

<論文>

## 実務環境における機械翻訳テキストのポストエディット:

### 観察と統計分析

立見みどり

(応用言語学研究者・翻訳者)

#### Abstract

Machine translation systems, when they are used in a commercial context for publishing purposes, are usually used in combination with human post-editing. But human post-editing often becomes a bottleneck in the whole MT workflow. Thus understanding the process of human post-editing is crucial in order to maximise the benefit of machine translation systems. This study observes professional Japanese post-editors' work and examines the effect of the amount of editing made during post-editing, source text characteristics, and post-editing actions, on the post-editing speed. The results indicate that a number of factors, such as sentence structure, sentence function types, use of product specific terms, and post-editing patterns and behaviour, have effect on the post-editing speed in an intertwined manner.

#### 1. 研究の背景

機械翻訳システム(Machine Translation System、以下「MT」)を使って出版目的の商用文書を翻訳する場合、多くは人手によるポストエディティング(Post-editing、「後編集」とも呼ばれる。以下「PE」)を行っている。近年では、MT と PE を組み合わせることで人手翻訳と比較してコストやスピードの面で生産性が向上することが報告されており (Guerberof 2008, O'Brien 2006a, Plitt & Masselot 2010,)、さらには人手翻訳よりも高い品質が実現するケースも報告されている (Fiederer & O'Brien 2009, Koehn 2009)。それでも MT ワークフローにおいて多くの場合、PE が時間とコストのボトルネックとなっていることは否めない (DePalma & Kelly 2009)。

このため、これまでもさまざまな企業や団体で PE 作業を効率化するための手法が考案されてきたが、担当者の経験則や直感によるものが多かった。最近では PE 作業の実態について詳しい調査が行われはじめた (Plitt & Masselot 2010) が、多くは英語と FIGS(フランス語、イタリア語、ドイツ語、およびスペイン語) の間の実験結果であり、英語と日本語間の翻訳の PE においてはまだ研究が進んでいない(最近の研究として (山田 2010) がある)。

MT は、大規模な多国籍組織の中でも IT(情報技術)業界において重要な役割を果たしている。その大きな理由として、IT 業界では大量の文書を高速に安価に翻訳しなければならないというニーズがあることに加え、IT 関連のマニュアルやヘルプの内容は繰り返しが多く文章構造も類

似しているため、MT による翻訳に適しているという点が挙げられる (DePalma & Kelly 2009)。

以上に述べた背景事情から、本研究では、IT 関連文書の翻訳の中でも戦略的に重要な言語ペアのひとつである英日翻訳 (DePalma & Kelly 2009, Japan Translation Federation 2009) の PE に焦点をあて、実務環境における PE の実態を正確に把握し定量化することを目指した。

また本研究は、アイルランドの産業を振興する政府機関である Enterprise Ireland と、アイルランドにローカライゼーションの拠点を置く IT 企業 Symantec 社との共同出資による Innovation Partnership プログラムの一環として、Symantec 社とアイルランドの大学 Dublin City University の産学協同プロジェクトとして行われた。シマンテック社は、40 以上の国に支社を置く、セキュリティ、ストレージ、およびシステム管理ソフトウェアの開発・販売企業である。シマンテック社で出版する文書の量は年間 2 千万ワード以上にのぼり、その一部がおよそ 30 の言語に翻訳されている。その大量の翻訳作業の時間とコストを削減するため、すでに導入されていた翻訳メモリ(TM)に加え、2004 年に規則ベースの MT システムが導入された。

## 2. PE の定義と PE 労力の尺度

本研究では PE を、「MT 結果を編集して、1) 原文の意味を正しく伝え、かつ 2) 訳文の文法に沿ったテキストにするための作業全般」と定義する。また、MT 結果がすでに 1) と 2) の条件を満たしていて編集の必要がない場合、そのことを確認するために原文や MT 結果を読んで検討する作業も PE に含める。

この研究の目的は、PE 労力の削減に役立てるため、人手による PE のプロセスを分析し、さまざまな要因が PE の労力量に与える影響を定量化することである。しかし PE の労力量を考える際、何を基準に測定するかは意見が分かれる。代表的な測定基準のひとつはテキストの修正量である。MT 結果と PE 結果の差をとって、それを PE 労力の量と解釈するのである。しかし、テキストの修正量は PE 労力のひとつの側面でしかない。同じ文字数の修正を行う場合でも、簡単な修正もあれば難しい修正もあるだろう。また、MT 結果が十分正しく一文字も修正する必要がない場合でも、MT 結果を読み、修正する必要がないと判断するには一定の時間と思考が必要となる。このような、テキスト修正量に反映されない認知上の負荷を考える必要がある。

しかし認知負荷を直接数量的に計測する方法は確立されていない。認知負荷の計測方法として、Klings (2001) の TAP (Think Aloud Protocol)、Jakobsen (1998) のポーズ分析、O'Brien の Translog (2006a) やアイトラッキング (2006b) など多々あるが、特殊な環境が必要な場合が多く、数値の解釈方法も一定していない。このため本研究では、より単純かつ明確に認知負荷を定量化する方法として「時間」を採用することにした。実務環境においては、時間という尺度が大きな意味を持つ。時間はコストに直結し (Wagner 1987)、製品の市場投入までの時間にも大きく影響するため、経済的な角度から PE を理解する際に重要となる (Klings 2001)。これまでも多くの PE 研究において、PE 労力の尺度として時間が使われてきた (Klings 2001, Guerberof 2008, O'Brien 2006a, Plitt & Masselot 2010)。また、時間は客観的かつ容易に計測できる尺度であり (Klings 2001)、単純な仕組みで日常業務の PE 環境に組み込むことができ非侵襲的な調査が可能となるため、調査手法としても優れている。単純な数値であるため、大量のデータを比較的

簡単に処理でき、分析手法として容易に利用できるという利点もある。日常的な感覚で定量化できることから、研究結果を産業の現場で解釈する際にも便利な尺度といえる。

以上のような理由から本研究では「PE の労力」を時間的な量によって計測する。また本研究ではテキストの単位としてセンテンスを用いるので、センテンスの長さを相殺するために、センテンスの長さ(語数)を時間(分)で割った「速度」を用いる。ここでの語数は、複数のポストエディタ(PE を行う人)による異なる PE 結果の文長に左右されずに分析を行えるよう、原文の語数を用いる。

さて、PE 労力の計測尺度として、テキストの修正量と PE 速度の 2 つを考えた。ここで問題となるのが、「テキストの修正量」と、その修正に要した「時間」(PE の速度)の関係である。そこで本研究では、まずテキスト修正量と PE 速度の間にどれほどの相関関係があるのかを調べ、その後テキストの修正量以外に PE 速度に重要な影響を与える要素は何なのかをセンテンス単位で実証的、定量的に調べるという手順を取った。次のセクションではさらに詳しく、研究を具体的に進めるために設定した課題を列挙する。

### 3. 研究の課題

前述した本研究の主目的「さまざまな要因が PE の労力量に与える影響を定量化すること」を具体的に実行するため、「PE の労力量」を依存変数とし、「さまざまな要因」を複数の独立変数に分割して、それぞれについて以下のように課題 (Research Question) を設定した。

RQ1: 編集時のテキスト修正量と PE 労力量の間にはどのような相関関係があるのか。

RQ2: 原文(英語)の特徴のうち、PE 労力量に影響を与える要素は何か。

RQ3: PE 作業タイプのうち、PE 労力量に影響を与える要素は何か。

RQ4: TM(翻訳メモリ)一致文の編集作業と MT の PE 作業の間にはどのような違いがあるのか。

### 4. 独立変数

「3 研究の課題」で列挙したように、本研究では、速度で表した「PE 労力量」を依存変数とし、複数の独立変数を用いてその関係を調べることを課題としている。以下に、本研究で用いる独立変数について説明する。

#### 4.1 テキスト修正量

テキストの修正量を定量的に分析するには、MT 結果と PE 結果のテキスト上の差異をなんらかの方法で数値化する必要がある。本研究では、主観性を排除して機械的に数値化するため、MT 用の自動評価測定(AEM)プログラムを使った。AEM は本来、MT 出力結果を、人が翻訳または編集した訳文(参照訳文)と比較しその差異を数値化して MT の品質の目安とするためのプログラムで、さまざまな種類があり、それぞれ異なるアルゴリズムに基づいて計算を行う。最も頻繁

に使われているプログラムはおそらく BLEU (Papineni et al. 2002)であるが、BLEU は文書単位での評価を目的としており、センテンス単位での評価には不向きとされている。そこで本研究では、センテンス単位の評価方法として開発された GTM (General Text Matcher) (Melamed et al. 2003, Turian et al. 2003) を採用した。GTM は、2 つのテキストの間で共通している語の数を数え、語順も考慮した上で、2 つのテキストがどの程度一致しているかを 0 ~ 1 の範囲の数値で表す。完全に一致していれば 1、共通している語がひとつもなければ 0 とされるため、直感的に理解しやすい。

しかし問題がひとつある。今述べたように、GTM は一致度を評価する単位として「語」を用いるが、この場合の「語」とは、スペースで区切られたひとまとまりのテキストを指す。日本語では英語のようにスペースで語を区切ることがないので、なんらかの方法で区切る必要がある。このような場合、日本語の言語処理では「語」の代わりに、形態素で区切った「形態」を使用することが多い。本研究でもこれに沿って、テキストを形態ごとに分割することにした。しかしどこまでをひとつの形態とするかは完全に一意に定義されているわけではない。本研究では、主観性を排除し、かつ素早く形態に分割するため、単一の形態素解析プログラム MeCab<sup>1</sup> を使用した。

#### 4.2 原文の特徴

Krings (2001) は、原文の文書単位での特徴を考え、言語ペア、内容の難易度、使用語彙といった原文の特徴が PE の労力量に影響を与えると結論した。より小さな単位で原文の特徴を検討したものとしては、制限言語 (controlled language) に関する一連の研究がある。Bernth & Gdaniec (2001) は「MTranslatibility」という造語を使い、原文の文法、曖昧さ、文体、句読点の使い方、スペリングに関して定めた規則に従って原文を執筆することによって MT 品質を向上させられるとして、いくつかの規則を提案した。制限言語への準拠がどれほど PE の労力を削減できるかについては、O'Brien & Roturier (2007) や Aikawa et al. (2007) の報告があるが、いずれも PE 効率向上の効果が認められたのは、極端に長い文を避ける、正しい文法に沿って書く、曖昧さを排除する、といった限られた数の規則である。本研究では、文の長さや構造に加え、各テキストの文書内での機能(「タイトル」や「手順説明文」など)、UI(画面用語)の有無など、さまざまな要素を原文の特徴として検討した。

#### 4.3 PE 作業タイプ

PE 中にどのような作業が行われているかについてはいくつか先行研究があり (Allen 2003, Schäfer 2003, LISA's QA Model)、いずれも MT 結果に存在するエラーに関連付けて分類している。たとえば LISA の QA モデルはエラーを次の 6 種類に分類している。

1. Mistranslation
2. Accuracy
3. Terminology
4. Language

## 5. Country

## 6. Consistency

しかし MT の出力結果を調べる際、特に意味の通らない訳文が生成された場合などは、どこからどこまでを 1 つのエラーとみなすべきか悩ましいことも多く、またそのエラーは誤訳なのか用語の選択ミスなのかなど、主観を入れずに分類するのは困難である。そこで本研究では、MT のエラーに関係なく、ポストエディタが実際に実行した作業のみに注目して PE 作業を分類するという手法を考案した。この手法の詳細と、この手法に基いて定義した PE 作業タイプについては「6.1 PE 作業の分類」で詳しく述べる。

### 4.4 ポストエディタ間の差異

本研究では主に、ポストエディタ間を比較するのではなく、各要素がどのポストエディタの PE 速度にも一様に与える影響力を調べることを目的としているので、定量分析においては各ポストエディタの平均速度を相殺する必要がある。このため、ポストエディタも独立変数のひとつとしてとして含めた。また現象をより深く理解するため、作業を観察し、その記録データに基づく定性分析も行った(観察および記録の方法については「5.4 PE 作業の観察」を参照)。

## 5. 実験デザイン

過去の PE 研究は、学生などのボランティアを被験者とし、コントロールされた環境で短いテキストを使って行われることが多かった。しかしそれでは実務環境で実際に何が起きているかを調べることは難しい。そこで本研究では、a) プロの実際の仕事を、b) 普段どおりの作業環境で、c) 現実的なテキストを使用して、d) できるだけ被侵襲的に観察することを目指して実験を計画した。

### 5.1 被験者

できるだけ実務環境に近い状況でデータを収集するため、プロの翻訳者 9 名に実験に参加してもらった(うち 8 名は PE 経験者)。被験者の翻訳や PE の経験は 1 年未満～ 5 年以上まで幅広いが、Symantec 文書の PE 経験者は含まなかった。

### 5.2 データの準備

実験用データには、すでに日本語化されている Symantec 社の英文ソフトウェアマニュアル(XML 形式)を使用した。過去の調査 (Plitt & Masselot 2010, TAUS 2010) を参考に、プロの翻訳者またはポストエディタが 1 ～ 2 日以内で PE できる分量として約 5,000 ワードを取り出し、約 5 分の 4 を規則ベースの MT (Systran) で翻訳した。Systran のオプションとして Symantec 社のユーザー辞書を使用した。また Symantec 社が日本語への翻訳時に標準で使用している、XML タグを保護するためなどのプリプロセッシング用スクリプトと、既知の訳文の問題(語尾のスタイルや用語の誤用など)を修正するためのポストプロセッシング用のスクリプトも使用し、Symantec 社が実際に翻訳ベンダーに PE を依頼するときに使用するデータに近い状態にし

た。残りの 5 分の 1 は TM(翻訳メモリ)一致文の編集作業と MT の PE 作業の違いを調べるため、Symantec 社の TM から、95~99%、85~94%、75~84% の 3 種類の一致率の訳文を同量程度ずつ取り出した。IT ドキュメントを MT システムで翻訳する際、既存の TM と組み合わせて使うことは標準的な手法であり、MT 結果の PE と TM 一致文の編集作業との間にどのような類似点や相違点があるのかは興味深いテーマである。

### 5.3 PE の方法

PE には SDL Trados Translator's Workbench および TagEditor(version 7)を使用した。これらは IT 関連の翻訳において普及率の高いソフトウェアであり、被験者 9 名のうち 8 名がこれらのソフトウェアの使用経験を持っていた。PE 速度を算出するための時間の計測にも同ソフトウェアのタイムスタンプ機能を活用した。PE は、被験者が普段から慣れている環境で作業できるように、各被験者が所属する翻訳ベンダーのオフィス内で、ベンダー所有のコンピュータを使って行われた。

PE の指示書としては、Symantec 社が普段から使用しているガイドラインを基にして、PE 時に準拠すべき簡単なガイドラインを供給した。PE 成果物の基準として、「原文の意味を正しく伝えていること」、また「日本語の文法に沿っていること」を定めた。さらに、それ以上の品質にする必要がないことを明確にするため、スタイルの洗練などの編集作業は避けるよう指示した。また、特に UI(User Interface:メニュー名やオプション名などのソフトウェア画面用語)について、日本語に翻訳されていても、英語のまま残されていても、変更せずそのままにしておくよう指示し(MT テキストには両者が混在している)、用語集は提供しなかった。その他製品に関する資料なども一切提供していない。ただし被験者は PE 中に、自分が普段使っているオンラインの辞書や参考資料を自由に使用することができた。

### 5.4 PE 作業の観察

定性分析用のデータを収集するため、PE の時間と編集後のテキストだけでなく、編集のプロセスを記録した。記録用のツールとして、コンピュータの画面上での動作(カーソル移動、文字の入力や削除、アプリケーションの切り替えなど)を動画として記録し再生できる BB Flashback というソフトウェアを使用した。この記録方法では外部カメラを使用しないため、ポストエディタ自身を撮影することはない。記録されるのは画面上で発生した PE 作業のみであり、ポストエディタ自身の身体的な動作や表情などは記録されない。これは非侵襲的な観察方法であり、ポストエディタが普段通りの条件で PE 作業を行うためにも好ましい方法であったと考える。

## 6. 結果と考察

### 6.1 PE 作業の分類

データ分析の第一段階は、PE 作業を分類することである。その考え方は「4.3 PE 作業タイプ」で述べたが、実際の分類手順は以下の通りである。

1) 形態素の分割と品詞の割り当て

MeCab を使用し、MT 結果と PE 結果のテキストをそれぞれ形態に分割して、品詞を割り当てる。

2) 物理的な作業タイプの特定

TER (Translation Edit Rate) (Snover et al. 2006) を使用して、それぞれの形態に対して行われた物理的な作業タイプを特定する。TER は 前述した GTM と同様、MT 結果を参照訳文と比較して MT 品質を評価するためのプログラムであるが、アルゴリズムが異なる。TER は、2 つのテキストを比較し、一方のテキストからもう一方のテキストに変更するために必要な最小限の作業を特定して、単語(日本語の場合は形態)ごとにその作業内容(挿入、削除、置換、移動)を示す。

1) と 2) により、以下のようなリストが作成される。

MT 結果:

それらに支援スタッフとポストの質問に連絡します。

PE 結果:

ご質問はサポートスタッフまでご連絡を。

MT 結果	品詞	PE 結果	品詞	物理的作業タイプ
それら	代名詞			削除
に	格助詞	ご	接頭詞	置換
質問	サ変接続名詞	質問	サ変接続名詞	移動: -5
支援	サ変接続名詞			削除
と	格助詞	は	副助詞	置換
ポスト	一般名詞	サポート	サ変接続名詞	置換
スタッフ	一般名詞	スタッフ	一般名詞	
の	格助詞	まで	副助詞	置換
に	格助詞	ご	接頭詞	置換
連絡	サ変接続名詞	連絡	サ変接続名詞	
し	サ変動詞			削除
ます	助動詞	を	格助詞	置換
。	句読点	。	句読点	

表 1. 形態への分割、品詞の割り当て、物理的作業タイプの設定

※ 「物理的作業タイプ」欄の上から 3 行目にある「-5」は移動の距離を示す(形態「質問」が、5 つ前に移動された)。

3) コーディング

この作業リストに基づいて、筆者が手作業で PE 作業の内容をコーディングし作業カテゴリーを定義する(カテゴリーをあらかじめ定めてからコーディングするのではなく、コーディングの過程で発見した事柄を基にカテゴリーを作成)。

この結果、PE 作業は大きく「文節レベル」の作業と「構造レベル」の作業に分類できることがわかった。さらに作業内容の特性に沿って細分化し、以下の 12 の作業タイプカテゴリーを定義した。[1] ~ [6] が文節レベル、[7] ~ [10] が構造レベルの PE 作業で、[11] と [12] は訳文そのものには無関係な作業である。

#### [1] 補足

文節レベルで、1) MT 結果に存在しない概念を PE で付け足す、または 2) MT 結果にあった代名詞をそれが指示する実際の名詞で置き換える(前方照応など)こと。

#### [2] 省略

文節レベルで、MT 結果に存在する概念を削除する(代名詞を省くなど)こと。

#### [3] 置換: UI(ユーザーインターフェース用語、画面用語)

MT 結果に存在する UI(画面用語)を、別の UI に置き換えること。(被験者は UI を修正しないよう指示されていたにも関わらず、修正が行われたケースが多々あったため、この項目を定義した)

#### [4] 置換: 用語

MT 結果に存在する UI 以外の専門用語を、別の用語に置き換えること。

#### [5] 置換: 一般語

MT 結果に存在する UI でも用語でもない語句を、別の意味を持つ語句に置き換えること。

#### [6] 文節レベルのスタイル変更

MT 結果に存在する UI でも用語でもない語句を、同じ意味を持つ別の言い方に置き換えること。

#### [7] 係り受けの変更

文節間の関係を変更すること(修飾節が修飾する対象を変更するなど)。

#### [8] 書き直し

文節内および文節間の範囲を超えて、意味の変化を伴う修正を行うこと。

#### [9] 構造レベルのスタイル変更

文節内の範囲を超えて、意味の変化を伴わない修正を行うこと。

[10] 句読点の変更

句読点を追加、削除、または移動すること。

[11] タグの編集

XML タグを追加、削除、または移動すること。

[12] コメントの記載

コメントを追加、削除、または移動すること(ポストエディタには、ガイドラインの指示によって解決できない問題が PE 中に生じた際、適宜処理するか、どうしてもわからない場合は時間をかけて考えることなく簡単なコメントを残すよう指示した)。

以上に定義した PE 作業タイプは、タイプによって PE 労力量(速度)に差が出るかどうかを調べるため、独立変数として使用した。

6.2 ポストエディタごとのテキスト修正量と PE 速度の相関関係

次に、各ポストエディタの PE 速度とテキスト修正量を集計した。その結果を以下に示す。

	PE 速度(英単語/分)				GTM スコア			
	平均	標準 偏差率	最小	最大	平均	標準 偏差率	最小	最大
ポストエディタ A	16.68	81%	0.92	84.00	0.72	28%	0.17	1
ポストエディタ B	17.39	204%	0.31	468.00	0.74	30%	0.14	1
ポストエディタ C	24.33	83%	2.03	154.29	0.64	34%	0.00	1
ポストエディタ D	17.81	91%	0.97	144.00	0.63	32%	0.00	1
ポストエディタ E	20.24	98%	0.89	210.00	0.70	30%	0.00	1
ポストエディタ F	19.93	137%	1.01	380.00	0.66	32%	0.14	1
ポストエディタ G	23.87	204%	0.10	660.00	0.73	26%	0.00	1
ポストエディタ H	23.11	86%	0.48	130.91	0.63	37%	0.00	1
ポストエディタ I	38.04	86%	1.17	195.00	0.78	26%	0.00	1
ポストエディタ平均	22.38	119%	0.88	269.58	0.69	30%	0.05	1

表2 PE 速度と GTM スコアの平均

この表からわかるように、PE 速度は、例外的に高速なためここでは外れ値<sup>2</sup>とみなすポストエディタ I を除くと、1 分あたり 10 ワード以内の範囲に分散している。最低速の A (16.68 語/分) の場合、時速にすると 1,000 ワード程度、最高速の C(24.33 語/分) の場合は時速 1,500 ワード程度である。これは、Plitt & Masselot (2010) が FIGS 言語の PE について計測した結果である時速 800 ~ 1,800 語に近い数値である。

GTM スコアは大まかに言って、PE 中に修正されなかったテキストの割合を示している。たとえば、GTM が 0.75 の場合、文中で修正されたテキストの割合は約 25% である。表によると、平均して 3 分の 1 以上のテキストを修正したポストエディタもいれば、約 4 分 1 しか修正していないポストエディタもいる。

標準偏差率は値の分散度を示している。値が大きいほど各ポストエディタ自身の分散度(個々のポストエディタにおけるセンテンス間での分散度)が高く、値の差異はポストエディタ間の分散度を示す。この表では、PE 速度の標準偏差率は、GTM スコアの標準偏差率と比較して、全体に値が高く、またポストエディタ間での差異も大きい。つまり、PE において行われるテキスト修正の量はセンテンス間でもポストエディタ間でも比較的差は小さいが、PE の速度はセンテンスごとに、またポストエディタごとに大きく異なるということである。

一方、PE 速度も GTM スコアもポストエディタ間である程度の差が見られることから、より多量のテキストを修正しがちなポストエディタもいるし、より PE に時間をかけているポストエディタもいることがわかる。ここで当然生じる疑問は、「テキスト修正量が比較的多いポストエディタは、テキスト修正量が比較的小さいポストエディタと比べて、PE 速度が遅いのだろうか」という点である。そこで、平均 GTM スコアと平均 PE 速度の相関関係を調べた。以下の分散チャートにその結果を示す。

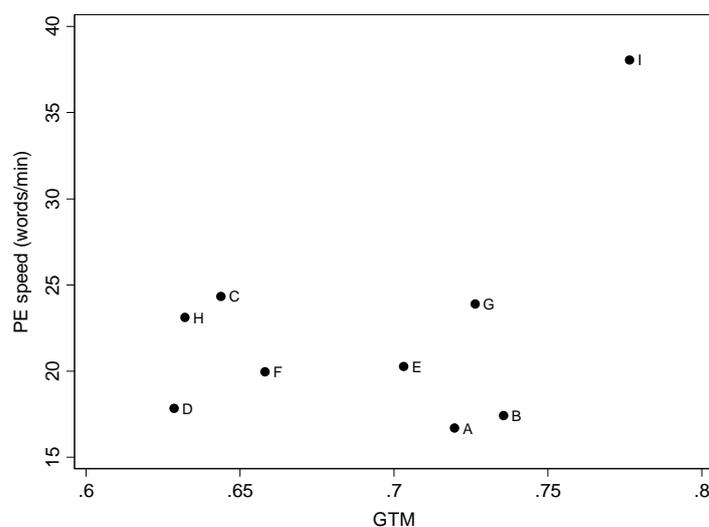


図 1 ポストエディタ間の平均 PE 速度と平均 GTM スコアの相関関係

この図に見られるように、外れ値である I を除いては、ポストエディタ間で比較した場合、平均 PE 速度と平均 GTM スコアとの間に相関関係はない。つまり、多くのテキストを修正しがちなポストエディタは、必ずしも PE が遅いわけではない。

### 6.3 センテンスごとのテキスト修正量と PE 速度の相関関係

次に、それぞれのポストエディタについて、センテンスごとのテキスト修正量と PE 速度の相関関係を調べた。以下に、最も相関関係が高かったポストエディタ A の分散チャートを示す。横軸

に GTM スコアをとり、縦軸に PE 速度をとって、各点は各センテンスを表している。点の集まりは緩やかに右方向へ向かって上昇している。これは、GTM 値が高くなる(修正量が少なくなる)につれ、PE 速度が高くなっていることを示している。ポストエディタごとにピアソン相関係数を計算して(すべて  $p < 0.001$ ) 平均をとった結果、 $r = 0.58$  となった。これは中程度の相関関係である。つまり、テキスト修正量は PE 速度の違いをある程度決定しているが、他にも速度に影響を与えている要因があることを意味している。

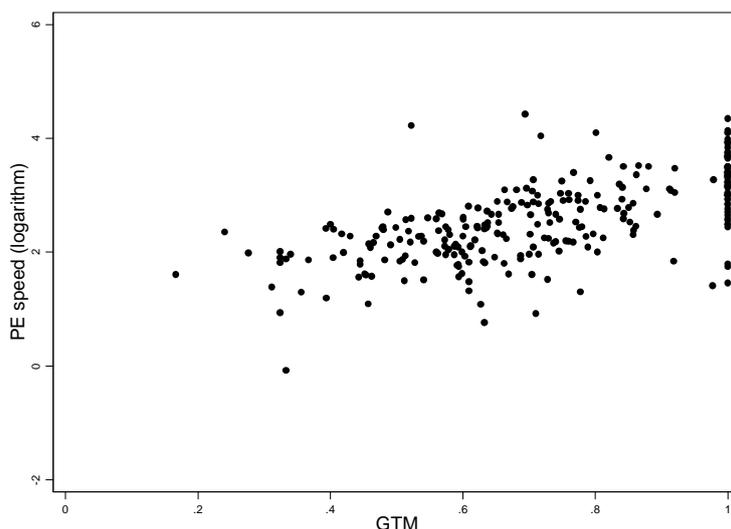


図 2 センテンスごとの PE 速度と GTM スコアの相関関係例

他の要因として考えられる事柄は先行研究によってもいくつか示されている。本研究ではそれら複数の要因が PE 速度に与える影響を、まず個別に検査し、その後重回帰分析の手法を用いて、複数の独立変数の影響を一度に測定した。重回帰分析を使用すると、複数の独立変数それぞれの影響力を、他の条件を一定にした上で計算することができる。翻訳研究において重回帰分析は確立された手法ではないが、人手による編集作業といった複数の要因が複雑に絡み合う分析には有効な手法ではないかとの考えから敢えて採用した。以下、分析の結果特に重要と思われた要因について説明する。

## 6.4 原文の特徴が PE 速度に与える影響

### 6.4.1 センテンスの構造

原文を構造別に「単文」、「重文・複文」、および「不完全文(語、体言止めの箇条書きやタイトルなど、文として完結していないもの)」の 3 種類に分けて、PE 速度に差異があるかどうか調べた(本研究では便宜上、不完全な文も「センテンス」と呼ぶ)。その結果、単文が最も速く、次に複文や重文が続き、不完全文が最も PE 速度が遅いことがわかった。これはセンテンスの長さなど他の条件を相殺しても同様であった。数値で示すと、単文と比較して、複文や重文は 13%、不完全文は 20% 遅かった。一方テキスト修正量を比較すると、単文が最も少なく、不完全文がそれよりわずかに多く、重文・複文は最も修正量が多かった。つまり不完全文は、重文や複文よりテキスト

修正量は少ないにも関わらず PE 速度は遅かったことになる。

さらに明らかになったことは、「6.3 センテンスごとのテキスト修正量と PE 速度の相関関係」で述べた修正量と PE 速度の相関関係の強さは、センテンスの構造によって異なるという点である。完全文(単文、複文、重文)ではより強く、不完全文では弱かった。数値で示すと、GTM が 0.1 上昇するごとに、単文、複文、重文では 24% 程度 PE 速度が上がるのに対し、不完全文では 17% しか上がらなかった ( $p < 0.001$ 。パーセント値は重回帰分析の結果得られた係数を基に計算した。これ以降も同じ)。つまり、テキスト修正量が PE 速度に反映される度合いは、完全文ではより高く、不完全文ではより低いことになる。

#### 6.4.2 センテンスの長さ

センテンスの長さについては、興味深い結果が得られた。先行研究により、原文が極端に短いまたは長い場合、MT 訳文の品質が低下することが示唆されており (Bernth & Gdaniec 2001、Underwood & Jongejan 2001)、その結果 PE の労力が増加するとの報告がある (Aikawa et al. 2007、O'Brien 2006a)。その理由として、非常に長いセンテンスは文法的にも意味的にも複雑さを増すこと、逆に非常に短いセンテンスやフレーズでは文脈や意味が曖昧になる可能性が高いことが挙げられる。本研究でも、文の長さだけに着目して PE 速度に与える影響力を測定したところ、原文が 5 語以下の場合の PE 速度は一様に遅く、6~11 語の場合に PE 速度が最も速く、それ以上になると、長くなるほど PE 速度が低下した。しかしこれは文の長さを単独の独立変数として影響力を調べた場合のみの結果であり、他の条件を勘案した重回帰分析では、センテンスの長さが PE 速度に有意に影響を及ぼすのは、不完全文のみであることがわかった。単文、重文、複文などの完全文では、センテンスの長さが PE 速度に統計的に有意な影響を与えることは証明されなかった。つまり、見かけ上はセンテンスの長さの違いによって PE 速度に違いが出ているように思われても、実際はおそらくセンテンスの長さとおる程度連動する別の要素(文の構造や機能など)のほうが強い影響を与えていることが示唆された。

#### 6.4.3 センテンスの機能

センテンスの機能として、セクションタイトル、表セルの内容、箇条書き、手順説明文、一般的な説明文などに分類し、PE 速度に有意な差があるかどうか調べた。その結果、手順説明文は、他の機能のセンテンスと比較して PE 速度が速いことがわかった。手順説明文は多くの場合、単文や複文であり、先のセクションで述べたように、これらのセンテンス自体が不完全文と比べて PE 速度が速い。しかしたとえば同じ単文でも、他の機能のセンテンスと比較して手順説明文は PE 速度が 37% 速かった(複文や重文では 22%)。ソフトウェアの操作手順を説明する文は多くの場合パターン化しているため、理解も簡単で、MT 結果の誤りを修正する際にも認知負荷が少ないことを示しているのかもしれない。

#### 6.4.4 UI の有無

UI(画面用語)は、人手による翻訳や編集作業の際に大きな負担となることが先行研究で示され

ている (Roturier & Lehmann 2009)。本研究でも、UI が含まれたセンテンスは、他の条件を一定にしても PE 速度が遅くなることがわかった。特に UI が日本語に翻訳されている場合に違いが顕著で、数値で示すと、翻訳済みの UI が 1 つ含まれるごとに、単文で 20%、複文や重文で 11% PE が低速になっていた (UI が含まれた不完全文は存在しなかった)。ここで特に興味深いのは、UI を変更したかしなかったかに関係なく、UI の存在自体が PE 速度に影響を与えていたことである(「5.3 PE の方法」で述べたとおり、そもそも被験者には UI を編集しないよう指示していた)。つまり、修正する必要性の有無に関係なく、そのセンテンスに UI が含まれているだけで、PE が遅くなっていたのである。この理由として、UI はマニュアルの中でも非常に重要な要素であるため、ポストエディタが PE の際に特に注意を払う必要性を感じて、認知負荷が高くなっていることが考えられる。

#### 6.5 PE 作業タイプが PE 速度に与える影響

作業タイプ別でみると、PE 速度に少しでも統計的に有意な影響を与えていたのは、「補足」、「置換:一般語」、「係り受けの変更」、「書き直し」、「句読点の変更」の 5 つで、残りの 5 つは PE 速度にほとんど影響を与えていなかった。影響を与えている 5 つの作業タイプもその影響の度合いはセンテンス構造に大きく依存し、特に単文だけに注目すると、PE 速度に影響を与えている作業タイプはひとつもなかった。不完全文に注目すると、「補足」作業が発生したときに特に PE 速度が有意に低下することがわかった。複文と重文では、「書き直し」と「句読点の変更」が統計的に有意に PE 速度を低下させていた。「句読点の変更」は、テキストの変更量は少ないが (通常は 1 文字のみの変更)、読みやすさの向上だけでなく、文全体の意味の変更にも関わることがある。これが認知負荷を増加させ、PE 速度の低下をもたらしていると考えられる。

#### 6.6 ポストエディタ間の差異

独立変数として組み込んだポストエディタ全員について、他の項目と比較して大きな統計的有意差が認められた。これは、全体的な PE 速度には、上記に述べたような原文の特徴や PE 作業タイプよりも、ポストエディタ間の差異の方が大きな影響を与えていることを意味している。なんであれ人手による作業を分析する際には付き物の問題であるが、PE も例外でなく、共通の要因と同じかそれ以上に、個人差による要因が労力に大きな影響をもたらしていることがわかった。しかし個人差の分析は本研究の対象外であるため、これ以上は言及しない。

#### 6.7 TM 一致文の編集作業との比較

次に、TM 一致文の編集作業と MT 結果の PE 作業の差異を比較する。まず、速度とテキスト修正量を比較してみた。時間的な比較については、Guerberof (2008) が英語からスペイン語への翻訳において、PE 速度は TM 80~90% 一致文の編集よりも高速であったと報告している。またテキスト修正量については、山田 (2010) が、英語から日本語への MT の PE では、TM 75% 一致文の編集と比較してテキスト修正量が多かったと報告している。本研究でも、TM 75% 一致文の編集と MT 結果の PE との比較では、PE のほうがテキスト修正量は多かった。しかし

その一方、編集速度は TM 75% 一致文の編集のほうが遅かった。つまり、MT 結果は、TM 75% 一致文と比較すると、修正の必要性は大きかったが、その修正にかかる時間は短かったということになる。

さらに、先に説明した、MT 結果の PE について行ったのと同じ重回帰分析を TM 一致文について行った結果、MT の PE と比較して、TM の編集ではテキスト修正量と PE 速度との相関関係が低かった。また、センテンスの構造、センテンスの機能(手順説明文かそれ以外か)、UI の有無を含め、PE 作業で統計的に有意な影響力が認められたほとんどの独立変数について、TM 一致文の編集では有意性が得られなかった。統計的に有意な影響力が見られたのは PE 作業タイプの「置換:一般語」と「書き直し」のみであった。ただしこの 2 つの項目が編集速度に与える影響力は、PE 速度に与える影響と比較して 3 ~ 4 倍とかなり大きいことがわかった。このことは、MT の PE と TM 一致文の編集では、その速度に影響を与える要因が大きく異なっていることを示している。

最後に、すべての形態の品詞を機能語と内容語に分け、どちらが多く編集されているかを比較してみた。MT 結果も TM 一致文も、テキストにもともと含まれる機能語と内容語の比率は 57:43 である。これに対し、MT の PE 中に変更が行われた箇所だけを取り出して機能語と内容語の比率を調べると 64:36 となり、TM 一致文の変更箇所では逆になって 36:64 であった。つまり、MT の PE では機能語が多く修正されており、TM の編集では内容語が多く修正されている。このことから、MT 結果の PE では、TM 一致文の編集と比較して、文法的な修正作業が多く、TM 一致文の編集では、MT の PE と比較して、語の修正が多いことがわかる。

## 6.8 PE プロセスの定性分析

次に、PE 作業のいくつかの際立った側面について、BB Flashback で記録した編集作業の画面動画を基に定性的分析を行った。この中から重要と思われる項目を 2 点挙げる。

### 6.8.1 PE 速度における外れ値の分析

ひとつ目は、PE が非常に高速または低速であったセンテンス(重回帰分析での外れ値)と、標準的な速度で PE されたセンテンスとの比較である。興味深いのは、PE が非常に高速であったセンテンスと非常に低速であったセンテンスの特徴がよく似ていることで、どちらも不完全文の占める割合が多く、文の長さが平均して標準より短く、またテキスト修正量は平均して少なかった。このうち特に PE が低速であった文について、BB Flashback の記録からその作業プロセスをたどってみた。すると、PE が低速なケースの特徴として、ポーズ(画面上で何も動作が行われていない時間)が多く発生していた。ポーズ時間の分析は、テキスト作成・編集作業において認知負荷の高さを示す重要な要素とされ (Schilperoord 1996) 翻訳研究にも取り入れられてきた (Carl et al. 2008, Jakobsen 1998, Koehn 2009, O'Brien 2006c.)。本研究では、PE が非常に低速なケースでは 20 秒以上の長いポーズが多く発生していた。これらの長いポーズは多くの場合、PE 開始直後や、PE 中にいったん PE 対象センテンスを離れて画面上で上下にスクロール移動した後

を突き止めることはできないため、推測になるが、この間ポストエディタは、そのセンテンスについてどのような PE 作業を行うべきかじっくり検討していたり、周囲のテキストを参照して得た情報をどのように PE に反映すべきか熟考していたりするのではないだろうか。だとすると、PE が非常に低速になる大きな原因は、テキストの修正量よりも、「考える」時間の長さなのではないだろうか。

### 6.8.2 センテンスの「再訪問」

ふたつ目は、一度 PE したセンテンスに戻って再度変更を加えるケースである。このうち興味深かったのは、初回 PE で UI や専門用語を変更した後、再度同じセンテンスに戻って元の用語に戻すというケースがいくつか見られたことである。さらに 3 度目の PE でもう一度初回 PE 時の変更内容に戻すというケースも観察された。このような、何度も用語を変更する PE 作業は、PE における用語の取り扱いの困難さを示している。「6.4.4 UI の有無」で述べた内容とも関連する点であり、PE 作業から用語の確認作業や修正作業を完全に排除することができれば、相当の効率向上が期待できることを示唆している。

## 7. 結論

以下に、定量分析で明らかになった点を列挙する。

- PE 速度は、PE におけるテキストの修正量と比較して、センテンス間でもポストエディタ間でも差異が大きい。
- PE 中に修正されるテキストの量と PE 速度との間には中程度の相関関係があり、これはどのポストエディタでもある程度共通している。
- ただしこの相関関係は、単文、複文、重文では比較的高く、不完全分では低い。
- センテンス構造別に PE 速度を比較すると、一般に単文が最も速く、次に複文や重文が続く、不完全分が最も遅い。
- 手順説明文は、他の機能のセンテンスと比較して PE 速度が速い。
- UI が存在するセンテンスは、変更の必要性に関係なく、PE 速度が遅くなる。またこれは、UI が原語のまま残されているより翻訳されている場合により顕著である。
- PE 作業のうち、不完全分では「補足」が、複文と重文では「書き直し」と「句読点の変更」が PE 速度を遅くさせている。
- MT の PE は、TM の 75% 一致文と比較して、テキストの修正量は大きいですが、速度は速い。
- MT の PE と TM 一致文の編集作業では、編集速度に影響を与える要因が大きく異なる。
- TM 一致文の編集作業では、MT の PE と比較して、語の変更作業が多い。
- MT の PE では、TM 一致文の編集作業と比較して、文法的な変更作業が多い。

このほか、定性分析では以下のような点が示唆された。

- PE を極端に低速化させるのは、修正量の多さよりも、熟考や調査のため作業が停止している「ポーズ」時間の長さである可能性が高い。
- UI や専門用語を何度も修正するケースが多々見られたことから、人手による編集プロセスから用語関連の作業を取り除くことができれば、編集作業の効率化が見込まれる。

以上、PE 作業時間をセンテンスごとに計測して、各種の独立変数に基づいて分析した本研究では、テキスト修正量だけに着目した調査ではわからない PE 作業の負荷の原因がいくつか明らかになった。なおこの論文は、2011 年 1 月に博士論文としてアイルランド Dublin City University に提出した論文の中から重要な点を抜粋してまとめたものである。研究の方法論、結果の詳細、結果に基づく考察の詳細については、論文 (<http://doras.dcu.ie/16062/> からダウンロード可能)を参照されたい。

.....  
【著者紹介】

立見 みどり (TATSUMI, Midori) : Dublin City University (アイルランド)、School of Applied Language and Intercultural Studies における本研究により博士号取得。現在は言語関連研究および翻訳業務に携わる。連絡先: [midori.tatsumi2@mail.dcu.ie](mailto:midori.tatsumi2@mail.dcu.ie)

.....

【註】

1. 京都大学情報学研究科-日本電信電話株式会社コミュニケーション科学基礎研究所 共同研究ユニットプロジェクトを通じて開発されたオープンソース 形態素解析エンジン。  
<http://mecab.sourceforge.net> からダウンロード可能(最終アクセス日:2012年2月29日)。
2. 外れ値の計算には、四分位数間領域 (IQR) による手法を使用した(第 1 四分点-1.5×IQR 以下、または第 3 四分点+1.5×IQR 以上)。

【参考文献】

- Aikawa, T., Schwartz, L., King, R., Mo, C. O., & Lozano, C. (2007). Impact of controlled language on translation quality and post-editing in a statistical machine translation environment. Paper presented at the *MT Summit XI*, Copenhagen, Denmark. pp. 1-7.
- Allen, J. (2003). Post-editing. In H. Somers (Ed.), *Computers and translation: A translator's guide* (pp. 297-317). Amsterdam: John Benjamins Publishing.
- Bernth, A., & Gdaniec, C. (2001). MTranslatibility. *Machine translation*, 16(3), 175-218.
- Carl, M., Jakobsen, A. L., & Jensen, K. T. H. (2008). Modelling human translator behaviour with user-activity data. Paper presented at the The 12th annual conference of the European association for machine translation, Hamburg, Germany. pp. 21-26.
- DePalma, D. A., & Kelly, N. (2009). *The business case for machine translation: How organizations*

- justify and adopt automated translation*. MA, USA: Common Sense Advisory, Inc.
- Fiederer, R., & O'Brien, S. (2009). Quality and machine translation: A realistic objective? *Journal of specialised translation*, (11), 52-74.
- Guerberof, A. A. (2008). Productivity and quality in the post-editing of outputs from translation memories and machine translation. *Localisation Focus: The international journal of localisation*, 7(1), 11-21.
- Jakobsen, A. L. (1998). Logging time delay in translation. *LSP texts and the process of translation*, 73-101.
- Japan Translation Federation. (2009). *White paper: The third research report on translation industry*. Japan: Japan Translation Federation.
- Koehn, P. (2009). A process study of computer-aided translation. *Machine translation*, 23(4), 241-263.
- Krings, H. P. (2001). [Koby G. S. (Ed.)], *Repairing texts: Empirical investigations of machine translation post-editing processes*. Kent, Ohio: The Kent State University Press.
- LISA. *LISA QA model version 3.1* LISA.
- Melamed, I. D., Green, R., & Turian, J. P. (2003). Precision and recall of machine translation. Paper presented at the HLT-NAACL 2003: Conference combining human language technology conference series and the North American chapter of the association for computational linguistics conference series, Edmonton, Canada. pp. 61-63.
- O'Brien, S. (2006a). *Machine-translatability and post-editing effort: An empirical study using translog and choice network analysis*. (PhD, Dublin City University).
- O'Brien, S. (2006b). Eye-tracking and translation memory matches. *Perspectives: Studies in translatology*, 14(3), 185-205.
- O'Brien, S. (2006c). Pauses as indicators of cognitive effort in post-editing machine translation output. *Across languages and cultures*, 7(1), 1-21.
- O'Brien, S., & Roturier, J. (2007). How portable are controlled language rules?: A comparison of two empirical MT studies. Paper presented at the MT summit XI, Copenhagen, Denmark. pp. 345-352.
- Papineni, K., Roukos, S., Ward, T., & Zhu, W. (2002). BLEU: A method for automatic evaluation of machine translation. Paper presented at the The 40th annual meeting of the association for computational linguistics (ACL), Philadelphia, pp. 311-318.
- Plitt, M., & Masselot, F. (2010). A productivity test of statistical machine translation post-editing in a typical localisation context. *The Prague bulletin of mathematical linguistics*, (93), 7-16.
- Roturier, J., & Lehmann, S. (2009). How to treat GUI options in IT technical texts for authoring and machine translation. *The journal of internationalisation and localisation*, 1, 40-59.
- Schäfer, F. (2003). MT post-editing: How to shed light on the “unknown task”. experiences at SAP. Paper presented at the Controlled language translation, EAMT-CLAW-03, Dublin City University, Dublin, Ireland. pp. 133-140.
- Schilperoord, J. (1996). *It's about time: Temporal aspects of cognitive processes in text production*.

Amsterdam: Rodopi.

- Snover, M., Dorr, B., Schwartz, R., Micciulla, L., & Makhoul, J. (2006). A study of translation edit rate with targeted human annotation. Paper presented at the 7th conference of the association for machine translation in the Americas, Cambridge, Massachusetts, USA. pp. 223-231.
- TAUS. (2010). *Postediting in practice*. De Rijp, The Netherlands: TAUS BV.
- Turian, J. P., Shen, L., & Melamed, I. D. (2003). Evaluation of machine translation and its evaluation. Paper presented at the MT summit IX, New Orleans, USA. pp. 386-393.
- Underwood, N. L., & Jongejan, B. (2001). Translatability checker: A tool to help decide whether to use MT. Paper presented at the MT summit VIII, Santiago de Compostela, Spain. pp. 363-368.
- Wagner, E. (1987). Post-editing: Practical considerations. *ITI conference. I: The business of translating and interpreting*, pp. 71-78.
- 山田 優 (2010) 機械翻訳は使えるのかーPost-editing の観点から検証する ー. 第 11 回 日本通訳翻訳学会年次大会, 東京.