

欧州におけるインタラクティブ・アダプティブ機械翻訳の動向

平成30年3月

国立研究開発法人 情報通信研究機構
(欧州連携センター)

目次

はじめに	2
要約.....	4
Summary.....	9
第1章 機械翻訳技術の市場動向.....	122
第1節 世界の機械翻訳市場の動向.....	133
第2節 欧州の機械翻訳市場の特徴.....	177
第2章 欧州におけるインタラクティブ・アダプティブ機械翻訳に関わるビジネスの普及の現 状と展望	19
第1節 インタラクティブ・アダプティブ機械翻訳とは何か?	20
第2節 リルトのインタラクティブ・アダプティブ機械翻訳.....	23
第3節 SDLのアダプティブ機械翻訳.....	28
第4節 インタラクティブ・アダプティブ機械翻訳の市場動向	311
第5節 インタラクティブ・アダプティブ機械翻訳に対する評価.....	355
第3章 欧州連合のホライゾン 2020 におけるアダプティブ・ニューラル機械翻訳システムの研 究開発動向 / MMT プロジェクトの動向	411
第1節 MMT プロジェクトの背景.....	422
第2節 MMT プロジェクトの概要.....	444
補足資料.....	488
SATT 2017 (School of Advanced Technologies for Translators) 視察レポート	488
TC ワールドカンファレンス 2017 視察レポート	588
META フォーラム 2017 視察レポート	666

はじめに

本報告書では、欧州におけるインタラクティブ・アダプティブ機械翻訳システムの動向について記す。

現在、機械翻訳分野では、統計的機械翻訳システムからニューラル機械翻訳システムへと、パラダイムの変換とも言われる機械翻訳のコア技術の刷新に大きく注目が集まっている一方で、「インタラクティブ・アダプティブ機械翻訳 (Interactive, adaptive MT)」、あるいは単に「アダプティブ機械翻訳 (Adaptive MT)」とも呼ばれる、新しい機械翻訳システムが実用化され、一定の評価を受けている。特に、新興企業であるリルト社 (米) と大手企業である SDL 社 (英) が同種のシステムの開発と商用化を進めていることが知られているが、欧州では欧州連合 (EU) の大型研究開発助成プログラムであるホライゾン 2020 の MMT (Modern MT) プロジェクトで、アダプティブ・ニューラル機械翻訳 (Adaptive neural MT) が開発されている。以上の 3 つのシステムに共通する技術的特徴は、機械翻訳エンジンと翻訳メモリ技術の融合であり、新しい自己学習型機械翻訳技術が導入されている。

3 つの新しい自己学習型機械翻訳技術

	機械翻訳システムの名称
リルト	インタラクティブ・アダプティブ機械翻訳 (Interactive, adaptive MT)
SDL	アダプティブ機械翻訳 (Adaptive MT)
MMT	アダプティブ・ニューラル機械翻訳 (Adaptive neural MT)

本報告書では、特にリルト、SDL、MMT プロジェクトにおける新しい自己学習型機械翻訳技術の研究開発及びビジネスの動向について記す。以上のため、第 1 章では、導入として機械

翻訳市場を概観する。第2章では、リルトインタラクティブ・アダプティブ機械翻訳と SDL のアダプティブ機械翻訳のビジネスの動向について示す。最後に、第3章では欧州固有の試みとして、ホライゾン 2020 の MMT プロジェクトの概要について記す。報告書末には、関係イベントを視察した際の視察レポートを補足資料として収録する。

注意事項

- 本報告書では、情報を入手したウェブサイトの URL を参考のため注に載せているが、これらの記事はウェブサイト管理運営者の判断で随時移動、修正、削除される可能性がある。したがって、本報告書の発表後、注に記された URL から情報源となった記事にアクセスできないことがありうる。
- 本調査では、関係者にインタビュー調査を実施したが、匿名性の保持のために名前や所属先を明記しない場合がある。

要約

- 本報告書では、まず、機械翻訳技術の市場を概観した後、リルトのインタラクティブ・アダプティブ機械翻訳、SDL のアダプティブ機械翻訳、そして、EU の MMT プロジェクトで開発されているアダプティブ・ニューラル機械翻訳について、概要、市場動向、評価などについて記す。

第1章 機械翻訳技術の市場の動向

世界の機械翻訳市場の動向

- タウスの『2017 年度機械翻訳市場レポート』（以下、『タウスレポートとする』）によれば、2017 年の世界機械翻訳市場は、約 1 億 3000 万ドルと見込まれている。
- 『タウスレポート』は、世界の機械翻訳市場の年間成長率を 6% とし、2020 年の市場規模を 2 億ドルと予測している。
- 『タウスレポート』は、ニューラル機械翻訳の登場を肯定的に捉えているものの、それが機械翻訳市場に非常に大きな影響を与えるとは考えていない。

欧州の機械翻訳市場の特徴

- 『機械翻訳における欧州の主導的役割』（EU のクラッカープロジェクトの調査報告書）によれば、機械翻訳に対する企業の需要の大部分は北米に由来している（77%）。北米企業が機械翻訳に対する需要を主導しているのは、IT 部門が伝統的に機械翻訳を積極的に利用しており、また、大規模の IT 企業が米国企業であるからである。
- 欧州は後編集機械翻訳（Post-edited machine translation : PEMT）システムの開発で主導的な役割を果たしている。世界で欧州の後編集機械翻訳供給企業数の割合は 56% であり、北アメリカは 26% である。

第2章 欧州におけるインタラクティブ・アダプティブ機械翻訳に関わるビジネスの普及の現状と展望

インタラクティブ・アダプティブ機械翻訳とは何か？

- 『タウスレポート』によれば、インタラクティブ・アダプティブ機械翻訳の本質的な機能はパラメータの適応機能にあり、翻訳者の後編集のフィードバック情報から、リアルタイムで学習し、翻訳を改善するという点にある。
- SDL とリルトの技術には共通点があり、実際、SDL がリルトを特許の問題で訴追していたが、両社は和解に至っている（2017年11月）。
- SDL とリルトの機械翻訳システムの他には、現在、EU の MMT プロジェクトでアダプティブ・ニューラル機械翻訳と呼ばれる機械翻訳が開発されている。
- 以上の3つのシステムに共通する技術的特徴は、機械翻訳エンジンと翻訳メモリ技術の融合であり、これらのシステムには新しい自己学習型機械翻訳技術が導入されている。

リルトのインタラクティブ・アダプティブ機械翻訳

- リルトは、米グーグル社で機械翻訳システムの開発を行っていたジョン・デニーロ（John DeNero）とスペンス・グリーン（Spence Green）が2015年にアメリカで設立したベンチャー企業である。
- インタラクティブ・アダプティブ機械翻訳システムの重要な要素は、リアルタイム・フィードバックループの実現である。それは、翻訳者が機械翻訳システムと相互作用し、機械翻訳システムが翻訳者による編集から連続的に学習し、翻訳の品質を向上させることにある。
- リルトの「メモリ（Memory）」とは、単なる翻訳メモリではなくて、翻訳メモリ、カスタム機械翻訳システム、用語ベースが結合したものである。

- リルトは、翻訳者が作業を楽しめるように、翻訳システムのインターフェイスの設計に多くの時間を使っている（直感的インターフェイスの実現）。
- リルトの機械翻訳システムはクラウドサービスで提供されており、同社のシステムを利用するには、何も PC にダウンロードする必要はなく、ブラウザからリルトのプラットフォームにサインインするだけで良い。

SDL のアダプティブ機械翻訳

- SDL 社は、1992 年にイギリスで設立された大手言語サービスプロバイダであり、機械翻訳システムの開発とともに、翻訳サービスを提供している。
- SDL は、翻訳者にとっての新しい重要な機械翻訳技術として、CAT (Computer-Assisted Translation) ツール、ニューラル機械翻訳、アダプティブ機械翻訳の 3 つを挙げており、アダプティブ機械翻訳を同社の重要な技術システムの 1 つとして考えている。
- SDL は、2016 年 11 月に SDL トラドス・スタジオ 2017 の提供を開始した。同 CAT ツール上で、アダプティブ機械翻訳エンジンを利用できる。CAT ツール自体はデスクトップアプリケーションであるが、アダプティブ機械翻訳エンジンはクラウドベースである。
- 現在、SDL のアダプティブ機械翻訳エンジンは統計的機械翻訳技術に基づくが、SDL のアダプティブ機械翻訳のホワイトペーパーによれば、今後、ニューラル機械翻訳技術を使って、アダプティブ機械翻訳の機能を実現することが目標とされている。

インタラクティブ・アダプティブ機械翻訳の市場動向

- リルトの顧客には翻訳・ローカライゼーション企業が多い。リルトの機械翻訳エンジンは他社の CAT ツール上で使うことができ、リルトは他社と対立するだけでなく、提携することも可能である。

- 以下が、インタビューした関係者の意見の中で注目すべき点である。
 1. リルトのインタラクティブ・アダプティブ機械翻訳と SDL のアダプティブ機械翻訳の詳しいユーザ数や市場規模は判明ではないが、まだユーザは多くなさそうである。
 2. これらのシステムに興味がある人は多く、また技術としてとても優れているので、今後、ユーザは増加する可能性が高い。
 3. SDL のアダプティブ機械翻訳に関しては、同社のトラドス・スタジオ 2017 に標準装備されているので、潜在的なユーザは多い。

インタラクティブ・アダプティブ機械翻訳に対する評価

- インタビュー調査では、リルトと SDL のシステムについて、肯定的な評価をする人が多いが、問題点を指摘する人もいた。
 - ・ 肯定的評価
 - 翻訳の品質向上
 - 翻訳速度の上昇
 - ・ 否定的評価
 - ユーザが間違った修正を行い、翻訳エンジンが誤って学習してしまうかもしれない。
 - 文体に関わる偏った修正を過度に与えすぎてしまい、翻訳エンジンの性能を減少させることがあるかもしれない。
 - 機械翻訳企業がユーザの後編集データを自社の機械翻訳エンジンの改善のために利用している可能性がある。
 - この最後の点に関しては、関係者間で意見が分かれた。ユーザの観点から、翻訳者にはこのような疑いを持つ人がいるが、技術的な観点から、機械翻訳の技術者は、フィ

ードバックデータの利用には費用がかかりすぎるので現実には不可能であるとしている。

第3章 欧州連合のホライゾン 2020 におけるアダプティブ・ニューラル機械翻訳システムの研究開発動向 / MMT プロジェクト

- 欧州では、EU のホライゾン 2020 によって助成される MMT (Modern MT) プロジェクト (研究期間 : 2015 年 1 月~2017 年 12 月、予算 : 369 万 5200 ユーロ[EU 拠出分は 299 万 4700 ユーロ]、コーディネータ : 伊トランスレイティッド社、参加組織 : 伊 FBK、英エジンバラ大学、蘭タウス) において、アダプティブ・ニューラル機械翻訳システムが開発されている。
- 2017 年 11 月、MMT のアダプティブ・ニューラル機械翻訳は、CAT ツールであるメイトキャット上で利用可能になった。
- MMT の技術的な目標の 1 つは、機械翻訳システムにユーザの翻訳メモリと後編集から自己学習する機能を与えることである。ユーザが後編集すると、その情報が MMT エンジンにフィードバックされ、リアルタイムで学習し、翻訳文の予測を改善する。
- MMT の アダプティブ・ニューラル機械翻訳は、ニューラル機械翻訳エンジンと翻訳メモリを結合することによって実現される。
- MMT エンジンを利用するには、翻訳する語の分量に応じて、ライセンスを購入する必要がある。
- MMT エンジンの API (Application Programming Interface) は公開されており、メイトキャット以外の CAT ツール上でも利用可能である。

Summary

This report describes the market trends and evaluation of Lilt's Interactive, adaptive MT, SDL's Adaptive MT, and Adaptive neural MT developed in the EU's MMT project, with an overview of the market of machine translation technology.

Chapter 1 Market overview of machine translation technology

Global Machine Translation Market

- The global machine translation market is estimated at about 130 million dollars in 2017 (according to the "Taus Machine Translation Market Report 2017" ["Taus report"]).
- The annual growth rate of the global machine translation market is estimated at about 6%, and its market size will be 200 million dollars in 2020 (according to the "Taus report").
- Although the "Taus report" has considered the appearance of neural machine translation to be positive for machine translation market, however, it has not believed that this new technology will have a major impact on the market.

Characteristics of the European machine translation market

- Most of the enterprise's demand for machine translation come from North America (77%). North American companies are leading in the demand for machine translation, because the IT department traditionally uses machine translation, and the largest IT companies are US companies (according to the EU CRACKER project's survey report, "Europe's leading role in machine translation" [Cracker report]).
- Europe plays a leading role in the development of Post-Edited Machine Translation (PEMT) systems. The percentage of European post-editing machine translation suppliers in the world is 56%, the percentage of North American suppliers, 26% (according to Cracker project).

Chapter 2 Current status and prospects of the business related to interactive, adaptive machine translation in Europe

What is interactive and adaptive machine translation?

- What is "interactive and adaptive machine translation"? "Taus report" considers Lilt's MT system (called Interactive, adaptive machine translation) and SDL's MT system (called Adaptive MT) as same kind of MT system, which has a self-learning function from translator's feedback in real time for improving translation prediction.
- There are technological common points between these two systems. Indeed, SDL prosecuted Lilt for patents, and they come to be reconciliated (in November 2017).
- Apart from SDL and Lilt, the same kind of MT system called "Adaptive neural machine translation" was developed in EU's MMT project.
- The convergence between machine translation engine and translation memory is a common character in these systems from a technological point of view. Interactive and adaptive machine translation is a new self-learning machine translation.

Lilt's Interactive, adaptive MT

- Lilt is a venture company founded in 2015 by John DeNero and Spence Green who had developed machine translation systems at Google.
- An important element of Interactive, adaptive machine translation consists in realization of real-time feedback loop. Namely, by interacting with translator, machine translation system continuously learns from the post-editing by translator to improve the quality of translation.
- Lilt's "Memory" is not just a translation memory. It combines translation memory, custom machine translation system and termbase.
- Lilt took great care to design the interface of its translation system, so that translators can enjoy their work (realizing intuitive interface).
- For using Lilt MT system, it is necessary to sign in Lilt platform from browser without downloading anything to the PC (cloud service).

SDL's Adaptive MT

- SDL is a major MT and language service provider founded in the UK in 1992. It develops machine translation system and gives translation services.
- For SDL, current important machine translation technologies for translators are CAT tool, neural machine translation and adaptive machine translation. So, it considers adaptive machine translation as one of the most important element of its machine translation system.
- SDL has started to offer SDL Trados Studio 2017 in November 2016. Adaptive MT engine is integrated in this new Trados Studio. Anyone who has a license for it can use Adaptive MT engine on this CAT tool. This CAT tool itself is a desktop application, but the Adaptive MT engine is cloud-based.
- Currently, SDL's Adaptive MT engine is based on statistical machine translation technology. However, according to SDL's Adaptive MT white paper, it aims to realize adaptive function by using neural machine translation technology.

Market trend of interactive and adaptive machine translation

- The most of Lilt's customers are translation and localization companies. It is noteworthy that Lilt's machine translation engine can be used on other companies' CAT tools, and Lilt can collaborate them.
- There are three important points in stakeholders' opinions.
 1. There aren't exact data of the number of user or the market size of Lilt's Interactive, adaptive MT and SDL's Adaptive MT, but they still are not likely to be large.
 2. Many people are interested in these systems, and their technologies look like very good, so users are likely to increase in future.
 3. SDL's Adaptive MT is integrated in SDL's Trados Studio 2017, so there are many potential users.

Evaluation of Interactive and adaptive machine translation

- In our interview survey, many people appreciated Lilt and SDL systems, but others pointed out problems.

- Positive opinions :
 - Improving translation quality.
 - Rising translation speed.
- Negative opinions :
 - If users incorrectly make corrections, translation engine may erroneously learn.
 - If users give too many corrections related to style, they may reduce the performance of translation engine.
 - Machine translation companies may use users' post edit data for improving their own machine translation engines.
 - Regarding this last point, opinions are divided among stakeholders. From user's point of view, translator may have such a suspicion, but from technological point of view, machine translation engineer denies it because it costs too much to use the feedback data.

Chapter 3 The R & D of Adaptive neural MT at the EU's Horizon 2020 / MMT Project

- In Europe, Adaptive neural MT was developed in MMT (Modern MT) project financed by the EU horizon 2020 (Research period: 1/2015-12/2017, Budget : 3 695 200 euro [the EU contribution: 2 994 700 euro], Coordinator : Translated [Italy], Participating organizations: FBK [Italy], University of Edinburgh [UK], TAUS [Dutch]).
- The machine translation system developed by the MMT project is called Adaptive neural MT.
- In November 2017, Adaptive neural machine translation of MMT become available on the CAT tool, Matecat.
- One of the technical goals of MMT consists in giving to machine translation system a self-learning function from user's translation memory and post-editing. After user's post-editing, these data are fed back to the MMT engine to learn in real time from them for improving the prediction of translation.
- In order to use the MMT engine, users need to purchase a license and pay as much as they use MMT.
- The API of the MMT engine is publicly available, so it can also be used on CAT tools other than Matecat.

第1章 機械翻訳技術の市場動向

- 第1章では機械翻訳技術の市場の動向を概観する。まず、タウス（オランダ）の『2017年度機械翻訳市場レポート』（以下、『タウスレポート』とする）に基づいて世界の機械翻訳市場を確認し、ついで、EUのクラッカー（CRACKER）プロジェクト¹が発表した調査報告書、『機械翻訳における欧州の主導的役割』（以下に、『クラッカーレポート』とする）に基づいて、欧州市場の特徴について記す²。

¹ クラッカープロジェクトは、直接研究開発を行うのではなく、機械翻訳関連の研究機関の提携、研究者のネットワーク作り、過去のEUプロジェクトの成果の延長を支援するプロジェクトである。全予算は、99万9995ユーロ（全額をEUが助成）、期間は2015年1月から2018年1月までの36ヶ月間、コーディネータはドイツ人工知能センターである。

² <https://www.taus.net/think-tank/reports/translate-reports/taus-machine-translation-market-report-2017>

第1節 世界の機械翻訳市場の動向

世界の機械翻訳市場の規模

- 『タウスレポート』は、2017年の世界機械翻訳市場を約1億3000万ドルと見積もっている。
タウスは2014年に、数年中に市場規模が2億5000万ドルまで成長すると予測していたが、これは予測を下回る結果であった。
- タウスによれば、この結果は、機械翻訳は「ソフトな (soft)」パワーとして変化をもたらす野心的な役割を担っているとはいえ、市場規模はまだ小さいことを示している。
- 『タウスレポート』では、世界の機械翻訳市場の年平均成長率を6%とし、2020年の市場規模を2億ドルと予測している。

機械翻訳技術供給企業の分類

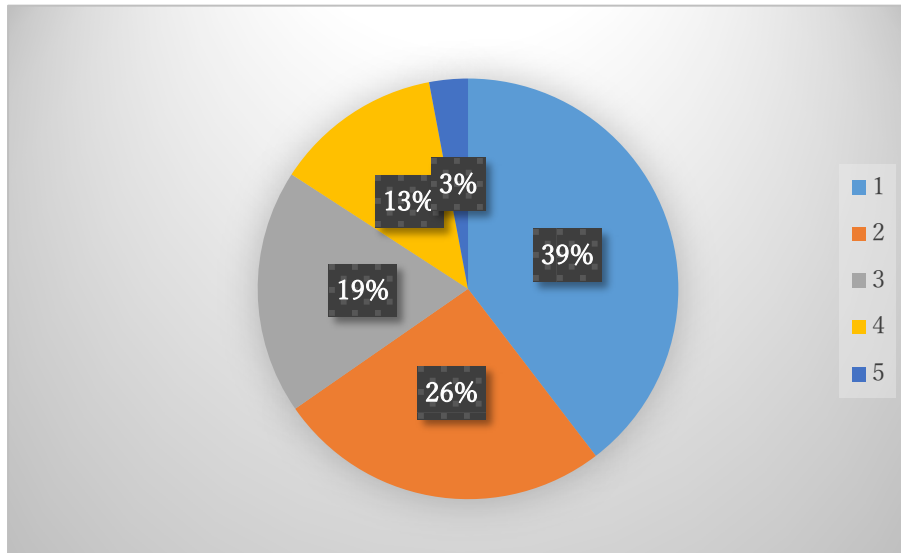
- 『タウスレポート』は、機械翻訳産業を分析するために5つの機械翻訳供給企業のグループを特定している。
 1. 少数のIT企業：米グーグル、米アマゾン、米マイクロソフト、露ヤンデックス、米フェイスブック、中バイドゥ：すべての人、あるいは顧客に無料機械翻訳サービス、あるいは有料サービスの提供：市場シェア率：39%³
 2. 純粋な機械翻訳プレーヤー：韓シストラン、アイルランド・カンタン MT、日本・みらい翻訳、シンガポール・OMNISCIENT、スペイン・TAUYOU、アイルランド・イコニックトランスレーションズ：市場シェア率：26%
 3. 顧客に応じて、機械翻訳を利用する言語サービスプロバイダ：英SDLなど：19%
 4. 他の言語技術と共に、企業向けに機械翻訳システムを販売している言語技術プロバイダ：

³ 『タウスレポート』では、40%と記されているが、5グループの割合を合計すると101%になってしまうので、便宜上、39%とした。

米スマートコミュニケーション、仏レヴェルソなど : 13%

5. 広告支援オンライン機械翻訳サービス : 米ワールドリンゴ、イスラエル・バビロン : 3%

図版 5 グループの市場シェア率



市場の集中

- 機械翻訳市場の 20%のプレーヤーが全収入の 80%以上（1 億ドル以上）を占め、多くの企業は 100 万ドル以下の収益である。
- この 20%以上のプレーヤーには、グーグル、シストラン、米レイセオン・BBN（米政府との大規模契約による）、マイクロソフト、SDL、スマートコミュニケーションズ、IBM、バビロン、米ライオンブリッジが含まれる。

機械翻訳ビジネスの動向

- 『タウスレポート』は、機械翻訳ビジネスの動向について、以下の 2 点を指摘している。
 1. この 3、4 年間、1 年の市場成長率は 6% 以下である。機械翻訳によって翻訳された語の数は増加しているが、機械翻訳市場に際立った牽引役が現れる見込みは少ない。顧客の価格低下の要望は根強い。また、顧客によって、機械翻訳は常に需要に適したソリューションとみなされているわけではない。顧客は翻訳が確実に効果的になされることを望んでおり、

機械翻訳が利用されなくとも構わない。

2. ニューラル機械翻訳がより大量、そして、低価格の翻訳への需要に対する技術的な回答になりうる可能性がある。

ニューラル機械翻訳の市場への影響

- 『タウスレポート』は、ニューラル機械翻訳の市場に対する影響について、以下の3つの動向を指摘している。
 1. 純粋な機械翻訳プレーヤーに属するシストランがニューラル機械翻訳を積極的に促進し、オープンソースの技術サポートを提供している。今後、多くの純粋な機械翻訳プレーヤーが、機械翻訳分野の専門家として競争力を保つためにニューラル機械翻訳を実用化していく。ニューラル機械翻訳の採用は、研究機関による新しい開発とオープンソースソフトウェアの広い利用可能性によって促進される。
 2. 大規模コンテンツの所有者（大手IT企業、ニュース・エンターテインメントメディア企業、動画・映画・ゲーム開発企業、金融・ビジネス情報企業、コンサルティング企業、官公庁、欧州委員会など）も翻訳されたコンテンツを使って、新しい市場を切り開く必要がある。これらの企業はニューラル機械翻訳に対応するローカライゼーション企業を買収する可能性がある。
 3. 多くの言語サービスプロバイダは、ニューラル機械翻訳の正確な利益と費用に確信が持てず、ニューラル機械翻訳が実際に全面的に翻訳の品質を改善するか、長期間の投資の実行可能性を確保するかどうか検討している。このような新しい技術に対する躊躇は、機械翻訳産業全体へわずかに否定的な影響を与えている。
- 『タウスレポート』は、ニューラル機械翻訳の登場を肯定的に捉えているものの、それが機械

翻訳市場に対して非常に大きな影響を与えるとは考えていない。

- 『タウスレポート』は、後述するように⁴、インタラクティブ・アダプティブ機械翻訳、あるいはアダプティブ機械翻訳に言及しているものの、これらのシステムの機械翻訳市場に対する影響については触れていない。

⁴ 本報告書第2章第1節を参考のこと。

第2節 欧州の機械翻訳市場の特徴

- 欧州の機械翻訳市場はどのような特徴を持つのか。欧州連合（EU）のホライゾン 2020 に助成されているクラッカー（CRACKER）プロジェクトは、2016年に欧州の機械翻訳市場動向に係る調査報告書、『機械翻訳における欧州の主導的役割』（以下に、『クラッカーレポート』とする）を発表している。以下に、同レポートに基づいて、世界の機械翻訳市場における欧州の特徴を記す。

機械翻訳の需要

- 『クラッカーレポート』は、機械翻訳の需要について、以下の点を指摘している。
 - ・ 機械翻訳に対する企業の需要の大部分は北米に由来している（77%）。欧州地域はだいぶ離れて2位に位置づく（19%）。
 - ・ 機械翻訳を利用している企業のうち、IT企業が63%を占め、サービス業が21%、非IT製品企業が16%を占める（世界の地域を問わず）。
 - ・ 機械翻訳を利用している企業のうち、9億ドル以上の年間収益がある企業は23%であり、9億ドル以下が77%である。
- 北米企業が機械翻訳への需要を主導しているのは、IT部門が伝統的に機械翻訳を主導して利用しており、大手IT企業が米国企業であるからである。

欧州の特徴

- 『クラッカーレポート』は、欧州の機械翻訳市場の特徴として、以下の点を指摘している。
 - ・ 欧州は、後編集機械翻訳（Post-edited machine translation : PEMT）システムの開発で主導的な役割を果たしている。世界で欧州の後編集機械翻訳供給企業数の割合は56%であり、北米は26%である。

- 『クラッカーレポート』によれば、以上のデータは、欧州の言語サービスプロバイダが伝統的に言語産業を支配している事実を反映している。
- 米 IT 企業の機械翻訳に対する需要は高いが、供給側に欧州の存在感が高いことは、ローカライゼーションにおいて英語への翻訳のパターンが長く確立されていることを反映している。
- 欧州企業にとって、機械翻訳は、翻訳一般と同じく輸出市場である。欧州企業は特に欧州で事業を行おうとする他国の企業、特に米国企業へ販売している。
- 以上のように、『クラッカーレポート』は、世界市場における欧州の機械翻訳産業の特徴を後編集機械翻訳企業の積極的な活動に見出している。

第2章 欧州におけるインタラクティブ・アダプティブ機械翻訳 に関わるビジネスの普及の現状と展望

- 本章では、特にリルトのインタラクティブ・アダプティブ機械翻訳システムと SDL のアダプティブ機械翻訳システムの動向について記す。まず、これらのシステムの共通性を確認した後、リルトと SDL のシステムについて記す。ついで、これらのシステムの市場動向、評価及び問題点について記す。

第1節 インタラクティブ・アダプティブ機械翻訳とは何か？

リルトのインタラクティブ機械翻訳と SDL のアダプティブ機械翻訳

- 「インタラクティブ・アダプティブ機械翻訳 (Interactive, adaptive MT)」は、具体的にどのような機械翻訳システムなのか。以下に、タウス (TAUS) の『2017 年度機械翻訳市場レポート』⁵ (以下、『タウスレポート』とする) における説明を示す。
- 『タウス 2017 レポート』の第 6 章第 3 節の「翻訳環境におけるアダプティブ MT」では、リルトのインタラクティブ・アダプティブ機械翻訳と SDL のアダプティブ機械翻訳が一緒に取り扱われ、同種の技術とみなされている。
- 『タウスレポート』は、現在人気があるアダプティブ機械翻訳とインタラクティブツール環境として、リルトと SDL トラドス・スタジオ 2017 を挙げている。確かに、リルトは自社の機械翻訳システムを「インタラクティブ・アダプティブ機械翻訳 (Interactive, Adaptive MT)」⁶と呼ぶ一方で、SDL は「アダプティブ機械翻訳 (Adaptive MT)」と呼ぶという名称の違いはある。だが実際のところ、基本的な点は共通し、どちらも「アダプティブ機械翻訳」というカテゴリーに属すると、『タウスレポート』では考えられている。
- 「アダプティブ機械翻訳」とは何か。『タウスレポート』は、以下のように定義している。

「アダプティブ機械翻訳 (adaptive MT) とは、翻訳者のフィードバックに基づいて、そのパラメータを適応させる (adapt) 最近開発されたシステムである。したがって、それは翻訳者の編集からリアルタイムで「学習 (learn)」し、改善し (improve)、後に行う翻訳と後編集の労力を減らす」⁷
- 以上の引用からも分かる通り、アダプティブ機械翻訳の本質的な機能、パラメータの適応機能

⁵ <https://www.taus.net/think-tank/reports/translate-reports/taus-machine-translation-market-report-2017>

⁶ <https://lilt.com/businesses>

⁷ 『タウス 2017 レポート』 P.13. 原文は以下の通り。

「Adaptive MT is a recently-developed system that adapts its parameters based on translators' feedback. It therefore "learns" and improves in real-time from the translators' edits, reducing subsequent translation and post-editing effort」

とは、翻訳者のフィードバックから、リアルタイムで学習し、翻訳を改善するという点にある。

- 『タウスレポート』は、アダプティブ機械翻訳と同様の機能を実行する技術として、「機械翻訳向けオンライン学習 (Online learning for MT)」と「インクリメンタル MT トレーニング (incremental MT training)」を挙げている。
- 他方で、SDLのウェブサイトでは、アダプティブ機械翻訳と翻訳メモリの違いについて、以下のように説明されており、前者は従来の翻訳メモリとは異なる技術であることが分かる⁸。
 - ・ 「翻訳メモリ (TM) は、翻訳を保存し、同様の内容を再度翻訳する必要がある場合に再利用できます。TM は完全一致またはあいまい一致を提供しますが、これらは承認済みの翻訳に関する類似性が基になっています。AdaptiveMT は、TM に一致する翻訳がない場合に使用するように設計されており、ポストエディットの結果を学習し、ユーザの用語集、スタイル、構文、コンテンツに沿った独自の MT 出力を提供します。よって機械翻訳 (MT) のミスは何度も修正する必要がなく、TM に一致する翻訳がないときの MT ポストエディット作業時間が大幅に削減されます」
- 機械翻訳技術のコンサルタントであるゲール・ヴォン・アッシュ氏⁹によれば、SDLでの開発当時は、まだアダプティブ機械翻訳システムを実用化することが技術的に難しかったので、SDLは同システムを商用化できなかったが、その後、リルトが現れ、類似するインタラクティブ・アダプティブ機械翻訳システムの商用化に成功した。そして、リルトの後、SDLもアダプティブ機械翻訳システムの開発を進め、近年商用化に至っている。
- したがって、SDL とリルトの技術には共通点があり、実際、SDL がリルトを特許の問題で訴追している。

⁸ <https://blog.sdltrados.com/adaptivemt-sdl-trados-studio-2017-transformational-mt-technology/>

⁹ ゲール・ヴォン・アッシュ氏へのインタビューを参考のこと (報告書末へ収録)。

リルトと SDL の係争

- 2017年4月、SDLはリルトを特許権の侵害で訴えている(場所はアメリカのカリフォルニア)
10。SDLによれば、リルトはSDLの3つの特許を侵害している。これらの特許は元々米ラン
ゲージ・ウェーバー社が保持していたもので、2010年にSDLが同社を買収することによって、
特許を獲得していた。
- 3つの特許とは、『テキスト・トゥ・テキストシステムのオンラインとオフライントレーニング
を可能にするアダプター』(2009年11月授与)¹¹、『オンライン・アダプテーションを介し
たパーソナライズされた機械翻訳』¹²(2015年10月授与)、『効果的なオンライン・ドメイン
アダプテーション』(2015年12月授与)である。
- 世界最大手のSDLとスタートアップ企業のリルトの係争は注目を浴びたが、2017年11月に
決着がついている¹³。SDL幹部によれば、SDLとリルトが双方満足する仕方で訴訟は終了し
ている。だが、和解の内容については公表されていない。

その他のアダプティブ機械翻訳システム

- では、SDLとリルトの他に、同種の機械翻訳システムは開発されているのか。機械翻訳技術
コンサルタントのゲール・ヴォン・アッシュ氏によれば、リルトとSDL以外の組織も、アダ
プティブ機械翻訳システムの開発を進めているが、商用化させていない。だが、本報告書第3
章で記すように、EUプロジェクトであるMMTプロジェクトでは、アダプティブ・ニューラ
ル機械翻訳エンジンが開発されている。
- リルト、SDL、MMTの3つのシステムの共通の技術的特徴は、機械翻訳エンジンと翻訳メモ
リ技術の融合であり、これらのシステムには新しい自己学習型機械翻訳技術が導入されている。

¹⁰ <https://slator.com/industry-news/sdl-sues-lilt-patent-infringement/>

¹¹ <https://slator.com/industry-news/sdl-sues-lilt-patent-infringement/>

¹² <https://www.google.com/patents/US9152622?dq=%229,152,622%22&hl=en&sa=X&ved=0ahUKEwi1-fDXk7TAhXIKZQKHVG3A7sQ6AEIGzAA>

¹³ <https://slator.com/industry-news/sdl-lilt-settle-patent-lawsuit/>

第2 節 リルトのインタラクティブ・アダプティブ機械翻訳

リルトの基本情報

- リルトは、米グーグル社で機械翻訳システムの開発を行っていたジョン・デニーロ (John DeNero) とスペンス・グリーン (Spence Green) が 2015 年に設立したベンチャー企業である¹⁴。
- 両氏は、機械翻訳システムによる適切な翻訳支援は翻訳者が作業をより楽しむことを可能にすると考え、それがインタラクティブ・アダプティブ機械翻訳 (Interactive, adaptive MT) を開発する動機になった。同社は、従来の CAT ツール上での後編集を超える機械翻訳ツールの開発を目指し、2014 年以前に、インタラクティブ・アダプティブ機械翻訳のプロトタイプは完成されていた。
- ビジネス情報サイトのオウラー (Owler) によれば、リルトの総資金は 600 万ドルである¹⁵。
- リルトは、2016 年度ゲーム・チェンジャーイノベーションコンテストで、翻訳産業において最も技術革新的なソリューションをもたらした新規参入者に対する賞を得ている¹⁶。

インタラクティブ・アダプティブ機械翻訳の概要

- リルトによれば、インタラクティブ・アダプティブ機械翻訳は、人間と機械のパートナーシップを実現するインテリジェント技術とされ、具体的には、翻訳者のフィードバックか、既存の翻訳メモリ、また双方からリアルタイムで学習し、適応するシステムである¹⁷。以下、同システムの概要について記す¹⁸。

¹⁴ <https://lilt.com/about>

¹⁵ <https://www.owler.com/company/lilt>

¹⁶

<https://www.taus.net/think-tank/news/press-release/lilt-and-boffin-win-awards-in-this-year-s-taus-game-changer-innovation-contest-hosted-by-paula-shannon-lionbridge>

¹⁷ <https://lilt.com/kb/memory/mt>

¹⁸ <https://lilt.com/businesses>

<https://lilt.com/translators>

インタラクティブ・アダプティブ機械翻訳の特性¹⁹

- リルトのコアコンセプト 1: リアルタイム・フィードバックループ
 - ・ インタラクティブ・アダプティブ機械翻訳システムの重要な要素は、リアルタイム・フィードバックループの実現である。
 - ・ インタラクティブ・アダプティブ機械翻訳システムのインタラクティブな側面とアダプティブな側面は、リアルタイム・フィードバックループの2側面に対応する。
 1. インタラクティブとは、機械翻訳システムが翻訳者の編集を観察し、示唆される翻訳文を刷新することである。インタラクティブは、別の言い方では、プレディクティブ (predictive) とも言われ、すでに翻訳された文を使い、残りの文を予測することである²⁰。
 2. アダプティブとは、機械翻訳システムが翻訳者から学習し、翻訳の品質を高めること。
- リルトのコアコンセプト 2 : メモリ
 - ・ リルトの「メモリ (Memory)」とは、単なる翻訳メモリではなくて、翻訳メモリ、カスタム機械翻訳システム、用語ベースが結合したものである。²¹
 - ・ メモリ内のデータは、機械翻訳システムをトレーニングし、翻訳メモリ、用語ベースを刷新するために利用される。
- クラウドサービス

https://www.youtube.com/watch?v=YByEqFvzMw0&utm_content=buffereca74&utm_medium=social&utm_source=twitter.com&utm_campaign=buffer

¹⁹ より詳しくは、以下のウェブサイト及び論文を参考のこと。

<https://labs.lilt.com/technology-for-interactive-mt-c62f48c9f15a>

"Hierarchical Incremental Adaptation for Statistical Machine Translation"

<https://www-i6.informatik.rwth-aachen.de/publications/downloader.php?id=986&row=pdf>

"Models and Inference for Prefix-Constrained Machine Translation"

<https://aclweb.org/anthology/P/P16/P16-1007.pdf>

²⁰ <https://lilt.com/kb/api/interactive-translation>

²¹ <https://lilt.com/kb/memory/memories>

- ・ リルトの機械翻訳システムはクラウドサービスで提供されており、同社のシステムを利用するには、何も PC にダウンロードする必要はなく、ブラウザからリルトのプラットフォームにサインインするだけで良い。
- 直感的インターフェイス
 - ・ リルトはインターフェイスの開発にこだわり、多くの時間を費やした。
 - ・ グーグルの検索エンジンに類似するインターフェイスを開発した。このインターフェイスには、すでに多くの人が使うのに慣れている。

インタラクティブ・アダプティブ機械翻訳のプライバシー

- データセキュリティはリルトの長所の1つである²²。
 - ・ 顧客のデータは他の顧客と共有されることはない。
 - ・ 顧客のデータは第三者に利用されることはない。
 - ・ 顧客のデータは暗号によって保護される。

対応言語²³

- 英語 ⇄ 中国語、仏語、ノルウェー語、ロシア語、デンマーク語、ドイツ語、ポーランド語、スペイン語、オランダ語、ポルトガル語、スウェーデン語

インタラクティブ・アダプティブ機械翻訳の利点

- 費用効果
 - ・ リルトシステムの利用によって、大幅に費用を削減することが可能になる。
 - ・ リルトの使用と他の言語サービスプロバイダを比較すると、語数が多ければ多いほど、リルトのシステムは費用を削減することを可能にする。

²² <https://lilt.com/security>

²³ <https://lilt.com/kb/languages>

- リルトのシステムの以下の点が費用を削減可能にする。
 - リアルタイムの翻訳メモリアップデート
 - 翻訳メモリのレバレッジ
 - 翻訳者の生産性の向上
 - 機械翻訳の予測精度の向上
 - 翻訳のワークフローの効率の向上
- 翻訳速度
 - e2f社の発表では、リルトを利用して、最高で1時間に1347語を翻訳できた²⁴。
- 翻訳の品質
 - BLEUによる翻訳品質評価テストでは、リルトのシステムは、通常の統計的機械翻訳システム、通常のニューラル機械翻訳システムと比べて、はるかに高い評価を取めている。

ニューラル機械翻訳システムの影響

- 2017年11月1日、リルトは従来の統計的機械翻訳システムに代えて、ニューラル機械翻訳システムを導入することを発表している²⁵。リルトによれば、同社の新しいニューラル機械翻訳システムは、人工知能とリアルタイム・フィードバックループを使う業界初のものである。
- リルトは、ニューラル機械翻訳の利点について、以下のように説明している。
 - 一般的に、ニューラル機械翻訳の翻訳品質は統計的機械翻訳よりも高い。さらに、リルトのシステムを利用すると、通常のニューラル機械翻訳システムよりも品質が向上する。
 - 翻訳者に対する語の予測精度がより高くなり、翻訳者はより生産的に作業ができる。

²⁴ <https://e2f.com/6153/>

²⁵ <http://www.prweb.com/releases/2017/11/prweb14865144.htm>

- 新しい言語ペアをつくるのにトレーニングコーパスがより少なくて済み、顧客により迅速に新しい言語ペアを提供できる。
- リルトの顧客であるゼンデスク社 (Zendesk) が実施した比較研究 (目隠し試験) によれば、リルトの新しいニューラル機械翻訳システムと従来のシステムでは、前者の方が 71% の場合、後者よりも優れているか、同程度の翻訳品質を実現すると試験参加者が答えている。

第3節 SDLのアダプティブ機械翻訳

- ついで、SDLのアダプティブ機械翻訳の概要について記す。

SDLの基本情報

- SDL社は、1992年にイギリスで設立された機械翻訳開発企業兼大手言語サービスプロバイダであり、機械翻訳システムの開発とともに、翻訳サービスを提供している。
- 顧客は世界中の多くの企業の他²⁶、データセキュリティを長所に行しているため、諜報機関や警察機関などの英政府系機関とも契約している。
- SDLの2017年の収益は3億8800米ドルの見込みである²⁷。
- SDLは、2017年11月の欧州委員会主催のイベント、「トランスレーティング・ヨーロッパフォーラム」²⁸で、翻訳者にとっての重要な新しい機械翻訳技術として、CATツール、ニューラル機械翻訳、アダプティブ機械翻訳の3つを挙げており、アダプティブ機械翻訳を同社の重要な技術システムの1つとして考えている。

アダプティブ機械翻訳の概要

トラドス・スタジオ 2017 とアダプティブ機械翻訳エンジン

- SDLは、2016年11月に同社のCATツールであるSDLトラドス・スタジオ2017（デスクトップアプリケーション）の提供を開始した²⁹。同CATツール上で、リアルタイムで翻訳者の編集から学習するアダプティブ機械翻訳（Adaptive MT）エンジンを利用できる（1ヶ月40万字まで無料）。

²⁶ <http://www.sdl.com/customers/>

²⁷ <https://slator.com/financial-results/sdl-2017-revenues-8-profits-restructuring/>

²⁸ https://ec.europa.eu/info/sites/info/files/tef2017_stevens_en.pdf

²⁹ <http://blog.sdl.com/company/sdl-trados-studio-2017-transformation-translation/>

アダプティブ機械翻訳の特性

- アダプティブ機械翻訳エンジンはクラウドベースであり、使用には SDL のランゲージ・クラウド (Language Cloud) アカウントが必要である。
- アダプティブ機械翻訳は、翻訳メモリに一致する翻訳がない場合に使用するよう設計されている。
- 現在、SDL のアダプティブ機械翻訳エンジンは統計的機械翻訳技術に基づく。
- SDL によれば³⁰、アダプティブ機械翻訳自身は「自己学習機械翻訳 (self-learning MT)」とも呼ばれ、SDL 内では新しいものではなく、2005 年以来開発してきた。アダプティブ機械翻訳システムによって、翻訳者は従来の翻訳作業の仕方を変えることなく、パーソナライズされた機械翻訳エンジンを持つことになり、ユーザが機械翻訳エンジンをアップデートすることができる。またプロジェクト毎にエンジンをつくることも可能である。ブログは、新しいアダプティブ機械翻訳エンジンは機械翻訳と翻訳メモリ技術の融合であるとしている。

アダプティブ機械翻訳の必要性

- SDL が行った調査によれば、現在の機械翻訳システムに対するユーザの不満は以下の点であり、アダプティブ機械翻訳はこれらの不満を解消する手段となる³¹。
 - 翻訳の品質が悪いので、編集に時間がかかる。
 - 機械翻訳が犯す同じ誤りを繰り返し修正しなければならない。
 - カスタム MT エンジンをトレーニングする時間がない。

アダプティブ機械翻訳の利点

- SDL があげるアダプティブ機械翻訳の利点は以下である。
 - 費用削減
 - 生産性の増大：同じ誤りを修正する必要性がないので、編集が減る。

³⁰ <https://blog.sdltrados.com/adaptivemt-sdl-trados-studio-2017-transformational-mt-technology/>

³¹ SDL は 2017 年 12 月に同社のアダプティブ機械翻訳を紹介するウェビナーを実施している。

<http://www.sdl.com/event/language/webinars/12-2017/spotlight-on-adaptivemt-next-generation-machine-translation-in-studio-2017.html>

- ・ 労力の削減：通常のトレーニングの必要がなく、すぐに高い品質の翻訳を享受できる。

アダプティブ機械翻訳のプライバシー

- SDL はアダプティブ機械翻訳のプライバシーについて、次のように説明している³²。
 - ・ SDL は元のテキストや翻訳されたテキストの複写を保持することはない。
 - ・ ユーザのデータは他のユーザに共有されることなく、利用されることもない。
 - ・ アダプティブ機械翻訳エンジンによって生まれた統計モデルはユーザ固有のものであり、ユーザのアダプティブ機械翻訳エンジンに保存され、他のユーザはアクセスすることはできない。
 - ・ アダプティブ機械翻訳エンジンは、米国にある米アマゾンの AWS サーバにホストされている。

アダプティブ機械翻訳の対応言語³³

- 英語 ⇄ 仏語、ドイツ語、イタリア語、スペイン語、オランダ語、ポルトガル語、日本語、中国語

今後の動向

- SDL のアダプティブ機械翻訳のホワイトペーパーでは³⁴、今後、ニューラル機械翻訳技術を使って、アダプティブ機械翻訳の機能を実現することが目標とされている。

³² <https://www.sdl.com/about/privacy/adaptivemt.html>

³³ <http://www.sdl.com/software-and-services/translation-software/machine-translation/language-cloud/language-pairs.html>

³⁴ <http://www.sdl.com/software-and-services/translation-software/machine-translation/language-cloud/adaptivemt/>

第4節 インタラクティブ・アダプティブ機械翻訳の市場動向

- ついで、リルトのインタラクティブ・アダプティブ機械翻訳と SDL のアダプティブ機械翻訳の市場の動向について記す。
- これらは新しいシステムであり、シンクタンクや調査機関はユーザ数、あるいは市場規模について本格的にまだ調査を実施していないようである。また、リルトや SDL もこれらの詳しいデータを公表しておらず、数量的に市場動向を明らかにするが困難である。

リルトの顧客情報

- リルトによれば、リルトのシステムにはフリーランスの翻訳者、言語サービスプロバイダ、企業という3つの顧客の枠があるが、これらすべての顧客数は毎月増加している。
- リルトはウェブサイト上で、顧客の情報を発表しており、同社の市場動向を知る上で参考になる。顧客は翻訳・ローカライゼーション企業が多いが、他業種の企業もいる。
 - クラウドワーズ³⁵ (米) : 企業の国際進出とローカライゼーションの支援企業
 - Arancho Doc³⁶ (イタリア) : 翻訳・ローカライゼーション企業
 - スマートリング³⁷ (米) : 翻訳・ローカライゼーション企業
 - Acclaro³⁸ (米) : 翻訳・ローカライゼーション企業
 - ログルス・グローバル³⁹ (米) : 翻訳・ローカライゼーション企業
 - e2f (米)⁴⁰ : 翻訳・ローカライゼーション企業

³⁵ <https://www.cloudwords.com>

³⁶ <http://www.aranchodoc.com>

³⁷ <https://www.smartling.com>

³⁸ <https://www.acclaro.com>

³⁹ <https://logrusglobal.com>

- ・ ゼンデスク⁴¹（米）：顧客管理サービス企業
 - ・ ハドソン湾会社⁴²（カナダ）：デパートなど小売業者
 - ・ スマートキャット社（米）⁴³：CAT ツールの開発と提供
- リルトの機械翻訳エンジンは他社の CAT ツール上での利用可能であり、他社と対立するだけでなく、提携することが可能である。例えば、すでに、米スマートキャット社⁴⁴の CAT ツール上でリルトのエンジンを利用できる。リルトは他社が利用できる REST API と Javascript ライブラリを持つ。したがって、リルトと他の CAT ツール企業間での提携が進めば、リルトのユーザは増加する可能性がある。

インタラクティブ・アダプティブ機械翻訳の市場動向に対する関係者の意見

- 以下に、インタビュー調査を行った際の機械翻訳技術分野の関係者の意見を紹介する。

注目すべき点

- 特に関係者の意見の中で、注目すべき3点について記す。
 1. リルトのインタラクティブ・アダプティブ機械翻訳と SDL のアダプティブ機械翻訳の詳しいユーザ数や市場規模は判明ではないが、まだユーザは多くなさそうである。
 2. これらのシステムに興味がある人は多く、また技術としてとても優れているので、今後、ユーザは増加する可能性が高い。
 3. SDL のアダプティブ機械翻訳に関しては、同社のトラドス・スタジオ 2017 に標準装備されているので、潜在的なユーザは多い。

⁴⁰ <https://e2f.com/7539/>

<https://e2f.com/case-study-lilt-travel-portal/>

⁴¹ <https://www.zendesk.com>

⁴² <https://www.thebay.com/webapp/wcs/stores/servlet/en/thebay>

⁴³ <https://www.smartcat.ai/blog/2017/05/04/lilt/>

⁴⁴ <https://www.smartcat.ai/blog/2017/05/04/lilt/>

各関係者の意見

- 欧州の言語技術の動向に詳しいジャーナリストは、次のように述べている。
 - 「インタラクティブ・アダプティブ機械翻訳については、まだ萌芽段階であり、採用している企業は多くないだろう。先日の言語技術産業サミットにおいても、(数十のプレゼンの中で) 直接的にインタラクティブ・アダプティブ機械翻訳について言及していた企業・研究機関はなかったと認識している。そのため、同技術に対する評価は難しいのではないか」
- イギリスの大学の翻訳の専門家は、次のように述べている。
 - 「アダプティブ機械翻訳システムの正確なユーザ数は知らないが、個人的な印象としては、まだユーザは多くないと思う。それには2つの理由がある」
 - 「第一に、アダプティブ機械翻訳システムを搭載した CAT ツールがまだあまりないから。アダプティブ機械翻訳システムが利用できるのは、現在までのところ、英 SDL 社のトラドス・スタジオ (CAT ツール) と米リルト社の機械翻訳システムにおいてだけであると思う」
 - 「第二に、ソーシャルネットワーキングにおける翻訳者のコミュニティでの会話のやり取りを見ていると、幾人かの翻訳者はアダプティブ機械翻訳システムを利用しているようだが、多くの翻訳者がまだ使っていないことが分かる。だが、興味がある翻訳者は多そうである」
 - 「アダプティブ機械翻訳システムは非常に優れた機械翻訳システムなので、将来的には、多くの CAT ツールで同システムを利用できるようになり、また多くの翻訳企業や翻訳者が利用して、大きく一般的に普及する可能性がある」
- 欧州の機械翻訳技術のコンサルタントは、次のように述べている。
 - 「現状として、リルトのインタラクティブ・アダプティブ機械翻訳システムと SDL のアダプティブ機械翻訳システムはとても普及していると思う。というのも、特に SDL に関しては、アダプティブ機械翻訳システムが同社の新しい「トラドス・スタジオ 2017」(製品名)

に標準装備されているからである。したがって、新しいトラス・スタジオのユーザはアダプティブ機械翻訳システムを自由に利用できる。しかも、ユーザは同システムを利用していることに気づいていないかもしれない」

- 「翻訳者は技術に対して保守的なところがあり、新しい技術を積極的に取り入れない傾向があるが、一度 SDL やリルトのシステムの性能を知れば、これらのシステムを使うようになり、将来的に、さらにこれらの機械翻訳システムの普及が進むであろう」
- フランスの大学の機械翻訳技術の専門家は、次のように述べている。
 - 「アダプティブ機械翻訳を使ったことがないが、非常に関心はある」
 - 「リルトのインタラクティブ・アダプティブ機械翻訳システムと SDL のアダプティブ機械翻訳システムを実際には利用したことがないが、非常に興味はある。他の多くの大学の同僚も興味を持っている。機械翻訳システムの伝播、自動修正、翻訳者からの自己学習という要素は、非常に興味深い課題であり、研究開発分野である」
 - 「アダプティブ機械翻訳システムを使って翻訳作業をしたことがない」。
 - 「SDL のアダプティブ機械翻訳エンジンを展示会で知り、その原理的要素を知ることができた。この技術は後編集作業を簡単にするのに私にはとても有望だと思われる」

第5節 インタラクティブ・アダプティブ機械翻訳に対する評価

- 最後に、リルトのインタラクティブ・アダプティブ機械翻訳と SDL のアダプティブ機械翻訳に対する現状の評価について記す。

リルトと SDL の機械翻訳システムに対する関係者の評価

- 以下に、インタビュー調査を行った際の機械翻訳技術分野の関係者の意見を紹介する（リルトと SDL のシステム両方について）。

注目すべき点

- リルトと SDL のシステムについて、肯定的な評価をする人が多いが、問題点を指摘する人もいる。
 - 肯定的評価
 - 翻訳の品質向上
 - 翻訳速度の上昇
 - 否定的評価
 - ユーザが間違った修正を行い、翻訳エンジンが誤って学習してしまうかもしれない。
 - 文体に関わる偏った修正を過度に与えすぎてしまい、翻訳エンジンの性能を減少させるかもしれない。
 - 機械翻訳企業がユーザの後編集データを自社の機械翻訳エンジンの改善のために利用している可能性がある。
 - この最後の点に関しては、関係者間で意見が分かれた。ユーザの観点から、翻訳者にはこのような疑いを持つ人がいるが、技術的な観点から、機械翻訳の技術者はフィードバックデータの利用には費用がかかりすぎるので現実には不可能であるとしている。

- ・ その他
 - すでに CAT ツールにはユーザの後編集から自己学習する機能は装備されており、それがリルトと SDL の新しいシステムと技術的にどのように違うのかよく分からない。

各関係者の意見

- イギリスの大学の翻訳の専門家は、次のように述べている。
 - ・ アダプティブ機械翻訳の優れた点
 - 「機械翻訳技術の最大の問題の1つは、翻訳の品質である。現在、翻訳の品質を向上させる最も有力な方法は、機械翻訳システムによって翻訳された文章を人間が修正することであり、翻訳者の後編集が非常に重要である。それゆえ、アダプティブ機械翻訳システムは潜在的にとっても有力である。このシステムは、翻訳者の介入という最も重要な要素をシステム内に取り入れているからである」
 - 「翻訳者の後編集から学習する機械翻訳システムは以前からあった。だが、アダプティブ機械翻訳システムは、翻訳者の修正からリアルタイムで学習し、翻訳エンジンをすぐに改善するという点でとても優れている」
 - ・ アダプティブ機械翻訳システムの問題点
 - 「特にアダプティブ機械翻訳システムに関しては、同システム固有の大きなリスクがあると思うが、まだ広く知られておらず、議論もされていない」
 - 「そのリスクとは、翻訳エンジンの改善には人間の翻訳者による翻訳文の修正が最も効果的であるが、翻訳企業が、アダプティブ機械翻訳システムの使用で生まれる修正データを自社の翻訳エンジンを改善するのに利用している可能性があることである」
 - 「どんな企業も翻訳者が行う1つ1つの個別の修正に興味は持たないが、多くの修正に見出されるパターンの検出には大きな関心を払っており、数年はかかるだろうが、この情報

を翻訳エンジンを改善するために使うことができる」

- 「アダプティブ機械翻訳システムは、ユーザ数が多くなれば、翻訳機械企業の翻訳エンジンを改善する非常に有効な手段ともなりうる」
- 欧州の機械翻訳技術のコンサルタントは、次のように述べている。
 - 「リルトの計算では、機械翻訳システムを使わずに翻訳者は 2500 語を 1 日に翻訳する。機械翻訳システムを利用すると 1 時間に 800 語を翻訳でき、機械翻訳システムを使わない場合と比べて、はるかに多くの量を翻訳できる（1 日に 6400 語）。さらにアダプティブ機械翻訳システムを使うのに慣れると、1 日に 1 万から 1 万 4000 語を翻訳できるようになれる。こうして、機械翻訳技術は翻訳作業の効率を向上させるので、翻訳者にとって非常に有用性が高く、翻訳者は特にアダプティブ機械翻訳システムから大きな利益を得る」
 - 「アダプティブ機械翻訳システムは、翻訳者による後編集からリアルタイムで学習し、翻訳文を改善するが、同システムを提供する機械翻訳企業がその後編集のデータを利用して、自社の一般的な機械翻訳エンジンを改善することに使っているのではないかという疑いを持つ人がいるようである。だが、そのようなことは実際にはありえない。というのも、一般的に言って、言語モデルを再トレーニングするには、時間と費用がとてかかるからである。翻訳モデルの再トレーニングには 5 万語から 20 万語、言語モデルには 5 億語が必要になり、非常に多くの時間と費用が必要となるので、もし翻訳者のフィードバックのデータを持っていたとしても、実際に使うことは難しい」
 - 「アダプティブ機械翻訳システムについては、このようなりスクはさらに少ないとも言える。アダプティブ機械翻訳システムは、翻訳者の後編集から学習し、翻訳者固有の翻訳メモリを変更するが、この変更データは個人のものであり、翻訳機械企業はその翻訳者のデータを他のデータから区別しているからである。アダプティブ機械翻訳システムが得る翻訳者のフィードバックデータは、決して企業の一般的な機械翻訳システムを改善すること

に利用されない。 ユーザのデータは、確かにサーバ上にあるが、それは他の目的に利用されない」

- フランスの大学の機械翻訳技術の専門家は、次のように述べている。
 - ・ 「インタラクティブ・アダプティブ機械翻訳システムと聞くと、まずプレディクティブ (predictive) と言う技術的側面を想像してしまう。現在、インタラクティブ、アダプティブ、プレディクティブという側面が同一のシステムに取り入れられており、それぞれが技術的にどう違うのかよく分からない」
 - ・ 「インタラクティブ・アダプティブ機械翻訳システムが有する翻訳者の後編集から学習し、翻訳を改善させる機能というのは、伝播 (propagation) と呼ばれているものに類似すると思う。現行の CAT ツールでも、翻訳者の後編集から学習し、同じ間違いを回避する機能はある」
 - ・ 「実際、アダプティブ機械翻訳の危険は、一方で、間違っただ修正を誤って翻訳エンジンにもたらすかもしれないこと、他方で、文体に関わる偏った修正を過度に与えずぎてしまい、翻訳エンジンの性能を減少させるかもしれないことである」

リルトの評価

- 上記では、リルトと SDL のシステムを区別せずに、関係者の意見を記した。ついで、リルトについての評価について記す。

リルト発表の評価

- リルト自身は、以下のようにユーザ評価を発表している。
 - ・ ユーザの声 (リルトの発表)
 - 「翻訳速度が上がった」。

- 「翻訳作業が楽しい（ユーザ経験の向上）」。
- 「カスタマーサポート体制が整備されている」。
- ・ ゼンデスク社は、リルトのシステムの長所について、人間の翻訳者に投資することが、同時に同社の機械翻訳エンジンの品質を向上させることにつながることを挙げている。（リルトの発表）
- 直感的インターフェイスについて
 - ・ リルトによれば、翻訳者が作業を楽しめるように、翻訳システムのインターフェイスの設計に多くの時間を使った。 スイスのチューリッヒ応用科学大学（ZHAW）がこの点について、SDL のインターフェイスと比較する研究を行い、結果を発表している。この研究は、CAT ツールを使う際の翻訳者の視線の動きを調べるものであったが、SDL のインターフェイス上では、翻訳者は視線をインターフェイスの左右に絶えず移さねばならず、翻訳者を疲れさせ、効率が良くないが、リルトのインターフェイス上では、視線をあまり動かす必要がなく、操作性が改善されている。

関係者の評価

- 欧州の言語技術の動向に詳しいジャーナリストは、次のように述べている。
 - ・ 「アメリカのリルト社のアダプティブ機械翻訳システムは非常に成功している。特に、インターネットやソーシャルネットワーキングに慣れ親しんでいる 20 代、30 代の若いユーザー（翻訳者）は、同社のシステムに満足している」
- イギリスの大学の翻訳の専門家は、リルトについて、次のように述べている。
 - ・ 「リルトでは、すべての情報が私的な情報であり、保護されるものとして扱われているとともに、多くのことが比較的オープンに公表されている。だが、証拠はないものの、SDL

はすべての情報を守秘義務がある情報としてみなしていないようであり、この点が明瞭ではない」

第3章 欧州連合のホライゾン2020におけるアダプティブ・ニューラル機械翻訳システムの研究開発動向 / MMT プロジェクトの動向

- 第2章では、リルトのインタラクティブ・アダプティブ機械翻訳と SDL のアダプティブ機械翻訳の動向について概観した。欧州では、EU のホライゾン 2020 によって助成される MMT (Modern MT) プロジェクトにおいて、アダプティブ・ニューラル機械翻訳システムが開発され、いわば EU 産のインタラクティブ・アダプティブ機械翻訳が実用化されている。以下に、同プロジェクトの概要などについて記す。

第1節 MMT プロジェクトの背景

- MMT プロジェクトは、EU の第7次枠組計画（FP7）で助成された MATECAT（メイトキャット）プロジェクトとほとんど同じ参加組織で実施されている。メイトキャットプロジェクトに参加していた仏メヌ大学は、MMT プロジェクトには参加しておらず、代わりにタウスが参加している（なお、メイトキャットプロジェクトに参加していたメヌ大学研究者はフランスのパリに設立されたフェイスブックの人工知能研究所に移動している）。

MMT プロジェクトの基本情報⁴⁵

プロジェクト名称	MMT (Modern Machine Translation)
研究期間	2015年1月～2017年12月
予算	369万5200ユーロ（EU 拠出分は299万4700ユーロ）
コーディネーター	伊トランスレイティッド
参加組織	伊FBK、英エジンバラ大学、蘭タウス

メイトキャットプロジェクトの基本情報⁴⁶

プロジェクト名称	MateCat (Machine Translation Enhanced Computer Assisted Translation)
研究期間	2011年11月～2014年10月
予算	336万8351ユーロ（EU 拠出分は265万ユーロ）
コーディネーター	伊FBK
参加組織	英エジンバラ大学、伊トランスレイティッド、仏メヌ大学

- メイトキャットプロジェクトでは、メイトキャット⁴⁷と呼ばれるオープンソースのCAT ツールを開発しており、すでに実用化されており、広く普及している。

⁴⁵ http://cordis.europa.eu/project/rcn/194327_en.html

⁴⁶ http://cordis.europa.eu/project/rcn/100744_en.html

⁴⁷ <https://www.matecat.com>

- メイトキャットプロジェクトでも、翻訳者の編集から学習する機械翻訳エンジンのインクリメンタルアダプテーションについての研究がなされていた⁴⁸。

⁴⁸ <https://hal.archives-ouvertes.fr/hal-01158054/document>

第2 節 MMT プロジェクトの概要

- MMT プロジェクトが開発したシステムは、アダプティブ・ニューラル機械翻訳 (Adaptive neural MT) と呼ばれている。
- 2017 年 11 月、MMT のアダプティブ・ニューラル機械翻訳は、メイトキャット上で一般に利用可能になった⁴⁹。
- プロジェクト開始当初は、統計的機械翻訳を利用する予定であったが、プロジェクト期間中に、ニューラル機械翻訳が登場し、最終的に後者を採用するに至っている。

MMT プロジェクトの目標

MMT プロジェクトの基本目標

- MMT は同時に、一般 (generic) MT エンジンとカスタム MT エンジンの良い側面をそれぞれ取り入れることを目標とする。
 - 一般 MT エンジンの長所
 - 一般的に翻訳の品質が高い。
 - スケーラビリティが高い。
 - 大量のデータに基づいている。
 - すぐに利用できる。
 - カスタム MT エンジンの長所
 - 特定のドメインで翻訳の品質がとても高い。

MMT の4つの技術的目標

1. 最初のトレーニング段階を必要とせず、ユーザデータと修正から連続的に学習する。
2. 自動的にコンテキストを取り扱い、ドメインに特有のシステムを構築する必要がない。

⁴⁹ <https://slator.com/sponsored-content/mmt-brings-adaptive-neural-mt-to-matecat-and-to-mymemory-plugin-for-sdl-trados/>

3. データとユーザのスケラビリティに対応することが可能である。
4. 大規模な IT 企業（グーグルなど）と機械翻訳産業のあいだにあるデータのギャップを埋めるために、データ収集インフラストラクチャをつくる。

アダプティブ・ニューラル機械翻訳の概要

- 上記のように、MMT の技術的な目標の 1 つは、機械翻訳システムがユーザの翻訳メモリと後編集から連続して学習することである。翻訳の品質は、ユーザの後編集活動から新しいトレーニングデータが追加されると向上していく⁵⁰。つまり、ユーザが後編集すると、その情報が MMT エンジンにフィードバックされ、エンジンがリアルタイムで学習し、次に後編集する際に、翻訳文の予測を改善する。
- 別の観点から言えば、MMT のアダプティブ・ニューラル機械翻訳は、ニューラル機械翻訳の利点であるコンセプトの抽象能力と翻訳の流暢さ、そして、統計的機械翻訳の情報を記憶するという利点を結合すること、より具体的には、ニューラル機械翻訳エンジンと翻訳メモリを結合することによって実現される。
- MMT の作業パッケージ 3 で、機械翻訳システムがユーザから学習する側面が研究されている。
 - 例えば、作業パッケージ 3 のレポート、「MMT The 2nd Report on Database and MT Infrastructure⁵¹」では、第 5 章第 4 節 (5.4) で、アダプティブ翻訳モデル、特にインクリメンタル・アップデート (Incremental updates) などについて説明されている。

MMT の利点

- アダプティブ・ニューラル機械翻訳、後編集の努力を最小限化することを可能にする。特に、

⁵⁰ より詳しくは、以下の論文やウェブサイトの情報を参考のこと。

"Multi-Domain Neural Machine Translation through Unsupervised Adaptation", <http://www.statmt.org/wmt17/pdf/WMT13.pdf>
http://cordis.europa.eu/result/rcn/190001_en.html

<https://www.metn17.metmeetings.org/abstract.php?id=984>

⁵¹ <https://www.modemmt.eu/wp-content/uploads/2016/12/D3.2-Second-Report-on-Database-and-MT-Infrastructure.pdf>

同じ語を繰り返す文にとっても有効であり、顧客が要求するスタイルに合わせて翻訳することを支援する。

- 一般・ニューラル機械翻訳、特殊・ニューラル機械翻訳、アダプティブ・ニューラル機械翻訳の翻訳品質を、BLEU を使って、欧州中央銀行（ECB）、国際連合の文書（UN-TM）などを使って比較した結果、アダプティブ・ニューラル機械翻訳は最も良い結果を得ている⁵²。

MMT のビジネスモデル

- MMT エンジンを利用するには、ユーザはライセンスを購入する必要がある。MMT は、ユーザがエンジンを利用した分だけ支払うビジネスモデルを取っている⁵³。
- MMT エンジンの API は公開されており、メイトキャット以外の CAT ツール上でも利用可能である。

MMT のプライバシーの保護

- 翻訳者は MMT エンジンを利用する際に、自己の翻訳メモリを利用することが可能である。MMT は、ユーザの翻訳と翻訳メモリを読むことになるが、その際、この情報は MMT エンジンが他のすべてのユーザに行うコンピューテーションを強化することに利用され、結果として、全てのユーザにとってより良い機械翻訳をつくることにつながる。翻訳者個人の翻訳メモリには誰もアクセスできず、その情報は保護される⁵⁴。
- ユーザの翻訳が公にされないことを確保するために、翻訳プロジェクトに翻訳メモリキーを追加する必要があり、こうすることによって、ユーザの翻訳にはユーザだけがアクセスできるようになる。

⁵² より詳しくは、論文「Multi-Domain Neural Machine Translation through Unsupervised Adaptation」を参考のこと。

<http://www.statmt.org/wmt17/pdf/WMT13.pdf>

⁵³ <https://www.matecat.com/support/advanced-features/modernmt-mmt-plugin/>

⁵⁴ 以上のプライバシーに関する情報は、2018年2月8日に行われたウェビナー末での質疑応答に由来する。

MMT の言語ペア

- 英語 ⇔ スペイン語、イタリア語、ポルトガル語、ドイツ語、仏語、オランダ語、ロシア語、アラビア語、中国語

補足資料

SATT 2017 (School of Advanced Technologies for Translators) 視察レポート

- 日程 : 2017 年 9 月 14 日
- 場所 : イタリア・トレント

イベントの概要

- イタリアのトレントで 2017 年 9 月 14 日から 15 日にかけて、SATT2017 が開催された（14 日だけ参加）⁵⁵。主催は FBK（ブルーノ・ケスラー財団：トレントに本拠地）である。
- SATT2017 は、School of Advanced Technologies for Translators（翻訳者のための先端技術スクール）の省略であり、今年で 2 回目の開催である。
- 参加者は 80 名ほど。翻訳を学ぶ学生が半分ほどであり、他にイタリアと欧州隣国のフリーランスの翻訳者と翻訳サービス事業者が参加（主催者からの情報）。
- スポンサーは、独ウィーローカライズ社（welocalize）、英 SDL 社、伊トランスレイティッド社（Translated）、スペイン・パンゲアニック社（Pangeanic）である。
- 1 日目は、主に最先端の機械翻訳技術と翻訳産業について、招待されたスピーカーがプレゼンテーションを行った。二日目は、メイトキャット（MATECAT）と SDL の CAT（Computer-Assisted Translation）ツールの利用方法を実際に参加者に教えるセミナーが行われた。
- スピーカーは、FBK やダブリン市民大学の ADAPT センターなどの研究機関の他、米リルト社（Lilt）やウィーローカライズ、トランスレイティッドなどの翻訳サービスの提供及び機械翻訳システムの研究開発を行う事業者であった。

⁵⁵ <https://satt2017.fbk.eu>

写真 会場内の様子



主なポイント

- セミナー全体を通して話題となったのが、ニューラル機械翻訳（Neural Machine Translation）である。数年来、同技術は従来の統計機械翻訳技術に代わるコア技術として研究開発が進み、現在産業界が導入を進めている。
 - ・ FBK で機械翻訳技術の研究開発を主導するマルチェロ・フェディリコ氏（FBK HLT-MT ユニット長兼トレント大学大学院助教授）によれば、ニューラル機械翻訳は機械翻訳システムの新時代を切り開いた今後最も注目すべき技術である。だが、ニューラル機械翻訳がすべての潜在的能力を発揮するにはまだ時間がかかり、従来のフレーズベース機械翻訳は産業ではまだ競争力がある。
- セミナーが翻訳者向けであったので、セミナー全体を通して、人工知能に基づく機械翻訳技術の進展は翻訳者の仕事を奪うわけではないが、仕事の内容と仕方が変わっていくだろうという主張がされた。
- FBK は、トランスレイティッド、エジンバラ大学と繋がりが深く、メイトキャット（MATECAT）⁵⁶と MMT⁵⁷という EU プロジェクトを共に実施している。
 - ・ 欧州には EU プロジェクトを通して、研究機関と民間企業がオープンソースの機械翻訳システムを提携して開発し、商用化するという流れがある。トランスレイティッドは EU プロジ

⁵⁶ メイトキャットプロジェクト参加組織：FBK（コーディネーター）、仏メヌ大学、伊トランスレイティッド、英エジンバラ大学、研究期間：2011年11月～2014年10月、予算：336万8351ユーロ（EU 拠出分は265万ユーロ）

⁵⁷ MMTプロジェクト参加組織：トランスレイティッド（コーディネーター）、FBK、英エジンバラ大学、蘭タウス、研究期間：2015年1月～2017年12月、予算：369万5200ユーロ（EU 拠出分は299万4700ユーロ）

ェクトで FBK、エジンバラ大学などの研究開発機関とともにオープンソースの CAT ツールを開発し、商用化するとともに、それを使って翻訳サービスを提供し、ビジネスを展開している。

- ・ メイトキャットプロジェクトでは、研究機関である FBK、エジンバラ大学などとトランスレイティッドが提携し、オープンソースで、無料のメイトキャットと呼ばれるインタラクティブな CAT ツールを開発し、製品化している。トランスレイティッドは、メイトキャットを使って、翻訳サービスを提供し、事業を行っている。
- ・ MMT プロジェクトでは、新しいオープンソース機械翻訳アーキテクチャに基づく、言語から独立した商用オンライン翻訳サービスを開発することを目標とする。同プロジェクトでは、ニューラル機械翻訳技術の研究を実施しており、2017 年 10 月から、メイトキャットに MMT が開発するニューラル機械翻訳システムをプラグインできるようになる。
- リルトのインタラクティブ・アダプティブ機械翻訳システムは、翻訳者の修正から学習し、リアルタイムでその修正を翻訳文の示唆に反映させるという点で、他の機械翻訳システムと技術的に一線を画す。

「翻訳技術と翻訳産業」のプレゼンテーションの概要

以下に、各プレゼンテーションの重要なポイントについてのみ記す。

ADAPT センター（アイルランド） / 機械翻訳はいかに翻訳者を支援できるか

- インターネットサービスの現状について：インターネット上のサービス（フェイスブック、ユーチューブ、ツイッターなど）が増加しており、大量のコンテンツを翻訳する必要が生まれている。だが、莫大な量のコンテンツに翻訳するのに必要なだけ翻訳者がおらず、十分な量のインターネットコンテンツが翻訳されていない（9 割が翻訳されていない）。
- より多くのコンテンツが翻訳された場合の利益は大きい。欧州における国際 e コマースの成長は、各国内部の e コマースの成長よりもだいぶ遅い。欧州デジタル市場をより確固としたものにするためには、言語の障壁を崩すとともに、欧州の言語的多様性を保持する必要がある。
- 機械翻訳市場について：2014 年のタウス (TAUS)⁵⁸の報告によれば、機械翻訳市場は 2 億 5000 万ユーロである⁵⁹。IT 企業は機械翻訳によって新しいサービスを提供し、また既存のサービスを強化することができる。
- 機械翻訳の新しいユースケースは、ユーザ生成コンテンツ (User Generated Content : UGC) である。UGC は、オンラインチャットやツイート、ブログ、ホテルや製品のレビュー、ソーシ

⁵⁸ タウス (TAUS) は、翻訳分野の国際的な産業団体であり、Translation Automation User Society の略称である。同団体は 2004 年に設立された。本拠地はオランダのアムステルダムである。

⁵⁹

<https://www.taus.net/think-tank/news/press-release/size-machine-translation-market-is-250-million-taus-publishes-new-market-report>

ャルメディアのポストなどである。

- ・ UGC は寿命が短い場合が多い。例えば、オンラインチャットの翻訳は、すぐに必要なくなり、消去される。
- ・ 企業は UGC の翻訳に予算を割くことができないが、翻訳すれば、ウェブサイトにも価値を付加できる。
- ・ UGC の特徴としては、対訳データの不在、言語スタイルの変則性、スペルミスなどを含むこと、速いリアルタイムの翻訳の必要性である。
- 機械翻訳の新しい技術として、ニューラル機械翻訳が登場した。グーグルが 2016 年 10 月に『グーグルのニューラル機械翻訳システム：人間と機械翻訳のギャップを埋めること』⁶⁰という論文を発表している。これにより、機械翻訳への注目がさらに高まるとともに、翻訳者からの抵抗も再燃した。
 - ・ だが、翻訳者は機械翻訳に好意的であるべきである。良い機械翻訳は翻訳者をより優秀にする。
- ニューラル機械翻訳は非常に将来的に有望である。

トランスレイティッド / ダヴィッド・カロセリ / メイトキャットにおける MMT

- MMT (Modern MT) プロジェクトは、ホライゾン 2020 から助成されている機械翻訳システムの研究開発プロジェクトである。
 - ・ 参加組織：トランスレイティッド (コーディネーター)、FBK、英エジンバラ大学、蘭タウス
 - ・ 研究期間：2015 年 1 月～2017 年 12 月
 - ・ 予算：369 万 5200 ユーロ (EU 拠出分は 299 万 4700 ユーロ)
- 現在機械翻訳に求められていること
 - ・ システムがユーザデータに適応し、後編集 (post-edit) から学習すること
 - ・ ユーザが高い生産性とより良いユーザ体験を楽しめること
 - ・ システムのインストールと展開が簡単なこと
 - ・ 新しいプロジェクト向けのセットアップが速いこと
 - ・ システムが効果的であること
 - ・ データとユーザに対してスケラブルであること
 - ・ MMT はこれらの要求に応える。
- MMT は、すぐに利用可能な基準となるシステムのモデルであり、新しいユーザデータの同化、リアルタイムの学習、その場での適応を可能とする。
- MMT は、オープンソースソフトウェアであり、インターネット上でリリースされている⁶¹。

⁶⁰ 原題は、『Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation』である。

⁶¹ <https://github.com/ModernMT/MMT>

- 新しい機械翻訳技術として、ニューラル機械翻訳が登場し、フレーズベースの機械翻訳は廃れていくだろう。
- ニューラル機械翻訳技術には、分野ごとに特殊化されたニューラル機械翻訳システム（specialized）と一般的なニューラル機械翻訳システム（generic）がある。前者は翻訳の品質は高いが、分野を横断して利用できず、高価であるという欠点があるが、後者は安価で保持が簡単であるが、翻訳の品質が劣るといった欠点がある。
- MMT プロジェクトが開発するソリューションは、アダプティブ・ニューラル機械翻訳（Adaptive Neural MT）システムであり、翻訳の品質がどの分野でも高く、安価である。
- MMT システムの実用化が進められており、2017年10月、EUプロジェクトで開発されたCATツールであるメイトキャット⁶²にプラグインすることが可能になる予定である。
- MMT のアダプティブ・ニューラル機械翻訳は、グーグル翻訳よりも翻訳の品質が高い。

リルト / カルメン・ヘーガー / 人間と機械のパートナーシップとしての翻訳

- リルト（LILT）は従来の分断されたサプライチェーンを統合されたワークフローへと変える。
- 後編集作業を好まない翻訳者は多い。プロの翻訳者の56%が後編集作業を好きかという質問に対して、好きではないと回答している（2015年の調査）。
- リルトを使う翻訳者からは好意的な声がある。リルトの機械翻訳システムは後編集ツールではない。
- リルトとグーグル翻訳のインターフェイスは似ており、リルトを使うことは簡単である。
- リルトのインタラクティブ・アダプティブ機械翻訳は、洗練された翻訳（インタラクティブな要素）と、迅速な分野適応と学習（アダプティブな要素）を可能にする。
- リルトの機械翻訳システムは後編集ツールの代替となる。リルトのシステムは、翻訳者の作業中に示唆を与えるが、彼がそれを承諾したこと、拒否したことをリアルタイムで学習し、訳語の示唆をすぐに洗練させていく。
- リルトのシステムでは、機械翻訳、翻訳メモリ、用語ベースが統合されており、それにより、リアルタイムで翻訳者の修正から学習して、翻訳文の示唆に反映させることが可能になる。
- リルトのシステムで、翻訳の生産性が高まっており、品質も向上している。
 - ・ リルトシステムを使った事例1：GetYourGuide⁶³のウェブサイトを経営する英語から複数の他言語へと翻訳する大型プロジェクトがあったが、生産性（1時間に翻訳される語数）がとても高かった。（翻訳メモリが0の状態から、10日間で100名の翻訳者が177万語の入力を行った）。

⁶² メイトキャットは、FBK（コーデイナー）、仏メーヌ大学、トランスレイティッド、エジンバラ大学のコンソーシアムが開発した。

http://cordis.europa.eu/project/rcn/100744_en.html

https://www.matecat.com/benefits/?gclid=Cj0KCQjwgiPOBRDnARIsAHA1X3RKp0isd3vaUVLEvdbd20axHp8rqA2hDX4qdAycv wNboT7VvyxEpDQaAi6yEALw_wcB

⁶³ <https://www.getyourguide.com>

- ・ リルトシステムを使った事例 2: IT 企業の FORTUNE-100 を翻訳したが、翻訳の品質が高かった。

ロゼッタファウンデーション / オルガ・ブラスコ / 翻訳産業の再編成：機械翻訳と人間の才能の価値のレ

(エ) ポリユーション [r(e)volution]

- オルガ・ブラスコ氏は、言語サービス事業者、技術系企業、NGO などへコンサルティングサービスを提供している（買収や合併を含むビジネス成長、組織の展開など）。
- ブラスコ氏は、非営利のローカライゼーション・翻訳振興組織であるロゼッタファウンデーション（The Rosetta Foundation：アイルランドのダブリンが本拠地）のボードメンバーである。
- ・ ロゼッタファウンデーションは、言語差別に基づく知識のギャップを根絶することを目的とし、そのため、法、教育、保健などの重要情報を、小数言語（使用者の少ない一言語）、翻訳サービスの対象になりにくい言語へと翻訳する活動を支援している。この活動は、現在世界中の多くの人々が母語しか理解せず、また、その母語が翻訳の対象となっていないので、多くの重要な情報にアクセスできず、言語により知識のギャップが生まれているという状況に対応するためのものである。同組織はボランティアの翻訳者を支援団体などへ仲介する役割を担う。現在、翻訳活動には 160 の団体と 6000 名の翻訳者が登録し、100 言語を翻訳している。
- ・ ロゼッタファウンデーションは、研究機関や企業、また個人からの寄付金、技術協力などによって、トロモンズ（Trommons）⁶⁴と呼ばれるプラットフォームを開発し、管理運営している。トロモンズは、翻訳者と翻訳者を必要とする団体を仲介するオンラインプラットフォームである。
- ・ ロゼッタファウンデーションは、2009 年 9 月に、アイルランドのリメリク大学ローカライゼーション研究センターと次世代ローカライゼーションセンター（現 ADAPT センター）のスピノフ団体として設立された。2010 年には、同組織はアイルランドで慈善団体として登録されている。同組織は、2017 年 6 月に別組織である「国境なき翻訳者団」⁶⁵の一部となった。
- ・ 国境なき翻訳者団は、2010 年に非営利団体としてアメリカで設立され、特に自然災害や伝染病、戦争などが起きた地域向けに翻訳活動を行っていた。同団体とロゼッタファウンデーションと基本的に同じ性格を持つことから、2つの団体の合併が検討され、決定された。
- 現在、様々な機械技術（ドローン、3D 印刷、自動運転、産業・医療ロボットなど）が発展しているが、人工知能を実現する技術として、機械学習（マシンラーニング）、ディープニューラルネットワーク技術が登場し、ニューラル機械翻訳技術が生まれた。
- ・ 人工知能は企業の再編を促している。例えば、グーグルの CEO は同社の将来は人工知能に

⁶⁴ <https://trommons.org>

⁶⁵ <https://translatorswithoutborders.org>

かかっているとしている。今後、人工知能は、例えば、グーグルマップが単に空港への行き方を教えるだけでなく、空港までの道のりで家族へのプレゼントを購入するために必要な情報を教えることを可能にする。

- ・ だが、イーロン・マスクやビル・ゲイツ、スティーブ・ホーキンスは規制のない人工知能の利用へ明白に懸念を表明している。
- 2016年の言語産業の世界市場は400億ドルであり、2020年までに450億ドルまで成長し、2018年までの成長率は年間6.5～7.5%と見込まれている。2016年の言語技術産業の市場は290億ドルである。
- 現在、翻訳すべきコンテンツは多様である（ユーザ生成コンテンツ、カスタマーサポート、教育、eコマースなど、また、動画、Text-to-speech、Speech-to-speech）。
- 新しい機械翻訳技術に必要な要素は、クラウドベースの単純さ、シームレスな相互互換性、自動化である。
- 機械翻訳産業の再編成が行われつつある。
 - ・ スタートアップ企業の登場
 - ・ オープンソースの機械翻訳システム
 - ・ 新しい標準としての機械翻訳
 - ・ リアルタイム翻訳
 - ・ ニューラル機械翻訳と機械学習への投資
 - ・ 新しい価格モデル
 - ・ ペタバイトのデータ（1ペタバイトは1000テラバイト）
 - ・ 企業の合併と吸収
- 2017年の機械翻訳システムの利用率
 - ・ ELIA（欧州の言語サービス企業の非営利組織）の2017年の調査では、企業の43%（2016年は41%）、個人の翻訳者の33%が機械翻訳を利用している。
 - ・ 企業が機械翻訳システムを利用しない理由は費用である。グーグル翻訳などは無料で利用できるが、品質の高いカスタマイズされた機械翻訳システムを利用するには費用がかかり、投資を怖がる企業がある。
- 国境なき翻訳者とロゼッタファウンデーションの翻訳者の一人は、メイトキャットのカスタマイズされたバージョンを使っている。
 - ・ 他のシステムへの統合が可能、簡単に利用可能であるから。
- ニューラル機械翻訳は真のブレークスルーである。
 - ・ 問題は、ニューラル機械翻訳が人間のような翻訳を提供するのにどのくらいの時間がかかるのかというのではなく、機械は何を行い、人間の役割は何かというものである。
- 人工知能が解決する今後の問題：
 - ・ 「この動画をスペイン語で説明せよ」
 - ・ 「ドイツ語のこれらの文書を探し出し、中国語で要約せよ」

- ・ 「英語でこのウイルス伝染の警告を読み、ハウサ語の音声で表現せよ」
- ・ 「この写真のなかの女性が着ているドレスを見つけ、買うこと」
- 機械にはない人間の価値とは、情動的知性、クリエイティブシンキング、顧客中心の問題解決、戦略的な意思決定、人間中心の設計、チームワークとチームの形成にある。

ウィーローカライズ / タンジャ・シュミット / 企業紹介

- ウィーローカライズ (Welocalize) 社の紹介
 - ・ 2015年に翻訳した語数：11億5000万語
 - ・ 翻訳した言語：175言語以上
 - ・ 従業員：1000名以上
 - ・ オフィス数（世界中で）：21
 - ・ 世界で第7番目、アメリカで第4番目に大きい機械翻訳サービス事業者
 - ・ 本社：ドイツのザールブリュッケン
- ウィーローカライズでは、品質、カスタマーサービス、グローバルチームワーク、イノベーションが4本の支柱であり、イノベーションがビジネスの中心にある。
 - ・ 同社はCIO (Chief Innovation Officer) のポストを作り、イノベーションに関するアイデアをすべての従業員から収集するシステムを設置している。
- ウィーローカライズのイノベーション分野
 - ・ 翻訳と翻訳管理技術
 - ・ 品質と生産性のアセスメントと評価
 - ・ 機械翻訳技術と機械翻訳インテグレーション
- ウィーローカライズの最近のイノベーション
 - ・ ニューラル機械翻訳ソリューション
 - ・ オープン・ニューラル機械翻訳ツールキットの試験的運用（アジア言語系、ロマンス語系、ゲルマン語系で良い結果が出ている）
 - ・ タウスと次世代のタウスレビューツールについて協力
 - ・ Text-to-Speech ソリューションとプログラム
- ニューラル機械翻訳について、今後さらに試験する。

パンゲアニック / マニュエル・ヘランツ / 企業紹介

- パンゲアニック (Pangeanic) 社 (スペイン) は、タウスとタウスデータ協会の創設メンバーである。
- パンゲアニックは、2009年にMOSES (欧州で開発されたオープンソースの統計機械翻訳システム) を商用化した言語サービス事業者である。2010年から後編集サービスも開始した。

- ソニー・ヨーロッパや米サイベース (Sybase) 社⁶⁶向けに翻訳エンジンを提供している。
- パンゲアニックは機械翻訳技術で大量のデータを処理することができるようになり、現在は翻訳サービスよりも、機械翻訳で知られている。
- 統計機械翻訳システムは、翻訳メモリの下位分節レベルへの一般化と見なしうる。
- 統計機械翻訳システムは、ソーステキストからの不明瞭な情報を取り出すことができず、さらにもっと情報を必要とする。
- 後編集を受け身に、また消極的な態度で行ってはいならない。
- 将来的に、言語に関わる仕事はなくなることはない。
- 人工知能は人間の敵ではないが、翻訳に関わる仕事はおそらく変化するだろう。

トランスレイティッド / アレッサンドロ・カテラン / 企業紹介

- トランスレイティッド (Translated) 社は、1999 年にイタリアのローマで設立された翻訳サービス及び機械翻訳システム開発事業者である。
- 機械翻訳利用の現状
 - ・ グーグル翻訳は1日に 1000 億語を翻訳し、言語サービス事業者は 3000 億語を翻訳している。
 - ・ 現在、3000 万人以上の翻訳者を必要としているが、現実には 30 万人のプロの翻訳者しかいない。
- トランスレイティッドの特徴 1 / T-RANK
 - ・ T-RANK は、大勢の翻訳者の中から、翻訳すべき文書に最もふさわしい翻訳者を選択するシステムである。翻訳者の実績などを考慮して、統計で客観的に選択する。
- トランスレイティッドの特徴 2 / マイメモリ (MyMemory) の利用
 - ・ マイメモリは、世界最大の翻訳メモリであり、無料である。マイメモリは、欧州連合、国際連合から翻訳メモリを集め、特殊分野の最も優れた多言語のウェブサイトと提携して作られている。
- トランスレイティッドの特徴 3 / メイトキャットの利用
 - ・ メイトキャットは、EU プロジェクトで助成されて開発された CAT ツールである。メイトキャットでは、関係するすべての人がすべてのプロセスを見ることができ、透明性が高い。

メイトキャットについて (FBK のマルチェロ・フェディリコ氏とトランスレイティッドのアレッサンドロ・カテラン氏の回答)

- メイトキャットは、EU プロジェクトで開発された CAT ツールである
- メイトキャットは、インタラクティブな CAT ツールであり、翻訳者の修正から学習し、翻訳

⁶⁶ サイベースは、データベース管理システムなどの製品を専門とする企業である。

文を示唆する。だが、リルトとは技術的に異なる。リルトは、機械翻訳、翻訳メモリ、用語ベースを統合しているので、翻訳者の修正をリアルタイムで学習し、それを翻訳文の示唆に直接反映させることができ、この点で、このようなリアルタイムのインタラクティブ性を持たないメイトキャットとは異なる。

- メイトキャットはオープンソースなので、リルトとはビジネスモデルが異なる。
- メイトキャットは、個人の翻訳者と企業が使っている。現在、2万5000人が登録している。90%が個人の翻訳者であるが、トランスレイティッド、ウィーローカライズなどの翻訳サービス事業及び機械翻訳技術研究開発事業者、そして、米ebayが利用している。
- メイトキャットは、現在、フレーズベースであるが、2017年10月からMMTプロジェクトで開発したニューラル機械翻訳システムを利用できるようになる。

TC ワールドカンファレンス 2017 視察レポート

- 日程 : 2017 年 10 月 24 日～25 日
- 場所 : ドイツ・シュトゥットガルト

全体の概要

- 2017 年 10 月 24 日から 26 日の 3 日間（視察は 2 日間）にかけて、ドイツのシュトゥットガルトで TC ワールドカンファレンス 2017⁶⁷が開催された。
- 主催はビジネスカンファレンスや見本市の組織を専門とする TC ワールド社である。同社は TC ワールドカンファレンスを欧州、インド、中国で開催している。同社の最大の顧客は、欧州テクニカルコミュニケーション協会（略称は TEKOM ヨーロッパ）であり、共に TC ワールドカンファレンスを毎年ドイツのシュトゥットガルトで開催している。
- TEKOM ヨーロッパは、テクニカルコミュニケーション分野で欧州最大の職業協会であり、2013 年 11 月にドイツで 12 組織により創設された（TEKOM オーストリア、TEKOM ベルギー、TEKOM ブルガリア、TEKOM デンマーク、TEKOM フランス、TEKOM ドイツ、TEKOM ハンガリー、TEKOM イスラエル、COM&TEC イタリア、TEKOM ポーランド、TEKOM ルーマニア、TEKOM トルコ）。同団体の本拠地はシュトゥットガルトであり、700 以上の企業と大学がメンバーである。
- テクニカルコミュニケーションとは？ : テクニカルコミュニケーションとは、安全で、効率的で、実際的な製品利用（技術システム、ソフトウェア、サービス）のために、情報製品を定義し、作成し、伝達するプロセスである⁶⁸。
- TEKOM ヨーロッパの目的
 - ・ 欧州でテクニカルコミュニケーションを推進し、開発を進めること
 - ・ テクニカルコミュニケーションの品質のために欧州標準を策定すること
 - ・ 欧州内で、商業と一般向けにテクニカルコミュニケーションの重要性を高めること
 - ・ テクニカルライターの職業プロフィールを強化し、調和させること
 - ・ 欧州内の大学と教育機関の提携を振興すること
 - ・ ユーザマニュアルと操作説明向けの国際標準の開発に参加すること
- TEKOM ヨーロッパは、日本のテクニカルコミュニケーション協会（JTCA）と緊密に提携している⁶⁹。
- TC ワールドカンファレンス 2017 では、機械翻訳を含む、テクニカルコミュニケーション分野の様々なカンファレンスが開催されるとともに、同分野の見本市が開かれ、企業がスタンドを立て、製品とサービスを紹介した。機械翻訳関連のカンファレンスでは、特に、ニューラル機械翻訳（NMT）の産業による取り入れを課題とし、翻訳品質の評価の研究を行い、NMT を従

⁶⁷ <http://conferences.tekom.de/tcworld17/tcworld17/>

⁶⁸ <http://www.technical-communication.org/technical-communication/defining-technical-communication.html>

⁶⁹ <http://www.technical-communication.org/cooperation-between-tekom-and-jtca.html>

来の統計的機械翻訳技術（SMT）などと比較するプレゼンテーションが多かった。

会場の外観



- 以下に、TC ワールドカンファレンス 2017 で行われた GALA ローカライゼーションフォーラムと言語技術カンファレンスの模様と機械翻訳技術開発企業へのスタンドでのインタビューの模様を収録する⁷⁰。

GALA ローカライゼーションフォーラムの概要

「翻訳産業向けの迅速に学習するアダプティブ機械翻訳」 / 伊 FBK⁷¹ マルチェロ・フェデリコ氏

- MMT (Modern MT) プロジェクトはホライゾン 2020 のプロジェクトである⁷²。
- MMT プロジェクトの開発要素
 - ・ すぐに利用できる基本モデル
 - ・ 新しいユーザデータを同化可能
 - ・ リアルタイムの学習
 - ・ その場での適応
- ニューラル機械翻訳システム (NMT) が登場し、これまで主流であったフレーズベース統計的機械翻訳システムは廃れていきつつある。
- NMT には、どのプロジェクトにも対応する一般 NMT (generic NMT) と、プロジェクト、ユ

⁷⁰ カンファレンスの幾つかのプレゼンテーション資料は、以下のウェブサイトから入手できる。

<http://conferences.tekom.de/tcworld17/conference-program/download-presentation-slides/>

⁷¹ FBK はイタリアのトレントにある研究機関である。

⁷² MMT (Modern MT) プロジェクトの概要

- 参加組織：トランスレイティッド（コーディネータ）、FBK、英エジンバラ大学、蘭タウス
- 研究期間：2015 年 1 月～2017 年 12 月
- 予算：369 万 5200 ユーロ（EU 拠出分は 299 万 4700 ユーロ）

ーザ毎にカスタマイズされた特殊 NMT (specialized NMT) がある。前者は安価であるが翻訳の品質が低く、反対に後者は、品質は高いが高価であるという長所と短所を持つ。

- MMT プロジェクトでは、アダプティブ NMT (adaptive NMT) を開発し、翻訳の品質が高く、安価なシステムを実現する。
- アダプティブ NMT は、すぐに全てのユーザに適応することができる機械翻訳システムである。
- アダプティブ MT は、一般 NMT を内蔵メモリによって強化したものである。
- 各翻訳者は自分の翻訳メモリのデータをアダプティブ NMT システム内のメモリに同調させる。
- ユーザが翻訳する際には、内蔵メモリ内のデータが利用される。
- 後編集の際には、常にそのデータが内蔵メモリをアップデートしていく。
- 複数のユーザが各自の翻訳メモリを一つの機械翻訳システムの内蔵メモリに同調できる。
- 各ユーザは自分の翻訳メモリを使って、翻訳作業を行う。
- 一般 NMT、特殊 NMT、アダプティブ NMT の翻訳品質を、BLEU を使って、欧州中央銀行 (ECB)、国際連合の文書 (UN-TM) などを使って比較した。アダプティブ NMT は最も良い結果を得た⁷³。
- 内蔵メモリのデータは、ニューラルネットワークの適応に限られて利用され、プライバシーは保護される。

「カンタン MT プラットフォーム上での SMT から NMT への移行：実験の結果」/ カンタン MT トニー・オドウド氏

- 機械翻訳技術としては、現在ルールベース機械翻訳 (RBMT)、統計的機械翻訳 (SMT)、ニューラル機械翻訳 (NMT) の 3 つの技術があり、現在は NMT が SMT を超える技術として注目が集まっている。
- NMT には、韓シストラン⁷⁴、米グーグル、英 SDL、米マイクロソフト、米フェイスブックなどの機械翻訳や IT 分野の大手企業が関心を寄せ、開発を進めているが、カンタン MT (アイルランド) はかなり早い段階から研究を行っている。
- カンタン MT は、SMT と NMT の翻訳品質を比較する実験を以下の条件で行った。
 - ・ 同一のトレーニングデータセット
 - ・ 同一のテスト参照セット
 - ・ 自動評価 : F-Measure、TER、BLEU
 - ・ ネイティブスピーカーのプロの翻訳者による評価
 - ・ NMT エンジンには、GPU プロセッサを使う KantanNeural を利用し、SMT エンジンには、CPU プロセッサを使う KantanMT を利用した。
- 英語から他の言語 (ドイツ語、中国語、日本語、イタリア語、スペイン語) への翻訳を対象と

⁷³ より詳しくは、論文「Multi-Domain Neural Machine Translation through Unsupervised Adaptation」を参考のこと。
<http://www.statmt.org/wmt17/pdf/WMT13.pdf>

⁷⁴ シストラン社は仏企業であったが、2014 年に韓国企業 CSLi によって買収されている。
<http://www.systransoft.com>

し、法、医療、技術の分野を指定した。

- トレーニングには、SMT よりも NMT に多くの時間がかかった。
- 自動評価では、SMT が NMT を上回った。
- ネイティブスピーカーのプロの翻訳者による評価 (A/B テスト) では、NMT が SMT を上回る結果が出た。A/B テストでは、評価者が翻訳文を見て、「同じ」、「SMT が良い」、「NMT が良い」という 3 つの選択から、1 つ選ぶ。
- 結論：ネイティブスピーカーのプロの翻訳者による評価では、すべてのケースにおいて、NMT が SMT の結果を上回った。自動評価の BLEU は NMT の品質を過小評価する可能性がある。

言語技術カンファレンスの概要

以下に、言語技術カンファレンスの機械翻訳に関わるプレゼンテーションの概要を記す。

「ニューラル技術登場後の商業向け機械翻訳」 / BEO 社⁷⁵ アンヌ・ベイヤー氏とDFKI⁷⁶ (ドイツ人工知能研究センター) アリョーシャ・ブルヒアルト氏⁷⁷

- 独 BEO は今後の機械翻訳システムとして、ルールベース機械翻訳 (以下、RBMT とする)、統計的機械翻訳 (SMT)、ニューラル機械翻訳 (NMT) のどれがいいか検討するため、DFKI と提携して、比較研究を行った。
- 機械翻訳技術の一般的現状：以前は、SMT として MOSES⁷⁸ (モーゼス) が使われていたが、現在ニューラル技術に注目が集まっている。
- ニューラル技術のアイデアは 1940 年代に遡るが、近年計算力が強化されるとともに、画像分類などの様々な分野での開発が進み、再び研究が進められている。
- NMT の現状：
 - ・ ネマツス⁷⁹ (Nematus) やオープン NMT⁸⁰ というオープンソースツールキットがある。
 - ・ グーグル、フェイスブック、ブッキングコムなどの産業が NMT を採用している。
 - ・ MT プロバイダ (韓シストラン、米ライオンブリッジ⁸¹、アイルランド・カンタン MT⁸²)

⁷⁵ <https://www.beo-doc.de/en/>

BEO は翻訳サービス及び機械翻訳システム開発企業であり、2003 年に創立され、本拠地はドイツのシュトゥットガルトである。

⁷⁶ https://www.dfki.de/web?set_language=en&cl=en

⁷⁷ 同氏は、DFKI の言語技術グループの研究者である (オフィスはベルリン)。

<https://www.dfki.de/web/kontakt/mitarbeiter?uid=albu02>

⁷⁸ MOSES は英エジンバラ大学の研究者によって開発された SMT システムである。MOSES は欧州で開発されたオープンソースの SMT システムとして広く利用されている。

<http://www.statmt.org/moses/>

⁷⁹ <https://github.com/EdinburghNLP/nematus>

⁸⁰ オープン NMT はハーバード NLP によって開発され、シストラン社からサポートを受けている。

<http://openmt.net>

⁸¹ ライオンブリッジは、米翻訳・ローカライゼーション企業である。

<https://www.lionbridge.com/en-us>

⁸² カンタン MT は、機械翻訳技術開発企業である。

<https://www.kantanmt.com/index.php>

から初期の NMT ソリューションが提供され始めている。

- RBMT、SMT、NMT を比較するために、自動評価尺度である BLEU などの他、DFKI はテストスイートアプローチを採用した⁸³。
 - ・ テストスイートは自然言語処理ではよく使われる評価尺度であり、機械翻訳にこの評価尺度を採用した。これにより、機械翻訳を質的、量的に評価できる。
 - ・ 欧州連合のホライズン 2020 から助成されている QT21 プロジェクト⁸⁴で、ドイツ語-英語ペアの以下の項目からなるテストスイートを作成した。
 - Ambiguity, Composition, Coordination&ellipsis, False friends, Function words, LDD&interrogative, MWE, NE&terminology, Negation, Subordination, Verb tense/aspect/mood, Verb valency
 - ・ QT21 プロジェクトで開発したテストスイートを用いて、以下の様々な機械翻訳システムを比較評価した。
 - O-PBMT: 古い SMT のグーグル翻訳 (オンライン / 2016 年 2 月)
 - O-NMT: 新しい NMT のグーグル翻訳 (オンライン / 2016 年 2 月)
 - RBMT: ルールベース MT
 - PBMT: オープンソース SMT (MOSES の利用)
 - DFKI-NMT: DFKI の NMT
 - RWTH-NMT: RWTH の NMT
 - ED-NMT: ネマツスのツールキットを使って構築されたエジンバラ大学の NMT
 - ・ NMT の中では、グーグル翻訳が高い結果を出した。
 - ・ さらに、テストスイートを使って、エジンバラ大学の NMT、SMT のグーグル翻訳、NMT のグーグル翻訳、独 DeepL⁸⁵の NMT を比較した。
 - ・ BEO の顧客のデータを使い、より具体的な場面に置きながら、機械翻訳の自動評価尺度である BLEU 及び METCOR とマニュアル評価で、MOSES の SMT とエジンバラ大学の NMT (WMT2016 で優勝している) を比較した。グーグル翻訳は優秀だが、セキュリティの問題で利用しなかった⁸⁶。

⁸³ DFKI のテストスイートの開発に関して、より詳しくは以下のウェブサイトを参考のこと。

<https://www.gala-global.org/publications/comparing-errors-neural-mt-vs-traditional-phrase-based-and-rule-based-mt>

⁸⁴ ホライズン 2020 の QT21 プロジェクトには欧州の有名な研究機関が参加している。

同プロジェクトの概要：研究期間 / 2015 年 2 月～2018 年 1 月 (36 か月間)、全予算 / 399 万ユーロ、プロジェクトコーディネーター / ドイツ人工知能センター (独)、RWTH アーヘン大学 (独)、アムステルダム大学 (蘭)、ダブリンシティ大学 (アイルランド)、エジンバラ大学 (英)、カールスルーエ技術研究院 (独)、フランス国立科学研究センター (仏)、プラハ・カレル大学 (チェコ)、FBK (伊)、シーフィールド大学 (英)、タウス (蘭)、テキスト&フォーム (独)、Tilde (ラトビア)、香港科学技術大学 (中)

⁸⁵ DeepL はドイツの機械翻訳システム開発企業である。

<https://www.deepl.com>

⁸⁶ BLEU とマニュアル評価の研究に関しては、以下の論文とウェビナーを参考のこと。

論文： https://ufal.mff.cuni.cz/eamt2017/user-project-product-papers/papers/user/EAMT2017_paper_32.pdf

ウェビナー： <http://www.qt21.eu/webinars/> の「Comparing Errors of Neural MT with Errors of "Traditional" Phrase-based and

- ・ 自動評価では、SMT が NMT を上回る結果だった。
- ・ マニュアル評価に関しては、ほとんどの点で NMT が高い評価を得た (Terminology と Tagging を除く)。
- 結論：
 - ・ NMT はまだ完全ではないが、SMT を超えて、情報を捉えることが可能である。
 - ・ NMT の開発に注目し続ける必要がある。
 - ・ 特殊なケース向けに生産性に係る研究が必要である。
 - ・ SMT と NMT では異なる間違いを犯すので、後編集を行う者がそれに慣れる必要がある。

「機械翻訳を使う際の翻訳者の生産性改善を自動的に計算すること」/ XTM インターナショナル：アンジェイ・ズイドロン氏

- 英 XTM インターナショナル社⁸⁷とダブリンシティ大学 (アイルランド) の共同研究。
- 諸言語は形態 (morphology) と文法で互いに異なる。英語、北京語、仏語は単純な形態を有し、ロシア語、フィンランド語、トルコ語は非常に豊かな形態を持つ。
- 機械翻訳 (MT) の発達の歴史：
 - ・ 1950 年代：ルールベース MT
 - ・ 2000 年代：統計的ワードベース MT
 - ・ 2008 年：統計的フレーズベース MT
 - ・ 2012 年：統計的ハイブリッドワード/フレーズ+文法
 - ・ 2015 年：統計的ディープラーニング
 - ニューラルネットワークなど
- MT を使うことによる生産性の改善をどのように査定できるのかという問題に取り組んだ。
- 機械翻訳の理論的境界：形態論的に類似しているほど、2つの言語は近似している。
 - ・ アメリカ英語とイギリス英語の形態論的な言語近似値：100%
 - ・ 英語とフランス語の近似値：80%
 - ・ 英語とドイツ語：60%
 - ・ 英語と中国語：40%
 - ・ 英語と日本語：25%
- PRTE (翻訳者の努力の割合減少評価：Estimating Percentage Reduction in Translator Effort) を、LC (言語近似値：Language Closeness)、TSS (トレーニングセットサイズ：Training Set Size)、DMS (ドメイン類似：Domain Similarity) によって計算した。
- 例 1：アメリカ英語をイギリス英語に翻訳：LC の値は 1 である。TSS が理想値の 1、DMS も理想値の 1 と想定する。PRET は、 $1 \times 1 \times 1 \times 100 = 100\%$ である。
- 例 2：アメリカ英語をフランス語に翻訳：LC の値は 0.8 である。TSS が 0.75 だとし、DMS が

rule-based MT」

⁸⁷ <https://xtm-intl.com>

XTM インターナショナル社は、機械翻訳システム開発企業であり、日本にも支社を持つ。

理想値の1と想定する。PRETは、 $0.8 \times 0.75 \times 1 \times 100 = 60\%$ である。

- 例3：アメリカ英語を日本語に翻訳：LCの値は0.2である。TSSが理想値の1、DMSも理想値の1と想定する。PRETは、 $2 \times 1 \times 1 \times 100 = 20\%$ である。

機械翻訳技術開発企業へのインタビュー調査の概要

- TCワールドカンファレンスと同時に、テクニカルコミュニケーション分野の見本市が開催され、関連企業がスタンドを立てた。以下に、キルグレイ・トランスレーションテクノロジー（ハンガリー）、メムソース（チェコ）、カンタンMT（アイルランド）、SDLのスタンドでのインタビューの様態を記す。

会場の様子



キルグレイ・トランスレーションテクノロジー（ハンガリー）

- キルグレイ・トランスレーションテクノロジーは、クラウドベースのCATツールである memoQ を開発しているソフトウェア開発企業である⁸⁸。
- 顧客は翻訳サービス企業などである。
- memoQにはどんな機械翻訳エンジンもプラグインでき、最新のバージョンは memoQ8.2 である。
- memoQのパートナー企業である Tauyou 社（スペイン）⁸⁹は、後編集から学ぶセルフラーニング機械翻訳システムを開発しており、memoQ上で利用できる。
- memoQデーが日本で2017年11月30日（東京）と12月4日（大阪）に開催される⁹⁰。

⁸⁸ <https://www.memoq.com/en/>

⁸⁹ <http://www.tauyou.com/newindex.html>

⁹⁰ <https://www.memoq.com/en/memoq-day-japan>

メモソース (チェコ)

- メモソース (MEMSOURCE) は、メモソースというクラウドベースの CAT ツールを開発しているソフトウェア開発企業である⁹¹。
- 顧客は翻訳サービス企業などである。
- 利用する機械翻訳エンジンは選択する必要がある。
- メモソース自身は機械翻訳エンジンを開発していない。
- リアルタイムに機械翻訳システムが学習するためには、ニューラル機械翻訳エンジンを利用する必要があるが、利用料金が高価である。

カンタンMT (アイルランド)

- カンタン MT (kantanMT) は、機械翻訳システムを開発している企業である。
- 機械翻訳技術の開発で有名なアイルランドのアダプトセンター (Adapt center) などと提携している。
- 同社は現在、ニューラル機械翻訳システムと統計的機械翻訳システムの二つを開発している。前者に注目が集まっているが、現状ではどちらのシステムがいいかは翻訳のプロジェクト次第である。
- カンタンアナリティクス (kantanAnalytics) という翻訳品質評価システム (BLEU などを使う) をアイルランドのダブリンシティ大学と共同で開発している⁹²。
- CAT ツールは開発していない。

SDL (英)

- SDL は機械翻訳システムと翻訳ソフトウェア (CAT ツールなど) の開発と翻訳サービスを提供する英企業である。
- SDL にはアップストアがあり、アプリケーションを購入し、他社製品を利用できる⁹³。
- SDL はトラドスタジオ (Trados Studio) という CAT ツールを開発している。
- SDL は、アダプティブ MT (AdaptiveMT)⁹⁴ という機械翻訳エンジンを開発しており、トラドスタジオと一緒に利用できる。
- アダプティブ MT は、後編集から学習するプライベート機械翻訳エンジンであり、ユーザは機械翻訳エンジンをパーソナライズする。
- アダプティブ MT は最近開発されたばかりであり、まだ少数の言語にしか対応していない。
- アダプティブ MT について、ウェブ上にホワイトペーパーが発表されている⁹⁵。
 - ・ 同ホワイトペーパーの最後では、今後の課題として、アダプティブ機械翻訳の学習機能をニューラル機械翻訳システム上に装備することが挙げられている。

⁹¹ <https://www.memsource.com>

⁹² <https://www.kantanmt.com/overview-analytics.php>

⁹³ <http://appstore.sdl.com>

⁹⁴ <http://www.sdl.com/software-and-services/translation-software/machine-translation/language-cloud/adaptivemt/>

⁹⁵ <http://www.sdl.com/download/adaptivemt-white-paper/119598/>

META フォーラム2017 視察レポート

- 日程 : 2017 年 11 月 14 日 (二日目のみの参加)
- 場所 : ベルギー・ブリュッセル

META フォーラムについて

- META (メタ)⁹⁶は、欧州連合 (EU) のプロジェクトである META-NET⁹⁷によって設立された言語技術関連の研究開発機関、企業などの組織の連合体である。
- META には 67 か国から 800 以上のメンバーが参加している。
- META-NET は 34 か国の 60 の研究機関から構成されるグループであり、議長はプラハ・カレル大学 (チェコ)、事務局はドイツ人工知能研究所 (DFKI) である。
- META フォーラムは、年に一度開催される META メンバーによる会議である。現在、META フォーラムは、ホライゾン 2020 プロジェクトの CRACKER プロジェクト⁹⁸によって助成されている。
- META フォーラム 2017 は「ヒューマンランゲージプロジェクト (Human Language Project : HLT) へ向けて」を題とし、言語技術について様々な角度からプレゼンテーション発表が行われた。二日目は特に機械翻訳分野の発表が多かった。
- EU には、FET フラッグシッププロジェクトという特定の研究分野へ長期間に渡って、多額の予算を助成する助成プログラムがあり、2013 年から、脳研究、グラフェンの研究に助成されており、2016 年に量子技術が第 3 のテーマとして選択された。欧州の言語技術研究者たちは、ドイツ人工知能センターが中心となって、EU の FET フラッグシッププロジェクトに応募する予定である。中心となる研究テーマは、深層自然言語理解 (Deep Natural Language Understanding)で、期間は 10 年間、2030 年まで、全予算は 10 億ユーロになる予定である。

⁹⁶ <http://www.meta-net.eu>

⁹⁷ <http://www.meta-net.eu/mission>

⁹⁸ <http://cracker-project.eu>

写真：会場の様子



全体のポイント

- META フォーラム 2017（二日目）で最も重要なキーワードは、ニューラル機械翻訳技術であった。欧州の機械翻訳技術研究のパネルディスカッションでは、議題としてニューラル機械翻訳が選ばれ、また、それ以外のセッションでも、多くのスピーカーが同技術に言及し、欧州における同技術への注目度の高さが伺える。
- 欧州では、ニューラル機械翻訳の開発が急速に進められ、産業にも採用されつつあるが、欧州の研究開発者は同技術の登場に柔軟に対応しており、統計的機械翻訳システムからの切り替えが大きく進んでいること、そして、シストラン社（韓）、SDL 社（英）、ティルド社（ラトビア）などの機械翻訳企業は、積極的にニューラル機械翻訳システムを取り入れていることが分かった。
- 欧州固有の事情として特に注目すべきことは、欧州委員会翻訳総局が、統計的機械翻訳システムに基づく機械翻訳サービスである「MT@EC」から「e トランスレーション」というニューラル機械翻訳システムに基づくサービスに 2017 年秋に切り替えていることである。 META フォーラムは毎年、際立った活動を行った言語技術関連組織に META プライズという賞を授与しているが、今年も欧州委員会翻訳総局が受賞者の一組織となった。

- eトランスレーションに技術を提供しているのはティルドである。ティルドによれば、ニューラル機械翻訳は、欧州の利用者が少ない言語（バルト3国の言語など）で流暢な翻訳文を実現する。
- リルトのインタラクティブ・アダプティブ機械翻訳、あるいはSDLのアダプティブ機械翻訳、また、この種の機械翻訳技術については、話題に昇ることがなく、ニューラル機械翻訳と比べると注目されていないことがわかった。
- だが、機械翻訳の自己学習する機能について言及する研究者もあり、FBK（伊）とカンタンMT社（アイルランド）の研究者は、翻訳精度を向上させるために、ユーザのフィードバックから学習する機能の重要性について述べた。
- FBKは、EUのMMTプロジェクトでニューラル・アダプティブ機械翻訳システムの開発を進めている。

META フォーラム 第2日目のセッションの概要

以下に、欧州言語技術産業、デジタル公共サービス向けの言語技術、EUに助成されている研究プロジェクトのセッションの概要を記す⁹⁹。

セッション：欧州言語技術産業

言語技術開発企業による製品や翻訳事業などの紹介が行われた。

アーティフィシアルソリューション社（スペイン）

- アーティフィシアルソリューションは、2001年に設立された（従業員は100名以上）。
- 同社は、テネオ（TENEIO）と呼ばれる会話ユーザインターフェイスを構築するプラットフォームを開発している¹⁰⁰（30言語以上に対応）。

⁹⁹ スピーカーのプレゼンテーション資料がウェブサイト上で入手可能である。

<http://www.meta-net.eu/events/meta-forum-2017/programme>

¹⁰⁰ <https://www.artificial-solutions.com/teneio>

- テネオは、音声認識及び音声合成機能とともに、人工知能が搭載されており、過去に行った会話を覚えて、ユーザはより複雑な質問をすることが可能である。
- アーティフィシシャルソリューションは、テネオを利用して、リラ (Lyra)¹⁰¹と呼ばれる音声認識機能モバイルアプリケーションを開発している。
- テネオは、モバイルアプリケーションのほか、スマートウェブサイト、スマートホーム、コネクティッドカー、IoT ソリューション、ウェアブル端末の開発に利用できる。

AX セマンティック社 (ドイツ)

- AX セマンティック¹⁰²は、2001年にシュトゥットガルトで設立された (従業員は50名以上)。
- 同社は、生のデータから情報を処理し、テキストを自動的に生成するソフトウェアを開発している (22言語に対応する)。
- 他言語のeコマース、リアルタイムの選挙速報ウェブサービスのテキストの生成などに利用されている。

ブッキングコム社 (アメリカ)

- ブッキングコムは、1996年にアムステルダムで設立され、2005年以来、プライライニンググループ (アメリカ) の参加である。世界中に198オフィスあり、1万5000名以上の従業員がいる。
- 同社は、世界一のホテル予約ウェブサイトであり、43言語に対応する。
- ブッキングコムでは、同社のホテル予約ウェブサイトの翻訳に、機械翻訳を利用している。目標は言語の壁を感じさせることなく、ユーザが部屋を予約できるようにすることである。現状では、3分の2のホテル予約が英語以外の言語でされており、このような状況に対応するために、ウェブサイトを適切にローカライズする必要がある。
- 現在、ホテルの概要説明は人間が翻訳している。カスタマレビューは翻訳されていないが、ユーザ経験の重要性が増大しており、翻訳する必要がある。

¹⁰¹ <https://www.heylyra.com>

¹⁰² <https://www.ax-semantic.com/en>

- 同社は、ニューラル機械翻訳を利用し、独自の機械翻訳評価プロセスを取り入れている（BLEU、妥当性/流暢さスコア、ビジネスセンシティブティ分析、A/B テスト）。

シストラン社（韓国）

- シストランは、機械翻訳システムのパイオニアであり、1960年代にアメリカで設立されたが、その後、フランスに本拠地を移した。現在は韓国に本拠地を持つが、研究開発はフランスのパリで実施されている。従業員は200名で、収益の25%以上を研究開発に使っている。同社の機械翻訳システムは140以上の言語ペアに対応する。
- 同社は現在ニューラル機械翻訳システムに力を入れており、同社のシステムを「ピュア NMT」と呼んでいる。
- 同社は、RNN（Recurrent Neural Network）と呼ばれる人間の脳に着想を得た自己学習技術を開発している。
- シストランは、米ハーバード大学のハーバード NLP グループと共に、「オープン NMT」¹⁰³というニューラル機械翻訳のオープンツールキットを開発している。

メモソース社（チェコ）

- メモソースは、2010年にプラハで設立された¹⁰⁴。
- クラウドベースの翻訳管理システムを開発し、提供している。顧客は、大半が言語サービス企業であるが、自動車、ゲーム、IT部門の企業も顧客である。

トランスレイティッド社（イタリア）

- ・ トランスレイティッド¹⁰⁵は、1999年に設立された翻訳及び機械翻訳開発企業である。
- ・ EUプロジェクトで、メイトキャット（MATECAT）というCATツールとMMTというニューラル機械翻訳エンジンの開発に参加している。

¹⁰³ <http://opennmt.net>

¹⁰⁴ <https://www.memsource.com/ja/>

¹⁰⁵ <https://www.translated.net/en/>

- ・ 同社は翻訳企業でもあり、T ランクというプロジェクトマネージャー向けのサービスも提供している。
- ・ 翻訳企業は多数の翻訳者を抱えているが、どのプロジェクトにどの翻訳者に割り振るかは重要な問題の一つである。
- ・ T ランクは、翻訳者の経験のデータベースであり、翻訳者を選択することを支援するシステムである。
- ・ プロジェクトマネージャーと T ランクを比較する実験を行ったが、T ランクの方が適切な翻訳者を選ぶ割合が高いという結果がでた。
- ・ T ランクはメイトキャットに搭載されている。

セッション： デジタル公共サービス向けの言語技術

欧州委員会の「e トランスレーション」サービスと欧州言語資源コーディネーション (ELRC) の活動の紹介が行われた。

欧州委員会翻訳総局

1. 欧州委員会翻訳総局では、MT@EC という名称で、モーゼス (MOSES) という統計的機械翻訳エンジンに基づく、24 の EU 言語に対応する機械翻訳サービスを提供していた (提供開始は 2013 年 6 月)。同サービスは諸々の EU 機関、ウェブサイト、行政業務で利用されていた。
2. 翻訳総局は、EURAMIS という翻訳されたセグメントのデータベースを保有している。
3. 2017 年 11 月に、機械翻訳システムを刷新し、MT@EC に代わって、ニューラル機械翻訳システムを利用する「e トランスレーション」¹⁰⁶というサービスの提供を開始した。

¹⁰⁶ <https://ec.europa.eu/cefdigital/wiki/display/CEFDIGITAL/eTranslation>

アテナ言語・音声処理研究院（ギリシア）

1. 欧州では、公共サービス向けの機械翻訳サービスを改善するために、欧州言語資源コーディネーション（ELRC）が言語資源を収集している。
2. ELRC は、ドイツ人工知能研究センター、評価・原語資源分配機関（ELDA）¹⁰⁷、アテナ言語・音声処理研究院（ギリシャ）、ティルド社が主導している。
3. ELRC は、言語資源を収集するために、「ELRC シェアリポジトリ」というウェブサイトを経営している¹⁰⁸。

ティルド（ラトビア）

- ・ ティルド社¹⁰⁹は、ラトビアで1991年に設立された機械翻訳技術開発企業であり、eトランスレーションに機械翻訳技術を提供している。元々は、バルト三国などの欧州の少数言語に対応する機械翻訳システムの開発で有名である。
- ・ 同社はニューラル機械翻訳システムの開発に力を入れている。
- ・ 同社は欧州委員会のほか、欧州諸国の政府、IBM、マイクロソフトなどとも提携している。

評価・原語資源分配機関

- ・ 評価・原語資源分配機関（ELDA）は、欧州言語資源協会（ELRA）とともに1995年2月に設立された¹¹⁰。ELDAは、ELRAの業務を展開し、実行する組織である。ELDAとELRAはともにフランスのパリに事務所を持つ。
- ・ ELDAは、欧州言語資源コーディネーション（ELRC）の技術ヘルプデスク、法律ヘルプデスクなどを担当している。

セッション：欧州研究エクセレンス

欧州研究評議会（ERC）によって助成されている研究開発機関によるプレゼンテーション発表が行われた。

¹⁰⁷ <http://portal.elda.org/en/about/elda/>

¹⁰⁸ <https://elrc-share.eu>

¹⁰⁹ <https://www.tilde.com>

¹¹⁰ <http://portal.elda.org/en/>

シーフィールド大学 (イギリス)

- シーフィールド大学 (ルシア・スペシア氏) は、マルチモーダル機械翻訳を開発している。特に、ニュースサイトなどを翻訳する際に、写真などの画像をテキストに添えて、機械翻訳システムによる翻訳を向上させる研究を MULTMT プロジェクト¹¹¹で行っている。

ミュンヘン大学 (ドイツ)

- ミュンヘン大学 (アレクサンダー・フレザー氏) は、機械翻訳システム向けのドメインアダプテーション技術の研究を DASMT プロジェクト¹¹²で行っている (2015年2月)。
- 当初は統計的機械翻訳システムを利用していたが、プロジェクトの途中で、ニューラル機械翻訳システムが登場し、熟考した末、後者のシステムの利用に切り替えた。同プロジェクトに助成している欧州研究評議会も、研究者のビジョンを信頼し、この方針変更を承認した。
- 統計的機械翻訳システムについて博士論文を書いたばかりの場合、どのようにニューラル機械翻訳システムに対応するかということが問題になっている。

¹¹¹ https://cordis.europa.eu/project/rcn/200709_en.html

¹¹² https://cordis.europa.eu/project/rcn/193728_en.html