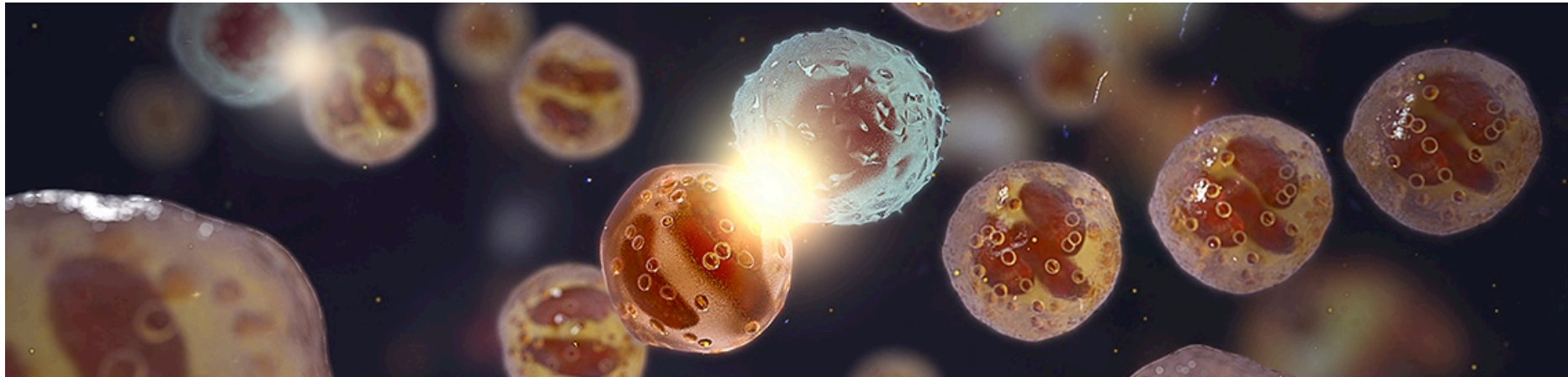


アストラゼネカにおけるAI翻訳(NMT)の導入

アストラゼネカ株式会社 研究開発本部
サイエンス&データテクノロジー統括部
田中 倫夫

6 March, 2019



免責事項

本内容は、あくまで発表者の見解によるものであり、アストラゼネカを代表するものではありません。



アストラゼネカ(AZ) 研究開発本部のミッション

- R&D Japan delivers AstraZeneca's pipeline in Japan
 - ... **with speed and quality**
 - ... **to maximise the value of our medicines for patients**



医薬品開発のトレンドは世界同時開発

- 国際多施設共同臨床試験
 - 日本で臨床試験(治験)を規制当局に提出するには
日本語の治験実施計画書(CSP)が必須
 - c.f. 臨床試験実施施設では英語でも可能
- 世界同時申請
 - 日本で新薬承認申請を規制当局に提出するには
日本語のサマリー文書(CTD)が必須



翻訳 – 2種類の用途

高い品質が求められるもの

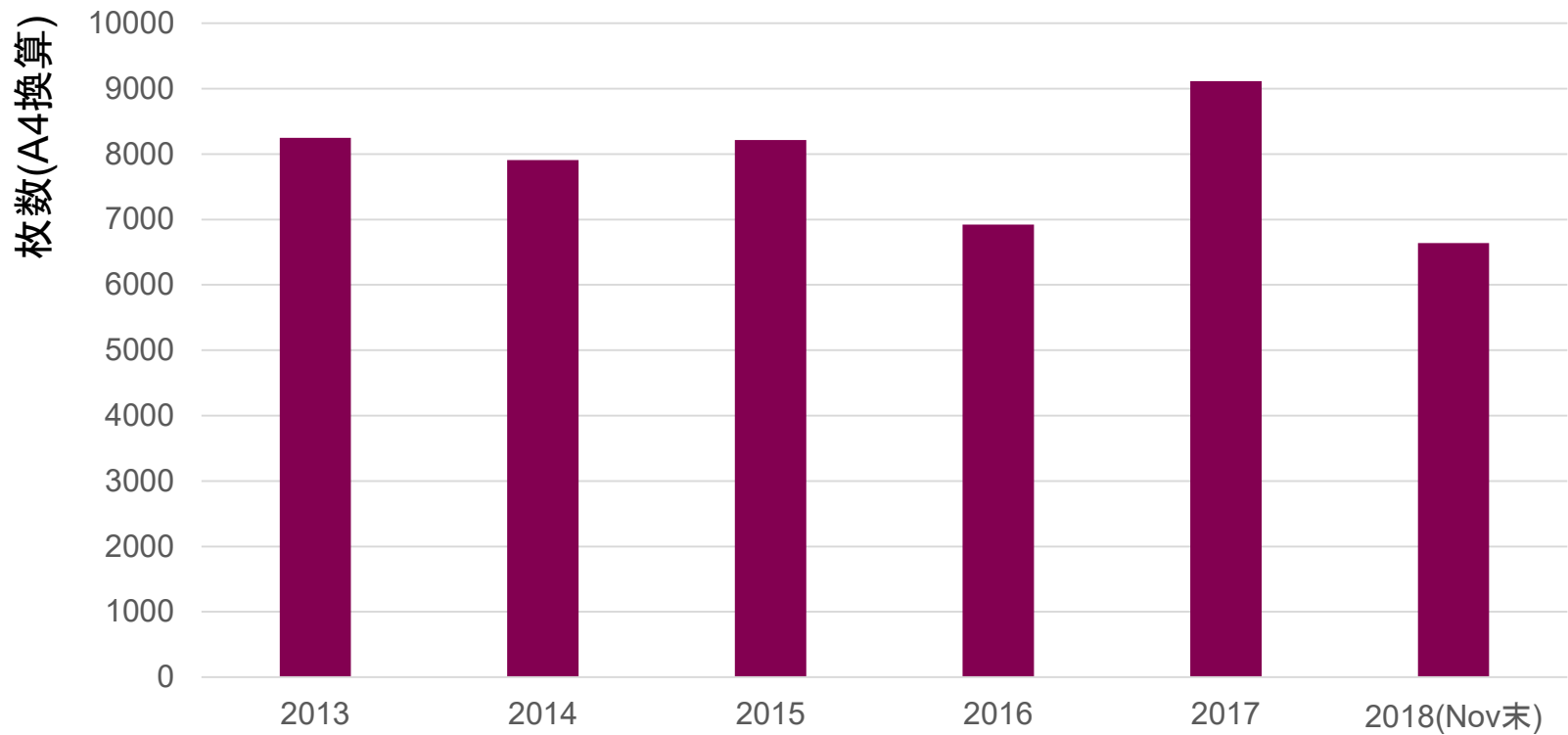
- 外部提出用途
- 臨床試験関連文書（CSP、同意説明文書等）
- 薬事文書（CTD、照会事項回答等）

意味が分かれば十分なもの

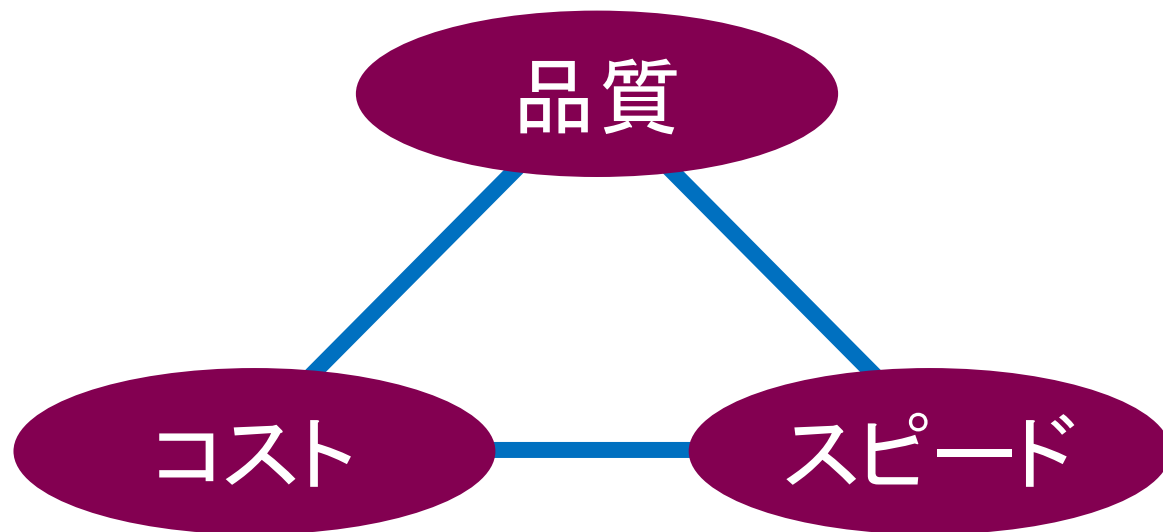
- 社内/個人用途
- 斜め読み



翻訳実績 (CSP、CTD、その他高い翻訳品質が求められる文書)



製薬業界の翻訳業務におけるビジネスニーズ



平成29年9月8日

総務省
国立研究開発法人情報通信研究機構

『翻訳バンク』の運用開始

—自動翻訳システムのさらなる高精度化に向けて、様々な分野の翻訳データを集積—

総務省と国立研究開発法人情報通信研究機構(NICT、理事長: 徳田 英幸)は、自動翻訳システムの様々な分野への対応や高精度化を進めるため、オール・ジャパン体制で翻訳データを集積する『翻訳バンク』の運用を開始します。これにより、社会・経済活動のグローバル化が進む中で我が国の国際競争力の強化に貢献します。

【ポイント】

- オール・ジャパンで様々な分野の翻訳データを集積する『翻訳バンク』の運用を開始
- ニューラル機械翻訳での活用等に向けて翻訳データを集積し、多分野対応や高精度化を実現
- 翻訳データを提供して頂く方にもメリットのある仕組みを導入

翻訳バンクのコンセプト

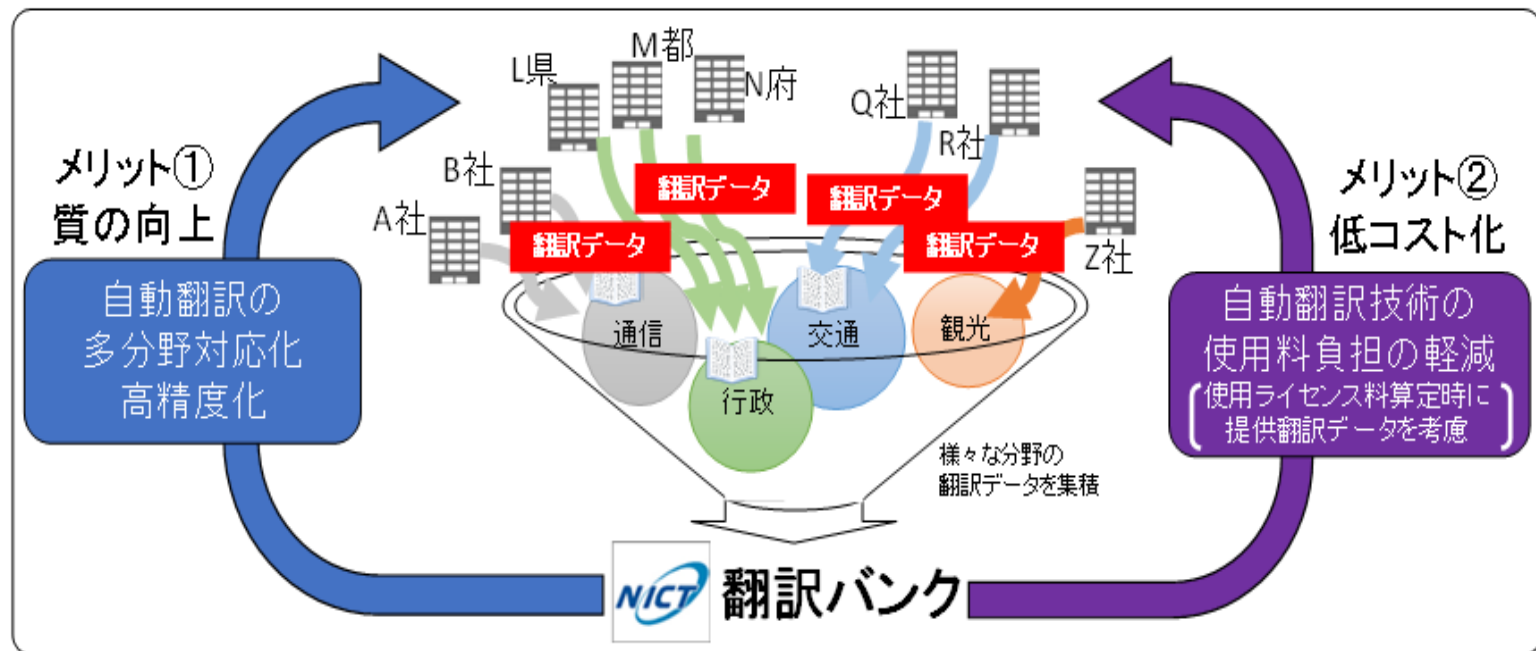


図 1 『翻訳バンク』のコンセプト



NICTとの共同研究

目的： 医薬品開発用にカスタマイズしたNMTによる翻訳物の品質評価



自動翻訳をみんなで育てるサイト
みんなの自動翻訳@TexTra テキストラ®
<https://mt-auto-minhon-mlt.ucri.jgn-x.jp/>



話題のニューラル機械翻訳(NMT)、
最新の自動翻訳研究に基づく、

「高精度自動翻訳エンジン」

をインターネット上で公開中!



汎用NMT*を凡そ100万文対
(5万ページ)の対訳データで
アダプテーション

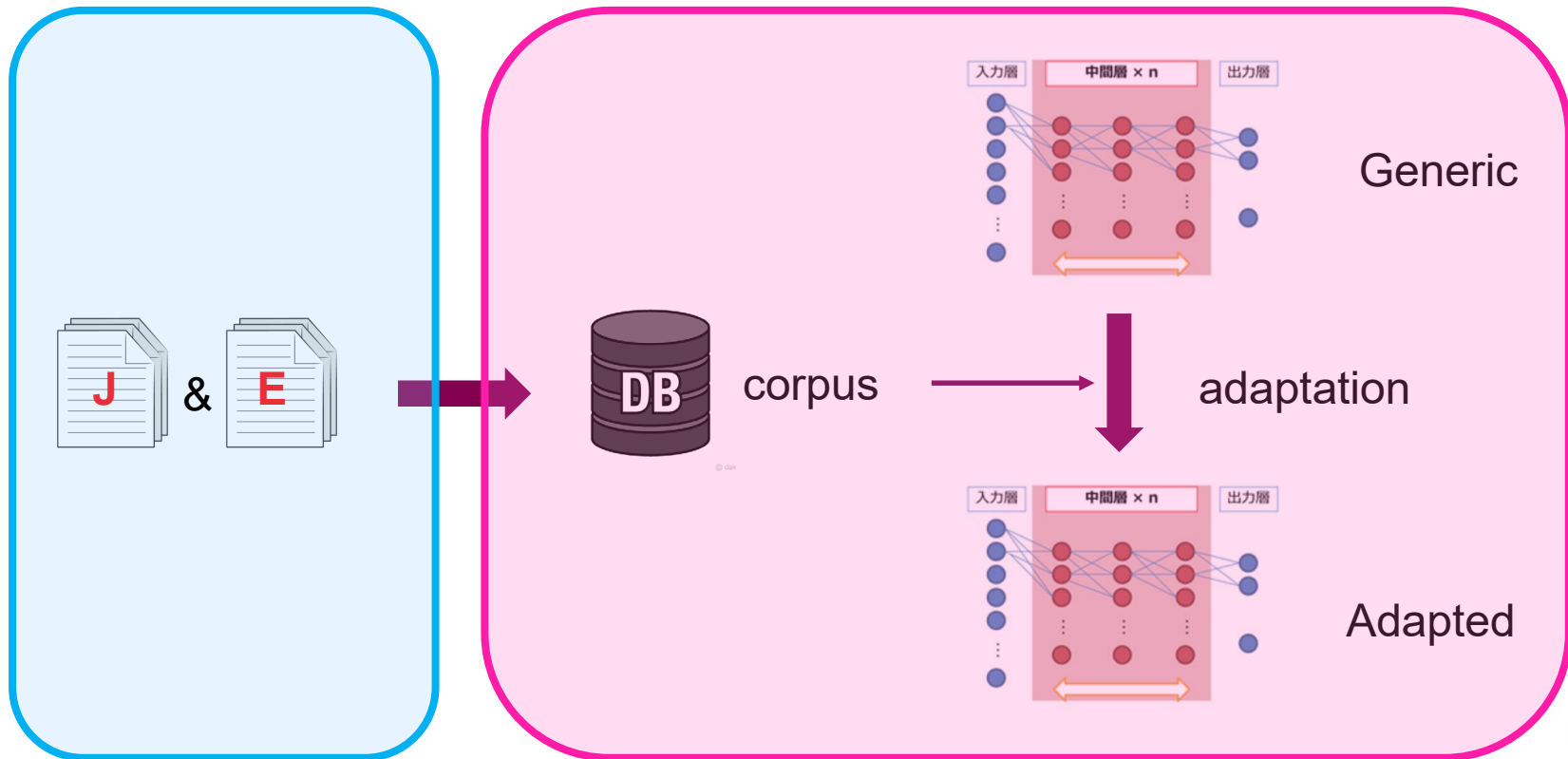
<https://mt-auto-minhon-mlt.ucri.jgn-x.jp/?oid=>



Study Method (1) – 翻訳エンジンのアダプテーション

AZ

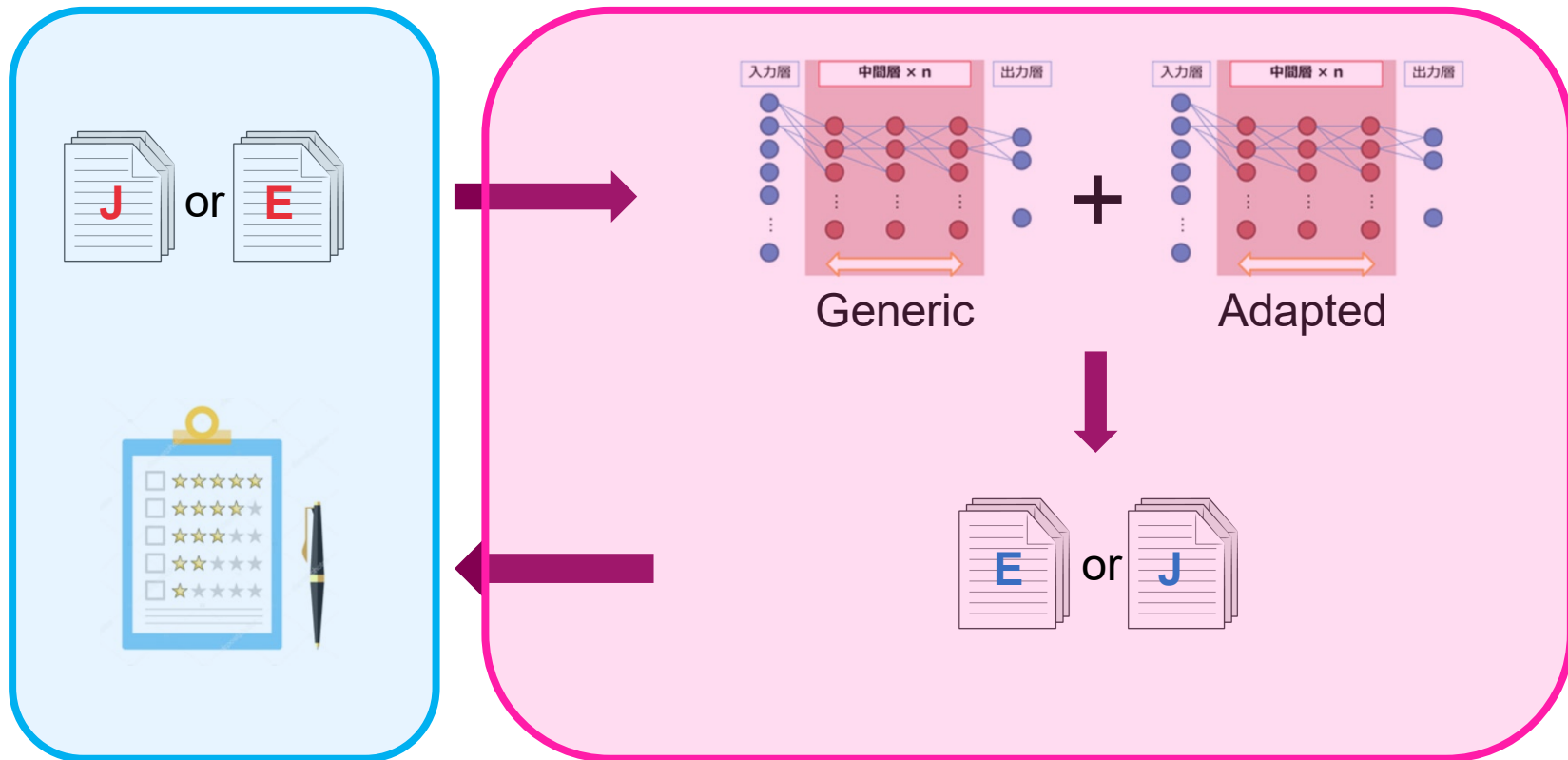
NICT



Study Method (2) – 翻訳の品質評価

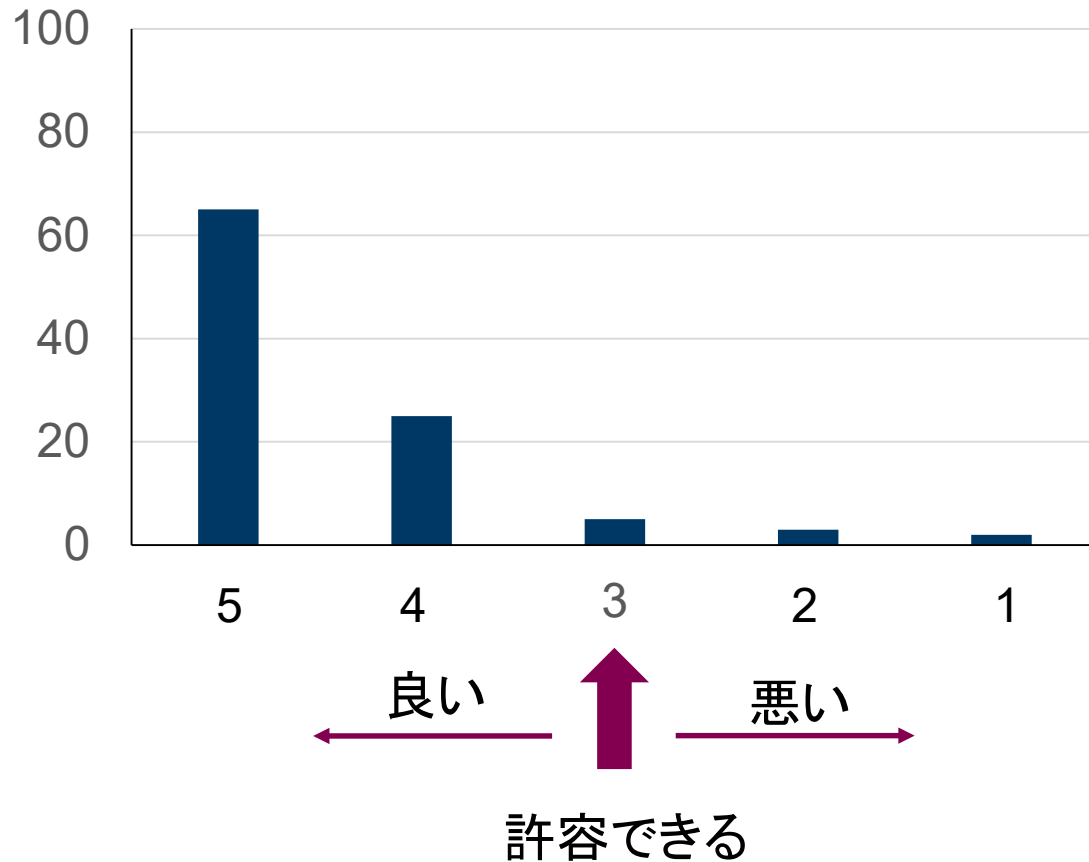
AZ

NICT

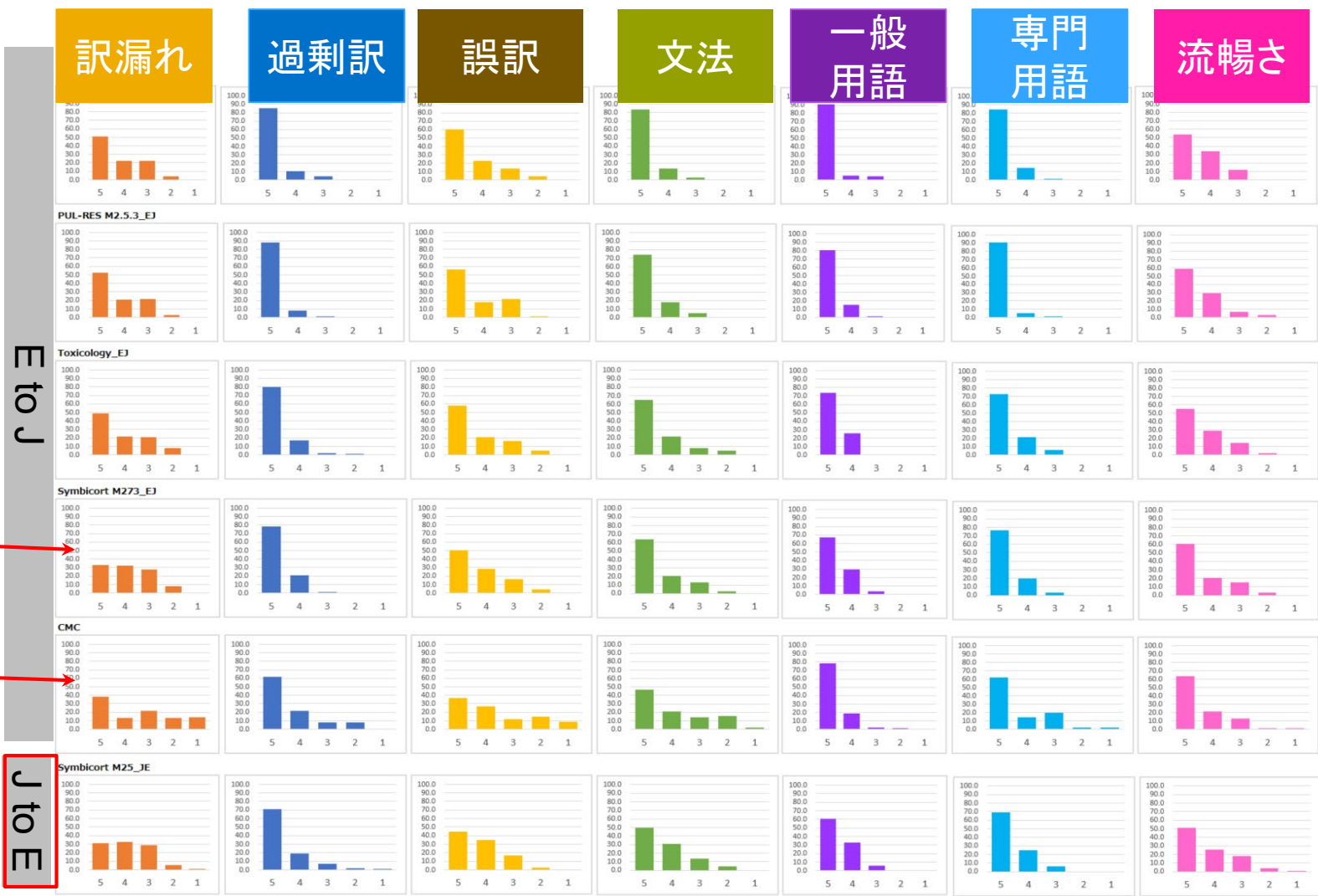


評価結果の見方

評価した全ての文に
対する割合(%)



評価結果 (CTD)



E to J

J to E

長文多い

特殊領域

BLEU スコア

評価結果

BLEU スコア	解釈*	日→英	英→日
< 10	ほとんど役に立たない		
10~19	主旨を理解するのが困難である		
20~29	主旨は明白であるが、文法上の重大なエラーがある	18.68	21.94
30~40	理解できる、適度な品質の翻訳		
40~50	高品質な翻訳	41.94	
50~60	非常に高品質で、適切かつ流暢な翻訳		50.52
> 60	人が翻訳した場合よりも高品質であることが多い		

* <https://cloud.google.com/translate/automl/docs/evaluating?hl=ja>

アダプテーション前
アダプテーション後



高い品質が確認できた（特に英→日）



- 全ての評価項目について許容できる品質を認めた
 - 翻訳者の品質に到達するにはポストエディット(PE)が必要
- 翻訳に要する時間
ポストエディットを実施しても、NMTを使うほうが早い



翻訳例1* (CSP)

原文	汎用NMT	アダプテーション後NMT
<p>The analysis of the primary endpoint will include all the data captured between weeks 0 and 16 of receiving injections.</p>	<p>一次エンドポイントの分析には、注射の0~16週間の間に取得されたすべてのデータが含まれます。</p>	<p>主要評価項目の解析には、投与0~16週後に得られた全てのデータが含まれる。</p>
<p>For the additional secondary outcome measure, change from baseline in mean ABC score will be measured for both treatment groups.</p>	<p>追加の二次アウトカム測定のために、平均ABCスコアにおけるベースラインからの変化は、両治療群の両方について測定されるであろう。</p>	<p>その他の副次的評価項目については、投与群別にABCスコアの平均値のベースラインからの変化量を測定する。</p>

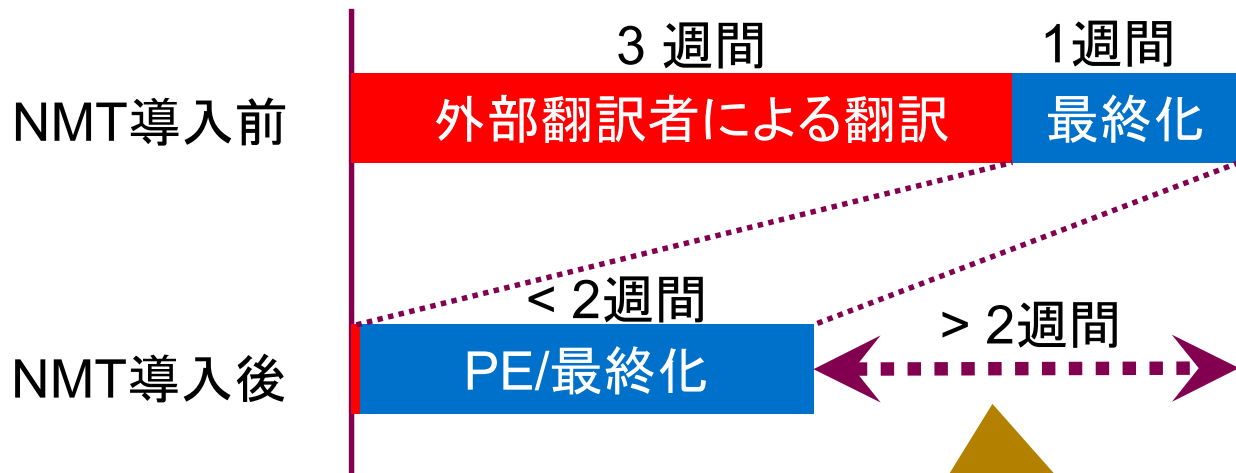


翻訳例2* (CSP)

原文	汎用NMT	アダプテーション後NMT
<p>A majority of subjects had AEs that were mild or moderate in severity and not related to investigational product (IP).</p>	<p>大多数の被験者は、重症度が軽度または中等度のAESを有し、治験薬(IP)とは無関係であった。</p>	<p>大多数の被験者では、軽度又は中等度のAEが認められ、治験薬との因果関係は否定された。</p>
<p>A detailed description of the chemistry, pharmacology, efficacy, and safety of the drug is provided in the Investigator's Brochure.</p>	<p>医薬品の化学的性質、薬理、有効性及び安全性に関する詳細な説明は、治験責任医師のパンフレットに記載されている。</p>	<p>治験薬の化学的、薬理的、有効性及び安全性の詳細な記述は治験薬概要書に記載されている。</p>



NMT使用実例; CSPの翻訳

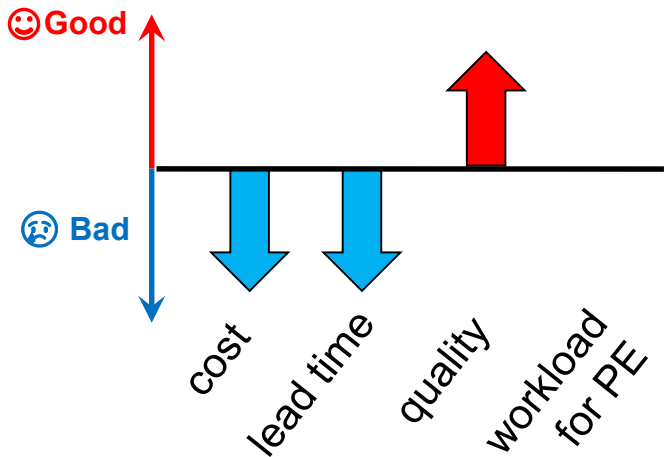


2週間の時間短縮を達成

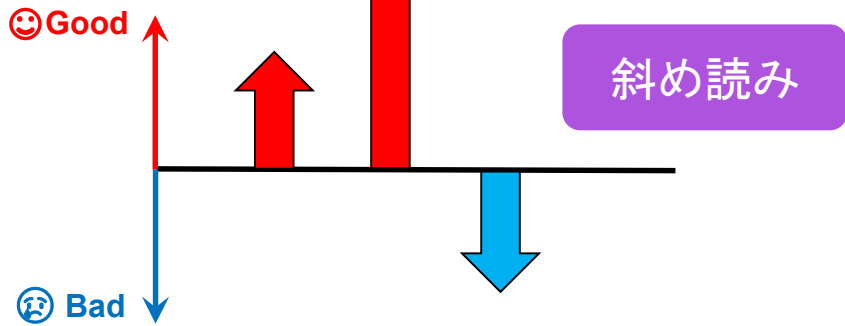


まとめ AI翻訳の有用性が確認された

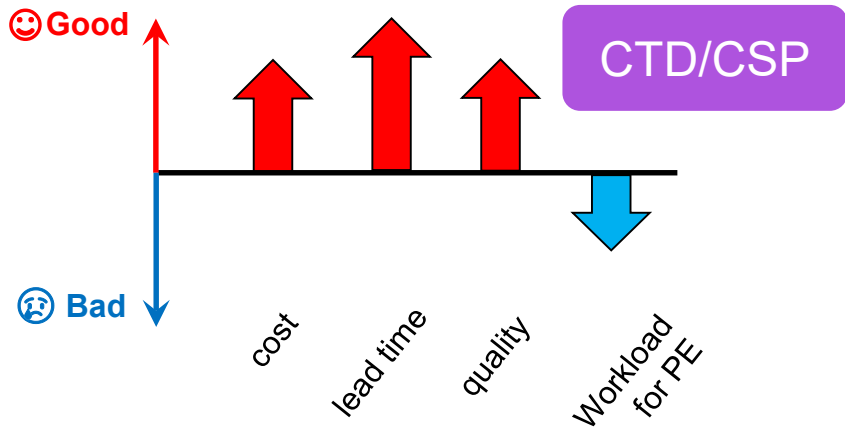
外部翻訳者(現状)



NMT (PEなし)



NMT (PEあり)



使用目的によっては
ポストエディットが必要



ご清聴ありがとうございました!



Confidentiality Notice

This file is private and may contain confidential and proprietary information. If you have received this file in error, please notify us and remove it from your system and note that you must not copy, distribute or take any action in reliance on it. Any unauthorized use or disclosure of the contents of this file is not permitted and may be unlawful. AstraZeneca PLC, 1 Francis Crick Avenue, Cambridge Biomedical Campus, Cambridge, CB2 0AA, UK, T: +44(0)203 749 5000, www.astrazeneca.com

