

翻訳バンクの概要説明

情報通信研究機構(NICT)

隅田英一郎

①

日本語と
多言語の
自動翻訳は
使える！

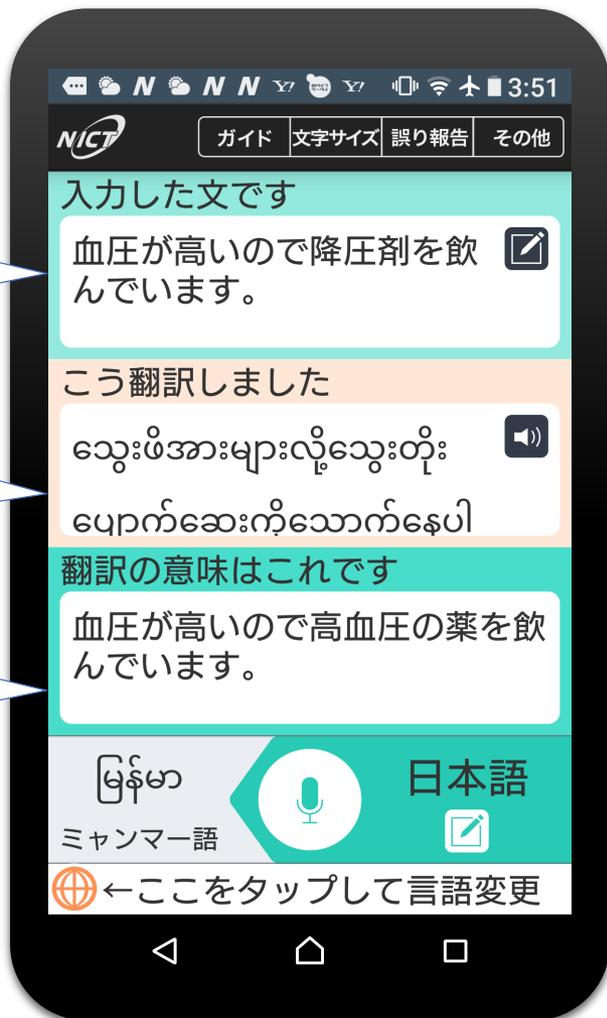
2019年1月15日で4,084,946 DLs

無料のVoiceTraをお試しく下さい

入力した文

翻訳結果

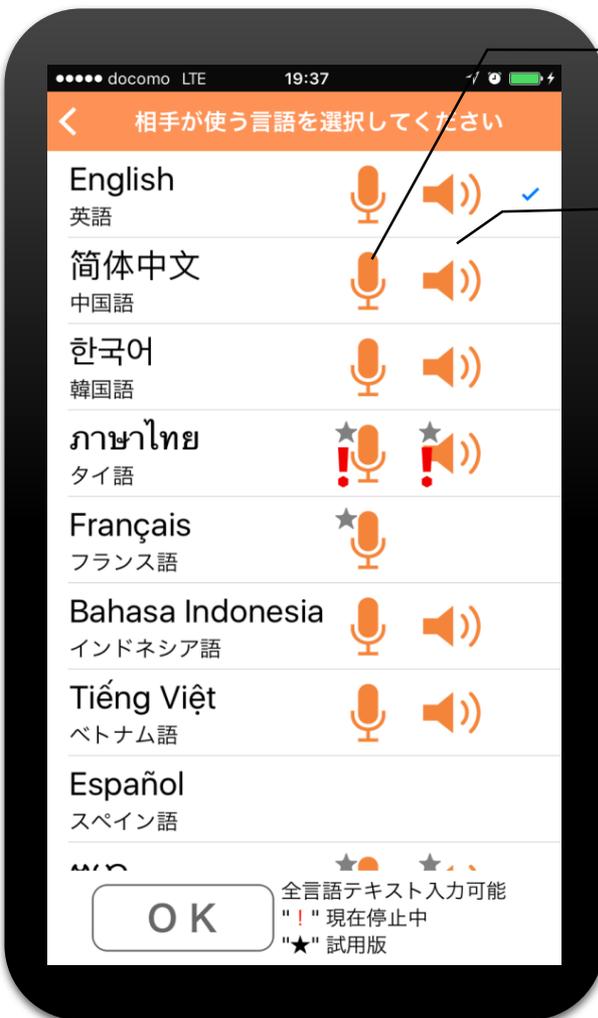
結果の確認



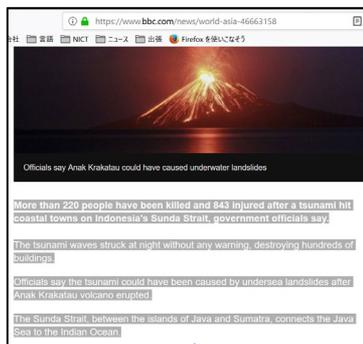
音声入力可
18言語

音声出力可
15言語

翻訳可能
31言語



ニューラル機械翻訳 (NMT) を 「みんなの自動翻訳@TexTra」で試せる



- WEBベースのコピペ
- OFFICEのアドインでファイル単位に
- CAT(MEMSOURCE,TRADOS,MemoQ)と連携
- APIでシステム構築



英語 → 日本語
↔

汎用NMT 【英語 - 日本語】
1

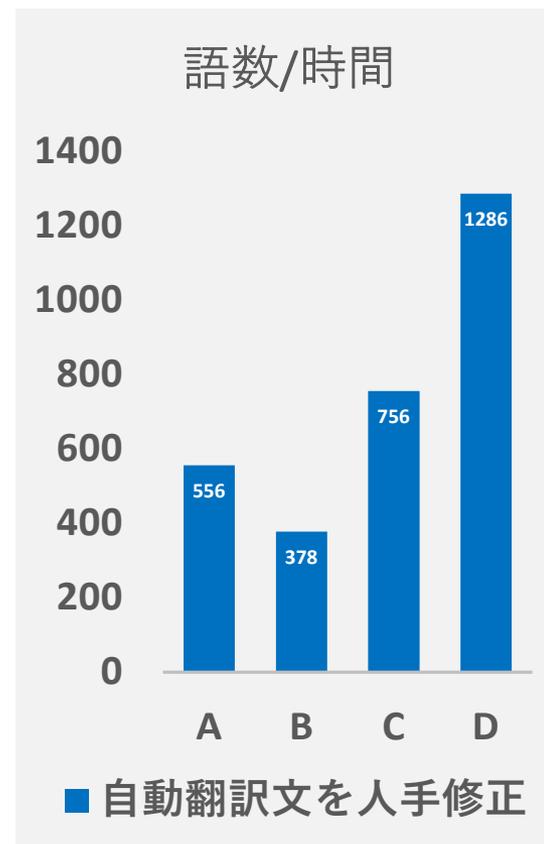
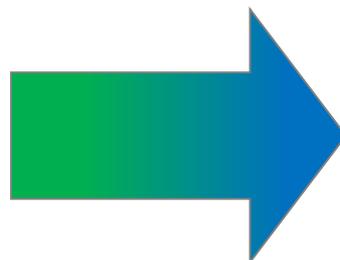
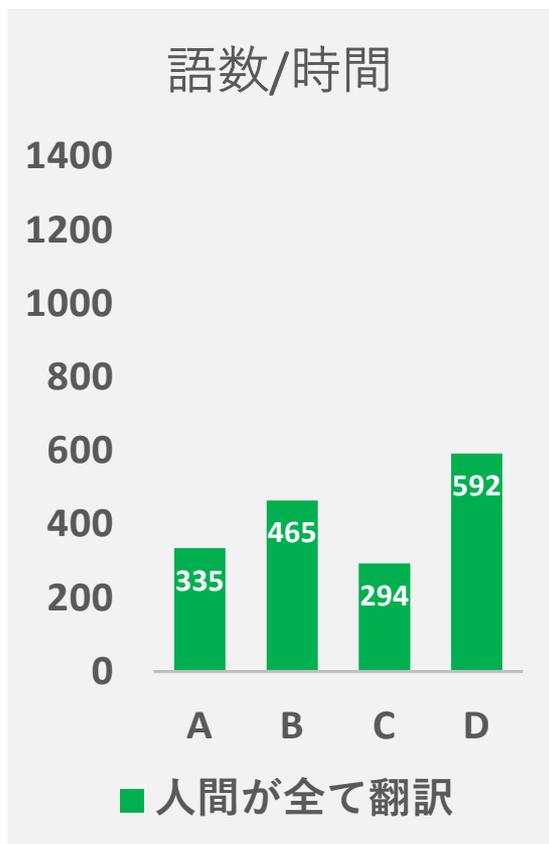
🗣️ 翻訳

More than 220 people have been killed and 843 injured after a tsunami hit coastal towns on Indonesia's Sunda Strait, government officials say. The tsunami waves struck at night without any warning, destroying hundreds of buildings. Officials say the tsunami could have been caused by undersea landslides after Anak Krakatau volcano erupted. The Sunda Strait, between the islands of Java and Sumatra, connects the Java Sea to the Indian Ocean.

インドネシアのスンダ海峡政府当局者によると、インドネシアの沿岸都市に津波が押し寄せた後、220人以上が死亡し、843人が負傷したという。津波は、何の警告もなく夜に襲ってきて、何100もの建物を破壊しました。アナク・クラカタウ火山噴火後の海底地滑りで津波が発生した可能性があるという。スンダ海峡は、Javaとスマトラの間にあり、ジャワ海をインド洋に接続している。

プロの翻訳者の効率も改善！

～人間 vs. (自動翻訳 + 人間)～



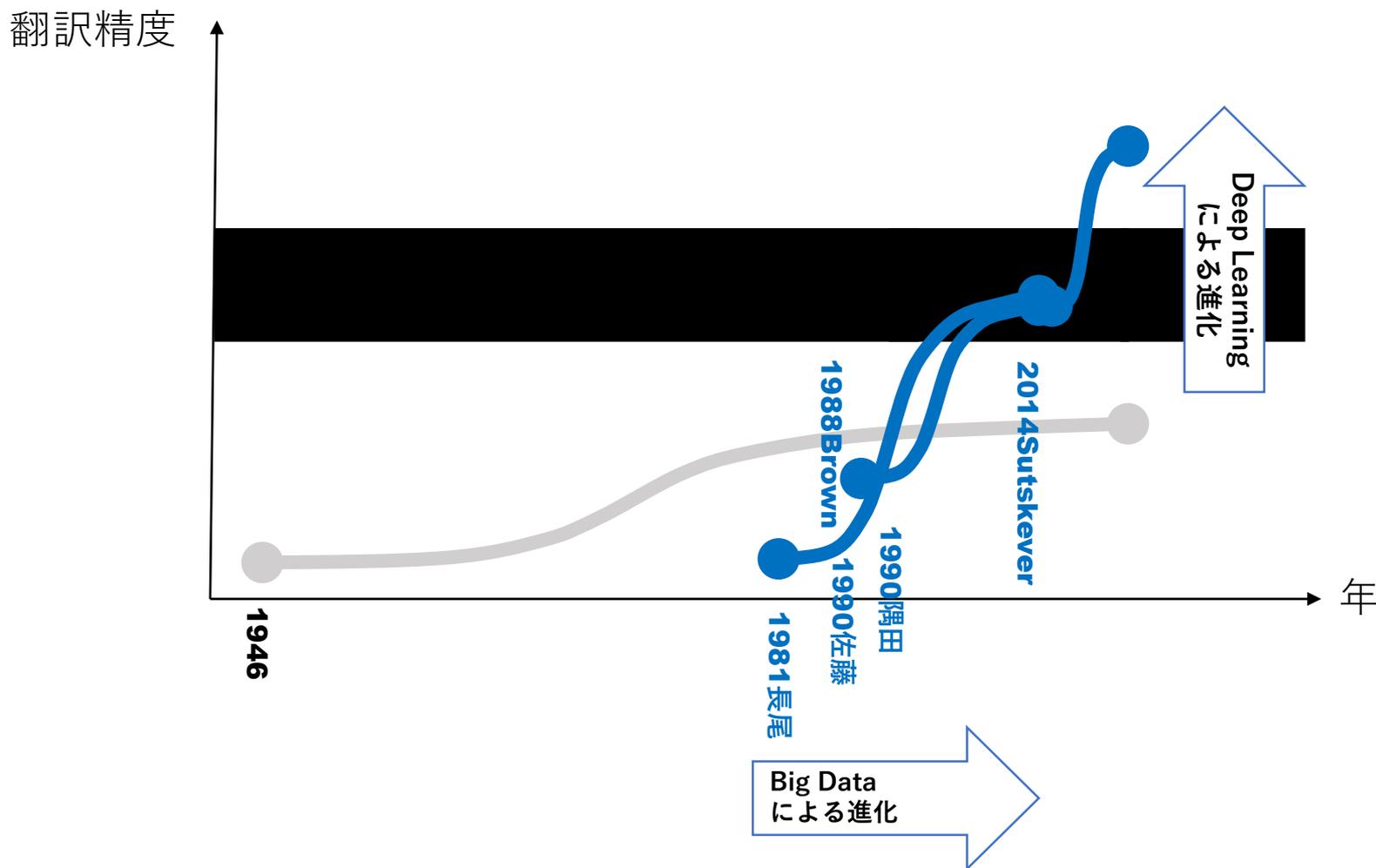
A、B、C、D、4人のベテラン翻訳者の協力の下、実験しました。

十印の三笠氏@JTF セミナー (2018/3/8) 資料から改変

②

自動翻訳は
典型的なAI
である。

自動翻訳の歴史



対訳 (Big Data) から深層学習した ニューラル機械翻訳 (NMT)

Big Data

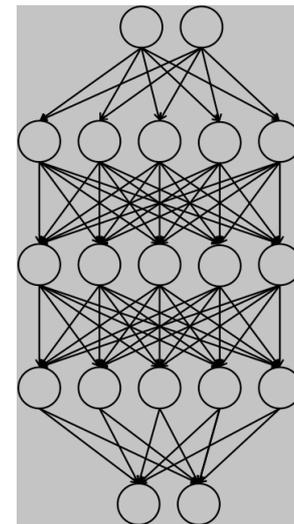
人類は、国家や地域、民族や世代など、あらゆる境界を越えて、相互の理解を深め、知恵を交わすなかで、発展してきました。コミュニケーションは人類社会を支えるもっとも重要な活動であり、情報通信技術はそのコミュニケーションを支える基礎であります。情報通信技術はまた、人類の高度な知的活動と経済活動を支える基盤でもあります。

Humanity has achieved progress as it has deepened its mutual understanding and shared its wisdom, overcoming barriers due to national, regional, ethnic, generational, and other differences. Communication is the most critical activity in human society, and information and communications technology (ICT) is the basis of that communication. ICT is also the infrastructure that supports humanity's advanced intellectual and economic activities.

<http://www.nict.go.jp/およびhttp://www.nict.go.jp/en/index.html>

深層学習

ニューラルネット



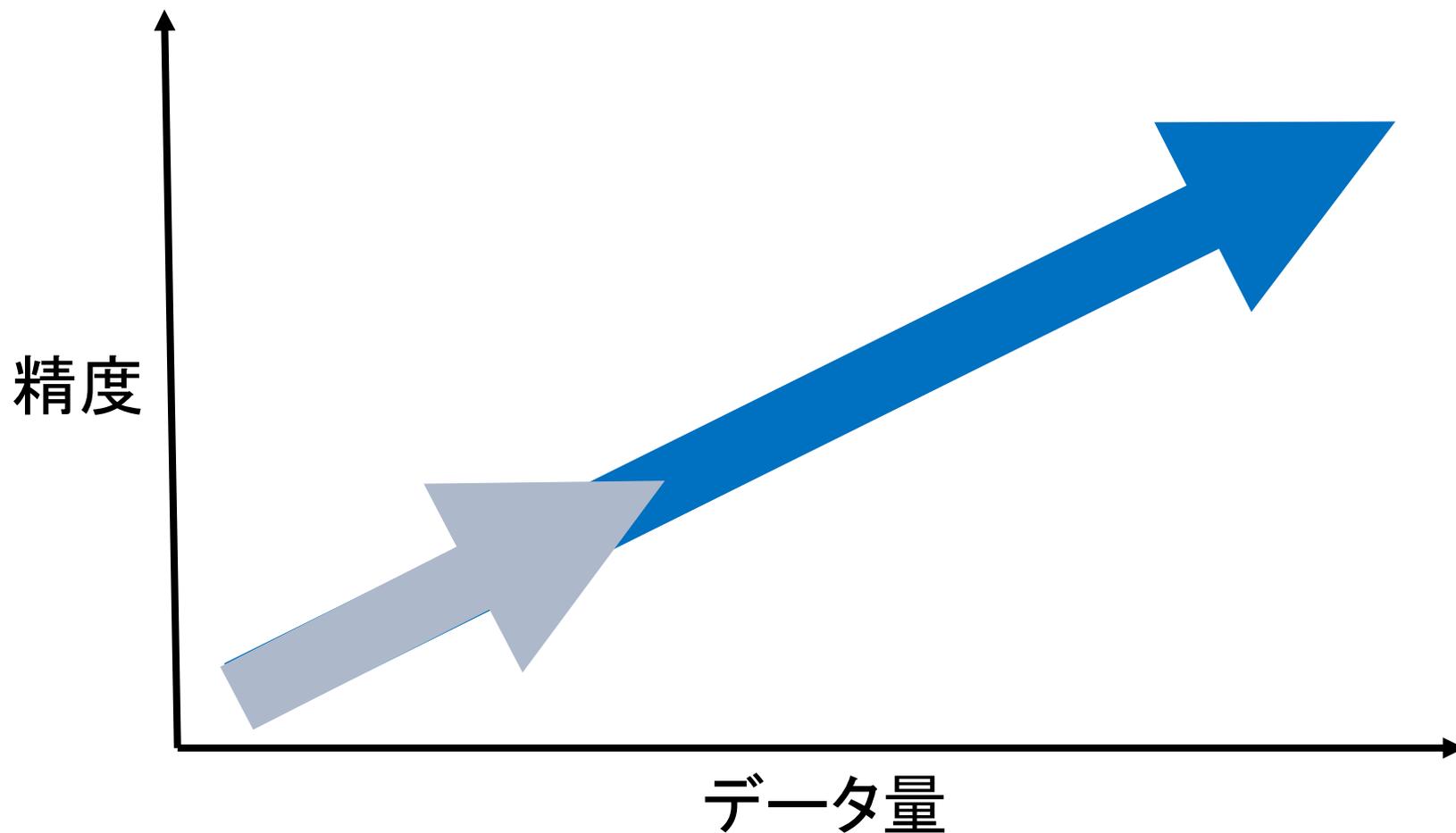
情報遺漏はない

翻訳精度は次式で決まる。

対訳データ x アルゴリズム

量 x 質

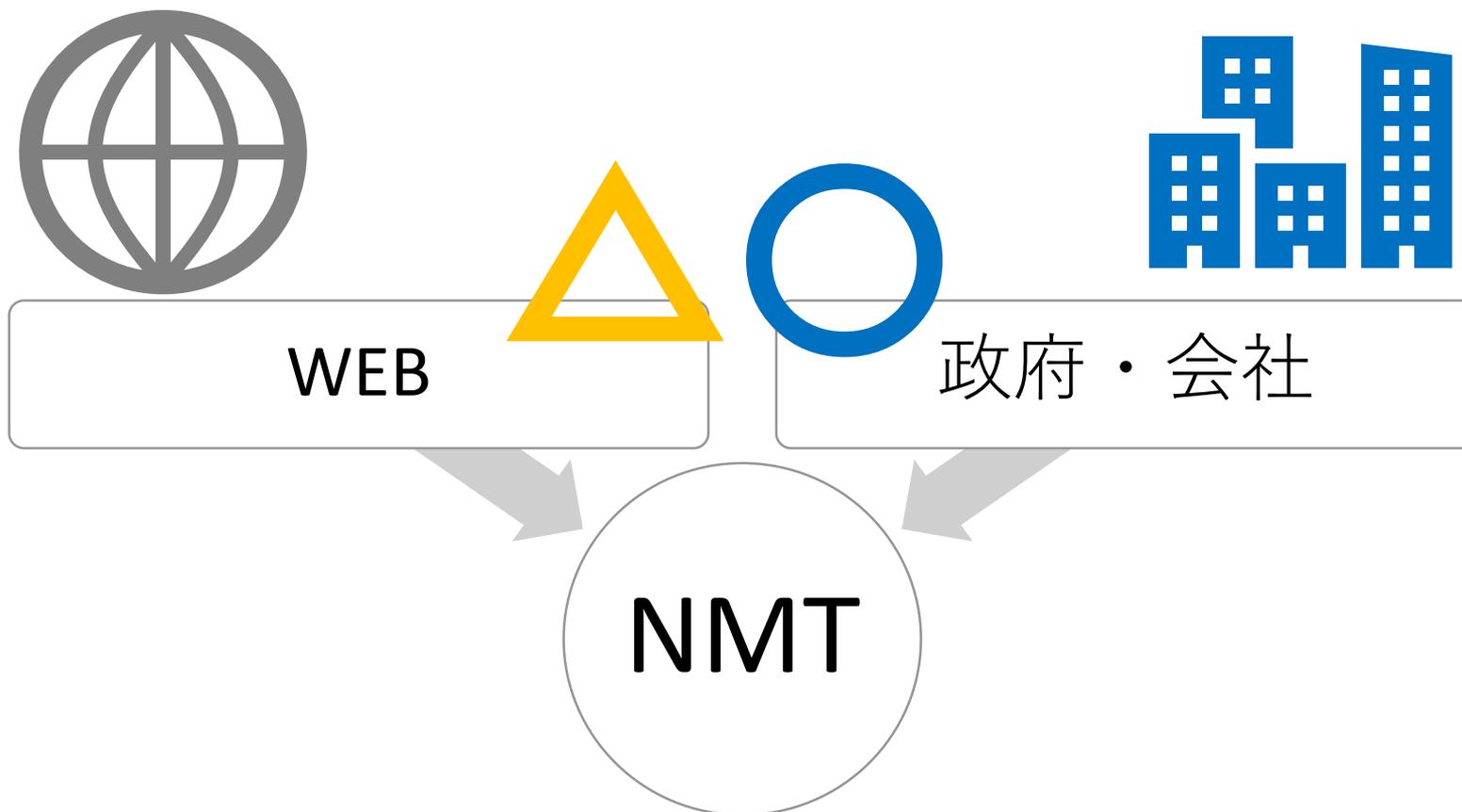
The more, the better.



量 * 質で決まるので**得手・不得手**が出る

日本語原文	二つのNMTの訳文	品質
電気がつきません	There is no electricity.	×
	The light doesn't work.	○
トイレが流れません	The toilet doesn't flow.	×
	The toilet doesn't flush.	○
地図上に行き方を書いてもらえませんか？	Can you write how to go on the map?	△
	Could you please write down how to get there on the map?	○

大量の良質データを集めたい



③

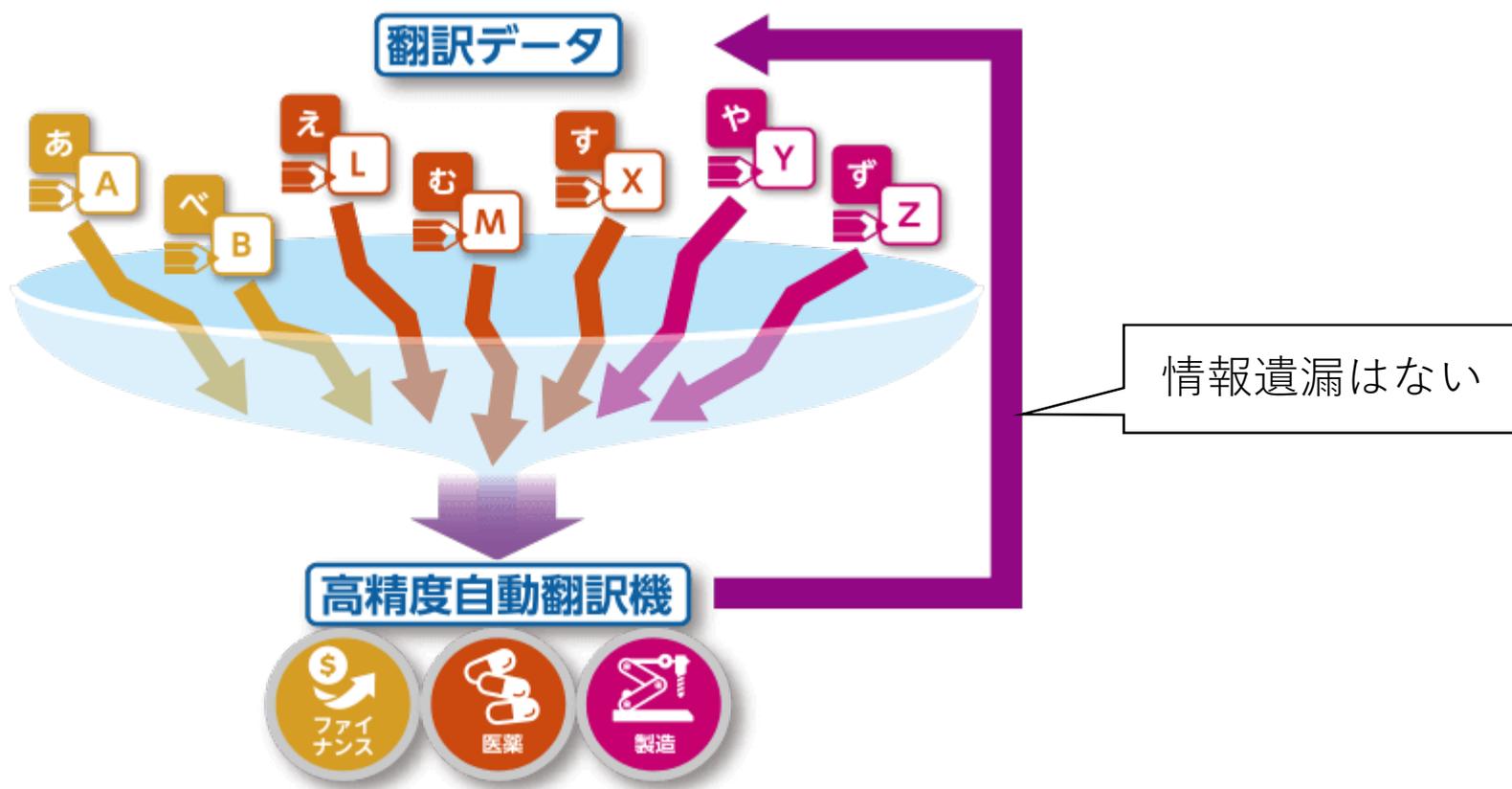
翻訳バンクで
大量の良質データを
蓄積し
システムを高精度化

大量の良質な翻訳データを 公的機関に集積

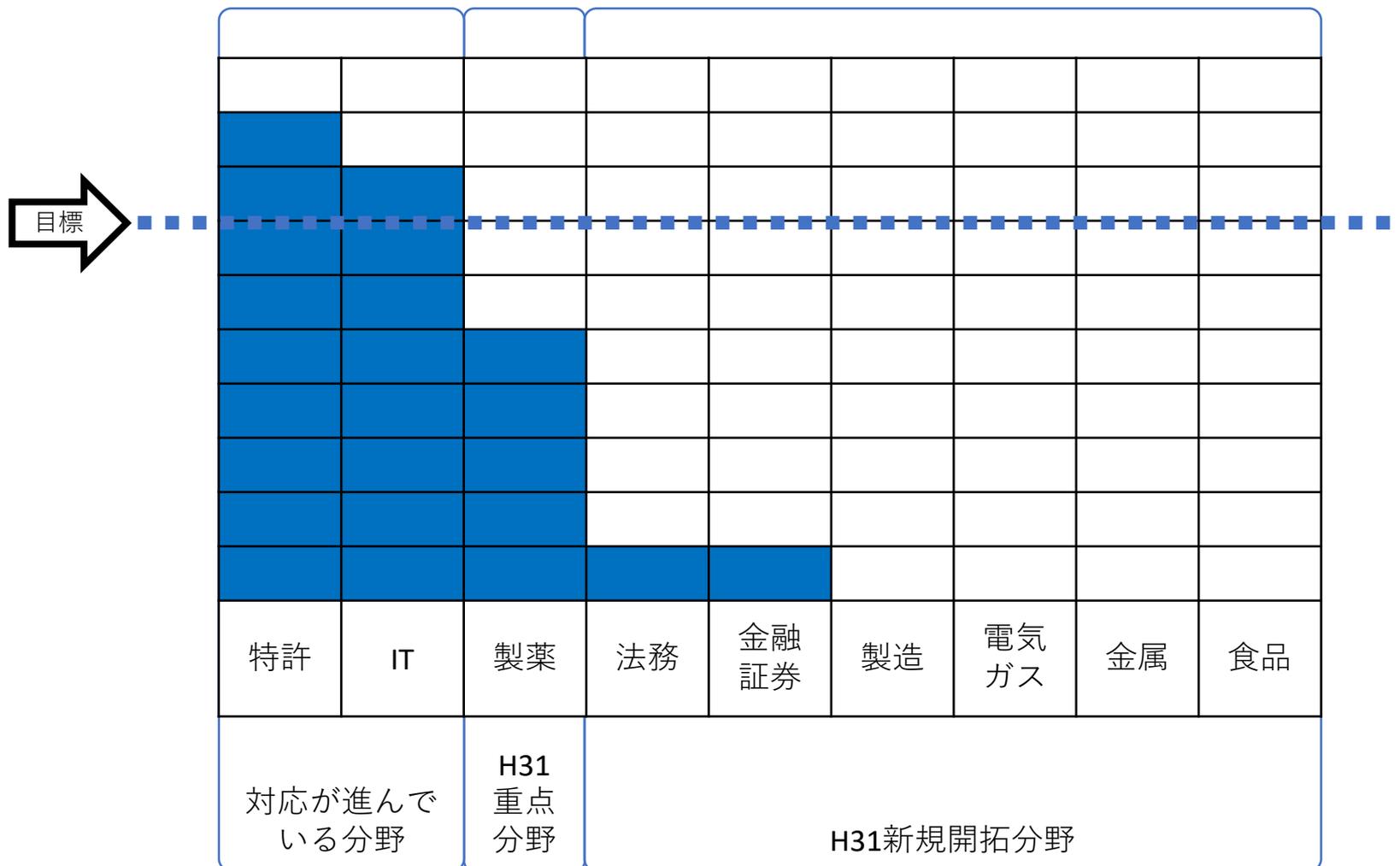
- 年間2000億円以上が翻訳に投じられている。
 - 5億文の翻訳が生産されていると推定。
- 当初目的に以外に使われることなく死蔵される。
 - 再利用で自動翻訳システム高精度化（新たな価値を生む。）



- ① 散在する**対訳**コーパスを一か所に集める。
- ② 高精度自動翻訳システムを作り配布。



各業界毎に最高精度の自動翻訳を構築



④

自動翻訳の
高精度化が
停まらな^い。

ポジティブ・スパイラル ～『翻訳バンク』という名のエコシステム～



当たり前前にしたい①



- 古紙
- ペットボトル
- 携帯電話中のレアメタル

再利用可能な保存を定着しよう！

現実

× 紙

× PDF

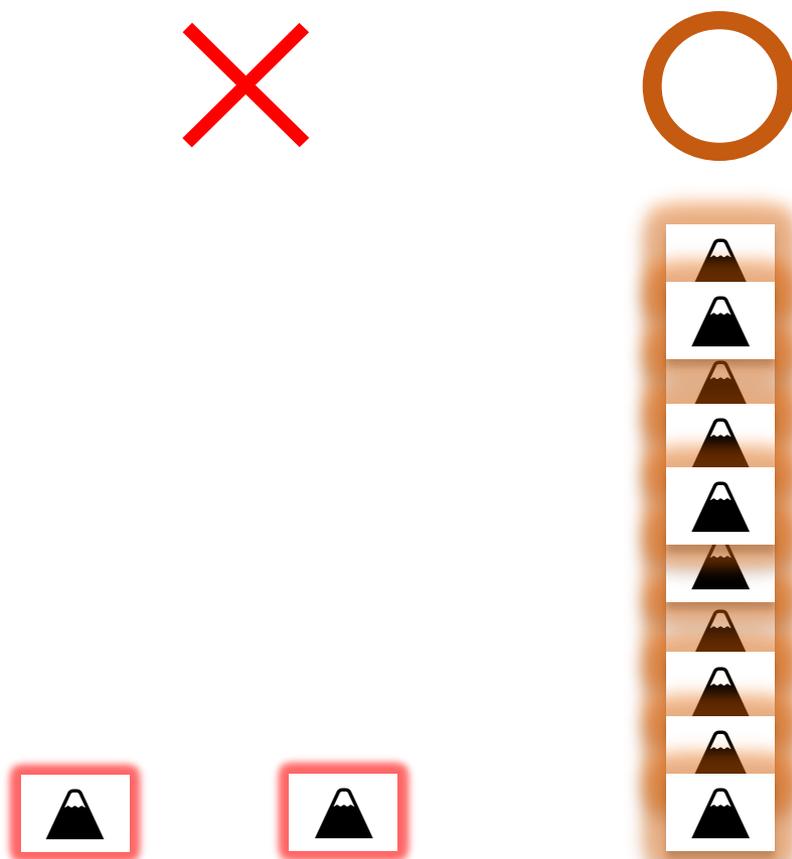
○ WORD

○ EXCEL

理想

◎ XML

最高精度の自動翻訳を作るベストな方法 ～競争でなく協調～



当たり前前にしたい②

- 翻訳の調達の仕様書に 1行加える

原文と翻訳文の対を翻訳バンクに提供。