真司	大阪工業大学 非常勤講師
委 員	杉尾 雄一	弁護士法人内田・鮫島法律事務所 弁護士・弁理士
委 員	的場 成夫	有限会社 夢屋 代表取締役 弁理士
委 員	長部 喜幸	一般財団法人日本特許情報機構 特許情報研究所 調査研究部長
委 員	久々宇 篤志	一般財団法人日本特許情報機構 特許情報研究所 研究企画課長
事務局	荻野 孝野	株式会社日本システムアプリケーション 言語処理グループ 主任研究員
事務局	塙 金治	一般財団法人日本特許情報機構 特許情報研究所 研究管理部
事務局	三吉 秀夫	株式会社日本システムアプリケーション

1. 背景

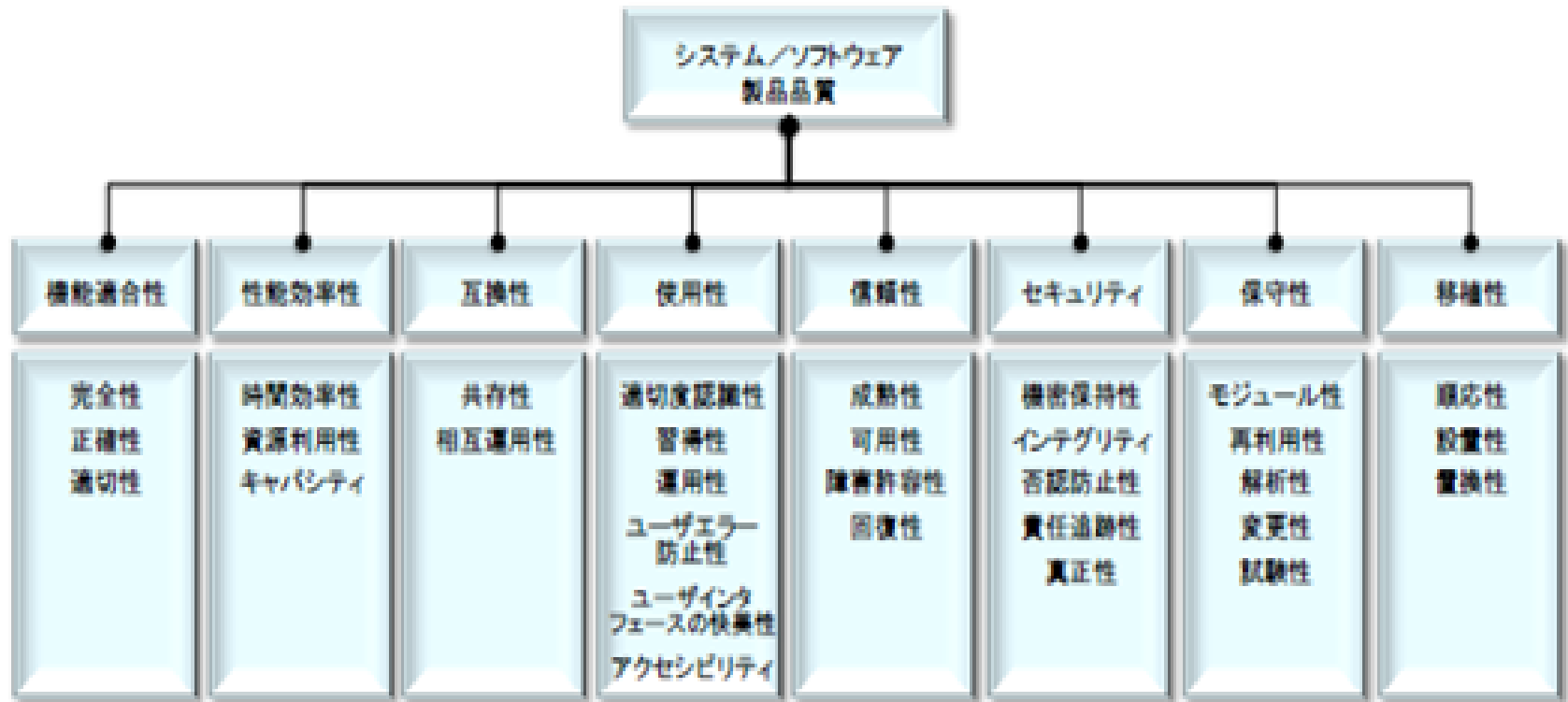
人、立場、シチュエーション等(例えば、出願前の事務所、中間処理時の出願人、侵害訴訟時の競合企業等)によって、特許文書の評価は異なっている



特許文書の品質の標準化が必要

特許に関する活動が加速

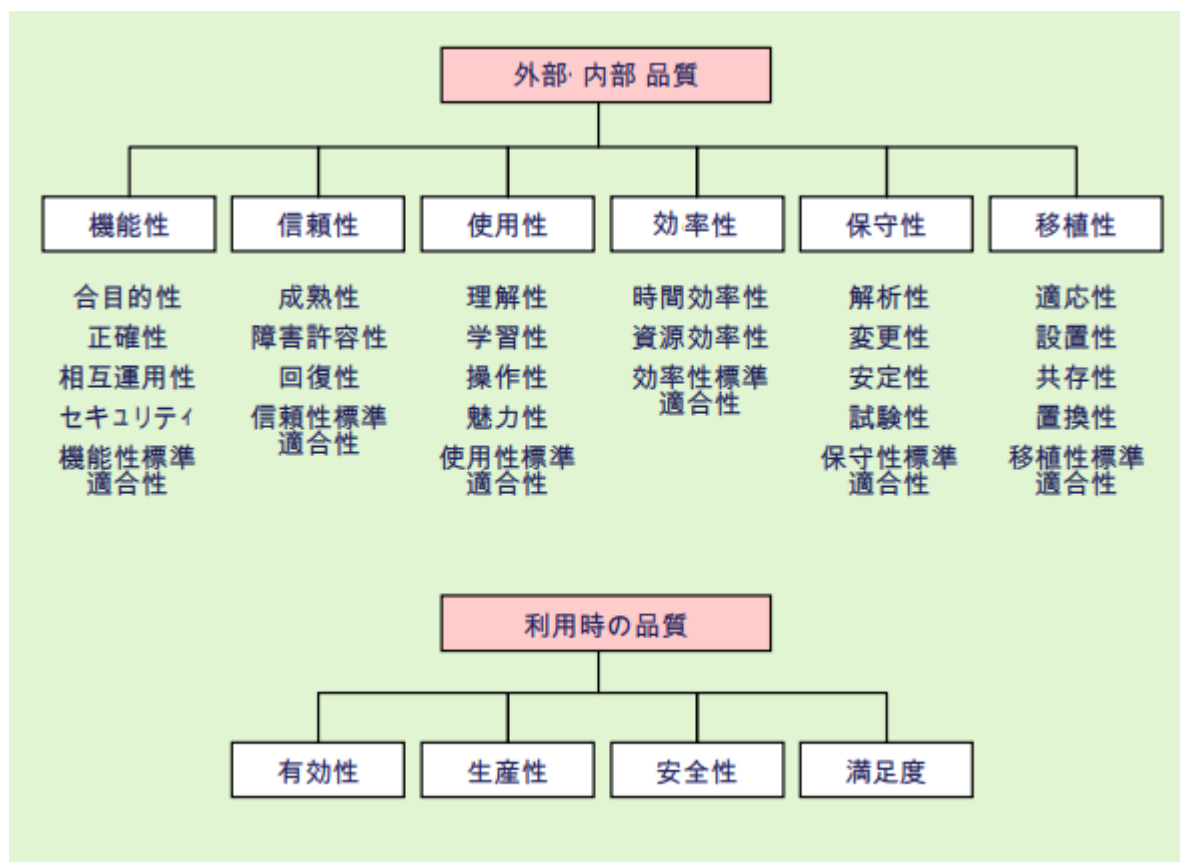
2. 関連研究(1)



システム・ソフトウェア製品の標準品質モデル

(参考)システム開発文書品質研究会(ASDoQ:Association of System Documentation Quality):『システム開発文書品質モデルVer.1.0』, 2015.05.
URL:http://asdoq.jp/common/fckeditor/editor/filemanager/connectors/php/transfer.php?file=/uid000003_E69687E69BB8E59381E8B3AAE383A2E38387E383AB5F3135303531312E706466

2. 関連研究(2)

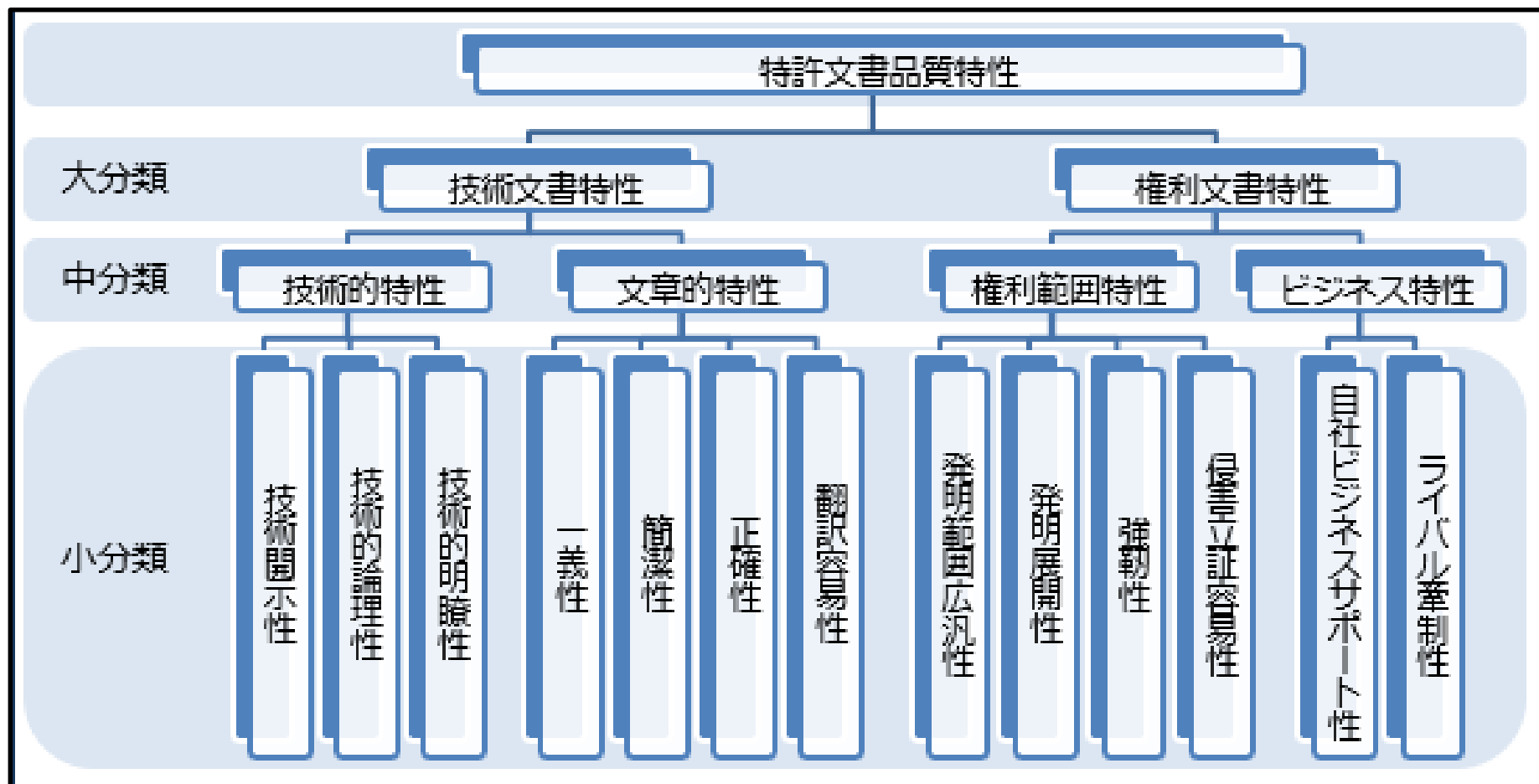


ソフトウェアのための標準品質モデル
(JIS X 0129-1 (ISO/IEC 9126-1)の品質モデル)

(参考)経済産業省 システム及びソフトウェア品質の見える化、確保及び向上のためのガイド, URL:

https://www.meti.go.jp/policy/it_policy/softseibi/metrics/product_metrics.pdf#search=%27%E3%82%BD%E3%83%95%E3%83%88%E3%82%A6%E3%82%A7%E3%82%A2+%E5%93%81%E8%B3%AA%27

3. 特許文書品質特性の概要(1)



3. 特許文書品質特性の概要(2)

品質特性			品質特性の説明
大分類	中分類	小分類	
技術文書特性	技術的特性	技術開示性	発明が十分に説明されている度合い。特許法が要求しているサポート要件、実施可能要件を満足している度合い。
		技術的論理性	論理的である度合い（例えば、クレームと明細書のストーリーとの技術的因果関係が明瞭であること、背景・従来技術・課題・解決手段・効果等のストーリーの筋が通っている度合い）。
		技術的明瞭性	発明の技術的な説明内容が明瞭である度合い。
	文章的特性	一義性	特許文書を構成する文章が一義的に捉えられる度合い。
		簡潔性	特許文書を構成する各文が簡潔である度合い。
		正確性	特許文書を構成する各文に誤りがない度合い。
		翻訳容易性	翻訳のし易さの度合い。
権利文書特性	権利範囲特性	発明範囲広汎	発明の本質が抽出されており、無用な限定が無い度合い。権利範囲の広さの度合い。
		発明展開性	発明が十分に展開されている度合い。
		強靱性	拒絶、無効になりにくい度合い。
		侵害立証容易性	侵害の立証が容易である度合い。
	ビジネス特性	自社ビジネスサポート性	自社製品をカバーしている度合い。
		ライバル牽制性	ライバルを牽制できている度合い。

4. 特許文書品質特性を用いた評価①

～ 定性的評価 ～

小分類	品質特性の説明	評価方法(定性的)
技術開示性	発明が十分に説明されている度合い。	・特許請求の範囲において、全ての実施形態が含まれるように記載されているか？ ・請求項に記載の発明特定事項(発明の本質)について、明細書中に十分な記載があるか？
技術的論理性	論理的である度合い	・背景・従来技術・課題・解決手段・効果のストーリーの筋が通っているか？ ・明細書の文章全体の構造が明確に把握できるか？
技術的明瞭性	発明の説明が明瞭である度合い	・一般名称、専門用語が正しい意味で使用されているか？ ・造語や略語は、定義が記載されているか？
一義性	多義的に解釈できない度合い	・定義の無い多義的な用語・文が使用されていないか？
簡潔性	簡潔な文の度合い	・用語の統一性が図れているか？ ・冗長用語、冗長表現(例:制御を実行する)が使用されているか？
正確性	誤記・技術的誤りが出現しない度合い	・用語が統一されているか？ ・明細書中の符号と図面の統一が統一されているか？
翻訳容易性	翻訳しやすい度合い	・主語があるか？ ・長文や複文が多いか？
発明範囲 広汎性	権利範囲の広さの度合い	・上位概念の用語が用いられているか？ ・明細書中に、発明特定事項の定義が限定的に記載されていないか？
発明展開性	発明が十分に展開されている度合い。	・独立項が多いか？(独立項が多いと、訂正要件を満たす可能性が高くなる。) ・クレームに包含される例が豊富であるか？
強靱性	拒絶、無効になりにくい度合い	・クレーム文言が明瞭でありクレーム文言と実施例との間に齟齬がないか？ ・クレームに包含される例が豊富であるか？
侵害立証 容易性	侵害の立証が容易である度合い	・発明特定事項を実施する主体が複数になっていないか？ ・クレームに包含される例が豊富であるか？
自社ビジネス サポート性	自社製品をカバーしている度合い	・自社製品・自社サービスに対応した特許請求の範囲、図面、フローチャートが存在するか？ ・収益を守るための請求項が記載されているか？
ライバル 牽制性	ライバルを牽制できている度合い	・他社製品・他社サービスと同一または近似した特許請求の範囲、図面が存在するか？ ・他社製品・他社サービスと同一または近似した用語が使用されているか？

4. 特許文書品質特性を用いた評価②

～ 定量的評価 ～

小分類	品質特性の説明	パラメータ(定量的)	
		数量	割合
技術開示性	発明が十分に説明されている度合い。	・実施の形態の数、実施例の数 ・図面数	・実施例でサポートされている請求項の割合 ・請求項の数値範囲をカバーするデータの割合
技術的論理性	論理的である度合い	・クレームツリーの枝にぶら下がっている請求項に番号飛びが存在する場合の飛びの数	・発明特定事項と効果の組が記載されている割合 ・請求項の発明特定事項の例示の記載率
技術的明瞭性	発明の説明が明瞭である度合い	・造語や略語の定義の数	・請求項の造語の定義率 ・請求項の発明特定事項の例示の記載率
一義性	多義的に解釈できない度合い	・多義的な用語の数 ・多義的な文の数	・単文率または複文率 ・能動態または受動態の使用率
簡潔性	簡潔な文の度合い	・長文、複文の数 ・重複記載の箇所の数	・冗長用語、冗長修飾詞の使用率 ・能動態の使用率
正確性	誤記・技術的誤りが出現しない度合い	・誤記、誤字、脱字の数 ・技術的誤りの数	・用語の統一率 ・誤記、誤字、脱字の発生率
翻訳容易性	翻訳しやすい度合い	・主語の無い文、長文、複文の数 ・助詞の誤使用の数	・長い複合詞の使用率 ・日本語特有の不明確になり易い用語の使用率
発明範囲 広汎性	権利範囲の広さの度合い	・独立項の発明特定事項数、文字数、発明特定事項に対する修飾数、格成分数* ・1つの請求項についての課題や効果の数	
発明展開性	発明が十分に展開されている度合い。	・クレーム数 ・クレームツリーのネストレベル(深さ)	
強靱性	拒絶、無効になりにくい度合い	・背景技術の先行文献数 ・実施例数	
侵害立証 容易性	侵害の立証が容易である度合い	・格成分数*	
自社ビジネス サポート性	自社製品をカバーしている度合い	・実施報償に関する情報、製品の売上情報等の自社固有の管理情報	
ライバル 牽制性	ライバルを牽制できている度合い	・契約情報(ライセンス契約を締結している等)	

* 格成分数・・・特許の広さ指標。特許請求の範囲の記載において各構成要素につきどれだけ限定がかけられているかを数値化したもの

5. 評価目的・評価シチュエーション別の重要度

品質特性			評価者 評価対象	特許書類の評価目的・評価のシチュエーション							
大分類	中分類	小分類		出願前		出願後・権利化前			権利化後		
				出願前の知財部の チェック	企業に提出前の 所内チェック	審査着手時	中間処理時	中間処理時	侵害訴訟	他社抑制効果	ライセンス契約
				出願人 (主に企業)	作成者側 (事務所)	審査官	出願人 (主に企業)	作成者側 (事務所)	競合企業	競合企業	競合企業・アライアンス企業
技術文書特性	技術的特性	技術開示性	明細書	A	B	A	C	C	A	B	C
		論理性	明細書 請求の範囲	A	B	A	C	C	B	C	C
		明瞭性	明細書	A	B	A	C	C	A	B	C
	文章的特性	一義性	明細書 請求の範囲	B	B	A	C	C	A	B	C
		簡潔性	明細書 請求の範囲	B	A	C	C	C	C	C	C
		正確性	明細書 請求の範囲	B	A	A	C	C	A	B	C
		翻訳容易性	明細書 請求の範囲	B	A	C	C	C	C	C	C
権利文書特性	権利範囲特性	発明範囲広汎性	請求の範囲	A	A	C	A	A	A	A	A
		発明展開性	請求の範囲	A	B	C	A	A	A	B	B
		強靱性	請求の範囲	A	B	A	A	A	A	B	C
		侵害立証容易性	請求の範囲	B	B	C	B	C	A	B	C
	ビジネス特性	ライバル牽制性	請求の範囲	B	C	C	A	C	A	A	B
		自社ビジネス貢献性	請求の範囲	A	C	C	A	C	A	A	A

6. 各特性の事例

大分類	中分類	小分類	対象特許	悪例(品質特性に関わる箇所に下線)	良例(品質特性に関わる箇所に下線)	コメント・理由
技術文書特性	技術的特性	技術開示性	2(タクシー)	図2に戻って、顧客Xが顧客端末501に表示されている空車状況情報から空車Bを選択すると、顧客Xが空車Bを選択した旨の空車選択通知とともに、顧客Xの電話番号およびメールアドレスを含む顧客特定情報が、基地局101のサーバ105の仲介コネクション106へ送信され(ステップS605)、これが <u>仲介テーブル107</u> へ記入される。	図2に戻って、顧客Xが顧客端末501に表示されている空車状況情報から空車Bを選択すると、顧客Xが空車Bを選択した旨の空車選択通知とともに、顧客Xの電話番号およびメールアドレスを含む顧客特定情報が、基地局101のサーバ105の仲介コネクション106へ送信され(ステップS605)、これが仲介テーブル107へ記入される。 <u>仲介コネクション106とは、顧客端末501と車載装置3とが情報を送受信する中継の装置である。仲介コネクション106は、例えば、クラウドサーバ、ASPサーバ等問わない。仲介テーブル107とは、顧客端末501と車載装置3とが情報を送受信するための情報が管理されるテーブルである。仲介テーブル107は、例えば、RDB、CSVファイル等であり、その構造は問わない。</u>	造語である「仲介コネクション106」「仲介テーブル107」の定義、および例示は記載しないと、技術開示性が低くなる。
		技術的論理性	1(半導体装置)	図3は、以上のようにして形成されたMOSTランジスタ100の、動作を説明する図である。(中略)図3(h)においては、ゲート電極112にゲート電圧VGとして0Vが入力されているため、 <u>MOS-FET</u> 100はオフしており、電流は流れない。	図3は、以上のようにして形成されたMOSTランジスタ100の、動作を説明する図である。(中略)図3(h)においては、ゲート電極112にゲート電圧VGとして0Vが入力されているため、 <u>MOSTランジスタ</u> 100はオフしており、電流は流れない。	「MOSTランジスタ」の用語が統一されておらず、符号100に対して、「MOS-FET」という異なる用語が使われている。
		技術的明瞭性	視力回復組成物	前記アントシアニンの <u>粒子径</u> が100nm以下にナノ化されたものであることを特徴とする請求項1～7のいずれか1項に記載の視力回復組成物。	前記アントシアニンの <u>重量平均粒子径</u> が100nm以下にナノ化されたものであることを特徴とする請求項1～7のいずれか1項に記載の視力回復組成物。	「粒子径」とは、平均粒子径を指すのかどうか不明瞭でない。また、平均粒子径を指したとしても、重量平均粒子径や数平均粒子径などのどれを指すのかどうか不明瞭でない。
	簡潔性	一義性	1(半導体装置)	例えば、図6(r)では、シリサイド化金属膜として、コバルト(Co)膜20を使用したか、これに代えて、 <u>より安価なニッケル(Ni)や白金(Pt)の膜</u> を使用してもよい。	例えば、図6(r)では、シリサイド化金属膜として、コバルト(Co)膜20を使用したか、これに代えて、 <u>より安価なニッケル(Ni)の膜</u> を使用することも可能である。また、 <u>白金(Pt)の膜</u> を使用してもよい。	「より安価な」がニッケルだけに保っているのか、白金にも保っているのか、一義的に解釈できない。
		2(タクシー)	…すると基地局101は、顧客Xの周辺にいる空車A、空車B、および空車Cを空車位置データベース102から特定して、それらの位置情報を取得する(ステップS202)とともに、これらの車両の属性情報を車両データベース103から取得し(ステップS203)、さらに顧客の周辺の地図データを地図データベース104から取得し(ステップS204)、 <u>以上の情報およびデータを合成して空車状況情報を作成し(ステップS205)、顧客Xの顧客端末501へ送信する(ステップS206)。</u>	基地局101は、顧客Xの周辺にいる空車A、空車B、および空車Cを空車位置データベース102から特定して、それらの位置情報を取得する(ステップS202)。 基地局101は、空車A、空車B、および空車Cの属性情報を車両データベース103から取得する(ステップS203)。 基地局101は、顧客の周辺の地図データを地図データベース104から取得する(ステップS204)。 基地局101は、 <u>取得した位置情報、属性情報、および地図データを合成して空車状況情報を作成する(ステップS205)。</u> 基地局101は、顧客Xの顧客端末501へ空車状況情報を送信する(ステップS206)。	複数の処理を一文で繋げるより、分割した方が簡潔である。必要でない接続詞は削除すべき。	
	文章的特性	正確性	3(視力回復組成物)	パソコンやスマートフォン等の普及により、これらの機器を長時間にわたって使用し続けることで目の疲れや痛み、更には視力の低下を訴える人が急増している。	<u>普及した</u> パソコンやスマートフォン等の機器を長時間にわたって使用し続ける人が <u>増えたことにより</u> 、目の疲れや痛み、更には視力の低下を訴える人が <u>急増している</u> 。	パソコンやスマートフォンの普及が、目の疲れの直接の原因ではない。それらの機器を長時間使用する人がいて、更にそのヒト達の中に、目の疲労を感じる人がいる。ということであるはず。

7. 特許文書品質の教育用テキスト(1)

目次案

第1章 特許文書品質特性の概要

第2章 特許文書品質特性の詳細

第3章 特許文書品質特性の事例(悪例→良例)

(1)電気 (2)情報 (3)化学 (4)機械

第4章 特許文書品質特性の利用方法(1)・・・定性評価(チェックリスト)

第5章 特許文書品質特性の利用方法(2)・・・定量評価(パラメータ)

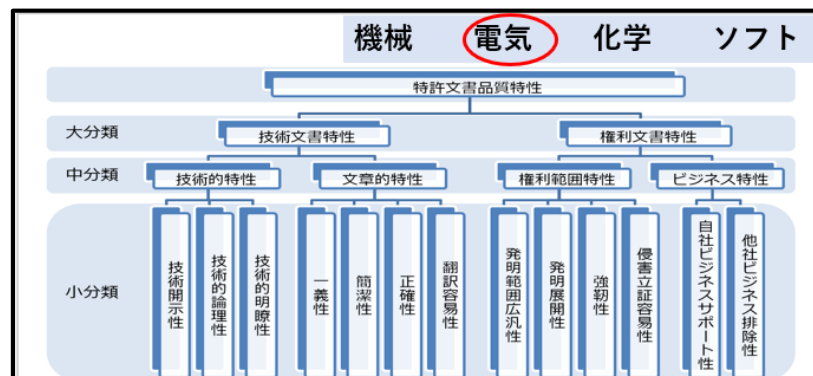
第6章 演習課題(悪例を多く含む仮想明細書に基づく)

第7章 終わりに

(付録) 仮想明細書(半導体, BM, 組成物, ライト)

7. 特許文書品質の教育用テキスト(2)

4つの技術分野ごと、13の特性ごとに、特性の理解を促す「悪例・良例」の事例を作成



特許文書中の項目 ; 請求項 明細書 (課題 実施形態)

【悪例】

【請求項1】

半導体基板上にゲート絶縁膜およびゲートポリシリコン膜を形成する工程と、前記ゲート絶縁膜およびゲートポリシリコン膜の側壁に、サイドウォール絶縁膜を形成する工程と、前記ゲートポリシリコン膜および前記サイドウォール絶縁膜をマスクとして、**砒素を、加速エネルギー1~1000keV、ドーズ量 $5 \times 10^{10} \sim 5 \times 10^{15} \text{cm}^{-2}$ の条件で、半導体基板の表面からイオン注入することにより**前記半導体基板の表面にドレイン領域およびソース領域を形成する工程と、前記半導体基板の全面を覆うように金属膜を堆積する工程と、前記半導体基板を急速熱処理する工程と、前記金属膜のうち未反応の部分を除去する工程と、を有する半導体装置の製造方法。



【良例】

【請求項1】

半導体基板上にゲート絶縁膜およびゲートポリシリコン膜を形成する工程と、前記ゲート絶縁膜およびゲートポリシリコン膜の側壁に、サイドウォール絶縁膜を形成する工程と、前記ゲートポリシリコン膜および前記サイドウォール絶縁膜をマスクとして、**イオン注入法により**前記半導体基板の表面にドレイン領域およびソース領域を形成する工程と、前記半導体基板の全面を覆うように金属膜を堆積する工程と、前記半導体基板を急速熱処理する工程と、前記金属膜のうち未反応の部分を除去する工程と、を有する半導体装置の製造方法。

【コメント・理由等】

イオン注入する元素やその注入条件の数値限定など、無意味に権利範囲を限定している。なお、この数値限定はこの発明のコアな部分ではなく、この数値範囲も一般的な値である。但しサポート要件、再現性を担保する上では、実施例等においてこれをカバーする記載か、データを入れておくべきである。

7. 特許文書品質の教育用テキスト(3)

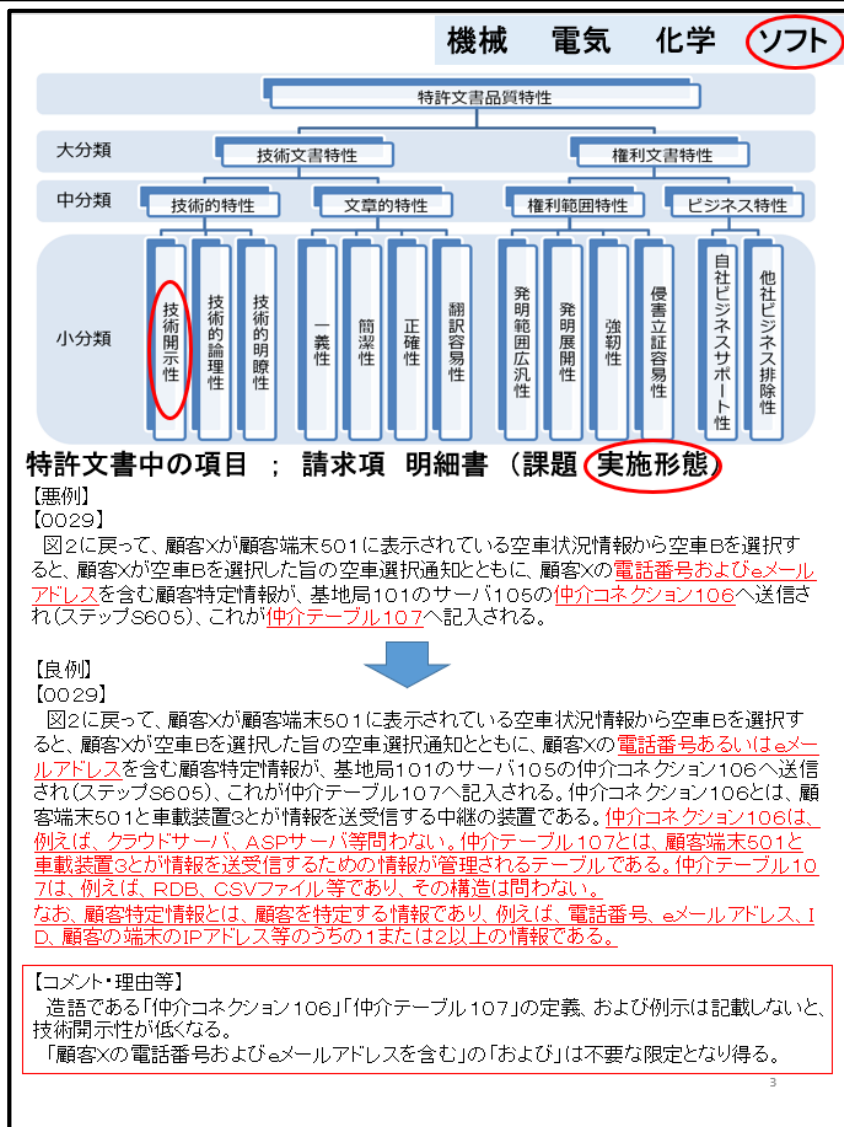
4つの技術分野ごとのテキストの作成も進める

特許文書の品質特性を
事例にて学ぶテキスト

特許文書の技術分野 ; 機械 電気 化学 ソフトウェア

7. 特許文書品質の教育用テキスト(4)

13の特性ごとに、特性の理解を促す「悪例・良例」の事例を掲載



7. 特許文書品質の教育用テキスト(5)

特許文書内での各事例の位置を明示

【発明を実施するための形態】

【実施例】

【0018】

図1は、本発明のタクシー捕捉システム1の全体構成を示す図である。基地局101、車載装置301、および顧客端末501が、インターネットのネットワーク111に接続されている。

【0019】

車載装置301は、ハードウェアとしては基本的に通常のカーナビゲーションシステムであって、本システムのためのアプリケーションプログラムをインストールしたものである。顧客端末501は、ハードウェアとしては一般のスマートフォンやタブレット端末であって、本システムのためのアプリケーションプログラムをインストールしたものである。車載装置301と顧客端末501は、それぞれ衛星121によるGPS(Global Positioning System)を使用して、自分の位置情報を時々刻々把握している。

【0020】

図2は、本発明のタクシー捕捉システム1の全体の動作を示す図である。道路を走行している、タクシーの空車A、空車B、および空車Cのそれぞれの車載装置301は、自車の位置情報を、GPSにより常時更新取得して基地局101に送り(ステップS401～S403)、基地局101は、それぞれの最新の位置情報を空車位置データベース102に保持する(ステップS201)。

【0021】

また、基地局101の車両データベース103には、予め、本システムに登録しているタクシーの車両の属性情報が格納されている。車両の属性情報として、「ハイブリッド小型車」、「黒塗り中型車」、「7人乗りワゴン車」、「〇〇のエアロ仕様の新車」といった車両自体の特徴や、「30年間無事故無違反」、「江東区の裏道に強い」といった運転手の自己PRなどの情報が格納され、さらに、各車両の車載装置301のIPアドレスが格納されている。

【0022】

顧客Xが顧客端末501上でアプリケーションプログラムを起動する(ステップS601)と、顧客端末501が顧客Xの位置情報を取得し(ステップS602)、これを基地局101へ送信する(ステップS603)。すると基地局101は、顧客Xの周辺にいる空車A、空車B、および空車Cを空車位置データベース102から特定して、それらの位置情報を取得する(ステップS202)とともに、これらの車両の属性情報を車両データベース103から取得し(ステップS203)、さらに顧客Xの周辺の地図データを地図データベース104から取得し(ステップS204)、以上の情報およびデータを合成して空車状況情報を作成し(ステップS205)、顧客Xの顧客端末501へ送信する(ステップS206)。顧客端末501のアプリケーションプログラムは、受信した空車状況情報を、顧客端末501の表示部に表示する(ステップS604)。

【0023】

図4は、顧客Xの顧客端末501の画面502に表示される空車状況情報を示す図である。顧客Xが、表示されたビル街の地図の中心付近、大通りの歩道上にいる。顧客Xは、タクシーに乗り込んで地図の上方である北方向へ行くこととしているとする。

【0024】

顧客Xのいちばん近くにいるのは、空車Aである。しかし、空車Aは、大通りを顧客Xが行きたい方向と逆の南向きに進んでいるから、これをUターンさせて北向きへと方向を変えさせるには時間がかかると予想される。

【0025】

一方、空車Bは、顧客Xから比較的遠い位置にいるが、あと2ブロック足らずを東進してから一度左折すれば顧客Xの位置に到達できるから、顧客Xは、空車Bを呼び寄せるのがいちばん早く乗車することができる選択であると判断することができる。

【0026】

そして、顧客Xが1人で急いで大通りを北へ行きたいのであれば、画面下方の車両の属性情報に表示されているような、江東区の裏道に強い運転手(空車A)や、7人乗りワゴン車(空車C)である必要はなく、運転手に特徴が見られない小型車である空車Bで十分である。

→ 強靱性

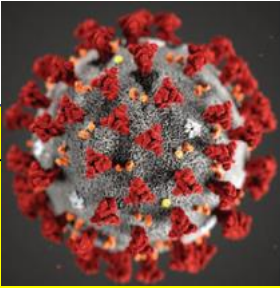
→ 技術的
論理性
正確性

→ 簡潔性

→ 技術的
論理性
正確性

8. 今後

- (1) 特許文書品質特性モデルの学習のためのテキスト完成
- (2) 特許文書品質特性モデルの普及に向けた方策の検討・実施
 - ・弁理士会の新人研修での上記テキストの使用の打診
 - ・日本知的財産協会等への講義・研修の打診
 - ・**パテント**、**日本知財学会**、知財管理、**特技懇**などへの投稿
 - ・解説書の無料配布(冊子, PDF)



COVID-19に絡む特許情報の分析

～ COVID-19(敵)を知り、己を守る～

孫氏の兵法

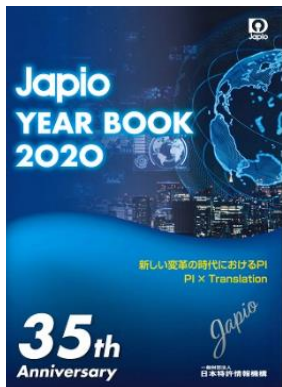
名誉あるプレゼンの機会に感謝

第6回特許情報シンポジウム

2021. 02. 26. PM

桐山 勉 (IP リサーチ フェロー)

(日本特許情報機構 特許情報研究所)



JapioYEARBOOK2020記事をベースにして

https://www.japio.or.jp/00yearbook/files/2020book/20_2_04.pdf

お断り：このプレゼン発表は、桐山の個人的な考え方であり、私が所属するいかなる組織の了解を取ったものではない。

自分事&利他

目次

1. はじめに
2. 新型コロナウイルス感染症の関連情報とは
3. **発症を**第5日目に**見逃さずに自覚**して**PCR検査**を第6日目に受ける
4. 免疫の不思議な力を学ぶ
5. 治療薬を学ぶ
6. ワクチンの早期接種を願う
7. コロナに関して**最も言いたいこと（1）（2）**
8. **参考：特許調査の極意らしきモノ**
9. **参考：私が利用する機械翻訳システム**
10. おわりに（謝辞）

1. はじめに

- 2020年は新型コロナに翻弄された。 2021年は上手に対応する年にしたい。
- **COVID-19全論文**(20万件)の**AI読解**。人間は万卷の書を読めない。
- **Open Science & Citizen Science時代**の情報収集。
- 正しく必要な**情報の選択**を基に、**自分の頭で科学的に論理的に考え、実践**する気構え。
- 完全主義を求めない。8 5 点の有用活用。80:20の法則。
- **機械翻訳の活用**により、日本語で学びのスピード・アップ。
- With/Afterコロナ時代の新常態に対応を！
- （参考）収集する情報の質を1 0 倍に高める。

収集する情報の質を10倍に高める。（私の実践法）参考

1. 情報源を選ぶ。

- 日経新聞、NHKスペシャル、TVニュース、専門家、特設サイト（WHO、CDC、他）、本。

2. 師を選ぶ（情報の洪水に溺れないための浮袋）

- ノーベル賞受賞者(山中伸弥、本庶祐)、ウイルス免疫学専門家（河岡義裕、宮坂昌之）

3. 録画し 人間は静止画像を1秒間に3-4枚しか認知できない、Snapshotで**見直す**。

- 自分の**メモ**に変換する。→**One Sheet**に描き直す。

4. 英語／日本語で**Web検索**する。芋づる方式。

5. 非特許情報を**特許情報で確認・検証**する。

10の5乗倍（10万倍）に高める可能性。

2. 新型コロナウイルス感染症の関連情報とは

- 特設サイトを選ぶ。専門家を選ぶ。NHKスペシャルから学ぶ。

- ①新型コロナウイルスの特性とは、②新型コロナウイルスの治療薬とは、③新型コロナウイルスの免疫とは、④PCR検査とは、⑤ワクチンとは、⑥善玉抗体とは、⑦感染力は、etc。

- 特性：罹患してから第5日目に症状を自覚できる。無症状感染力。重症化と軽症の分岐点は第15日目。基礎疾患者は、自覚症状が出てから3日以内に初期対応が必須である。（私の自己防衛策）（理由：獲得免疫の作用機能の低下）

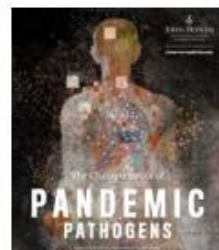
図1 COVID-19（敵）の正体を学び、One-Sheetに描く（メモ魔の習慣・趣味）

情報収集(原理原則に沿って)

No	定点観測地点	観測具体事例	備考
1	政府・官庁	WHO、国立感染症研究所、米国CDC、	ワクチン開発資料
2	学術・学会 展示会 専門誌	Nature Review Lancet 各種医学関係の専門誌	COVID-19特設 コーナー
3	新聞情報	新型コロナウイルス感染症、治療薬、ワクチン、血漿分画製剤、検査キット（PCR、抗原、抗体）日経新聞をSDI的に監視・切り抜き	日経新聞2020年3月以降に継続監視
4	Web情報	Open science & Citizen Science, Open Source Investigation等を参照する	治療薬、ワクチン、検査キット
5	Wiki情報	既存薬候補薬の治験、検査キット、会社情報	
6	シンクタンク	Johns Hopkins Univ. Oxford Univ.国立感染症研究所、大阪大学	最先端専門家集団組織
7	特許情報	ワクチン開発企業TOP20社の検索結果（DB3種以上、Derwent Innovation-Themescape, etc	新規研究

新聞・TVで得た情報をGoogle翻訳で英語表現を知り、英語でGoogle検索を実施する習慣。日本語と英語の芋づる方式検索。（74歳の老兵の新常態）

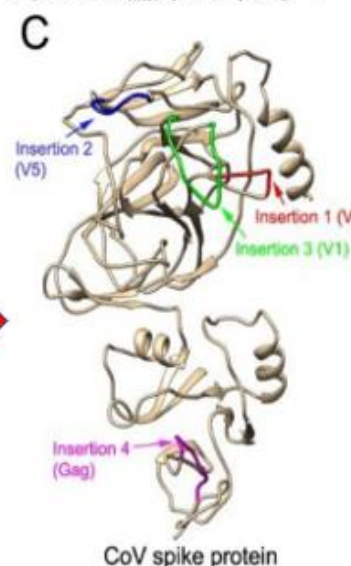
警告論文



Nature



オリジナル論文から学ぶ。



ACS



NCBI



新聞・TV から学ぶ



5/20の日経新聞記事と桐山が既に集めた情報をもとに、桐山が

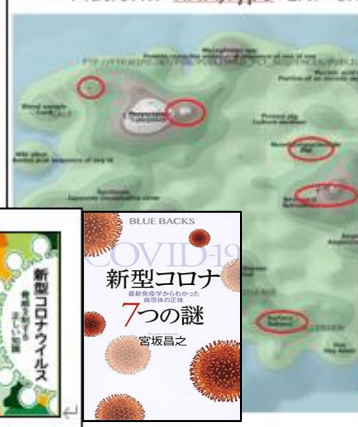
開発元 企業名	開発国	取り組み	CEPI 支援 日本	WHO 特許制 限5/19
Moderna	米国	P1, 7月大規模治験		強制無
Inovio	米国	P1, 今秋実用化		強制無
J & J	米国	7月からP1?, 9月 治験, 大量生産設備		強制無
シノバックス	中国	P2, シノバックス協力臨 床試験		中国 賛成
シノファーム北京	中国	P1, 20年末21年初 めに実用化		中国 賛成
シノファーム武漢	中国	P1, 6 成, 2 に実		
シノバックス	中国	P2, 産開		
Oxford Univ	英	21年		
Viontech-Pfizer	独-米	年内		



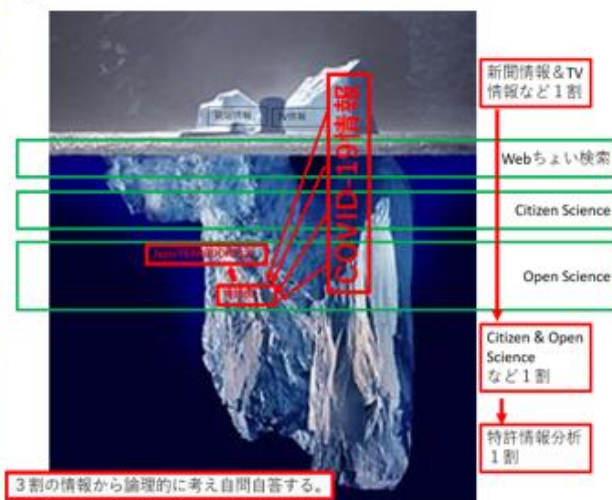
特許情報から学ぶ

Moderna社の開発

Platform=RNA, Type=LNP-en



Open-Science & Citizen Science



新型コロナ感染症の特性とは

- 無症状でも他人に感染させる。（**無症状感染力**）
- 発症が自覚できるのは、罹患後の第5日目である。
- 基礎疾患者には、インターフェロンの働きを低下させ、獲得免疫の活動を約十分の一にする。
- 発症して（第5日目）から、自宅待機中に5日以内に急激に悪化する人が多い。（ハッピー・**ハイポキシア症**＝**低酸素血症**）
- 罹患して第15日目が、軽症と重症の分岐点である。
- 重症化した人は、サイトカインストーム（免疫暴走）に陥っている。
- 感染者の**約8割**は軽症ですみ、**約2割が重症化**する可能性があり、**約数%の人が重篤化**する傾向がある。（→人間の弱点）
- まだ、特効薬が見つかっていない。

受診の目安 (2021年1月)



忽那賢志 | 感染症専門医

1/8(金) 8:06

新型コロナウイルスの感染力



ツイート



シェア



B! ブックマーク

宮坂昌之名誉教授の言葉
「このウイルスは10時間で1000倍に増殖する。」
(震えるような恐怖)

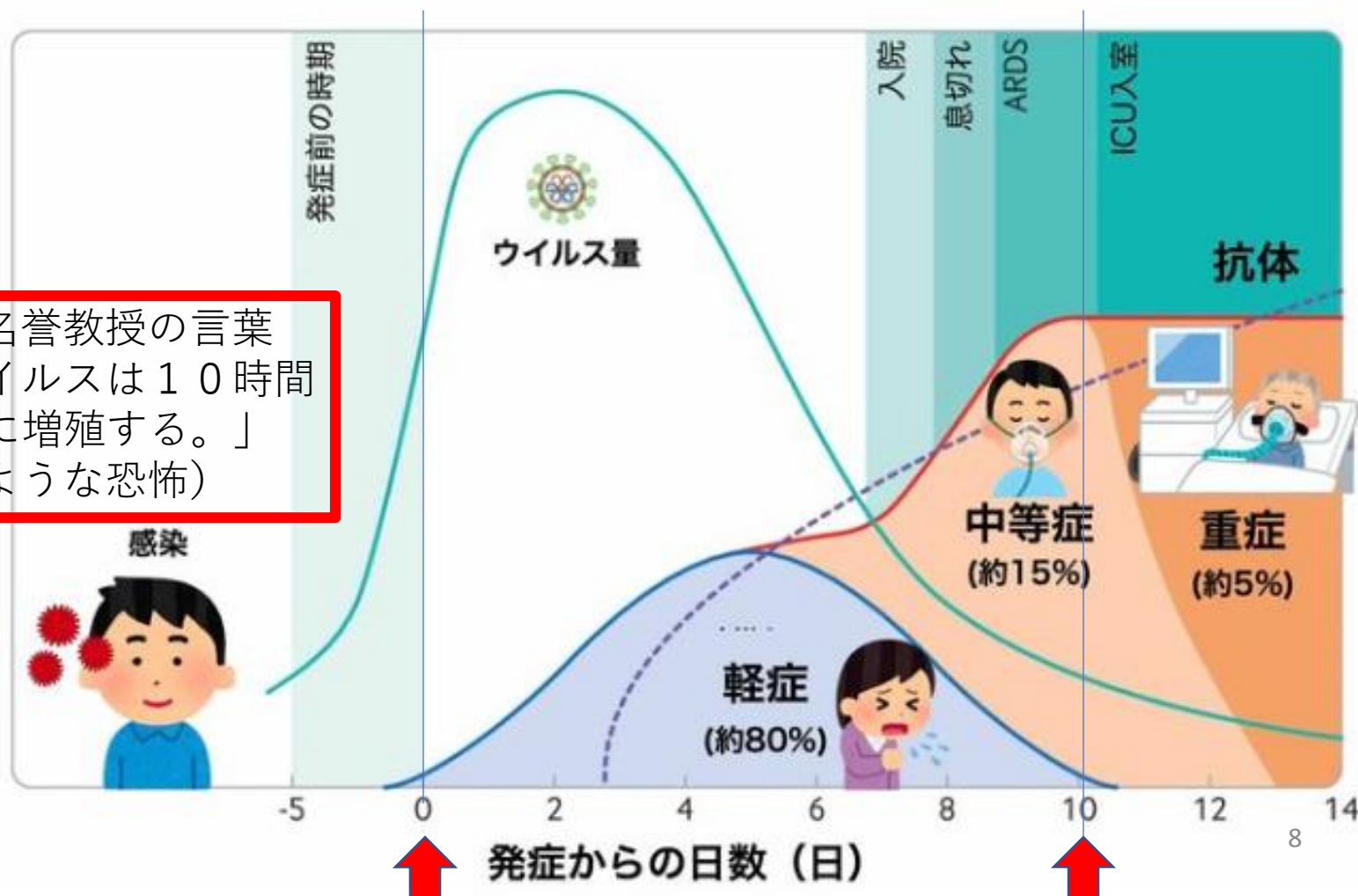


表 1 COVID-19に関する必要な情報

No	必要な情報法	備考
1	各国の 感染者数 の推移分析	人口率に対する
2	各国の対策状況	WHO,CEPI,Gavi
3	既存薬 の治験結果	免疫の仕組み
4	抗原・抗体・ PCR検査 キット	迅速・高精度
5	コンタクト・トレーシング	感染力と時期
6	IT技術を応用したBigデータ	端末移動状態
7	健康認証の政策（QRコード）	ドローン観察(中国)
8	ワクチン薬 の開発	
9	血清・血漿の分別と治療	スーパ-抗体者
10	人工呼吸器・人工心肺器	
11	専門知識の共有化	DB=CORD-19 全論文AI解析
12	重症患者の搬送・治療	
13	軍隊の医薬関連部隊の動き	
14	専門家会議とブレーン	
15	慈善団体の支援	関連Report

3. 罹患を第5日目に見逃さずに、PCR検査を受ける

検査は徹底的に、拡大できないのか？ まずは、**検査方法**を理解する。

新型コロナウイルス感染症の各種検査方法と関連企業

検査法	検査物質	現在 罹患	過去に 罹患	検査 場所	検査機器メーカー	関連特許群事例
PCR 法	ウイルス遺伝物質	○	X	粘膜	ロッシュ プレジジョン・システム・サイエンス	特開 2010-032513、特許 05775252、特表 2006-503589 WO2006/12796、特許 04193997、特許 04612037
PCR 法	〃	○	X	唾液	島津製作所 メディカロイド みから HD・・・富士レビオ 杏林製薬	特開 2020-080807、特開 2020-080806、特開 2007-074960 WO2018/159565、特開 2020-060880 WO2012/005238、WO2014/125879、特開 2012-018051
LAMP 法	ウイルス遺伝物質	○	X	唾液	キャノン・ディカル 栄研化学	特許 06400392、特許 06605238、特許 06081071
SATIC 法	ウイルス遺伝物質	○	X	唾液	塩野義製薬・日大	WO2018/168895、WO2018/020831、WO2016/152936
抗原検査	ウイルスたんぱく質	○	X	粘膜 唾液	みから HD・・・富士レビオ Quidel	WO2018/020831、特開 US2018/0187237、WO20
抗体検査 (通常抗体)	IgM 抗体	X	○	血液	日本ベイトクファーマ セルスベクト ？(免疫法) ロッシュ(ECLIA 法)	特許 06588602、特許 06

私は、毎朝、
検温、血中酸素濃度、
匂い・味の感覚を
チェックする。

日本のPCR検査は、米国、ドイツ、英国の**10分の一以下の規模**。（現状認識と理解）

PCR検査だけじゃない。ロッシュ社の測定器だけじゃない。

アイディア事例：**世田谷モデル**。保坂展人区長、児玉竜彦名誉教授。

誰でも、いつでも、何回でも、無料でPCR検査：実施中、米国ニューヨーク、独1州

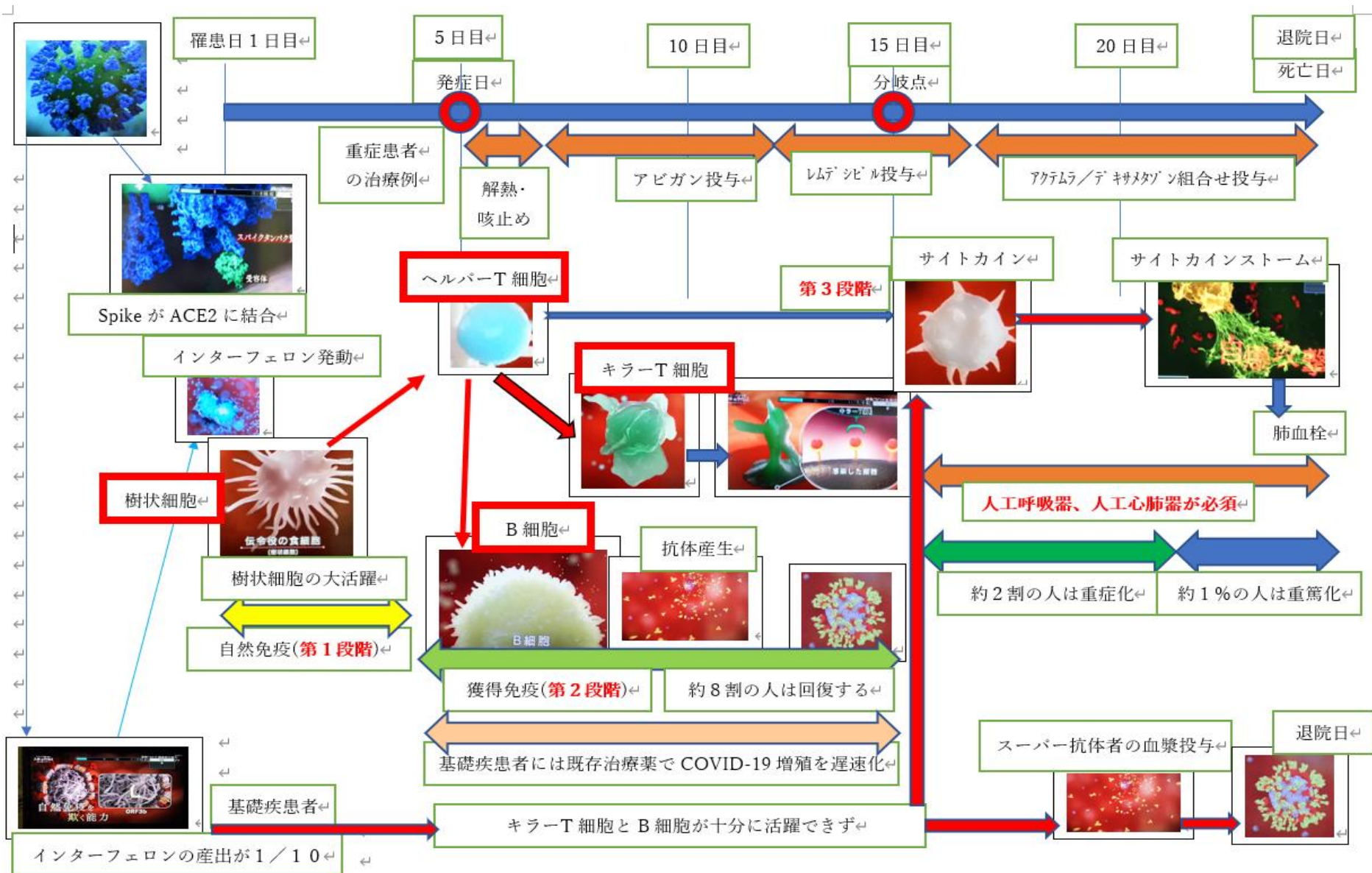
和歌山県モデル：3日前まで濃厚接触者追跡、PCR検査を実施

自覚症状が出る前に、他人を感染させる特性：重要な知見。要注意。

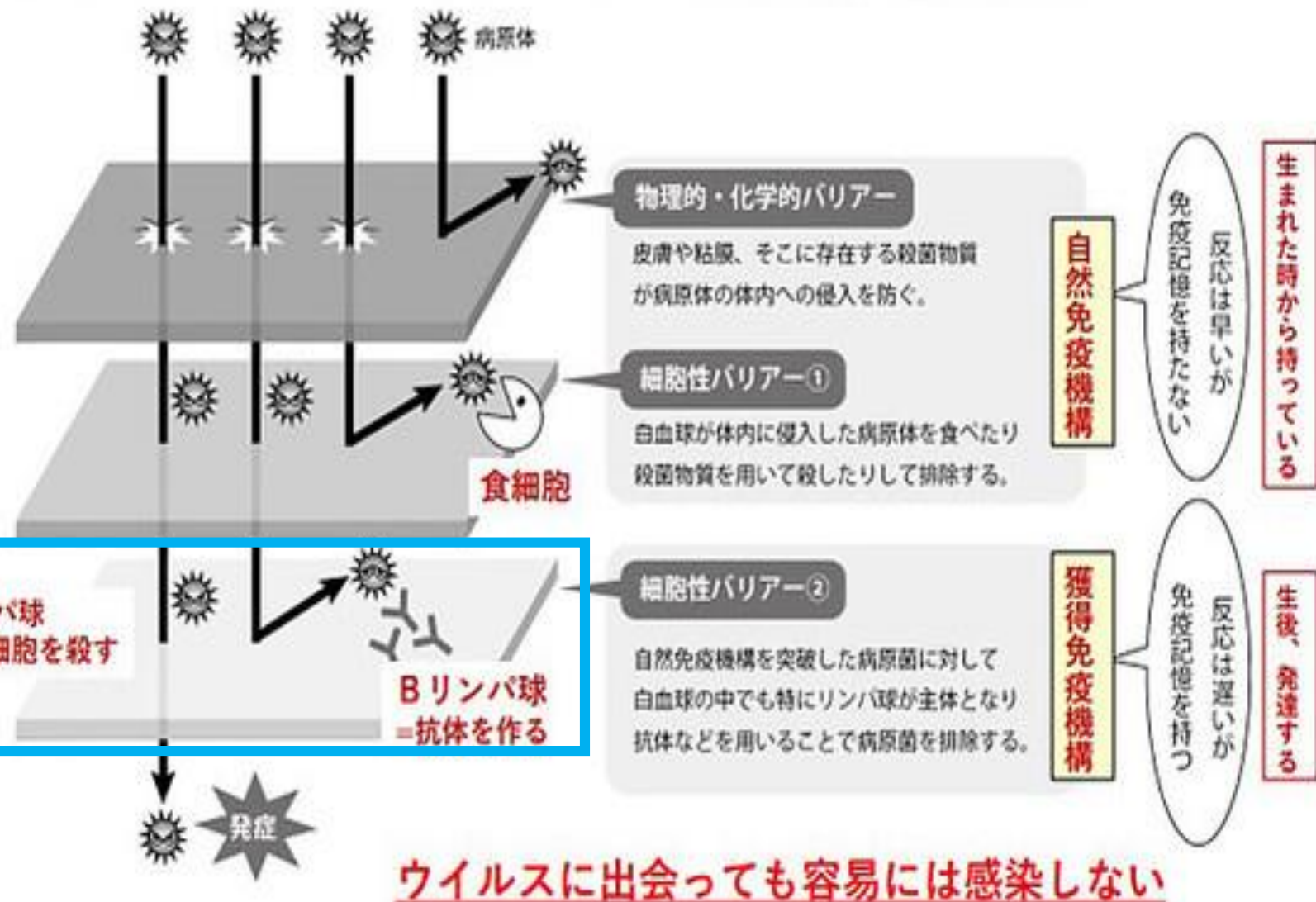
基礎疾患者の獲得免疫の働きを10分の一以下に低下させる特性：重要な知見。

長野県松本市モデル：信州大学付属病院（重症者受入れ）＋中核病院（軽症者受入れ）

4. 免疫の不思議な力を探る

備考：NHKスペシャル(7/4)「タモリ x 山中一人間vsウイルス」完全映像化技術より学ぶ₁

病原体を防ぐからだのメカニズムは二段構え＝自然免疫＋獲得免疫



出典：宮坂昌之名誉教授資料から転載。

<https://news.yahoo.co.jp/articles/c9c9ca76dc0b5fc50e9f2fb642935a9af83bffa0>

「サイエンスアゴラ2020」のシンポジウム「研究者と語ろう～新型コロナウイルス（COVID-19）免疫学的視点×ウイルス学的視点～」(2020年11月21日開催) からー

5. 治療薬を学ぶ

追記：5-ALA(8量体+Fe) 長崎大学に注目

既存治療薬候補の One-Sheet ・ *印・ 日本政府が COVID-19 治療薬として承認済

治療薬販売	治療薬	効能薬	開発元・製造元	桐山が注目する特許例
	アビガン (*)	新型インフルエンザ治療薬	富士フイルム 富山化学	WO2016120301A、US10098879、 特表 2018-505183、EP3250206A
	レムデシベル *	Ebora 出血熱治療薬	Gilead Science	US8957199、US2018/0071302、 US2015/0150893、US2015/0080289
	カレトラ	HIV プロテアーゼ阻害薬	Abbvie	WO2016004389A、US2017/0058027、 US2016/0280791、WO2013102042A
	フサン	セリンプロテアーゼ阻害薬・急性膵炎薬	鳥居薬品	特開平 09-040583、特公平 02-036239、 WO2007/013696
	シクレソニド	喘息治療薬	帝人ファーマ	WO10/126025、WO06/016715、 WO06/098519、WO17/199903
	アクテムラ	抗リウマチ薬	中外製薬、大阪大学、ロッシュ、	特開 2017-081937、特許 04012293、 特許 03281099、特許 03274532
	ヒドロキヂクロキノン	抗マラリア・免疫調整薬	Sanofi (仏)	US9625444、US8334392、 US8304440、US2011/0250620
	サリルマブ	関節リウマチ薬 遺伝子組み換え	Sanofi (仏) REGENERON	US9943594、US9556277、 US9308256、US9028819
	デキサメタゾン * - *	抗炎症薬	古くからある一般薬	
	?	サダミ駆虫薬、 スパコン富岳予測	石原産業	JP04300009、 WO2001/014340、014341、

6. ワクチンの早期接種を願う

開発国	ワクチン開発	Vaccine platform	Type	開発 Phase	ワクチン製造	桐山注目感染症特許(赤色-早期公開出願)	
	Oxford 大学	Non-Replicating Viral Vector	ChAdOx1-s	Phase3	 アストロゼネカ、セラム、 エマージェント・バイオソリューションズ	WO2020148554A, WO2020043869A, WO2019021013A, WO2019086900A, WO2019034887A	3
	CanSino Sinovac Wuhan	同上 Inactivated Inactivated	Adenovirus Type Vector Inactivate Inactivate	Phase3 3種類とも		CN111218459, US2020/0255862, WO2020143634A, CN111481531A, CN110982795A, CN110982794A, CN111411086A, CN111249456A, CN111057684A	4
	Moderna (NIAID)	RNA	LNP-encapsulated mRNA	Phase3	 ロンザ(スイス)	US10709779, US10272150, US2019/0008938, US2018/0311336, US2018/0303929	2
	CUREVAC	RNA RN Aoptimizer	mRNA CVnCoV	Phase2	 CureVac	US2020/0155668, US10682426, US10653799, US2020/0155668, US10588959, US10568958,	
	Sanofi Pasteur /GSK	Protein Submit	S protein Baculovirus production	Phase1/2		US2019/0194626, EP2880154B, US2019/0192649, WO2019192997A, WO2017140905A,	
	Novavax	Protein Submit	Glycoprotein Nanoparticle vaccine adjuvanted	Phase2	 エマージェント・バイオソリューションズ、AGC、富士フィルム	US2020/0085934, EP3595709A, US2017/0202948, WO2017041100A, WO2019023196A, EP3658118A	
	J&J Janssen	Non-Replicating Viral Vector	Ad26COVS1	Phase3	 J&J	US2020/0231577, US2020/0164057, US2020/0038500, US2020/0164057, US10517944, US2020/0030437,	5
	Anges 大阪大学 タカラバイオ	DNA	DNA plasmid vaccine +Adjuvant	Phase1/2	 タカラバイオ AGC、富士フィルム HD	US9889186, US9695219, US9376470, US10226529, US2016/0296613, US10172927, WO2020045368A,	6
 	BioNTech/ Pfizer	RNA	3-LNP-mRNAs	Phase 3	 Pfizer	US2018/0142198, US2018/028686, US2017/0079916, WO2020045368A, WO2019230654A, WO2018181997A,	1
	Inovio	DNA	DNA plasmid vaccine	Phase1/2	 Richter-Helm BioLogics	US2020/0222527, US2019/0022213, US2019/0153040, US2019/0135899, WO2019148086A, WO2019152600A,	

中国の人民解放軍の軍医・研究者 (Chen Wei=陈薇) 少将の特許に注目。

7. 最も伝えたいこと (1) 私が学んだこと

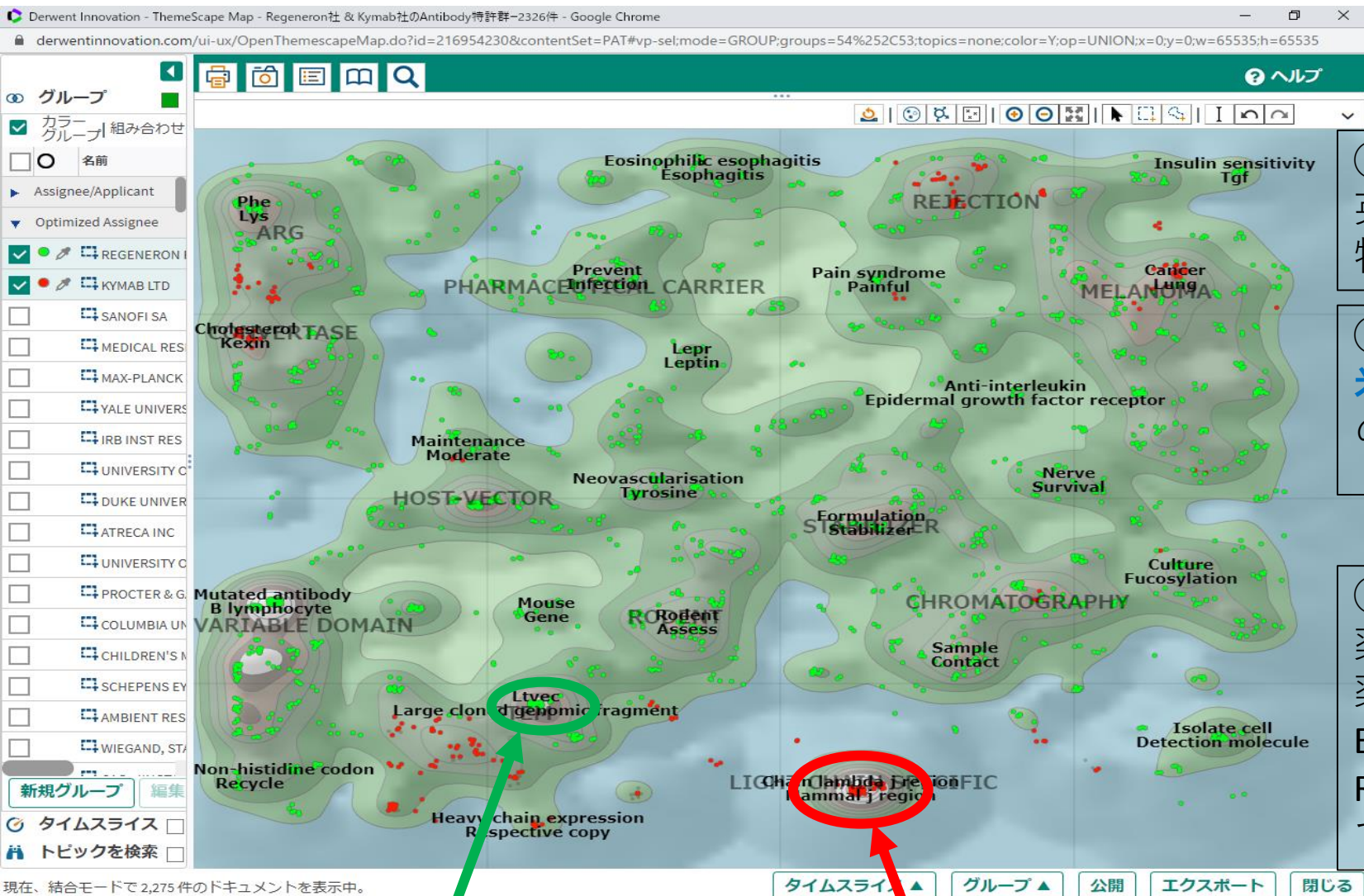
1. COVID-19の早期出願が急増。(次頁)
2. 人工抗体薬を緊急使用承認を日本にも。(桐山願望)
3. **人工抗体薬**の特許群(後述)
4. 1918年びスペイン風邪とCOVID-19の比較
5. 感染症から守る生活習慣：①**Stay Home**、②**手洗い**、うがい、洗面、マスク着用、③検温、**血中酸素濃度測定**、④免疫力の維持(適度な運動、散歩vs食事vs睡眠)、⑤正しい情報の選択入手と、自分の頭で論理的に考える習慣。
6. **見直されるべき情報**：①222nmの紫外線、②湿度(加湿器)、③交差免疫、④ワクチン接種、⑤5-ALA

7-2. 人工抗体薬の緊急使用承認を日本にも。(桐山希望)

- 前米Trump大統領が軍医から受けたから治療。
Regeneron Pharmaceutical社の人工抗体薬。
- 米FDA（食品安全管理局）は、11月にEli Lilly社
とRegeneron Pharmaceutical社の人工抗体薬を**緊急使用承認**を出した。
- 新型コロナ感染者の回復者の**スーパー抗体者の
血清血漿**が欧州で重傷者を救った事例がある。
- 日本の免疫学・ウイルス学の**専門家達**（宮坂昌
之名誉教授、河岡義裕教授ら）も中和抗体の作
用を示唆している。



7-3. 人工抗体薬の特許群



①米 R社と英K社との特許係争

②米 R社と 米Kelle社と の特許係争

③人工抗体
薬の二大製
薬会社は
Eli Lillyと
Regeneron
である。

EP2264163B EP1360287B WO2002066630A	米Regeneron社 Murphy特許
---	-------------------------

米Regeneron社
Murphy特許

英Kymab社
Bradley特許

US10064398
US9505827
US9447177
US9434782

7-4. 1918年のスペイン風邪とCOVID-19の比較

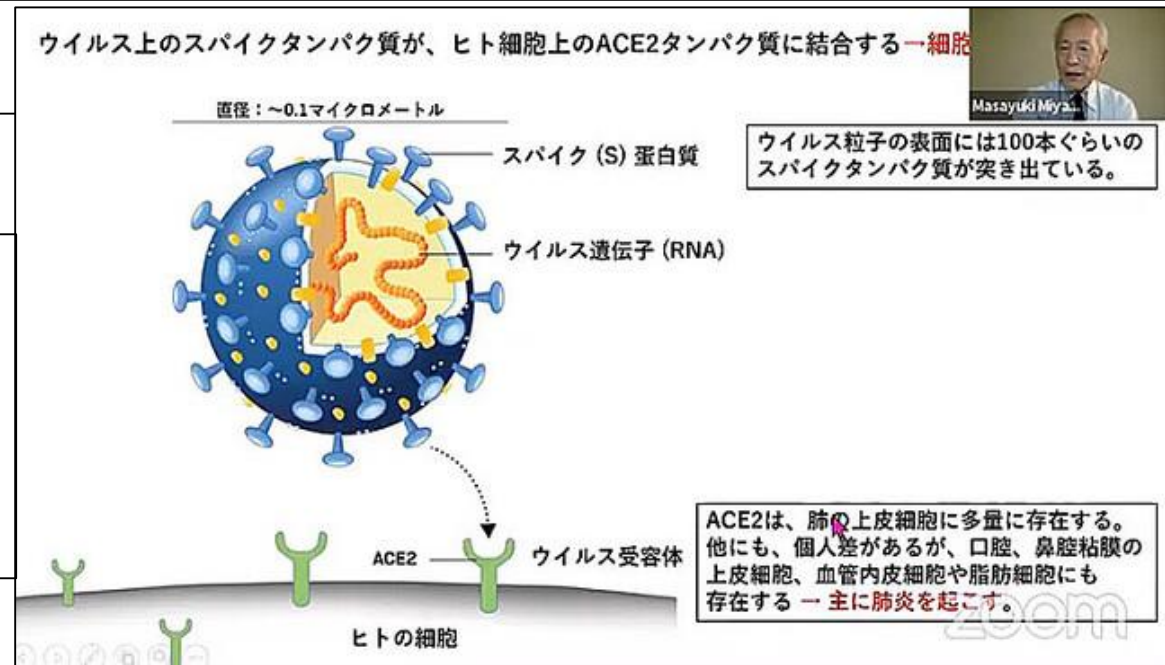
表2 □1918年のスペイン風邪と COVID-19 の比較

No	項目	新型コロナウイルス	スペイン風邪
1	時代	2019 年 12 月から	1918 年ー1921 年
2	死者数	約 76.4 万人(8/15 時点)感染者 2116 万人、 (日本死者数 1092 人、日本は例外的に少ない。) 米国の死者数はベトナム戦争戦死者の 3 倍以上。	死者数 1700 万人～5000 万人、 第一次世界大戦中 。 詳細は不明。
3	原因	約 1 か月後には、 ウイルス と判明。 全遺伝子情報が解明された。 RNA 型 である。	原因が当初不明。 1954 年にウイルスと判明。
4	検査方法	PCR 検査、抗原検査、抗体検査、他。	当時は検査法が 無し 。
5	ワクチン 抗ウイルス剤	ワクチン設計は約 1 か月後。6 月には P-3 相テスト。 年内に部分実現。抗 ウイルス 剤の開発も平行に進む。	当初全く無し。2009 年に 新型インフルエンザ 流行時にワクチン接種。
6	科学的根拠	PCR 検査 と第 5 日目に発症を多数確認。	当時は全く分からず。
7	政府の対処	各国一時、ロックダウン又は全面自粛	戦争中で対処法□未公開
8	基本姿勢	大人として 理性ある行動 で向き合う必要がある。 根拠ある楽観姿勢に変えて 行く必要がある。	原因も解らず、恐怖が先行した。 理性も持てず。
9	経済・物流	リアルの世界で、物流で経済を支えている 。 全面的な自粛は一定期間行ったが・・・。	1918 年は、恐怖と混乱のみ。

7-5. 感染症から守る生活習慣

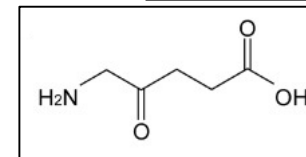
1. **Stay Home** (不要不急の外出自粛、在宅勤務、**3密x**)
(「**君子危うきに近寄らず**」がベストな正攻法と考える。)
2. **手洗い**、うがい、洗面、**マスク**着用
3. 検温、**血中酸素濃度測定**、
4. 免疫力の維持 (**適度な運動**(**筋肉・骨**)、散歩vs食事vs睡眠、)
5. 正しい情報の選択入手と、自分の頭で科学的に論理的に考える習慣。

新型コロナウイルスは
脂質の球状被膜に
覆われているので、
**石鹸の手洗いで、
分解・殺菌できる。**
(世界は教科書でできている。)



7-6. 見直されるべき情報(私案)

1. 222nmの紫外線（殺菌）、デビッド・ブレナー教授。
2. 湿度（高い方が良い。加湿器。）
3. 交差免疫、
4. ワクチン接種、
5. 善玉抗体薬、
6. Brain fog (脳の霧)
7. 5-アミノレブリン酸（5-ALA）



正しい情報の選択入手と、自分の頭で科学的に論理的に考える習慣。

桐山の場合には、非特許情報を特許情報で再確認と検証を行う。

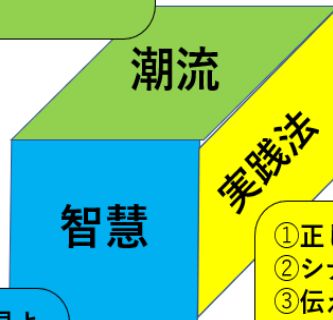
個人的な人生方程式（参考）

Enhanced, Augmented & Empowered with my **Experiences through (AI+DX+NMT+IP Information)** 自分の実践能力を強化する方法。

7. 最も伝えたいこと (2)

潮流と智慧と実践法

- ①AI & DX
- ②気候変動（脱炭素、電動化）
- ③コロナ禍（緊急事態宣言）
- ④Open Science & Citizen Science

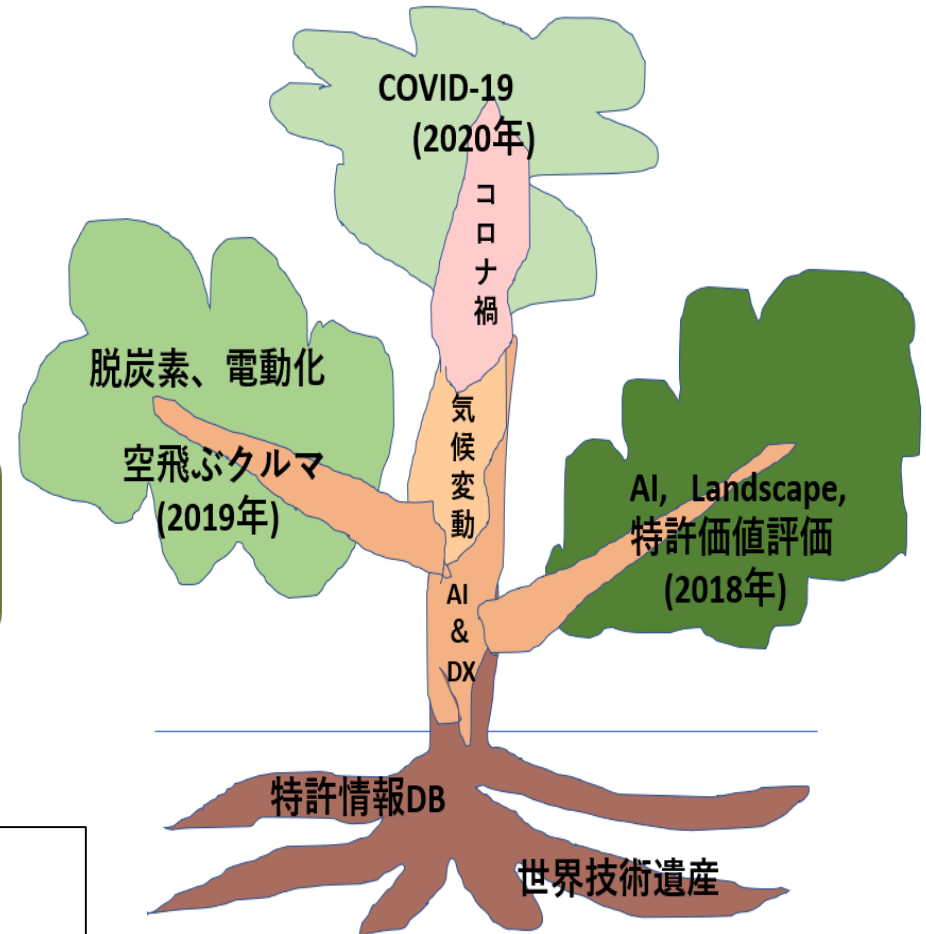


- ①正しい情報と論理的思考
- ②シナリオ作りと選択
- ③伝え方（共感、One Word Value）
- ④人生方程式
- ⑤梃子の原理

- ①先を読み、世界を見よ
- ②本質(正体)を掴め
- ③万巻の書は読めない
- ④師と読書
- ⑤自分事と利他

私のニックネーム：Cubic from Japan,
図解好き(ダヴィンチ)

特許情報の分析の木を描く



8. 参考：特許調査の極意らしきモノ

- **著名人の本・記事**から学ぶ：中村栄、和田玲子、酒井美里、野崎篤志、桐山勉、他（例えば、INFOSTA月刊誌－7月特集号）
- **有料**セミナー・研究会から学ぶ
- ど素人の分野の特許調査は代行検索を引き受けてはいけない。
（**検索原則の禁止行為**）
- **先ずは非特許情報**をかなりの程度学んでから、特許調査を行う。
- 調査する**対象技術の特許分類**を理解する
- **暗黙の検索禁止行為**の5か条、**やってはいけない検索式作成法**の5か条から学ぶ。
- **AI時代のインフォプロ像**：**C-Suiteの心に突き刺さるストーリー**。
- 裏側で、**AI技術が組み込まれた有料検索システム**を利用する。
(cf. 人生方程式)

想定Q&A

- **検索原則の禁止行為**とは、
 - ど素人の分野では代行の特許調査を引き受けてはいけない。
- **暗黙の検索禁止行為の5か条**とは、
 - ①コストパフォーマンス無視、②再現率と適合率を無意識、③三段でなく一段検索のみ、④技術内容の未確認、⑤特許分類を学ばず、
- **やってはいけない検索式作成法の5か条**とは、
 - ①異なる概念要素の積と和の混合、②不適切な二重絞り込み、③上位概念と下位概念の無意識、④特許分類の階層構造の無視、⑤ごちゃまぜ検索

右の本の第27章―「賢い検索式とはどんなものか」-P58 -P59に掲載。



想定Q&A INFOSTA-2020年-7月号

- 「AI時代のインフォプロ」-2020年7月号特集
- 情報調査・分析およびインフォプロの今後、野崎篤志
- 「AI系調査ツールとの付き合い方」、酒井美里
- IP軍師たるインフォプロを目指す、桐山勉
- AI時代のIPランドスケープを迫る知財アナリスト、和田玲子、中村栄
 - 解析シナリオ構築力のレベルアップを目指して

想定Q&A JapioYERBOOK2020

- 新たな価値提供分野におけるIPランドスケープの貢献、中村栄
 - IPL de Connect, ハード面のスキル、ソフト面のスキル。

新型コロナウイルスの事例

- コロナウイルスの本を 3 冊読みこむ
- 3 か月以上、日経新聞記事から切り抜き記事を蓄積し、その分野の知識・常識とある程度の専門用語を理解する。
- 非特許情報を特許情報で確認・検証する。この逆の「非特許を学ばずに特許情報を検索してはいけない」。禁止行為である。難しく言うと、非特許情報と特許情報の融合 (Integration)を図る。
- 新型コロナウイルス関係の分野を細分化して、非特許情報とその関連特許情報を蓄積する。一定期間の孵化（勉強）期間の後に、特許分析に入る。
- 何故、新型コロナのことを調べ、特許分析をするか。自分事として、自分を守るのが目的である。他人事みたいな調査をしない。（裏と表で考える）
- 正しい必要情報の集め方、学び方を理解する。

9. (参考)私が利用する機械翻訳システム

限られたオリジナル特許群に対して、色々な機械翻訳エンジンで比較検討をする。（大量の特許群ではできない）

- Japio-GPG/FX—AI翻訳サービス

- ## • みんなの自動翻訳@TexTra

- Google 翻訳

Japio-GPG /FX		AI翻訳サービス	
		☒ 原文表示 ☒ 代表国表示 ☒ 全図面表示 再翻訳	
<<		Excel ▼ 出力	
クリア	ハイライト	最初 ◀ 前へ 1 次へ ▶ 最後	
ウェアラブル	☐ 検索		
ナビ	☐ 検索		
	☐ 検索		
	☐ 検索		
発明人	☐ 検索		
発明者	☐ 検索		
I P C	G01C21/20[200601][G01C21/20]		
C P C	G01C21/206; G01C21/200		
ファミリーID	60266032		
優先権主張番号	WO2016CN81408 [2016-05-09]		
ファミリー	WO2017193255A1; CN107529339A		
書誌情報 出願番号 CN201680008104.4 [CN201680008104] [2016-05-09] 文献番号 CN107529339A [CN107529339A] [2017-12-29] 発明の名称 可穿戴装置による室内導航方法 出願人 発明者 I P C G01C21/20[200601][G01C21/20] C P C G01C21/206; G01C21/200 ファミリーID 60266032 優先権主張番号 WO2016CN81408 [2016-05-09] ファミリー WO2017193255A1; CN107529339A			
発明の名称 可穿戴装置及其室内导航方法		発明の名称（訳文） ウェアラブル装置及びその室内ナビゲーション方法	
要約 一種可穿戴装置の室内导航方法。方法为：通过根据接收的导航指令，在预设的室内电子地图中查找目标位置(S10)；获取用户当前所处位置，根据所述当前位置、所述用户当前所处位置，确定由所述用户当前所处位置去至所述目标位置的最近路径(S20)；根据所述最近路径进行导航。导航过程中途经该最近路径上的各个位置时，将所述用户在所述最近路径上所处的位置以及该位置附近的位置显示出来，使该用户沿所述最近路径到达目的地(S30)。若按，则生成并播放提示信息，使该用户沿所述最近路径继续在国内(S40)。还提供一种可穿戴装置。所提供的可穿戴装置及其室内导航方法，能够实现在室内导航，以最短的时间指引用户到达室内目标位置，减少用户在室内寻找目标位置的时间，室内导航导航效果好。		要約（訳文） ウェアラブル装置の室内ナビゲーション方法。方法は以下のとおりである：受信したナビゲーションコマンドに基づき、予め設定された室内電子地図から目標位置を探索する(S10)、ユーザの現在位置を取得し、前記目標位置、前記ユーザの現在位置に基づき、前記ユーザの現在位置から前記目標位置までの最適なルート決定する(S20)、前記最適ルートに基づいてナビゲーションを行い、ナビゲーション途中に、ユーザの所在位置をリアルタイムに取得し、前記ユーザの所在位置が前記最適ルート範囲外であるかかを判断する(S30)、そうであれば、提示情報を生成し且つ再生し、ユーザに前記最適ルート範囲内に居るように指示する(S40)。ウェアラブルデバイスも提供される。提供されるウェアラブル装置及びその室内ナビゲーション方法は、室内ナビゲーションを実現することができる。最短の時間でユーザが室内目標位置に到着するようにガイドし、ユーザが室内で目標位置を待つ時間を減少させ、室内ナビゲーションの体験効果が高い。	
請求項 1. 一種可穿戴装置の室内导航方法，其特征在于，所述可穿戴装置的室内导航方法包括： 步骤一，根据接收到的导航指令，在预设的室内电子地图中查找目标位置； 步骤二，根据所述当前位置、所述用户当前所处位置，确定由所述用户当前所处位置去至所述目标位置的最近路径； 步骤三，根据所述最近路径进行导航。导航过程中途经该最近路径上的各个位置时，将所述用户在所述最近路径上所处的位置以及该位置附近的位置显示出来，使该用户沿所述最近路径到达目的地； 步骤四，若按，则生成并播放提示信息，使该用户沿所述最近路径继续在国内。		請求項（訳文） 1. ウェアラブル装置の室内ナビゲーション方法であって、前記のウェアラブル装置の室内ナビゲーション方法には以下を含むことを特徴とする：ステップ1、受信したナビゲーションコマンドに基づき、予め設定された室内電子地図から目標位置を探索する；ステップ2、前記ユーザの現在位置、前記ユーザの現在位置に基づき、前記ユーザの現在位置から前記目標位置までの最適なルート決定する；ステップ3、前記最適ルートに基づいてナビゲーションを行い、ナビゲーション途中に、ユーザの所在位置をリアルタイムに取得し、前記ユーザの所在位置が前記最適ルート範囲外であるかかを判断する；ステップ4、そうであれば、提示情報を生成し且つ再生し、ユーザに前記最適ルート範囲内に居るように指示する。	

私の機械翻訳の評価テンプレート

機械翻訳された日本語文ベースで、差分比較を検討する。

英語文 中国語文	人手翻訳	Japio-GPG/FX AI翻訳サービス	みんなの自動翻訳 @TexTra	Google翻訳
オリジナル ①	正解案 A1	文章群 B1	文章群 C1	文章群 D1
評価	基準 評価源	差分比較、類似 度比較、主観的 疑似数値法比較、 等。 B1/A1	差分比較、類似度 比較、主観的疑似 数値法比較、等。 C1/A1	差分比較、類似 度比較、主観的 疑似数値法比較、 等。 D1/A1
オリジナル ②	正解案 A2	文章群 B2	文章群 C2	文章群 D2
評価	基準 評価源	オリジナル①と 同様に評価 B2/A2	オリジナル①と 同様に評価 C2/A2	オリジナル①と 同様に評価 D2/A2

事例 1：米バイデン大統領の就任演説の文章群の中で、桐山が興味を持った段落英文とその日本語訳文章群で比較評価を行う。

私の機械翻訳の評価テンプレート

英語なら少しは理解できるので、英語ベースで差分比較する方法を選択する。

中国文	Shareresearch 英語文訳	Derwent 英語文	Google翻訳	CNIP 英語版文
オリジナル ①	正解案 A1	文章群 B1	文章群 C1	文章群 D1
評価	基準 評価源	差分比較、類似 度比較、主観的 疑似数値法比較、 等。B1/A1	差分比較、類似 度比較、主観的 疑似数値法比較、 等。C1/A1	差分比較、類似度 比較、主観的疑似 数値法比較、等。 D1/A1
オリジナル ②	正解案 A2	文章群 B2	文章群 C2	文章群 D2
評価	基準 評価源	オリジナル①同様に B2/A2	オリジナル①同様に C2/A2	オリジナル①同様に D2/A2

事例 2：中国の mRNA ワクチンの特許 CN111218459 の英訳比較にてチェックする。
つまり、英語訳文が全く同一になるとは限らない。
分母の Ani を Derwent 英語訳にするか、Shareresearch 英訳分にするか、選択する。

事例 2：中国人民解放军军事科学院军事医学研究院の陈薇（Chen Wei）少将が発明者である CN111218459B に桐山は注目している。

10. おわりに

- 謝辞

- Clarivate Analytics様 (Derwent Innovation)
- 日立製作所様 (Sharersearch)
- アイ・ピー・ファイン様 (THE調査力AI)
- 日本特許情報機構様 (Japio-GPG/FX-特許AI機械翻訳)
- アジア特許情報研究会皆様
- INFOSTA-専門部会ーパテントドキュメンテーション部会皆様

お断り：このプレゼン発表は、桐山の個人的な考え方であり、私が所属するいかなる組織の了解を取ったものではない。

ご清聴を有難うございました。 (謝謝)



自動翻訳技術の概要： なにができるか／ できるようになってきているか

情報通信研究機構（N I C T）

内山将夫

特許情報シンポジウム

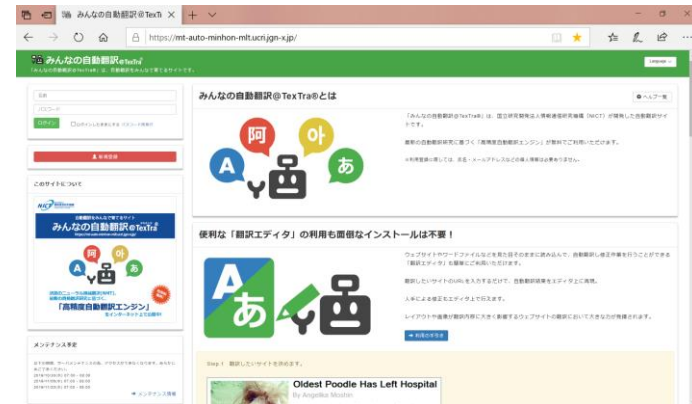
2021年2月26日

本講演の目的

- 自動翻訳技術の概要をお伝えする
- 情報通信研究機構の取り組みの一部をご紹介します

自動翻訳の歴史

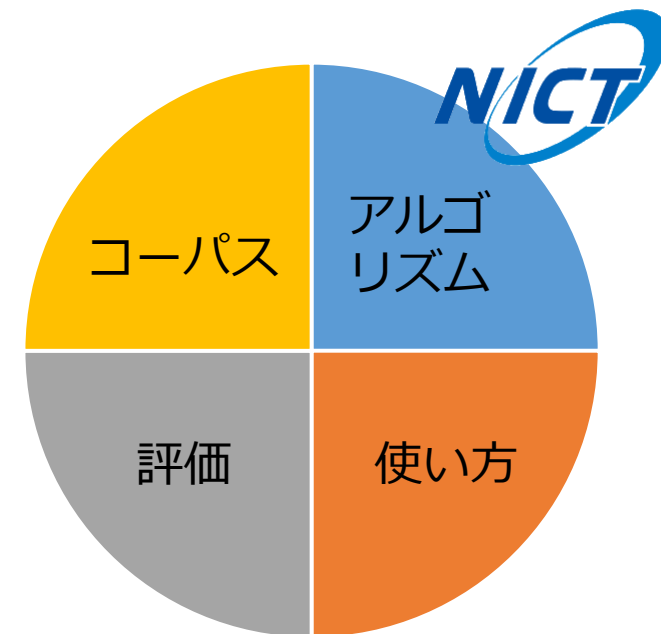
- 計算機が発明されてすぐに自動翻訳の研究が開始(1950年代)
- 自動翻訳の考え自体は 1949 年にWarren Weaverが提案 (cf. Wikipedia)
- 半世紀以上の研究開発を経て、自動翻訳が一般に普及
 - みんなの自動翻訳@TexTra等のW e b 翻訳サービス
 - VoiceTra等のスマホアプリ
 - ポケトーク等の音声翻訳専用機



自動翻訳技術のタイプ

- 規則ベース自動翻訳
文法規則や辞書を人間が記述
上記に基づき自動翻訳を実施
- コーパスベース自動翻訳（MT）
対訳コーパスから自動翻訳エンジンを自動学習
任意言語対に対して適用可能
ニューラル機械翻訳（NMT）はこちら

コーパスベースMTの構成要素



- 自動翻訳アルゴリズム

1980年代からの発展⇒NMT

- 学習データとしての対訳コーパス

異なる言語で同一意味の文章の組からなるデータベース

- MTの評価手法

人間がどのように感じるかの観点からのMTの良さ

- MTの使い方

MTの普及により、MTの適切な使い方の周知が必要

対訳コーパスの一例

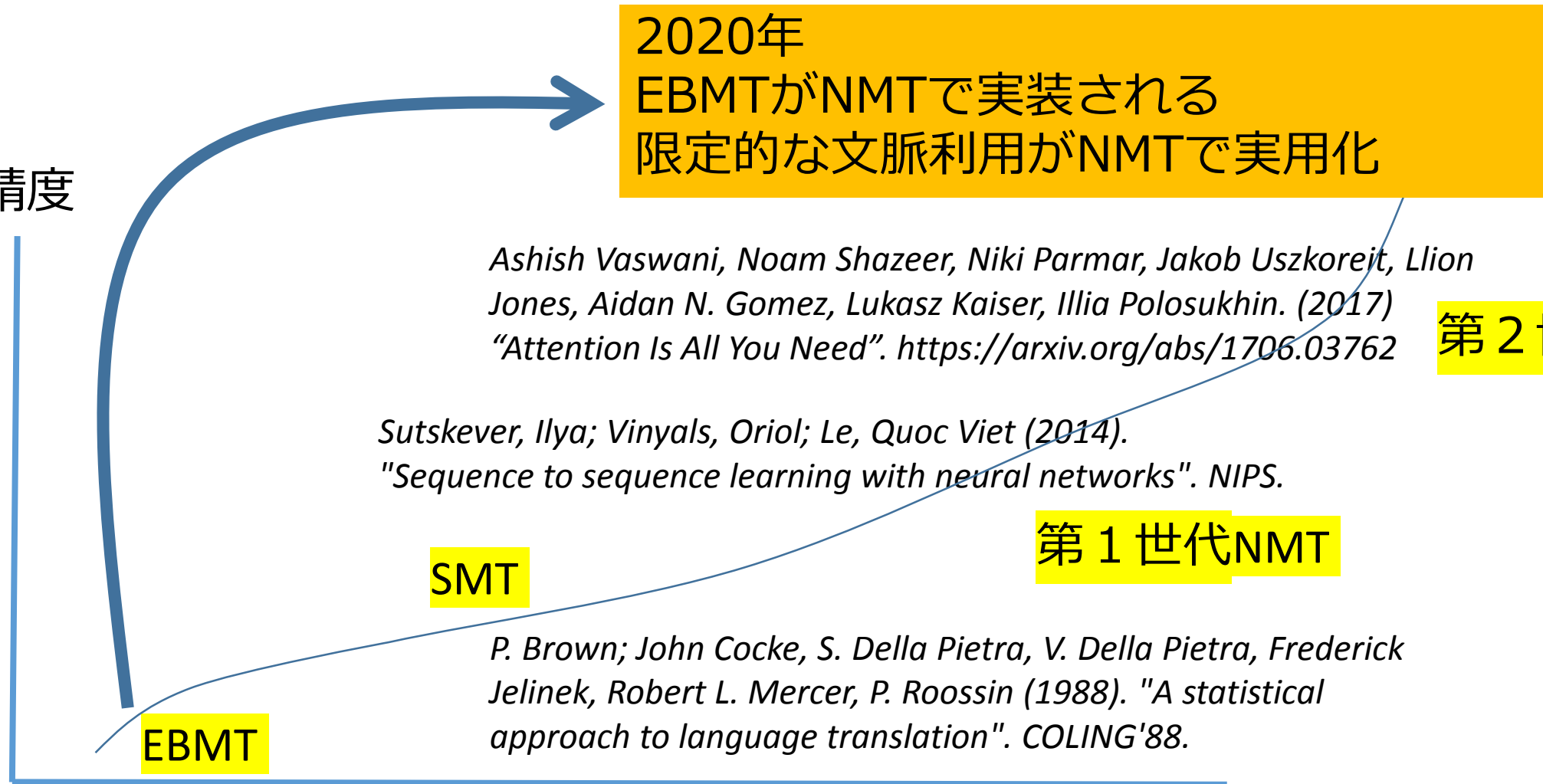


- Italy have defeated Portugal 31-5 in Pool C of the 2007 Rugby World Cup at Parc des Princes, Paris, France.
- Italia berhasil mengalahkan Portugal 31-5 di grup C dalam Piala Dunia Rugby 2007 di Parc des Princes, Paris, Perancis.
- フランスのパリ、パルク・デ・フランスで行われた2007年ラグビーワールドカップのプールCで、イタリアは31対5でポルトガルを下した。
- អ៊ីតាលីបានឈ្នះលើព័រទុយហ្គាល់ 31-5 ក្នុងប្លង់C នៃពិធីប្រកួតពានរង្វាន់ពិភពលោកនៃកីឡាបាល់ឱបឆ្នាំ2007ដែលប្រព្រឹត្តទៅប៉ារីសខេត្តប្រ៊ីន ក្រុងប៉ារីស បារាំង។
- Itali telah mengalahkan Portugal 31-5 dalam Pool C pada Piala Dunia Ragbi 2007 di Parc des Princes, Paris, Perancis.
- ပြင်သစ်နိုင်ငံ ပါရီမြို့ပါဒက်စ် ပရင်စက် ၌ ၂၀၀၇ခုနှစ် ရုပ်ဘီ ကမ္ဘာ့ ဖလား တွင် အီတလီ သည် ပေါ်တူဂီ ကို ၃၁-၅ ဂိုး ဖြင့် ရေကူးကန် စီ တွင် ရှုံးနိမ့်သွားပါသည်။ ။
- Ý đã đánh bại Bồ Đào Nha với tỉ số 31-5 ở Bảng C Giải vô địch Rugby thế giới 2007 tại Parc des Princes, Pari, Pháp.
- อิตาลีได้เอาชนะโปรตุเกสด้วยคะแนน31ต่อ5 ในกลุ่มC ของการแข่งขันรักบี้เวิลด์คัพปี2007 ที่สนามปาร์กเดอแพรงส์ ที่กรุงปารีส ประเทศฝรั่งเศส
- Natalo ng Italya ang Portugal sa puntos na 31-5 sa Grupong C noong 2007 sa Pandaigdigang laro ng Ragbi sa Parc des Princes, Paris, France.

コーパスベースMTアルゴリズムの進展



翻訳精度



Makoto Nagao (1984). "A framework of a mechanical translation between Japanese and English by analogy principle". In A. Elithorn and R. Banerji. *Artificial and Human Intelligence*. Elsevier Science Publishers

年代

みんなの自動翻訳@TexTra

<https://mt-auto-minhon-mlt.ucri.jgn-x.jp/>



The screenshot shows a web browser window with the URL <https://mt-auto-minhon-mlt.ucri.jgn-x.jp/>. The page has a green header with the site name and a language selector. On the left, there is a login section with fields for '名前' (Name) and 'パスワード' (Password), a 'ログイン' (Login) button, a checkbox for 'ログインしたままにする' (Keep me logged in), and a link for 'パスワード再発行' (Reset password). Below this is a '新規登録' (New registration) button. At the bottom left, there is a section 'このサイトについて' (About this site) with the NICT logo and a small description. The main content area is titled 'みんなの自動翻訳@TexTra®とは' (What is everyone's automatic translation@TexTra®) and includes a 'ヘルプ一覧' (Help list) link. It features a graphic with speech bubbles containing the characters 'ア', 'ア', and 'あ' around a robot head icon. The text explains that the site is a free automatic translation engine developed by NICT, based on the latest research. It also states that no personal information is required for registration.

みんなの自動翻訳@TexTra®
「みんなの自動翻訳@TexTra®」は、自動翻訳をみんなで育てるサイトです。

Language ▾

名前
パスワード
ログイン
☐ ログインしたままにする
パスワード再発行

新規登録

このサイトについて

NICT
国立研究開発法人情報通信研究機構

みんなの自動翻訳@TexTra®とは
ヘルプ一覧

「みんなの自動翻訳@TexTra®」は、国立研究開発法人情報通信研究機構（NICT）が開発した自動翻訳サイトです。

最新の自動翻訳研究に基づく「高精度自動翻訳エンジン」が無料でご利用いただけます。

※利用登録に際しては、氏名・メールアドレスなどの個人情報は必要ありません。

文脈翻訳の事例

自動翻訳

日本語 → 英語

汎用NT 【日本語 - 英語】 1

翻訳

彼は学校の先生です。
数学を教えています。
彼女は学校の先生です。
数学を教えています。

He is a school teacher.
He teaches mathematics.
She is a school teacher.
She teaches mathematics.

翻訳不要記号を挿入

☐ 文章を結合する
☐ 文章を区切る
☐ 辞書引き不要

文脈利用翻訳

前1文

<

↓ 訳文コピー

✕ リセット

専門的文章の翻訳例

In Transformer-based neural machine translation (NMT), the positional encoding mechanism helps the self-attention networks to learn the source representation with order dependency, which makes the Transformer-based NMT achieve state-of-the-art results for various translation tasks.

However, Transformer-based NMT only adds representations of positions sequentially to word vectors in the input sentence and does not explicitly consider reordering information in this sentence.

In this paper, we first empirically investigate the relationship between source reordering information and translation performance.

The empirical findings show that the source input with the target order learned from the bilingual parallel dataset can substantially improve translation performance.

Thus, we propose a novel reordering method to explicitly model this reordering information for the Transformer-based NMT.

The empirical results on the WMT14 English-to-German, WAT ASPEC Japanese-to-English, and WMT17 Chinese-to-English translation tasks show the effectiveness of the proposed approach.

Transformer-based Neural Machine Translation(NMT)において、位置符号化機構は自己注目ネットワークが順序依存性を持つソース表現を学習するのを助ける。

これは、Transformer-based NMTが種々の翻訳タスクに対して最先端の結果を達成することを可能にする。しかし、Transformer-based NMTは入力文中の単語ベクトルに連続的に位置の表現を加えるだけで、この文中の情報の順序変更を明示的に考慮しない。

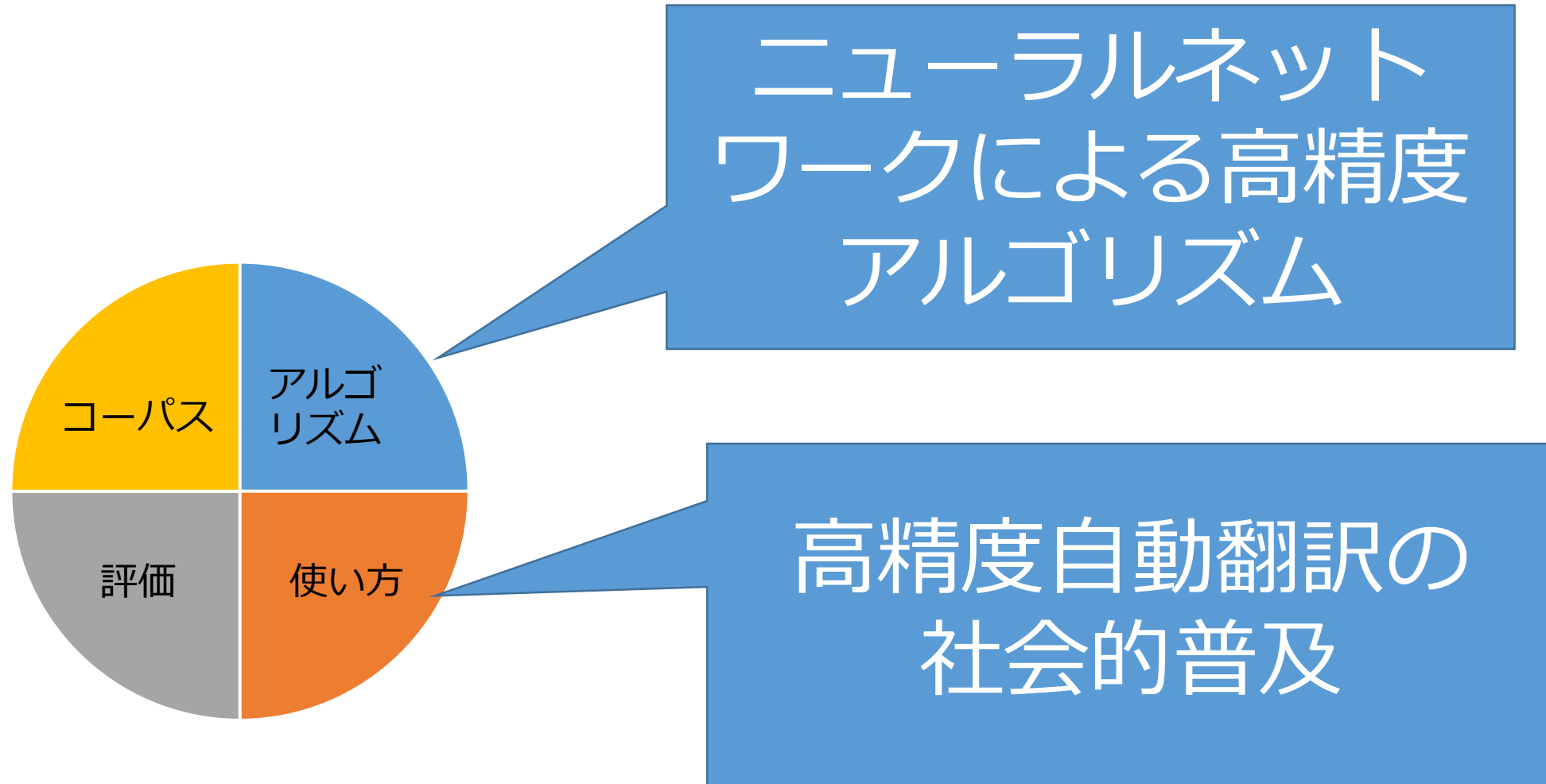
本論文では、まずソース順序変更情報と翻訳性能との関係を経験的に検討する。

経験的発見は、バイリンガル並列データセットから学習したターゲット順序を持つソース入力が翻訳性能を大幅に改善できることを示す。

そこで、Transformer-based NMTに対して、この順序変更情報を明示的にモデル化する新しい順序変更法を提案した。

WMT14English-to-German,WAT ASPEC Japanese-to-English,WMT17Chinese-to-English翻訳タスクの経験的結果は、提案アプローチの有効性を示した。

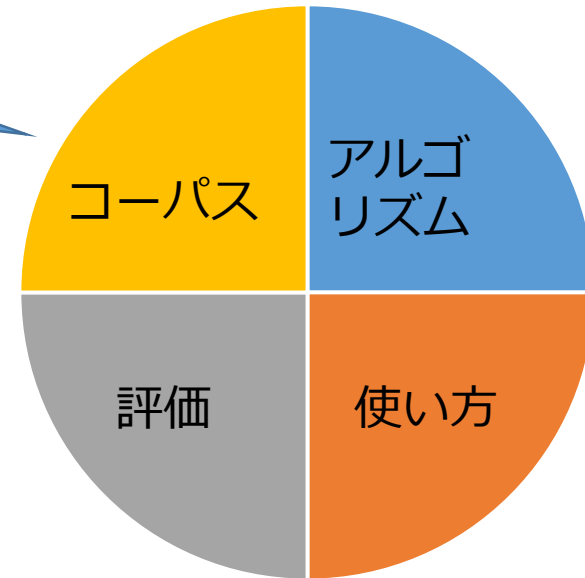
NMTにより新しくなった点



NMTでも以前と同じ点

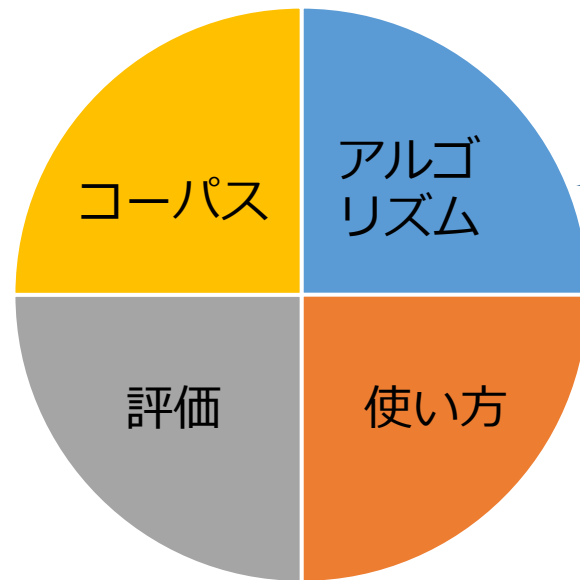
大規模対訳
コーパスが必要

人間による評価
が必要



NMT アルゴリズムの概要

- 基本訓練
- アダプテーション
- EBMT



ニューラルネット
ワークによる高精度
アルゴリズム

NMTの基本構造（第1・2世代ほぼ共通）

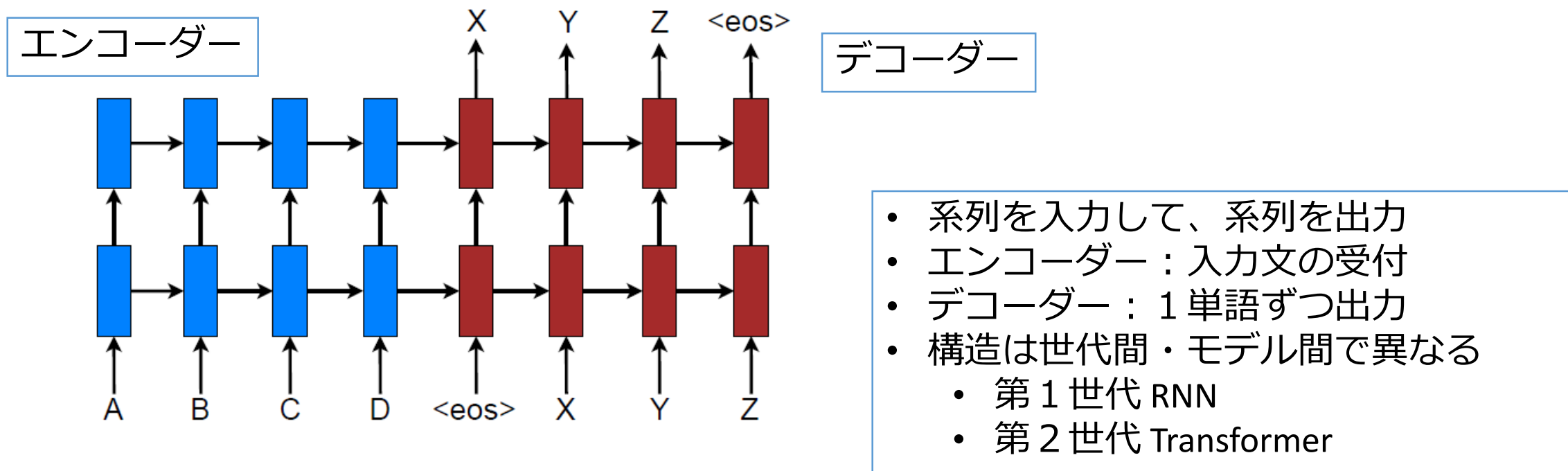


Figure 1: **Neural machine translation** – a stacking recurrent architecture for translating a source sequence A B C D into a target sequence X Y Z. Here, <eos> marks the end of a sentence.

Thang Luong; Hieu Pham; Christopher D. Manning. (2015)
Effective Approaches to Attention-based Neural Machine Translation. EMNLP

第1世代と第2世代の自動評価尺度BLEUの比較

	汎用日英	汎用英日	特許日英	特許英日	
2017	22.1	24.2	40.7	41.7	第1世代
2018	21.2	24.5	40.0	41.6	第1世代
2019	20.8	23.2	37.7	40.1	第1世代
2020	27.6	31.5	45.6	46.8	第2世代
2021	28.1	31.0	47.3	46.5	第2世代

汎用・特許ともに、
第1世代(RNN)から第2世代(Transformer)にかけて、
大きくBLEU値が向上

訓練・アダプテーション・EBMT

【訓練】 1文ずつNMTモデルのパラメタを調整する

- 大雑把には：
 - 入力文を翻訳
 - 参照訳文と比較
 - 翻訳文と参照訳文の違いに応じてNMTのパラメタを更新
 - 以上を大規模に繰り返す（数億回になることもある）

【アダプテーション】（fine tuning とも言います）

- 訓練済みNMTモデルに、上記訓練を特定分野データで追加
- 訓練済みモデルをベースにするので、比較的少量データで高精度

【EBMT】（NICT開発・詳細未発表）

- 入力文と類似した対訳文をデータベースから検索
- 十分に類似した文があるときには、それを参考に自動翻訳
- 類似文がない場合には、ベースのNMTで自動翻訳

アダプテーションとEBMTの特徴

● アダプテーション+EBMT

ベースモデルをアダプテーションしたモデルを利用してEBMTを実施します。訓練データが1万文以上のときにお勧めします。

○ アダプテーション

ベースモデルを訓練データでアダプテーションします。訓練データと似ている文の自動翻訳精度向上が見込めますが、訓練データと似ていない文の自動翻訳精度がベースモデルより低下する場合があります。訓練データが1万文以上で、少しノイズがあるときにお勧めします。当初から「アダプテーション」として当サイトで提供していた訓練方法です。

○ EBMT

Example-based Machine Translationです。訓練データ中に入力文と似ている文がある場合には、それを参考に自動翻訳します。入力文との類似文がない場合はベースモデルで翻訳します。訓練データが1万文未満のときにお勧めします。

みんなの自動翻訳からお試しできます

アダプテーションの有効性の事例

自動車法規文の自動翻訳をニューラル技術で高精度化

～トヨタとの共同研究を通じ、英日・中日翻訳の実用度が向上～

- 自動車業界からトヨタが翻訳バンクに協力、翻訳データを提供
- 自動車法規を対象とした翻訳をニューラル英日翻訳システムで
24%実用度向上

法規原文	人による翻訳	ニューラル翻訳
This Australian Design Rule prescribes requirements for the number and mode of installation of lighting and light signalling devices on motor vehicle other than L-group vehicles.	本オーストラリア設計規則はLグループ車両以外の自動車への灯火および灯火信号装置の数と取り付け方法に関する要件を規定する。	本オーストラリア設計規則は、Lグループ車両以外の自動車への灯火および灯火信号装置の数および取り付け方法に関する要件を規定する。

アダプテーションで専門用語に適応



原文	第 1 世代汎用 NMT	第 2 世代汎用 NMT	第 1 世代金融 NMT	第 2 世代金融 NMT
本取引は、相互売買による取引実績が豊富な本投資法人ならではの取引であると考えており、本投資法人は今後も戦略的かつ継続的に資産入替を実施していく方針です。	This transaction is considered to be a transaction unique to this investment corporation, which has a rich record of transactions through mutual trade. This investment corporation will continue to change assets in a strategic and ongoing manner.	We believe that this transaction is unique to the Investment Corporation, which has a rich track record of transactions through mutual transactions. The Investment Corporation will continue to implement this transaction strategically and continuously.	The Transaction is considered as a transaction that is unique to the Investment Corporation, which has a wealth of transaction experience through mutual transaction. The Investment Corporation will continue to implement strategic and continuous asset replacement in the future.	The Investment Corporation believes that the Transaction is unique to the Investment Corporation, which has a wealth of transaction records through mutual transactions. Therefore, the Investment Corporation will continue to implement strategic and continuous asset replacement.

金融分野でのアダプテーション・EBMT

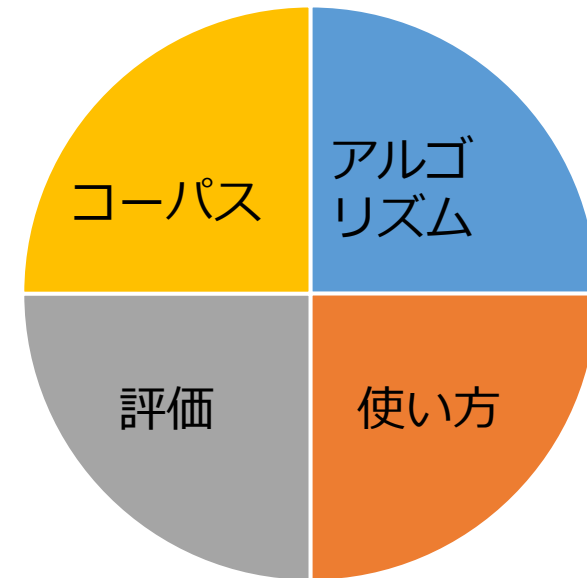
	汎用	ADAPT	EBMT	ADAPT+ EBMT
訓練と重複 あり1000	31.0	38.1	54.5	55.2
訓練と重複 無し514	29.3	35.2	47.7	49.1
訓練と重複 のみ416文	34.4	44.1	68.3	67.5

適時開示文書は、異なる文書であっても重複する文が多いので、EBMTの効果が著しく高い

M T の評価の概要

- M T の評価はM T を改善する原動力
- 誤訳報告⇒誤訳の改善
- 適切な評価⇒適切なM T

人間による評価
が必要



自動評価

- 簡便
- M T 訳と参照訳を比較
- 評価値を自動計算
- 開発にはとても役に立つ！
 - システム改修後、自動評価値に大きな変化がないことを確認
 - 大きな数値差（BLEUなら5程度）があれば品質差もほぼある

自動評価尺度 B L E U

- M T 訳と参照訳の類似度の尺度
- 1 0 0 0 文～5 0 0 0 文程度の文章で計算
- 一般的に B L E U 3 0 ～4 0 以上であれば高精度

Papineni, Kishore and Roukos, Salim and Ward, Todd and Zhu, Wei-Jing. (2002)
BLEU: a Method for Automatic Evaluation of Machine Translation. ACL

みんなの自動翻訳でのBLEU計算法

- 自動翻訳＞自動翻訳評価＞「＋新規登録」



The screenshot shows a web browser window with the URL <https://mt-auto-minhon-mlt.ucrjgn-x.jp/content/eval/edit/index.html>. The page title is "みんなの自動翻訳@TexTra". The main heading is "自動翻訳評価 登録". Below the heading, there are several input fields and checkboxes:

- 名前**: A text input field with the placeholder "名前を入力してください。".
- 言語方向**: Two buttons, "日本語" and "英語", with a double-headed arrow between them.
- 自動翻訳**: Two tabs, "標準" and "共有".
- チェックボックス**: A list of checkboxes for different translation methods, all currently unchecked:
 - ☐ 汎用NT 【日本語 - 英語】 ⓘ
 - ☐ 汎用NMT 【日本語 - 英語】 ⓘ
 - ☐ 特許NT 【日本語 - 英語】 ⓘ
 - ☐ 特許NMT 【日本語 - 英語】 ⓘ
 - ☐ 特許請求項NMT 【日本語 - 英語】 ⓘ
 - ☐ 対話NMT 【日本語 - 英語】 ⓘ
 - ☐ ミニ対話NMT 【日本語 - 英語】 ⓘ

自動評価尺度BLEUでの比較

	第 1 世代	第 2 世代
汎用	2 5 . 1	2 9 . 0
金融	3 4 . 6	3 8 . 4

第 1 世代 << 第 2 世代

汎用 << 金融

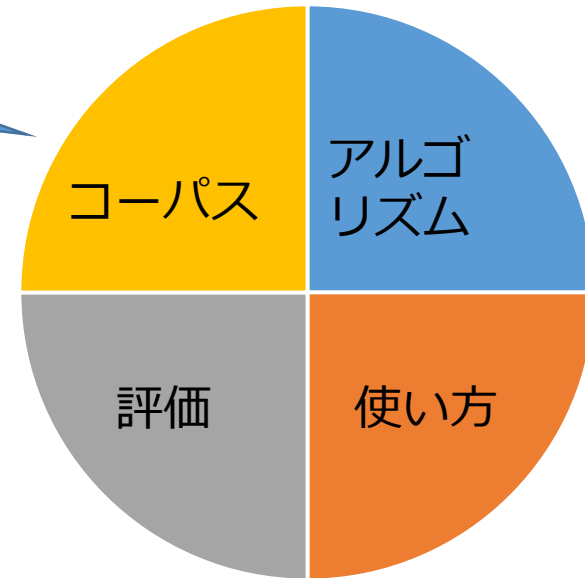
AAMT 2019, Tokyo: パネルディスカッション「機械翻訳もある総合的な翻訳サービスの模索～金融・IR分野を例に～」の内山のスライドの 1 枚より引用
日英翻訳の例について、みんなの自動翻訳で BLEU を計算

人間評価

- 原文とMT訳を人間が比較
- 人間が誤訳かを分析
- MTの研究開発に必須
- 第2世代NMTでは、おおむね正しい翻訳ができるようになり、以前は見逃されていた誤訳が目立ち始めた
 - 「N I C T」⇒「N I C T」（なぜか空白が挿入）
- ピンポイントな誤訳を検出する自動評価方法は知られていない
- ピンポイントな誤訳は、ピンポイントなルールなどで改善

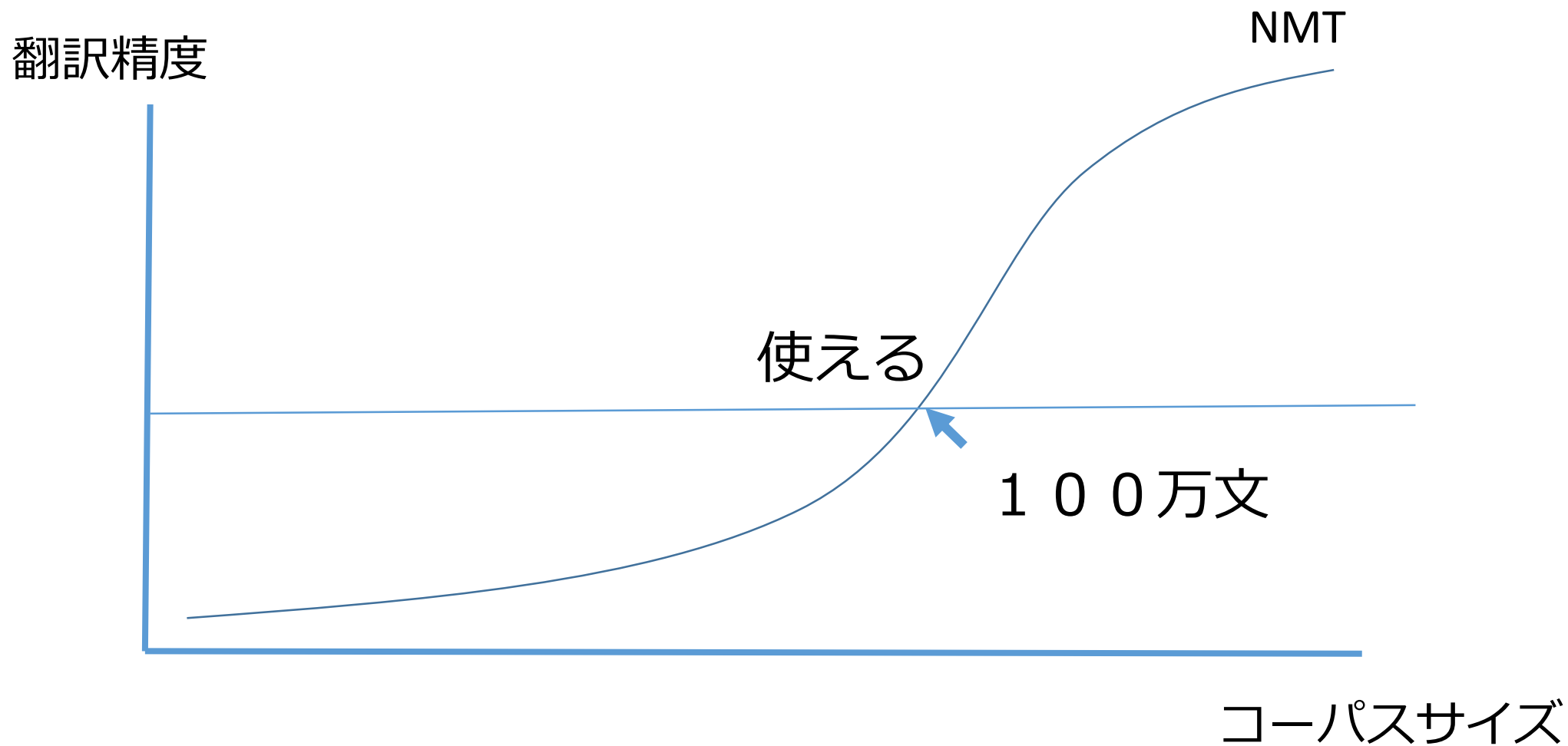
NMTでの対訳コーパスの重要性

大規模対訳
コーパスが必要

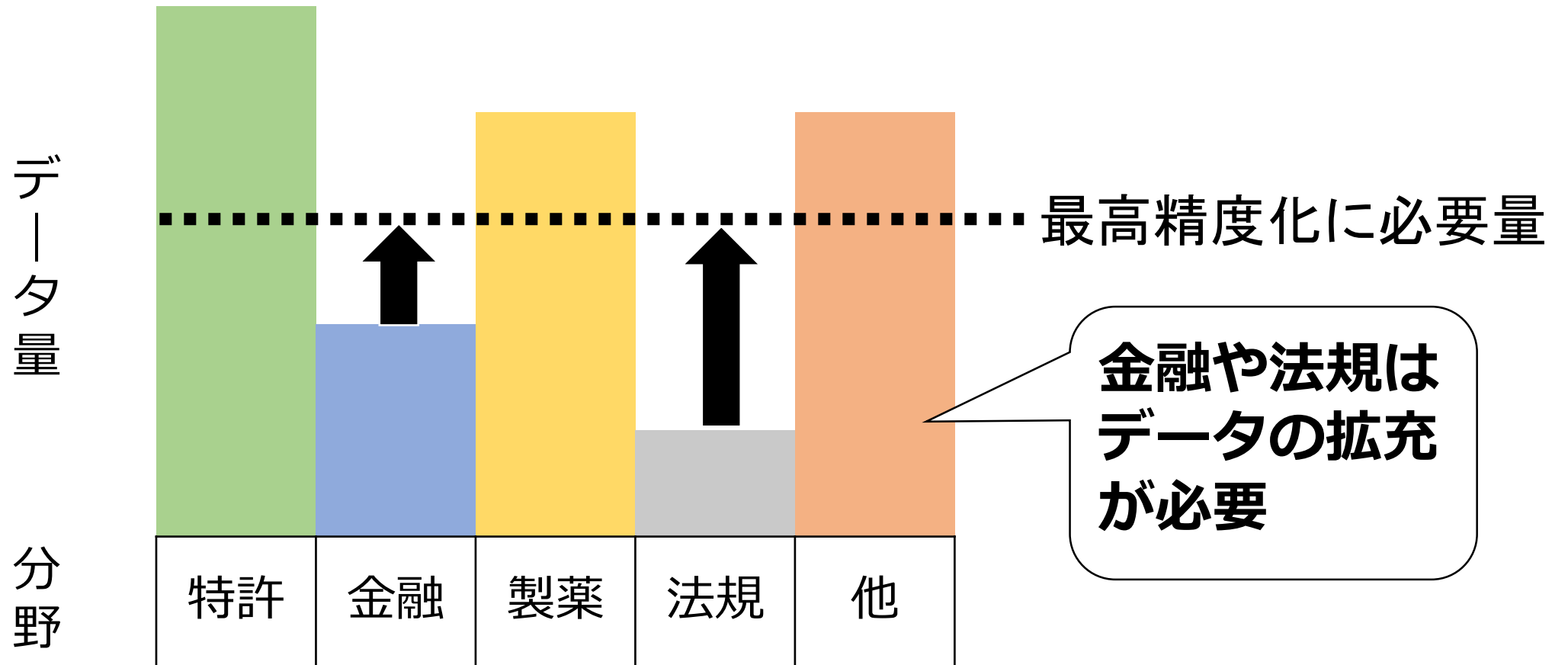


- 28

大規模対訳コーパスでNMTの性能向上



高精度 N M T には対訳データが重要



分野にマッチした対訳が重要 特許対訳は普通の文には不向き

This is **the** desk.



特許対訳

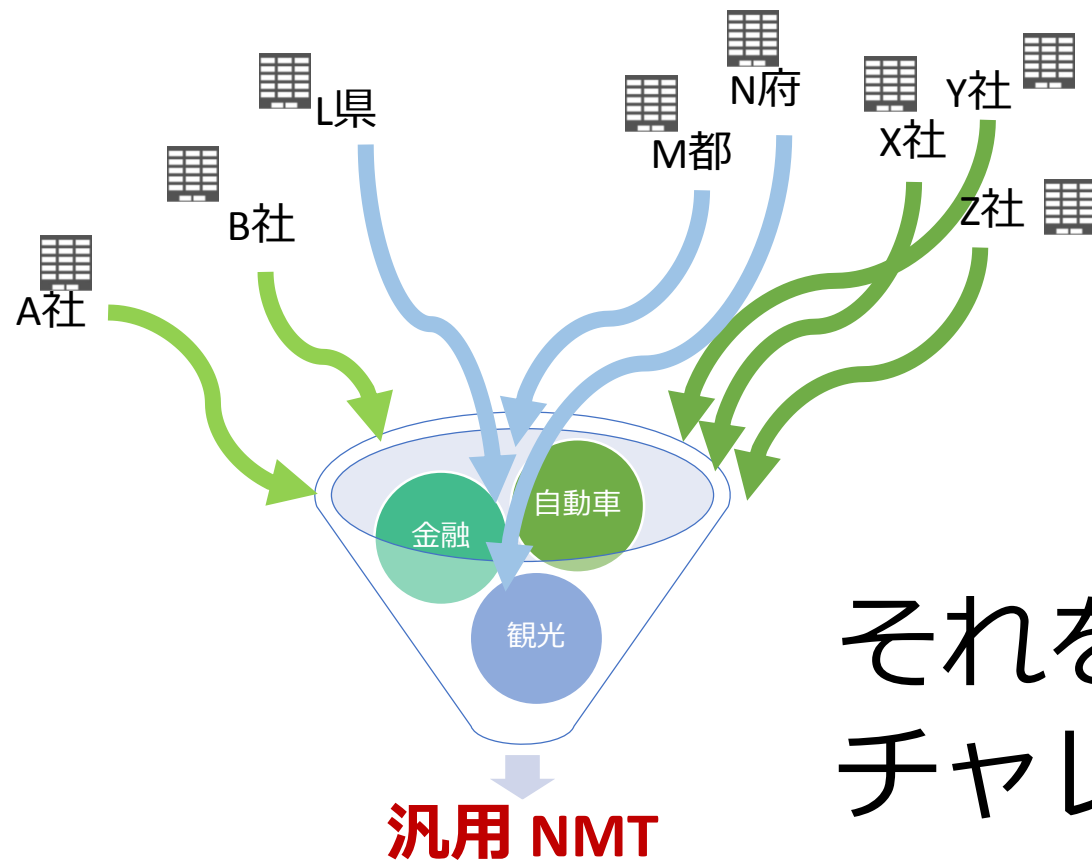


特許NMT



これは**前記**机である。 ???

汎用対訳コーパスが汎用 NMT には必要

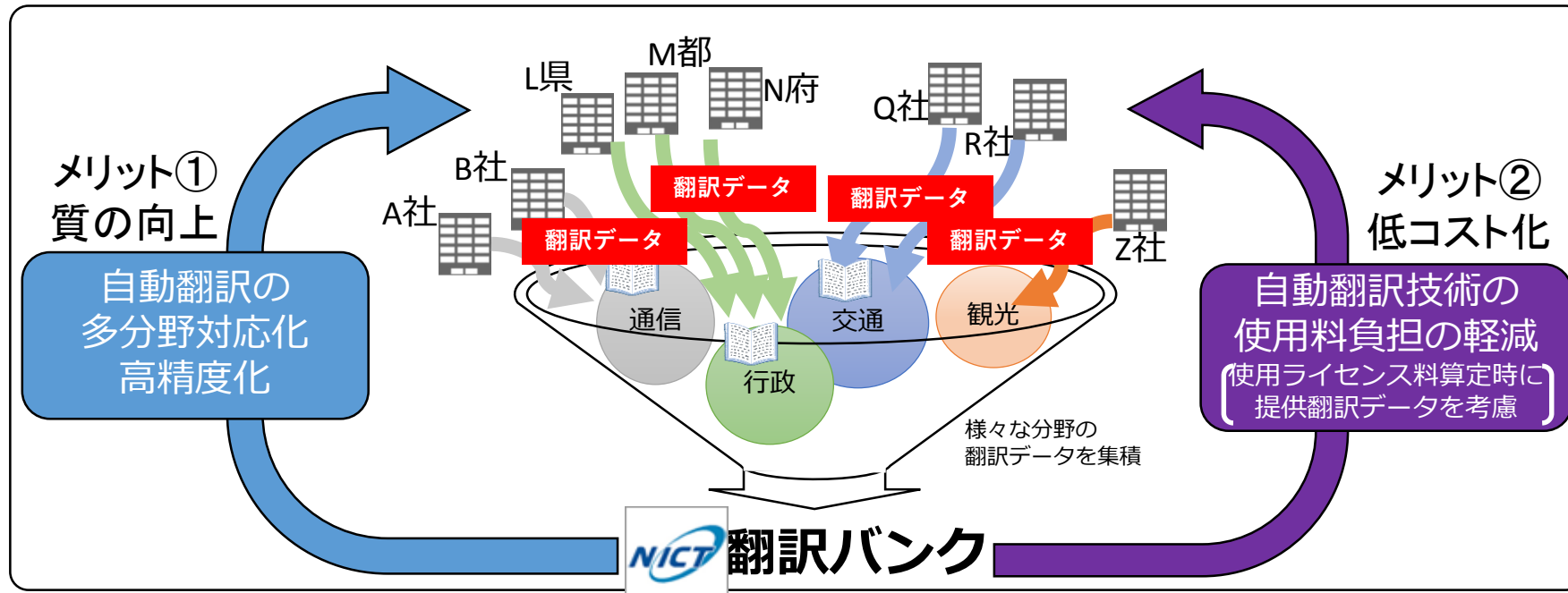


対訳コーパスは
散らばっている

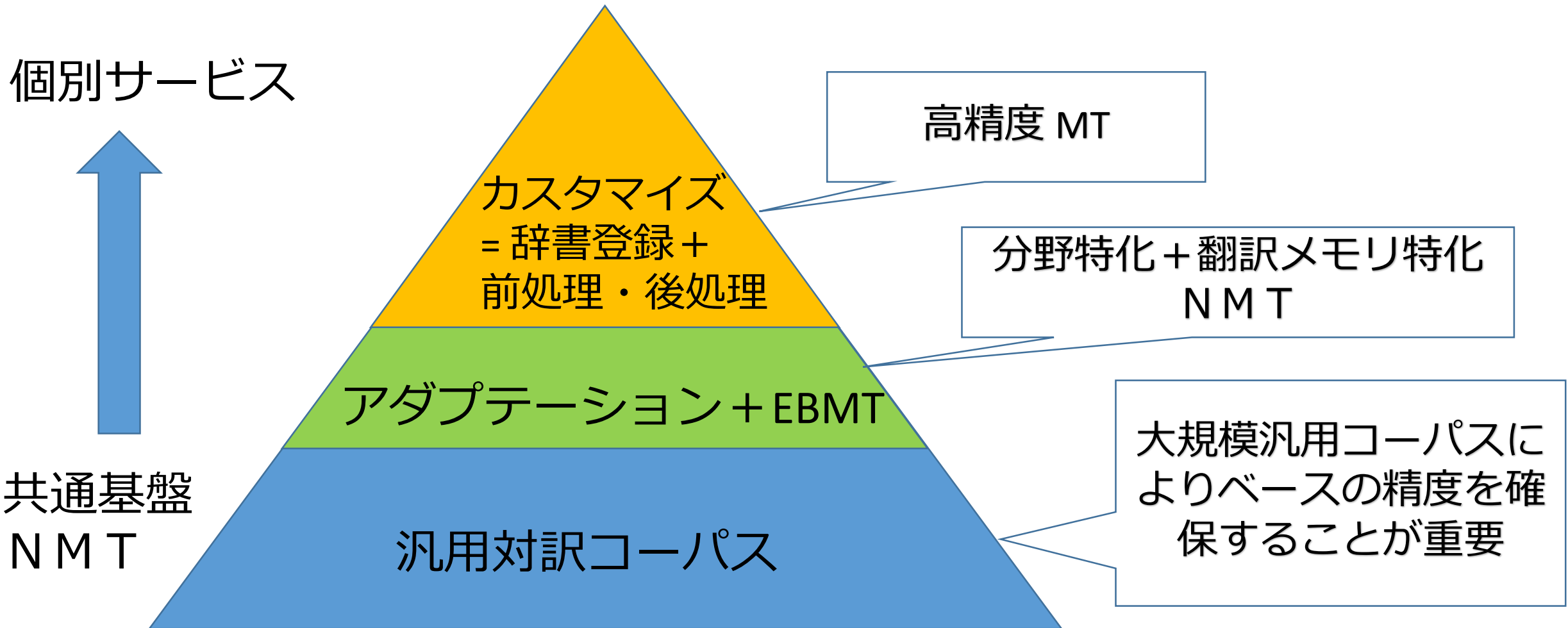
それを集めるのが
チャレンジ

翻訳バンクで高品質対訳データ

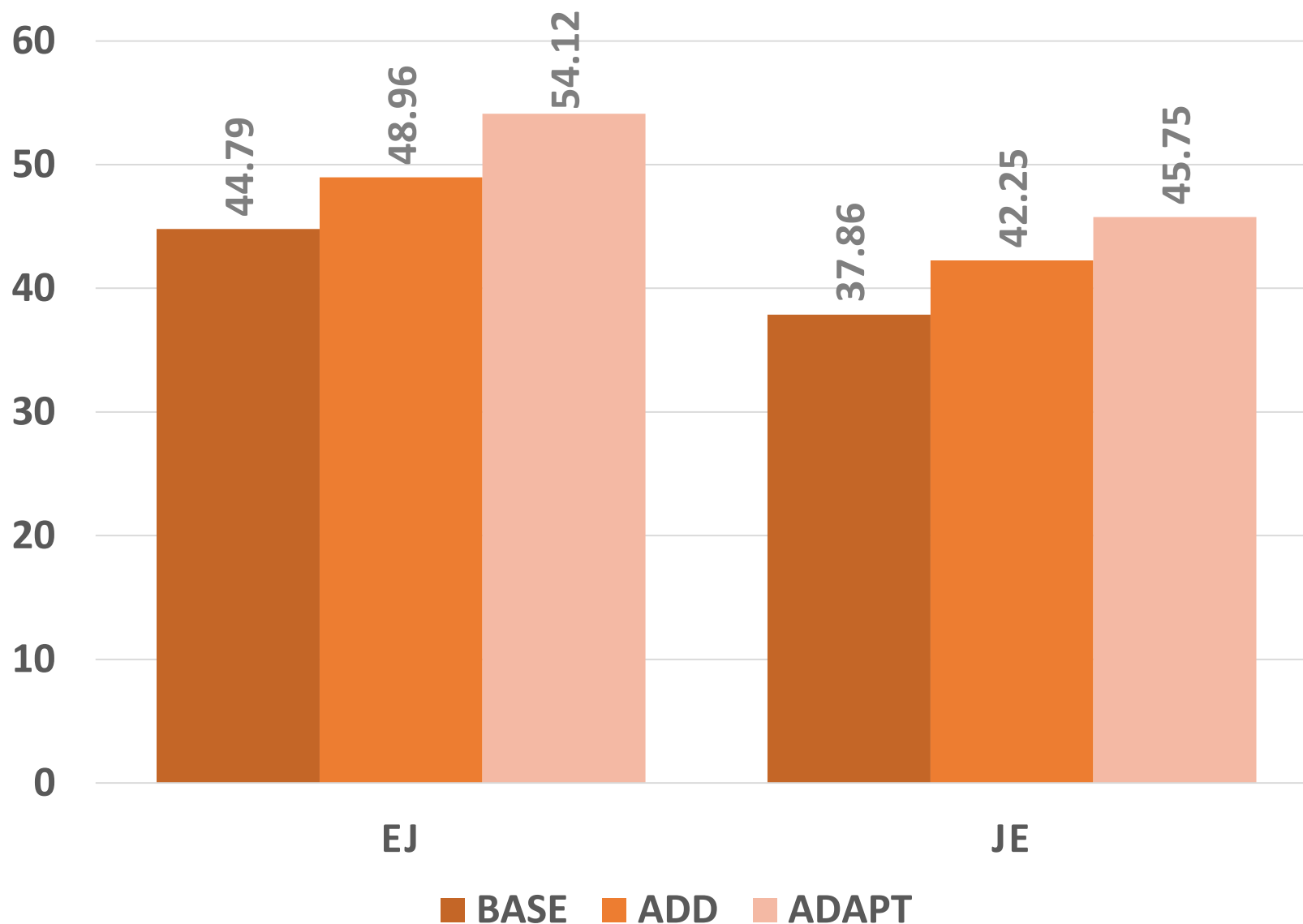
NICTの自動翻訳技術の使用ライセンス料の算定の際に、提供が見込まれる翻訳データを勘案して負担を軽減する仕組みを導入



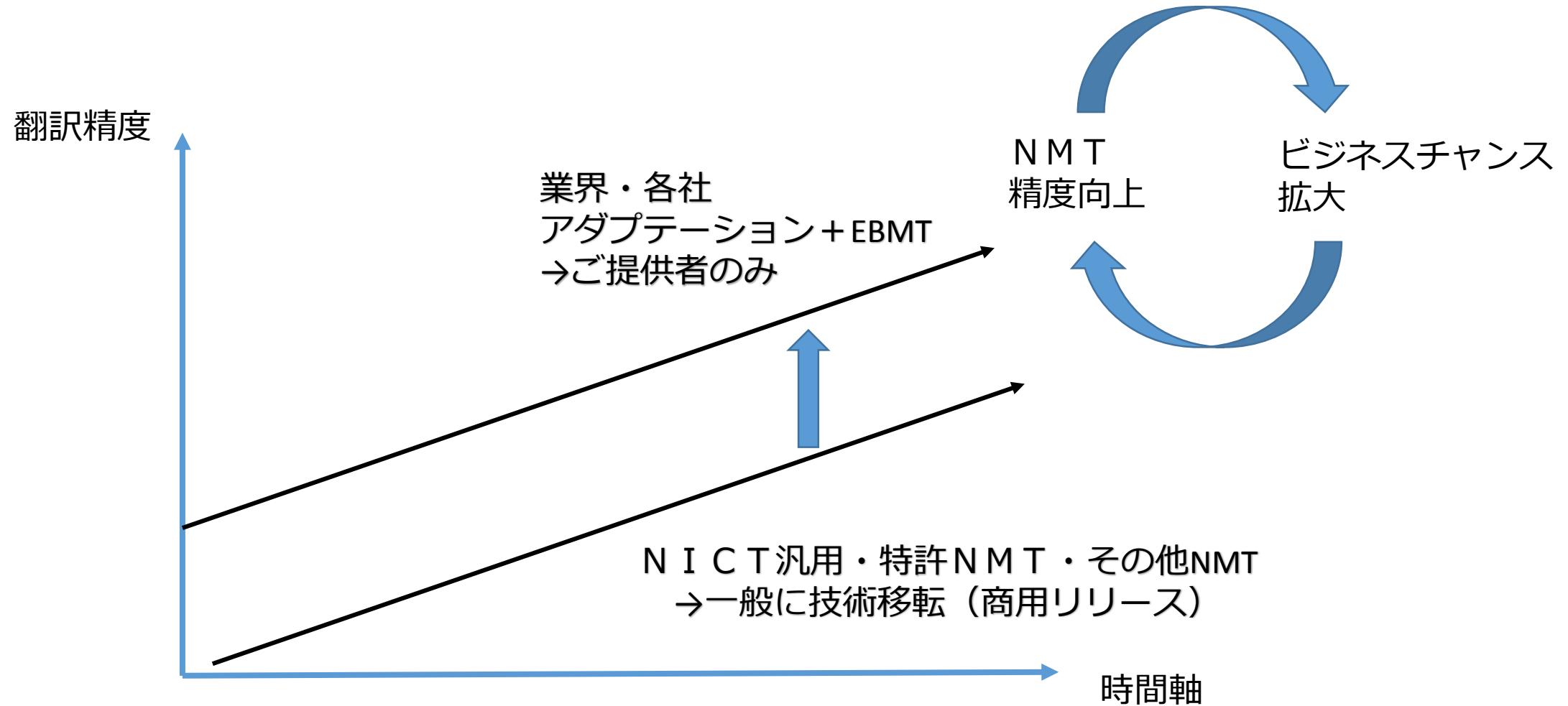
高精度MT = 汎用NMT + アダプテーション + EBMT + カスタマイズ



翻訳精度BLEUの2段階向上



対訳は2度使う



翻訳バンクへのご協力方法

1 『みんなの自動翻訳@TexTra®』の利用



2 NICTとの翻訳データの提供に関する契約の締結



3 ライセンス契約に伴う翻訳データの提供



翻訳データ量でライセンス料の負担軽減

NMTの安全な使い方



実用化の一方、誤訳の影響の拡大可能性

自動翻訳の特性を理解した
適切な使い方が必要

M Tの安全な使い方

前提：自動翻訳においては、自動翻訳結果の訳文の正確性は保証されません。



どのようにM Tを使うかが重要

* M Tサービスを利用するときには、その利用規約を参照してください。

自分がM Tを使う場合

- 自動翻訳結果に間違いが含まれる可能性を理解する
- 理解したうえで、利用する
- 英語の特許等を日本語で読む
- 日本語のメールを英語に自動翻訳して編集する
- 両言語がわかる人がM Tを使えば、非常に便利で安全な道具

M T の出力を他人が使う場合

- M T の出力を誰か（企業・ボランティア等）が修正する場合
 - 修正する企業等を信頼して利用する
- M T の出力をそのまま利用する場合
 - 何とかして翻訳を正確にしたい
 - 主語・述語が明瞭な文章を入力する。
 - 翻訳結果を再度日本語に翻訳して、意味が一致するかを確かめる
 - 複数の自動翻訳エンジンの翻訳結果が一致するかを確かめる
 - 日本語入力を言い換えた日本語入力について上記を実施する
 - 入力文とM T 訳文の双方を利用者に提示する。
 - M T の出力のため、間違いが含まれる可能性があることを知らせる。

まとめ

