

第 5 回特許情報シンポジウム

論文資料集

2018 年 12 月 7 日

ビジョンセンター浜松町

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

まえがき

特許情報の翻訳は、商用機械翻訳システム開発の初期の段階から重要なターゲットと考えられてきた。1987年に箱根で開催された第1回機械翻訳サミットにおいて、日本特許情報機構（JAPIO）から「発明の名称」の日英機械翻訳の試行について報告されている。その後、システムの改良や専門用語対訳辞書の整備が進み、現在では、「明細書」や「審査書類」の機械翻訳がそれぞれ公知例調査や審査結果の参照などに利用されている。言語対も日英、英日だけでなく日中、中日へと広がってきている。

アジア太平洋機械翻訳協会（AAMT）では、JAPIOの依頼と支援を受け、2003年度からAAMT/JAPIO 特許翻訳研究会を設置し、辻井潤一委員長のリーダーシップの下、特許の機械翻訳に関わるさまざまな技術や事例の調査研究を行ってきた。特許情報シンポジウムはその活動の一環として2010年、2012年、2014年、2016年に開催され、今回は第5回目である。前4回と同様に、研究者、開発者、利用者、あるいは政策担当者が議論する場を提供することによって、翻訳を中心とする特許情報処理の技術開発と利用を促進することが本シンポジウムの目的である。この目的に沿って以下のようなプログラムを編成した。

午後前半に、合計4組9名の招待講演者に、特許庁における機械翻訳への取り組みについてのご発表、ニューラル翻訳時代における機械翻訳システムの社会実装についてのご発表、ニューラル機械翻訳の研究動向についてのご発表、ユーザーから見たNMTの使い勝手と活用の展望に関するご発表をお願いした。これらの講演を通じて、機械翻訳の技術と特許分野における利用の現状について理解を深めていただけるものとする。

投稿ベースの一般発表も興味深い4編の論文を採用することができた。特許明細書の翻訳に関する論文に加えて、請求項の記載の構造化に関するご提案、特許明細書の高品質翻訳手法、効果的な特許検索特手法に関する論文をご発表いただく。

以上のように、特許情報処理に関するさまざまな立場からのさまざまな内容の発表からなるプログラムを編成することができた。快くお引き受けいただいた招待講演者と興味深い論文を投稿していただいた方々に感謝の意を表する次第である。この分野に関心をもつ参加者の皆様が意見を交換し、理解を深めることにより、特許情報処理がますます発展することを期待する。

2018年12月 第5回特許情報シンポジウム
実行/プログラム委員会
委員長 宇津呂武仁

特許庁における機械翻訳活用の 現状と今後の課題

平成30年12月7日

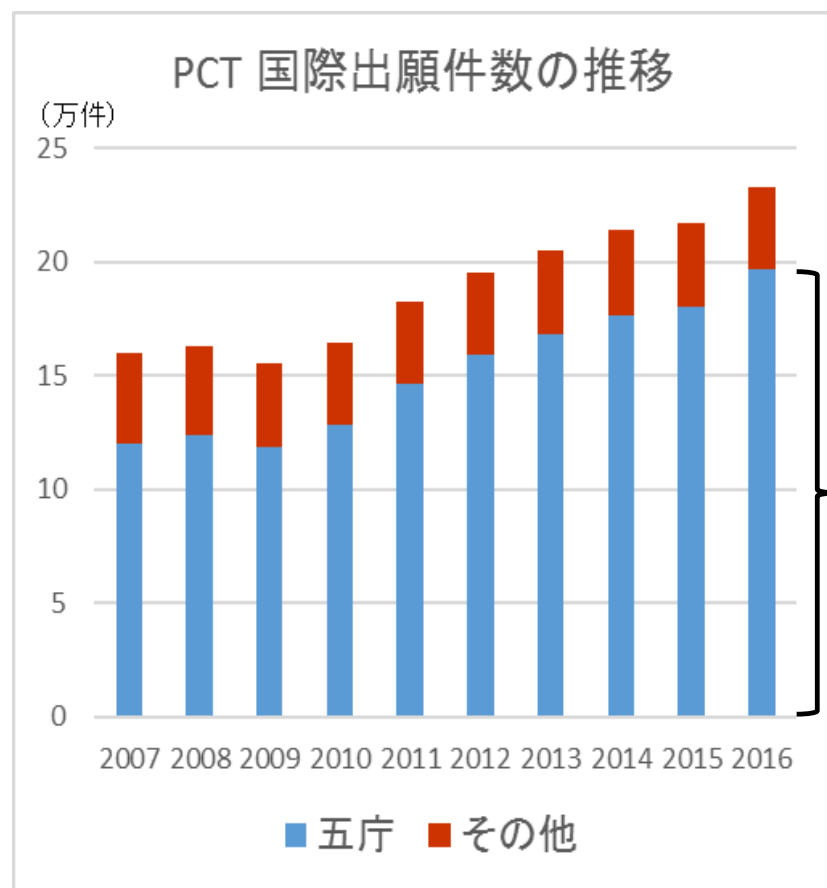
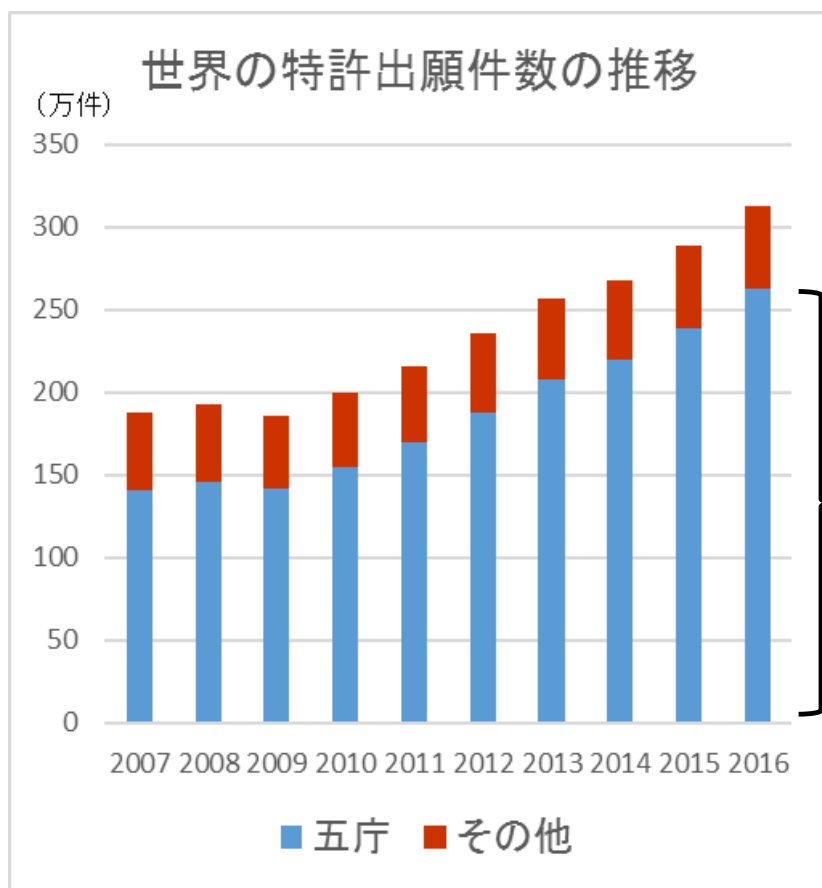
特許庁総務部総務課 特許情報室

松谷 洋平

1. 背景
2. 特許庁の機械翻訳の取組の現状
3. 海外の動向
4. 特許庁の今後の取組
5. まとめ

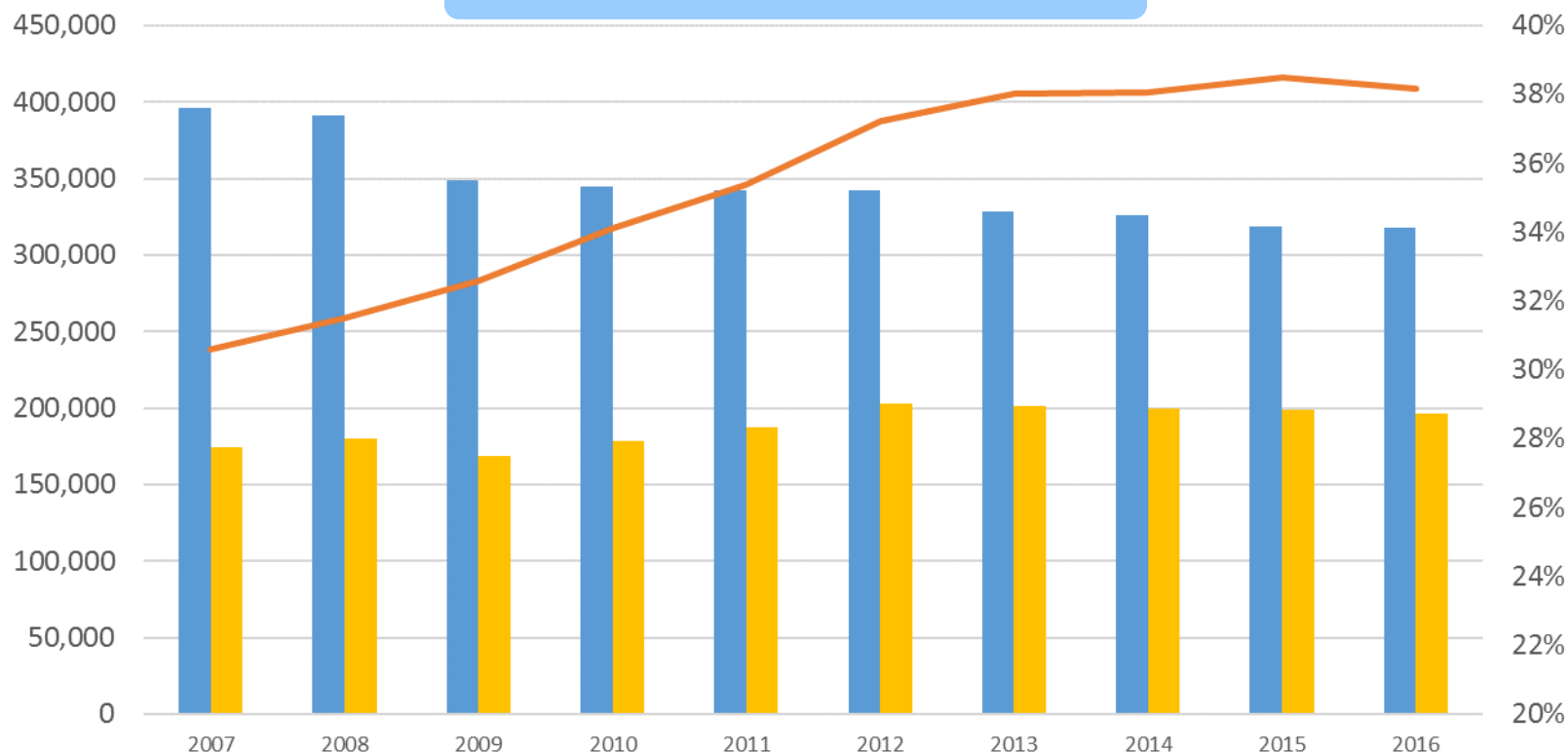
1. 背景
2. 特許庁の機械翻訳の取組の現状
3. 海外の動向
4. 特許庁の今後の取組
5. まとめ

- 世界の出願件数は、年々増加しており、2016年には313万件。
- PCT出願件数も、2016年には23.3万件となり増加傾向を維持。
- その中でも五庁（日米欧中韓）の出願件数は共に全体の80%以上を占める。



- 海外出願割合は過去10年間、増大傾向。
- ここ数年は落ちついているものの、海外市場の拡大が見込まれるなか、今後も海外出願の重要性に変わりはない。

国内及び海外出願件数の比較

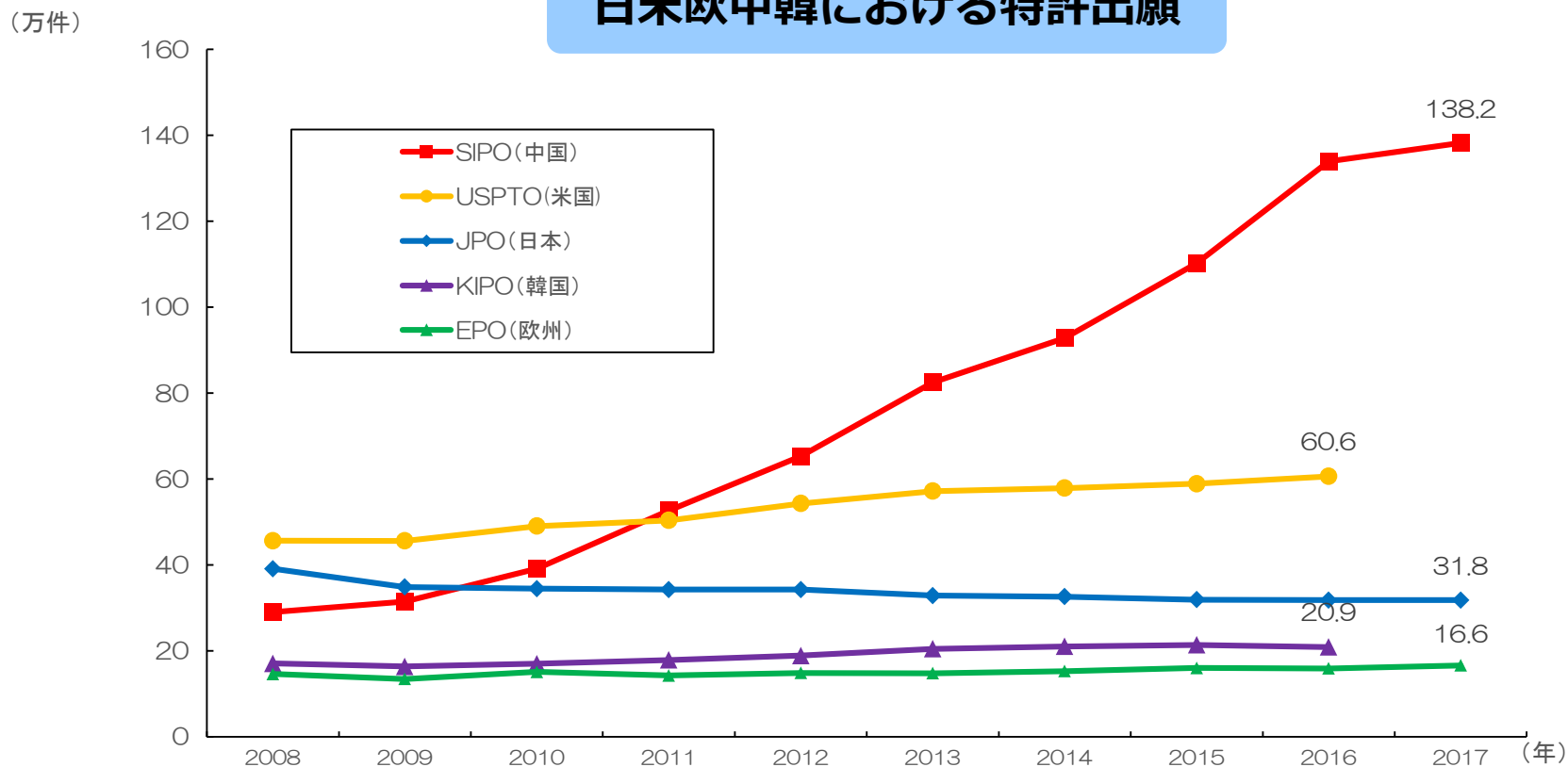


■ 国内出願件数 ■ 日本から海外への出願件数 — 海外出願割合 (出典) 特許庁作成

※海外出願割合は、国内出願件数、日本から海外への出願件数の合計に対する、日本から海外への出願件数の割合。

- 日米欧中韓（五大特許庁）において中国出願の件数が依然として突出。
- 日本の出願件数は、特許出願の量から質への転換が図られていたが、下げ止まりの傾向にある。

日米欧中韓における特許出願



7

(出典)

WIPO IP Statistics Data Center、特許行政年次報告書2017および各国ウェブサイトを基に特許庁作成

(出願年)

- JPOの「世界をリードする審査」を核に、日本の審査結果を国際的に発信し、海外知財庁での利用促進を図る。

世界最速

権利化までの期間※： 平均 1 4 か月以内
一次審査期間： 平均 1 0 か月以内

(平成35年度目標)

※出願人が補正等を行うことに起因して特許庁から再度の応答等を出願人に求めるような場合や、特許庁に応答期間の延長や早期の審査を求める場合等の、出願人に認められている手続を利用した場合を除く。

最高品質

品質管理の取組強化

外国文献調査の充実

面接審査関連施策の拡充

第四次産業革命への対応



日本の審査結果の発信強化

制度・運用調和

P P H

C P G

日米協働調査

P C T 協働調査

ドシエ共有

途上国支援

デジタル・ガバメント実行計画

（行政の在り方そのものをデジタル前提で見直すデジタル・ガバメントの実現）

- 「デジタル・ガバメント推進方針」などに示された方向性を具体化し、実行することによって、安心、安全かつ公平、公正で豊かな社会を実現するための計画
- また、IT宣言・官民データ計画に掲げられた重点分野の一つである電子行政分野を深掘りし、詳細化した計画

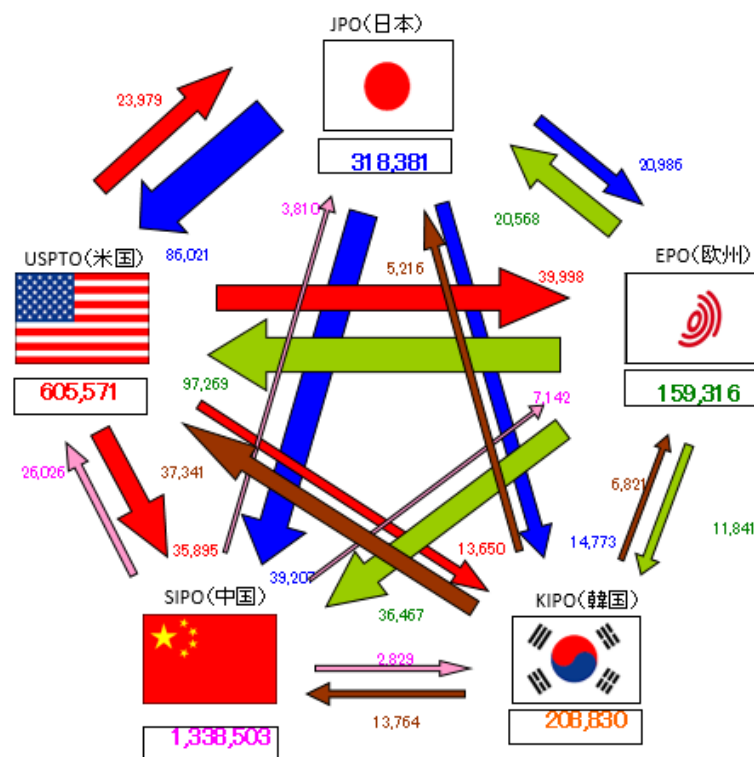
経済産業省デジタル・ガバメント中長期計画

（経済産業省におけるデジタル・ガバメントの推進）

- 行政自身の効率化のみならず、民間における生産性を向上させ、社会全体の生産性革命を強力に推進していく指針
- **特許庁の取組：**
 1. AI技術を活用したシステム構築に向けた体制整備
 2. 特許情報提供サービスの迅速化
 3. 特許庁における業務・システム最適化

- 出願人の国際的知財ポートフォリオ作成に要する、出願人、特許庁双方のイニシャルコストを下げる

【日米欧中韓の特許出願状況(2016年)】



(出典) 特許庁調べ

○出願人の負担

- ・各国からの重複した拒絶理由などへの対応
→ 出願コストの増大

○特許庁の負担

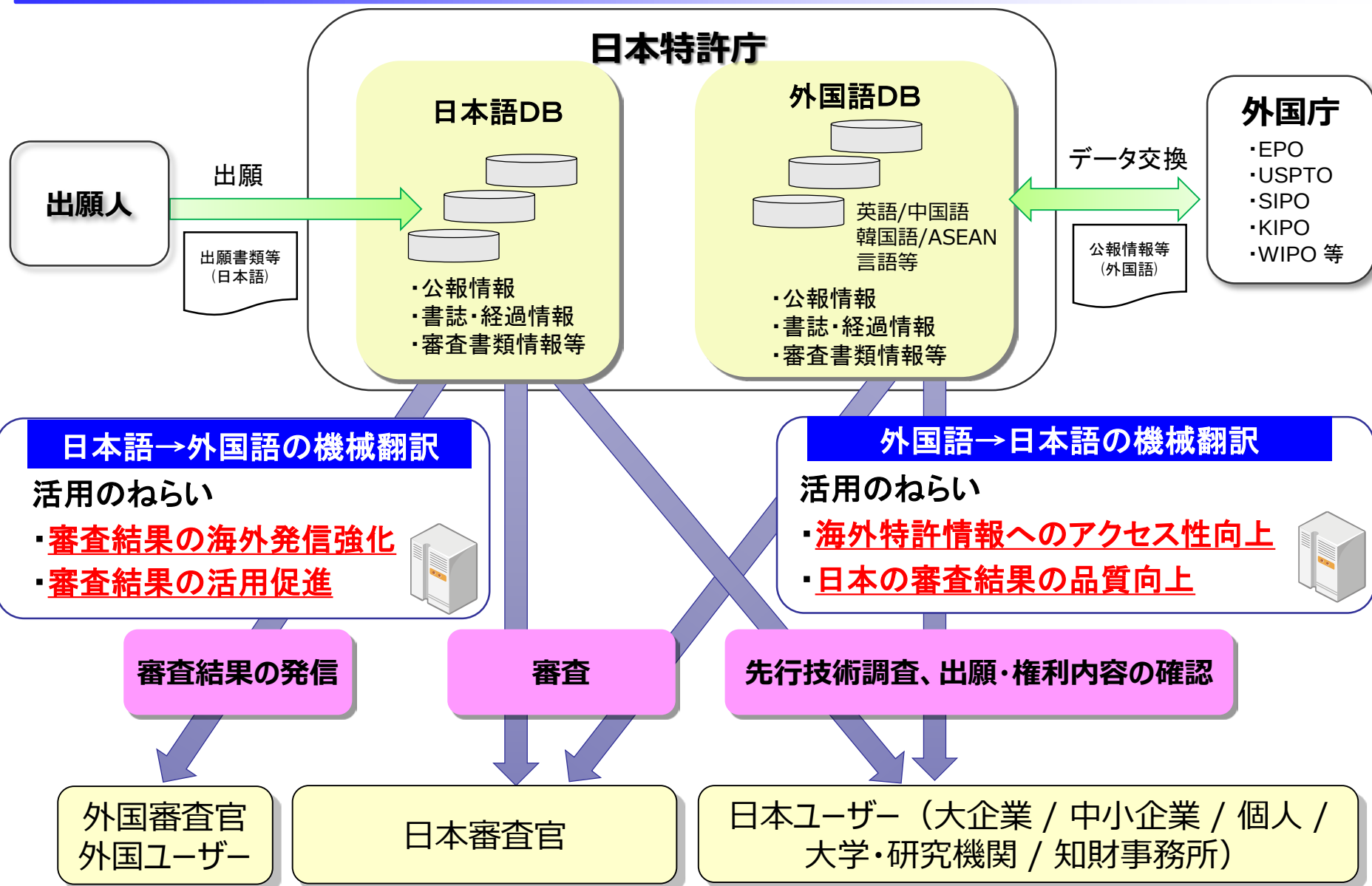
- ・同内容の発明について、各国が各々審査
→ 審査遅延、審査コスト

出願人、特許庁双方の負担軽減のため

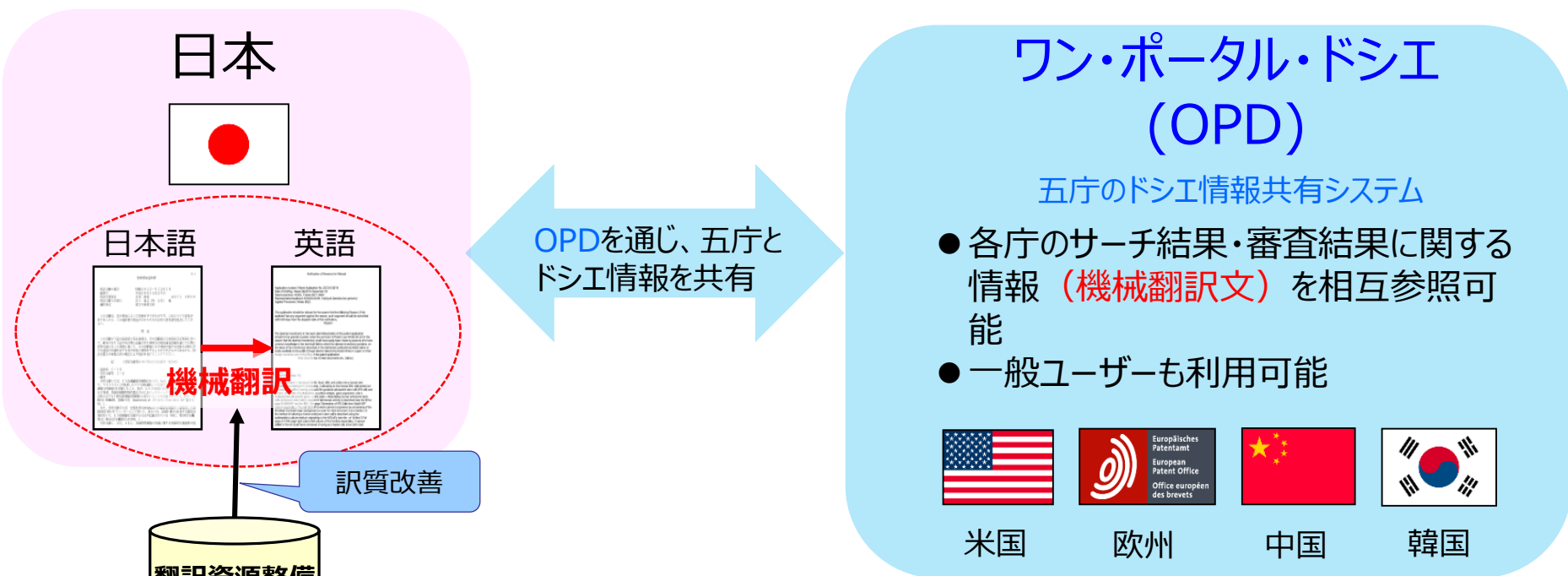
- ・制度・運用の調和
- ・出願手続の利便性向上
- ・審査のワークシェアリングが必要

1. 背景
- 2. 特許庁の機械翻訳の取組の現状**
3. 海外の動向
4. 特許庁の今後の取組
5. まとめ

特許庁における機械翻訳活用のねらい

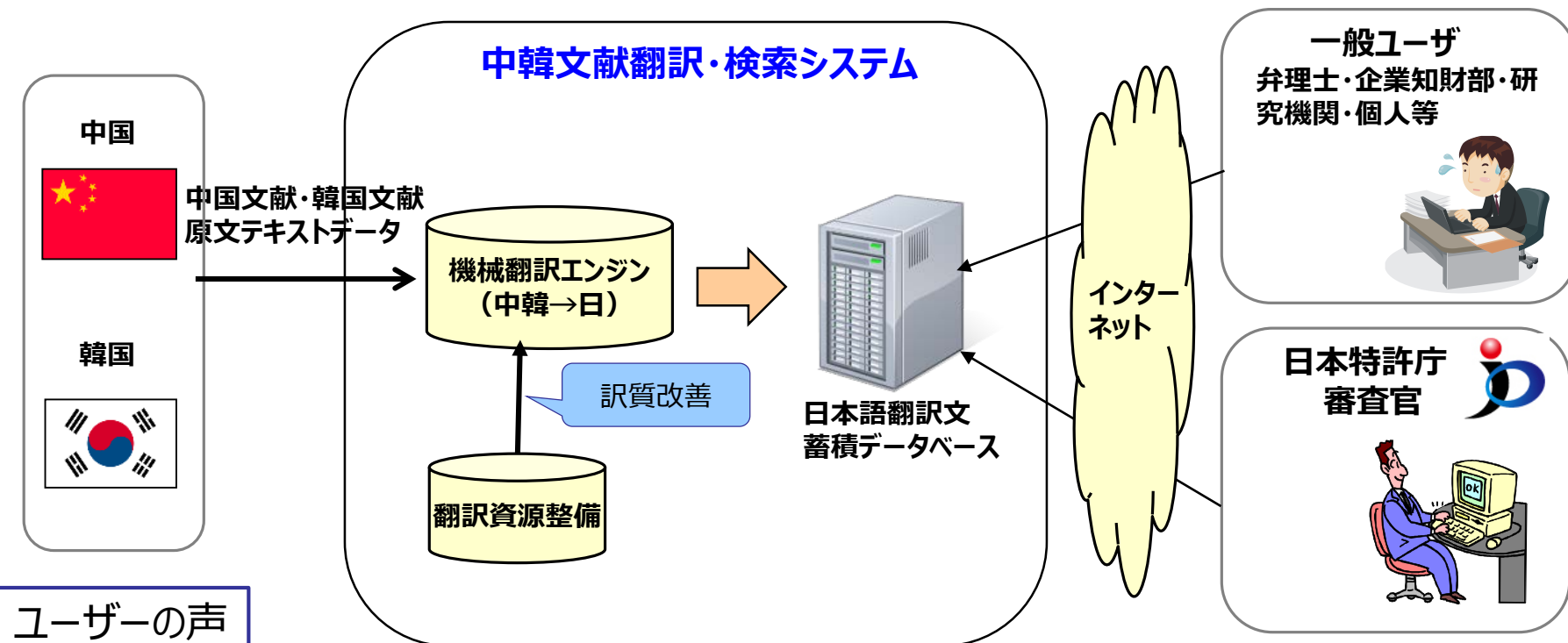


- 海外の審査官、一般ユーザーに向けて、我が国の「世界最速・最高品質の特許審査」によるドシエ情報（特許出願の手续や審査に関連する情報）を日英機械翻訳文で提供。



- 辞書データ・対訳コーパスの作成
- 誤訳の分析
- 未知語の辞書登録
- 翻訳メモリの構築

- 急増する中国・韓国語特許文献の検索の利便性を向上させるために、中国・韓国語特許文献の全文を日本語に機械翻訳し、日本語でテキスト検索及び照会できるシステムを提供。



ユーザーの声

- ✓ 中国での係争が増えていることから、中韓文献翻訳・検索システムを利用することがある。（国内ユーザ）
- ✓ 中国は生薬関連の特許が多いため、生薬関連の先行技術を調査するのに、中韓文献翻訳・検索システムを利用している（国内ユーザ）

- 14

- インターネットと分離された環境
に構築されているので未公開
案件も安心して翻訳できる

English → Japanese ▼ ■ Claim mode

JPO NICT Translation

インター
に構築
案件

FIG. 542 illustrates a value of the RL of the antenna structure 50 when the antenna structure 50 has no second connecting section 573. Curve 543 illustrates a value of the RL of the antenna structure 50 when the antenna structure 50 has the ground portion 59 and the second connecting section 573. It can be derived from FIG. 4 that the ground portion 59 in a low-frequency mode of the antenna structure 50. The second connecting section 573 is configured to broaden a bandwidth of the antenna structure 50 in a high-frequency mode. FIG. 5 illustrates that when the capacitance value of the first capacitor C1 is about 4 pF, a bandwidth of the first low-frequency mode can be about 704 MHz to about 894 MHz. When the capacitance value of the first capacitor C1 is about 20 pF, a bandwidth of the first low-frequency mode can be about 750 MHz to about 960 MHz (see curve 553). A high-frequency mode can be about 1710 MHz to about 2690 MHz. As shown in FIGS. 3-5, the antenna structure 50 and the wireless communication device 100 employing the antenna structure 50 can be utilized in common wireless communication systems with exceptional communication quality.

TRANSLATE

Loading: 21/27 Lines

Download Translated Text

Line	Original	Translated
1	This application claims priority to Chinese Patent Application No. 201310713082.X filed on Dec. 23, 2013, the contents of which are incorporated by reference herein.	本出願は2013年12月23日に出願された中国特許出願第201310713082.X号(これらの内容は、参照により本明細書に組み込まれる優先権を主張する。
2	The subject matter herein generally relates to an antenna structure and a wireless communication device using the antenna structure.	本明細書の主題は、一般に、アンテナ構造を使用してアンテナ構造および無線通信装置である。
3	Antennas are important elements of wireless communication devices, such as mobile phones or personal digital assistants. Many wireless communication devices further employ metal housings for improving heat dissipation or other purposes.	アンテナは、無線通信装置の重要な要素、例えば、移動電話や携帯情報装置は、さらに、放熱又は他の目的を改良するための金属ハウジングを使用する。
4	Implementations of the present technology will now be described, by way of example only, with reference to the attached figures.	本技術の実施態様は、単なる例として、添付の図面を参照して説明する。
5	FIG. 1 is an isometric view of an embodiment of a wireless communication device employing an antenna structure.	図1は、アンテナ構造を使用する無線通信デバイスの一実施形態の等角図である。
6	FIG. 2 is a circuit view of a matching circuit of the wireless communication device of FIG. 1.	図2は、図1の無線通信装置の整合回路の回路図である。
7	FIG. 3 is a return loss (RL) graph of the antenna structure of the wireless communication device of FIG. 1.	図3は、図1の無線通信装置のアンテナ構造のリターンロス(RL)のグラフである。
8	FIG. 4 is another RL graph of the antenna structure of the wireless communication device of FIG. 1.	図4は、図1の無線通信装置のアンテナ構造の別のRLのグラフである。
9	FIG. 5 is a RL graph of the antenna structure of FIG. 1, when changing a capacitance of a first capacitor of the matching circuit of FIG. 2.	図5は、図1のアンテナ構造のRLのグラフであり、図2の整合回路の第一コンデンサの容量を変化させたときのRLのグラフである。
10	It will be appreciated that for simplicity and clarity of illustration, where appropriate, reference numerals have been repeated among the different figures to indicate corresponding or analogous elements. In addition, numerous specific details are set forth in order to provide a thorough understanding of the embodiments described herein. However, it will be understood by those of ordinary skill in the art that the embodiments described herein can be practiced without these specific details. In other instances, methods, procedures and components have not been described in detail so as not to obscure the related relevant feature being described. Also, the description is not to be	図示の単純化及び明瞭化のために、適切な場合には、異なる図面に示す類似の要素を示すために参照番号が理解されるであろう。本明細書に記載される実施形態は、理解を与えるために多数の特定の詳細が記載されているが、本明細書に記載の実施形態は、必ずしも図面に示すように実施されるべきものではない。むしろ、本明細書に記載の実施形態は、特定の要素を省略し、または図面に示すよりも詳細に示すことができる。また、本明細書に記載の実施形態は、必ずしも図面に示すように実施されるべきものではない。むしろ、本明細書に記載の実施形態は、特定の要素を省略し、または図面に示すよりも詳細に示すことができる。また、本明細書に記載の実施形態は、必ずしも図面に示すように実施されるべきものではない。むしろ、本明細書に記載の実施形態は、特定の要素を省略し、または図面に示すよりも詳細に示すことができる。

翻訳文
なっ

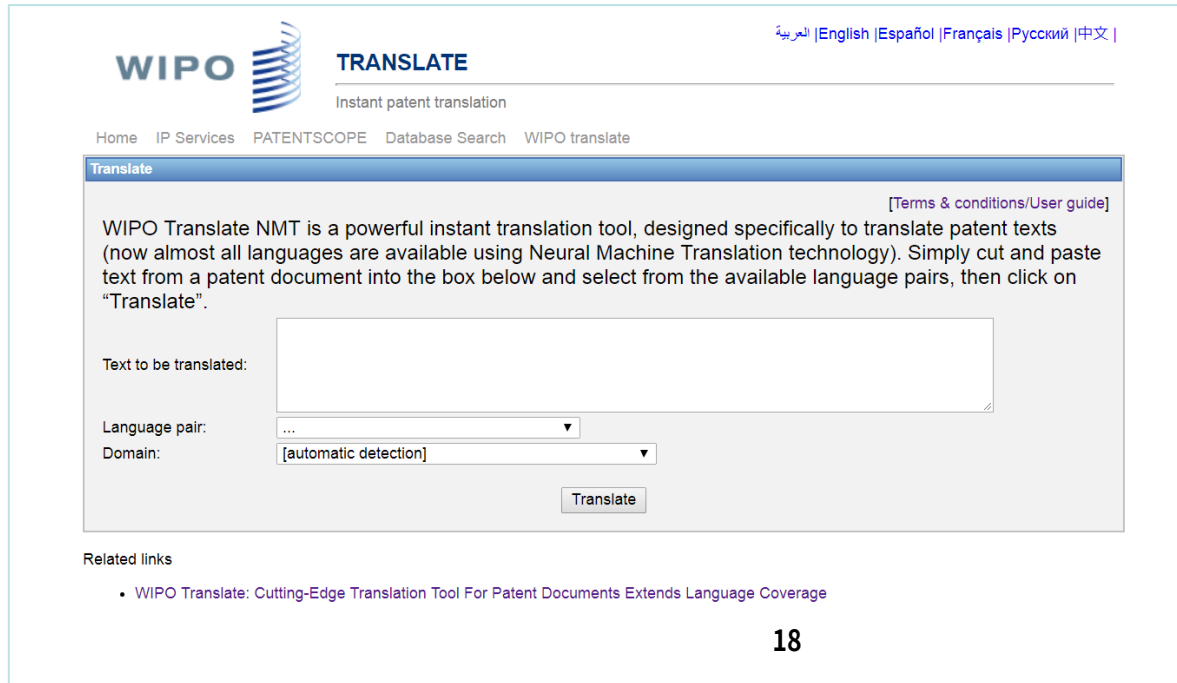
翻訳速度

翻訳文が特許特有の文章になっていて違和感がない

翻訳速度が速くてよい

1. 背景
2. 特許庁の機械翻訳の取組の現状
- 3. 海外の動向**
4. 特許庁の今後の取組
5. まとめ

- WIPO（世界知的所有権機関）が開発した機械翻訳ツールであり、WIPOが提供する特許データベース検索サービスPATENTSCOPEで使用可能。2016年9月より、ニューラル機械翻訳を採用し翻訳精度を大幅に向上させた。
- WIPOは各国知財庁にWIPO Translateのソースコードの無償提供を発表。
- WIPO TranslateはMarian NMTをベースに、PATENTSCOPEに蓄積された6500万の特許文献等のデータを学習。PCT言語（（英、仏、独、露、西、葡、中、韓、日、アラビア）⇔英）の言語対に対応。



The screenshot shows the WIPO Translate web interface. At the top, there is a navigation bar with the WIPO logo, the word "TRANSLATE", and a list of languages: العربية | English | Español | Français | Русский | 中文. Below this is a sub-header "Instant patent translation". The main content area has a blue header "Translate" and a link "[Terms & conditions/User guide]". The text describes the tool: "WIPO Translate NMT is a powerful instant translation tool, designed specifically to translate patent texts (now almost all languages are available using Neural Machine Translation technology). Simply cut and paste text from a patent document into the box below and select from the available language pairs, then click on 'Translate'." There is a large text input box labeled "Text to be translated:". Below it are two dropdown menus: "Language pair:" and "Domain:". The "Domain:" dropdown is currently set to "[automatic detection]". A "Translate" button is located at the bottom right of the input area. At the bottom, there is a section "Related links" with a single link: "WIPO Translate: Cutting-Edge Translation Tool For Patent Documents Extends Language Coverage".

- 2010年、Googleと特許の機械翻訳に関する協力を発表。2013年、同協力の下、機械翻訳サービス“Patent Translate”を開始。2018年現在、32言語をカバー（↔英、独、仏語）。
- 2017年5月、ニューラル機械翻訳の採用を発表。EPOがGoogleに特許文献のデータを提供し、特許に最適化された翻訳エンジンを開発。



The screenshot shows the Patent Translate web interface. At the top, it says "Patent Translate Powered by EPO and Google". On the left, there is a language selection menu with "French" and "German" selected. The main content area displays a patent abstract in German, titled "ZUSAMMENFASSUNG WO2018198161". The text describes an endoscope image processing apparatus. On the right side, there are options to "Print" the translation (PDF only or PDF original and translation) and a feedback section titled "Please help us to improve the translation quality." with a rating scale from "Human translation" to "Very bad".

1. 背景
2. 特許庁の機械翻訳の取組の現状
3. 海外の動向
- 4. 特許庁の今後の取組**
5. まとめ

翻訳資源の充実化

- 辞書データ・対訳コーパスの作成
- 定型文データの整備、新語対応

技術動向の調査

- 最新の機械翻訳技術の把握
- リソース等に関する調査

NICTとの協力

庁内で試行利用

急成長をとげる翻訳技術への即時対応

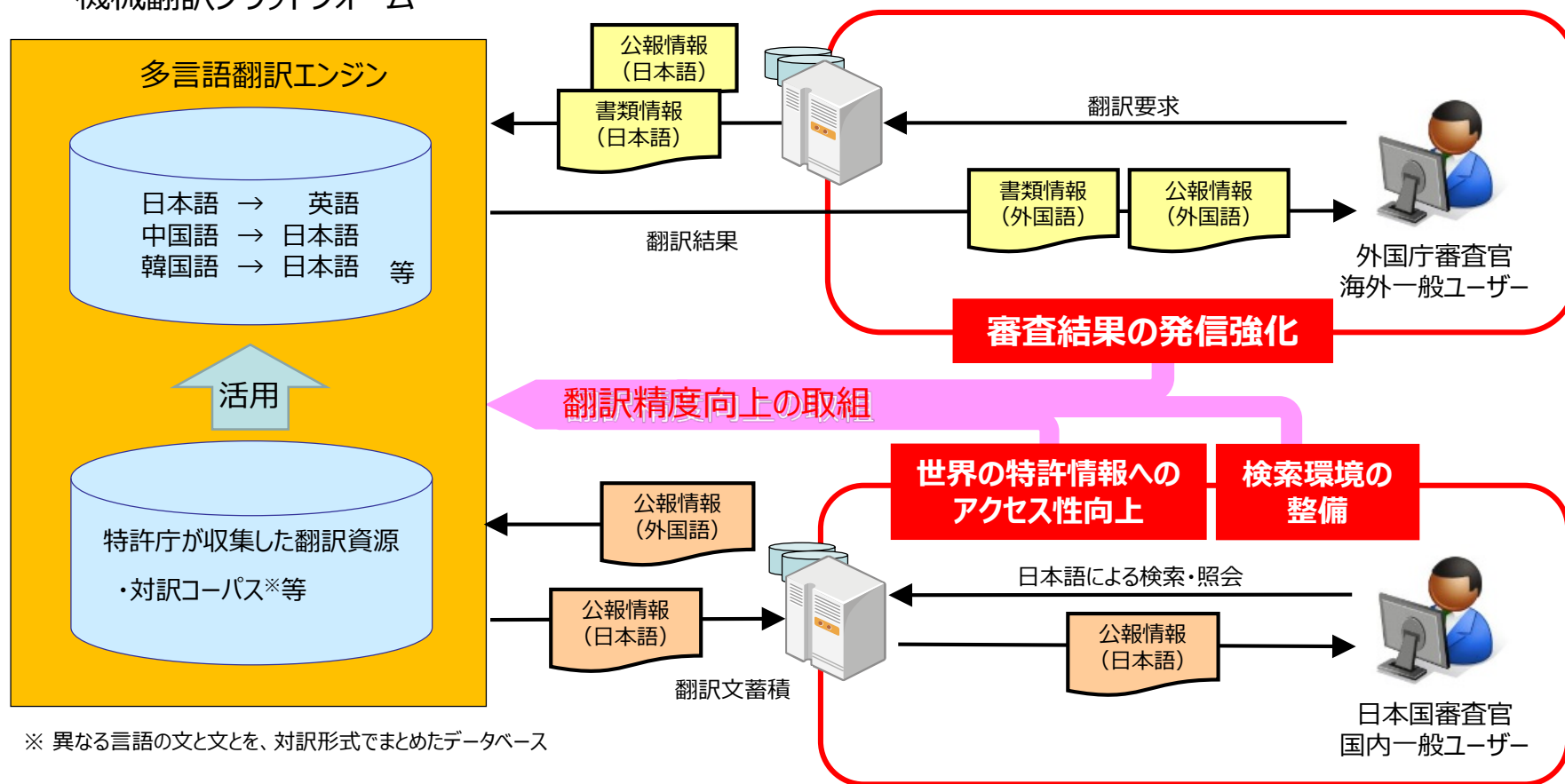
機械翻訳プラットフォームの開発

- 庁外のさまざまなシステムからの翻訳要求に対応
- ニューラル機械翻訳の採用

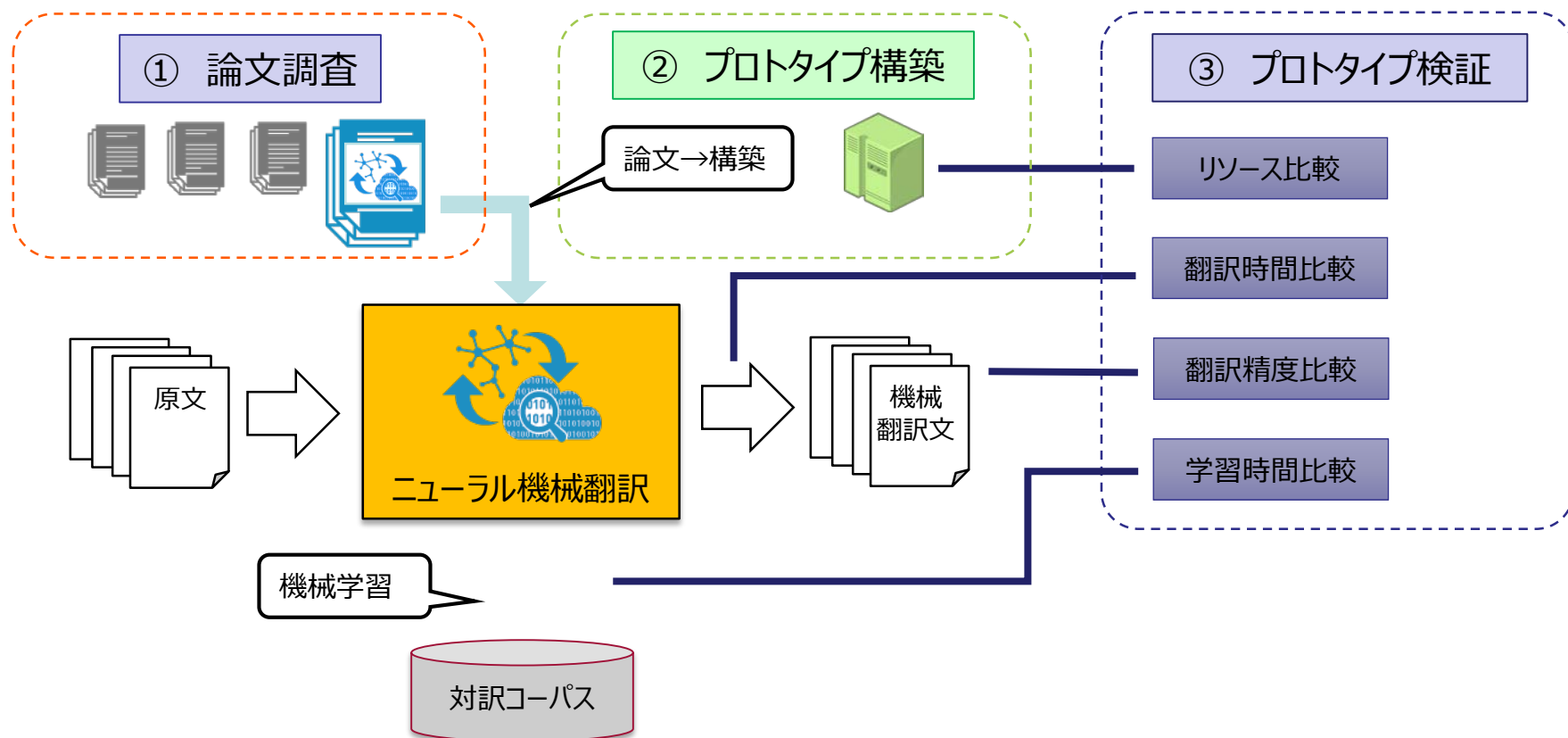
- 海外への情報発信強化
- 外国情報へのアクセス性向上

- 世界の特許情報へのアクセス性向上に対するユーザーニーズが高まる中、審査結果の発信強化や検索環境の整備を推進する観点から、多言語に対応可能とし、より利便性の高い機械翻訳文を提供するため、特許庁保有の翻訳資源（対訳コーパス等）を活用した翻訳エンジンに基づく機械翻訳プラットフォームのリリースを予定。

機械翻訳プラットフォーム



- 現在研究されている最先端のニューラル機械翻訳について詳細を調査し、今後の機械翻訳システムの検討の為の基礎資料とする。



1. 背景
2. 特許庁の機械翻訳の取組の現状
3. 海外の動向
4. 特許庁の今後の取組
5. **まとめ**

- 特許庁は、急増する外国文献へのアクセス性向上、及び、日本の高品質な審査結果の海外発信強化のため、機械翻訳システムを活用している。
- 海外でもWIPOやEPOなど、ニューラル機械翻訳を活用した機械翻訳サービスを提供する機関が出てきている。
- 特許庁は、ニューラル機械翻訳による翻訳エンジンを採用した「機械翻訳プラットフォーム」を提供予定。
- 今後は、翻訳資源が不足している審査書類等の訳質向上のため、翻訳資源作成及び調査分析に係る事業を実施予定。

**特許庁は、機械翻訳研究の一層の進展を
期待しています！！！！**

ニューラル翻訳時代の機械 翻訳システムの社会実装

豊橋技術科学大学
井佐原均

我が国の機械翻訳

- 1950年頃に研究が始まる
- 1980年代は日本が世界のトップ
 - 多くの企業の参入、国際会議での発信
- **日本の(機械翻訳ビジネスの)衰退**
 - 当時の機械翻訳の精度は低かった(特に日本語と他言語の間)
 - **機械翻訳システムを翻訳サービスの一環と考える視点の欠如**
- 1990年頃からデータを利用した機械翻訳の発展
 - 用例翻訳システム、統計翻訳システム
 - 大量の言語データの存在とコンピュータの性能向上

我が国の機械翻訳(続き)

- 統計翻訳からニューラル翻訳への移行
 - 深層学習・ニューラルネットを用いた機械翻訳システム
 - 単語の置き換えから、文から文への置き換え(翻訳)へ。
 - Google や Microsoft の発表で話題に。
 - 一般的な人よりも翻訳能力が高い。流暢度は完ぺき。忠実度も人間並み。
- 機械翻訳を利用したいというユーザが増加。

産業界での国際競争力の強化に向けて

- 国際競争力強化には**産業文書**(ノウハウ文書、取扱説明書など)の**多言語化**が必須
 - 企業の海外展開(海外拠点)の支援、製品の輸出、知財貿易など、波及効果は膨大
 - 産業に関する情報を表現する日本語をこれまで以上に、**迅速に、正確に、低コストで翻訳**することが重要
- ◆ より多くの製品(**サービス**)に、より多くのビジネスチャンス(時間)を

物の販売からサービスの販売へ

機械翻訳
システム

大量の対訳データを使って
システムが自動学習

産業界での国際競争力の強化に向けて

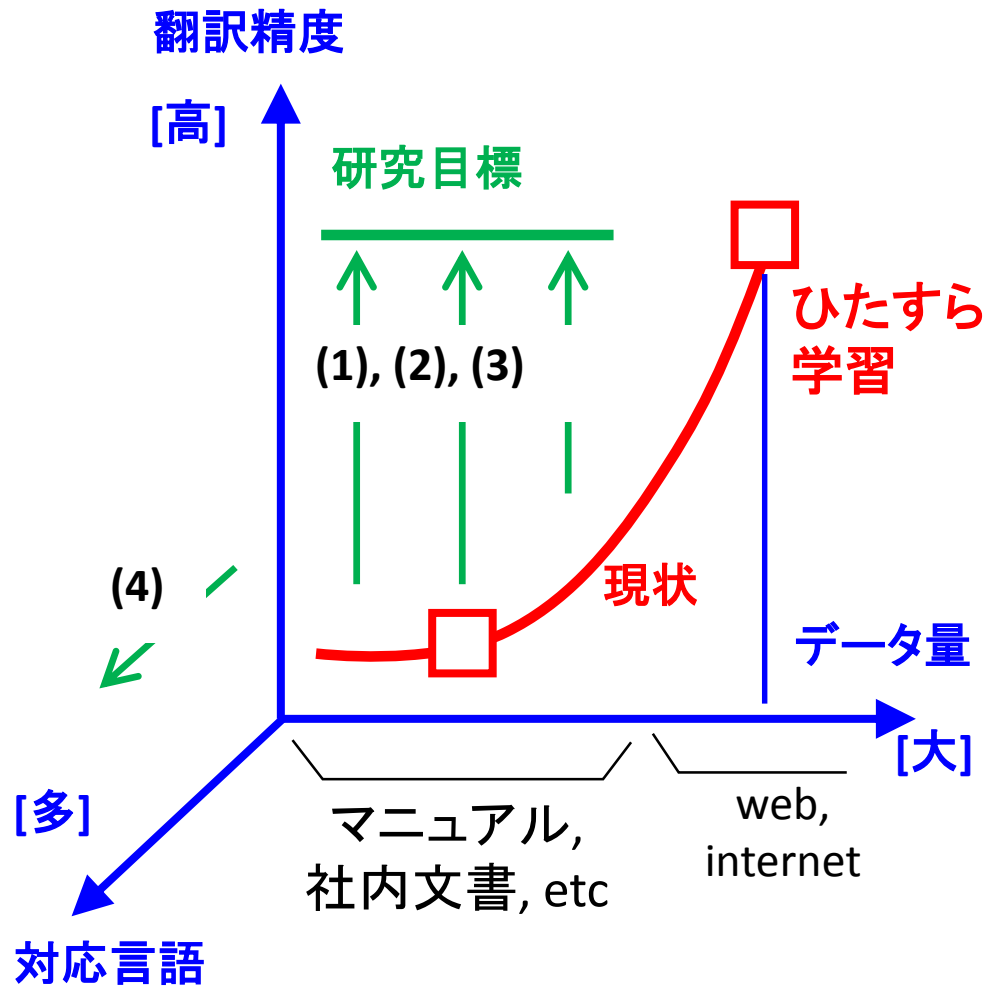
データの蓄積、再利用、共有

- **企業間のデータ共有**
 - 機械翻訳を用いた翻訳プロセスの完成度を高める
 - 自社文書の提供(公開)が、より大きなビジネスチャンスに繋がる
- 既に翻訳した文書を活用することにより、新たな文書の翻訳コストを削減できる
- コスト削減によって、これまで訳していなかった大量の文書が翻訳可能に

非競争領域での協力により、
競争領域での個々の競争力の
強化に集中できる。

欧米企業
で実証済

機械翻訳の現状と研究目標



【日本語⇄英語】

- (1) 個別分野への適用
- (2) 語彙数の増加に対応できるエンジン
- (3) 誤翻訳検出
英語に特化して精度向上

【次段階】 多言語に展開

- (4) 英語⇄多言語
・留学生の活用

【将来ビジョン】

日本文化の多言語化
 例: 小説、俳句、落語等

本学の機械翻訳研究の考え方

- それぞれの位置のプレイヤーとの連携
 - 機械翻訳エンジンベンダー
 - 日本マイクロソフト
 - 応用システム開発
 - ブロードバンドタワー
 - ○○
 - ユーザ
 - 港区（自治体情報の自動翻訳の実証）
 - ブリックス（通訳サービスへの応用）
 - グローバルデザイン（自治体向け翻訳サービス）
 - ○○



豊橋技科大の機械翻訳(共同)プロジェクト

- 豊橋技科大、日本マイクロソフト、ブロードバンドタワー

6月22日 15時 日本経済新聞

訪日客に即時翻訳 マイクロソフトなど

日本マイクロソフト(MS)、ブロードバンドタワー(BBタワー)、豊橋技術科学大学の3者は21日、機械学習技術を使った訪日外国人向けリアルタイム翻訳サービスの開発で協力すると発表。訪日外国人が母国語で日本人と対話しても自然にコミュニケーションできる水準をめざし、3年後をメドに実用化する。

旅館や店舗などがシスデムやアプリから呼び出して使える翻訳サービスの形で提供する。外国語の音声を送ると画面に翻訳した日本語を返し、50カ国語に対応するMSの翻訳サービスを基盤として活用し、観光や医療、災害など、訪日外国人の翻訳ニーズが強い分野で正確な翻訳ができるようにする。豊橋技科大が特定分野に絞って収集した大量の対訳データを機械学習し、分野ごとの正確な翻訳用辞書をつくる。それをMSの翻訳サービスに組み込んで高品質な翻訳を実現する計画だ。対訳データを持つ企業や自治体との調整を始めており、富崎県が観光情報の対訳データを提供することが決まった。

3者は東京五輪が開催される20年には機械翻訳サービス市場が約1000億円に上ると見込む。

掲載日 2016年6月22日 日本経済新聞 朝刊 15ページ ©日本経済新聞社 印刷複製無断禁じます。

2016年6月22日、日本経済新聞

ニュース

豊橋技科大と日本MS、ブロードバンドタワーが多言語翻訳サービスを共同開発

八木 玲子=日経コンピュータ

2016/06/21 日経コンピュータ

目次へ戻る

シェア 共有 1 ブックマーク 2 Pocket ツイート 保存する

豊橋技術科学大学、日本マイクロソフト、ブロードバンドタワーは2016年6月21日、AI(人工知能)や機械学習を使った多言語翻訳技術を共同で開発すると発表した(写真1)。インバウンド(訪日外国人)が必要とする観光や医療、災害の情報をリアルタイムに翻訳するサービスを、2019年をめどに開発/提供する。2020年に開催される東京オリンピック・パラリンピックでの多言語コミュニケーションの実現を目指す。

翻訳エンジンや機械学習技術には、主にマイクロソフトの技術を活用する(写真2)。豊橋技術科学大学は、「対訳コーパス」などと呼ぶ日本語とその外国語訳がセットになったデータ群の構築を主に担当する。様々な自治体や企業と連携してデータを収集し、匿名化や分野ごとの重要語句の抽出、辞書化を行う。こうした大規模データを管理/活用するための基盤としては、マイクロソフトのクラウドサービス「Microsoft Azure」を利用する。

ブロードバンドタワーは、2015年に設立した子会社エーアイスクエアでAI関連技術を開発している。同社技術や、豊橋技術科学大学が構築する対訳コーパス、マイクロソフト

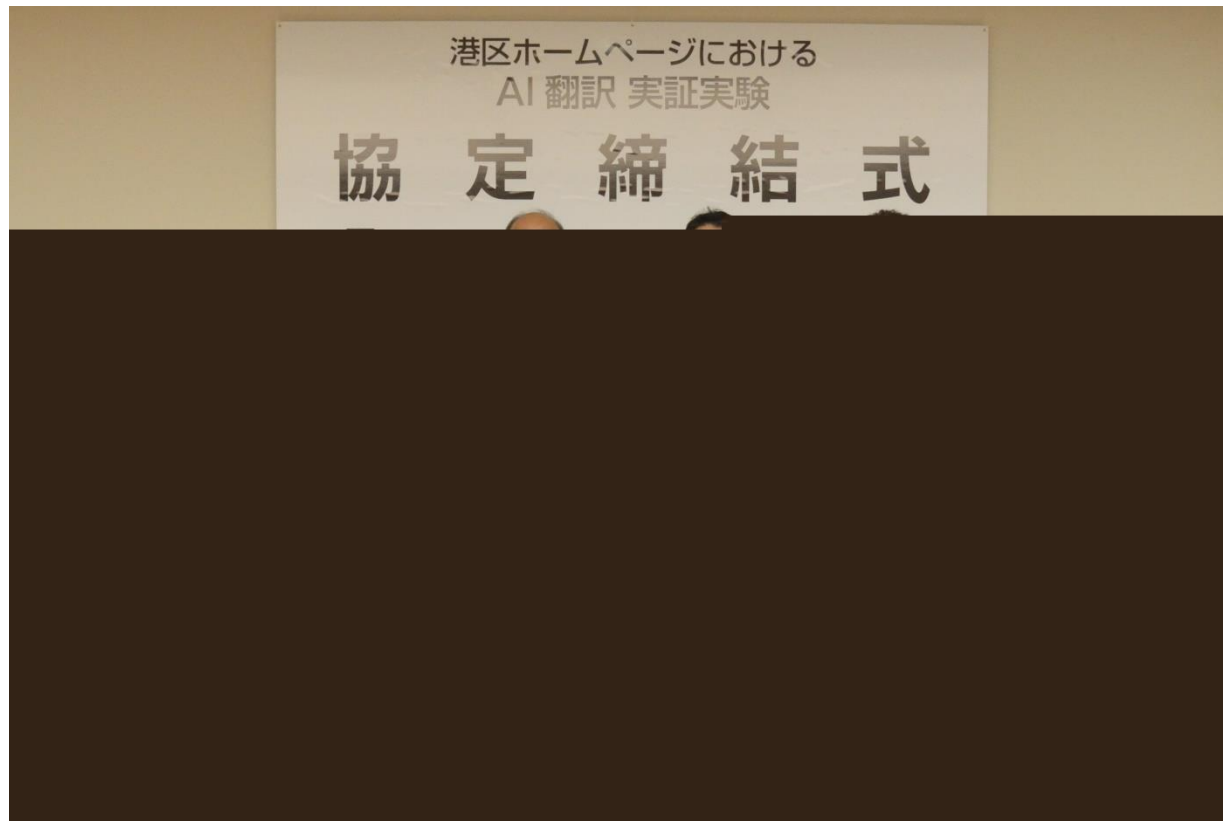
写真1 ●発表会の登壇者
左から、豊橋技術科学大学 情報メディア基盤センター センター長の川佐厚均教授、同大学の泉邦彦副学長、日本マイクロソフト 執行役 CTOの柳原彰氏、ブロードバンドタワーの藤原洋代表取締役会長兼社長、エーアイスクエアの石田正樹代表取締役、日本マイクロソフト 技術統括室 業務執行役員 NTOの田丸健三郎氏
[画像のクリックで拡大表示]

写真2 ●3者が協力して翻訳サービスを開発する
[画像のクリックで拡大表示]

2016年6月21日、日経コンピュータ・ウェブ版

豊橋技科大の機械翻訳(共同)プロジェクト

- 豊橋技科大、港区、日本マイクロソフト、グローバルデザイン社による機械翻訳実証プロジェクト(2018年7月に報道発表)



港区ホームページにおける「AI翻訳実証実験」

- 港区の課題

- 港区ホームページは、日本語の全ページを機械翻訳システムにより自動翻訳し、外国人住民等に対する正確で充実した区政情報の提供に努めている。
- 機械翻訳システムによる翻訳は、主語の省略や比喻表現が多い等、日本語の特徴的な文章構造により、特に英語の翻訳精度が低く、その精度向上が求められている。

- 4者協働

- 行政分野における精度の高いAI翻訳を実現するため、4者の持つ環境や技術力を総合し、港区ホームページをフィールドとして、行政向けAI翻訳システムの開発及び実証実験を行い、その効果を検証。

港区ホームページにおける「AI翻訳実証実験」

• 実証実験

1. これまで発行した冊子やパンフレット等の日英対訳文をAIに学習させた上で、行政分野で頻繁に使用する用語や港区特有の用語を登録することのできる辞書登録機能を付加し、行政分野に特化したAI翻訳システムを開発。
2. 完成したAI翻訳システムを、港区ホームページの一部に適用し、翻訳精度を検証。

• 分担

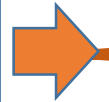
- 日本マイクロソフト株式会社
 - AI翻訳システムの提供及び技術支援、システムへの単語登録機能の開発
- 国立大学法人豊橋技術科学大学
 - 実証実験後の効果検証

機械翻訳と人間通訳を融合したスマホアプリ

- ブリックス、豊橋技科大、日本マイクロソフトが協働
- 簡単な内容は自動翻訳で対応し、難しい内容になると人間通訳に変わる。
- 人間と機械の長所を生かしたアプリ。
 - 機械：低コストでいつでもどこでも即時に利用可能。十分に流暢だが、固有名詞(用語)や完璧性を求められるに課題。
 - 人間：正確な翻訳。文脈(状況)に依存した翻訳が可能。
- 早期の事業化を予定

機械翻訳の精度向上

普通の文は
普通の人以上に
翻訳できる



翻訳機構の改良

大量データの利用

誤訳や訳の欠落
の検出・訂正

製品、企業、業種、
自治体、観光地

分野に特化した文書
の翻訳への対応



対訳データの充実

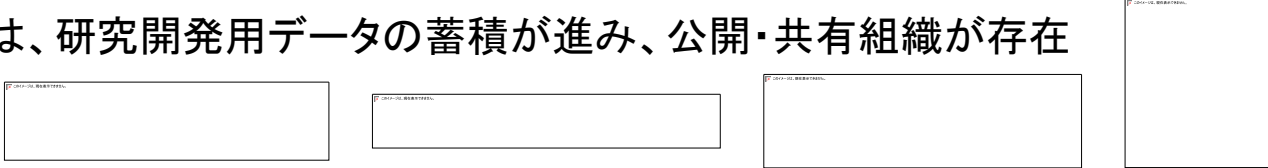
対訳データの利用
法の開発

分野固有の語句（低頻度）
高品質（公式文書）

□ 翻訳をサービスと
とらえる視点

研究開発用のデータの共有

現状と課題

- AIの学習、研究、実装全てにおいて求められるデータ
 - 海外では、研究開発用データの蓄積が進み、公開・共有組織が存在
- 
- **圧倒的に不利**な状況にある日本固有のデータ(日本語データ、国内画像データ)
 - データ利用・共有の有効性に対する**認識の欠如**。実証を困難にする**規制**。
 - 海外データによる偏ったAIモデル ➡ **我が国の競争力の激減(AI後進国化)**

計画と効果

- データの利活用のルールを策定し、データの流通を促進することにより

企業 ■ データはあるが、AIで紐解くことができない	と	研究機関・AIスタートアップ・研究者・学生 ■ AIの知見・頭脳はあるがデータがない
--------------------------------------	---	--

 を結び付ける。
- 法整備(個人情報、知的財産権の検討)、データ共有サービス基盤の実装
- 現在の実社会の課題を解決するだけでなく、全く新しい産業・サービスの創出

ニューラル機械翻訳の 研究動向

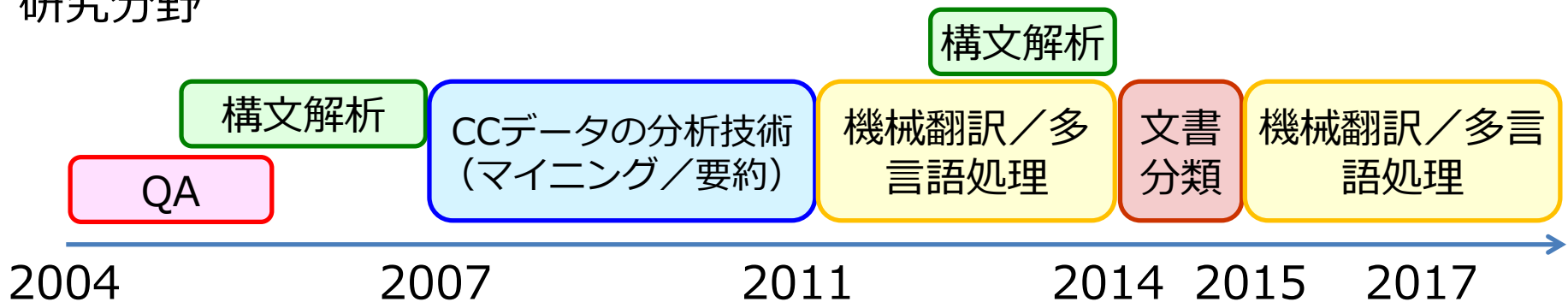
愛媛大学 大学院理工学研究科
田村 晃裕

第5回特許情報シンポジウム 2018年12月7日

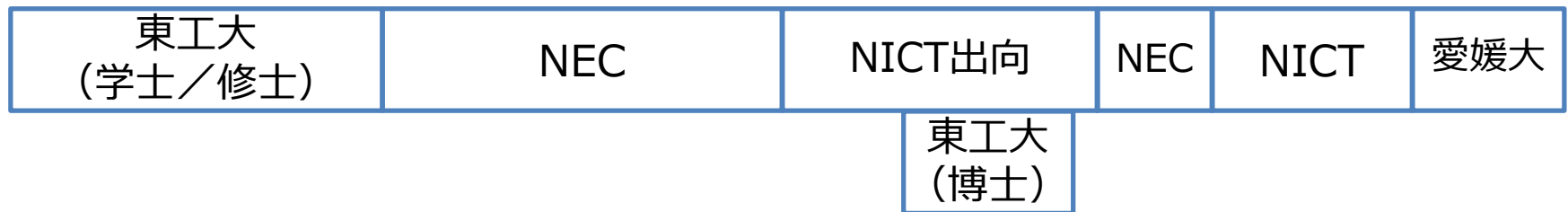
自己紹介 (1/2)

- ・ 田村 晃裕 (たむら あきひろ)
- ・ 愛媛大学大学院理工学研究科 電子情報工学専攻 助教

研究分野



所属



自己紹介 (2/2)

- 機械翻訳の研究に関して
 - 非対訳コーパスからの対訳対抽出 (2011年～現在)
 - 統計的機械翻訳 (SMT) の高度化 (2011年～2016年)
 - ニューラル機械翻訳 (NMT) の高度化 (2016年～現在)
 - 特に, 機械翻訳で文構造を活用する試み

【主な研究成果】

- Bilingual Lexicon Extraction from Comparable Corpora Using Label Propagation, EMNLP 2012.
- Distortion Model Considering Rich Context for Statistical Machine Translation, ACL 2013.
- Part-of-Speech Induction in Dependency Trees for Statistical Machine Translation, ACL 2013.
- Recurrent Neural Networks for Word Alignment Model, ACL 2014.
- Bilingual Segmented Topic Model, ACL 2016.
- Unsupervised Word Alignment by Agreement Under ITG Constraint, EMNLP 2016.
- Neural Machine Translation with Source Dependency Representation, EMNLP 2017.
- CKY-based Convolutional Attention for Neural Machine Translation, IJCNLP 2017.
- Forest-based Neural Machine Translation, ACL 2018.
- Neural Machine Translation Incorporating ⁴²Named Entity, COLING 2018.

目次

- NMTの概要
- 代表的なNMTモデル
- NMTの研究動向
- 現在の主な取り組み

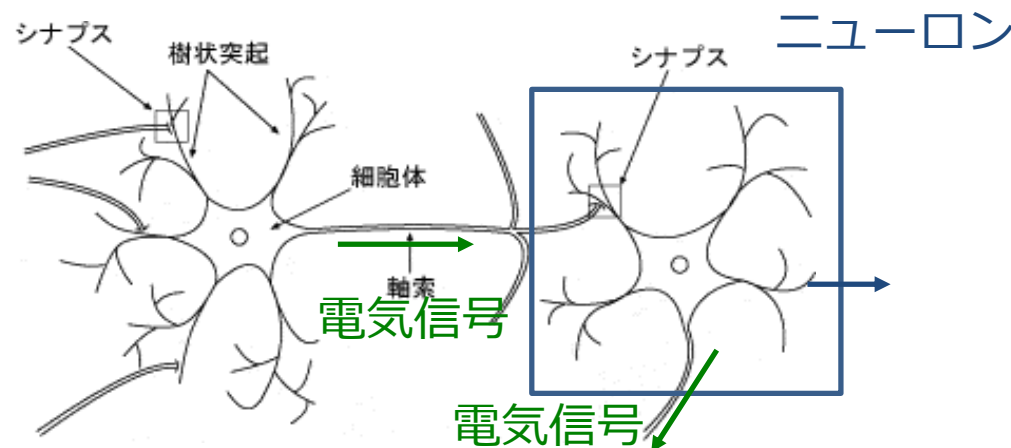
目次

- NMTの概要
- 代表的なNMTモデル
- NMTの研究動向
- 現在の主な取り組み

NMTとは？

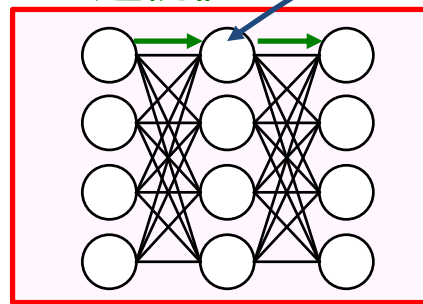
- ニューラルネットワーク（NN）を用いた機械翻訳
 - NN：人間の脳の仕組みを模倣した計算モデル

人間の脳



連続値

I live in Tokyo



NMT

45

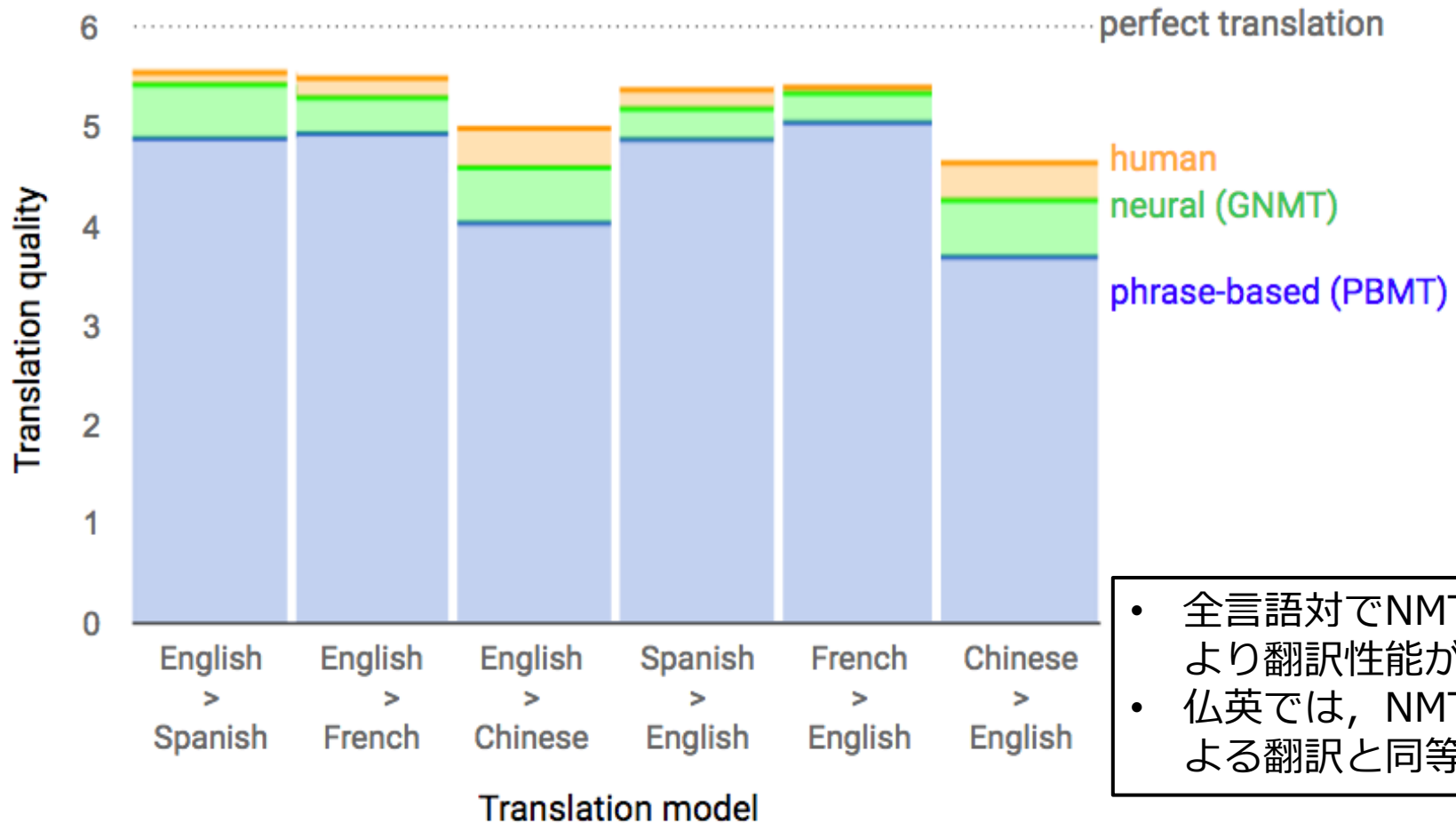


私は東京に住んでいる

重みパラメータを持つ
対訳データから重みを学習
学習した重みを使って翻訳

NMTの性能 (1)

2016年時点



- 全言語対でNMTはSMTより翻訳性能が高い
- 仏英では, NMTは人による翻訳と同等

NMTの性能 (2)

- WAT

BLEU(%)

	ASPEC (科学技術論文)				JPC (特許)			
	日英	英日	日中	中日	日英	英日	日中	中日
Moses (Phrase)	18.45	29.80	27.96	35.16	30.80	34.26	32.03	38.51
Moses (Hiero)	18.72	32.56	27.71	35.91	32.23	36.61	31.57	39.52
Moses (S2T)	20.36	33.44	28.65	36.52	34.40	37.65	32.35	39.90
AIAYN (NMT)	28.06	42.55	34.97	47.30	-	-	-	-
Online (2016)	16.69	19.81	10.72	12.87	35.77	37.89	23.57	27.91
Online (2016/11)	22.04	28.22	15.91	20.64	49.35	51.65	33.92	43.76

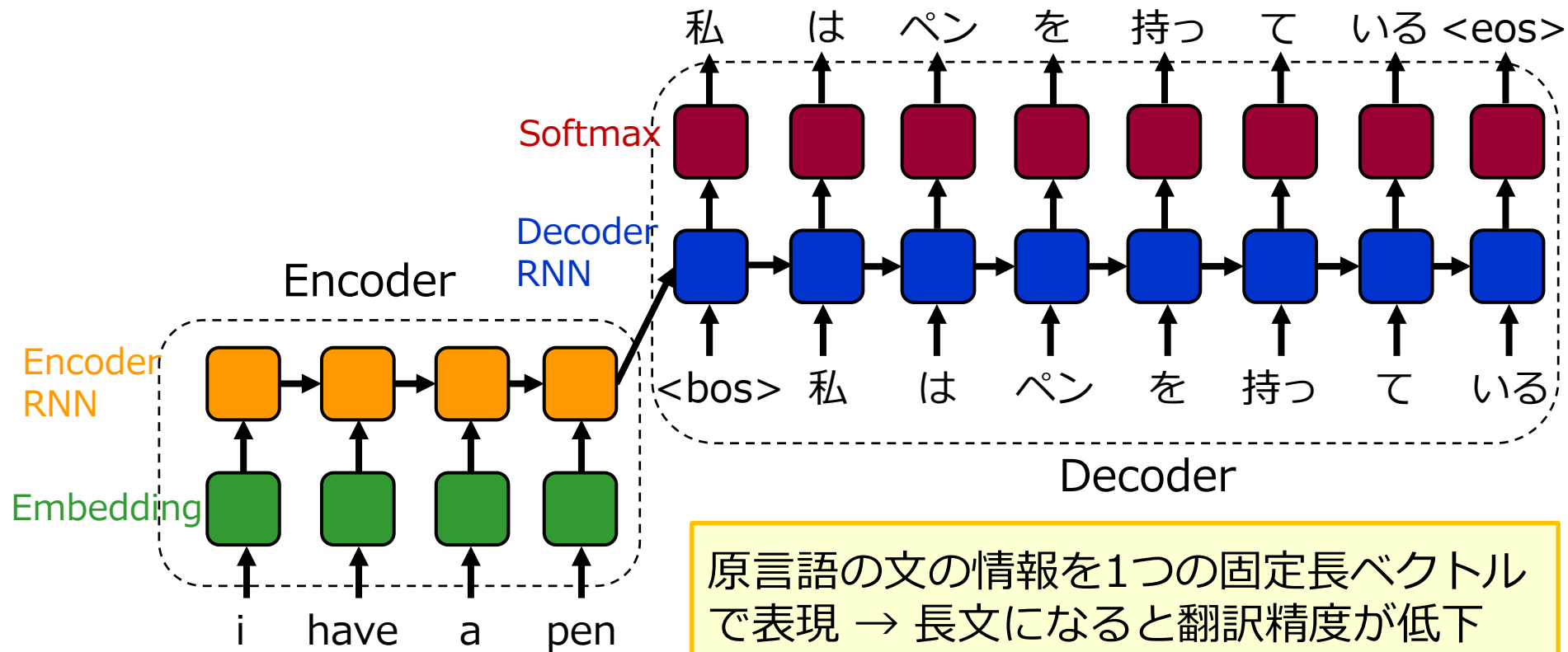
Nakazawa et al., "Overview of the 4th Workshop on Asian Translation", WAT 2017.

目次

- NMTの概要
- 代表的なNMTモデル
- NMTの研究動向
- 現在の主な取り組み

Sequence-to-Sequenceモデル [Sutskever et al., 2014]

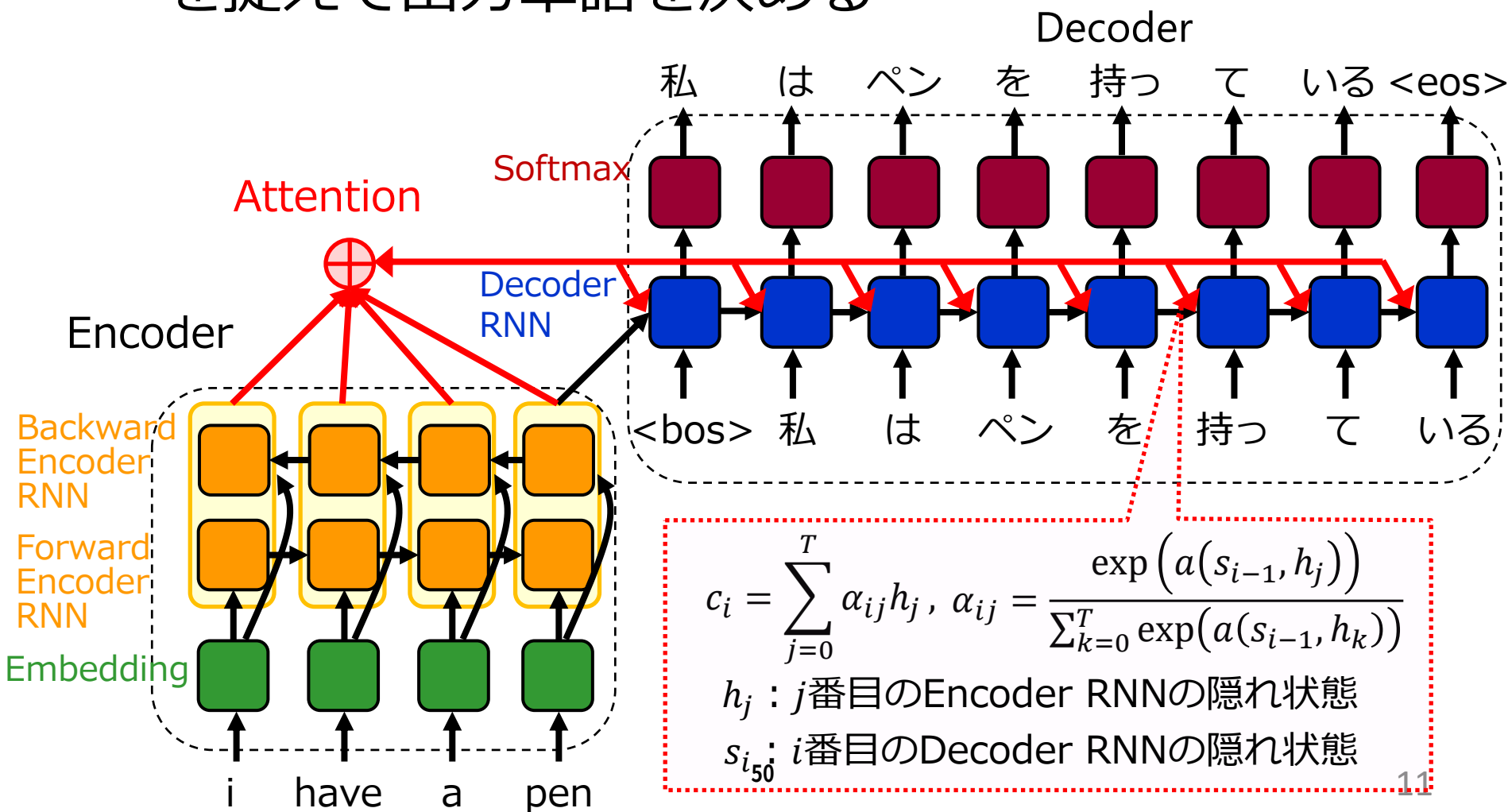
- Encoder用のRNNとDecoder用のRNNの2種類のRNNで翻訳
 - Encoderでは, 原言語の文から中間表現 (ベクトル) を生成
 - Decoderでは, 中間表現から出力文を生成



原言語の文の情報を1つの固定長ベクトルで表現 → 長文になると翻訳精度が低下

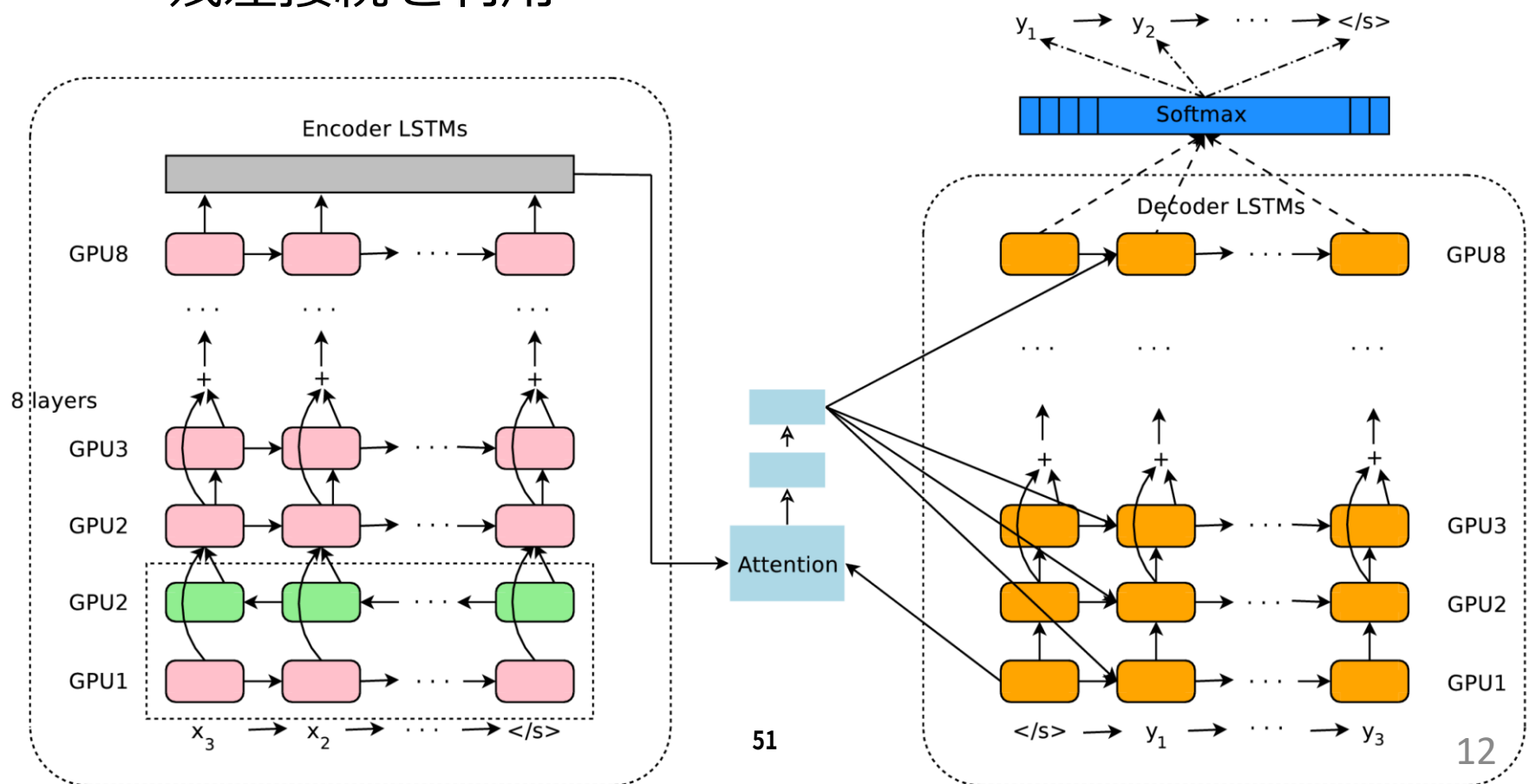
Attentionの導入 [Bahdanau et al., 2015]

- デコード時に原言語の文の中で着目すべき部分を捉えて出力単語を決める



Google's NMT (GNMT) [Wu et al., 2016]

- Attention-based NMTの拡張
 - Encoder, Decoderとして8層のLSTMを利用
 - 残差接続を利用



Convolutional Sequence-to-Sequenceモデル

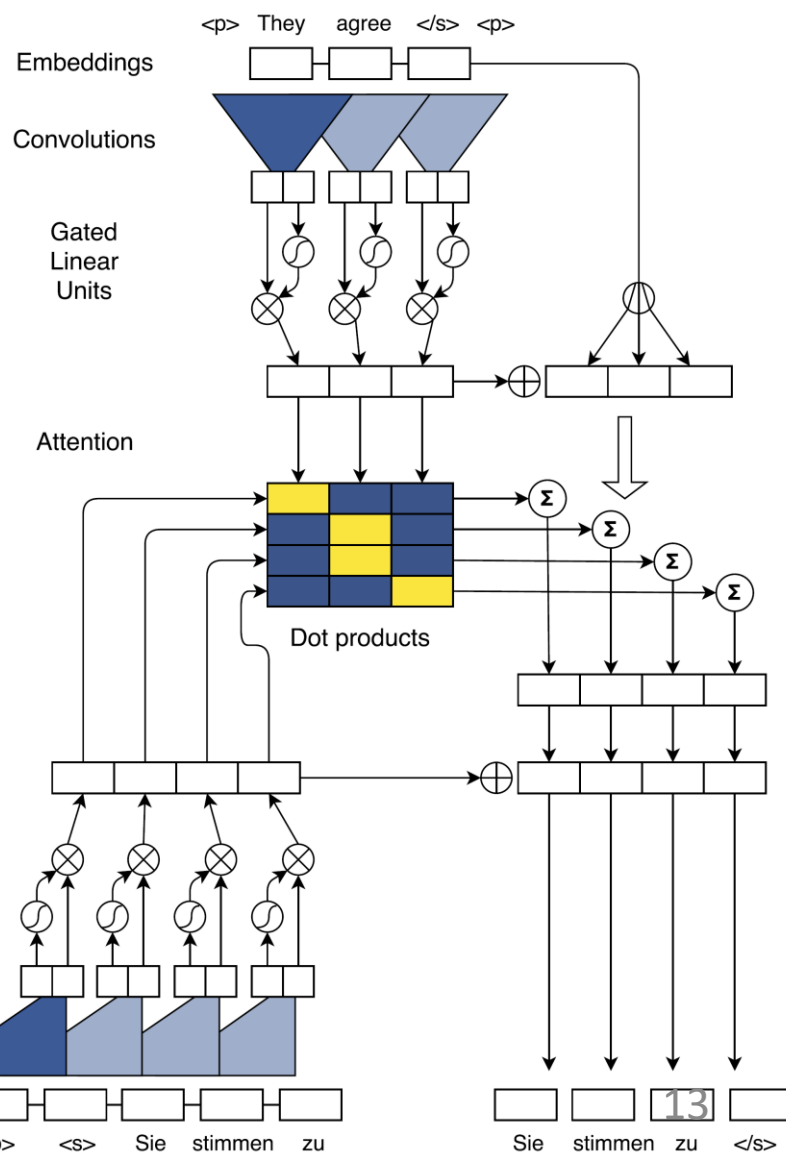
[Gehring et al., 2017]

- EncoderとDecoderにCNNを利用
 - 各CNNは並列演算可能
 - RNNベースのSeq2Seqモデルに比べて高速

WMT2014 英仏

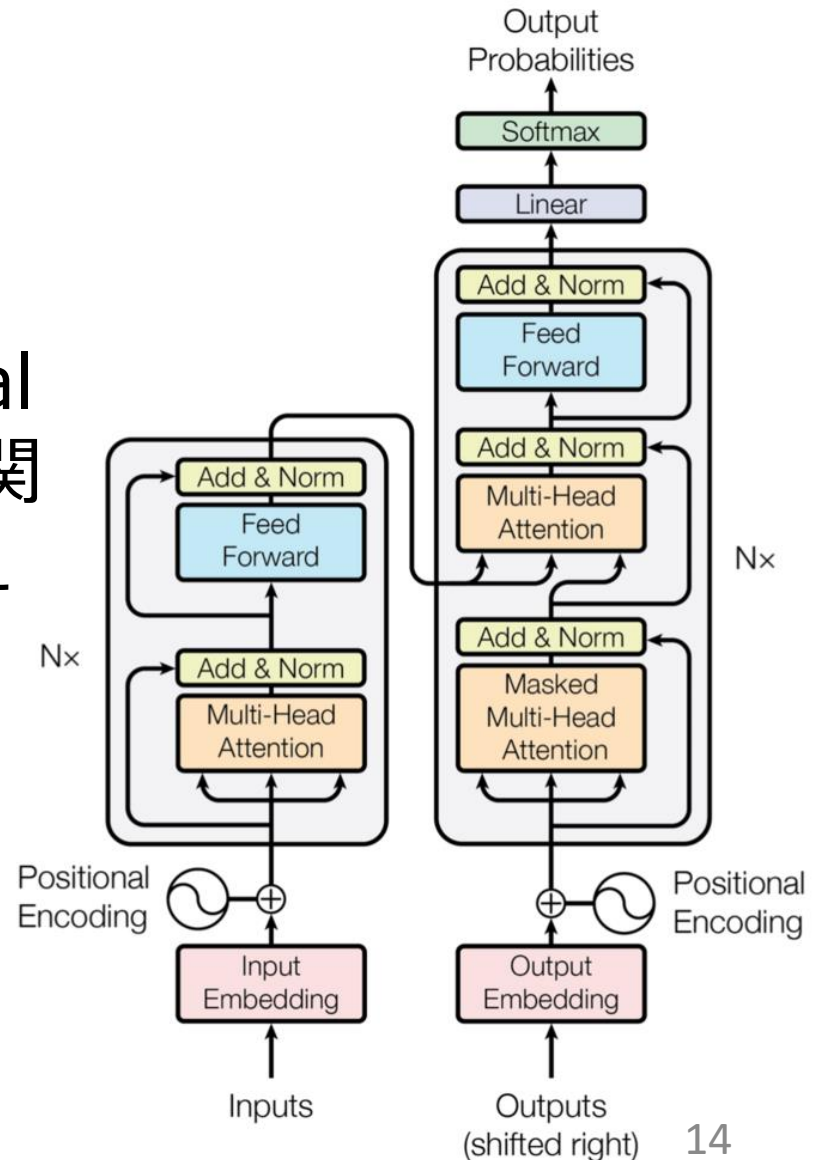
	BLEU	Training	Decoding for Dev
GNMT	39.92	96GPUで6日	3,028秒
Conv S2S	40.51	8GPUで37日	587秒

各CNNでは特定サイズの窓内の文脈のみを利用 → 長距離の関係性を捉えられない



Transformerモデル (1) [Vaswani et al., 2017]

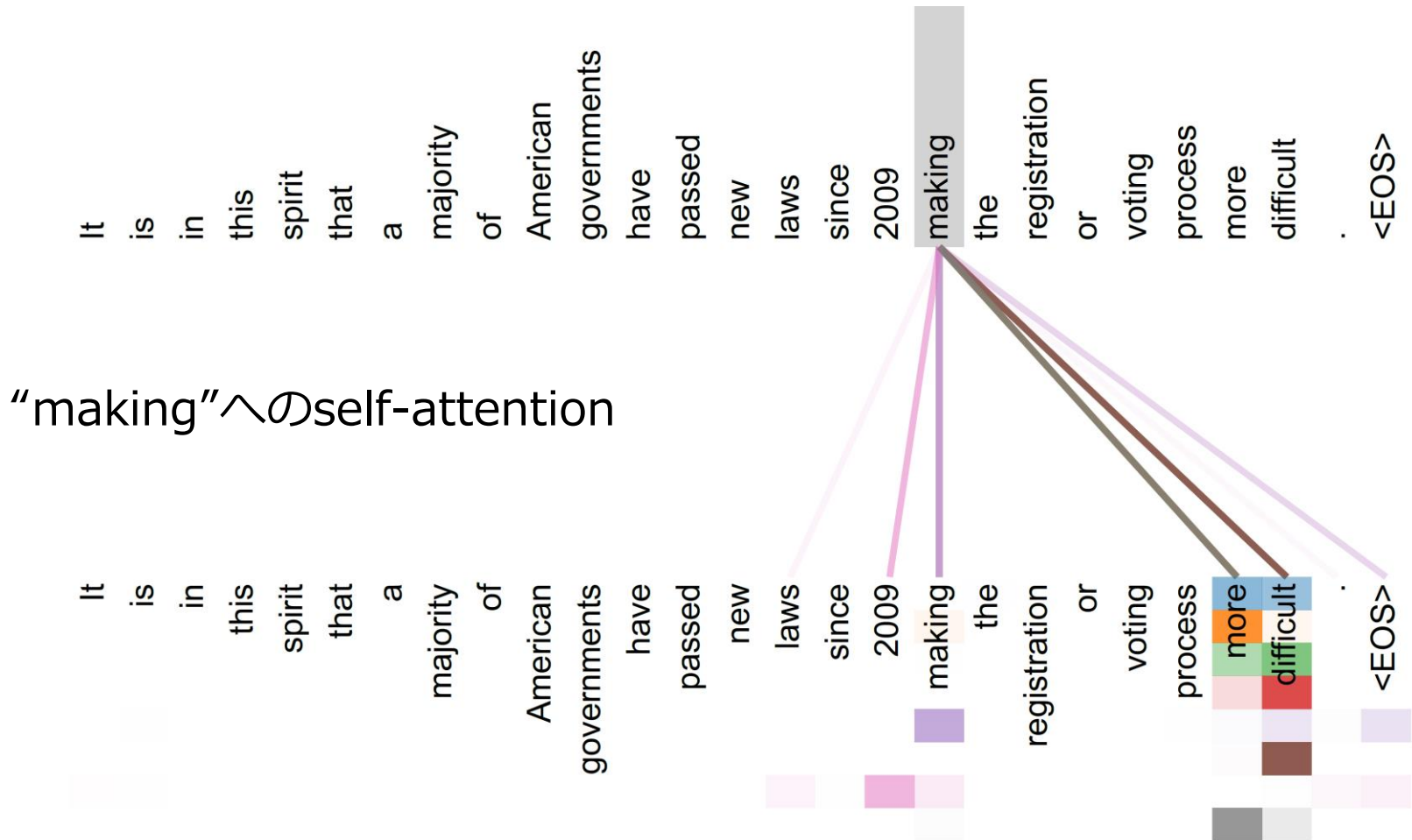
- Feed-Forwardネットワークのみ
→ 並列処理可能で高速
- Self-AttentionとPositional Encodingにより単語間の関連や単語の順序関係を捉える



Model	BLEU		Training Cost (FLOPs)	
	EN-DE	EN-FR	EN-DE	EN-FR
GNMT + RL	24.6	39.92	$2.3 \cdot 10^{19}$	$1.4 \cdot 10^{20}$
ConvS2S	25.16	40.46	$9.6 \cdot 10^{18}$	$1.5 \cdot 10^{20}$
GNMT + RL Ensemble	26.30	41.16	$1.8 \cdot 10^{20}$	$1.1 \cdot 10^{21}$
ConvS2S Ensemble	26.36	41.29	$7.7 \cdot 10^{19}$	$1.2 \cdot 10^{21}$
Transformer (base model)	27.3	38.1	$3.3 \cdot 10^{18}$	
Transformer (big)	28.4	41.0	$2.3 \cdot 10^{19}$	

Transformerモデル (2) [Vaswani et al., 2017]

- 長距離の関連を捉えられる

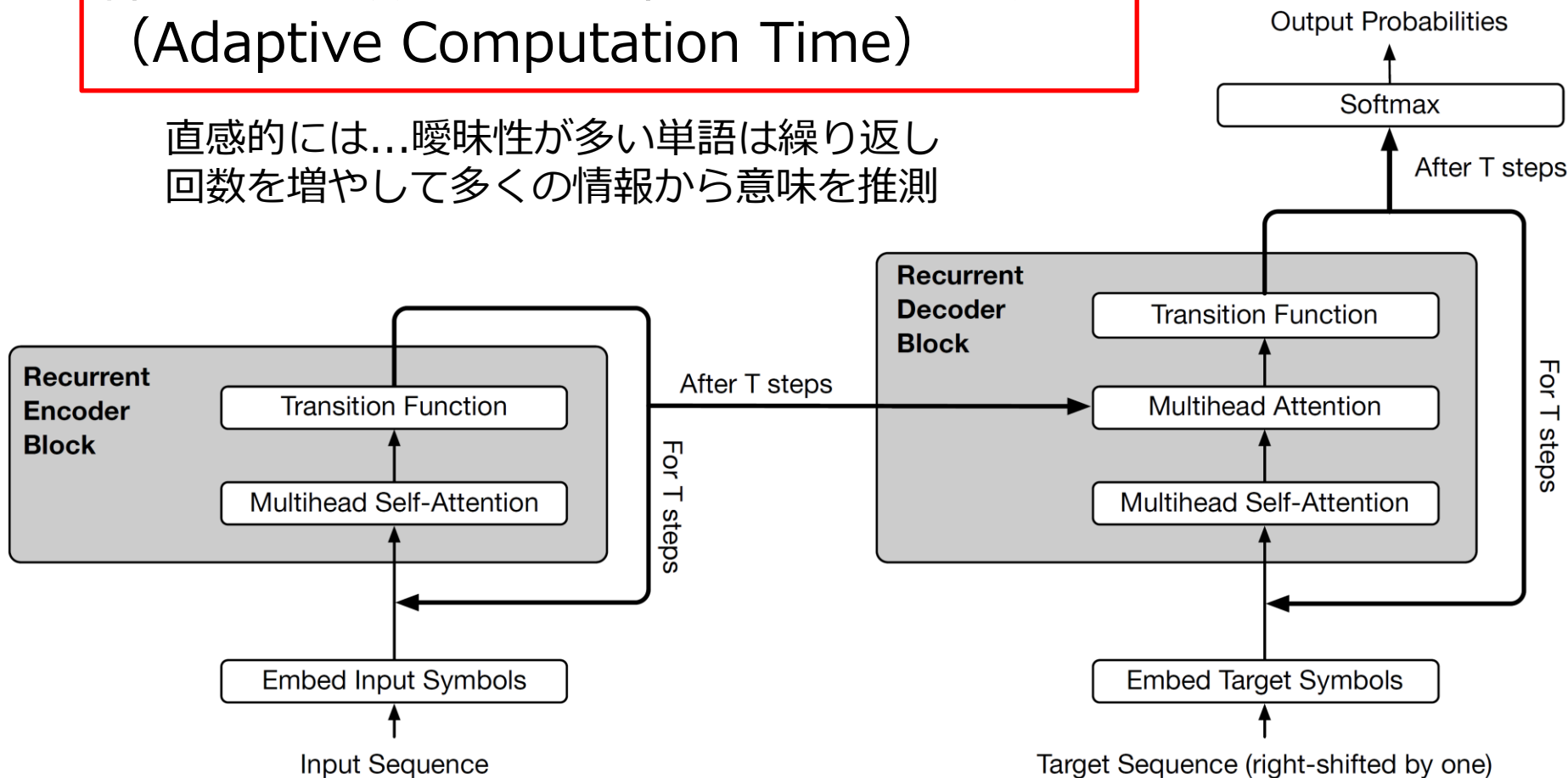


“making … more difficult”⁵⁴という構文を捉えられている

Universal Transformers [Dehghani et al., 2017]

繰り返し回数 T は入力単語毎に動的に決定
(Adaptive Computation Time)

直感的には...曖昧性が多い単語は繰り返し
回数を増やして多くの情報から意味を推測



WMT2014 英独 (BLEU)

Transformer : 28.0 → Universal Transformer : 28.9

目次

- NMTの概要
- 代表的なNMTモデル
- **NMTの研究動向**
- 現在の主な取り組み

NMTの特長と課題

- 特長
 - End-to-Endなモデル
 - 高い翻訳精度，流暢な翻訳文
- 課題
 - 訳抜けや重複訳が起こりやすい
 - 実用上，語彙数を制限する必要あり
 - モデルの解釈性が低い

研究の方向性 (1)

- 課題への対処
 - 訳抜け, 重複訳への対応
 - 未知語への対応
 - モデルの解釈
- 高精度化
 - 文構造の活用
 - 言語学的情報の活用
 - モノリンガルデータの活用
 - など

研究の方向性 (2)

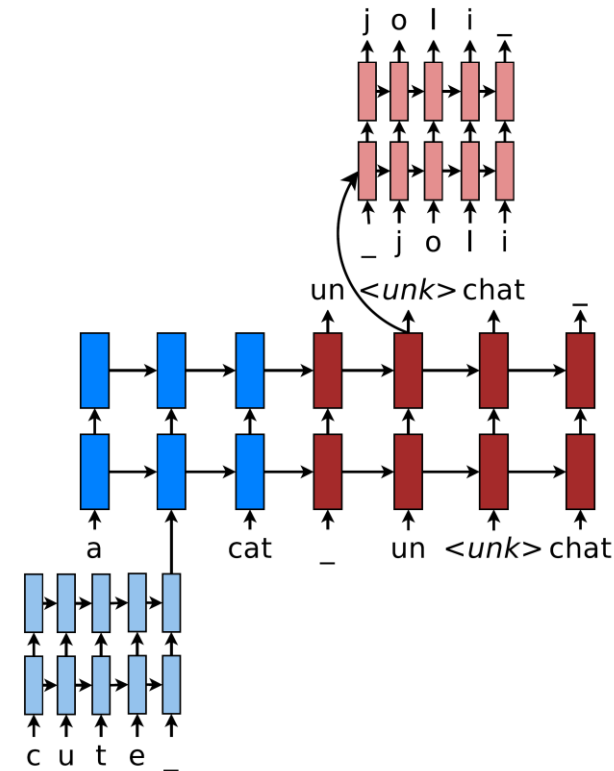
- モデルの軽量化
- 低資源／ゼロ資源における翻訳
- 文書レベルの翻訳
- マルチモーダル翻訳

訳抜け，重複訳への対応

- Coverage Vector [Tu et al., 2016], Coverage Embedding [Mi et al., 2016]
 - 原言語のどの単語が翻訳されたかを管理してattentionの計算時に利用
- Reconstruction [Tu et al., 2017]
 - 翻訳結果を原言語に翻訳しなおして元の文に戻るよう
に学習
- Phrase-based Forced Decoding [Zhang et al., 2017]
 - SMTでは訳抜け，重複訳が起こりにくいことを利用
 - NMTの翻訳候補をSMTのフレーズテーブルでリランキン
グ

未知語への対応

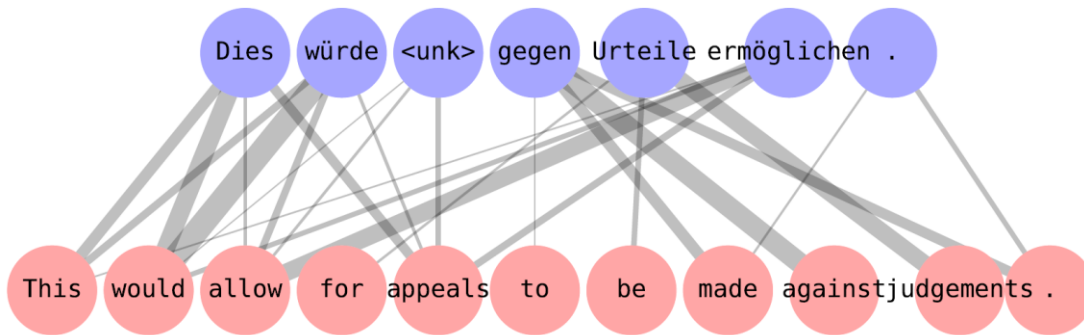
- Copying Mechanism [Luong et al., 2015; Jean et al., 2015]
 - 単語アライメント結果やattentionに基づいて未知語に対応する原言語の単語を特定し, 外部資源 (対訳辞書等) で翻訳したものに置き換え
- 文字単位にする [Luong and Manning 2016; Lee et al., 2017; Cherry et al., 2018]
- Sub-word (Byte Pair Encoding) [Sennrich et al., 2016; Wu et al., 2016; Kudo, 2018]
 - 全文字, 頻度の高い文字列を語彙集合として使用



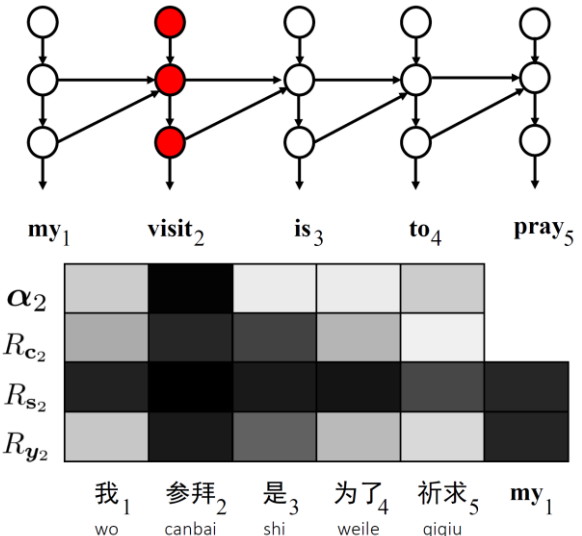
[Luong and Manning, 2016]

モデルの解釈

- 翻訳の解釈を可視化 [Alvarez-Melis and Jaakkola, 2017; Ding et al., 2017]



[Alvarez-Melis and Jaakkola, 2017]

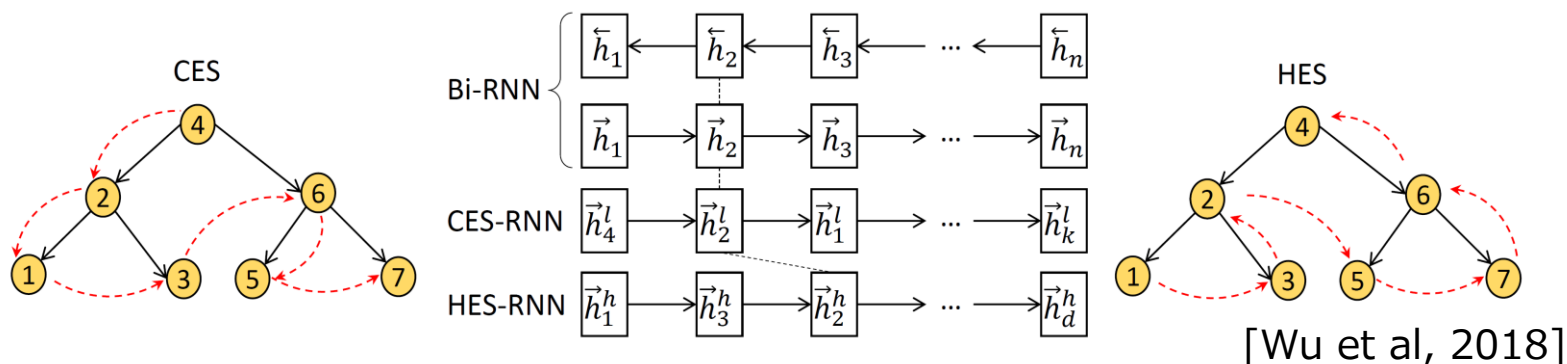


[Ding et al, 2017]

- ニューロンの役割を解析 [Bau et al., 2018]
 - 複数のNMTモデル間で発火するニューロンの相関を測ることで, 出力の各性質 (時制, 数, 性) に関するニューロンを特定

文構造の活用 (1)

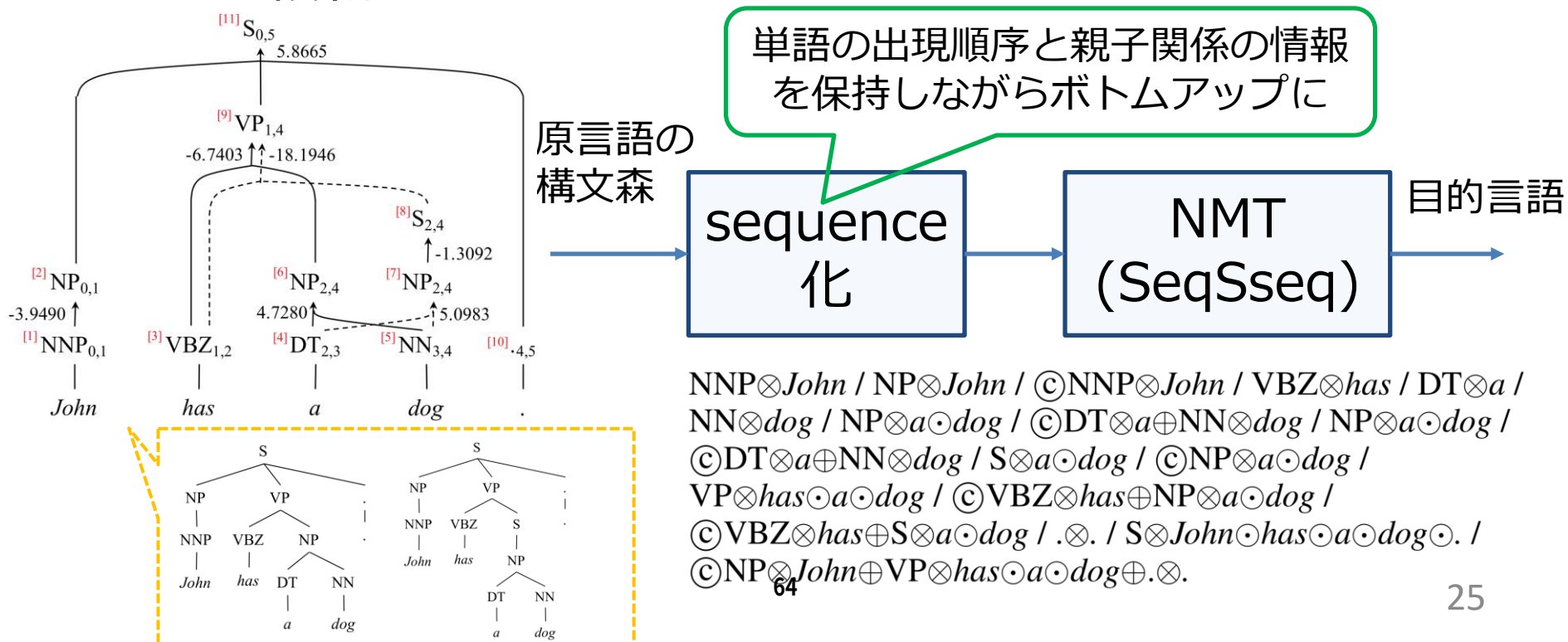
- 文構造をsequence化し, Seq2Seqモデルで扱う
 - 深さ優先走査 [Aharoni and Goldberg, 2017; Li et al., 2017]
 - transition action系列 [Wu et al., 2017]
 - top down & bottom up [Wu et al., 2018]



- 木構造用のNN (Tree-structured NN) を Encoder/Decoderで使用 [Eriguchi et al., 2016; Chen et al., 2017; Wang et al., 2018; Zareemoodi and Haffari, 2018]

文構造の活用 (2)

- Forest-based NMT [Ma et al., 2018]
 - 従来のSyntax-based NMTは構文解析の第一位候補を利用 → 構文解析誤りの影響を受けやすい
 - 構文森 (forest) を使うことで構文解析誤りの悪影響を緩和



言語学的情報の活用

- 言語学的素性（見出し語，品詞，格など）の結合ベクトルを Encoder に入力 [Sennrich and Haddow, 2016]

原言語の単語の埋め込みベクトル

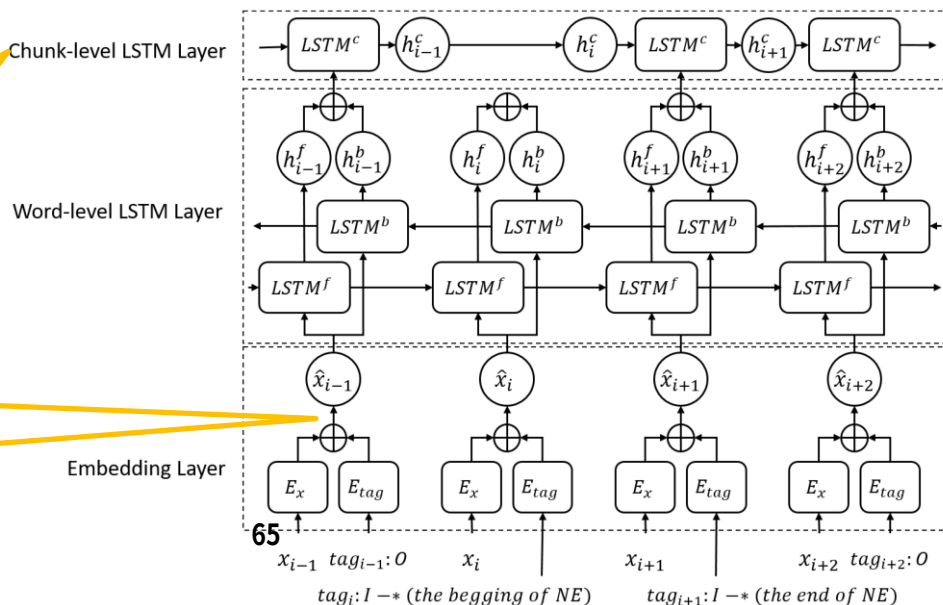
$$h_j = f_{enc}(h_{j-1}, (E_{word}(x_j) [; \sum_{k=1}^{|F|} E_k(x_{jk})]))$$

言語学的素性の埋め込みベクトル

- 原言語の固有名詞の情報を Encoder で活用 [Ugawa et al., 2018]
 - タグ情報：翻訳時の曖昧性解消の手がかり（Apple：アップル社/りんご）
 - チャンク情報：複合名詞の翻訳の手がかり

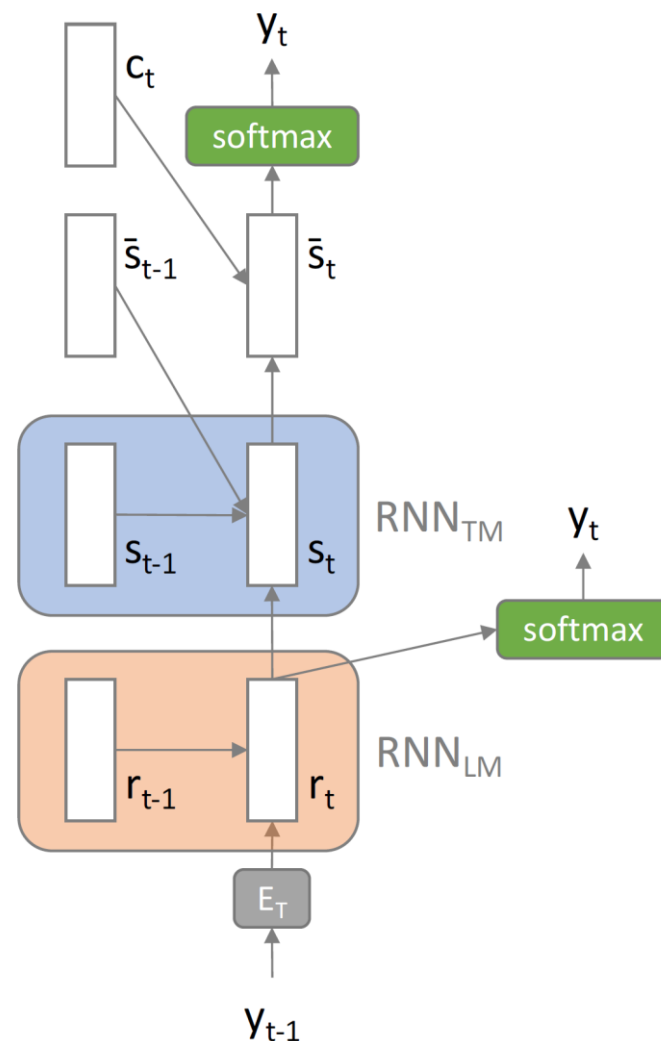
「Chunk層 - Word層」の階層構造

単語ベクトルと固有名詞ベクトルを線形結合



モノリンガルデータの活用 (1)

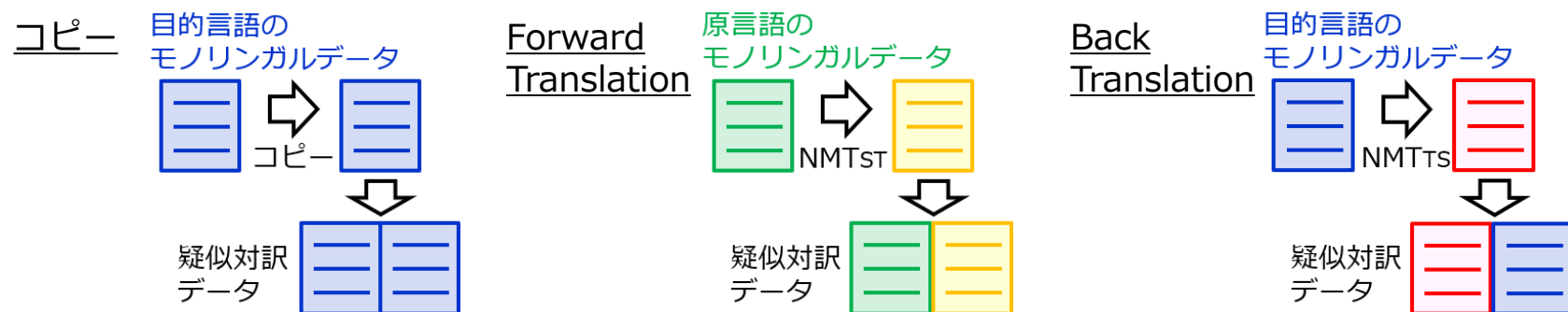
- モノリンガルデータから学習する言語モデルを活用
 - NMTの出力を言語モデルのスコアでリランキング [Luong et al., 2015]
 - NMTモデルに言語モデルの層を追加
 - モノリンガルデータを言語モデル層の初期化に使用 [Skorokhodov et al., 2018]
 - マルチタスク学習 [Domhan and Hieber, 2017]



[Domhan and Hieber, 2017]

モノリンガルデータの活用 (2)

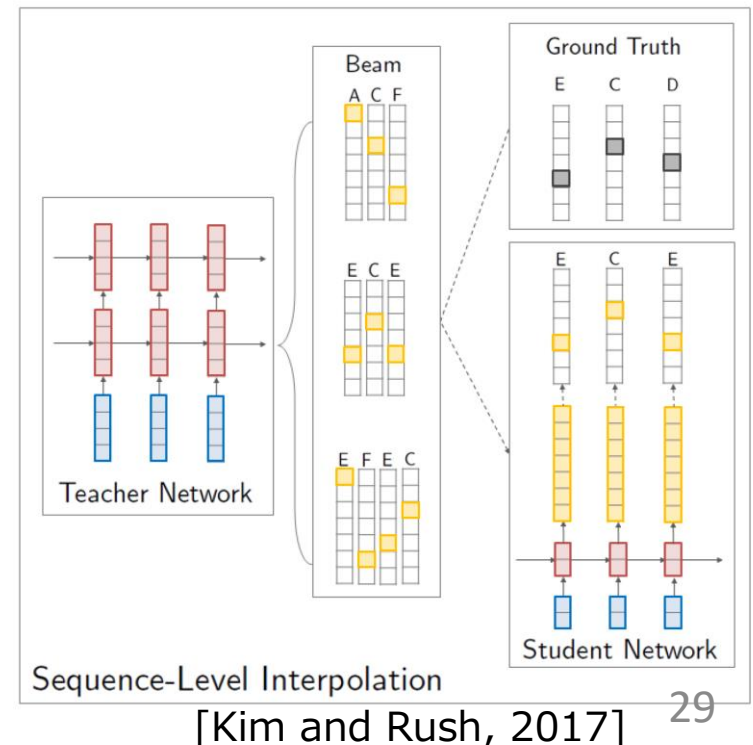
- データ拡張 (Data Augmentation) : 疑似の対訳文を生成してNMTの学習に使用
 - コピー [Currey et al., 2017]
 - Forward Translation [Chinea-Rios et al., 2017; Park et al., 2017]
 - Back Translation** [Sennrich et al., 2016; Hoang et al., 2018; Fadaee and Monz, 2018; Edunov et al., 2018]



- [Burlot and Yvon, 2018] : NMTにおけるモノリンガルデータの様々な使い方を比較 → Back Translationが良い！
- [Edunov et al., 2018] : WMT2014 (英独) で35.0(BLUE) 最高性能！
※ ちなみに, Transformer 28.4⁷, DeepL 33.3

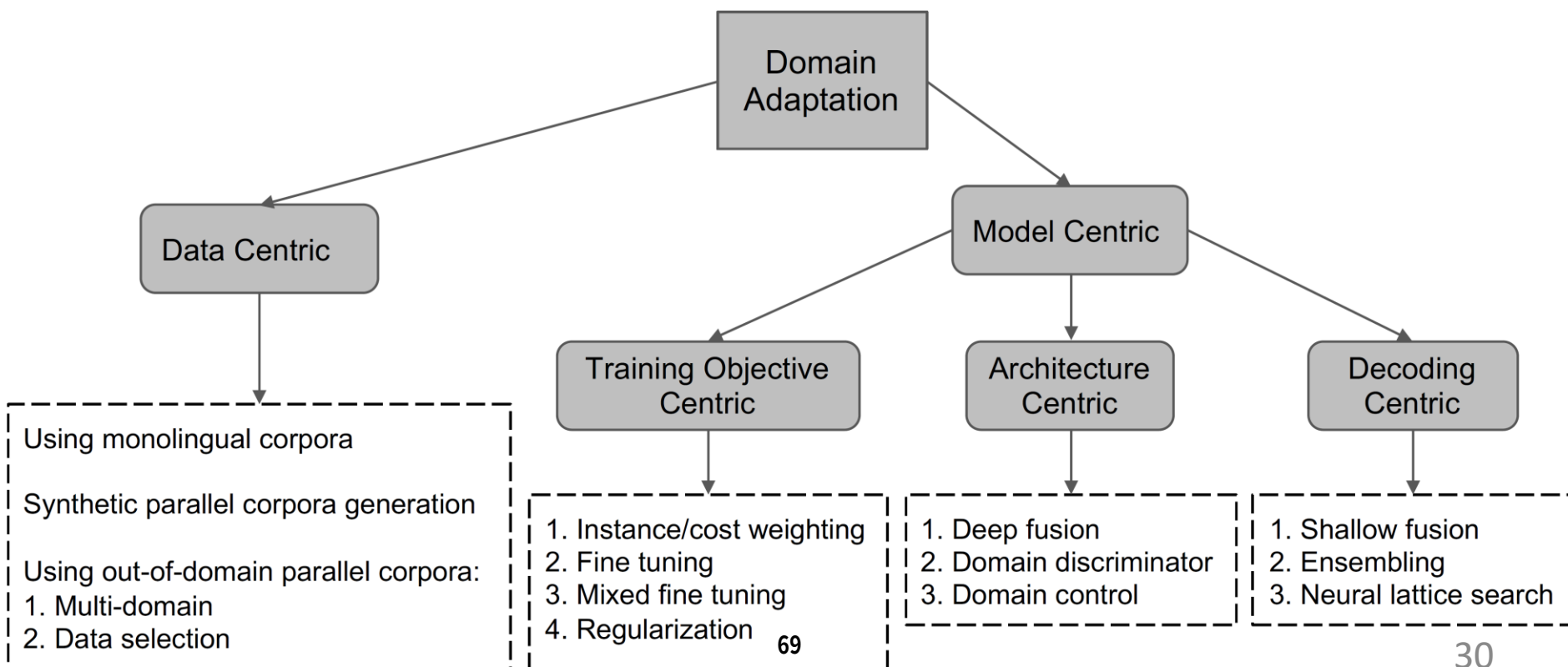
モデルの軽量化

- プルーニング [Han et al., 2015; See et al., 2016; Zhang et al., 2017]
 - Han et al. (2015)
 - 重みが閾値以下のコネクションを間引く
 - 入力か出力が0のニューロンを間引く
 - 間引いた後のモデルを再学習
- 知識の蒸留 [Kim and Rush, 2016; Freitag et al., 2017; Zhang et al., 2018]
 - 大きいモデルを教師モデルとして準備して, 小さいモデルを教師モデルの予測と似るように学習



低資源／ゼロ資源における翻訳 (1)

- 分野適応：翻訳対象分野以外の対訳データ, 翻訳対象分野のモノリンガルデータを使って翻訳性能改善
 - Chu and Wang, “A Survey of Domain Adaptation for Neural Machine Translation”, COLING 2018.



低資源／ゼロ資源における翻訳 (2)

• ゼロ資源翻訳

<2es> How are you? → Cómo estás?

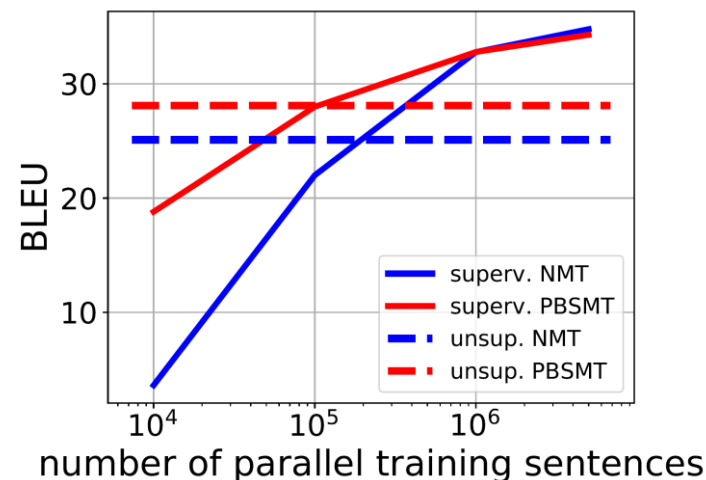
<2en> Ich bin ein Student. → I am a student.

– 翻訳対象言語対以外の対訳データを利用

- ゼロショット学習 [Johnson et al., 2017] : 原言語側に翻訳先言語のタグを付与し, 多言語対訳データを同時学習させることで学習データにない言語対の翻訳も可能に

– 原言語, 目的言語のモノリンガルデータを利用

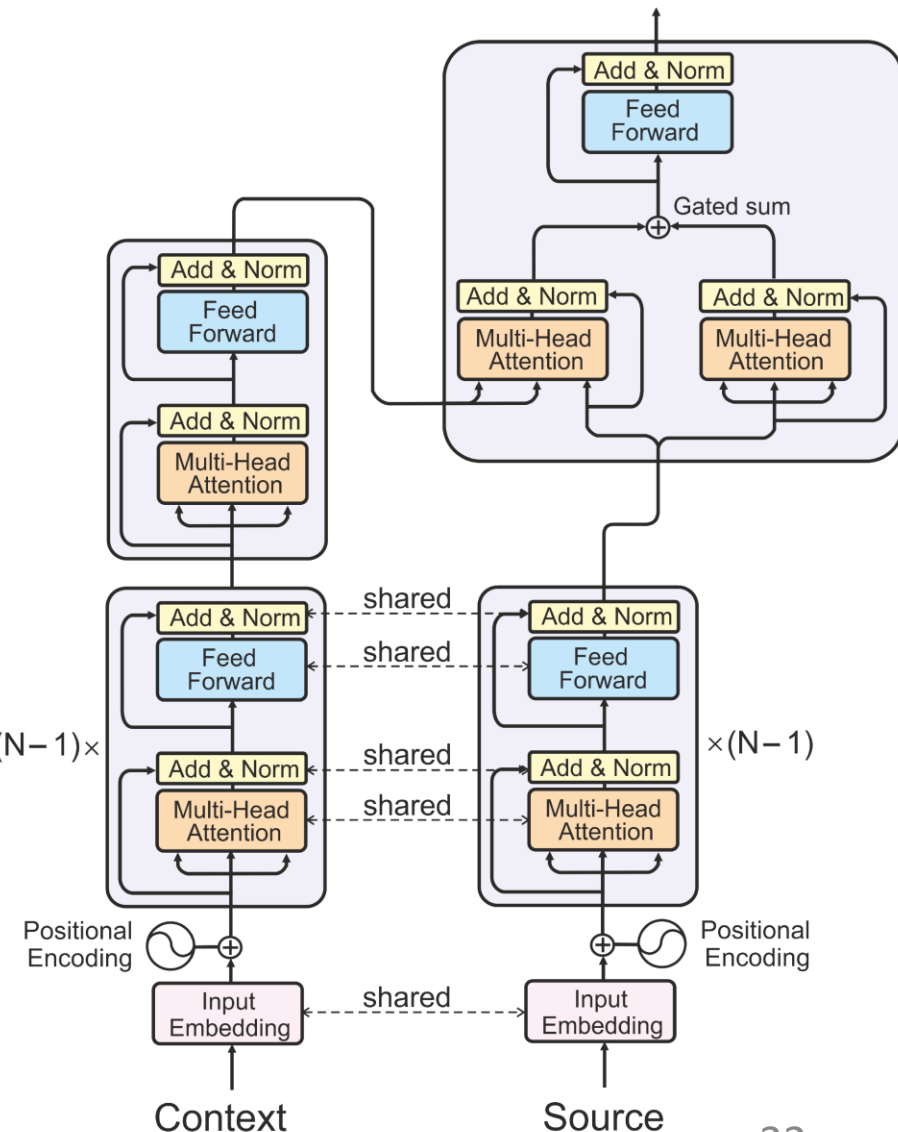
- 教師なしNMT + 教師なしSMT [Lample et al., 2018] : 初期の貧弱な翻訳モデル (モノリンガルデータの結合データから獲得した多言語エンベディング, 翻訳単語対) を言語モデルとBack Translationを繰り返し適用して高精度化



- 従来の教師なしMTの性能を10ポイント (BLEU) 以上上回る
- WMT2014 (英仏) で学習データ10万文の教師ありSMT/NMTと同等の性能

文書レベルの翻訳

- 文脈を考慮することで翻訳性能を改善
 - 一貫性の高い翻訳, 照応解析, 曖昧性解消の手がかりなど
- 手法
 - 翻訳対象の文に前方の文を結合させてNMTの入力とする [Tiedemann and Scherrer, 2017; Agrawal et al. 2018]
 - 文脈を捉えるためのEncoderやAttentionを追加してデコード時にそれらを考慮 [Jean et al., 2017; Wang et al., 2017; Tu et al., 2017; Bawden et al., 2018; Voita et al., 2018; Maruf and Haffari 2018; Miculicich et al., 2018]
 - DiscoMTで色々なモデルが発表されている

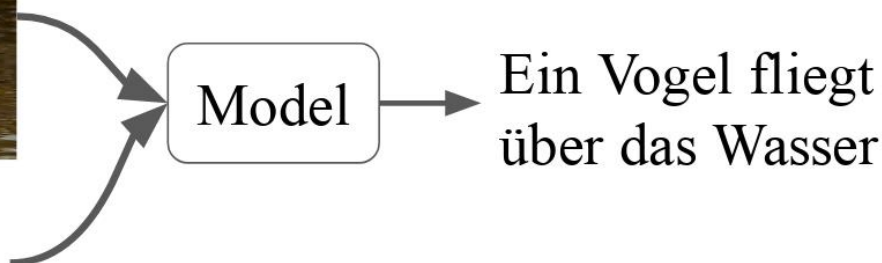


[Voita et al, 2018]

マルチモーダル翻訳 (1)



A bird flies
over the water

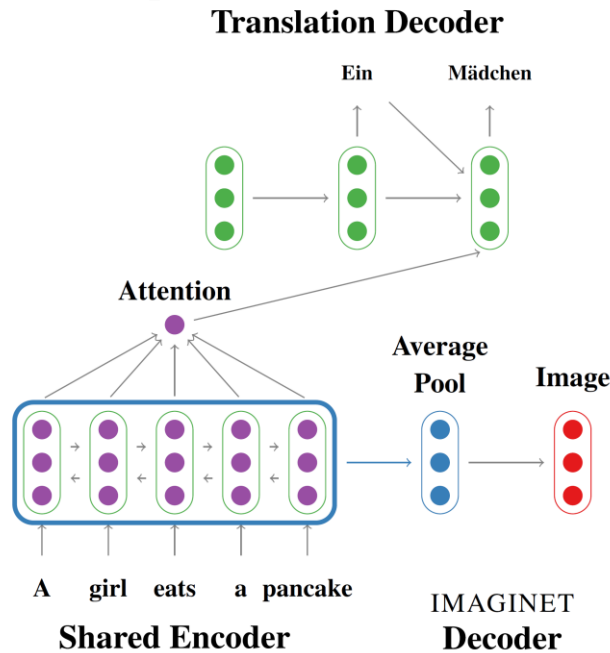


Ein Vogel fliegt
über das Wasser

- 翻訳対象の文に関する画像を考慮することで翻訳性能を改善
 - 語義曖昧性解消の手がかり, 画像を介した表記の揺れ解消, 前置詞を決める際の手がかりなど

マルチモーダル翻訳 (2)

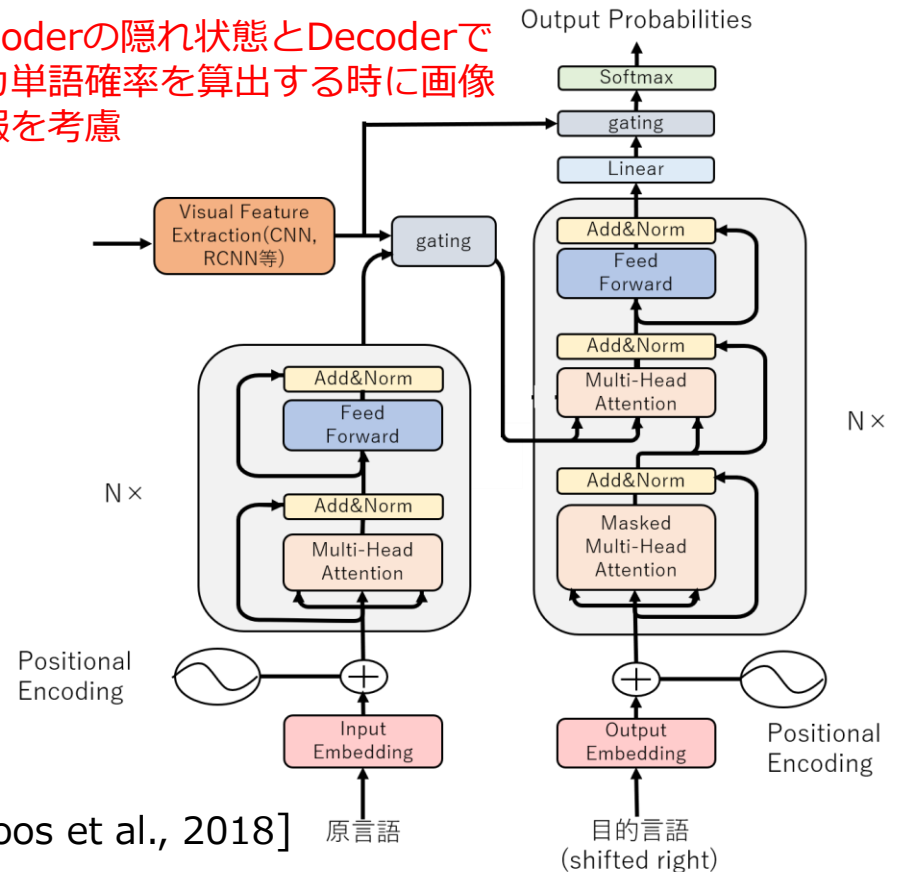
- マルチタスク学習 [Luong et al., 2016, Elliott and Kadar, 2017]



[Elliott and Kadar, 2017]

- 同時学習 [Huang et al., 2016; Calixto et al., 2017; Zhou et al., 2018; Grönroos et al., 2018]

Encoderの隠れ状態とDecoderで出力単語確率を算出する時に画像情報を考慮



[Grönroos et al., 2018]

- WMTのMultimodal Machine Translation Shared Taskで色々なモデルが発表されている 73

目次

- NMTの概要
- 代表的なNMTモデル
- NMTの研究動向
- 現在の主な取り組み

構文情報の活用

- Transformerで原言語の文の構造を考慮

	英日 (ASPEC)	英中 (NIST)
Transformer	34.09	34.84
Tree-based Transformer	35.81	36.40
Forest-based Transformer	36.41	36.93

マルチモーダル翻訳

- Transformerベースのマルチモーダル機械翻訳の試作版を実装, 評価
- 画像を使うことで翻訳精度が改善できることを確認
 - 英独 (Multi30k) : 34.30 → 35.26 BLEU



Obst (果物)のAttention



翻訳元: two young boys putting fruit on the bike .

参照訳(正解)	zwei jungen packen obst auf das fahrrad . (訳) 2人の男の子が自転車に 果物 を 詰め込んで います。
ベースライン	zwei kleine jungen legen auf das bike . (訳) 2人の少年が自転車に 乗って います。
試作モデル	zwei jungen stellen obst auf das fahrrad . (訳) 2人の男の子が自転車に 果物 を 入れました 。

まとめ

- 次々に新しいNMTモデルや技術が開発されている
- 現在の注目技術
 - Transformerモデル
 - Byte Pair Encoding
 - Back Translation
- 今後注目したい動向
 - 教師無しNMT
 - 文書レベルの翻訳
 - マルチモーダル翻訳₇₇

参考文献 (1)

- Ruchit Agrawal, Marco Turchi, and Matteo Negri, "Contextual Handling in Neural Machine Translation: Look Behind, Ahead and on Both Sides", EAMT 2018.
- Roei Aharoni and Yoav Goldberg, "Towards String-to-Tree Neural Machine Translation", ACL 2017.
- David Alvarez-Melis and Tommi S. Jaakkola, "A causal framework for explaining the predictions of black-box sequence-to-sequence models", EMNLP 2017.
- Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio, "Neural Machine Translation by Jointly Learning to Align and Translate", ICLR 2015.
- Anthony Bau, Yonatan Belinkov, Hassan Sajjad, Nadir Durrani, Fahim Dalvi, and James Glass, "Identifying and Controlling Important Neurons in Neural Machine Translation", arXiv 2018.
- Rachel Bawden, Rico Sennrich, Alexandra Birch, and Barry Haddow. "Evaluating Discourse Phenomena in Neural Machine Translation", NAACL 2018.
- Franck Burlot and Francois Yvon, "Using Monolingual Data in Neural Machine Translation: a Systematic Study", WMT 2018.
- Iacer Calixto, Qun Liu, and Nick Campbell, "Doubly-Attentive Decoder for Multi-modal Neural Machine Translation", ACL 2017.
- Huadong Chen, Shujian Huang, David Chiang, and Jiajun Chen, "Improved Neural Machine Translation with a Syntax-Aware Encoder and Decoder", ACL 2017.
- Colin Cherry, George Foster, Ankur Bapna, Orhan Firat, and Wolfgang Macherey, "Revisiting Character-Based Neural Machine Translation with Capacity and Compression", EMNLP 2018.

参考文献 (2)

- Mara Chinae-Rios, Alvaro Peris, and Francisco Casacuberta, "Adapting Neural Machine Translation with Parallel Synthetic Data", WMT 2017.
- Chenhui Chu and Rui Wang, "A Survey of Domain Adaptation for Neural Machine Translation", COLING 2018.
- Anna Currey, Antonio Valerio Miceli Barone, and Kenneth Heafield, "Copied Monolingual Data Improves Low-Resource Neural Machine Translation", WMT 2017.
- Mostafa Dehghani, Stephan Gouws, Oriol Vinyals, Jakob Uszkoreit, and Lukasz Kaiser, "Universal Transformers", arXiv 2018.
- Yanzhuo Ding, Yang Liu, Huanbo Luan, and Maosong Sun, "Visualizing and Understanding Neural Machine Translation", ACL 2017.
- Tobias Domhan and Felix Hieber, "Using Target-side Monolingual Data for Neural Machine Translation through Multi-task Learning", EMNLP 2017.
- Sergey Edunov, Myle Ott, Michael Auli, and David Grangier, "Understanding Back-Translation at Scale", EMNLP 2018.
- Desmond Elliott and Akos Kádár, "Imagination Improves Multimodal Translation", IJCNLP 2017.
- Akiko Eriguchi, Kazuma Hashimoto, and Yoshimasa Tsuruoka, "Tree-to-Sequence Attentional Neural Machine Translation", ACL 2016.
- Marzieh Fadaee and Christof Monz, "Back-Translation Sampling by Targeting Difficult Words in Neural Machine Translation", EMNLP 2018.
- Markus Freitag, Yaser Al-Onaizan, and Baskaran Sankaran, "Ensemble Distillation for Neural Machine Translation", arXiv 2017.

参考文献 (3)

- Jonas Gehring, Michael Auli, David Grangier, Denis Yarats, and Yann N. Dauphin, "Convolutional Sequence to Sequence Learning", ICML 2017.
- Stig-Arne Gronroos et al., "The MeMAD Submission to the WMT18 Multimodal Translation Task", WMT 2018.
- Song Han, Jeff Pool, John Tran, and William Dally, "Learning both Weights and Connections for Efficient Neural Network", NIPS 2015.
- Vu Cong Duy Hoang, Philipp Koehn, Gholamreza Haffari, and Trevor Cohn, "Iterative Back-Translation for Neural Machine Translation", WNMT 2018.
- Po-Yao Huang, Frederick Liu, Sz-Rung Shiang, Jean Oh, and Chris Dyer, "Attention-based Multimodal Neural Machine Translation", WMT 2016.
- Sebastien Jean, Kyunghyun Cho, and Yoshua Bengio, "On Using Very Large Target Vocabulary for Neural Machine Translation", ACL 2015.
- Sebastien Jean, Stanislas Lauly, Orhan Firat, and Kyunghyun Cho, "Does Neural Machine Translation Benefit from Larger Context?", arXiv 2017.
- Melvin Johnson et al, "Google's Multilingual Neural Machine Translation System: Enabling Zero-Shot Translation", TACL 2017.
- Yoon Kim and Alexander M. Rush, "Sequence-Level Knowledge Distillation", EMNLP 2016.
- Taku Kudo, "Subword Regularization: Improving Neural Network Translation Models with Multiple Subword Candidates", EMNLP 2018.
- Guillaume Lample, Myle Ott, Alexis Conneau, Ludovic Denoyer, and Marc'Aurelio Ranzato, "Phrase-Based & Neural Unsupervised Machine Translation", EMNLP 2018.
- Jason Lee, Kyunghyun Cho, and Thomas Hofmann, "Fully Character-Level Neural Machine Translation without Explicit Segmentation", TACL 2017.

参考文献 (4)

- Junhui Li, Deyi Xiong, Zhaopeng Tu, Muhua Zhu, Min Zhang, and Guodong Zhou, "Modeling Source Syntax for Neural Machine Translation", ACL 2017.
- Minh-Thang Luong, Ilya Sutskever, Quoc V. Le, Oriol Vinyals, and Wojciech Zaremba, "Addressing the Rare Word Problem in Neural Machine Translation", ACL 2015.
- Chunpeng Ma, Akihiro Tamura, Masao Utiyama, Tiejun Zhao, and Eiichiro Sumita, "Forest-Based Neural Machine Translation", ACL 2018.
- Sameen Maruf and Gholamreza Haffari, "Document Context Neural Machine Translation with Memory Networks", ACL 2018.
- Haitao Mi, Baskaran Sankaran, Zhiguo Wang, and Abe Ittycheriah, "Coverage Embedding Models for Neural Machine Translation", EMNLP 2016.
- Lesly Miculicich, Dhananjay Ram, Nikolaos Pappas, and James Henderson, "Document-Level Neural Machine Translation with Hierarchical Attention Networks", ACL 2018.
- Toshiaki Nakazawa et al., "Overview of the 4th Workshop on Asian Translation", WAT 2017.
- Jaehong Park, Jongyoon Song, and Sungroh Yoon, "Building a Neural Machine Translation System Using Only Synthetic Parallel Data", arXiv 2017.
- Abigail See, Minh-Thang Luong, and Christopher D. Manning, "Compression of Neural Machine Translation Models via Pruning", CoNLL 2016.
- Rico Sennrich and Barry Haddow, "Linguistic Input Features Improve Neural Machine Translation", WMT 2016.
- Rico Sennrich, Barry Haddow, and Alexandra Birch, "Neural Machine Translation of Rare Words with Subword Units", ACL 2016.
- Rico Sennrich, Barry Haddow, and Alexandra Birch, "Improving Neural Machine Translation Models with Monolingual Data", ACL 2016.

参考文献 (5)

- Ivan Skorokhodov, Anton Rykachevskiy, Dmitry Emelyanenko, Sergey Slotin, Anton Ponkratov, "Semi-Supervised Neural Machine Translation with Language Models", LoResMT 2018.
- Ilya Sutskever, Oriol Vinyals, and Quoc V. Le, "Sequence to Sequence Learning with Neural Networks", NIPS 2014.
- Minh-Thang Luong, Michael Kayser, and Christopher D. Manning, "Deep Neural Language Models for Machine Translation", CoNLL 2015.
- Minh-Thang Luong, Quoc V. Le, Ilya Sutskever, Oriol Vinyals, and Lukasz Kaiser, "MULTI-TASK SEQUENCE TO SEQUENCE LEARNING", ICLR 2017.
- Minh-Thang Luong and Christopher D. Manning, "Achieving open vocabulary neural machine translation with hybrid word-character models", ACL 2016.
- Jorg Tiedemann and Yves Scherrer, "Neural Machine Translation with Extended Context", DiscoMT 2017.
- Zhaopeng Tu, Yang Liu, Lifeng Shang, Xiaohua Liu, and Hang Li, "Neural Machine Translation with Reconstruction", AAAI 2017.
- Zhaopeng Tu, Yang Liu, Zhengdong Lu, Xiaohua Liu, and Hang Li, "Context Gates for Neural Machine Translation", TACL 2017.
- Zhaopeng Tu, Zhengdong Lu, Yang Liu, Xiaohua Liu, and Hang Li, "Modeling Coverage for Neural Machine Translation", ACL 2016.
- Arata Ugawa, Akihiro Tamura, Takashi Ninomiya, Hiroya Takamura, and Manabu Okumura, "Neural Machine Translation Incorporating Named Entity", COLING 2018.
- Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin, "Attention Is All You Need", NIPS 2017.

参考文献 (6)

- Elena Voita, Pavel Serdyukov, Rico Sennrich, and Ivan Titov, "Context-Aware Neural Machine Translation Learns Anaphora Resolution", ACL 2018.
- Longyue Wang, Zhaopeng Tu, Andy Way, and Qun Liu. "Exploiting Cross-Sentence Context for Neural Machine Translation", EMNLP 2017.
- Xinyi Wang, Hieu Pham, Pengcheng Yin, and Graham Neubig, "A Tree-based Decoder for Neural Machine Translation", EMNLP 2018.
- Shuangzhi Wu, Dongdong Zhang, Nan Yang, Mu Li, and Ming Zhou, "Sequence-to-Dependency Neural Machine Translation", ACL 2017.
- Shuangzhi Wu, Dongdong Zhang, Zhirui Zhang, Nan Yang, Mu Li, and Ming Zhou, "Dependency-to-Dependency Neural Machine Translation", TASL 2018.
- Yonghui Wu et al., "Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation", arXiv 2016.
- Poorya Zareemoodi and Gholamreza Haffari, "Incorporating Syntactic Uncertainty in Neural Machine Translation with a Forest-to-Sequence Model", COLING 2018.
- Dakun Zhang, Josep Crego, and Jean Senellart, "Analyzing Knowledge Distillation in Neural Machine Translation", IWSLT 2018.
- Jingyi Zhang, Masao Utiyama, Eiichiro Sumita, Graham Neubig, and Satoshi Nakamura, "Improving Neural Machine Translation through Phrase-based Forced Decoding", IJCNLP 2017.
- Xiaowei Zhang, Wei Chen, Feng Wang, Shuang Xu, and Bo Xu, "Towards Compact and Fast Neural Machine Translation Using a Combined Method", EMNLP 2017.
- Mingyang Zhou, Runxiang Cheng, Yong Jae Lee, and Zhou Yu, "A Visual Attention Grounding Neural Model for Multimodal Machine Translation", EMNLP 2018.



第5回特許情報シンポジウム
AAMT-Japio特許翻訳研究会
2018年12月7日(金)
於 浜松町ビジョンセンター

ユーザーから見た NMTの使い勝手と活用の展望



特定非営利活動法人(NPO)
日本知的財産翻訳協会(NIPTA)
[https://www.nipta.org/
bureau@nipta.org](https://www.nipta.org/bureau@nipta.org)

NIPTAの活動と 特許機械翻訳研究会の紹介

奥山 尚一

日本知的財産翻訳協会 (NIPTA) 理事長

久遠特許事務所代表弁理士

AIPPI, Second Vice President

okuyama@quon-ip.jp

日本知的財産翻訳協会 (NIPTA)の活動

- 知的財産翻訳検定の実施
- ジャーナルの発行
- セミナーの開催
- 研究活動
- <https://www.nipta.org>

日本語と翻訳

- これからも翻訳の重要性は変わらない
- 日本は、営々と翻訳をして外の知識と文化を自家薬籠中のものとして、独自の文化圏を作ってきた
- その結果、日本語は世界に類を見ない優れた言語に育っている
 - 仮名があって、漢字の世界があって、ローマ字入力ができる、外来語も簡単に取り込めるって
 - まじ？

特許翻訳

- 日本全国で10,000人を超える特許翻訳者
- 技術翻訳と、「特許語」とも言うべき独特な国際的ルールの間方をマスターすることが必要
- 優秀な翻訳者は慢性的に不足
- 翻訳者にとっても発注者にとっても基準あるいは目標となるものがあることが望ましい
- → 知的財産翻訳検定

知的財産翻訳検定とは

- 2004年12月より実施
- 全国どこからでも受験が可能！電子メール(Eメール)を使用
- 春には、和文英訳 秋は、英文和訳
- 英語の試験は、1級、2級、3級の3つの級別出題
 - 【1級】分野別で記述式
 - (1)知財法務実務
 - (2)電気・電子工学、(3)機械工学、(4)化学、(5)バイオ
 - 【2級】技術選択なし／記述式
 - 【3級】技術選択なし／記述+選択式問題
- 中文和訳(秋のみ) 技術選択なし／記述式
- 独文和訳(秋のみ) 技術選択なし／記述式

試験時間

- 【午前】
 - 1級(化学、機械工学) 9:00～12:00
 - 3級 9:00～11:00
 - 中国語(※秋のみ実施) 9:00～12:00
- 【午後】
 - 1級(知財法務実務、バイオテクノロジー、電気・電子工学) 14:00～17:00
 - 2級 14:00～17:00
 - ドイツ語(※秋のみ実施) 14:00～17:00
- 和文英訳・英文和訳の1級合格候補者には面接試験
- 午前・午後の言語・級・分野は併願受験が可能

受験料(税込み)

- 受験料は、受験言語および受験級により異なる
- <和文英訳・英文和訳>
 - 【1級】15,000円
 - 【2級】10,000円
 - 【3級】5,000円
- <中文和訳>※秋のみ実施
 - 10,000円
- <独文和訳>※秋のみ実施
 - 10,000円
- 各級合格者には認定証(1級は日本語版と英語版)を交付
- クレジットカードの支払いが可能

1 級の目標

- 知財分野における専門職業翻訳者として推薦できる
 - (1)知財法務実務
 - (2)電気・電子工学、(3)機械工学、(4)化学、(5)バイオ
- 出題内容
 - 法務の(1)は、次の二つの設問〈1〉〈2〉の両方に答える
 - 〈1〉知的財産権利化のプロセスにおいて用いられる文書の翻訳（例えば中間処理文書、審決、審決に対する訴え、など）
 - 〈2〉知的財産権の活用・展開に際して用いられる文書の翻訳（例えば侵害訴訟関連文書、ライセンス契約書、侵害警告書、など）
 - 技術系の(2)から(5)については、技術理解力、特許明細書翻訳力を問う
- いずれも完全記述式
- 二人の委員が採点する

2級と3級

- 2級
 - 特許明細書翻訳の基本を理解し実務に堪える力がある
 - 一般的な技術内容(技術分野選択なし)
 - 特許明細書翻訳力を問う出題
 - 完全記述式
- 3級
 - 入門者・初心者レベル
 - 「特許語」(特許の世界のローカルルール。普通の日本語とも英語とも違う)についての基礎知識を問う(技術分野選択なし)
 - 一部記述式、一部マークシートから選択式(選択式については日英混合)

特許機械翻訳研究会

- 2016年11月11日に事件が起きた
 - Googleが日本語のNMTをアップ(中国語は9月)
 - ネット上で大騒ぎ
 - 今年の6月にはオフライン版を公表
 - 各言語セットは35～45Mバイト
- 2017年10月に当研究会をスタート
 - 種々のNMTの品評会ではない
 - NMTは使えるのか、どう使うのかを考える
 - オープンな集まり
 - NIPTAに限らず、JTF、JAT、ソフト開発者にも声をかける
 - NICT、JAPIO、JST、研究者の方々もゲスト参加
 - 隔月で勉強会
 - いろいろわかってきた！

研究会のメンバー

岡本光弘
梶木正紀
柴田知恵
上野哲也
平林千春
新田順也
湯浅豊裕

浜口宗武
奥山尚一
加藤勇樹
菊地公一
宮本伸也
渡部孝明
田村和彦

Joseph P. Farrar
河目志津加
河野弘毅
柴田純一郎
宮城三次

(順不同)

参考文献

- 機械翻訳の新しいパラダイム:ニューラル機械翻訳の原理
 - 中澤 敏明
 - https://www.jstage.jst.go.jp/article/johokanri/60/5/60_299/html/-char/ja/
- AI vs. 教科書が読めない子どもたち
 - 新井紀子
- 人工知能は人間を超えるか (角川EPUB選書)
 - 松尾豊

技術分野別 NMT特許翻訳評価分析

1. 化学分科会

上野 哲也
株式会社MK翻訳事務所

【化学分科会メンバー】

- 梶木正紀 特許翻訳者(MK翻訳事務所代表)
- 上野哲也 特許翻訳者(MK翻訳事務所)
- 柴田知恵 特許翻訳者(知財コーポレーション)
- 岡本光弘 特許翻訳者(トランスユーロ株式会社翻訳部)
- 平林千春 弁理士・特許翻訳者(平林特許・翻訳事務所)

【評価時期】

- 2018年4月現在

【利用した翻訳エンジン】

- この時期に利用可能であったNMTエンジン3種。

【評価対象】

★英和

NIPTA知財翻訳検定試験第15回

- ・問1(背景技術)
- ・問3(実施例)
- ・問4(クレーム)

★和英

NIPTA知財翻訳検定試験第22回

- ・問1(クレーム)
- ・問4(実施例)

【評価対象】

★NIPTA知財翻訳検定試験第15回

・問1 (背景技術): 169 words

The use of photovoltaic cells for the direct conversion of solar radiation into electrical energy is well known, see Swanson, U.S. Pat. No. 4,234,352 for example. Briefly, the photovoltaic cell comprises a substrate of semiconductive material having a p-n junction defined therein. In the planar silicon cell the p-n junction is formed near a surface of the substrate which receives impinging radiation. Radiated photons create mobile carriers (holes and electrons) in the substrate which can be directed to an electrical circuit outside of the cell. Only photons having at least a minimum energy level (e.g., 1.1 electron volt for silicon) can generate an electron-hole pair in the semiconductor pair. Photons having less energy are either not absorbed or are absorbed as heat, and the excess energy of photons having more than 1.1 electron volt energy (e.g. photons have a wavelength of 1.1 micrometer and less) creates heat. These and other losses limit the efficiency of silicon photovoltaic cells in directly converting solar energy to electricity to less than 30%.

【評価対象】

★NIPTA知財翻訳検定試験第15回

・問3(実施例): 173 words

5-Bromo-1,3-difluoro-2-[(R)-3-methylpentyl]benzene (3)

0.4 g of 5 % polyvinyl chloride (dry) is added to a solution of 1.8 g (5.9 mmol) of 5-Bromo-1,3-difluoro-2-[(R)-3-methylpent-1-enyl]benzene in 0.50 ml of heptane, and the mixture is hydrogenated at atmospheric pressure for 20 hours. The solvent is removed under reduced pressure, and the entire amount is converted into 4.

3,5-Difluoro-4-[(R)-3-methylpentyl]phenylboronic acid (4)

3.3 ml (5.5 mmol) of 1.6 M BuLi are added dropwise at -78 degrees C to a solution of 1.4 g (5 mmol) of 3 in 5 ml of dry diethyl ether. After 30 minutes, 0.6 ml (5.5 mmol) of trimethyl borate is added dropwise. The mixture is allowed to warm to room temperature overnight, 5.2 ml of water, 5.2 ml of methyl tert-butyl ether and 3 ml of conc. HCl are subsequently added. The organic phase is washed with water (2 x 3 ml) and sat. NaCl (1 x 3 ml) and dried using magnesium sulfate, and the solvent is removed under reduced pressure. The residue is filtered through SiO₂ (heptane/dichloromethane 1:1), (yield: 0.8 g, 70%).

【評価対象】

★NIPTA知財翻訳検定試験第15回

・問4(クレーム): 90 words

1. A multi-polymer hydrogel article comprising a first polymeric, water-swellaable material and a second polymeric material, wherein a first region of the article substantially comprises the first polymeric, water-swellaable material, a second region adjacent the first region comprises a mixture of the first polymeric, water-swellaable material and the second polymeric material, and a third region adjacent the second region substantially comprises the second polymeric material, and wherein the second polymeric material exhibits an increasing concentration gradient moving from the first region, through the second region, to the third region.

【評価対象】

★NIPTA知財翻訳検定試験第22回

・問1(クレーム): 380文字

【請求項1】

内燃機関の排気ガスを浄化するためのハニカム構造体であって、
円筒状に形成された金属製の外周壁部(20)と、該外周壁部内に設けられた複数枚の金属製の板状部(21)とを有し、該板状部によって、上記外周壁部内の空間が、上記外周壁部の軸方向に貫通した複数の分割空間(S)に分割された金属フレーム(2)と、

個々の上記分割空間に收容されたハニカムユニット(3)とを備え、
該ハニカムユニットは多角形格子状のセル壁(30)を有し、上記ハニカムユニットには、上記セル壁によって区画され上記排気ガスの流路になる複数のセル(31)が形成されており、

上記ハニカムユニットは、少なくともセリウムとジルコニウムとを含む複合酸化物からなる助触媒と、該助触媒の粒子を互いに結合する無機バインダーとを含有する助触媒含有基材(33)によって形成されていることを特徴とするハニカム構造体。

【評価対象】

★NIPTA知財翻訳検定試験第22回

・問4(実施例): 371文字

熱収縮率

長さ100mmの直線をそれぞれMD方向およびTD方向に対して平行に試験片(フィルム; 15mm×120mm)上に描いた。この試験片を標準状態(温度23°C×湿度50%)に2時間放置し、その後試験前の直線の長さを測定した。続いて、150°Cまたは180°Cまたは200°Cの雰囲気を設定した熱風循環式オーブン内で一角を支持した状態にて30分間放置した後、取り出して標準状態に2時間放置冷却した。その後各方向の直線の長さを測定し、試験前の長さから変化量を求め、当該試験前の長さに対する変化量の割合として熱収縮率R2を求めた。R2について正の値は収縮したことを意味する。

◎; R2の絶対値 $\leq$ 1.5%(最良);

○; 1.5% $<$ R2の絶対値 $\leq$ 2.0%(良);

△; 2.0% $<$ R2の絶対値 $\leq$ 2.5%(実用上問題なし);

×; 2.5% $<$ R2の絶対値(実用上問題あり)。

【評価項目】

- [1] 重要語句の訳抜け(例:クレーム構成部材)
- [2] 軽微な語句の訳抜け(例:符番抜け)
- [3] 係り受けの間違い(修飾句、入れ子構文等)
- [4] 係り受けの間違い(形容詞、副詞等)
- [5] 「係り受け」以外の誤訳・不適切訳(例:～以上・以下の取り違え)
- [6] 専門用語・重要語句の間違い
- [7] 軽微な語句の間違い(例:符番違い、前置詞の使い方)
- [8] 翻訳(MT出力)に起因するもの
- [9] 原文に起因するもの
- [10] 文章になっていない、文章が破綻している
- [11] 主語の取り違え
- [12] 用語のバラつき
- [13] フローティング(原文との脈略がない語句が急に出てくる)

【評価手法】

(1) 各評価者が、機械翻訳された各セグメント中に何個ミスが含まれるかを評価項目ごとにカウントする。

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
原文	MT訳	(a.1) 訳抜け		(a.2) 誤訳・不適切訳 (意味のシフト)				(a.3) 不明瞭訳		(a.4) MT 翻訳特有なエラー						
		重要	軽微	係・修入	係・形副	係受以外	専/重要	軽微	MT 出	原文	非文	主語	ばらばら	フロー		
5-Bromo-1,3-difluoro-2-[(R)-3-methylpentyl]benzene (3)	5-ブロモ-1,3-ジフルオロ-2-[(R)-3-メチルペンチル]ベンゼン (3)															
5-Bromo-1,3-difluoro-2-[(R)-3-methylpentyl]benzene (3)	5-ブロモ-1,3-ジフルオロ-2-[(R)-3-メチルペンチル]ベンゼン(3)。															
5-Bromo-1,3-difluoro-2-[(R)-3-methylpentyl]benzene (3)	5-ブロモ-1,3-ジフルオロ-2-[(R)-3-メチルペンチル]ベンゼン(3)															
0.4 g of 5 % polyvinyl chloride (dry) is added to a solution of 1.8 g (5.9 mmol) of 5-Bromo-1,3-difluoro-2-[(R)-3-methylpent-1-enyl]benzene in 0.50 ml of heptane, and the mixture is hydrogenated at atmospheric pressure for 20 hours.	5-ブロモ-1,3-ジフルオロ-2-[(R)-3-メチルペンタ-1-エニル]ベンゼン1.8g (5.9ミリモル) の溶液に、5%ポリ塩化ビニル (乾燥) 0.4gを加え、ヘプタン0.50mlを加え、混合物を大気圧下で20時間水素化する。															
0.4 g of 5 % polyvinyl chloride (dry) is added to a solution of 1.8 g (5.9 mmol) of 5-Bromo-1,3-difluoro-2-[(R)-3-methylpent-1-enyl]benzene in 0.50 ml of heptane, and the mixture is hydrogenated at atmospheric pressure for 20 hours.	0.4gの5%塩化ポリビニル(乾燥)を0.50mlのヘプタン中の1.8g(5.9mmol)の5-ブロモ-1,3-ジフルオロ-2-[(R)-3-メチルペンタ-1-エニル]ベンゼンの溶液に加え、混合物を大気圧で20時間水素化する。															
0.4 g of 5 % polyvinyl chloride (dry) is added to a solution of 1.8 g (5.9 mmol) of 5-Bromo-1,3-difluoro-2-[(R)-3-methylpent-1-enyl]benzene in 0.50 ml of heptane, and the mixture is hydrogenated at atmospheric pressure for 20 hours.	0.4gの5%ポリ塩化ビニル(乾燥)を、0.50mlのヘプタン中の5-ブロモ-1,3-ジフルオロ-2-[(R)-3-メチルペンタ-1-エニル]ベンゼン1.8g(5.9ミリモル)の溶液に添加し、そして混合物を大気圧で20時間水素化する。															
0.4 g of 5 % polyvinyl chloride (dry) is added to a solution of 1.8 g (5.9 mmol) of 5-Bromo-1,3-difluoro-2-[(R)-3-methylpent-1-enyl]benzene in 0.50 ml of heptane, and the mixture is hydrogenated at atmospheric pressure for 20 hours.	混合物を室温で一晩温め、5.2mlの水、5.2mlのメチルtert-ブチルエーテルおよび3mlの濃塩酸を添加する。															
The mixture is allowed to warm to room temperature overnight, 5.2 ml of water, 5.2 ml of methyl tert-butyl ether and 3 ml of conc.	混合物を室温で一晩加温し、水5.2ml、メチルtert-ブチルエーテル5.2mlおよび濃塩酸3mlを加える。															
The mixture is allowed to warm to room temperature overnight, 5.2 ml of water, 5.2 ml of methyl tert-butyl ether and 3 ml of conc.	混合物を室温まで一晩、5.2mlの水、5.2mlのメチルtert-ブチルエーテルおよび3mlの濃度まで加温する。															
HCl are subsequently added.	続いてHClを加える。															
HCl are subsequently added.	続いてHClを添加する。															
HCl are subsequently added.	続いてHClを添加する。															
The organic phase is washed with water (2 x 3 ml) and sat.	有機相を水 (2×3ml) および飽和食塩水で洗浄する。															
The organic phase is washed with water (2 x 3 ml) and sat.	有機相を水(2×3ml)および飽和NaHCO3で洗浄する。															
The organic phase is washed with water (2 x 3 ml) and sat.	有機相を水(2×3ml)および飽和水で洗浄する。															

専門用語・重要語句
の間違いが2個

2

塩化ビニル

ていない

107

専門用語・重要語句
の間違いが2個

ポリ塩化ビニル

訳出できていない

【評価手法】

(2) 全評価者の評価結果を平均する。

原文	MT訳	(a.1) 訳抜け		(a.2) 誤訳・不適切訳 (意味のシフト)					(a.3) 不明瞭訳		(a.4) MT 翻訳特有なエラー				
		重要	軽微	係・修入	係・形副	係受以外	専/重要	軽微	MT 出	原文	非文	主語	ばらつき	フロー	
5-Bromo-1,3-difluoro-2-[(R)-3-methylpentyl]benzene (3)	5-ブロモ-1,3-ジフルオロ-2-[(R)-3-メチルペンチル]ベンゼン(3)	0	0	0	0	0		0	0	0	0	0	0	0	
0.4 g of 5 % polyvinyl chloride (dry) is added to a solution of 1.8 g (5.9 mmol) of 5-Bromo-1,3-difluoro-2-[(R)-3-methylpent-1-enyl]benzene in 0.50 ml of heptane, and the mixture is hydrogenated at atmospheric pressure for 20 hours.	5-ブロモ-1,3-ジフルオロ-2-[(R)-3-メチルペンタ-1-エニル]ベンゼン1.8g (5.9ミリモル)の溶液に、5%ポリ塩化ビニル(乾燥)0.4gを加え、ヘプタン0.50mlを加え、混合物を大気圧下で20時間水素化する。	0	0.25	0.4	0.5	0		0	0.25	0.25	0	0	0	0.25	
0.4 g of 5 % polyvinyl chloride (dry) is added to a solution of 1.8 g (5.9 mmol) of 5-Bromo-1,3-difluoro-2-[(R)-3-methylpent-1-enyl]benzene in 0.50 ml of heptane, and the mixture is hydrogenated at atmospheric pressure for 20 hours.	0.4gの5%塩化ポリビニル(乾燥)を0.50mlのヘプタン中の1.8g(5.9mmol)の5-ブロモ-1,3-ジフルオロ-2-[(R)-3-methylpent-1-enyl]ベンゼンの溶液に加え、混合物を大気圧で20時間水素化する。	0	0.25	0	0	0	1.16666667	0	0.75		0	0	0	0	
0.4 g of 5 % polyvinyl chloride (dry) is added to a solution of 1.8 g (5.9 mmol) of 5-Bromo-1,3-difluoro-2-[(R)-3-methylpent-1-enyl]benzene in 0.50 ml of heptane, and the mixture is hydrogenated at atmospheric pressure for 20 hours.	0.4gの5%ポリ塩化ビニル(乾燥)を、0.50mlのヘプタン中の5-ブロモ-1,3-ジフルオロ-2-[(R)-3-メチルペンタ-1-エニル]ベンゼン1.8g(5.9ミリモル)の溶液に添加し、そして混合物を大気圧で20時間水素化する。	0	0.25	0	0	0		0.8	0	0	0	0	0	0	
The mixture is allowed to warm to room temperature overnight, 5.2 ml of water, 5.2 ml of methyl tert-butyl ether and 3 ml of conc.	混合物を室温まで一晩温め、5.2mlの水、5.2mlのメチルtert-ブチルエーテルおよび3mlの濃塩酸を添加する。	0	0.25	0	0	0		0.25	0.3333	0.5	0.6	0	0	0.4	
The mixture is allowed to warm to room temperature overnight, 5.2 ml of water, 5.2 ml of methyl tert-butyl ether and 3 ml of conc.	混合物を室温に一晩加温し、水5.2ml、メチルtert-ブチルエーテル5.2mlおよび濃塩酸3mlを加える。	0	0.25	0	0	0		0.25	0.3333	0.5	0.6	0	0	0.4	
The mixture is allowed to warm to room temperature overnight, 5.2 ml of water, 5.2 ml of methyl tert-butyl ether and 3 ml of conc.	混合物を室温まで一晩、5.2mlの水、5.2mlのメチルtert-ブチルエーテルおよび3mlの濃塩酸まで加温する。	0.3333	0.25	0.25	0.3333333333	0		0.4	0.3333	0.5	0.6	0	0	0	
HCl are subsequently added.	続いてHClを加える。	0	0	0	0	0		0	0.3333	0.4	0	0	0	0	
HCl are subsequently added.	続いてHClを添加する。	0	0	0	0	0		0	0.3333	0.4	0	0	0	0	
HCl are subsequently added.	続いてHClを添加する。	0	0	0	0	0		0	0.3333	0.4	0	0	0	0	
The organic phase is washed with water (2 x 3 ml) and sat.	有機相を水(2x3ml)および飽和食塩水で洗浄する。	0	0	0	0	0		0	0.5	0.4	0	0	0	0.4	
The organic phase is washed with water (2 x 3 ml) and sat.	有機相を水(2x3ml)および飽和NaHCO3で洗浄する。	0	0	0	0	0		0.5	0	0.4	0	0	0	0.5	
The organic phase is washed with water (2 x 3 ml) and sat.	有機相を水(2x3ml)および飽和水で洗浄する。	0	0	0	0	0		0.5	0	0.4	0	0	0	0.5	
NaCl (1 x 3 ml) and dried using magnesium															

【気になる点①:クレームの訳出が苦手】

人間による英和翻訳:

1. A multi-polymer hydrogel article comprising a first polymeric, water-swallowable material and a second polymeric material, wherein a first region of the article substantially comprises the first polymeric, water-swallowable material, a second region adjacent the first region comprises a mixture of the first polymeric, water-swallowable material and the second polymeric material, and a third region adjacent the second region substantially comprises the second polymeric material, and wherein the second polymeric material exhibits an increasing concentration gradient moving from the first region, through the second region, to the third region.



第一の水膨潤性ポリマー材料と第二のポリマー材料を含むマルチポリマーハイドロゲル物品であって、

前記物品の第一領域が、実質的に前記第一の水膨潤性ポリマー材料を含み、

前記第一領域に隣接する第二領域が、前記第一の水膨潤性ポリマー材料および前記第二のポリマー材料の混合物を含み、

前記第二領域に隣接する第三領域が、実質的に前記第二のポリマー材料を含み、

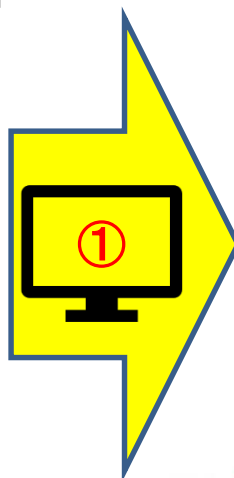
前記第二のポリマー材料が、前記第一領域から前記第二領域を経て前記第三領域に進むにつれて増加する濃度勾配を示す、

マルチポリマーハイドロゲル物品。

【気になる点①:クレームの訳出が苦手】

NMTその1による英和翻訳:

1. A multi-polymer hydrogel article comprising a first polymeric, water-swallowable material and a second polymeric material, wherein a first region of the article substantially comprises the first polymeric, water-swallowable material, a second region adjacent the first region comprises a mixture of the first polymeric, water-swallowable material and the second polymeric material, and a third region adjacent the second region substantially comprises the second polymeric material, and wherein the second polymeric material exhibits an increasing concentration gradient moving from the first region, through the second region, to the third region.



物品の第1の領域が実質的に第1のポリマー性の水膨潤性材料を含み、

第1の領域に隣接する第2の領域が第2のポリマー性水膨潤性材料を含むマルチポリマーヒドロゲル物品であって、

第2のポリマー材料を実質的に含み、

第2のポリマー材料は、第1の領域から第2のポリマー材料を通して移動する漸増する濃度勾配を示す。

第2の領域から第3の領域に移動する。



クレームの体を
なしていない...

【気になる点①:クレームの訳出が苦手】

NMTその2による英和翻訳:

1. A multi-polymer hydrogel article comprising a first polymeric, water-swallowable material and a second polymeric material, wherein a first region of the article substantially comprises the first polymeric, water-swallowable material, a second region adjacent the first region comprises a mixture of the first polymeric, water-swallowable material and the second polymeric material, and a third region adjacent the second region substantially comprises the second polymeric material, and wherein the second polymeric material exhibits an increasing concentration gradient moving from the first region, through the second region, to the third region.



111

1. 第1のポリマー、水膨潤性材料および第2のポリマー材料を含む多ポリマーヒドロゲル物品であって、前記物品の第1の領域が実質的に第1のポリマー、水膨潤性材料を含み、前記第1の領域に隣接する第2の領域が前記第1のポリマー、水膨潤性材料および前記第2のポリマー材料の混合物を含み、前記第2の領域に隣接する第3の領域が実質的に前記第2のポリマー材料を含み、前記第2のポリマー材料が、前記第1の領域から前記第2の領域を通して前記第3の領域に移動する増加する濃度勾配を示す、多ポリマーヒドロゲル物品。

- ・第1ポリマーと水膨潤性材料が別物に見える。
- ・「移動する」が変。

【気になる点①:クレームの訳出が苦手】 NMTその3による英和翻訳:

1. A multi-polymer hydrogel article comprising a first polymeric, water-swallowable material and a second polymeric material, wherein a first region of the article substantially comprises the first polymeric, water-swallowable material, a second region adjacent the first region comprises a mixture of the first polymeric, water-swallowable material and the second polymeric material, and a third region adjacent the second region substantially comprises the second polymeric material, and wherein the second polymeric material exhibits an increasing concentration gradient moving from the first region, through the second region, to the third region.



1. **第1のポリマーの水膨潤性材料**と第2のポリマー材料とを含むマルチポリマーヒドロゲル物品であって、前記物品の第1の領域は、実質的に、前記第1のポリマーの水膨潤性材料を含み、前記第1の領域に隣接する第2の領域は、前記第1のポリマーの水膨潤性材料と前記第2のポリマー材料との混合物を含み、前記第3の領域に隣接する第2の領域は、実質的に、前記第2のポリマー材料を含み、前記第2のポリマー材料は、前記第1の領域から前記第2の領域を通して前記第3の領域に向かって**移動する勾配勾配**を示す、マルチポリマーヒドロゲル物品。



112

- ・「第1のポリマーの水膨潤性材料」が変？
- ・「移動する」が変。
- ・「勾配勾配」??

【気になる点①:クレームの訳出が苦手】

NMTによる和英翻訳:

【請求項1】

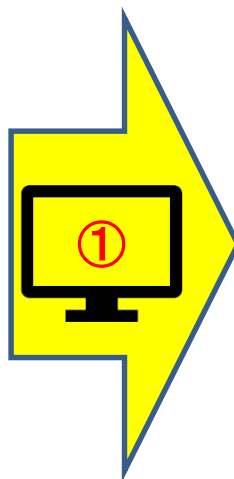
内燃機関の排気ガスを浄化するためのハニカム構造体であって、

円筒状に形成された金属製の外周壁部(20)と、該外周壁部内に設けられた複数枚の金属製の板状部(21)とを有し、該板状部によって、上記外周壁部内の空間が、上記外周壁部の軸方向に貫通した複数の分割空間(S)に分割された金属フレーム(2)と、

個々の上記分割空間に収容されたハニカムユニット(3)とを備え、

該ハニカムユニットは多角形格子状のセル壁(30)を有し、上記ハニカムユニットには、上記セル壁によって区画され上記排気ガスの流路になる複数のセル(31)が形成されており、

上記ハニカムユニットは、少なくともセリウムとジルコニウムとを含む複合酸化物からなる助触媒と、該助触媒の粒子を互いに結合する無機バインダーとを含有する助触媒含有基材(33)によって形成されていることを特徴とするハニカム構造体。



113

A honeycomb structure for purifying exhaust gas of an internal combustion engine.

A metal frame (2) comprising : a metal outer peripheral wall (20) formed in a cylindrical shape ; and a plurality of metal plate portions (21) provided in the outer peripheral wall portion, wherein a space in the outer peripheral wall portion is divided into a plurality of divided spaces (S) penetrating axially of the outer peripheral wall portion.

The honeycomb unit (3) **as claimed in any one of the preceding claims, comprising a honeycomb unit (10).**

The honeycomb unit **according to claim 30,** wherein the honeycomb unit has a polygonal lattice cell wall (), and the honeycomb unit is formed with a plurality of cells (31) which are partitioned by the cell walls to serve as flow paths of the exhaust gas.

A honeycomb structure, characterized in that the honeycomb unit is formed by a catalyst-containing substrate (33) containing a co-catalyst comprising a composite oxide containing at least cerium and zirconium and an inorganic binder which bonds the particles of the co-catalyst together.



29

【気になる点②:ピリオドを含む略語が苦手】

The mixture is allowed to warm to room temperature overnight, 5.2 ml of water, 5.2 ml of methyl tert-butyl ether and 3 ml of **conc.** HCl are subsequently added.



混合物を室温まで一晩温め、5.2mlの水、5.2mlのメチルtert-ブチルエーテルおよび3mlの濃塩酸を添加する。**続いてHClを加える。**



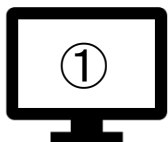
混合物を室温に一晩加温し、水5.2ml、メチルtert-ブチルエーテル5.2mlおよび濃塩酸3mLを加える。**続いてHClを添加する。**



混合物を室温まで一晩、5.2mlの水、5.2mlのメチルtert-ブチルエーテルおよび3mlの**濃度まで加温**する。**続いてHClを添加する。**

【気になる点②:ピリオドを含む略語が苦手】

The organic phase is washed with water (2 x 3 ml) and **sat.** NaCl (1 x 3 ml) and dried using magnesium sulfate, and the solvent is removed under reduced pressure.



有機相を水(2×3ml)および飽和食塩水で洗浄する。**NaCl(1×3ml)**で**洗浄し**、硫酸マグネシウムを用いて乾燥させ、溶媒を減圧下で除去する。



有機相を水(2×3ml)および飽和**NaHCO₃**で洗浄する。**NaCl(1×3ml)**および硫酸マグネシウムを用いて乾燥させ、溶媒を減圧下で除去する。



有機相を水(2×3ml)および飽和水で洗浄する。**NaCl(1×3ml)**を硫酸マグネシウムを用いて**乾燥し**、そして溶媒を減圧下で除去する。



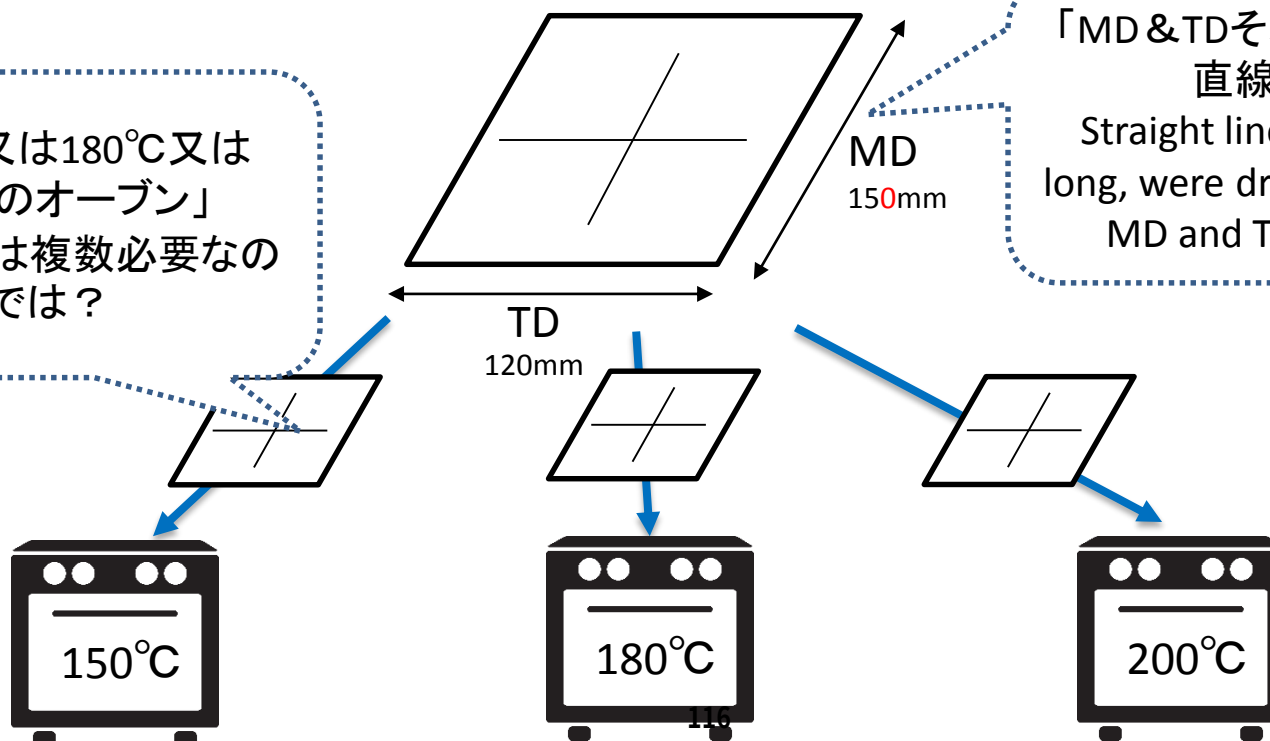
このほか、anhyd.(無水)やPat.なども苦手。

【気になる点③: 単複の見分けが苦手】

長さ100mmの直線をそれぞれMD方向およびTD方向に対して平行に試験片(フィルム; 15mm × 120mm)上に描いた。この試験片を標準状態(温度23°C × 湿度50%)に2時間放置し、その後試験前の直線の長さを測定した。続いて、150°Cまたは180°Cまたは200°Cの雰囲気中に設定した熱風循環式オーブン内で一角を支持した状態にて30分間放置した後、取り出して標準状態に2時間放置冷却した。その後各方向の直線の長さを測定し、...

「150°C又は180°C又は200°Cのオーブン」
= 試験片は複数必要なのでは？

「MD & TDそれぞれに平行」= 直線は複数？
Straight lines, each 100 mm long, were drawn parallel to the MD and TD, respectively.



【気になる点③: 単複の見分けが苦手】



A **straight line** having a length of 100 mm was drawn on a **test piece** (film; 15 mm × 120 mm) parallel to the MD direction and the TD direction. The **specimen** was allowed to stand for 2 hours in a standard state (temperature 23 ° C. × humidity 50%), and then the length of the **straight line** before the test was measured.

- ・用語の不統一 (test piece, specimen) (次スライド)
- ・「×」の使い方



A **straight line** of 100 in length was drawn on a test piece (film ; 15 × 120) parallel to the MD and TD directions, respectively. The test piece was left in a standard state (23 ° C. × 50% humidity) for 2 hours, and then the length of a **straight line** before the test was measured.

- ・単位抜け
- ・既出概念の初出扱い



A **straight line** of the length 100 mm was drawn on the specimen (film; 15 mm × 120 mm) parallel to the MD and TD directions, respectively. The **specimens** were left under standard conditions (temperature 23° C × 50% humidity) for 2 hours, after which the length of the straight line before testing was measured.

- ・初出概念の既出扱い
- ・単複の不一致

【気になる点④：用語の統一が苦手】

試験片



A straight line having a length of 100 mm was drawn on a **test piece** (film; 15 mm × 120 mm) parallel to the MD direction and the TD direction. The **specimen** was allowed to stand for 2 hours in a standard state (temperature 23 ° C. × humidity 50%), and then the length of the straight line before the test was measured.

熱収縮率



Thermal shrinkage

中略

Thereafter, a length of a straight line in each direction was measured, a change amount was obtained from a length before the test, and a **heat shrinkage ratio** R 2 was obtained as a ratio of a change amount to a length before the test.

気になる点⑤: 化合物名の訳出が苦手】

Polyvinyl
chloride



塩化ポリビニル

5-Bromo-1,3-
difluoro-2-[(R)-3-
methylpent-1-
enyl]benzene



5-ブロモ-1,3-ジフルオロ
-2-[(R)-3methylpent1enyl]
ベンゼン

【気になる点⑥: 原文との脈略がない語句が急に出てくる】

★「フローティング」と呼ばれる

該ハニカムユニットは多角形格子状のセル壁(30)を有し、上記ハニカムユニットには、上記セル壁によって区画され上記排気ガスの流路になる複数のセル(31)が形成されており、



The honeycomb unit **according to claim 30**, wherein the honeycomb unit has a polygonal lattice cell wall (), and the honeycomb unit is formed with a plurality of cells (31) which are partitioned by the cell walls to serve as flow paths of the exhaust gas.

The organic phase is washed with water (2 x 3 ml) and sat. NaCl (1 x 3 ml) and dried using magnesium sulfate, and the solvent is removed under reduced pressure.



有機相を水(2×3ml)および飽和 **NaHCO₃** で洗浄する。NaCl(1×3ml)および硫酸マグネシウムを用いて乾燥させ、溶媒を減圧下で除去する。

【MT初心者が使ってみた感想(まとめ)】

★長所

- ・訳文自体は読みやすい場合が多い。
- ・仕事でそのまま使えるレベルではないが、
全体的には上手く訳せている。ときには驚くほど上手。

★短所

- ・クレームの訳出はまだ苦手(特に和→英)。
- ・ピリオドを含む略語が苦手。
- ・単複の見分けが苦手。
- ・用語の統一が苦手。
- ・化合物名の訳出が苦手。
- ・フローティング(原文との脈略がない語句が急に出てくる)

【MTを既に活用しているMK翻訳事務所様からの提言】

- 機械翻訳を使用する場合は、CATツールを使用した方が効果的である。
- ツールを使用して、セグメント分割を行う（セグメントを短くする）。
- その上で再度、機械翻訳を走らせる。
- ポストエディットが終われば、セグメント結合を行う。
- 機械翻訳は、作業を10～20%程度軽減してくれるものだと割り切る（自分を信じる）。

技術分野別 NMT特許翻訳評価分析

2. 機械工学分科会

湯浅豊裕

WIS知財コンシェル株式会社 翻訳事業部
日本知的財産翻訳協会(NIPTA)理事

【機械工学分科会メンバー】

奥山 尚一（久遠特許事務所 代表・弁理士）

浜口 宗武（株式会社知財コーポレーション 代表取締役会長）

新田 順也（エヌ・アイ・ティー株式会社 代表取締役）

鈴木 彰浩（トランスユーロ株式会社）

湯浅 豊裕（WIS知財コンシェル株式会社）

1. NMTをつかってみました。

知的財産翻訳検定1級（機械工学）の過去問を
NMTにかけてみました。

1級に合格できる品質ではありませんが、
ツールとして利用することはできます。

2. 得意なところ

概して、**自然なことば(流暢)に翻訳**されています。

Having two stems 11 enables easier access if the component 1 is located in a hard-to-reach place or is placed close to a wall.

2つのステム11を有することにより、コンポーネント1が届きにくい場所に配置されている場合、または壁の近くに配置されている場合にアクセスが容易になる。

2. 得意なところ

流暢である(読みやすい)ことの**利点**

チェックがしやすいです。

流暢さとは関係ありませんが、

タイピングの負荷が減ります。

3. 苦手なところ(形式的なところ)

訳語が揺れることがあります。

蒸着工程が繰り返される過程でチャンバ20と噴射ノズル10は、繰り返し熱によって膨張と収縮が繰り返される繰り返し荷重に露出し、(中略)、噴射ノズル10とチャンバの外壁21との間には微細な隙間が形成される可能性がある。

In the process of repeating the vapor deposition process, the chamber 20 and the spray nozzle 10 are exposed to repeated loads in which expansion and contraction are repeated by repeated heat, and (中略) there is a possibility that a minute gap is formed between the injection nozzle 10 and the outer wall 21 of the chamber.

3. 苦手なところ(形式的なところ)

脈絡のない語が出てくることがあります。

As will be appreciated from the above, turning one of the stems 11 will cause the ball 7 to rotate between a closed position, where the bore 9 is perpendicular to the main bore 4, and an open position, where the bore communicates with the main bore.

上記から理解されるように、ステム11の1つを回転させることにより、ボア9が主ボア4に対して垂直である閉位置と、ボアが主ボア4に連通する開位置との間でボール7が回転する。**ボア。**

3. 苦手なところ(形式的なところ)

訳抜けすることがあります。

The invention will now be described, by way of example, with reference to the accompanying drawings, in which: **Figure 2a** is a side-sectional view of a valve component in accordance with a first embodiment of the invention; Figure 2b is a side perspective view of the valve component shown in Figure 2a.

本発明を、添付の図面を参照して例として以下に説明する。本発明の第1の実施形態による弁構成要素の側断面図である。図2bは、図2aに示される弁構成要素の側面斜視図である。

3. 苦手なところ(形式的なところ)

傾向と対策

エラーにもパターンがあります。

さまざまなプレエディットやポストエディットの方法論
が提案されています。

4. 苦手な点(実体的なところ)

字面を追った解釈をすることがあります。

ゴルフクラブ(以下クラブ)には飛んで曲がらない(曲がりにくい)ことが最も要求されている。

It is most demanded that a golf club (hereinafter referred to as a club) does not bend by bending (hardly bending).

4. 苦手な点(実体的なところ)

技術(内容)を理解していないと翻訳できない箇所は
苦手なようです。

人間も苦手？

5. 機械翻訳とCATツールを一緒につかう

CATツールにより、

NMTの苦手なところを補うことができます。

便利な機能を何か1つ見つけて使ってみてください。

5. 機械翻訳とCATツールを一緒につかう

長い文を**短い文に分割**する



機械翻訳にかける



パズルのように組み立てる。

www.BANDICAM.COM

memoQ プロジェクト 文書 準備 翻訳 レビュー 編集 表示 クイックアクセス

ソースを空のターゲットにコピー 数字のみのセグメントを自動記入 前翻訳 翻訳をクリア 翻訳 検索と置換 検索 次を検索 直接 詳細 セグメントをロック/ロック解除 セグメントの状況を変更 QA 変更と融合 正規表現タグ化ツール 用語を抽出 確定して更新 略語の検索 コメントをクリア

プロジェクトホーム コミュニケーション CATxMT sample.docx

ソース ターゲット 並べ替え 並べ替えない

10.	【2-1】 長い文をMTIにかけて、PEする。	0%	
11.	図4の工程(c)に示すように、無機バリア膜21を覆うように、例えば、膜厚10nm以上1000nm以下程度で有機アルミニウム膜を、トリメチルアルミニウムとエチレングリコールとを、例えば、13.56MHzの高周波電力で励起してプラズマ状態で反応させる分子層堆積法で成膜して、無機有機複合バリア膜22を形成する(分子層堆積工程)。	0%	
	As shown in step (c) of FIG. 4, an organic aluminum film is formed to cover the inorganic barrier film 21, for example, with a film thickness of about 10 nm or more and 1000 nm or less, trimethylaluminum and ethylene glycol are mixed at, for example, 13.56 MHz And is formed by a molecular layer deposition method excited by high frequency power and reacted in a plasma state to form an inorganic-organic composite barrier film 22 (molecular layer deposition step).	0%	
12.		0%	
13.	【2-2】 長い文を短いかたまりに分けて、それをMTIにかけて、PEする。	0%	
14.	図4の工程(c)に示すように、	0%	
	As shown in step (c) of FIG. 4,	0%	
15.	無機バリア膜21を覆うように、	0%	
	In order to cover the inorganic barrier film 21,	0%	
16.	例えば、	0%	
	For example,	0%	
17.	膜厚10nm以上1000nm以下程度で有機アルミニウム膜を、	0%	
	An organic aluminum film having a film thickness of about 10 nm or more and 1000 nm or less,	0%	
18.	トリメチルアルミニウムとエチレングリコールとを、	0%	
	Trimethylaluminum and ethylene glycol,	0%	
19.	例えば、	0%	
	For example,	0%	
20.	13.56MHzの高周波電力で励起してプラズマ状態で反応させる分子層堆積法で成膜して、	0%	
	A film is formed by a molecular layer deposition method exciting with a high frequency power of 13.56 MHz and reacting in a plasma state,	0%	
21.	無機有機複合バリア膜22を形成する	0%	
	Thereby forming the inorganic-organic composite barrier film 22	0%	
22.	分子層堆積工程)。	0%	
	(Molecular layer deposition step).	0%	

翻訳結果

図4の工程(c)に示すように、...

As shown in step (c) of FIG. 4, an organic aluminum film is formed to cover the inorganic barrier film 21, for example, with a film thickness of about 10 nm or more and 1000 nm or less, trimethylaluminum and ethylene glycol are mixed at, for example, 13.56 MHz and is formed by a molecular layer deposition method excited by high frequency power and reacted in a plasma state to form an inorganic-organic composite barrier film 22 (molecular layer deposition step).

図4の工程(c)に示すように、無機バリア膜21を覆うように、例えば、膜厚10nm以上1000nm以下程度で有機アルミニウム膜を、トリメチルアルミニウムとエチレングリコールとを、例えば、13.56MHzの高周波電力で励起してプラズマ状態で反応させる分子層堆積法で成膜して、無機有機複合バリア膜22を形成する(分子層堆積工程)。

As shown in step (c) of FIG. 4, an organic aluminum film is formed to cover the inorganic barrier film 21, for example, with a film thickness of about 10 nm or more and 1000 nm or less, trimethylaluminum and ethylene glycol are mixed at, for example, 13.56 MHz and is formed by a molecular layer deposition method excited by high frequency power and reacted in a plasma state to form an inorganic-organic composite barrier film 22 (molecular layer deposition step).

postedit sample_20180605... 9:39 10:00

ご参考にしていただけたところがありましたら幸いです。

ありがとうございました。

技術分野別 NMT特許翻訳評価分析

3. 電気・電子工学分科会

菊地 公一

DEN IPソリューションズ株式会社 代表取締役
弁理士

【電気・電子工学分科会メンバー】

- ・ 田村和彦 弁理士・平木国際特許事務所
- ・ 宮本伸也 日本ビジネス翻訳株式会社代表取締役社長
- ・ 渡部孝明 株式会社翻訳センター特許営業部大阪オフィス
- ・ 菊地公一 弁理士・DEN IP ソリューションズ株式会社

◎ NMT使ってみてどうだった？（実用性評価・分析結果）

[A社]

英日：同じ翻訳文が繰り返し出現する現象が起こった。

日英：未訳（日本語が残る）があった。

[B社]

英日：「です・ます」調の翻訳文があった。

日英：並列関係が不適切な翻訳文があった。

[C社]

英日：単語の「意味」としては正しいが、「翻訳」としては不適切な箇所があった。

日英：文章として成立していない翻訳文があった。

《総評》

実用性の点から観て各社の出力結果に大きな差は見られなかった。

原文の複雑さに依拠するにせよ、最終成果物としてはPostEditは必要と思われる。尚、PostEditの作業は項目化して進める方が望ましい。

◎ NMT使ってみてどうだった？（実用性評価・分析結果）

PostEdit作業の項目化（例）

（共通）

- ✓ 用語の統一
- ✓ 表現の統一
- ✓ 未訳の検出
- ✓ 湧き出し・フローティングの検出
- ✓ 全角・半角

（英日）

- ✓ 文体の統一：「です・ます」調→「である」調

（日英）

- ✓ 冠詞：初出/既出
- ✓ 単数/複数

◎ こんなところ何とかならない？（開発者側への提言）

（1）湧き出し・フローティング・脱訳について

注記・警告らしき表示が出るとPostEdit時に注意を払えるため便利

（2）用語登録オプション

登録する用語に関して品詞指定が出来れば便利
 さもないと反ってNMTの品質低下を招く？

（例文）送風機は、ファンの回転数を増減させるインバータを備える。

* 用語登録なし

The blower includes an inverter for *increasing or decreasing* the number of revolutions of the fan.

* 用語登録あり（増減：variation）

The blower includes inverters *for variation* the rotational speed of the fan.

◎ NMTは果たして脅威？（翻訳サプライヤ側への提言）

『脅威』と感じたところで時代の流れは止まらない

テクノロジーの進歩はむしろ好機と捉えるべきではないか？

タイピングの重労働からの解放

今後は翻訳文の質的な面を中心に注力することが可能

* プロの翻訳者からプロのPostEditorへの転身

◎ ユーザ側で対応可能な改善策(翻訳サプライヤへの提言)

1. ネスト構造文を避ける

- ✓ MTに不向きな日本語をPreEditする
- ✓ 例えば、修飾句の挿入は1つまで

2. 文理解釈の容易な日本語にPreEditする

- ✓ 特許ライティングマニュアル(JAPIO)を参考にPreEdit
- ✓ (例)色deチェックの利用

3. 用語のバラつき等を抑える

- ✓ 翻訳支援ツールと組み合わせる
- ✓ (例) Memsourceの利用

◎弁理士、特許出願人から見た要望、懸念
(NMT、ツール・人によるチェックのどこで担保するか)

✓ 訳抜けはゼロに (⇔ NMTでは訳抜け多発)

→ 審査や権利行使などで不利益。
1語の訳抜けで特許が無効になり得る。

✓ 発明の内容を正確に反映

→ 本来の発明を権利でカバーしたい

◎機械翻訳の導入で変わること、変わらないこと

- ✓ 機械訳をリライトすると、機械の訳にひっぱられる？

例： ～(排出)するおそれがある There is a danger of discharging...

リライト1： There is a possibility to discharge ...

訳例： ... possibly discharge ...

- ✓ 人が、原文から英文を作る能力は必要

- ✓ 翻訳作業は効率化される。

ただし、原文の読み込み・理解、翻訳作業後の見直しは省略できない。

→ 効率化された分、翻訳料金は下がる？

原文の読み込みなどに注力して、料金維持+質向上？

◎ 5年後10年後は？（近未来予測）

PostEditを中心とした業務

- ✓ 品質はむしろ向上する方向へ向かうかも？

スクラッチからの翻訳はマイナー言語を除くと少なくなる？

- ✓ 但し、課題はPostEditorの担い手
- ✓ MT以上の高品質さが求められる
- ✓ 適正な市場価格の形成

以上

技術分野別 NMT特許翻訳評価分析

4. 知財法務実務分科会

宮城 三次

一色国際特許業務法人パートナー弁理士

知的財産法務実務 ワーキンググループ

- メンバー(五十音順、敬称略)
 - － 河野弘毅
 - 機械翻訳コンサルタント・JTF理事
 - － 河目志津加
 - 特許翻訳者
 - － 柴田純一郎
 - 弁理士・ソニーピクチャーズエンタテインメント株式会社法務部・NIPTA試験委員
 - － Joseph Farrar
 - 米国特許弁護士・一色外国法事務弁護士事務所・元NIPTA試験委員
 - － 宮城三次
 - 弁理士・一色国際特許業務法人パートナー・NIPTA試験委員

はじめに

- 知的財産法務実務というカテゴリー
 - 知的財産権利化のプロセスにおいて用いられる文書の翻訳(例えば中間処理文書、審決、審決に対する訴え、など)
 - 知的財産権の活用・展開に際して用いられる文書の翻訳(例えば侵害訴訟関連文書、ライセンス契約書、侵害警告書、など)
- 特許明細書以外の文書の翻訳全般

機械翻訳エンジンの評価手法

- 評価対象
 - Google翻訳 (Google, Inc.)
 - 汎用NMT (日本語・英語) (国立研究開発法人情報通信研究機構)
- 評価用原文
 - 日本知的財産翻訳検定協会 (NIPTA) が実施している検定試験の過去問から、機械翻訳エンジンの翻訳性能を客観的に評価するのに適当と考えられる問題文を選定

機械翻訳エンジンの評価手法

• 評価指標

- 評価対象である翻訳エンジンに翻訳させた結果を、以下の指標を用いて評価
- 訳抜け
 - 重要語句の訳抜け(例: クレーム構成部材)
 - 軽微な語句の訳抜け(例: 符番抜け)
- 誤訳・不適切訳(意味のシフト)
 - 係り受けの間違い(修飾句、入れ子構文等)
 - 係り受けの間違い(形容詞、副詞等)
 - 「係り受け」以外の誤訳・不適切訳(例: ～以上・以下の取り違い)

機械翻訳エンジンの評価手法

- 評価指標(つづき)
 - 不明瞭訳
 - 翻訳(MT出力)に起因するもの
 - 原文に起因するもの
 - 機械翻訳特有の誤り
 - 文章になっていない、文章が破綻している
 - 主語の取り違い
 - 用語のバラつき
 - フローティング(原文との脈略がない語句が急に出てくる)

評価用原文

- (イ)原告は、各基礎出願明細書の図2から、螺旋溝27の傾斜角度が「約20度」であることが読み取れ、当該「約20度」を中央値として「±10度」とすることで、「10度から30度」の範囲が設定できることは、図面および明細書の記載から当業者にとって自明な事項である旨主張する。
- しかしながら各基礎出願明細書の図2は、前示のとおり、クランプロッドの下摺動部分の拡大展開図、すなわち、4つのガイド溝26の配設態様を示した特許図面にすぎず、設計図面のように各ガイド溝に係る寸法値や角度値まで看取し得るほど正確に描かれた図面とは認められない。
- したがって、そのような特許図面から「螺旋溝27」の具体的な傾斜角度（「約20度」）を読み取ることはできないというべきである。

評価用原文(つづき)

- また、仮に、上記図2から、螺旋溝27の具体的な傾斜角度を「約20度」と読み取ることが可能であったとしても、図2は、螺旋溝27の傾斜角度を「中央値」として示した図とは認められないから、当該「約20度」を「中央値」と解すべき理由はない。
- さらに、そのような「約20度」を中央値と設定した上に、「±10度」することで上限・下限値を設定し螺旋溝27の傾斜角度の範囲とすることも、合理性はない。すなわち、各基礎出願明細書に、「30度」を上限とし「10度」を下限とする傾斜角度の範囲が記載されたものと認めることはできないし、これを自明な事項とすることもできないというべきである。

評価用原文(つづき)

- (5) 以上総合すると、本件特許発明1の「前記の第2端壁(3b)に緊密に嵌合支持されると共に上記の3つ又は4つのガイド溝(26)が設けられた上記の第2摺動部分(12)について、その第2摺動部分(12)の外周面を展開した状態における上記の旋回溝(27)の傾斜角度(A)を10度から30度の範囲内に設定し、かつ、上記の隣り合うガイド溝(26)(26)の隔壁の最小厚さ(T)を、同上のガイド溝(26)の溝幅(W)よりも小さい値に設定した」(構成①, ②)との事項は、優先権主張の基礎とされた第1ないし第3基礎出願の明細書等に記載されていないことは明らかである。

原文第1段落の対比

- (イ) 原告は, 各基礎出願明細書の図2から, 螺旋溝27の傾斜角度が「約20度」であることが読み取れ, 当該「約20度」を中央値として「±10度」とすることで, 「10度から30度」の範囲が設定できることは, 図面および明細書の記載から当業者にとって自明な事項である旨主張する。
- (A) The plaintiff can read that the angle of inclination of the spiral groove 27 is "about 20 degrees" from FIG. 2 of each basic application specification, it can be read as "± 10 degrees" with the "about 20 degrees" as the median value That the range of "10 degrees to 30 degrees" can be set, it is obvious to those skilled in the art from the drawings and the description of the specification.

代表的な誤り

1. 主語と述語の対応を読み違っている
 2. 上記に伴う意味不明訳の発生
 3. 重要語句の脱落
- → 原文を機械翻訳によりフィットするように書き換えたら...

原文の修正方針

- 入れ子構成を解消する
 - A...原告は, (B, C) 旨主張する。
 - B...(C)ことは, 図面および明細書の記載から当業者にとって自明な事項である
 - C...各基礎出願明細書の図2から, 螺旋溝27の傾斜角度が「約20度」であることが読み取れ, 当該「約20度」を中央値として「±10度」とすることで, 「10度から30度」の範囲が設定できる

原文の修正方針

- C部を分割する
 - C1...各基礎出願明細書の図2から、螺旋溝27の傾斜角度が「約20度」であることが読み取れる。
 - C2...当該「約20度」を中央値として「±10度」とすることで、「10度から30度」の範囲が設定できる

原文の修正方針

- C2の表現（中央値として、「±10度」とする）を言い換える
 - C2...当該「約20度」を中央値と設定して、その中央値から10度増減することで、「10度から30度」の範囲が設定できる

原文の修正方針

- 以上をまとめると、例えば以下のMT入力用変換文を得る。
 - 原告は、下記を主張する。
 - 以下のことは、図面および明細書の記載から当業者にとって自明な事項である。
 - 各基礎出願明細書の図2から、螺旋溝27の傾斜角度が「約20度」であることが読み取れる。したがって、当該「約20度」を中央値と設定して、その中央値から10度増減することで、「10度から30度」の範囲が設定できる。

修正後原文の翻訳結果

- Plaintiff insists on the following.
- The following will be obvious to those skilled in the art from the drawings and the description of the specification.
- From FIG. 2 of each of the basic patent specifications, it can be seen that the inclination angle of the spiral groove 27 is "about 20 degrees". Therefore, by setting "about 20 degrees" as the median value and increasing / decreasing by 10 degrees from the median value, the range of "10 degrees to 30 degrees" can be set.
(Google翻訳, 2018-07-20)

まとめ・今後の課題

- 判決文のように一文の長さが非常に長く、係り受けが複雑な文書では、主述の読み間違いといった基本的な誤りが生じやすい。
- 原文について、短文化、係り受けの整理などのプレエディットを行うことで、前記のような誤りを予防することができる。
- 機械翻訳のメリットである高速処理により、プレエディットを反復して行いながら質の高い翻訳文を効率的に生産可能となることが期待される。
- 知財権利化過程の種々の文書（拒絶に対する意見書等）、各種契約書についてもプレエディットの効果を確認していきたい。

NMTを翻訳者の秘書にする ～「NMTの間違いを指摘する」ツールを開発～

新田 順也

翻訳者

色deチェック等翻訳マクロ開発者

Microsoft MVP for Office Apps and Services受賞者

ブログ「みんなのワードマクロ」管理人

nitta@n-i-t.jp

誤訳のまとめ

(共通)

- ・専門用語の間違い(誤訳、略語の誤判定)
- ・訳揺れ(文章間で異なる訳語)
- ・訳もれ(単語、文章単位、肯定/否定)
- ・並列関係の間違い
- ・フローティング(単語、フレーズが追加される)
- ・数字の誤記(日付、数値、参照符号の分離)
- ・主語や係り受けの間違い
- ・MT出力に起因する間違い(繰り返し)
- ・全角/半角の表記

(英日)

- ・常体と敬体が混在

(日英)

- ・単数/複数の間違い

NMTの活用方法とは？

- すべてが完璧な訳文ではないが使える訳文もある
- プリエディットやポストエディットをすれば使える訳文が増える
- 文脈を考慮した訳文作成のために翻訳者が必須
- 機械翻訳が困難な文章は人間が翻訳すればよい



**適切なツールの活用により
機械翻訳をもっと活用できる**

「NMTは誤訳をする」という前提で開発したツール

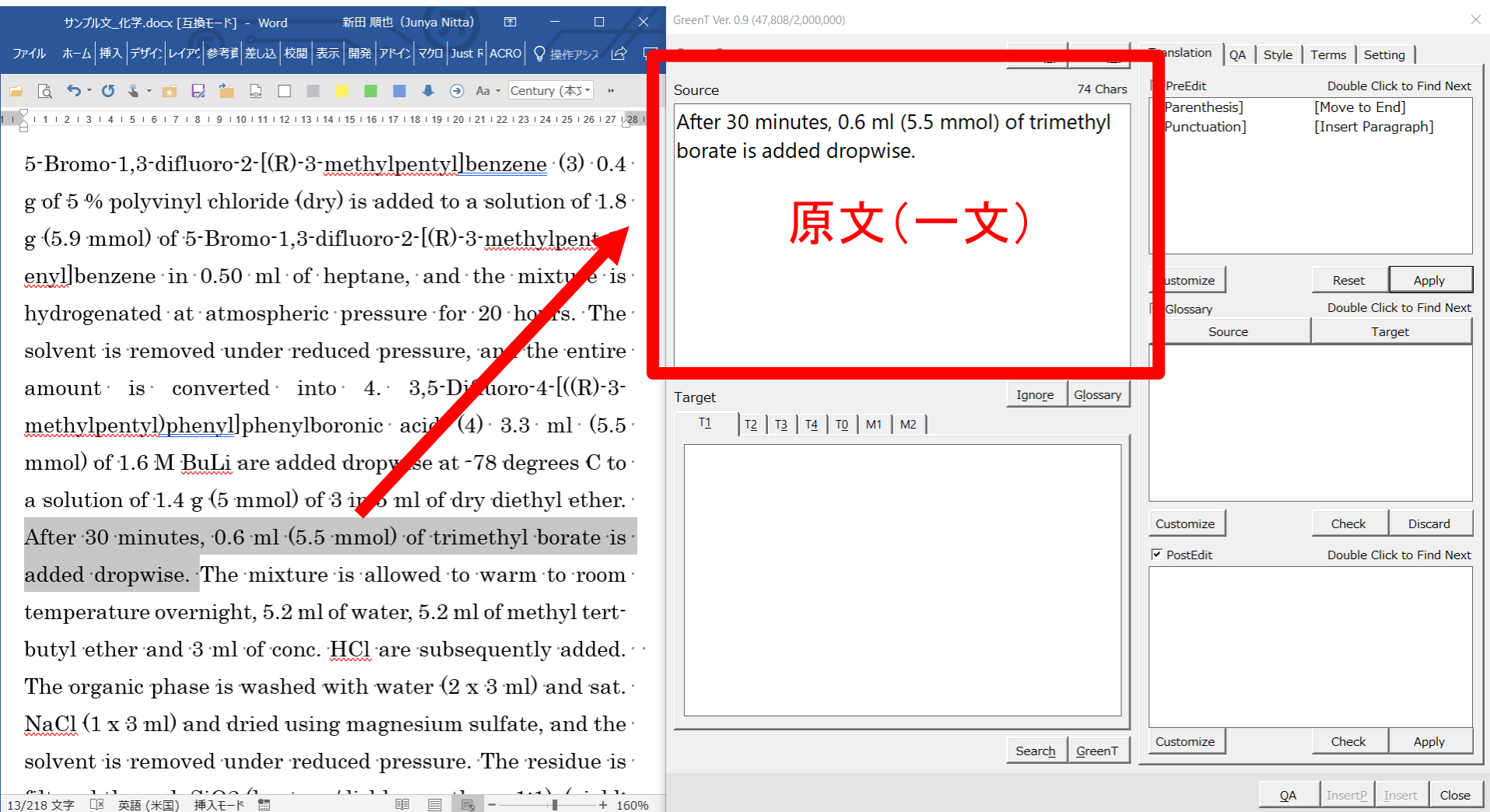
【運用方法】

- ・文章単位の機械翻訳出力を翻訳者が修正
- ・複数エンジンによる訳文を相互に比較
- ・人間翻訳用の上書き翻訳を支援

【機能】

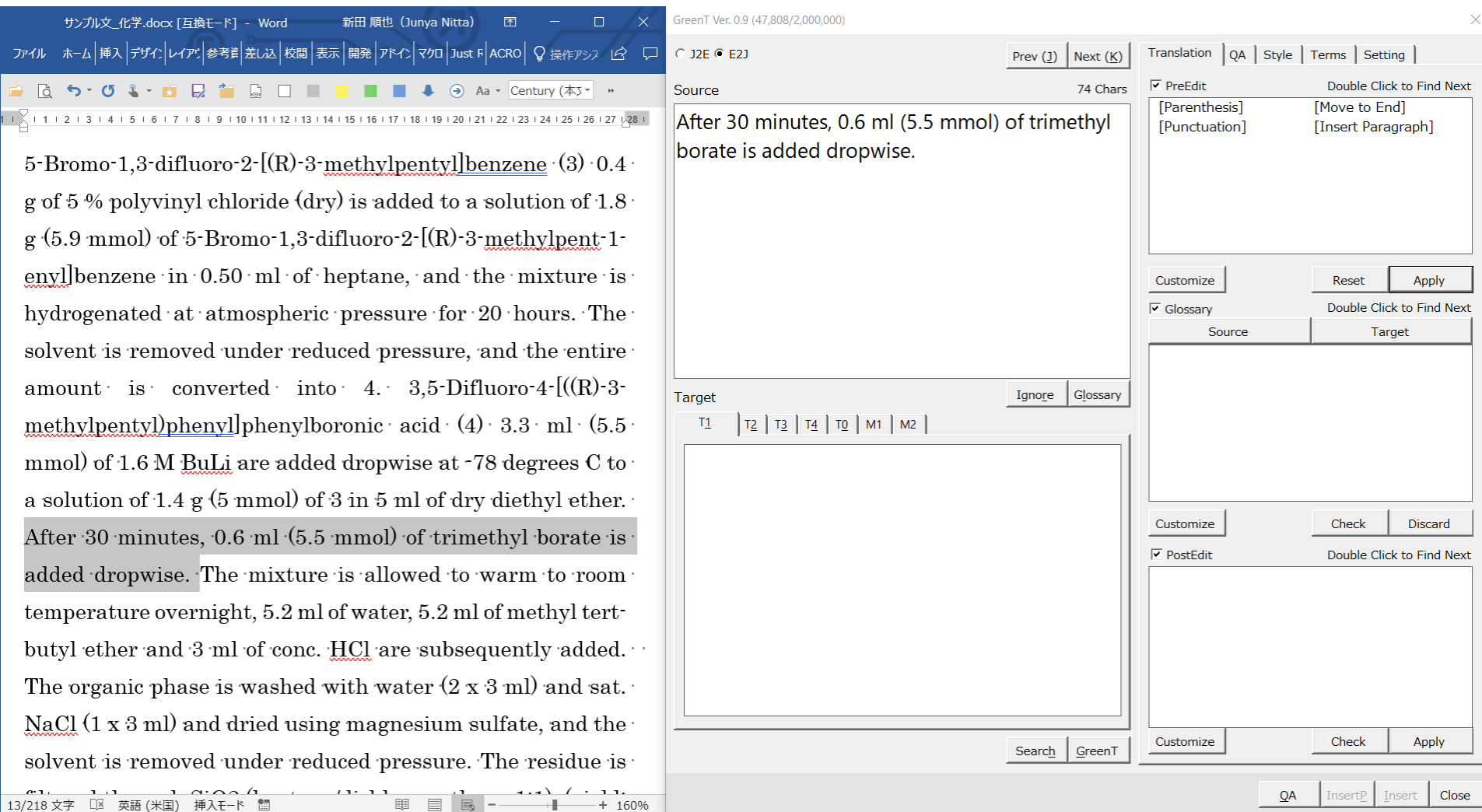
- ・プリエディットの支援(検出、修正、分解)
- ・用語の統一(専門用語、日付の表記ルール)
- ・ポストエディットの支援(訳揺れ修正、表記統一、NGワード検出)
- ・QAツールによる訳文チェック(用語、数字、初出/既出の判定)

GreenTの仕組み



The screenshot displays the GreenT software interface. On the left, a Microsoft Word document titled 'サンプル文_化学.docx' is open, showing a chemical synthesis procedure. A red arrow points from a specific sentence in the document to the GreenT window. The sentence is: 'After 30 minutes, 0.6 ml (5.5 mmol) of trimethyl borate is added dropwise.' This sentence is highlighted in the document and is also the text shown in the 'Source' field of the GreenT window. The GreenT window has a red border and contains the text '原文(一文)' (Original Text) in red. The interface includes various tabs like 'Translation', 'QA', 'Style', 'Terms', and 'Setting'. The 'Source' field is labeled '74 Chars'. The 'Target' field is empty. The 'Glossary' tab is selected, showing a list of terms: T1, T2, T3, T4, T5, T6, M1, M2. The 'Search' button is visible at the bottom of the GreenT window. The Word document shows the following text: '5-Bromo-1,3-difluoro-2-[(R)-3-methylpentyl]benzene (3) 0.4 g of 5 % polyvinyl chloride (dry) is added to a solution of 1.8 g (5.9 mmol) of 5-Bromo-1,3-difluoro-2-[(R)-3-methylpentyl]benzene in 0.50 ml of heptane, and the mixture is hydrogenated at atmospheric pressure for 20 hours. The solvent is removed under reduced pressure, and the entire amount is converted into 4. 3,5-Difluoro-4-[(R)-3-methylpentyl]phenylphenylboronic acid (4) 3.3 ml (5.5 mmol) of 1.6 M BuLi are added dropwise at -78 degrees C to a solution of 1.4 g (5 mmol) of 3 in 5 ml of dry diethyl ether. After 30 minutes, 0.6 ml (5.5 mmol) of trimethyl borate is added dropwise. The mixture is allowed to warm to room temperature overnight, 5.2 ml of water, 5.2 ml of methyl tert-butyl ether and 3 ml of conc. HCl are subsequently added. The organic phase is washed with water (2 x 3 ml) and sat. NaCl (1 x 3 ml) and dried using magnesium sulfate, and the solvent is removed under reduced pressure. The residue is'.

GreenTの仕組み



The screenshot displays the GreenT software interface, which is used for translating chemical synthesis procedures. The main window shows a Word document titled "サンプル文_化学.docx [互換モード] - Word" by 新田 順也 (Junya Nitta). The document contains a chemical synthesis procedure. The GreenT window is open, showing the "Source" text area with the following text:

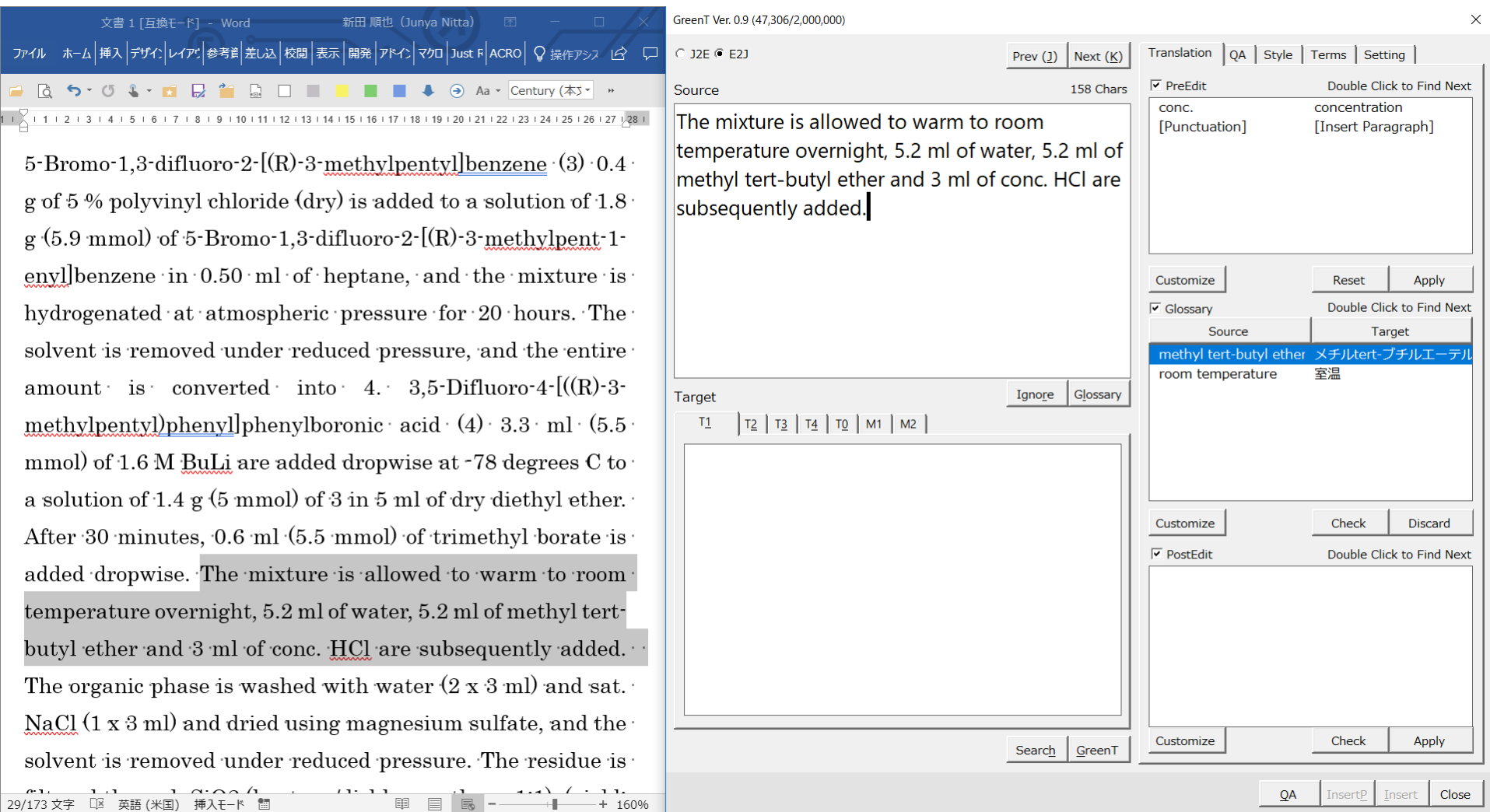
After 30 minutes, 0.6 ml (5.5 mmol) of trimethyl borate is added dropwise.

The "Target" area is empty, with a "Glossary" button visible. The right sidebar contains several panels for translation and editing:

- Translation** panel: Includes checkboxes for "PreEdit" and "Glossary", and buttons for "Customize", "Reset", "Apply", "Check", and "Discard".
- QA** panel: Includes checkboxes for "PostEdit" and "Glossary", and buttons for "Customize", "Check", and "Apply".
- Style** panel: Includes buttons for "Move to End" and "Insert Paragraph".
- Terms** panel: Includes a table with "Source" and "Target" columns.
- Setting** panel: Includes a "Double Click to Find Next" button.

The bottom status bar shows "13/218 文字" and "英語 (米国) 挿入モード". The GreenT window title bar indicates "GreenT Ver. 0.9 (47,808/2,000,000)".

GreenTの仕組み

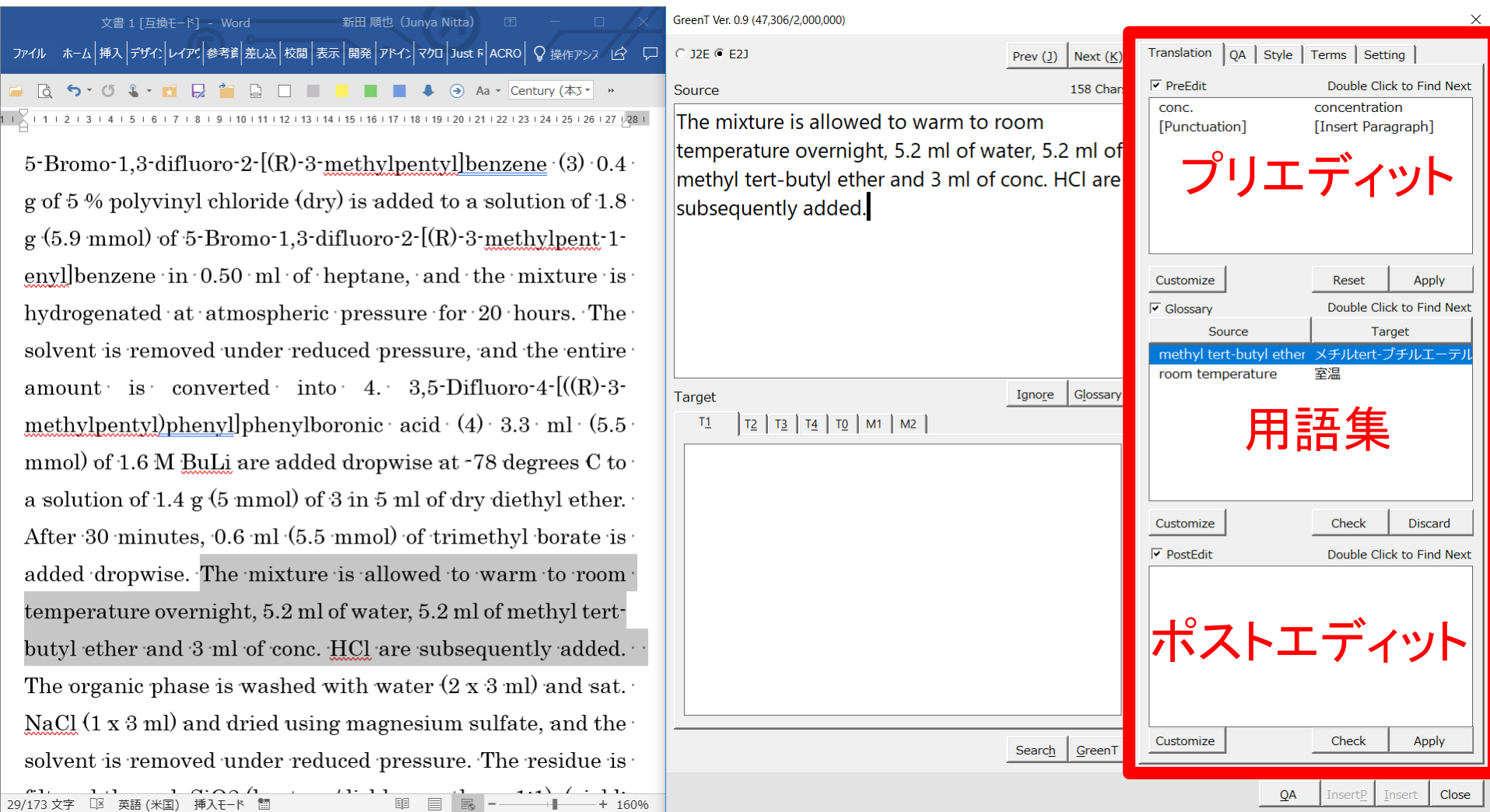


The screenshot displays the GreenT software interface, which is used for translating chemical synthesis procedures from English to Japanese. The main window shows a Microsoft Word document with a chemical synthesis procedure. The procedure involves the reaction of 5-Bromo-1,3-difluoro-2-[(R)-3-methylpentyl]benzene with 5% polyvinyl chloride (dry) in heptane, followed by hydrogenation at atmospheric pressure for 20 hours. The mixture is then allowed to warm to room temperature overnight, and 5.2 ml of water, 5.2 ml of methyl tert-butyl ether, and 3 ml of concentrated HCl are subsequently added. The organic phase is washed with water (2 x 3 ml) and saturated NaCl (1 x 3 ml), and the solvent is removed under reduced pressure. The residue is then dried using magnesium sulfate.

The GreenT window (Ver. 0.9) is open, showing the source text on the left and the target text on the right. The source text is: "The mixture is allowed to warm to room temperature overnight, 5.2 ml of water, 5.2 ml of methyl tert-butyl ether and 3 ml of conc. HCl are subsequently added." The target text is: "メチルtert-ブチルエーテル 室温". The GreenT window also includes a Glossary section with a table of terms and their translations.

Source	Target
methyl tert-butyl ether	メチルtert-ブチルエーテル
room temperature	室温

GreenTの仕組み



The screenshot displays the GreenT software interface, which is designed for chemical synthesis procedures. The main window shows a document titled "文書 1 [互換モード] - Word" by Junya Nitta. The document content is a chemical synthesis procedure for 5-Bromo-1,3-difluoro-2-[(R)-3-methylpent-1-enyl]benzene. The procedure involves adding 5-Bromo-1,3-difluoro-2-[(R)-3-methylpent-1-enyl]benzene to a solution of polyvinyl chloride (dry) in heptane, followed by hydrogenation at atmospheric pressure for 20 hours. The solvent is removed under reduced pressure, and the entire amount is converted into 4,3,5-Difluoro-4-[(R)-3-methylpent-1-enyl]phenylboronic acid. The procedure continues with the addition of 1.6 M BuLi, followed by a solution of 1.4 g (5 mmol) of 3 in 5 ml of dry diethyl ether. After 30 minutes, 0.6 ml (5.5 mmol) of trimethyl borate is added dropwise. The mixture is allowed to warm to room temperature overnight, 5.2 ml of water, 5.2 ml of methyl tert-butyl ether and 3 ml of conc. HCl are subsequently added. The organic phase is washed with water (2 x 3 ml) and sat. NaCl (1 x 3 ml) and dried using magnesium sulfate, and the solvent is removed under reduced pressure. The residue is...

The GreenT application window (Ver. 0.9) is overlaid on the right side of the document. It features a "Translation" tab and a "Glossary" tab. The "Translation" tab shows the source text and the target text, with a "Double Click to Find Next" button. The "Glossary" tab shows a list of terms and their translations, with a "Double Click to Find Next" button. The "Glossary" tab is currently active, showing the following entries:

Source	Target
methy tert-butyl ether	メチルtert-ブチルエーテル
room temperature	室温

The "Glossary" tab also includes a "Customize" button, a "Check" button, and a "Discard" button. The "PostEdit" tab is also visible, showing a "Double Click to Find Next" button and a "Customize" button.

- ・ピリオドを含む略語の復元
- ・文節区切りの誤りの修正
- ・長文を短文に修正
- ・不明確な表現の検出
- ・羅列表記の場合、コンマや読点で改行
- ・括弧内の表現やwhich節の分離

ピリオドを含む略語の復元

The mixture is allowed to warm to room temperature overnight, 5.2 ml of water, 5.2 ml of methyl tert-butyl ether and 3 ml of **conc.** **HCl** are subsequently added.

混合物を室温まで一晩温め、5.2mlの水、5.2mlのメチルtert-ブチルエーテルおよび3mlの濃塩酸を加えた。続いてHClを加える。



The mixture is allowed to warm to room temperature overnight, 5.2 ml of water, 5.2 ml of methyl tert-butyl ether and 3 ml of **concentrated** HCl are subsequently added.

混合物を室温に一晩放置してから、5.2mlの水、5.2mlのメチルtert-ブチルエーテルおよび3mlの濃HClを続いて添加する。

用語や表現の統一

- ・専門用語
- ・日付の表記ルール

色deチェック
→ Iro de Check

October 26, 2018
→ 平成30年10月26日

ポストエディット(訳文修正)の支援

- ・常体/敬体の修正案の提示
- ・誤訳箇所の修正案の提示
- ・繰り返し表現の判定
- ・肯定表現/否定表現の判定
- ・表記揺れの修正案の提示
- ・NGワードの検索

センサー
→ センサ

です
→ である

常温 常温
→ 常温

英文中に
全角文字があります

QAツールによる訳文チェック

- ・用語集の適用
- ・数字の一致
- ・初出/既出の判定

electric vehicle
は初出です

10aが
抜けています

「噴射ノズル」の
訳語が違います

GreenTの仕組み解説

GreenTの仕組み

○ J2E ● E2J

Prev (J) Next (K)

Translation QA Style Terms Setting

Source 158 Chars

The mixture is allowed to warm to room temperature overnight, 5.2 ml of water, 5.2 ml of methyl tert-butyl ether and 3 ml of **conc.** HCl are subsequently added.

Target

T1 T2 T3 T4 T0 M1 M2

Ignore Glossary

179

Search GreenT

☒ PreEdit Double Click to Find Next

conc.	concentrated
conc.	concentration
[Punctuation]	[Insert Paragraph]

プリエディット

Customize Reset Apply

☒ Glossary Double Click to Find Next

Source	Target
methyl tert-butyl ether	メチルtert-ブチルエーテル
room temperature	室温

Customize Check Discard

☒ PostEdit Double Click to Find Next

Customize Check Apply

GreenTの仕組み

☐ J2E
 ☒ E2J

Prev (J)
 Next (K)

Source
 158 Chars

The mixture is allowed to warm to room temperature overnight, 5.2 ml of water, 5.2 ml of methyl tert butyl ether and 3 ml of concentrated HCl are subsequently added.

Target
 Ignore Glossary

T1 T2 T3 T4 T0 M1 M2

180

Search GreenT

Translation QA Style Terms Setting

☒ PreEdit
 Double Click to Find Next

[Punctuation] [Insert Paragraph]

Customize Reset Apply

☒ Glossary
 Double Click to Find Next

Source Target

methyl tert-butyl ether メチルtert-ブチルエーテル
 room temperature 室温

Customize Check Discard

☒ PostEdit
 Double Click to Find Next

Customize Check Apply

GreenTの仕組み

☐ J2E ☒ E2J

Prev (J) Next (K)

Source 158 Chars

The mixture is allowed to warm to room temperature overnight, 5.2 ml of water, 5.2 ml of methyl tert-butyl ether and 3 ml of concentrated HCl are subsequently added.

Target Ignore Glossary

T1 T2 T3 T4 T0 M1 M2

181

Search GreenT

Translation QA Style Terms Setting

☒ PreEdit Double Click to Find Next

[Punctuation] [Insert Paragraph]

Customize Reset Apply

☒ Glossary Double Click to Find Next

Source	Target
methyl tert-butyl ether	メチルtert-ブチルエーテル
room temperature	室温

用語集

Customize Check Discard

Customize Check Apply

GreenTの仕組み

☐ J2E ☒ E2J
 Prev (J) Next (K)

Source 158 Chars

The mixture is allowed to warm to room temperature overnight, 5.2 ml of water, 5.2 ml of methyl tert-butyl ether and 3 ml of concentrated HCl are subsequently added.

Target
 Ignore Glossary

T1 T2 T3 T4 T0 M1 M2

182

Translation QA Style Terms Setting

☒ PreEdit Double Click to Find Next

[Punctuation] [Insert Paragraph]

Customize Reset Apply

☒ Glossary Double Click to Find Next

Source	Target
methyl tert-butyl ether	メチルtert-ブチルエーテル
room temperature	室温

Customize Check Discard

☒ PostEdit Double Click to Find Next

Customize Check Apply

Search GreenT

諷文



GreenTの仕組み

☐ J2E
 ☒ E2J

Prev (J)

Next (K)

Source

158 Chars

The mixture is allowed to warm to room temperature overnight, 5.2 ml of water, 5.2 ml of methyl tert-butyl ether and 3 ml of concentrated HCl are subsequently added.

Target

Ignore Glossary

T1 T2 T3 T4 T0 M1 M2

合成物を一晩室温に戻し、5.2mlの水、5.2mlのメチルtert-ブチルエーテルおよび3mlの濃HClを続いて添加する。

Translation

QA

Style

Terms

Setting

☒ PreEdit
 Double Click to Find Next

[Punctuation]

[Insert Paragraph]

Customize

Reset

Apply

☒ Glossary
 Double Click to Find Next

Source

Target

methyl tert-butyl ether

メチルtert-ブチルエーテル

room temperature

室温

Customize

Check

Discard

☒ PostEdit
 Double Click to Find Next

合成物

混合物

Customize

Check

Apply

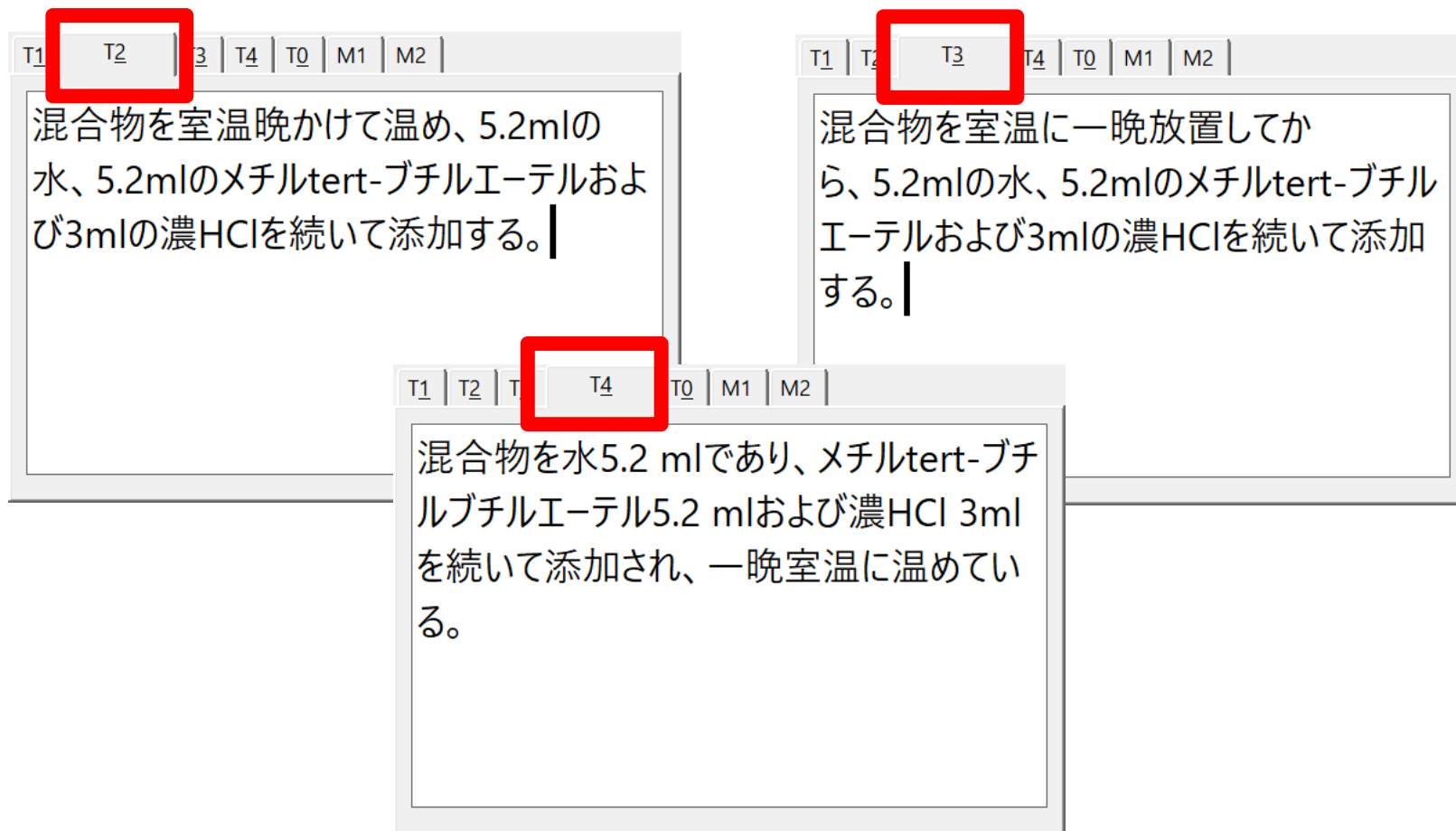
185

Search GreenT

修正対象

修正済

GreenTの仕組み



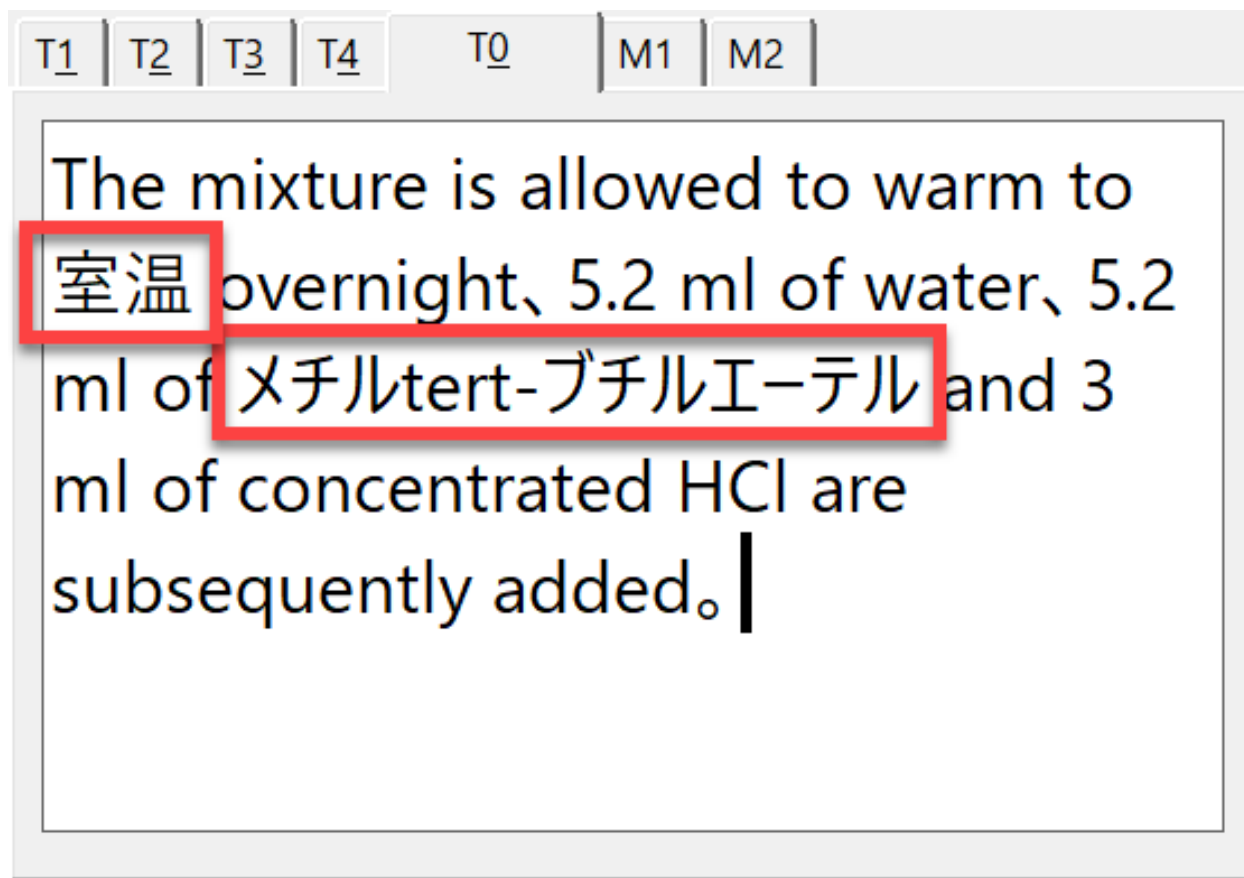
異なる訳文を提案可能

GreenTの仕組み

T1	T2	T3	T4	T0	M1	M2
混合物を室温に一晩放置してから、5. 2 m l の水、5. 2 m l のメチル t e r t – ブチルエーテルおよび 3 m l の濃 H C l を続いて添加する。						

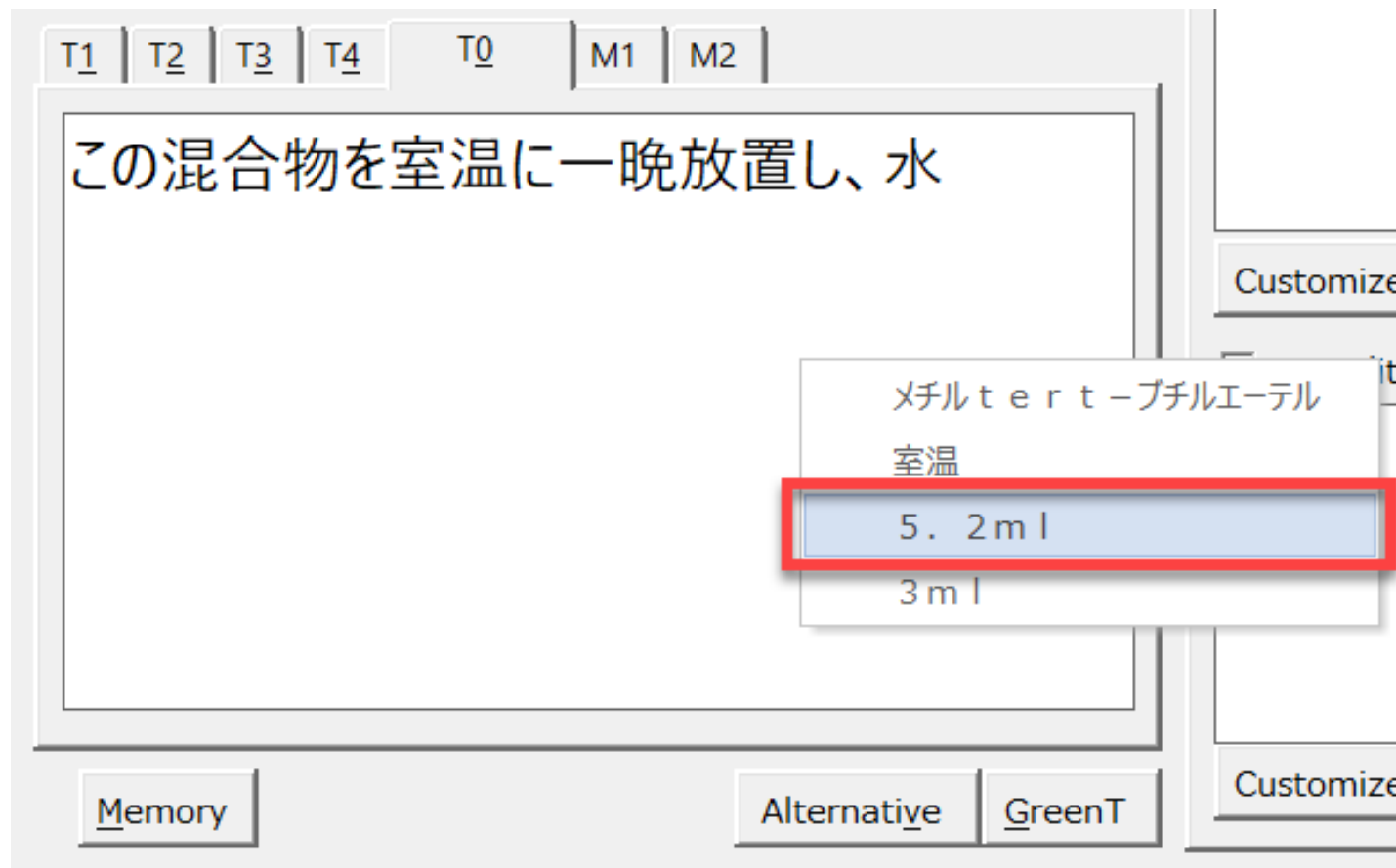
英数字全角で出力可能

GreenTの仕組み



用語集で置換→翻訳者が翻訳

GreenTの仕組み



用語集と数字をカーソル位置に挿入
全角化も可能

GreenTの仕組み

○ J2E ● E2J

Prev (J) Next (K)

Source 158 Chars

The mixture is allowed to warm to room temperature overnight, 5.2 ml of water, 5.2 ml of methyl tert-butyl ether and 3 ml of concentrated HCl are subsequently added.

Target Ignore Glossary

1 2 3 4 5 6 7 8 9 10 11 12

混合物を室温まで一晩放置した後に、5.2 mlの水、5.2 mlのメチルtert-ブチルエーテルおよび3 mlの濃塩酸を加える。

最終的な訳文

191

Translation QA Style Terms Setting

☒ PreEdit Double Click to Find Next

[Punctuation] [Insert Paragraph]

Customize Reset Apply

☒ Glossary Double Click to Find Next

Source	Target
methyl tert-butyl ether	メチルtert-ブチルエーテル
room temperature	室温

Customize Check Discard

☒ PostEdit Double Click to Find Next

Customize Check Apply

Search GreenT

GreenTの仕組み

QA実施

☐ J2E
 ☒ E2J

Prev (J)
 Next (K)

Translation
 QA
 Style
 Terms
 Setting

Source158 Chars

The mixture is allowed to warm to room temperature overnight, 5.2 ml of water, 5.2 ml of methyl tert-butyl ether and 3 ml of concentrated HCl are subsequently added.

Target

Ignore

Glossary

T1

T2

T3

T4

T0

M1

M2

混合物を室温まで一晩放置した後に、5.2 mlの水、5.2 mlのメチルtert-ブチルエーテルおよび3 mlの濃塩酸を加える。

Glossary Check

☐ methyl tert-butyl et メチルtert-ブチルエーテル
 ☐ room temperature 室温

Number Check

	Source	Target
☐ 5.2ml	2	2
☐ 3ml	1	1

Antecedent Check

GreenTの仕組み

○ J2E ● E2J Prev (J) Next (K) Translation OA Style Terms Setting

Source 158 Chars

The mixture is allowed to warm to room temperature overnight, 5.2 ml of water, 5.2 ml of methyl tert-butyl ether and 3 ml of concentrated HCl are subsequently added.

Target Ignore Glossary

T1 | T2 | T3 | T4 | T5 | M1 | M2

混合物を室温まで一晩放置した後に、5.2 mlの水、5.2 mlのメチルtert-ブチルエーテルおよび3 mlの濃塩酸を加える。

193

Glossary Check

- ☐ methyl tert-butyl et メチルtert-ブチルエーテル
- ☐ room temperature 室温

用語集

Number Check Source Target

☐ 5.2ml	2	2
☐ 3ml	1	1

Antecedent Check

Search GreenT

GreenTの仕組み

○ J2E ● E2J

Prev (J) Next (K)

Translation QA Style Terms Setting

Source 158 Chars

The mixture is allowed to warm to room temperature overnight, 5.2 ml of water, 5.2 ml of methyl tert-butyl ether and 3 ml of concentrated HCl are subsequently added.

Target

Ignore Glossary

T1 T2 T3 T4 T5 M1 M2

混合物を室温まで一晩放置した後に、5.2 mlの水、5.2 mlのメチルtert-ブチルエーテルおよび3 mlの濃塩酸を加える。

194

Search GreenT

Glossary Check

- ☐ methyl tert-butyl et メチルtert-ブチルエーテル
- ☐ room temperature 室温

Number Check

	Source	Target
☐ 5.2ml	2	2
☐ 3ml	1	1

英数字記号

Antecedent Check

ツールの活用により、

原文の修正（プリエディット）がしやすくなる
訳文の修正（ポストエディット）がしやすくなる
プリエディット/ポストエディットのノウハウを蓄積できる

用語を統一しやすくなる
英数字記号の誤記を発見しやすくなる
和文特許明細書での半角文字がなくなる
訳文の破綻が減り、編集しやすい訳文が増える

まとめ

機械翻訳が困難な文章は人間が翻訳すればよい

今後の予定

ツールの公開

プリエディット/ポストエディットの検出/修正の精度向上

各種翻訳エンジンの連携強化

TMツールとの連携

ご清聴ありがとうございました

特定非営利活動法人(NPO)
日本知的財産翻訳協会(NIPTA)
<https://www.nipta.org/>
bureau@nipta.org

吉川 潔 ２０１８年１２月７日（金）

- ① はじめに
- ② 特許関連資料を翻訳するときの注意事項
- ③ 図面を参照して、特許関連資料を翻訳
機械翻訳（MT）の弱点
- ④ MTを特許関連資料の翻訳に効果的に利用する事例
- ⑤ インターネットから利用できるMTを
技術系（特に特許）資料で試訳した結果（２６０文例）
- ⑥ まとめと今後の方針

- ① はじめに
 - ◆ 発表者は、新潟で、都内の特許事務所から原稿をメールで受信し、翻訳後に返信という仕事を３０数年間、テレワーク（SOHO）で働いてきた。
 - ◆ その間に、国際出願で高名な弁理士から指導を受けた。
その多くは、学校の教科書や翻訳用の参考書に非記載である。
 - ◆ それらが、機械翻訳（MT）の大きな弱点になる。
AAMTが翻訳の現場から遊離と感じているので、
この発表を翻訳者からのノウハウとして役立ててほしい。
 - ◆ 特許情報シンポジウムにおける、発表者の過去の内容
2010年 第一回 翻訳ソフトの実用化のための提案
2012年 第二回 基本的な誤訳例の対策を提案
2014年 第三回 MTを優れた辞書として活用する方法
2016年 第四回 特許明細書の翻訳の注意事項－１
2018年 第五回 々 々 々 ー２
(翻訳者からのノウハウ)

今回の発表は、第四回の「特許明細書の翻訳の注意事項」の続編である。
そこで、今回の発表の一部が前回と重複している。
第一回と第二回で翻訳ソフトの誤訳を指摘した。新規のニューラル翻訳では大半が解決しているが、⑤で指摘するように、新たな誤訳を誘発している。

- ② 特許関連資料を翻訳するときの注意事項
日本人は単数と複数に鈍感だが、アングロサクソンは敏感である。日本語が曖昧でも、技術的背景を考慮して、単数と複数を区別して英訳することが要求される。

- ②－１ 単数と複数の区別を正確に英訳
(カメラとタブレットは電池を備えている)を英訳
A camera and a tablet contain a battery.
一つのカメラと一つのタブレットが一つの電池を共有 ？
A camera and a tablet contain batteries.
一つのカメラと一つのタブレットが幾つかの電池を共有 ？
Each of a camera and a tablet contains a battery
一つのカメラと一つのタブレットの各々が一つの電池を持つ
三段目が日本語原稿を正確に訳すと想定すると、
日本語原稿にない「each of」を追記しなければならない

②-2 特許明細書の請求項

下記の日本語をインターネットから利用できる、ニューラル翻訳の
三つの機械翻訳 (MT) を用いて、試訳したら一つだけ正訳だった。

請求項 1 に係る発明と請求項 1 を引用する請求項 5 に係る発明は不明確

- The invention according to claim 1 and
the invention according to claim 5 citing claim 1 are unclear

次に、下記の日本語を、上記の○印の MT で試訳したら、△である

請求項 1 に係る発明と請求項 1 を引用する請求項 2～9 に係る発明は不明確

- △ The invention according to claim 1 and the invention
according to claims 2 to 9 citing claim 1 are unclear

請求項 2～9 が、個々に請求項 1 を引用するのか、

全体的に請求項 1 を引用するのか不明瞭である。

「請求項 1 を各々が引用する請求項 2～9 に係る発明」が本来の意味。

不足語句 (each of which) を補充して正確に英訳しなければならない

- The invention according to claim 1 and the inventions
according to claims 2 to 9 each of which cites claim 1 are unclear

②-3 「and/or」 (及び/又は)

「それは、炭素繊維やガラス繊維から作られたシートである」を英訳

It is a sheet made of carbon fiber and glass fiber.

「炭素繊維とガラス繊維の両方から作られたシート」の意味

It is a sheet made of carbon fiber or glass fiber.

「炭素繊維またはガラス繊維の一方から作られたシート」の意味

It is a sheet made of carbon fiber and/or glass fiber.

「炭素繊維及び又はガラス繊維から作られたシート」の意味

(両方でも一方でも可)

「炭素繊維またはガラス繊維あるいはその両方から作られたシート」

日本語原文からは、上記の三つのどれか分からないので、前後関係から推定して訳す

②-4 誤記を訂正して訳す

「トランジスタの順方向電圧は 4.0 V である」を英訳する時に、

4.0 V の誤記を訂正して英訳する。

4.0 V → 0.4 V

A forward voltage of a transistor is 0.4 V.

トランジスタの順方向電圧は 0.4 V は、半導体分野の基礎知識である。

翻訳の分野は、理工系～法文系まで様々だが、各々の分野で原稿ミスに気付いて、
それを訂正して翻訳できる専門的な基礎知識が必要である。

②-5 分詞構文の主語 分詞構文の主語が省略してある場合に、

本文の主語が分詞構文の主語になるのが英文法

一例として、下記に注意

When travelling in Hokkaido, snow still remained. →

雪が北海道に旅行したら、雪が残っていた

When I travelled in Hokkaido, snow still remained. →
私が北海道に旅行したら、雪が残っていた

機械翻訳 (MT) を利用する時も、日本語原稿の主語に注意！

私が手紙を航空便でハワイに送ると、いつごろハワイに到着しますか

When I send a letter to Hawaii by airmail, when will I arrive at Hawaii ?
いつごろ私はハワイに到着しますか？

私が手紙を航空便でハワイに送ると、いつごろ手紙はハワイに到着しますか

When I send a letter by airmail to Hawaii,
when will the letter arrive at Hawaii ?

日本語原稿で主語が明確に記してあれば、機械翻訳 (MT) は正訳する

②-6 数量表現

競技者は30数人です 三つ共 × There are 30 players.
正訳 Players are 30-odd persons.

試験の点は60点台である

二つが × The point of the exam is 60 points.
一つが ○ The test point is a 60-point stand.

環境は1990年代の初頭以来、大きく変わった

三つが正訳： Environment has significantly changed since the early 1990s.

ニューラル翻訳は基本的な数量表現の誤訳は減少したが、微細部で誤訳がある。
技術系資料や特許関連の翻訳は、正確な数量表現が要求される。
数量表現は様々であるが、辞書や参考書を総動員すると対応できる。
対応可能なことは、数量表現の範囲は無限大でなく、有限を意味する。
それらを可能な限り集めて、MTに記憶させるべきである

③ 図面を参照して、特許関連資料を翻訳 → 機械翻訳 (MT) の弱点

「スイッチC、Dがノードと抵抗A、B間に接続」を英訳するときに、
配置関係の記述が不鮮明なので、正確に翻訳できない。

図面を見て、

「スイッチCがノードと抵抗Aの間」「スイッチDがノードと抵抗Bの間」
に接続していることを確認して英訳する。

すなわち

A switch C is connected between a node and a resistor A, and
a switch D is connected between the node and a resistor B.

翻訳者は、クライアントから翻訳原稿を受け取る時に図面も受け取る。

翻訳原稿と図面を照合しながら、翻訳原稿の不備を補って翻訳する。

図面をOCR (光学式読取装置) で参照しながら、原稿の不備を補って
翻訳することは、現状のMTは難しい。MTの弱点である。

④ 機械翻訳 (MT) を特許関連の翻訳に効果的に利用する事例

原稿の意味が分からない又は表現が難しい和文英訳のヒントの入手に役立つ

「これは法令で読み替えて準用される - - -」が原稿に現れ当惑した。

(～ を読み替えて準用 ～) は法曹分野では常套句らしいが、門外漢は？

「これは法令で読み替えて準用される」をMTに入力したら、
This is applied mutatis mutandis in Law と出力
使用数をインターネットの検索機能で調べてから、英文に用いた
「~ is applied mutatis mutandis in ~」の使用例は 16400

このように、苦手分野の慣用句や常套句を含んだ一文を
MTに入力すると、MTの出力からヒントを入手できて、
使用数もインターネットから調査できるので非常に助かる
MTは各分野の専門事項に弱い、苦手の分野で「お助けマン」になる。
MTは各分野の翻訳で70点のオールマイティである

⑤ インターネット上の三つのMTを利用して

技術系（特に特許）資料を試訳した結果（260文例）
本日の主題は「特許翻訳の注意事項」なので、MT試訳は詳述しません。
前編と後編あわせてA4・38頁を、希望者に電子メールで送ります
主な内容

前編（27頁）：

MTに期待できないこと、期待できること
インターネット上で利用できる三つのMTの試訳結果（170文例）
パソコンを用いる地域資料の多国語翻訳
スマホを用いる多国語の通訳とプロセス
筆者（発表者）の翻訳歴とMTとのかかわり

後編（11頁）：

MTの効果的な活用方法
インターネット上で利用できる三つのMTの試訳（続編 90文例）
連語（複数の名詞や形容詞が連なる語句）で部分的な訳抜けが多い
例えば、 It is reversely easy to make a sample.
× サンプルを作るのは逆です。
○ サンプルを作るのは逆に簡単です

⑥ まとめと今後の方針

今回の発表は、30数年の翻訳実務のなかで、翻訳会社や特許事務所
から注意され指導を受けた注意事項を要約した発表です。
クライアントは、翻訳者が原稿のミスを訂正して翻訳することを要求します。
翻訳者は、専門分野の知識を豊富にして対処しなければなりません。

同様に、MTは、翻訳原稿のミスを訂正して翻訳しなければなりません。
専門知識を背景にした翻訳や、図面を参照しながらの翻訳は、
AI（人工知能）に準じたニューラル翻訳といえども、現状では難しい。

MTは個別の分野の専門知識に弱い、全分野で70点のオールマイティです。
MTに翻訳原稿の丸投げの全面依存は無理です。
マイナスの部分を補充してMTを活用すると、「お助けマン」になります。

普通の人（学校を卒業してから英語に無縁の人）がMTを利用するときに、
誤訳に出会うと、利用を止めて、食わず嫌いになりがちです。
普通の人々が、上記の欠陥（マイナス30点の部分）を認識したうえで、

MT を有意義に利用できるように、使用方法を説明すべきです。

MTを現場の翻訳実務で試訳して発信する翻訳者が少ないです。

そのために、改善に必要な実験（試訳）データが不足しています。

これは、理論先行で実験が伴わないことになります。

私は、ピーク時に比し翻訳量は減少ですが、実務で気付いた文例を試訳し、特に数量表現に関係した語句を集約し、MTの研究者や技術者に提供します。

人生100年ならば、残の25年のライフワークとして、

機械翻訳（MT）の実用化を手伝う意欲に燃えています。

◎ 追記： 機械翻訳（MT）の開発に貢献した桜井督三氏の紹介

30数年前、八王子市のIBS社の桜井督三社長と知り合い翻訳を始めた。

同氏は、翻訳原稿をメールで送受信する全国規模の翻訳網を日本で最初に構築し、SOHO、テレワークという労働形態を実践し普及させた功労者である。

初期のMTを社内に設置し、各社の試訳依頼に協力した。同社のMTを頂点に、外部の翻訳関係者が利用できるシステム、クラウドを発案し、関係者の見学希望に応じていた。27年前のAAMTの設立総会で、翻訳業界を代表して挨拶した。

しかし、同社の経営は苦しく、破産し、桜井氏は行方不明になった。

私は、同氏のおかげで翻訳者として自立できた。14年前の新潟県の水害や中越地震も、在宅翻訳のおかげで克服できた。その恩に報いるために、同氏の夢だったMTの実用化を願い、MTを試訳してきた。それがMTの実用化に役立つことを願っている。

第5回特許情報シンポジウム

平成30年12月7日

請求項の記載を 構造化する提案

弁理士 博士 小池誠

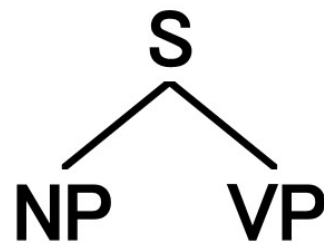
ヒューマンインタフェース

- 2つのアプローチ
 - 1. 機械が人間に合わせる
 - ✓ 機械が人間に分かる言語を使う
 - ✓ 人工知能、自然言語処理
 - 2. 人間が機械に合わせる
 - ✓ 人間が機械に分かる言語を使う
 - ✓ 本発表のアプローチ

機械翻訳

- トランスファー方式
- 実例型機械翻訳
- 統計的機械翻訳
- ニューラル機械翻訳

構文解析



機械翻訳の課題:特許文献

- 特許文献のなかで請求項が最も重要
 - ✓ 権利範囲確定(特許法70条1項)
 - ✓ 審査対象(リパーゼ事件最高裁判決)
- 請求項の翻訳に求められる品質は特に高い

機械翻訳の課題: 請求項

- 請求項は一文で記載するので、一文が長い！（特許法36条5項）
- 現状の人工知能では、構文解析の性能が必ずしも十分でない
- 構文解析のエラーは機械翻訳において翻訳ミスに発展

特許文献の特徴

- 構造化されている；墨付きカッコ【】
- ✓【書類名】【0001】【請求項1】
- ✓特許法施行規則23条、24条、24条の4等
- 人間が機械に合わせる
- 電子出願、インターネット出願に対応
- ✓工業所有権に関する手続等の特例に関する法律

提案1;セミコロン

- 各請求項は、構成要件に分けて記載する
- 個々の構成要件が終了する記号として
セミコロン;を用いる
- 機械が、請求項中のセミコロンを認識する
- ✓米国特許法スタイル

物の発明

- 【請求項1】 要素1と;
要素2と;
を含む装置。
- セミコロンは機械の認識のため
 - ✓構文解析:**NP1\NP2**
 - 改行は人間の認識のため

方法の発明

- 【請求項1】○を動詞1する工程と；
△を動詞2する工程と；
を含む方法
- セミコロンは機械の認識のため
 - ✓ 構文解析：VP1∧VP2
 - 改行は人間の認識のため

請求項の構文解析

- NP1∧NP2 VP1∧VP2
- NP_i(*i*=1 or 2) 又はVP_i(*i*=1 or 2)について構文解析するのは、クレーム全体の構文解析より遥かに容易
- 構文解析のエラーが減る

提案2;ジェプソン

- 請求項中の『特徴』という用語は、通常、ジェプソンクレームを意味する
 - 請求項中の『特徴』という用語は、構成要件が終了する記号と認識
 - 請求項中の『特徴』という用語は、セミコロンと同様に扱う
- ✓ヨーロッパ特許法スタイル

物の発明;例示

【請求項1】 要素1と;
要素2と;
を含む装置において、
要素3を更に含むことを**特徴**とする装置。
構文解析;**NP1\NP2\NP3**

方法の発明;例示

【請求項1】 動詞1する工程と;

動詞2する工程と;

を含む方法において、

動詞3する工程を更に含むことを**特徴**とする方法。

構文解析;**VP1^VP2^VP3**

提案3;ジェプソン

- 請求項中に『特徴』という用語が使われている場合、同じ請求項中に装置、方法など前庭部となる用語が2回、使われていることを確認する
- ジェプソンクレームであることを機械で確認

セミicolon違反

- 機械で構文解析する便宜を図る趣旨
- 方式違反に過ぎない
- 立法論としては、先行技術文献情報開示要件(特許法48条の7、49条5号)と同様な取り扱いが適切では
- 拒絶理由であるが、特許異議理由、特許無効理由でない

まとめ

- 請求項の記載において、セミicolon『;』『特徴』を区切り記号と認識する
- ✓ 請求項の記載方式を僅かに改正するだけで大きなインパクト！
- ✓ 請求項の構文解析が格段に向上し、機械翻訳の精度も向上する
- 請求項の記載方式の国際的協調

用語集による原文の一括置換と 正規表現による一括並べ換えを用いた 特許明細書の高品質翻訳

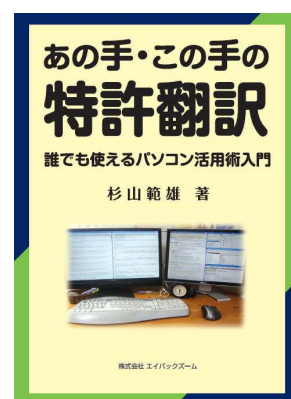
杉山範雄
(株)杉山特許翻訳

2018年12月7日
第5回特許情報シンポジウム



略歴

- 電気メーカー・特許事務所に勤務
- 2002年：フリーランス特許翻訳者（電気／電子／材料）
- 2011年：「あの手・この手の特許翻訳」出版
- 2016年：(株)杉山特許翻訳 設立
- 日本知的財産翻訳検定1級試験委員
- 工学博士



機械翻訳の現状

翻訳ミスの発生

1. 訳抜け
2. 訳ゆれ
3. 数字ミス etc.

⇒ ポストエディットの負担が大きい



ノンタイピング人間翻訳のご紹介

翻訳手順（タイピングは最少限）

1. 原文の一括置換（用語集の利用）
2. 置換語の一括並べ換え（正規表現の利用）

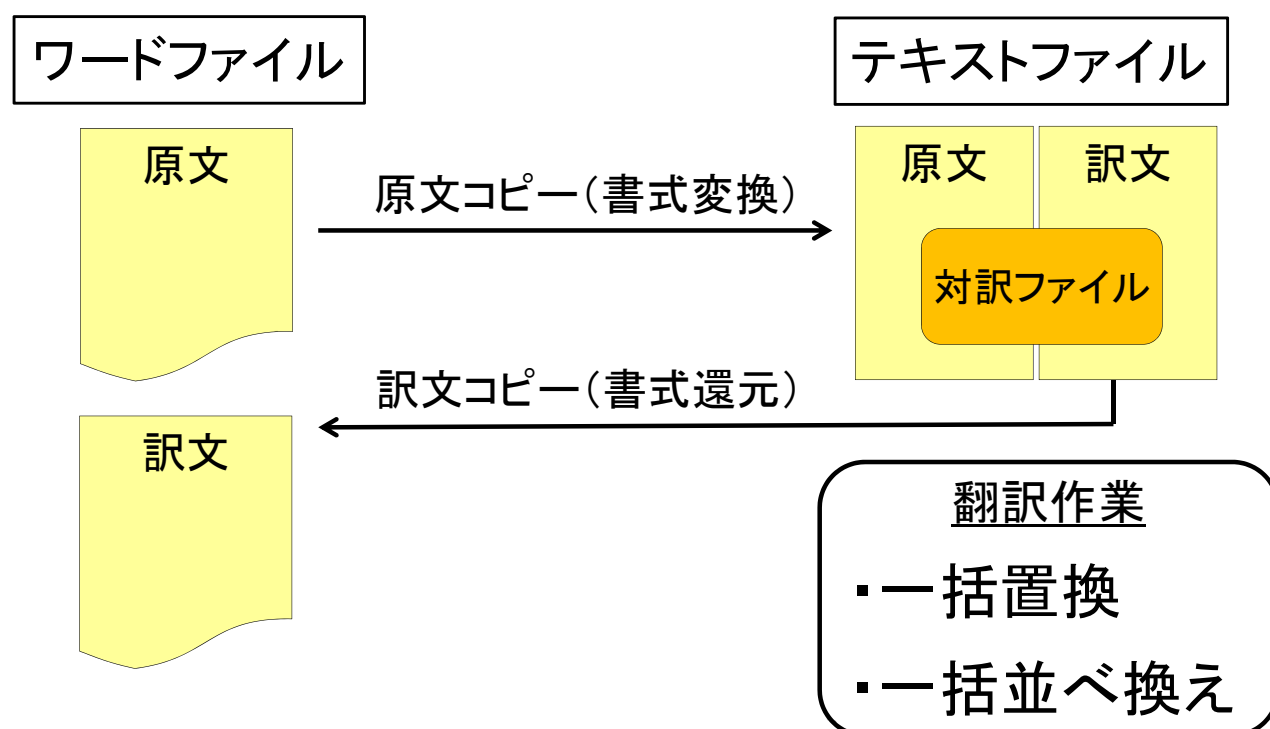
効果

1. 原文情報を直接利用 ⇒ 訳ぬけ・数字ミスなし
2. 用語を一括置換 ⇒ 用語が統一、訳ゆれなし
3. 用語を一括並べ換え ⇒ 訳文が統一、訳ゆれなし
4. 並換えを正規表現で ⇒ 数字（番号、符号など）が違う訳文も統一

⇒ 翻訳ミス（タイピング由来）が基本的に皆無



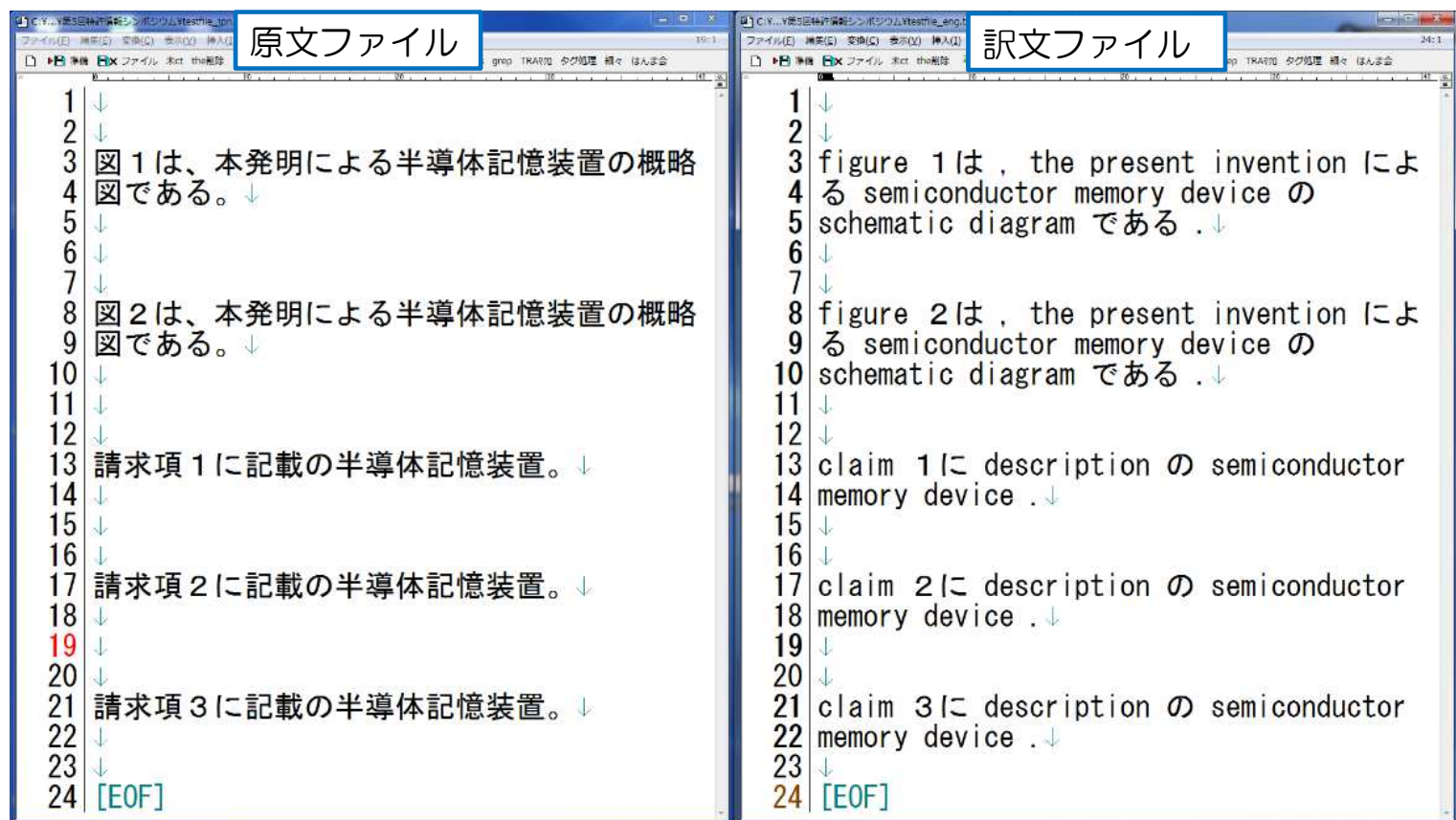
翻訳フロー（秀丸エディタ+マクロ）



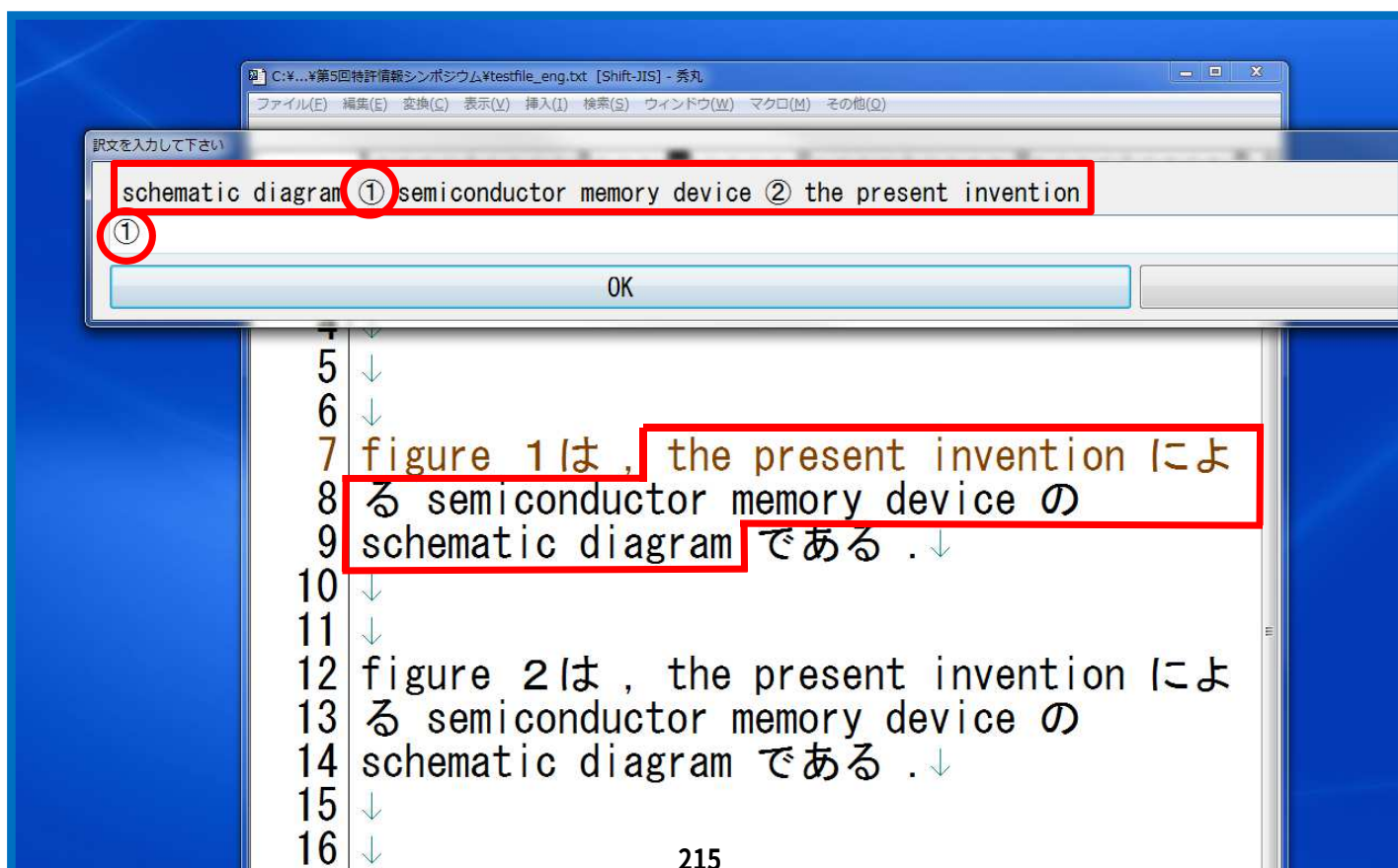
置換用用語集(日英)

3136	演算時間	operating time↓
3137	演算増幅器	operational amplifier↓
3138	演算動作	arithmetic operation↓
3139	演算部	computing portion↓
3140	縁部	edge↓
3141	遠い側	far side↓
3142	遠隔操作式	remote controlled↓
3143	遠心式	centrifugal↓
3144	遠心分離器	centrifuge↓
3145	遠心力	centrifugal force↓
3146	鉛直方向	vertical direction↓
3147	鉛直方向の成分	vertical component↓
3148	塩化アルミニウム	aluminum chloride↓
3149	塩化クロミル	chromyl chloride↓
3150	塩化クロム	chromium chloride↓
3151	塩化クロム溶液	chromium chloride solution↓
3152	塩化亜鉛	zinc chloride↓

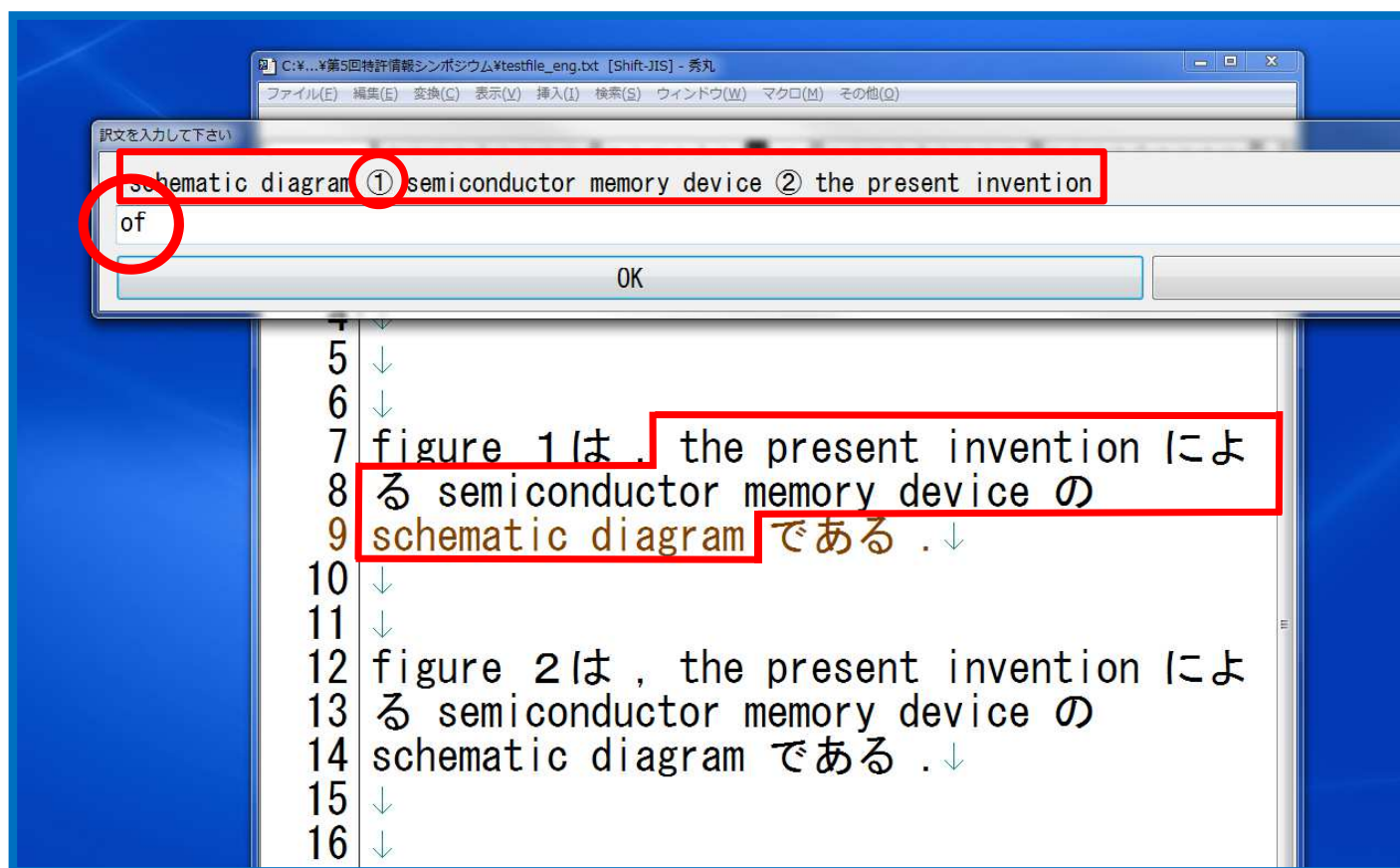
原文の一括置換後（置換ソフト使用）



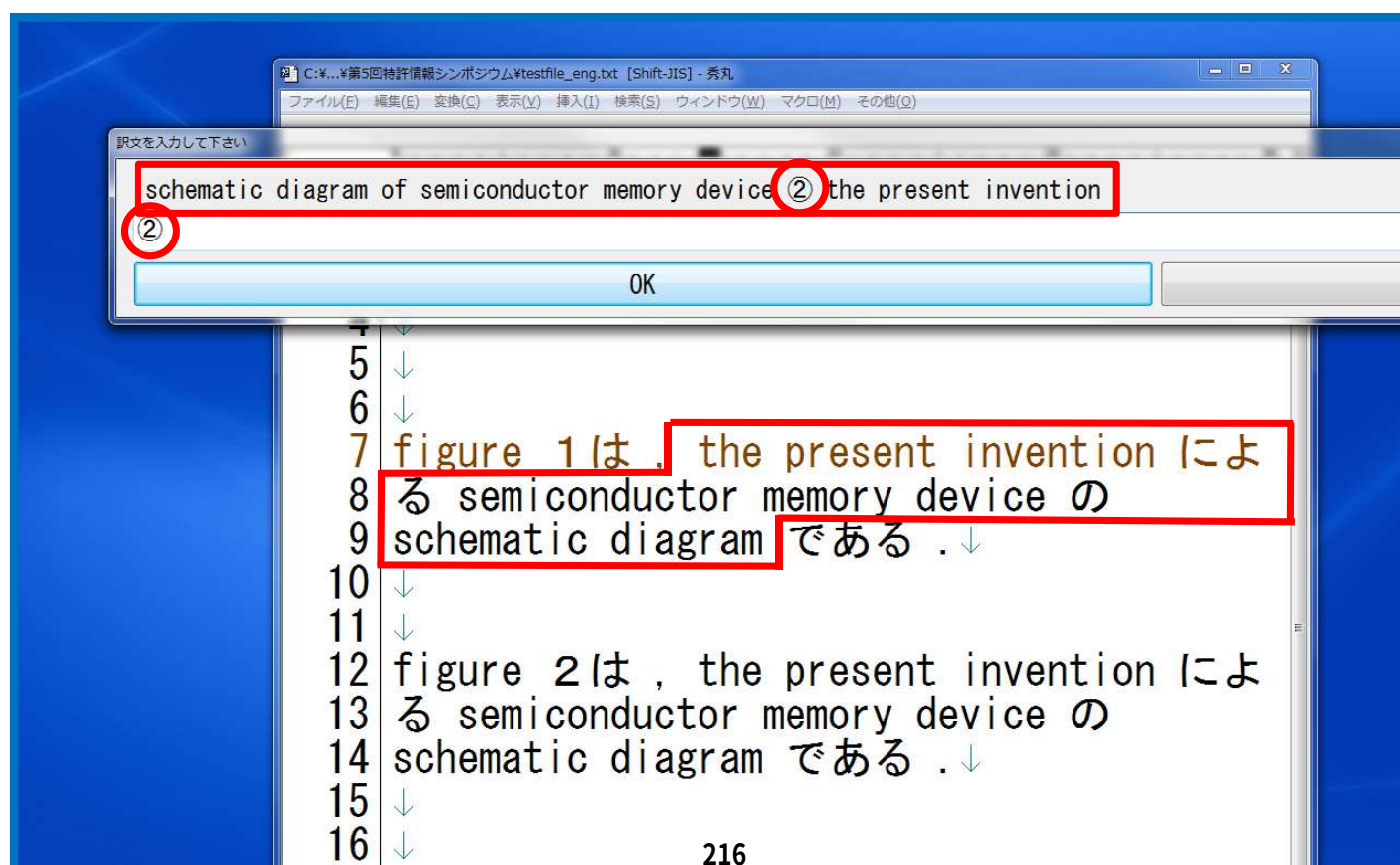
置換語の一括並べ換えマクロ（1）：範囲選択(順番反転)



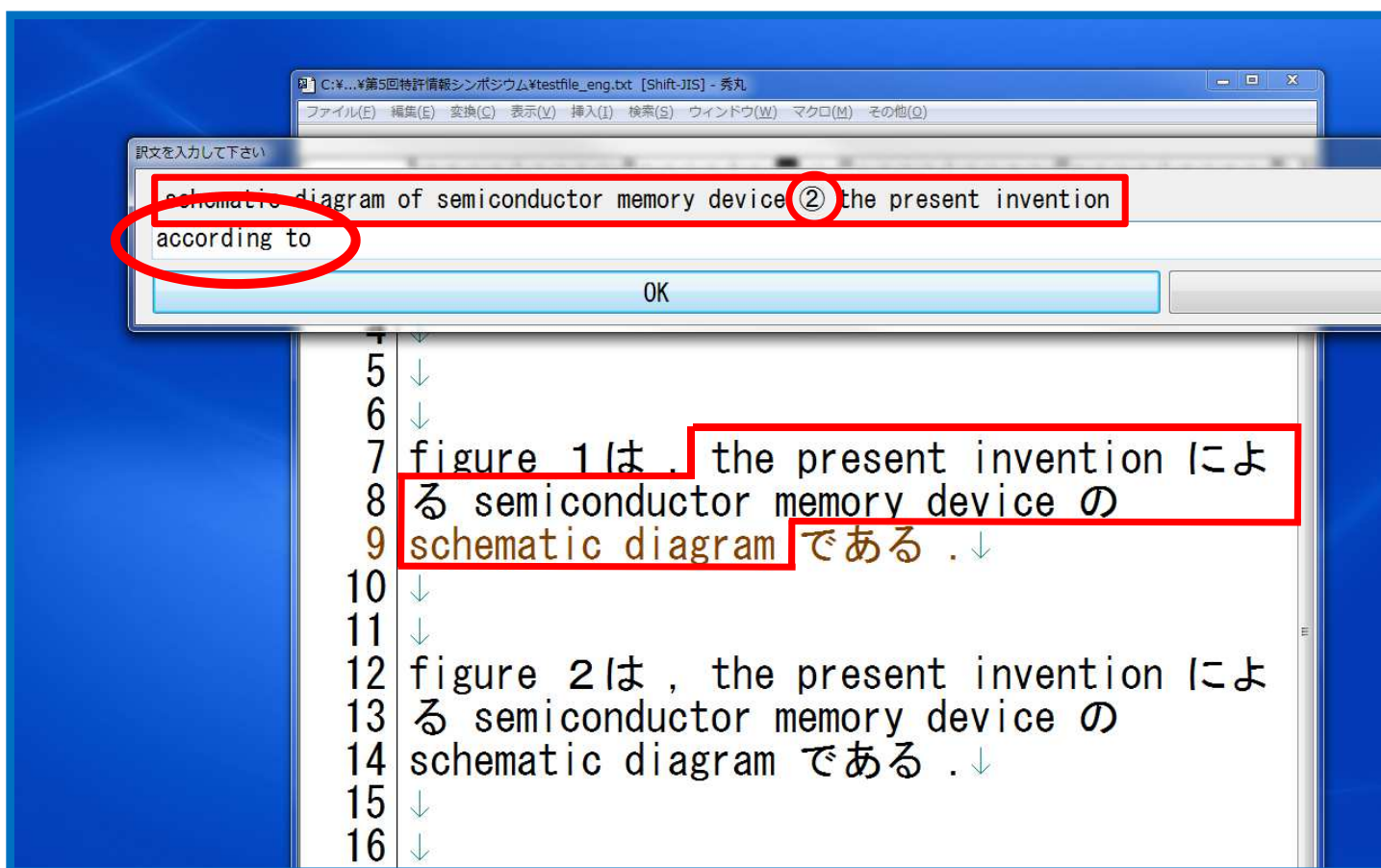
置換語の一括並べ換えマクロ（２）：訳語挿入



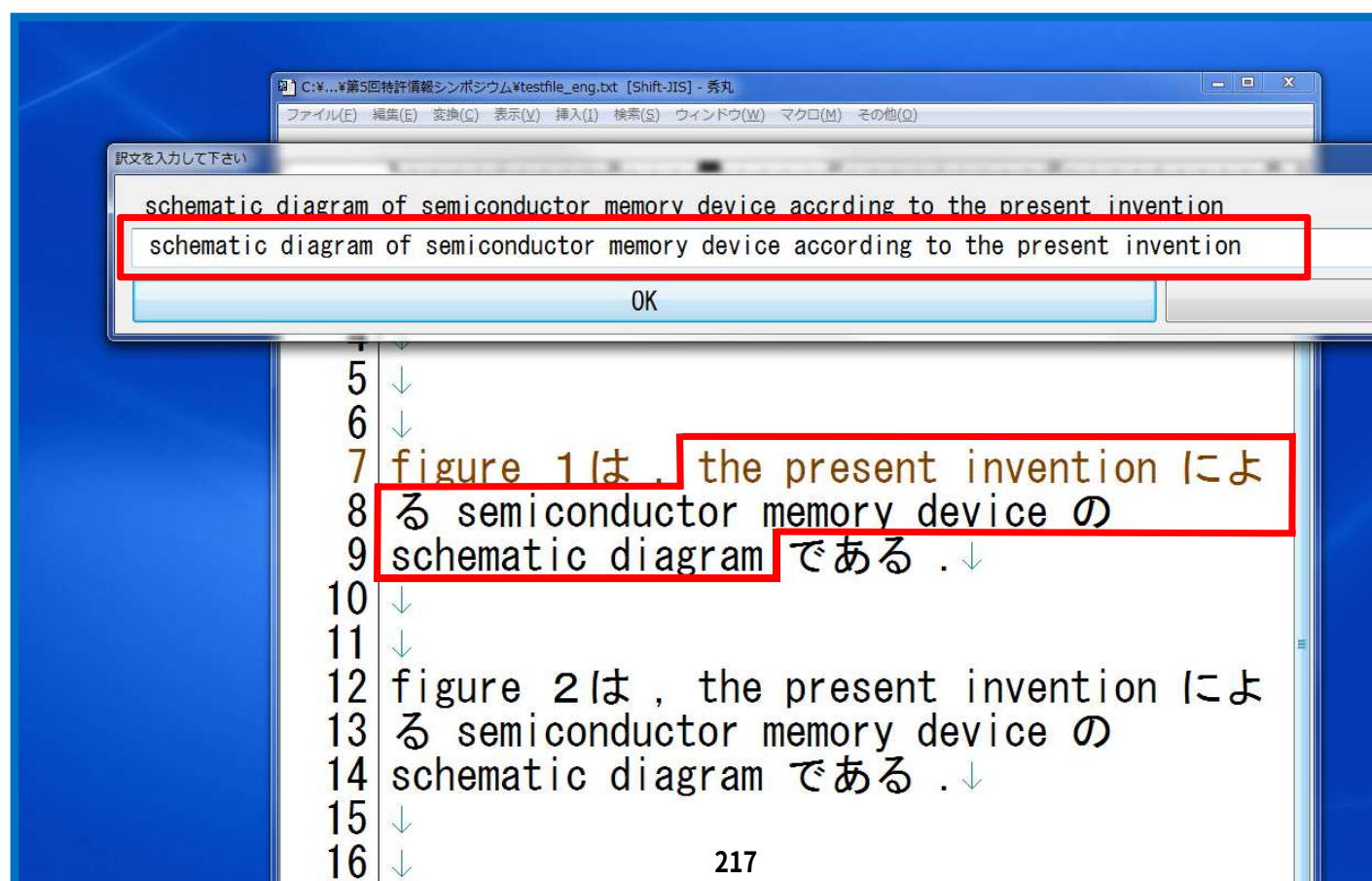
置換語の一括並べ換えマクロ（３）：次の訳語へ



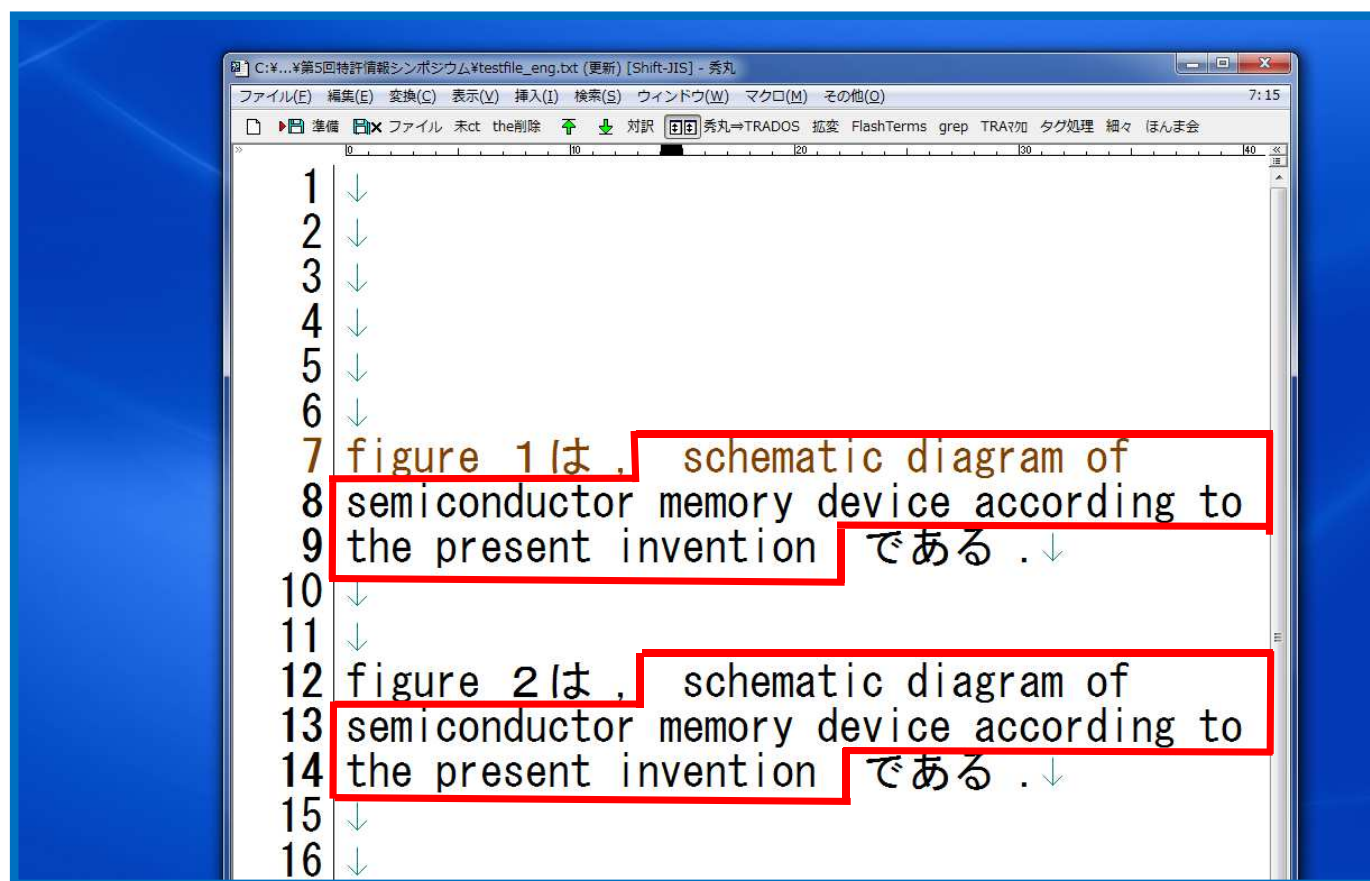
置換語の一括並べ換えマクロ（４）：訳語挿入



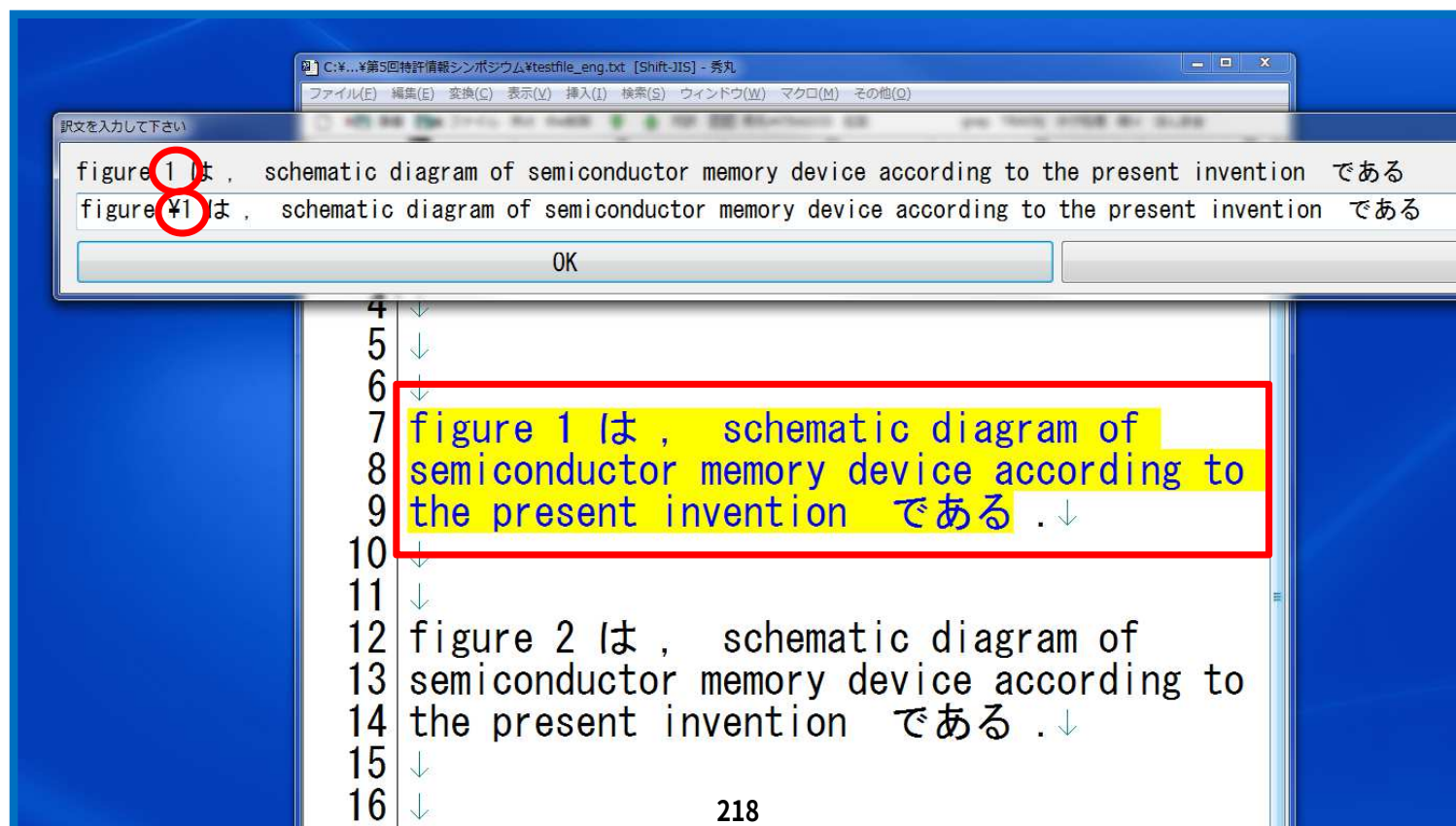
置換語の一括並べ換えマクロ（５）：訳文完成



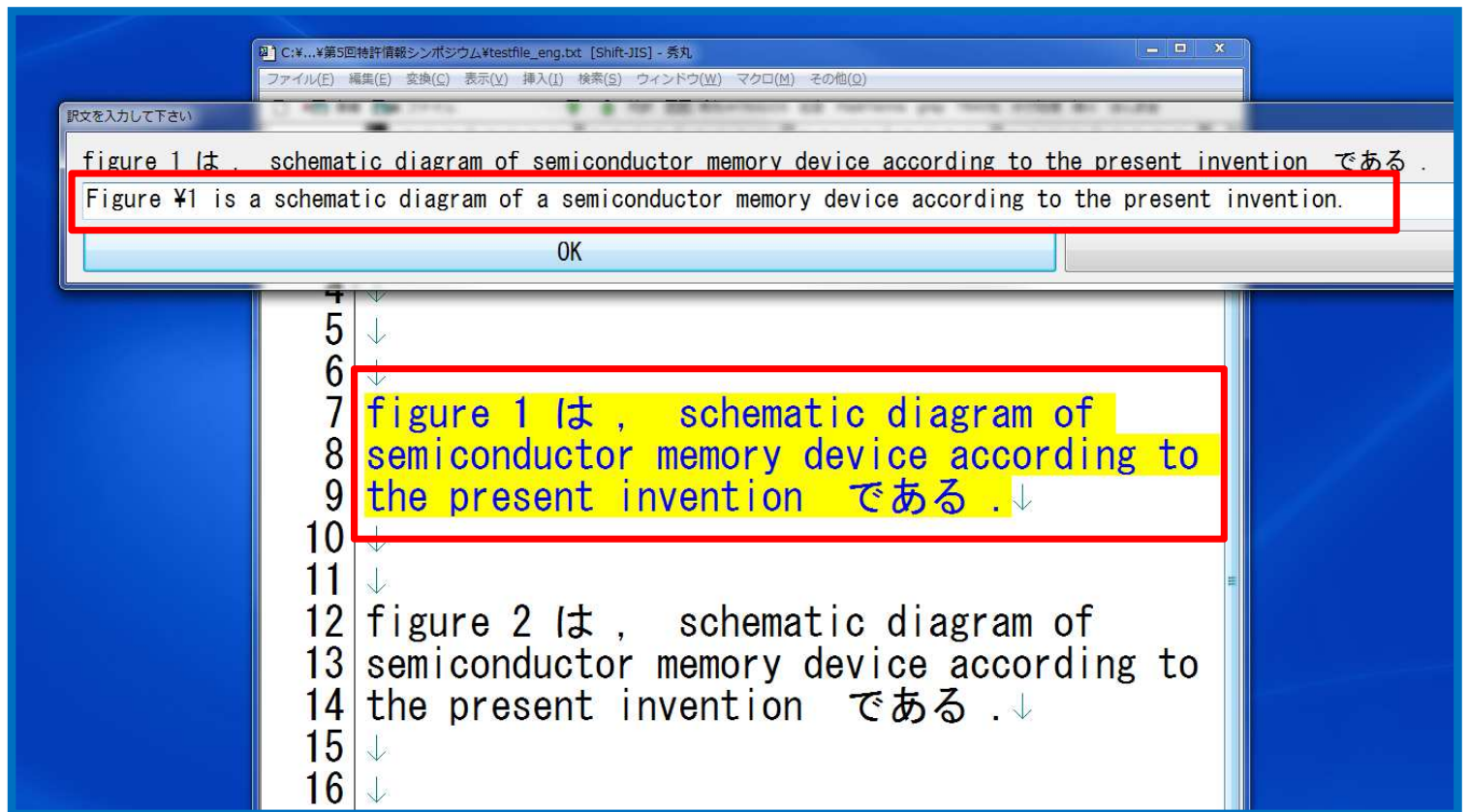
置換語の一括並べ換えマクロ（6）：一括並べ換え



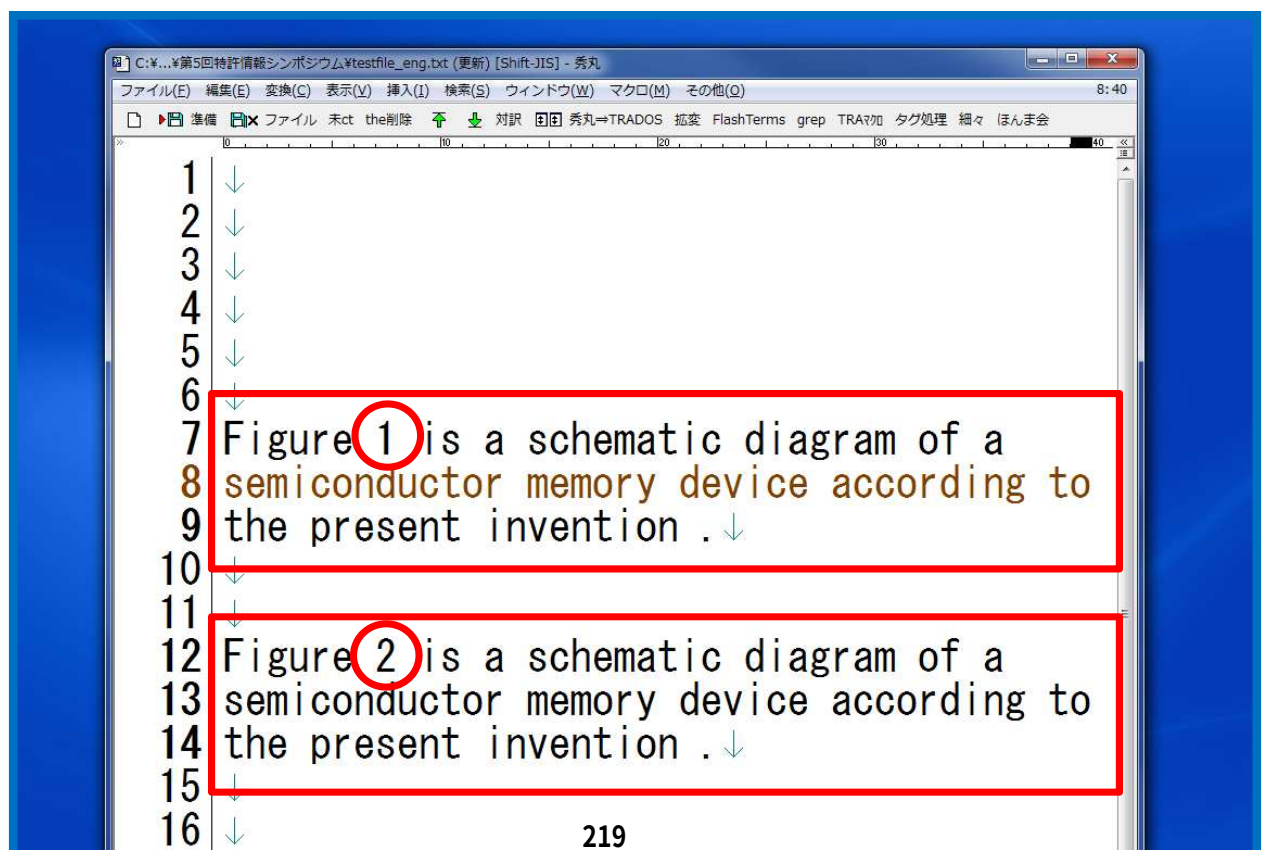
置換語の一括並べ換えマクロ（7）：範囲選択（正規表現）



置換語の一括並べ換えマクロ（８）：訳文完成



置換語の一括並べ換えマクロ（９）：一括並べ換え



知的財産リスクマネジメントにおける効果的な特許検索

ヤフー株式会社 知的財産部

宮崎 祐 河野 京子

企業の知的財産マネジメントにおける特許情報活用の重要性はますます高まっている。自社のサービスや製品が他社特許を侵害する特許リスクも、特許検索ツールにより他社特許を抽出することにより、未然に防ぐことが可能になる。しかし、この他社特許をどの程度実施すれば十分なのか明確な基準はなく、膨大な特許情報の中から、どのくらいの特許情報を抽出し、どのくらい時間をかけて調査すべきかの問題がある。また、サービスの開発規模に合わせて、効率的な調査時間をどのようにすればいいかなどの問題がある。本論文では、他社特許検索における問題を洗い出し、効率的な方法を考察する。

1. はじめに

近年、高額な特許侵害訴訟が提起されるなど、企業経営リスクを考える中でも、知的財産権に関わるリスクを考慮する機運が高まってきている。このような状況下では、知的財産権リスクを最小限にする手法が特許訴訟に巻き込まれないようにするためにも、非常に重要となる[1]。その中の重要検討項目の1つに、いかに正確に必要とされる他社特許情報を検索できるかどうかというクリアランス調査[2]が挙げられる。しかし、実際に実務として新サービスや新製品に対して、クリアランス調査を実施する場合、下記の問題がある。

- (1) どこまで稼働をかけて特許調査すべきか。
- (2) 調査担当者の技能によってどのような配慮をすべきか。
- (3) サービスや製品の規模に応じて、どのような調査をすべきか。

このような問題点を解決するために、ク

リアランス調査の実情をデータとして把握する必要があるが、抽出された自社製品やサービスに該当する他社特許がすべて洗い出されているのかどうかを判定することは難しいし、把握はほぼ不可能である。また、これまで、この問題点の解決のための手法も行われていない。

そこで、今回は、あらかじめ他社特許に該当するような機能が含まれる仮想のサービスを設定し、検索実施担当者が、この仮想サービスに関する仕様を頼りに、クリアランス調査を行うという実験を実施した。被験者には、特許庁の分類を活用した絞り込み件数と絞り込みにかかった所用時間、クレームと全文に関するキーワード検索での絞り込み数と検索にかかった所用時間、クレームと全文に関して実際に文章を読み込んで絞り込む作業による絞り込み件数と絞り込みにかかった所用時間をデータとして記録してもらった。実験の主催者は絞り込んだ特許群の中に予め該当するようにした正解の他社特許が入っているかどうかを確認し、分析を実施した。

2. 実験概要

上記問題を解決するために、他社特許検索の実情をデータとして把握する必要があるが、抽出された自社製品やサービスに該当する他社特許がすべて洗い出されているのかどうかを判定することは難しいし、把握はほぼ不可能である。

そこで、今回は、必要なデータを取得するために、あらかじめ存在する他社特許が該当するような仮想サービスを設定し、その仕様にもとづいて調査を実施するという実験形態をとった。

この仮想サービスは、図1に示すようなユーザIDに基づいて、ユーザに対する適合度を考慮したターゲティング広告システムを想定し、システムの機能数は、単純な処理は除いて、6つの機能と想定した。また、4つの特許が該当するようなサービスとなるように仕様を調整した。

今回の実験に参加した4人の被験者には、表1のような調査票を配布し、FI、クレーム、全文等の検索条件と所要時間を記録するように依頼し、実験を行った。

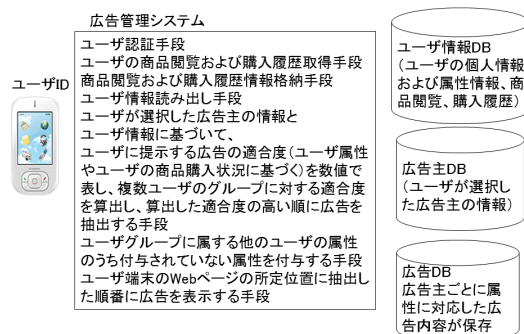


図1 仮想サービス概要

また、この仮想クリアランス調査は、最

初にFIによる絞り込みをおこない、絞り込んだ特許明細書の中から、さらにキーワードで絞り込みを行い、最終的には、特許明細書を読み込むことにより該当しているかどうかの確認を実施する流れをとった。ここで、FIとは、国際的に用いられている特許分類であるIPCを日本の特許庁がさらに細分化し、調査効率が上がるようにした分類である[3][4]。この手順により、図2のような関係で該当特許まで到達する。ここで、サービス機能空間とは、特許明細書のクレームの記載に対応する実サービスの機能要件の集合を指す。

表1 調査票

	分類	キーワード		明細書内容確認	
	FI	クレーム	全文	クレーム	全文
実施内容	FI番号を記載してください	クレームで限定したキーワードを記載してください	全文で限定したキーワードを記載してください	クレームを読んだ明細書の特許番号を記載してください	全文を読んだ明細書の特許番号を記載してください
対象件数					
所用時間					



図2 特許情報の絞り込みの関係性

また、そのデータから上記(1)～(3)の問題点を解決するための下記の分析指標を考案した。

(a) クリアランス調査である他社特許検索全体における効率のよい特許調査方法を分析するため、上記各絞り込み段階での効率を表す指標を考案し、他社特許調査における問題点を浮き出すような分析ができるようにした。

(b) 1時間あたりの他社特許の発見率を表す指標を考案し、サービス分野や規模に応じた目安の特許調査時間を算出できるようにした。

3. 実験結果

3. 1 FI 絞り込み段階での他社特許見逃し率

今回は、3段階で絞り込みを実施し、求めている特許明細書を求めるという実験であるが、各絞り込み段階で正解を見逃している状況と、その絞り込みにかかった時間を考慮した効率を調査するために、見逃し率と絞り込み効率という指標を導入し、分析した。

FIによる絞り込みに関する見逃し率 f_r は、調査の第一段階での絞り込みであるFIの絞り込みのミスで他社特許を捕獲できなくなった割合を表す指標で、下記の式により求めた。

$$f_r = \frac{f_n}{w_n}$$

ここで、 f_n はFIの設定により見逃した他社特許数であり、 w_n は仮想サービスに含まれる他社特許の全体数であり、今回の実験の場合は、4件になる。

また、FI絞り込み効率 f_e は、正解を得るための効率を表す指標で、下記の式により求めた。

$$f_e = w_c \times f_t$$

ここで、 w_c は絞りこまれた特許数をそこに含まれている正解の特許数で割った単位正解数あたりの他社特許数であり、 f_t は、そのかかった時間を表す。

表2と表3に各担当者ごとの f_r と f_e を求めた結果を示す。表において、各担当者ごとの f_r を比較すると、担当者AとDが見逃し率は0となり、第1段階での絞り込みをクリアできているが、担当者Bは、見逃しが比較的多いことがわかる。

表2 FI 絞り込み段階での見逃し率 f_r

	担当者A	担当者B	担当者C	担当者D
見逃し件数	0	2	1	0
見逃し率(FI)	0	0.5	0.25	0

*他社特許全体数は4

表3 FI 絞り込み効率 f_e

	担当者A	担当者B	担当者C	担当者D
FI通過数(正解数)	4	2	3	4
絞りこみ件数	6,699	14,425	149,852	112,215
単位正解数あたりの絞り込数	1,675	7,213	49,951	28,054
FI調査時間(分)	30	10	10	5
FI絞り込効率	837	721	8,325	2,338

f_e に関しては、担当者ごとに比較すると、担当者AはFI選定に時間をかけ、絞り込み効率を落とす要因を作っているが、その分、絞り込み範囲件数は、非常に絞られて精度を上げており、全体としては、絞り込み効率が一番よくなっている。逆に担当者Dは、時間をかけずに、大きく網をかけて絞り込みを実施して、絞り込み効率を上げている。

担当者Bは、あまり時間をかけずに絞り込み、絞り込み範囲も狭く、よく絞り込まれているが、見逃し率も考慮すると、もう少し

し時間をかけてもよかったと思われる。

3. 2 キーワード絞り込み段階での他社特許見逃し率

キーワードによる絞り込みに関する見逃し率 k_r は、調査の第二段階での絞り込みであるキーワードの設定ミスで他社特許を捕獲できなくなった割合を表す指標で、下記の式により求めた。

$$k_r = \frac{k_n}{w_f}$$

ここで、 k_n はキーワードの設定により見逃した他社特許数であり、 w_f はF I 設定により絞り込まれた他社特許数である。

また、キーワードによる絞り込み効率 k_e は、F I のときと同様に、正解を得るための効率を表す指標で、下記の式により求めた。

$$k_e = k_c \times k_t$$

ここで、 k_c は絞りこまれた特許数をそこに含まれている正解の特許数で割った単位正解数あたりの他社特許数であり、 k_t は、そのかかった時間を表す。

表4と表5に各担当者ごとの k_r と k_e を求めた結果を示す。表において、各担当者ごとの k_r を比較すると、担当者Dが見逃し率は0となり、第2段階での絞り込みをクリアできているが、第1段階で効率がよかった担当者Aが、第2段階ではかなり見逃していることがわかる。

k_e に関しては、担当者ごとに比較すると、担当者Bはキーワード選定に他の人の2倍以上時間をかけており、絞り込み効率を落とす要因を作っているが、その分、絞込範囲件数は、非常にしぼれて精度を上げてい

る。逆に担当者Cは、時間をかけずに、キーワード検索を実施して、絞り込み効率を上げている。

表4 キーワード検索段階での見逃し率 k_r

	担当者A	担当者B	担当者C	担当者D
見逃し件数	3(3)	2(0)	3(2)	1(1)
FI通過数	4	2	3	4
見逃し率(キーワード)	0.75 (0.75)	0.5 (0)	0.75 (0.67)	0.25 (0.25)

()内はFIの設定を除いて、キーワードのみの効果を算出

表5 キーワード検索絞り込み効率 k_e

	担当者A	担当者B	担当者C	担当者D
キーワード通過数(正解数)	1	2	1	3
絞りこみ件数	40	220	162	119
単位正解数あたりの絞込数	40	110	162	40
キーワード調査時間(分)	40	80	20	60
キーワード絞込効率	27	147	54	40

3. 3 特許明細書確認段階での他社特許見逃し率

特許明細書確認による絞り込みに関する見逃し率 s_r は、調査の第三段階での絞り込みである特許明細書確認の絞り込みのミスで他社特許を捕獲できなくなった割合を表す指標で、下記の式により求めた。

$$s_r = \frac{s_n}{w_k}$$

ここで、 s_n は特許明細書確認により見逃した他社特許数であり、 w_k はF I とキーワード検索により絞りこまれた他社特許の全体数である。

また、特許明細書確認効率 s_e は、正解を得るための効率を表す指標で、下記の式により求めた。

$$s_e = s_c \times s_t$$

ここで、 s_c は絞りこまれた特許数であるが、この段階は、最終段階であるため、担当者が正解として抽出した数となり、そこに含まれている正解の特許数で割った単位正解数あたりの他社特許数であり、 s_t は、そのかかった時間を表す。

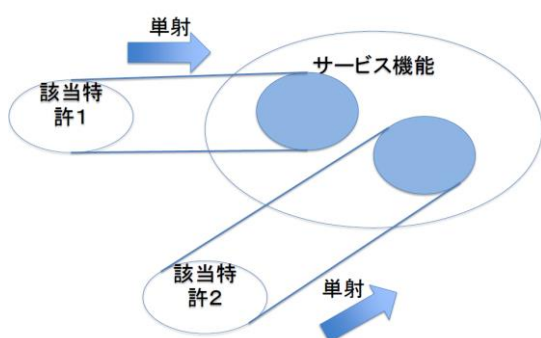


図3. サービスの特許への該当性

表6と表7に各担当者ごとの s_r と s_e を求めた結果を示す。表において、各担当者ごとの s_r を比較すると、担当者AとCが見逃し率は0となり、第3段階での絞り込みをクリアできているが、これまでのほとんど見逃さずに絞り込んできた担当者Dが、最後の読み込みで見逃していることがわかる。

s_e に関しては、担当者ごとに比較すると、担当者Aは明細書を時間をかけて読んでおり、効率を落としている。逆に担当者Cは、担当者Cは、短時間で、読んでいたにもかかわらず、見逃しがなく、明細書読み込みスキルは高いことがわかる。

表6 明細書確認段階での見逃し率 s_r

	担当者A	担当者B	担当者C	担当者D
見逃し件数	0	1	0	2
キーワード通過数	1	2	1	3
見逃し率(明細書)	0	0.5	0	0.67

表7 特許明細書確認効率 s_e

	担当者A	担当者B	担当者C	担当者D
明細書通過数(正解数)	1	1	1	1
読み込み対象明細書数	40	220	162	119
明細書調査時間(分)	200	120	90	123
1件あたりの明細書読み込み時間(分)	5	0.5	0.6	1
明細書確認効率	200	120	90	123

以上の調査指標を基に、単位時間あたりで他社特許を発見する他社特許発見効率 d_e を下記の式により求めた。

$$d_e = \frac{d_n}{w_t}$$

ここで、 d_n は最終的抽出できた他社特許数であり、 w_t は検索全体にかかった時間を表す。

表8に担当者全員の他社特許発見効率 d_e を記載した。全員が全正解数4件のうち1件を見つけることになったが、短時間でみつけたのは担当者Cであり、調査効率は高いことがわかる。

表8 他社特許発見効率 d_e

	担当者A	担当者B	担当者C	担当者D	平均
特許発見数	1	1	1	1	1
全体調査時間	270	210	120	188	197
他社特許発見効率	0.2	0.3	0.5	0.3	0.3

各調査段階での他社特許調査効率を表9にまとめる。

表9 各調査段階での見逃し率のまとめ

	FI	キーワード	明細書
見逃し率	0.25	0.42	0.3
調査時間(分)	14	50	190

この結果より、実際に内容を読んで確認する段階である明細書を読み込段階においても見逃しが発生する可能性が高いことがわかる。

3. 4 特許調査にかかる目安の時間の算出

この仮想サービスによる実験により求めた値から、実際のサービスにおける他社特許検索の目安となる検索作業時間を求める。

まず、次の指標を準備する。1機能あたりに含まれる他社特許数を示す他社特許含有率 c_i は、下記の式により求めた。

$$c_i = \frac{a_n}{e_n} \quad (1)$$

ここで、 a_n は調査対象となるサービスに含まれると推定される他社特許数であり、 e_n は調査対象となるサービスの機能要素数を表す。この指標は、仮想サービス設計により、想定される機能と実際に対応づけ可能な特許数から求めるもので、分野により調整する必要がある。

また、対象となるサービスの規模に応じて変化する他社特許が含まれる割合を示すである他社特許発見率 d_r は下記の式により求めた。

$$d_r = \frac{e_n \cap s_n}{e_n} \quad (2)$$

ここで、 s_n は調査により抽出した件数を示す。また、 $e_n \cap s_n$ は、図3による対応関係により導出される要素の交わりを示す。

サービスの規模に応じて変化する単位時間あたりで他社特許を発見する他社特許調査効率 c_e は、下記の式により求まる。

$$c_e = \frac{e_n \cap s_n}{w_t} \quad (3)$$

したがって、必要とされる調査時間 n_t は、下記の式で求まる。

$$n_t = \frac{e_n \times c_i}{c_e} \quad (4)$$

この他社特許発見率 d_r は個人、およびチームのスキルの指標とすることができ、他社特許調査効率 c_e は調査しなければならないサービスの規模に応じた調査時間を求めるために活用することができる。

実際の当社システムに実装されているサービスで、表10のような要素機能を持つシステムに適用してみた。

表10 実システムにおける機能要素と検索結果

機能要素番号	類似件数	該当
1	3	
2	8	
3	2	
4	2	1
5	0	
6	4	
計	19	1

式(1)において、仮想実験の結果を用いるため、仮想実験の前提でとした $a_n = 4$ 、 $e_n = 6$ を用いる。今回適用した実際のサービスも機能要素は6個としており、他社特許含有率 $c_i = 0.67$ と求まり、式(2)において $e_n \cap s_n = 1$ 、 $e_n = 4$ とすると他社特許発見率は $d_r = 0.25$ と求まる。

また、式(3)において上記と同様に $e_n \cap s_n = 1$ とし、表8で求めた他社特許発見効率 $d_e = 0.3$ の値を c_e の値と見なし、式(4)において $e_n = 4$ 、 $c_i = 0.25$ 、 $c_e = 0.3$ と

すると必要とされる調査時間は $n_t = 3$ と求まり、仮想実験におけるレベルの調査結果を得るためには、調査時間は約3時間となり、実際に開発対象機能に対して、特許調査ポイントとして抜き出した機能6つ（＝サービス規模）に対して、3時間の調査を実施した結果、該当特許1件を検出することができた。

この時間を規準にサービスの重要性等を考慮し、目安となる時間を設定することが可能となる。

4. 実務上の考察

本仮想実験を実施後に被験者へヒアリングを実施すると、担当者が思っていた予想以上に、該当特許は見つけられていないことがわかった。本実験の実施前は、95%は該当特許を見つけれられるくらいの感触をもっている担当者もいたため、実際の結果に驚いている。

FIは短時間で絞り込む手段としては、非常に有効であり、さらにFI自体を調べる時間を短縮するために、自社のサービス別によく使うFIを整理しておき、同じ分野の特許を特定して、同じFIで検索をかけると見つけ易くなる。

さらに、本実験のような簡易な実験を実施することより、担当者の意識改善や検索による弱点等の早くに役立ち、各メンバーの検索スキル向上のための施策提案のための基礎資料となる効果も期待できる。

5. おわりに

本稿では、仮想実験の結果により得た指

標が、知的財産リスクマネジメントににおける問題点を解決する指標となっていることを確認した。この結果により、クリアランス調査業務の効率向上のための現状把握や改善検討項目の抽出等に有効であることを確認できた。

参考文献

- [1] 京本 富田. “知的財産リスクのマネジメント”. 知財マネジメント研究 2 PP28-38 2004
- [2] 野崎. “弁理士が知っておきたい国内外特許情報調査の基礎知識”. パテント Vol. 67 No. 1 2014
- [3] 奥村 藤井 谷川 岩山 難波 山本 内 山. “特許情報処理:言語処理的アプローチ コロナ社 PP56-67 2012
- [4] 特許庁
日本の特許分類 (FI・F ターム) について
<https://www.jpo.go.jp/seido/tokkyo/seido/bunrui/fi/index.html>

実行・プログラム委員会

委員長：	宇津呂武仁	筑波大学
副委員長：	須藤 克仁	奈良先端科学技術大学院大学
顧問：	辻井 潤一	産業技術総合研究所
	小林 明	日本特許情報機構
委員：	石附 直弥	日本特許情報機構
	今村 賢治	情報通信研究機構
	越前谷 博	北海学園大学
	江原 暉将	元・山梨英和大学
	大塩 只明	日本特許情報機構
	木下 聡	日本特許情報機構
	清藤 弘晃	日本特許情報機構
	黒橋 禎夫	京都大学
	後藤 功雄	NHK放送技術研究所
	綱川 隆司	静岡大学
	中澤 敏明	東京大学
	二宮 崇	愛媛大学
	横山 晶一	元・山形大学
事務局：	株式会社インターグループ	

————— 禁 無 断 転 載 —————

平成30年度AAMT/Japio特許翻訳研究会
第5回特許情報シンポジウム論文資料集

発 行 日 平成30年12月

発 行 一般財団法人 日本特許情報機構（Japio）
〒135-0016 東京都江東区東陽4丁目1番7号
佐藤ダイヤビルディング
TEL: (03) 3615-5511 FAX: (03) 3615-5521

編 集 AAMT/Japio特許翻訳研究会
アジア太平洋機械翻訳協会（AAMT）

印 刷 株式会社 インターグループ