2-3　自動同時通訳技術
2-3　Automatic Simultaneous Interpretation Technology

2-3-1　みんなの自動翻訳 @TexTra®
2-3-1　Min'na no Jidou Hon'yaku @ TexTra®

内山 将夫
UTIYAMA Masao

みんなの自動翻訳 @TexTra®は、情報通信研究機構で研究開発している自動翻訳エンジンのWeb サイトである。当サイトは、無料で利用可能（非商用限定）で、日本語・英語・中国語・韓国語の 4 言語を中心に多言語に対応しており、一般的な Web 翻訳と同様に、誰でも簡単に使用できる。当サイトは、翻訳バンクと連携して、翻訳バンクにご提供いただいた対訳を利用して、自動翻訳エンジンの高精度化を実現している。

"Min'na no Jidou Hon'yaku @ TexTra®" is a website where machine translation engines are researched and developed by the National Institute of Information and Communications Technology. The services on the website are free and available for non-commercial use. The machine translation engines on the site are mainly available for four languages, Japanese, English, Chinese, and Korean. The site also supports other languages. The site corporates with "Hon'yaku Bank" and uses the translation data provided to Hon'yaku Bank to improve the accuracy of the machine translation engines.

1　みんなの自動翻訳 @TexTra®の特徴

みんなの自動翻訳 @TexTra®（以下、みんなの自動翻訳）は、情報通信研究機構（NICT）で研究開発している自動翻訳エンジンの Web サイトである：

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

当サイトは、無料で利用可能（非商用限定）である。また、NICT の技術移転先企業では、当サイトと同等の商用サービスを提供している。

当サイトは、日本語・英語・中国語・韓国語の 4 言語を中心に多言語に対応しており、一般的な Web 翻訳と同様に、誰でも簡単に使用できる。

さらに、翻訳支援エディタをインストールなしで利用できるほか、Word や Ecxel などのオフィスソフトウェアのアドインが利用可能である。また、各種翻訳支援ツール（Trados 等）から自動翻訳エンジンを呼び出すことができ、自分のプログラムから直接 WebAPI を呼ぶこともできる。

このように、みんなの自動翻訳は、簡便に活用することができる。

2　みんなの自動翻訳の歴史

みんなの自動翻訳は、2014 年 6 月 19 日にオープンした。オープン当初は、統計的機械翻訳技術に基づく自動翻訳技術を提供していた。その後、2017 年 6 月にニューラル機械翻訳（NMT）を導入した。表 1 には、みんなの自動翻訳に関する NICT からの発表を中心にリストしている。これらはみんなの自動翻訳サイトから確認できる。

3　翻訳バンクによる NMT の高精度化

表 1 を時系列で確認すると、NMT の導入により、自動翻訳が一気に実用化する様子が分かると思う。

この実用化には、NMT だけでなく、「翻訳バンク」（https://h-bank.nict.go.jp/）が重要であった。「翻訳バンク」とは、外部機関からの寄付で NICT に翻訳データを集積し、自動翻訳の多分野化・高精度化を進める取組である。翻訳バンクでは、図 1 のように、NICT の自動翻訳技術の使用ライセンス料の算定の際に、提供が見込まれる翻訳データを勘案して負担を軽減する仕

表 1　みんなの自動翻訳の歴史

2014/6/19	第 9 回 AAMT 長尾賞を受賞
2014/6/19	*「みんなの自動翻訳@ TexTra®」を一般公開* [1]
2014/7/28	NICT と特許庁が多言語特許文献の高精度自動翻訳の実現に向けて協力合意
2015/3/30	科学技術文献データベースの作成に「高精度自動翻訳システム」を導入
2016/4/1	NICT と特許庁の特許文献の機械翻訳に関する協力の継続について
2017/1/18	高精度でセキュアな英文特許自動翻訳の提供開始
2017/6/28	*ニューラル機械翻訳で音声翻訳アプリ VoiceTra® が更なる高精度化を実現* [2]
2017/9/8	*『翻訳バンク』の運用開始－自動翻訳システムの更なる高精度化に向けて、様々な分野の翻訳データを集積－* [3]
2018/6/28	国立研究開発法人情報通信研究機構と MSD 株式会社 AI 多言語音声翻訳アプリ「VoiceTra®」(ボイストラ) に、医学事典「MSD マニュアル」の 10 言語翻訳データを活用することで合意
2018/7/26	“VoiceTra®” の音声翻訳技術が “POCKETALK® W” に採用
2019/4/23	自動車法規文の自動翻訳をニューラル技術で高精度化 ～トヨタとの共同研究を通じ、英日・中日翻訳の実用度が向上～
2019/9/30	IR・金融分野向け自動翻訳エンジンの性能向上を確認 ～日本財務翻訳との共同研究により実用化へ～
2019/10/7	大規模翻訳データによる製薬業界向け AI 自動翻訳システムの最適化 ～情報通信研究機構 (NICT) と R&D Head Club の共同開発～
2020/1/15	金融特化型 AI 自動翻訳システムを共同開発
2020/2/14	*第２回日本オープンイノベーション大賞総務大臣賞「ビッグデータで AI 翻訳を高精度化し翻訳産業に革命を起こす翻訳バンク」* [4]
2020/12/2	オープンソースのコミュニティに NICT「みんなの自動翻訳」を提供
2022/3/11	金融分野向けの高精度 AI 翻訳システムを開発 ～金融庁による翻訳文書の大量収集と、NICT による深層学習の連携で高精度化～
2022/4/27	情報セキュリティマネジメントシステムの国際規格 ISO/IEC27001 認証を取得 ～世界水準のデータ管理体制を構築し、音声翻訳技術の研究開発における「AI データ」収集を強化～

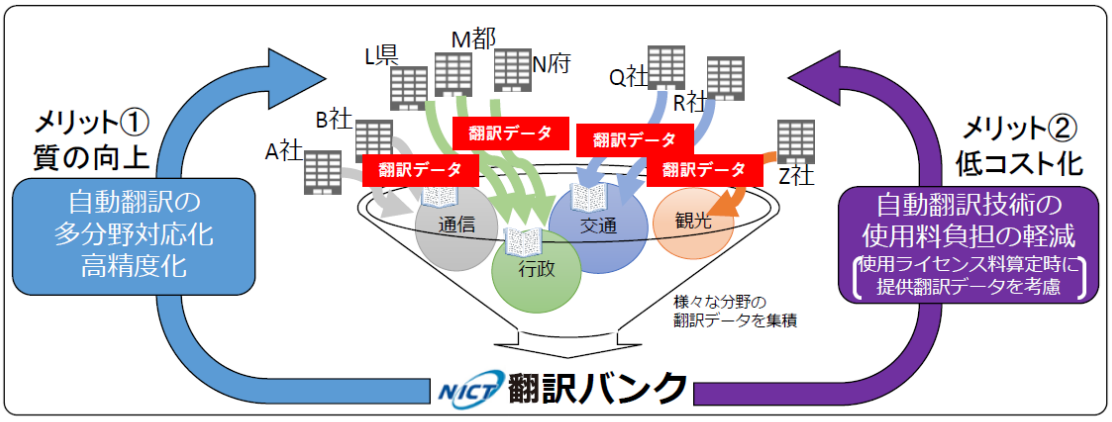

図 1　翻訳バンクの概要

組みも導入している。

NMT は優れた自動翻訳アルゴリズムではあるが、学習(訓練)データである対訳データがなければ自動翻訳エンジンはできない。NICT では、みんなの自動翻訳をオープンする前から、特許庁などと協力して、対訳データの収集に努めてきた。翻訳バンクはこの取組を加速するものである。

まず、高精度 NMT には、図 2 に示すように、対訳データが重要である。

次に、集積された対訳データを活用して、NICT の汎用 NMT を構築することにより、高精度な共通基盤としての NMT が実現可能である。さらに、「アダプテーション」という技術を適用することにより、さらに高精度な NMT を構築できると考えている。

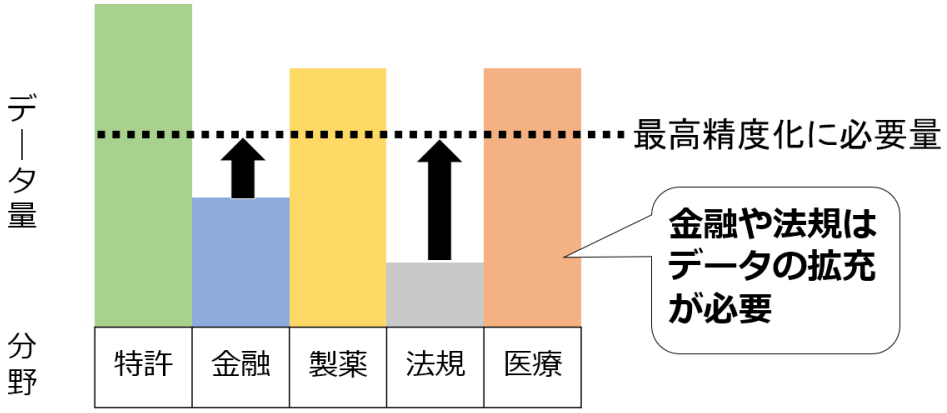

図２　分野ごとの対訳データ量

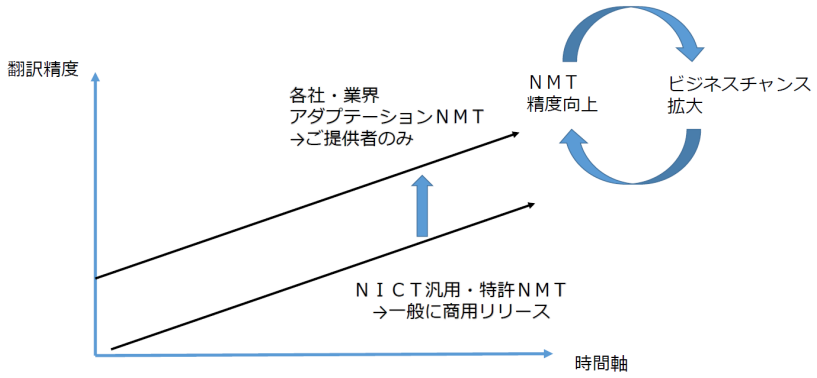

図３　NICT 汎用 NMT とアダプテーション NMT との関係

「アダプテーション」というのは、NMT においては、既に訓練された NMT を、翻訳対象の分野に応じて追加で訓練することである。

まず、NMT の訓練自体は、非常に大まかにいえば、「現時点の NMT で原文を翻訳してみて、その結果が参照訳文と異なる度合いに応じて、NMT のパラメタを調整すること」で実施される。この過程を大量の対訳文に対して何回か繰り返すことにより、共通基盤となる NICT 汎用 NMT が構築される。

次に、上記パラメタ調整が完了した汎用 NMT に対して、翻訳対象分野の対訳文を利用して、同様なパラメタ調整を追加実施するのがアダプテーションである。

このように、アダプテーションでは、すでに学習済みの汎用 NMT パラメタを翻訳分野に応じて追加調整するので、比較的少量の対訳データで当該分野に対する高精度な翻訳エンジンを構築できる。

また、翻訳バンクに対訳データをご提供いただくことで、NICT 汎用 NMT の底上げが期待できるが、それに加えて、アダプテーション NMT については、対訳データご提供者のみが利用できる。そのため、NICT 汎用 NMT とアダプテーション NMT には、図３の関係が成り立つ。

このように、翻訳バンクでは、提供者が NMT の精度向上の恩恵を一番受ける枠組みとなっている。

4　最近の発展

NMT の研究は日進月歩で進んでいる。たとえば、これまで、タグを含む文を機械翻訳するのは困難であったが、最近の研究により、タグを含む文の翻訳精度が向上している。たとえば、みんなの自動翻訳では、オプションで「XML 翻訳を利用する」と設定することにより、タグを含む文の自動翻訳を精度よく実現することが可能である。タグを含む文章としては、HTML の文章が考えられるが、みんなの自動翻訳では、次のように翻訳された。

【原文】「アダプテーション」 というのは、NMT においては、既に訓練された NMT を、翻訳対象の分野に応じて追加で訓練することです。

【訳文】 "adaptation" means, in NMT, that an already trained NMT is additionally trained according to the field to be translated.

また、みんなの自動翻訳の Word アドインでは、本

機能を利用して、テキスト中における強調や斜体などの書体情報を保存しての自動翻訳が可能である。

NMT の技術を自動翻訳以外に活用することも可能である。みんなの自動翻訳では、同一言語内の翻訳としての観点から、長い文章を短くする短文化エンジンを公開している。短文化エンジンは、入力文を 75 % 程度の文字数に短文化するようになっている。これにより、議事録などの音声書き起こしのテキストを本エンジンにより、簡潔なテキストにすることができる。たとえば、次のように短文化される。

【原文】早速質問に入らせていただきたいと思います。
【短文】早速質問したいと思います。

さらに、VoiceTra® で対応している全言語について、全言語方向の自動翻訳を 1 モデルで実施できるようになった。そのため、本モデルを技術移転することにより、旅行会話を対象とした場合には、全言語方向について 1 モデルで自動翻訳が可能となりる。

これによるメリットは、起動する翻訳エンジン数が少なくなることである。VoiceTra® では 31 言語に対応しているため、もし 1 言語方向に 1 エンジンとすると、言語方向としては 31 × 30 言語方向あるので、900 エンジン以上が必要となり、翻訳エンジンのメンテナンスが困難になる。一方、1 モデルで全言語方向の自動翻訳に対応できることにより、1 エンジンで全言語方向の自動翻訳が可能になる。

5　今後の展開

NICT では、みんなの自動翻訳に最新の NMT 研究成果を展開すると同時に、翻訳バンクにより集積された対訳データを活用することにより、実用的で高精度な NMT を社会還元することを目標としている。自動翻訳エンジンの性能向上には、翻訳アルゴリズムの研究開発だけでなく、自動翻訳エンジンの訓練のための対訳データが非常に重要である。そのため、翻訳バンクへの対訳データのご提供について、皆様のご協力をよろしくお願いいたします。

【参考文献】

1　「みんなの自動翻訳@ TexTra®」を一般公開
https://www.nict.go.jp/info/topics/2014/06/140619-1.html
2　ニューラル機械翻訳で音声翻訳アプリ VoiceTra が更なる高精度化を実現
https://www.nict.go.jp/press/2017/06/28-1.html
3　『翻訳バンク』の運用開始
https://www.nict.go.jp/press/2017/09/08-1.html
4　内閣府：「第 2 回 日本オープンイノベーション大賞」受賞取組・プロジェクトの概要について
https://www8.cao.go.jp/cstp/openinnovation/prize/2020 taishogaiyo.pdf

内山 将夫（うちやま まさお）
ユニバーサルコミュニケーション研究所
先進的音声翻訳研究開発推進センター
先進的翻訳技術研究室
上席研究員
博士（工学）
機械翻訳
【受賞歴】
2020 年　内閣府 第 2 回日本オープンイノベーション大賞総務大臣賞
2016 年　電気通信普及財団 第 31 回電気通信普及財団賞（テレコムシステム技術賞）
2014 年　一般社団法人アジア太平洋機械翻訳協会 第 9 回 AAMT 長尾賞