

#1 Global Summit on Machine Translation

Virtual event for machine translation users,
providers, and researchers

August 16 - 20, 2021



機械翻訳サミット2021参加報告

—ユーザー・プロバイダーセッションを中心に

情報通信研究機構 田中英輝

機械翻訳サミット (MT Summit) 2021概要

機械翻訳サミット (MT Summit) について

歴史

1989年の第1回箱根開催から数えて18回目

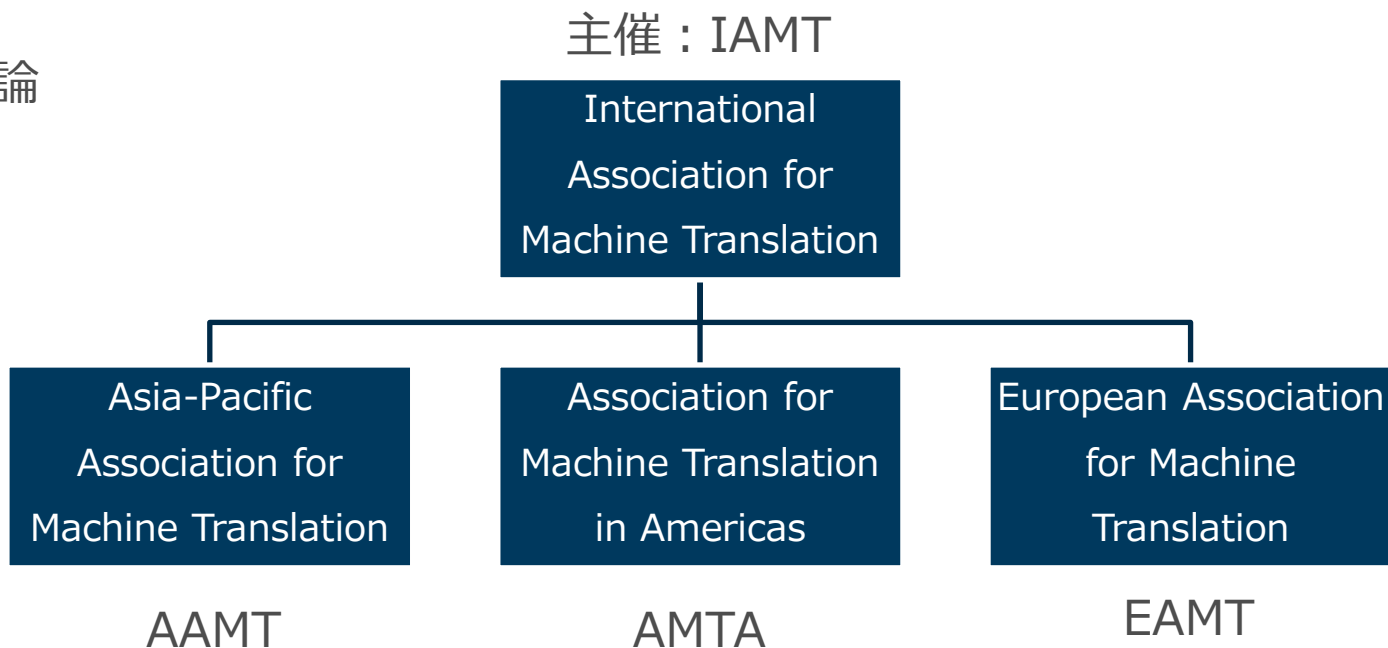
特徴

機械翻訳の研究者、開発者、ベンダー、ユーザー（大学、産業界、フリーランサー）が参加

機械翻訳の包括的議論

運営組織

3組織で隔年開催



開催形式

遠隔会議のみ

会議の構成

4 チュートリアル、4 ワークショップ、本会議、ラウンドテーブル

参加者

過去最大、669名（2019年の本会議参加者は290名）

論文トラック構成

3トラックで論文発表 合計52本(2019年は65本)
ユーザ・プロバイダ(24)、リサーチ(24)、政府(4)

発表一覧

分野	概要	https://aclanthology.org/
MTシステム構築	ヨーロッパの多言語翻訳システム構築プロジェクト	2021.mtsummit-up.5/
MTシステム構築	ヨーロッパの希少4言語のシステム構築、データ収集	2021.mtsummit-up.25/
MTシステム構築	Booking.com社の文体適応化手法、タグの利用	2021.mtsummit-up.27/
MTシステム構築	EUの手話通訳PJ紹介。音声手話テキスト混在の翻訳の計画	2021.mtsummit-up.20/
PE	HTとMTの関係。MT支援でHTは良くなるか。被験者は少ないが面白い	2021.mtsummit-up.13/
PE	MT出力とPE結果の差の観察でPE誤りを発見する方法	2021.mtsummit-up.10/
PE	実験室、実際条件でHTとPEの効率を比較。後者の向上を測定	2021.mtsummit-up.30/
PE	PEに音声を使う試み。作業が早くなるという実験結果を得た	2021.mtsummit-up.26/
PE	PEの教育応用。ポーランドの状況報告と小規模調査	2021.mtsummit-up.15/
学習データ	学習データのフィルタリング。対訳の良さの評価を自動評価で実施	2021.mtsummit-up.9/
カスタマイズ	英語-アラビア語システムの構築、カスタマイズ手順	2021.mtsummit-up.24/
カスタマイズ	MTプロバイダーが顧客のためにカスタマイズした事例のインタビュー	NA
自動・人手評価	世界知的所有権機関（WIPO）へのMT導入のための調査	2021.mtsummit-up.11/
自動・人手評価	8種の自動評価を5言語に適用して人手評価と比較	2021.mtsummit-up.29/
自動評価	memsourceで実用している自動評価からその限界を指摘	2021.mtsummit-up.21/
自動評価	翻訳ワークフローに自動評価を導入した場合の効率改善を実験	2021.mtsummit-up.22/
自動評価	翻訳自動評価システムhLEPORをPythonにポーティング	2021.mtsummit-up.28/
ターミノロジー	MTを使う場合のターミノロジーの発見、管理、利用について	2021.mtsummit-up.7/
ターミノロジー	4つのMTシステムのターミノロジーの翻訳性能	2021.mtsummit-up.8/
タグ付き文書の翻訳	タグ付き文書の翻訳法：単語アラインメントを使う方法を実施	2021.mtsummit-up.19/
生翻訳の利用	VMwareでの生MTサービスの導入報告	2021.mtsummit-up.16/
生翻訳の利用	ニュースの生MT配信実験	2021.mtsummit-up.17/
人手評価	機械翻訳システムの人手評価の課題	2021.mtsummit-up.12/
元文書分析	MTが苦手な元文書の条件検討	2021.mtsummit-up.14/

MTと周辺

上流システム

Translation Management System, Contents Management System

利用

入力

人手翻訳

機械翻訳

人手翻訳

生出力

後編集 (PE)

PE出力

開発

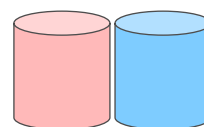
MTシステム

エンジン

モデル

ネット上の
言語データ

コンテンツ
所有者



データ獲得

大規模
対訳コーパス



傾向

- 「利用」分野で多い話題
- 生翻訳の自動評価と利用、後編集
- 自動評価は特に今回目立った
- 「開発」分野で多い話題
- カスタマイズ、システム構築



「利用」分野から生翻訳の利用、後編集、自動評価の話題を紹介

- Field Experiments of Real Time Foreign News Distribution Powered by MT
- A Rising Tide Lifts All Boats? Quality Correlation of Human and Machine Assisted Translation
- Deploying MT Quality Estimation on a large scale: Lessons learned and open questions

Field Experiments of Real Time Foreign News Distribution Powered by MT

Keiji Yasuda, Ichiro Yamada, Naoaki Okazaki, Hideki Tanaka,
Hidehiro Asaka, Takeshi Anzai, Fumiaki Sugaya

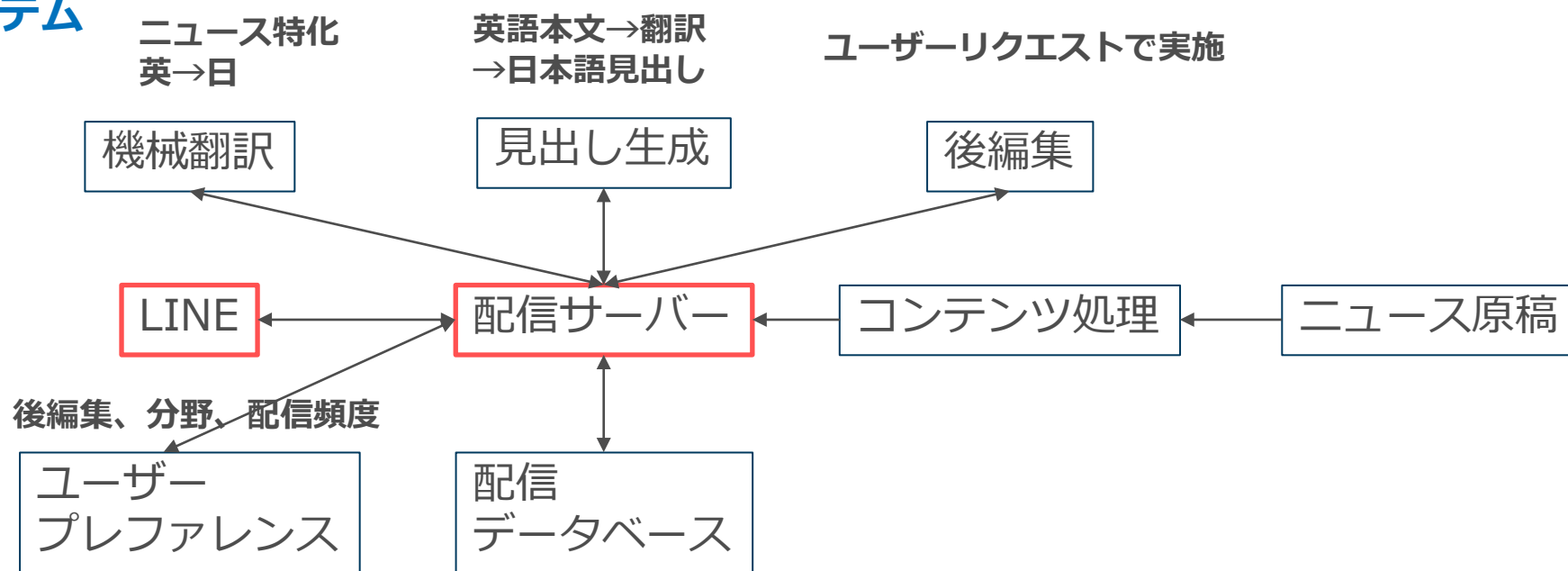
<https://aclanthology.org/2021.mtsummit-up.17/>

概要

内容

- 機械翻訳の生翻訳の利用
- 英語ニュースを日本語に翻訳してLINEでリアルタイム配信
- ユーザによる後編集リクエスト、配信分野（10）、配信頻度の指定機能を実装
- 2週間の大規模配信実験とユーザー評価

システム



実証実験

実験

- 被験者：139名
- 期間：2020/12/9-22
- 記事数：40,000

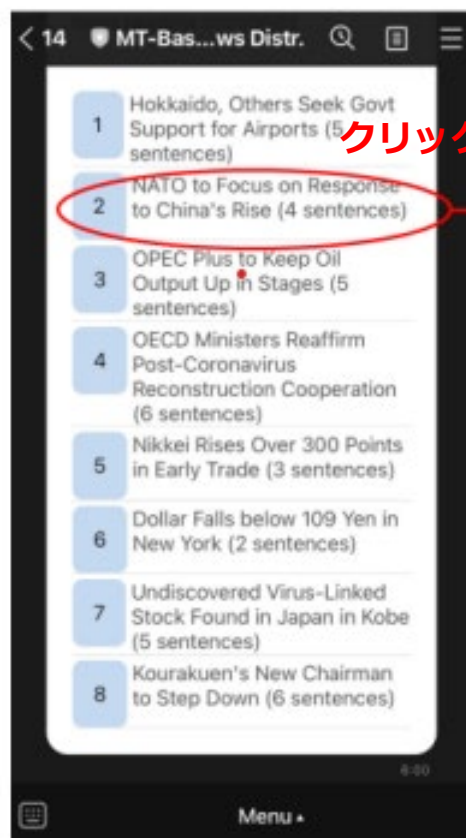
被験者のプレファレンス

- 後編集リクエスト
- 配信分野
- 配信頻度

被験者からのフィードバック

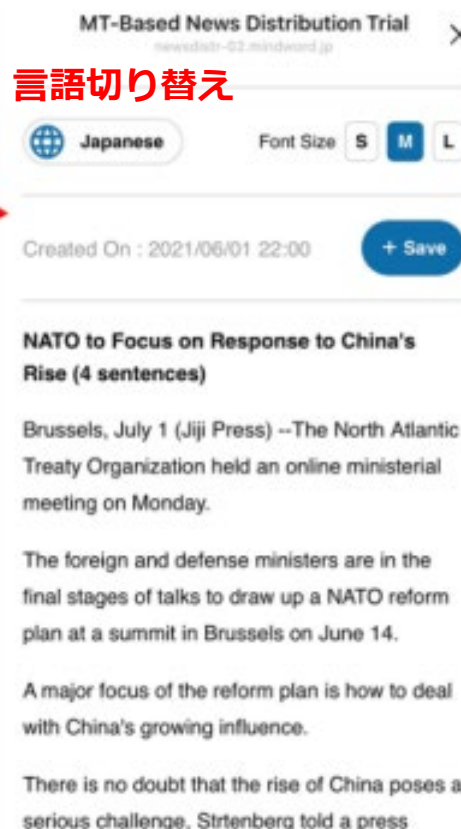
- 翻訳、見出し性能

配信画面
見出しリスト



本文

言語切り替え



実際は日本語表示（報告のため日英翻訳の英語画面を掲載）

評価結果

翻訳、見出し生成の評価

翻訳		見出し	
理解可能	74.7%	適切	79.3%
どちらでもない	14.8%	どちらでもない	10.5%
理解不能	10.5%	不適切	10.2%

分野配信比率と後編集比率

分野	配信比率	後編集比率
政治	23.3%	38.0%
経済	17.4%	18.9%
スポーツ	15.6%	17.3%
健康	14.5%	10.4%
社会	14.6%	10.0%

平均後編集時間：2時間45分

後編集による理解可能性向上率：88.9%

実サービスへの期待

- 継続利用：希望、59.7%、希望せず、11.5%

感想

- ニュースに特化した英日翻訳システムを構築
- 完璧な翻訳はできない
- リアルタイムサービスに使うためのアイデア
- 情報収集→生翻訳
- 精読→後編集
- 今後の期待
- 翻訳精度の向上、新しい話題への追従
- 生翻訳利用の拡大

A Rising Tide Lifts All Boats? Quality Correlation of Human and Machine Assisted Translation

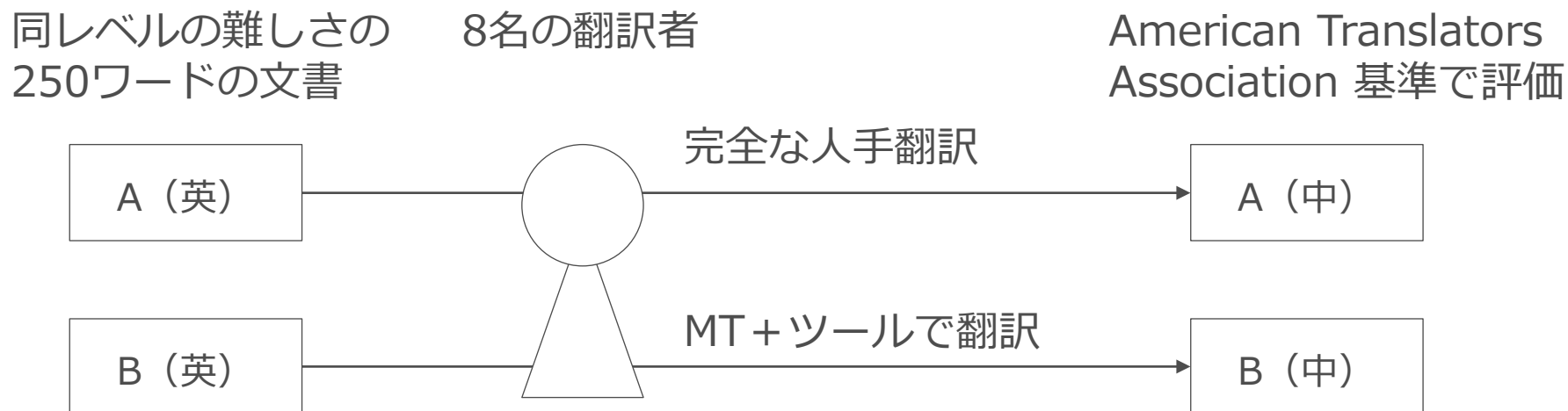
Evelyn Yang Garland, Rony Gao
<https://aclanthology.org/2021.mtsummit-up.13/>

概要

リサーチクエスチョン

1. 人手翻訳と機械支援翻訳の評価スコアには（正の）相関があるか
翻訳が上手な（下手な）人は機械を使っても上手に（下手に）翻訳
2. 翻訳と後編集は全く違うスキルか
3. 機械支援翻訳（MT+ツール）はいつも生の機械翻訳結果より良いか

調査手法（英→中翻訳実験）



結果 1 : 人手翻訳と機械支援翻訳の相関

リサーチクエスチョン 1

相関

X軸 文書Aのスコア (人手翻訳)

Y軸 文書Bのスコア (機械支援翻訳)

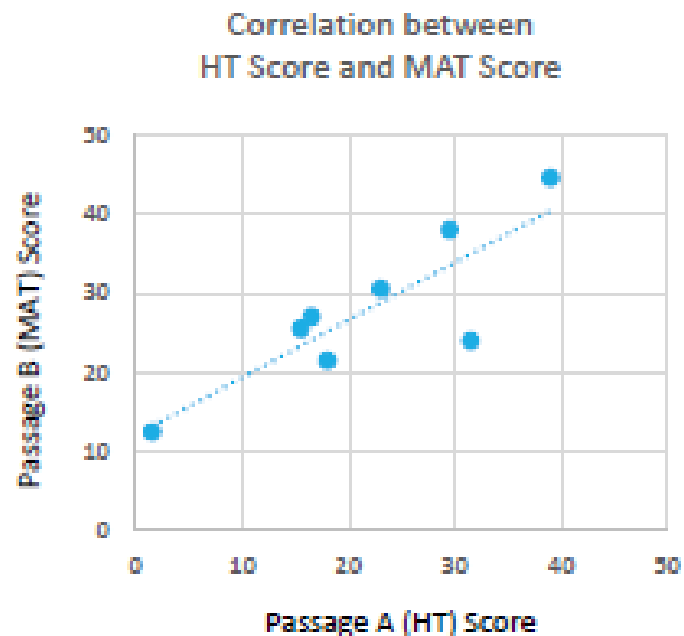
人手翻訳と機械支援翻訳のスコアには高い正の相関あり

リサーチクエスチョン 2

人手翻訳と後編集は別のスキルか？

機械支援翻訳は基本的に後編集

リサーチクエスチョン 1 の相関の高さから、両者は同じスキルと考えられる



各軸のスコアは誤りの多さを示すペナルティとなっている

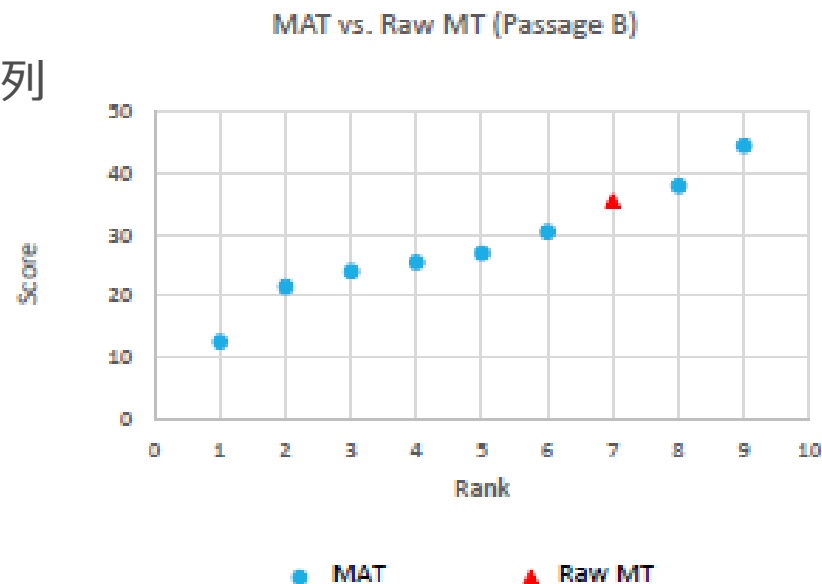
結果 2 : 機械支援翻訳は生のMTの出力より常に良いか

リサーチクエスチョン3

X軸：翻訳者を文書Bの得点が良い順に左から羅列

7位が生の機械翻訳出力

8位と9位の人々の機械支援翻訳は生出力に劣る



「機械支援翻訳は生のMTの出力より常に良い」とは言えない！

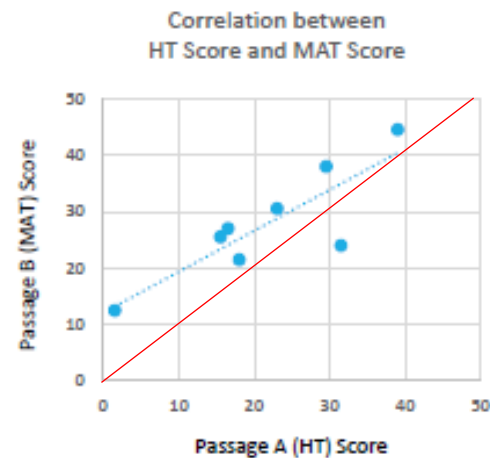
本研究の課題

- 被験者の数が少ない
- 2つの文書A,Bの難しさが同一かは不明
- 人手翻訳と機械支援翻訳の差の議論ができない
- 一つの評価基準（ATA）のみ利用

赤線：AとBが同一得点
ほとんどの点が線より上
B（機械支援翻訳）の方がペナルティ大！

感想

- 本研究：機械翻訳システムの定量的な有用性評価
- → 翻訳者への訴求効果は大きい
- 今回は翻訳の誤りの評価
- → 作業時間の変化の評価
- 今後知りたいこと
- → 機械翻訳を使うことで、個々人の能力が増強されるか



Deploying MT Quality Estimation on a large scale: Lessons learned and open questions

Aleš Tamchyna

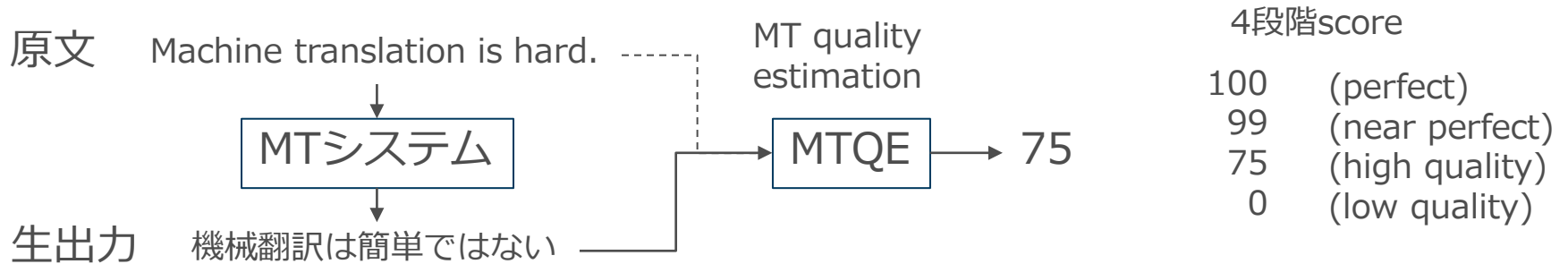
<https://aclanthology.org/2021.mtsummit-up.21/>

Memsource社での自動評価 (MTQE)

内容

- Memsource社の翻訳自動評価機能の紹介
- 3年の運用経験から見た問題点

自動評価の流れ

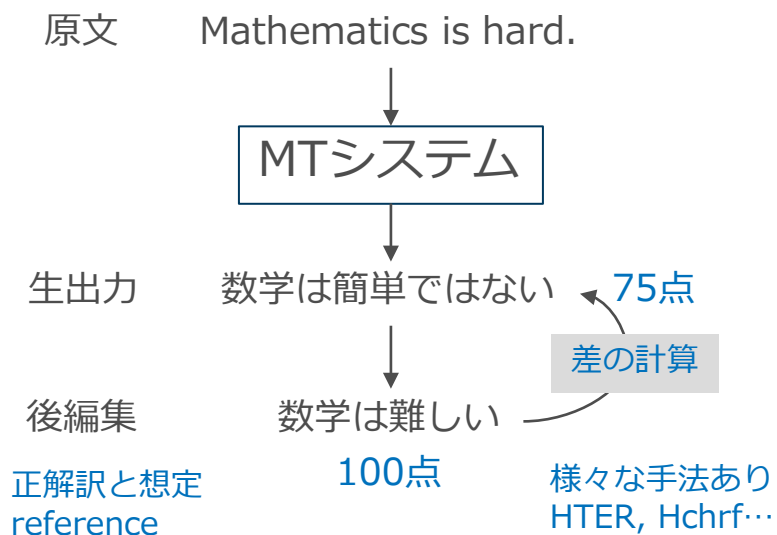


自動評価の使い道

- MTによって得られるコスト低減予測 (後編集利用の判断)
- 後編集か、最初から翻訳するかを選択支援 (翻訳者支援)
- 後編集が不要な場合を指摘 (翻訳者支援)

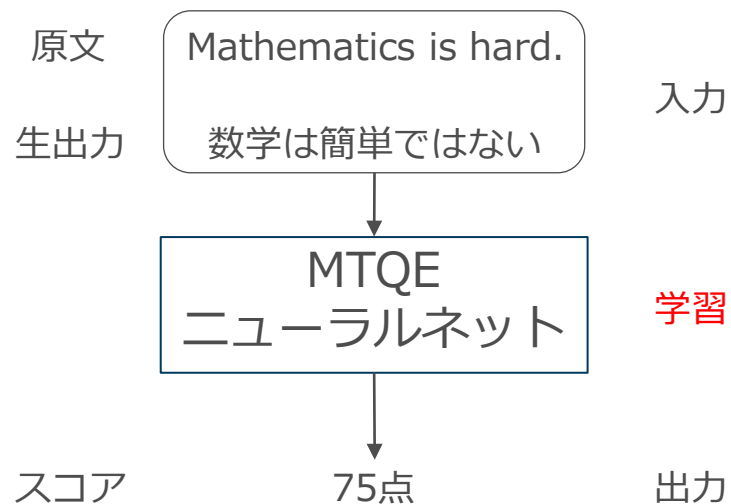
自動評価システムの構築

後編集による学習データ構築



日々の後編集作業を活用

MTQEシステムの学習



大量の「原文、生出力、スコア」データでMTQEを学習

実績

- 2018年に運用開始
- 月に1000万セグメント（文）を評価
- 130以上の言語ペアに対応
- 評価モデルを毎月更新

まとめ

課題

- 学会で高評価の評価手法が現場では必ずしも有用とは限らない
- 学習データの数、ドメイン数、品質評価の目標（固定、変化）の違いによる
- 後編集の品質ばらつき（referenceとしての適格性？）
- 顧客の予算の影響（light PEの混在）
- 翻訳者（後編集者）による後編集程度のばらつき

感想

- QEシステム利用の先進的・現実的なアプローチで参考になる点が多い
- 後編集作業を利用したエコシステム
- 後編集データはQEシステムだけでなくMTの学習にも利用可能
- 今回と同じ問題が発生するか？

おわりに

追悼

長尾眞

John Hutchins

Merle Tenney

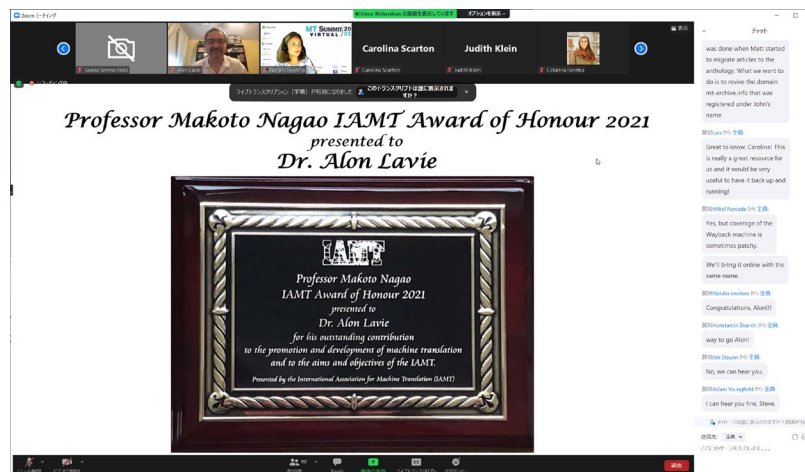


長尾眞IAMTアワード

Alon Lavie (Unbabel)

次回開催 2023年

AAMT主催：マカオ



完全オンライン開催

- 日本との時差は13時間
- 実時間での参加は疲労が大きい。一方、参加費が抑えられた
- 多くのセッションで100名以上参加
- 新しい試み：ユーザートラック発表の論文はオプション
- 2023年マカオ開催の参考にしたい

報告傾向（再掲）

- 「利用」分野で多い話題
- 生翻訳の自動評価と利用、後編集
- 自動評価は特に今回目立った
- 「開発」分野で多い話題
- カスタマイズ、システム構築