

機械翻訳サミット2019参加報告

—ヨーロッパの機械翻訳の研究開発・利用状況を中心に
NHKエンジニアリングシステム 田中英輝

0

機械翻訳サミット (MT Summit) 2019概要

1

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17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

主催：IAMT

International
Association for
Machine Translation

Asia-Pacific
Association for
Machine Translation

AAMT

Association for
Machine Translation
in Americas

AMTA

European Association
for Machine
Translation

EAMT

2

MT Summit 2019 (8/19-8/23) 概要

開催地

アイルランド、ダブリンシティ大学、The Helix



3

MT Summit 2019 概要

会議の構成

8/19-8/23の5日間で4チュートリアル、9ワークショップ、本会議

参加者

過去最大規模

参加人数

本会議	チュートリアル	ワークショップ
290	97	188

論文トラック構成

4 トラックで論文発表 合計65本

論文数

リサーチ	プロジェクト	ユーザー	翻訳者
27	9	14	15

4

ヨーロッパの機械翻訳研究・市場の特徴

EUの存在

ファンドの存在

Connecting European Facility (CEF)

Horizon

行政文書の翻訳（産業翻訳に加えて）

EU用のシステム開発が盛ん

MT（eTranslationシステム）、自動後編集システム、TM

多言語

公用語 24言語（加盟28カ国）

少数言語への配慮

言語データ獲得（例、ノルウェー語、アイルランド語、クロアチア語）

機械翻訳（例、スペイン語→カタラン語翻訳）

翻訳学の伝統

幅広い研究

翻訳者教育、MTとヒューマンファクター

5

プロジェクトトラック概観 ECファンディングの特徴

広範囲の資金投下

翻訳システム開発

iADAATPTA

CEF 2016-EU-IA-0132

INTERACT

Horizon 2020 743211

GoURMET

Horizon 2020 825299

TM開発

NEC TM DATA

CEF 2017-EU-IA-0149

自動後編集と自動評価システム開発

APE_QUEST

CEF 2017-EU-IA-0151

言語データ収集

PRINCIPLE

CEF 2018-EU-IA-0050

ParaCrawl

CEF 2018-EU-IA-0063

ミドルウェア開発

MICE

CEF 2017-EU-IA-0196

6

CEFとHorizon

CEF (Connection European Facility)

2019年の募集

機械翻訳のみで400万ユーロ (4.8億円)



Horizon2020

欧州規模の研究、開発

2014-2020年で800億ユーロ (10兆円)

ホライズン2020とは

Horizon 2020は全欧州規模で実施される、研究及び革新的開発を促進するための欧州研究・イノベーション枠組み計画 Framework Programme (FP) です。2014-2020年に渡り約800億ユーロ (約10兆円)に上るEUからの公的資金が投入されています。



7

以下の論文紹介

NMTのユースケースとして興味深かった下記2本の論文を紹介

Hungarian translators' perceptions of Neural Machine Translation
in the European Commission

Ágnes Lesznyák

MTPE in Patents: A Successful Business Story

Valeria Premoli, Elena Murgolo and Diego Cresceri

用語

- MT: Machine Translation
- NMT: Neural Machine Translation
- TM: Translation Memory
- PT: Post Editing
- CAT: Computer Aided Translation

8

概要

ECのDirectorate-General for Translation (DGT) でのNMT使用実態調査

Hungarian translator's perception of neural machine translation in the European Commission

Ágnes Lesznyák

European Commission Directorate-General for Translation

9

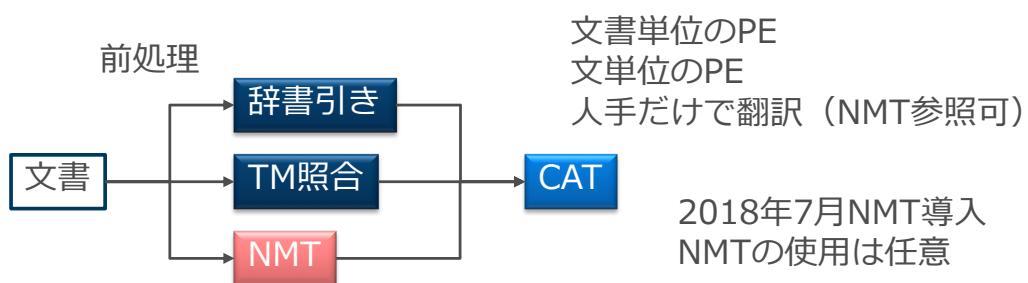
背景

eTranslation

European CommissionのMTサービス
ECのConnecting Europe Facility(CEF)ファンドでDGTで開発
SMTとNMTを提供、EU公用語の24言語、ノルウェー語、アイスランド語をカバー
NMTになってハンガリー語の翻訳性能が向上

Directorate-General for Translation (DGT) 翻訳総局

他のCommission Directorates-Generalから送られた文書を翻訳
49名の翻訳者と9名のアシスタントが所属



10

インタビュー概要

目的

NMTに対する翻訳者の関わりの質的な分析

期間

2018年6月～2019年1月

翻訳者

大半は翻訳を専攻、一部は法律、経済、エンジニアリングを専攻
インタビュー時には8ヶ月から14ヶ月のMT利用経験あり
大半がハンガリー語→英語

質問

NMTをどのタイミングで使うと有効か
NMTをどう使っているか
ソース、MT出力のどちらを最初に読むか
TMにある人手翻訳との違いは何か
NMTを使う利点は何か
翻訳を見て、NMTによる翻訳であることが分かるか

11

議論

調査結果

翻訳者によって回答がまちまち
NMTを有用と思う人もいれば、使うのをやめた人もいる

NMTの有用性に関連する要因

翻訳文書間の類似性

EUの文書は以前の文書との類似性が高い（法律など）

翻訳者は前の翻訳を利用しなくてはならず、TMが有利

NMTの全翻訳はむしろ有害

TMの照合率

高い場合、低い場合に使う人がいて一貫性がない

一般に低い場合NMTが有効だが、NMTの学習も不十分で低品質な可能性が高い

未翻訳

TMマッチがない場合に利用

緊急性

緊急時は翻訳の質が悪くとも翻訳を出すことがDGTの使命

12

議論

NMTの有用性に関連する要因 続き

ジャンル、分野

特に関連なし

ソース言語、分野への慣れ

ソース言語、分野の理解力不足の場合、NMTに頼りがち

翻訳エラー

予測不能なエラー

翻訳単位ごとに質が大きく違うのですべて確認しなくてはならない

文長

短いほど質が高い傾向あり

30~40単語を境に質が変わる

長いと誤訳がなくとも内容の把握が困難

誤訳と表面的流暢性

エラーを見つけるのは困難

13

議論

翻訳エラー続き

固有名の誤訳

存在しない (fictional) wordへ翻訳される

ターミノロジー

一貫性のない翻訳、特に法律用語で問題

NMT出力であることは容易に分かる (多くの翻訳者。3名は区別不能と回答)

利用方法

多くの翻訳者は文書単位のPEを行わない！

NMT出力を自身の翻訳に挿入

翻訳作業を翻訳者がコントロール

ソース文書を十分読んで頭で翻訳した後に、NMT出力を利用

コピペーストせずタイプし直す翻訳者も存在

文書をPEする翻訳者は少数派

30~40ワード以上の出力は消去して自分で翻訳などの対応あり

NMT出力を先に読むのは少数派

誤訳にまどわされないため

14

議論

利点とリスク

生産性

タイピングの減少による生産性向上 (半数)

タイピングは頭の中での翻訳に必須

翻訳作業の認知的に重要な作業

生産性向上は文書によるとの意見も

今回実際の翻訳速度は測定していない

心理的、認知的効果

翻訳忘れの防止

特に疲労時や原文の意味がわかりにくい時

素晴らしい翻訳に感動することもある

PE

すべて翻訳するより認知的負荷が軽い (全員が賛同ではない)

むしろPEの認知的負荷は重い

翻訳能力、クリエイティビティの低下につながる

注意不足の場合誤訳を見逃し高リスク

15

翻訳者のインタビューより

多くはNMTを有用と考えるが日々の業務には使っていない
NMTの訳質の不安定性が原因
便利なツールではあるが、問題もあり人の置き換えにはならない
自分がコントロールした状態で利用する
PEしない
誤訳が見逃されるなど新たなリスクも生じ、さらなる分析が必要
今回矛盾する回答も多かったが、新規ツールの評価では普通
ツールへの習熟が進めば有用性は増す可能性あり

16

概要

NMTの出現で翻訳品質が向上
MTPEへの関心の高まり
2年前にパテントのMTPEのビジネスをゼロから立ち上げ、成功

MTPE in Patents: A Successful Business Story

Valeria Premoli Elena Murgolo Diego Cresceri

Aglatech 14 Creative Words

17

背景とシステム構築

背景

著者らの会社：機械工学とライフサイエンスのパテント翻訳を専門
NMTの出現で顧客からPEの問い合わせが増加
ゼロからシステム、体制を構築することにする

データ収集とMT開発

機械工学分野のデータを自社TMから収集
MTプロバイダーにシステム開発を依頼

初期システム構築と評価

MTとTMと結合
2名の翻訳者でMTの性能を評価→高品質であることが判明



18

社内チームの構築

Techチームの立ち上げ

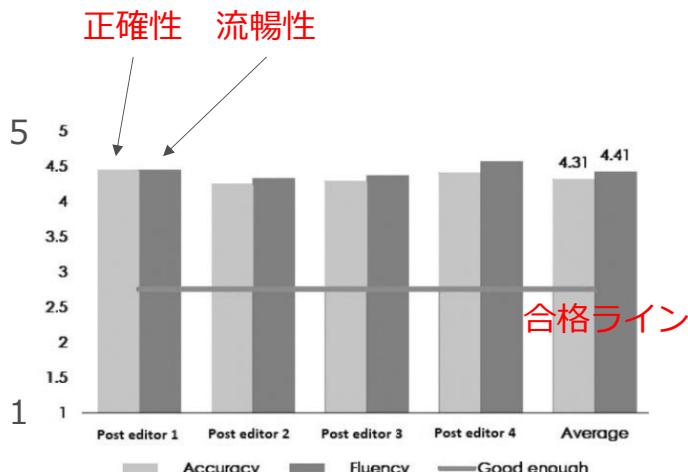
自社内の専門チームを立ち上げ
PE担当者：新卒で技術に興味のある翻訳学専攻の学生を3名採用
PE担当者：社員1名
コーディネータ：社員1名

Techチームの教育

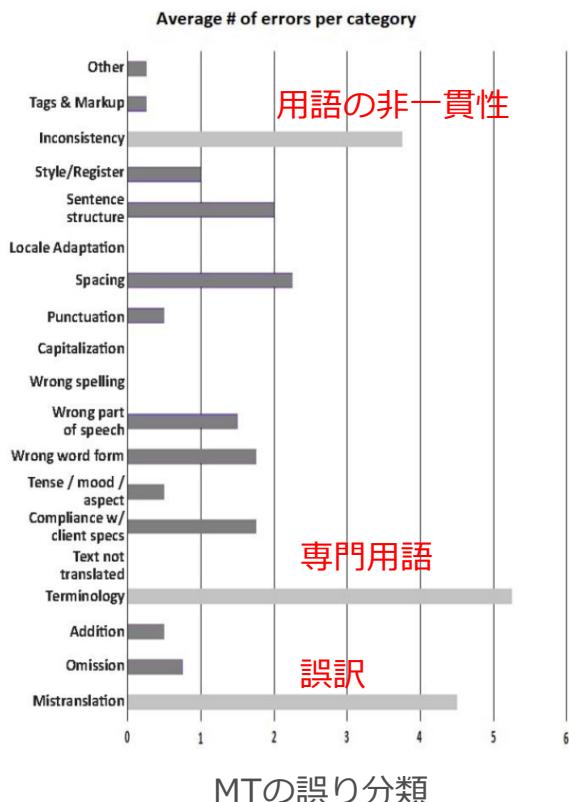
初期システムの評価翻訳者2名で数ヶ月の教育を実施
CATツールの技術的説明
MT出力の言語的な問題
実際のパテントでのPEを実施

19

社内チームによるNMT評価



TechチームPE担当者のMT評価



20

外部人材利用への展開

外部人材の獲得 外部LSPとの協業

2017年早期よりPE実施

新卒の言語学者を採用しPEを教育

12名をパテントPE担当に任命

遠隔でパテント翻訳を教育

パテントの構造

専門用語とスタイル

12名から増加中

外部人材の獲得 翻訳経験者（フリーランス）

ソーシャルメディアを通じた募集

若手言語学者：好反応

5年+の翻訳経験者：PEはレベルの低い仕事、かなり批判的

PE業務に携わった翻訳経験者

PEの単価は安く賃金への不信

MT翻訳性能への疑問

21

翻訳経験者のPE実施

活用に向けて

MTPEへの不信が強い

アンケート実施：PEの機会がないため経験なし、MTPEへの本当の不信は少数、教育希望者は多い

ミラノとボローニヤでトレーニングセッションを実施

内容に満足、PEを始めたいと思った人（23/30）

翻訳経験者の理解促進が必要

22

結論

現状と課題

現状ではMTPEは一部の分野に限定

機械工学、化学分野のイタリア語→英語翻訳

相当量を処理

2018年、MTで5.5M単語を生成

2M単語：外部でPE（12名）

3.5M単語：社内でPE（4名）

MTPEへの顧客の要望は増加傾向 → 多言語展開を検討

成功のための著者らの信条

PEは翻訳者の置き換えではない

翻訳者の価値を落とすものではない

翻訳者の機械化ではない

23

おわりに

24

本日のメッセージ

- MTサミット概要
- ヨーロッパの機械翻訳の研究・市場の特徴
 - ✓ EUの存在
 - ✓ 多言語
 - ✓ 翻訳学の伝統
- プロジェクトトラック概観
 - 広範囲の資金投下
- NMT応用事例
 - ✓ DGFのハンガリー語→英語翻訳
 - ベテラン翻訳者のインタビュー
 - 多様な（矛盾する）回答
 - NMTをあまり利用しない（使用は任意）
 - 使っていても文書単位のPEを行わない
 - ✓ イタリア語→英語のパテント翻訳
 - 新卒のPE担当者採用で成功
 - 翻訳経験者への拡大が課題

25

ご静聴ありがとうございました。

論文は「MTSummit2019」で検索。

Proceedingのページで入手可能。

本調査は動向調査は、国立研究開発法人情報通信研究機構の 委託研究の一環で行いました。