

特別研究報告書

クラウドソーシングを用いた翻訳プロセスに  
おける報酬分配

指導教員 石田 亨 教授

京都大学工学部情報学科

西村 拓哉

平成 24 年 2 月 2 日

## クラウドソーシングを用いた翻訳プロセスにおける報酬分配

西村 拓哉

### 内容梗概

経済のグローバル化の影響を受けて日本企業がアジアなどの諸外国に進出する機会が増加しており、それに伴いドキュメントの翻訳に対する需要も増加している。しかし、プロの翻訳者の数は限られているため処理できるドキュメントの数には限界があり、また翻訳にかかるその金銭的成本は非常に大きい。そのため、プロの翻訳者がより効率よく翻訳作業を行える仕組みを作ることが求められている。

その問題に対して、クラウドソーシングと後編集 (Post-editing) を用いて翻訳を行う “Crowd-Assisted Translation (CAT)” というシステムが提案されている。クラウドソーシングとはネット上の無名の労働者に業務を委託する雇用形態であり、後編集とは機械翻訳の作成した翻訳文を人間が改訂して翻訳を行う翻訳プロセスである。CAT はそれらを組み合わせて翻訳を行う。まず、機械翻訳の作成した文を基に翻訳スキルの低い複数の翻訳作業者が協力して下訳を作成する。この時、翻訳作業者は他の作業や機械翻訳が作成した翻訳文を自由に参照しながら、より良い翻訳文を作成する。その後、翻訳スキルの高い翻訳専門家が翻訳作業者の作成した下訳を後編集して最終的な翻訳文を作成する。

翻訳作業者はクラウドソーシングの形で雇用されるため、報酬を支払う必要がある。その分配を考える必要がある。代表的なクラウドソーシングプラットフォームである Amazon Mechanical Turk (AMT) では、同一種類の作業であれば誰がしても同じ額の報酬が得られるようになっている。しかしその報酬分配を CAT でそのまま利用することはできない。なぜなら AMT で行われる作業は画像へのタグ付けなどの単純な作業であるのに対して、CAT において翻訳作業者が行う作業は翻訳という複雑な作業を他者と共同で行うものであり、その作業内容一人一人の作業内容を評価して報酬に反映する必要があるからである。

本研究の目的は作業内容一人一人の作業内容を評価する方法の設計である。そのためには、以下のような課題がある。

#### 1. 評価方法の設計方針の決定

翻訳作業者の作業内容の評価方法には様々な方針が考えられる。評価を行うためには広く受け入れられる評価方針を設定する必要がある。

## 2. 評価方法の設計

評価方針に基づいて実際に各翻訳作業者の作業内容を評価する方法を設計する必要がある。

本研究では上記の課題点を解決するために、各翻訳文の貢献量という尺度を定義し、それをを用いて各翻訳作業者の作業内容を評価した。

まず、翻訳作業者の作業をその作業によって減少した翻訳専門家の作業時間を基準に評価するという方針を決めた。翻訳作業者の作業の目的は、良い翻訳文を作成することで翻訳専門家の後編集作業にかかる作業時間を減少させることである。そこで、翻訳専門家の後編集作業にかかる時間のある翻訳文が作成されていない場合と作成された場合とで比較し、前者の値から見た後者の値の減少量をその翻訳文の貢献量と定義し、それによって評価することとした。

次に、実際に貢献量を算出する方法を設計した。貢献量を算出するためには各翻訳文の後編集にかかる作業時間を知る必要があるが、それを計測することは出来ない。そこで本研究では翻訳文の品質と後編集の作業速度に相関があることを利用することを考えた。まず後編集の作業速度を翻訳文品質の機械的評価指標による評価値から近似的に算出し、次にそれから作業時間を算出した。その後作業時間から各翻訳文の貢献量を計算する方法をアルゴリズムとして設計した。

最後に作業内容の評価方針が広く受け入れられるものであるかについて評価実験を行い、受け入れられるものであることを確かめた。

本研究の貢献は、以下の2点の課題点を解決したことである。

### 1. 評価方法の設計方針の決定

翻訳作業者の作業の目的が翻訳専門家の作業時間の削減である点を踏まえ、翻訳作業者によって作成された翻訳文による作業時間の減少量によって作業内容を評価することとした。また、翻訳文による作業時間の減少量をその翻訳文の貢献量であると定義した。また、実験によって、この方針が広く受け入れられるものであることを確かめた。

### 2. 評価方法の設計

GTM という翻訳文の品質評価尺度を用いることで翻訳専門家の作業時間を近似的に算出し、それをを用いて各翻訳文の貢献量を算出するアルゴリズムを設計した。

# Reward Distribution in Translation Process Using Crowdsourcing

Takuya NISHIMURA

## Abstract

In recent years, many Japanese companies expand to other countries in Asia because of economic globalization, so demand for translation of the document is increasing. However, the number of professional translators is limited. So the number of document they can translate in a given time is also limited and cost of translation of document is very high. Therefore, we need new translation process in which professional translators can work more efficiently.

For the problem, “ Crowd-Assisted Translation (CAT) ” is proposed. CAT is translation process using crowdsourcing and post-editing. Crowdsourcing is one of business model, it means sourcing tasks to crowd people in Web. Post-editing is translation process by translator revises machine translation. CAT uses them in combination. Firstly, workers – amateur translators – make draft translation by revising machine translation cooperatively. Worker makes translated sentences according to machine translation or sentences translated by other translators. Secondly expert – professional translator – accomplishes translation by revising draft translation.

Workers are hired as crowdsourcing, so CAT must pay reward to them. In Amazon Mechanical Turk (AMT) which is typical crowdsourcing platform, workers receive same reward when they do same task. However, in CAT we can't use same reward distribution model. Because tasks sourced to workers in AMT are very simple like labeling to image, although worker handles complicated task in CAT. So, we must evaluate each work of workers and reflect them to reward distribution.

Purpose of this research is designing model of evaluating each work. In order to realize it, we have to solve the following two problems:

1. Define evaluation policy

We can choose evaluation policy from many options. We must choose policy accepted by many people.

2. Design how to evaluate as the policy

We need evaluating method realizing the policy we choose.

In this research, we evaluate works by contribution amount, which means decrement of expert work time by the work.

Firstly, decide the evaluation policy. Purpose of workers is decreasing work time of expert, so we decide evaluating works by margin of work time of expert's post-editing between before and after of the work. And we call decreased work time by the work as contribution amount of the work.

Secondly, design how to evaluate as the policy. We need to know work time of expert's post-editing in order to calculate contribution amount, although we can't measure them. So, we calculate decrement of post-editing work time from one sentence by using General Text Matcher, one of scale evaluating translation sentence quality, correlate with work speed of post-editing. We calculate work speed and work time by scores of GTM. And we make algorithm which calculate contribution amount of each work.

Finally, we validate the policy we choose is accepted by many people.

Contributions of this research are as follows:

1. Define evaluation policy

We choose the evaluating policy which evaluates works by margin of work time of expert's post-editing between before and after of the work. And validate the policy we choose is accepted by many people.

2. Design how to evaluate as the policy

We calculate contribution amount by calculating decrement of post-editing work time by using GTM.

# クラウドソーシングを用いた翻訳プロセスにおける報酬分配

## 目次

第1章	はじめに	1
第2章	背景	3
2.1	クラウドソーシング	3
2.2	後編集 (Post-editing)	3
2.3	CAT(Crowd-Assisted Translation)	4
2.3.1	CAT の概要	4
2.3.2	機械翻訳による下々訳作成	5
2.3.3	翻訳作業者による前編集	6
2.3.4	翻訳専門家による後編集	6
2.3.5	報酬の支払い	7
2.3.6	実装	7
2.3.7	動作例	8
第3章	翻訳作業者の作業の評価	8
3.1	評価の動機とアプローチ	8
3.2	評価の前提	9
3.3	評価の方針と貢献量の定義	10
3.4	貢献量の計算	10
3.4.1	貢献量計算の仮定	10
3.4.2	後編集にかかる作業時間の算出	11
3.4.3	貢献量の計算アルゴリズム	13
第4章	実験及び評価	15
4.1	実験の目的	15
4.2	実験の方法	15
4.3	結果と分析	17
第5章	既存の報酬分配への貢献量の反映の提案	18
5.1	基礎報酬と賞与報酬による報酬分配	18
5.2	作業者の評判情報による報酬分配のアルゴリズム	19

第6章	おわりに	21
	謝辞	24
	参考文献	24
	付録：実験に用いたアンケート用紙	A-1

## 第1章 はじめに

近年，経済のグローバル化の影響を受けて日本企業がアジアなどの諸外国に進出する機会が増加している．それに伴い，ドキュメントの翻訳に対する需要も増加している．例えば，日本企業の製品の操作マニュアルを現地の言語に翻訳したり，現地の労働法などの法律を翻訳したりする必要がある．しかしプロの翻訳者の数は限られているため処理できるドキュメントの数には限界があり，また，プロの翻訳者に依頼して翻訳を行うための金銭的成本は非常に大きい．そのため，プロの翻訳者がより効率よく翻訳作業を行える仕組みを作ることと同じ時間で処理できるドキュメントの数を増やし，同時に翻訳にかかる金銭的成本を下げるのが求められている．

現在，ウェブ上の無名の労働者に業務をアウトソーシングするクラウドソーシングが流行し始めている．クラウドソーシングを利用した翻訳サービスも登場しており，日本国内では“Conyac<sup>1)</sup>”などが存在する．それらは安価な翻訳を実現してはいるものの，プロでない翻訳者が翻訳を行う事からその品質は低いものであり，また翻訳ミスなどがあったときに責任を取るのがクライアントの側であるため，企業が公的文書を翻訳するのに利用するのは難しい．

これとは別に現在研究されている翻訳プロセスの一つに Post-editing がある．Post-editing とは機械翻訳が作成した質の低い翻訳文を人間が改訂することで翻訳を行うものである．既存の研究の中には，機械翻訳をモノリンガルが一度改訂し，その後バイリンガルが改訂することでバイリンガルの作業効率を上げることが出来るものがある [1]．

クラウドソーシングと Post-editing という2つのアイデアを基に作られた翻訳システムとして“Crowd-Assisted Translation (CAT)”が提案されている．これは，機械翻訳が作成した下訳を翻訳スキルの低い複数の翻訳作業者が協力しながら前編集して下訳を作成し，それを翻訳スキルの高い翻訳専門家が改訂することで翻訳を行う．翻訳作業者はクラウドソーシングされた作業者が担い，翻訳専門家はプロの翻訳者が担う．CAT では翻訳作業者をを用いて質の高い下訳を準備出来るため，翻訳専門家の作業効率を上げることが出来るのではないかと考えられる．また，ク時間当たりの雇用コストが低い翻訳作業者をを用いて，時間当たりの金銭的成本が高い翻訳専門家の作業時間を短くすることが出来

---

<sup>1)</sup> Conyac(<http://www.conyac.cc/>)



るため、全体の金銭的成本を下げる事が出来るのではないかと考えられる。また、最終的に翻訳専門家が後編集を行う事から、翻訳ミスなどは保証することになるため、先に述べたクラウドソーシングを用いた翻訳サービスの問題点も解決される。

CATにおける翻訳作業者はクラウドソーシングの形で雇用されるため、雇用者は作業者に報酬を支払う必要があり、その分配方法を決定する必要がある。代表的なクラウドソーシングプラットフォームである Amazon Mechanical Turk では、ある種類の作業であれば誰がしても同じ額の報酬が得られるようになっている。しかし、その報酬分配方法を CAT でそのまま利用することは出来ない。CAT で報酬分配を行うためには以下の課題点がある。

#### 1. 評価方法の設計方針の決定

翻訳作業者の作業内容は複雑であり、その評価方法には様々な方針が考えられる。評価を行うためには広く受け入れられる評価方針を設定する必要がある。

#### 2. 評価方法の設計

実際にそれに基づいて各翻訳作業者の作業内容を評価する必要がある。

本研究では上記の課題点を解決するために、各翻訳文の貢献量という尺度を定義し、それを用いて各翻訳作業者の作業内容を評価した。

本稿では第2章で本研究の背景としてクラウドソーシングと Post-editing について述べた後、それらを組み合わせた翻訳システムとして現在提案されている“Crowd-Assisted Translation(CAT)”について述べ、その中で翻訳作業者の作業を評価する必要性について述べる。第3章では翻訳作業者の作業の評価の課題とアプローチ、前提について述べた後にその評価の方針を述べ、実際にどのようにその評価を算出するかを設計する。第4章では評価の方針がどの程度受け入れられるものであるかと、評価の算出がどの程度正確に出来ているかの検証実験を行いその結果を分析する。第5章では今後への提案として、既存のクラウドソーシングの報酬分配に翻訳作業者の作業の評価を反映する方法を2つ提案する。第6章では本研究における成果や今後の課題について述べる。

## 第2章 背景

### 2.1 クラウドソーシング

インターネットの普及に伴い，ウェブ上の不特定多数の人間に事業を委託するクラウドソーシングが盛んに行われるようになってきており，研究も盛んである [2]．クラウドソーシングの代表的プラットフォームとして Amazon Mechanical Turk などが存在する<sup>1)</sup>．

クラウドソーシングは翻訳の分野でも用いられ始めている．日本では Conyac などのサービスが存在する．Conyac の具体的サービス内容について述べる．クライアントが原稿を Conyac にコミットすると，クラウドソーシングによって集められた作業員にその原稿の翻訳がタスクとして提示される．それに対して作業員の中から先着で 2 人までがそれぞれ翻訳を行い，Conyac に翻訳稿をコミットする．クライアントはそれらの翻訳稿を全て受け取ることが出来る．このようなプロセスによって，安価で素早い翻訳サービスを実現している．しかしながら，Conyac ではクラウドソーシングによって集められた作業員によって全ての翻訳が行われるため，仮に誤訳があった場合にその責任の所在はクライアントということになる<sup>2)</sup>．企業の発行する文書など，誤訳によって大きな問題が発生しうるような原稿の翻訳には不向きであるといえる．

クラウドソーシングを用いた翻訳については研究も行われており，クラウドソーシングによって複数の翻訳文を集め，さらにそれらを改訂した後に最良の翻訳文を選択する方法などが提案されている [3]．しかしそれらにおいても翻訳文の作成にはプロの翻訳者が関与していないため，前述の問題は解決されていない．

### 2.2 後編集 (Post-editing)

後編集 ( Post-editing ) とは，機械翻訳の作成した文を人間が修正することで翻訳を行う翻訳プロセスである．後編集を行う人を後編集者 (Post-editor) と呼ぶ．後編集を用いると従来の翻訳よりも少ないコストで翻訳が可能であるといわれている．後編集の作業フローを表したものが図 1 である．

<sup>1)</sup> Amazon Mechanical Turk(<https://www.mturk.com/mturk/welcome>)

<sup>2)</sup> 利用規約に“翻訳結果等 全てのコンテンツに関する正確性、完全性、有用性、適法性等を自らの責任と判断により本サービスを利用するもの”と明記されている



図 1: 後編集の翻訳プロセス

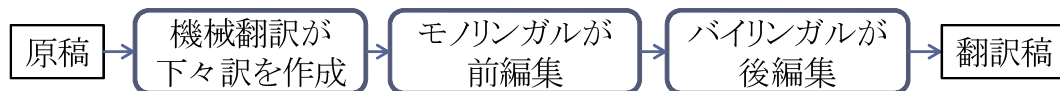


図 2: Lin の提案した翻訳プロセス

また、機械翻訳の作成した文を後編集者が翻訳する前に他の翻訳者が一度翻訳を行ったり修正を行うこともあり、その作業のことは前編集 (Pre-editing) と呼ばれる。

後編集についてはさまざまな研究がなされている。例えばある機械翻訳器が作成した翻訳文を別の機械翻訳が前編集することで下訳の品質を向上させるプロセスなどが提案されている [4]。

Lin による 2010 年の研究 [1] では、機械翻訳の作成した文をモノリンガルが前編集した後にバイリンガルが後編集するという翻訳プロセスが提案されている。その研究内ではモノリンガルが前編集に時間をかければかけるほどバイリンガルの作業時間を削減することが出来ると言われている。その翻訳プロセスを表したものが図 2 である。

## 2.3 CAT(Crowd-Assisted Translation)

### 2.3.1 CAT の概要

Crowd-Assisted Translation (以下、CAT と称する) はクラウドソーシングと後編集によって翻訳を行う翻訳サービスである。CAT の基本的なアイデアは、クラウドソーシングを用いて翻訳専門家の作業時間を減少させ、品質の保証された翻訳を維持しつつ、翻訳にかかる金銭的成本を下げることである。翻訳プロセスは、後編集の応用であり、機械翻訳の作成した下々訳に対して翻訳スキルの低い複数の翻訳作業者が協力しながら前編集を行うことで下訳を作成し、翻訳のプロである翻訳専門家が更に後編集を行うというプロセスになる。これらを表した概略図を図 3 に示す。

動作フローは以下のようなものになる。

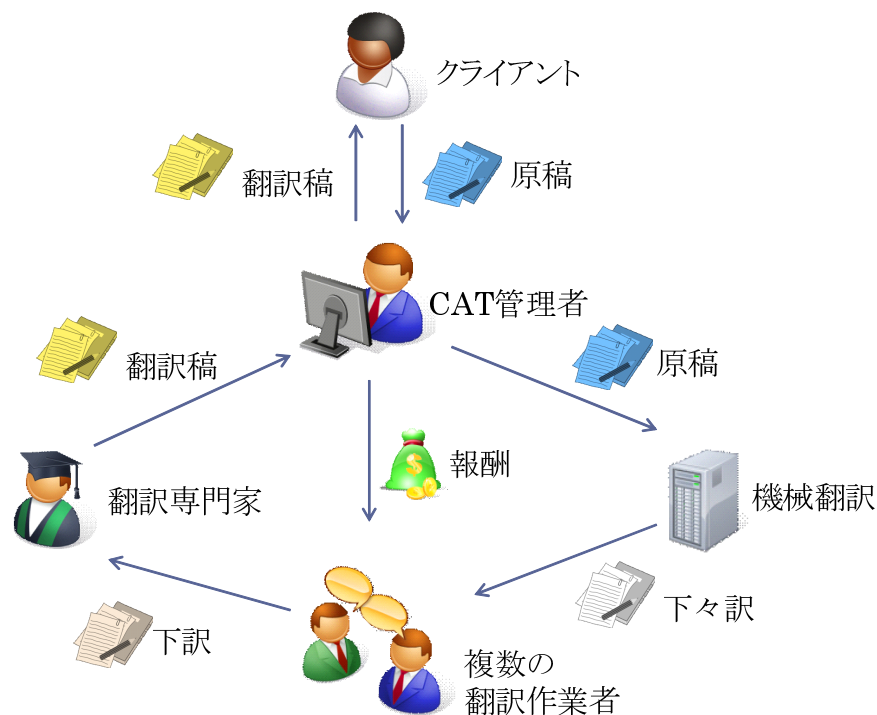


図 3: CAT の概略図

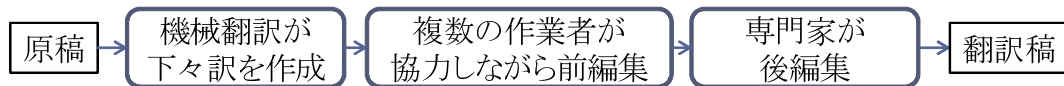


図 4: CAT の動作フロー図

1. クライアントによって原稿がコミットされると、最初に機械翻訳によって各原文に対する翻訳文が作成され、下々訳としてシステムにコミットされる。
2. 複数の翻訳作業員が協力して後編集を行い、下訳を作成する。
3. 時間経過などの終了条件を満たした段階で翻訳作業員による作業が打ち切られ、それまでに作成された全ての翻訳文が下訳となる。
4. 翻訳専門家が下訳に対して後編集を行う。
5. 翻訳専門家の作業が完了したら、翻訳稿の完成となり、システムからアウトプットされクライアントに届けられる。

CAT の動作フローを表したものが図 4 である。

### 2.3.2 機械翻訳による下々訳作成

クライアントがシステムに原稿をコミットすると、原稿が一文単位に分割された後、機械翻訳によって翻訳文が作成される。動作フローは以下ようになる。

1. 原稿がシステムにコミットされる。
2. 句点を基準に，システムが原稿を一文単位に分割する。
3. 機械翻訳がすべての文に対して翻訳文を作成する。
4. 翻訳文が下々訳としてシステムにコミットされる。

### 2.3.3 翻訳作業による前編集

機械翻訳に対して，複数の翻訳作業者が協力しながら前編集を行い，下訳を作成する．その作業フローは以下のようになる．

1. 原稿とシステムにコミットされた下々訳を基に翻訳作業用ページがシステムによって作成される。
2. 最初の翻訳作業者が下々訳と原稿を参照しながら，翻訳文を作成する。
3. 2番目以降の翻訳作業者が，その作業より以前に作業した翻訳作業および機械翻訳の作成した全ての翻訳文と原稿を参照しながら翻訳文を作成する。
4. 一定時間が経過したら作業を打ち切る．ここまでで作成された翻訳文をまとめたものが下訳となり，システムにコミットされる。

この作業は前編集の一種ではあるが，特徴として前編集の基として参照する翻訳文が一つだけではなく，機械翻訳や他の翻訳作業によって作成された翻訳文全てである点が挙げられる．翻訳作業者は複数の翻訳文の中から最良であると思う文を選び，それを基に改訂を行うことが出来る．よって，翻訳作業による前編集は集合知的側面を強く持ち，単独での作業と比較してより品質の高い下訳が作成されると期待できる．

実際にこの作業が行われているの画面例を図5に示す．この画面では日英翻訳が行われている．原文となっているところには翻訳前の日本語の文が表示されている．過去に投稿された翻訳文のところには機械翻訳および他の翻訳作業によって作成された翻訳文が全て表示されている．翻訳作業者はそれらを参考にしながら，編集欄に翻訳文を入力し，保存することでシステムに投稿する．

この作業の際には，機械翻訳や他の翻訳作業によって作成された翻訳文を翻訳文入力欄にコピーする機能を利用することが出来る．この機能は翻訳作業者間の協力を円滑にする．

### 2.3.4 翻訳専門家による後編集

翻訳専門家は下訳を参照しながら最終的な翻訳文を作成していく．通常の後編集作業と異なり，下訳は各原文につき一文ではなく，機械翻訳と翻訳作業者



図 5: CAT の画面例

によって作成された全ての翻訳文である．翻訳専門家はそれの中から最良であると思われる翻訳文を参考にしながら翻訳文を作成していくことになる．実際の作業は翻訳作業者と同様に、図 5 のような画面で行われる．

### 2.3.5 報酬の支払い

翻訳が完了した後に、翻訳作業者に対してはその作業内容に応じて報酬が支払われる．AMT を始めとした既存のクラウドソーシングでは同一の作業に対しては同一の報酬が支払われる．CAT においてそれを踏襲することは望ましくない．なぜならば、AMT で行われるタスクは画像へのタグ付けなど非常に単純なものに限られているのに対して、CAT で翻訳作業者が行う、他社と協力しての翻訳作業というのは非常に複雑であり、個人の能力やかけた時間、あるいはその前後関係などによって作業内容が大きく変化する．そのため、翻訳作業者の作業を評価し、それを反映した報酬分配を行うことが望ましい．

現在の CAT で報酬分配機能は未実装である．

### 2.3.6 実装

CAT は言語グリッドプロジェクト内で開発された Toolbox 上で動作するモジュールとして実装されている．言語グリッドは、辞書や機械翻訳などの言語資源を言語サービスとして登録し、共有可能にするインターネット上の多言語サービス基盤であり [5]，Toolbox とは言語グリッドを用いた多言語コミュニケー

表 1: CAT による日英翻訳の具体例

順序	作成者	文
		キャラクターグッズを欲しがると年齢層もこの年齢層に重なっていた。
4	専門家	It was also the children of this age group who are into the character goods.
3	作業員 2	It was also the children of this age group who importune character-licensed merchandises.
2	作業員 1	The age group which would want to get character-licensed merchandises was also the children under twelve.
1	機械翻訳	The age bracket which wants a character goods also fell on this age bracket.

ション支援ツールで、言語グリッドの機能を用いたサービスを Web ブラウザから誰でも簡単に使えるようにしたものである [6] .

### 2.3.7 動作例

実際に CAT によって翻訳された文の例を表 1 に示す .

表 1 を例に CAT での翻訳について説明する . まず , CAT に原稿が投稿されると , 句点を基準に 1 文毎に分割され , 機械翻訳器によって翻訳文が作成される . 原稿が分割され 1 文になったものを原文と呼ぶ . 表 1 では 1 番上が原文であり , 1 番下の文が機械翻訳器による翻訳文である . 次に , 翻訳作業員の作業が始まる . 1 人目の翻訳作業員は機械翻訳器による翻訳文を参考に翻訳文を作成し , それ以降の翻訳作業員はそれまでに作成された全ての翻訳文を参考に翻訳文を作成する . 表 1 では下から 2 番目の文が 1 人目の翻訳作業員による翻訳文であり , 下から 3 番目の文が 2 人目の翻訳作業員による翻訳文である . 1 人目の翻訳作業員が機械翻訳を基に作成し , 2 人目は 1 人目の作成した翻訳文を基に作成していることが分かる . 最後に , 翻訳専門家がそれまでに作成された全ての翻訳文を下訳として , 後編集を行う . 表 1 では下から 4 番目の文が翻訳専門家の作成した翻訳文である . 2 人目の翻訳作業員の作成した翻訳文を基に作成していることが分かる .

## 第 3 章 翻訳作業員の作業の評価

### 3.1 評価の動機とアプローチ

雇用者はクラウドソーシングによって確保された労働力に対して報酬を支払う必要がある . CAT の翻訳作業員に対してもそれは同様である . クラウドソーシングの代表的プラットフォームである Amazon Mechanical Turk の多くのタ

スクにおいては、単純な出来高制の報酬体系が採用されている。つまり、同一のタスクに対しては、それを誰が行っても、あるいはそれを同一人物が複数回行った場合も良い場合は何回行っても、一回当たりの報酬は一定である。これは Amazon Mechanical Turk で扱われる作業が極めて単純なものに限られ、個人によって行った作業内容が大きく変化することなく、また作業者に協力が発生しないからだと考えられる。

CAT における翻訳作業者の作業内容はこれとは大きく異なる。CAT において翻訳作業者が行う他者と協力しながらの翻訳という作業は、同じ量を行ったとしても翻訳作業者のスキルによって大きくその成果が左右されるため、単純な出来高制度を採用することは不平等につながる。そのため、翻訳作業者一人一人の作業を評価し、それを報酬へと反映する必要がある。

本研究では、作業者の作業内容を評価する方法の設計を目的とし、以下のような手順で設計する。

1. 評価方法の設計方針の決定

翻訳作業者の作業内容は複雑であり、その評価方法には様々な方針が考えられる。評価を行うためには広く受け入れられる評価方針を設定する必要がある。

2. 評価方法の設計

実際にそれに基づいて各翻訳作業者の作業内容を評価する必要がある。

### 3.2 評価の前提

翻訳作業者の作業の評価について、以下のような前提のもとで行った。

- 作業の評価は原文一文一文単位で行う

通常の翻訳では、翻訳元言語での一文が必ずしも翻訳先言語においても一文になるとは限らない。しかし CAT での翻訳は、最初に原稿が文単位に分割され、その後翻訳は分単位で行われる。本研究は CAT を念頭においたものとなっているため、作業の評価も一文単位で行うものとする。

- 作業の評価は機械翻訳や翻訳専門家に対しても行われる

協力して翻訳作業を行うという観点でいえば、機械翻訳や翻訳専門家も協力している一員であるといえる。そのため、本研究では貢献量は機械翻訳や翻訳専門家も含めて計算するものとする。



### 3.3 評価の方針と貢献量の定義

翻訳作業者の作業の評価を行うためには、何を基準として評価するかを決定する必要がある。基準には様々なものが考えられるが、本稿では翻訳作業者の作業の目的に着目した。翻訳作業者が協力して行う前編集の目的は、翻訳専門家の後編集にかかる作業時間の短縮である。そこで、翻訳作業者の作業によって減少した翻訳専門家の作業時間を貢献量と呼び、それに基づいて評価を行うこととした。本稿では貢献量を以下のように定義する。

- 翻訳文 A によって減った翻訳専門家の作業時間とは、翻訳文 A が作成される以前に翻訳専門家が後編集する場合にかかる作業時間から、翻訳文 A が作成された後にかかる作業時間に等しい

翻訳専門家が後編集を行う際に、品質の高い翻訳文から後編集を行う方が作業時間が短いと考えられる。翻訳文 A より前に作成された翻訳文と比較して翻訳文 A の方が品質が高い場合、翻訳文 A が作成される前より作成された後の方が翻訳専門家が後編集にかかる作業時間が短くて済むと考えられる。

- 翻訳文 A の貢献量とは、翻訳文 A によって減った翻訳専門家の作業時間に比例する

翻訳作業者が協力して行う Pre-editing 作業の目的は翻訳専門家の作業時間の減少にある。すなわち、翻訳作業者の作成した翻訳文の貢献とは翻訳専門家の作業時間の減少である。そこで、何も無い状態からの翻訳専門家の作業時間全体を 1 とした時に、翻訳文が減少させた作業時間を翻訳文の貢献量と定義する。

### 3.4 貢献量の計算

本節では各翻訳文の作業への貢献量を実際に計算する方法を設計する。

#### 3.4.1 貢献量計算の仮定

以下のような仮定のもとで翻訳作業者の貢献量を計算するアルゴリズムを設計する。

- 翻訳作業者および翻訳専門家はそれまでの翻訳文の中で最も質の良い翻訳文を参照する

翻訳作業者および翻訳専門家は自分が改訂作業をする以前に作成された全

での翻訳文を参照することが出来る．実際に改訂の際に参照するのはその中で最も質の良い翻訳文であると考えられる．

- 翻訳文 A によって減った翻訳専門家の作業時間とは，翻訳文 A が作成される以前に翻訳専門家が後編集する場合にかかる作業時間から，翻訳文 A が作成された後にかかる作業時間に等しい

翻訳専門家が後編集を行う際に，品質の高い翻訳文から後編集を行う方が作業時間が短いと考えられる．翻訳文 A より前に作成された翻訳文と比較して翻訳文 A の方が品質が高い場合，翻訳文 A が作成される前より作成された後の方が翻訳専門家が後編集にかかる作業時間が短くて済むと考えられる．

### 3.4.2 後編集にかかる作業時間の算出

貢献量を 3.4.1 の仮定に基づいて計算しようとした場合，翻訳文を翻訳専門家が後編集するのに必要な作業時間を実際に計測する以外の方法で算出する必要がある．なぜなら，翻訳文 A の貢献量を算出するためには翻訳文 A がない時とある時の翻訳専門家の作業時間を比較する必要があるが，実際の翻訳プロセスの中で翻訳専門家がその両方で翻訳を行うことはないからである．翻訳文 A が翻訳作業員及び機械翻訳器の作成した全ての翻訳文の中で最も質の良い翻訳文であるとすれば，翻訳専門家は翻訳文 A を参照しながら後編集を行い，それ以外の場合では翻訳文 A 以外の翻訳文を参照しながら後編集を行う形になる．

本研究では翻訳専門家が後編集するのに必要な作業時間を算出するために翻訳文の品質評価指標を利用することを考えた．

過去の関連研究から，後翻訳の下訳となる翻訳文の品質とその労力には相関があることが知られている．[7]．しかし，その関連研究では翻訳文の品質は人手で評価を行っているが，CAT で実際に運用することを想定するとその方法を用いることは出来ない．そこで，翻訳文の品質を機械的に評価できる，翻訳文の品質評価指標を用いる．

翻訳文の品質評価指標について説明する．機械翻訳が盛んになるにつれ，各機械翻訳器の性能を論じる必要が出てきたために，機械翻訳器によって作成された翻訳文の品質の評価が行われるようになった．手動での評価が中心ではあったが，同時に機械的に評価する指標も必要になったため，様々な指標が設計された．Levenshtein が 1965 年に提案したレーベンシュタイン距離 (あるいは編集距離)[8] や Snover らが 2006 年に提案した Translation Edit Rate [9]，Papineni

らが2002年に提案したBLEU [10]などが代表的な指標として知られている。これらはいずれも複数もしくは単一の正解文と、対象となる翻訳文を比較し、その近似度を計算することで、対象となる翻訳文の品質を評価する。

本研究では、翻訳文への評価を行う指標として、General Text Matcher (以下GTM と称する) を用いた。GTM はTurianらによって2006年に提唱された翻訳文の品質指標であり [11]、上記のものと同じく正解文との近似度によって対象となる翻訳文の品質を評価する。

GTMの具体的内容について述べる。GTMは自然言語処理で良く用いられる適合率 (precision) と再現率 (recall) という尺度を用いて翻訳文の近似度を計算する方法である。候補となる情報の集合である  $Y$  と、参照される情報の集合である  $X$  があったとき、適合率と再現率はそれぞれ

$$precision(Y|X) = \frac{|X \cap Y|}{|Y|};$$
$$recall(Y|X) = \frac{|X \cap Y|}{|X|};$$

となる。これを翻訳文の近似度計算として用いるためには、 $|X|$ 、 $|Y|$ 、 $|X \cap Y|$  の3つの定義をする必要がある。GTMではその定義として、正解文と対象の翻訳文のそれぞれの単語数を  $|X|$  と  $|Y|$  として、2文の間で単語を単位とした最長マッチングを行い、それによってマッチングした単語の数を  $|X \cap Y|$  と定義している。

さらにこれらから F 値 (F-measure) と呼ばれる尺度を算出し、これを GTM による評価とする。F 値とは、一般的に適合率を上げると再現率が下がり、再現率を上げると適合率が下がる傾向にあるため、それらの調和平均を取ったものである。

$$F - measure = \frac{2 * precision * recall}{precision + recall};$$

となり、この F 値が GTM の評価値となる。

GTMは正解文との近似度を計算するため、GTMによって算出されたスコアが高い翻訳文ほど、後編集にかかる作業時間が少ないことになる。

Tatsumi が2010年に発表した“Post-Editing Machine Translated Text in A Commercial Setting: Observation and Statistical Analysis”では、後編集に関する実験を行い、様々な統計的データをまとめている。その中では後編集にかかる作業時間と各種の翻訳文評価指標による評価値の相関関係に関する実験が

行われており、結果も述べられている。実験は3人の後編集者に435文の翻訳文をそれぞれ翻訳してもらい、後編集の作業速度(語/分)とGTM, Translation Edit Rate, BLEUの3つの指標による評価値の相関値を算出している。その結果によれば、3つの指標の中でGTMが最も相関係数が高く、3人の後編集者についてそれぞれ0.60, 0.56, 0.53という結果であった。

次に、GTMを利用して作業時間を算出する。

まず、作業速度を算出することを考える。GTMによる評価値と作業速度には相関があるが、比例関係にはない。そこで、作業速度を従属変数、GTMによる評価値を独立変数とする一次式に近似する。 $workspeed$ を作業速度、 $GTMscore$ をGTMによる評価値とおくと、式は

$$workspeed = \epsilon + \beta * GTMscore;$$

となる。 $\epsilon$ と $\beta$ の値を最小二乗法を用いて求める。最小二乗法とは、モデル関数に対して計測データの誤差が正規分布になっているという前提のもとに、残差の二乗和が最少となるように関数の係数を決定する方法である。実際にこれを行うためには、実験を行いGTMのスコアと後編集の作業速度の対データを記録する必要があるが、本研究はそこまでは至ることが出来なかった。

次に、作業速度から作業時間を算出する。作業時間を $worktime$ 、翻訳文の長さを $length$ とおくと、

$$worktime = \frac{length}{workspeed};$$

となる。

### 3.4.3 貢献量の計算アルゴリズム

3.4.1でおいた仮定を基に、貢献量の計算方法を設計していく。作業員及び専門家は作成された中で最良の翻訳文を参照する点を考慮すると、ある翻訳文Aの貢献量はそれまでの翻訳文の中で最良の翻訳文を改訂するのに必要な作業時間から、翻訳文Aを改訂するのに必要な作業時間を引いたものに比例することになる。

仮に翻訳文Aを改訂するのに必要な作業時間がそれまでの翻訳文の中で最良の翻訳文を改訂するのに必要な作業時間よりも長かった場合、翻訳文Aの作成によって翻訳専門家の作業時間は変化しないことになり、翻訳文Aの貢献量は0ということになる。

---

**Algorithm 1** 各翻訳文の作業への貢献量の算出 Contribution(Score)

---

```
score /* 翻訳文のスコアの配列 */
score[i] /* i 番目に作成された翻訳文のスコア */
C /* 各翻訳文の貢献量の配列 */
C[i] /* i 番目に作成された翻訳文の貢献量 */
n /* 作成された翻訳文の総数 */
currentMaxScore ← 0
for i = 1 to i = n do
  if currentMaxScore < score[i] then
    C[i] ← score[i] - currentMaxScore
    currentMaxScore ← score[i]
  else
    C[i] ← 0
  end if
end for
return C
```

---

表 2: CAT による日英翻訳から貢献量を算出した具体例

順序	作成者	文	後編集に かかる時間	スコア	貢献量
		キャラクターグッズを欲しが る年齢層もこの年齢層に重 なっていた。	-	-	-
4	専門家	It was also the children of this age group who are into the character goods.	0 分	1	0.4
3	作業員 2	It was also the children of this age group who importune character-licensed merchandises.	12 分	0.6	0.1
2	作業員 1	The age group which would want to get character-licensed merchandises was also the children under twelve.	15 分	0.5	0.2
1	機械翻訳	The age bracket which wants a character goods also fell on this age bracket.	21 分	0.3	0.3

これらの計算を行うために、翻訳文のスコア ( $score$ ) を以下のように定義する。

$$score = 1 - \frac{\epsilon}{\epsilon + \beta * GTMscore};$$

これは、作業時間を文の長さで正規化し、さらにそれを最大値が 1 となるように正規化したものを 1 から引いたものである。スコアの差分は、翻訳文を後編集するのにかかる作業時間の差と比例する。

以上のようにして設計した貢献量の計算アルゴリズムが Algorithm1 である。

このアルゴリズムを利用して表 1 の翻訳例の各翻訳文の貢献量を算出したものが表 2 である。ただし、GTM の評価値から作業速度を算出する近似式が本研究では求められていないため、それぞれの翻訳文からの後編集にかかる翻訳専門家の作業時間を仮にしている。また、後編集ではなく原文からの翻訳には

30 分かかるとする。

機械翻訳器による翻訳文のスコアが 0.4 であるということは、この翻訳文がある状態からの翻訳は何もない状態で翻訳を行った場合と比較して 7 割の作業時間で済むことを示している。つまり、機械翻訳器による貢献量は 0.3 である。二人の翻訳作業員それぞれのスコアは 0.5, 0.6 であり、貢献量はそれぞれ 0.2 である。つまり、翻訳作業員それぞれの働きによって翻訳専門家の作業時間は 2 割と 1 割減少していることになり、機械翻訳器による翻訳文から後編集する場合に比べて、翻訳作業員の働きにより翻訳専門家の作業時間が 3 割削減されていることになる。

## 第 4 章 実験及び評価

### 4.1 実験の目的

本研究では翻訳作業員の作業の目的は翻訳専門家の後編集にかかる作業時間の短縮であるとし、ある翻訳文の作成前後で減少した翻訳専門家の作業時間をその翻訳文の貢献量と定義し、それによって翻訳作業員の作業の評価を行った。本章ではその評価の方針が多くの人にとって受け入れられるものであるかどうかを実験によって検証する。

### 4.2 実験の方法

3 つの評価方針を被験者に提示し、それぞれの評価が正当な評価であるか否かをアンケートで聞くことで検証を行った。

3 つの評価方針は以下である。

1. 貢献量によって評価する。実験中で被験者に伝える際には、作業によって減少した翻訳専門家の作業時間に比例して評価するとした。提案手法である。
2. 全ての作業を均一に評価する。AMT では原則すべての作業に対して同一の報酬であることを考えると、この評価方針を採用しているといえる。
3. 前後関係を無視し、作成された翻訳文がどれだけ翻訳専門家の作業時間を減少させているかに比例して評価する方針。前後関係を無視し、純粋に翻訳文の品質によって評価を行う。当然ながら、後から作業した翻訳作業員ほど高い評価を得られることになる。

具体的な実験方法は以下である。

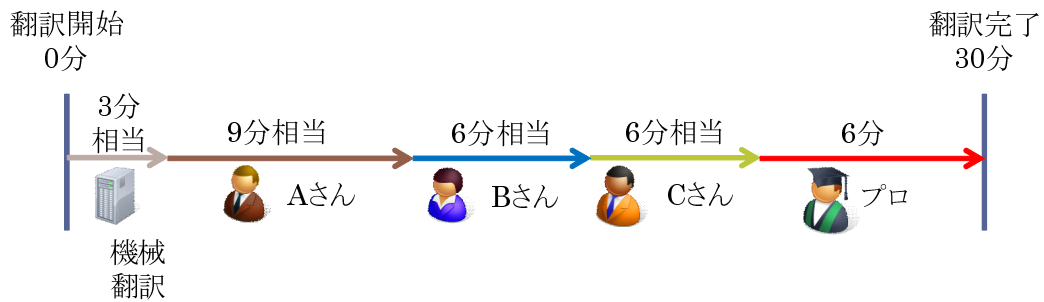


図 6: 実験で被験者に示した機械翻訳と各翻訳作業者の作業による翻訳専門家の作業時間の変化を示す図の一例

表 3: 実験実施データ

目的	翻訳作業者の作業の評価方針が多くの人に受け入れられるものであるか検証
被験者	15 名
実験方法	被験者へのアンケート
評価方法	対象の評価方法が正当な評価であるか 5 段階で回答
評価対象	1:貢献量（翻訳専門家の作業時間の減少量に比例）による評価 2:全ての作業を均一に評価 3:前後関係を無視し，作成された翻訳文の品質によって評価
提示データ	機械翻訳と各翻訳作業者の作業が翻訳専門家の作業時間をどれだけ減少させたかの図

1. CAT の翻訳プロセスの概要について被験者に説明する．特に，機械翻訳器と翻訳作業者が翻訳作業を行うことによって，翻訳専門家の作業時間を減少させることが CAT の目的であることを説明する．
2. この実験の目的は機械翻訳と各翻訳作業者の作業内容を 3 つの評価方針で評価し，どれが最も正当な評価方針であるかを調べることである旨を説明する．
3. 3 つの評価方針について説明する．
4. 図 6 のような，機械翻訳と各翻訳作業者の作業が翻訳専門家の作業時間でいうとどれだけの量に相当する翻訳作業を行ったかを示す図を 3 つ被験者に見せる．併せてそれぞれの場合についての機械翻訳と各翻訳作業者への評価の比を 3 つの評価方針によって算出したものを，被験者に示す．
5. 被験者にそれぞれの評価方針について，適切な評価であると感じるかどうかを 5 段階で回答してもらう．

これらをまとめたものが表 3 である．また，実際に利用したアンケート用紙は本稿に付録として添付されている．

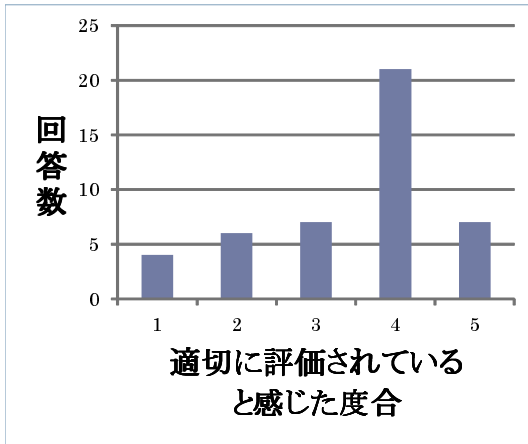


図 7: 貢献量で評価

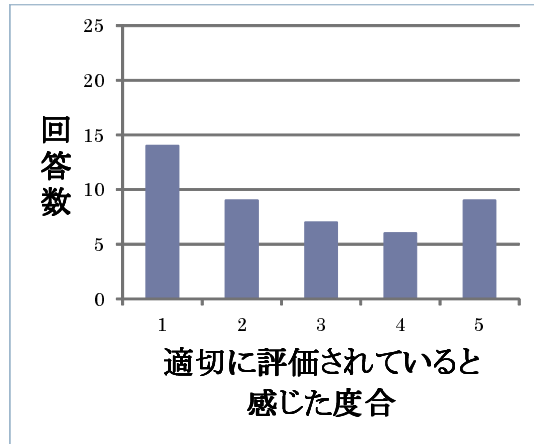


図 8: 全ての作業を均一に評価

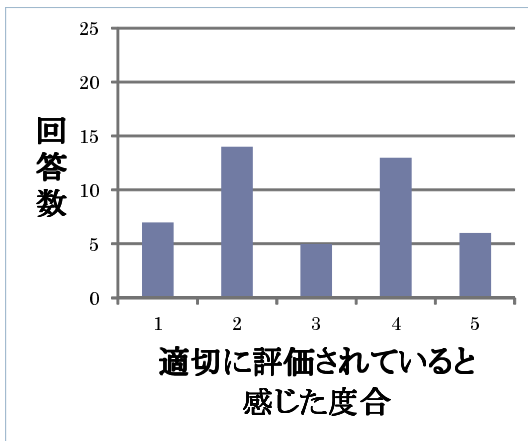


図 9: 前後関係を無視して評価

### 4.3 結果と分析

実験結果として、適切であると感じた各度合への回答数を棒グラフとして示す。それぞれ、図 7 が貢献量に比例して評価する方針、図 8 が全ての作業を均一に評価する方針、図 9 が前後関係を無視して評価する方針への回答の結果である。

平均値は表 4 のようになった。この実験の結果から、提案手法である貢献量で評価する方針が今回の 3 つの方針の中で最も多くの人から受け入れられる評価方針であることが分かった。



表 4: 各評価方針に対して適切だと感じた度合の平均値

	適切に評価されていると感じた度合の平均値
貢献量で評価	3.47
全ての作業を均一に評価	2.71
前後関係を無視して評価	2.93

## 第5章 既存の報酬分配への貢献量の反映の提案

翻訳作業者の作業の評価を実現出来れば、次はそれを報酬分配へと反映する必要がある。クラウドソーシングにおける報酬分配について、どのようなものが望ましいかについては先行研究もあるものの [12]、定説はないのが現状である。

そこで、本稿ではクラウドソーシングにおける既存の報酬分配を2つ取り上げ、それぞれに対して3.4で算出した翻訳文の貢献量を反映する方法を提案する。本研究ではその評価にまでは至れなかったが、貢献量が作業者の作業内容を反映した報酬分配に活かせることを示すことを本章の目的とする。

また、本章では以下の仮定をおく。

- 原文一文への報酬総額は文字数によってあらかじめ決定される  
報酬分配を行うためには報酬の総額がいくらであるかが決定している必要がある。本章では、原文一文一文に対して、その文字数によって報酬総額が決定されているものと仮定する。一文字当たりの報酬額は翻訳元言語と翻訳先言語の組み合わせによって決定されるものとする。現在行われている翻訳サービスの多くも原則としてこの方法で報酬総額が決定している。

### 5.1 基礎報酬と賞与報酬による報酬分配

全ての作業に対して均一の基礎報酬を分配した後に、クライアントの設定した条件を満たした作業にのみ別途賞与報酬を分配するものである。Amazon Mechanical Turk などでも採用されている報酬分配である。

これをCATで利用する際には、以下のような問題点がある。

- 基礎報酬は必ず支払われるため、作業数が多くなるとあらかじめ定められていた報酬総額より支払うべき報酬総額の方が上回ってしまう  
この報酬分配は作業参加者全員に一定の報酬を支払う必要があるため、たくさんの作業者が参加し、たくさんの作業が行われるほど必要な報酬総額が増加してしまう。しかし本研究の前提条件として、各原文あたりの報酬

総額は文字数によって固定されているため、問題が生じる。

- 賞与の基準を設定する必要がある

各作業に対して賞与を与えるかどうかの基準を設定し、それを満たしているか自動的に判断できるようなシステムを設計し実装する必要がある。

この点を考慮した上で、報酬分配を設計する。

Amazon Mechanical Turk などでもこの報酬分配が採用されているが、問題にならないのは作業数の上限が固定されているからである。そこで、CAT の報酬分配でもそれを採用することを考える。つまり、一つの原文に対して報酬を支払う翻訳文の数を制限するのである。一つの原文に対して報酬を支払う翻訳文の数を  $n$  とした場合、(基礎報酬と賞与報酬の合計)\* $n$  がその原文の総額報酬を上回っていなければ、どのような作業が行われたとしても問題なく報酬分配を行うことが可能になる。

この上で賞与を支払うかどうかの基準について貢献量を用いて判断するようになれば、作業者の作業内容を反映した報酬分配を実現できる。

これをアルゴリズムとして記述したものが Algorithm 2 である。

実際にこの報酬分配を CAT で利用するためには1つの原文に対して報酬を支払う翻訳文の数や、基礎報酬と賞与報酬の金額、賞与の条件などを決定する必要がある。これらはクライアントの希望や、今後の CAT の研究結果から決定されることになるであろう。例えば、統計的に1つの原文に対して  $n$  個の翻訳文が作成されれば翻訳専門家の作業時間の減少量が収束する、などといったことが判明すれば、1つの原文に対して報酬を支払う翻訳文の数は  $n$  個に決定されるであろう。

この報酬分配は作業者についての事前情報を必要としないため、翻訳作業者の担い手が雑多になることが想定される場合などに有効であると考えられる。

## 5.2 作業者の評判情報による報酬分配のアルゴリズム

作業者のこれまでの作業履歴からランク付けを行い、そのランクに従って一作業当たりの報酬総額を変化させる報酬分配である。過去の研究として、クラウドソーシングの報酬分配を評判情報を用いて行うことは雇用者と作業者にとって効率的な報酬分配の仕組みであるとした研究がある。[13]

これに貢献量を反映する際には、以下のような問題点がある。

- 基礎報酬は必ず支払われるため、作業数が多くなるとあらかじめ定められ

---

**Algorithm 2** 基礎報酬と賞与報酬による報酬分配のアルゴリズム Dist-by-base-and-bonus( $C$ )

---

```
 $C$  /* 各翻訳文の貢献量の配列 */
 $C[i]$  /*  $i$  番目に作成された翻訳文の貢献量 */
 $R$  /* 各翻訳文の報酬の配列 */
 $R[i]$  /*  $i$  番目に作成された翻訳文の報酬 */
 $n$  /* 作成された翻訳文の総数 */
 $m$  /* 1つの原文に対して報酬を支払う翻訳文の数の上限 */
 $a$  /* 基礎報酬額 */
 $b$  /* 賞与報酬額 */
 $\epsilon$  /* 賞与の基準となる貢献量 */
 $sum$  /* 決定されている報酬総額 */
if  $(a + b) * m > sum$  then
    return false
end if
for  $i = 1$  to  $i = n$  do
    if  $i \leq m$  then
        if  $C[i] \geq \epsilon$  then
             $R[i] \leftarrow a + b$ 
        else
             $R[i] \leftarrow a$ 
        end if
    else
         $R[i] \leftarrow 0$ 
    end if
end for
return  $R[ ]$ 
```

---

ていた報酬総額より支払うべき報酬総額の方が上回ってしまう

一作業当たりの報酬額がランクによって固定であるため、作業数が多くなると報酬総額以上の報酬を支払う必要が生じるという問題点がある。

- ランク付けの基準の設定が必要である

作業者のランクの変動を作業の内容に基づいて判断する基準が必要であり、その基準を判断する仕組み設計し実装する必要がある。

これらの点を考慮した上で、報酬分配を設計する。

作業数が多くなると支払うべき報酬が報酬総額を超えてしまう問題に対しては、基礎報酬と賞与報酬による報酬分配と同様に、一つの原文に対して作成出来る翻訳文の数に上限を設けることで対処することを考える。一つの原文に対して翻訳文を作成できる回数を  $n$  回とした場合、(一作業当たりの報酬額として最も高額な報酬額) $*n$  がその原文の総額報酬を上回っていなければ、どのような作業が行われたとしても問題なく報酬分配を行うことが可能になる。

作業者のランク付けの基準としては、作成した翻訳文の貢献量を基準とすることで、作業内容に基づいたランク付けが可能となる。

報酬分配についてアルゴリズムとして記述したものが Algorithm 3 で、作業者のランク付けについて記述したものが Algorithm 4 である。

実際にこの報酬分配を実装するためには、作業者の信頼情報の管理機能を実装した上で、ランクの昇格条件や評判情報の上昇の基準となる貢献量、およびそれぞれのランクでの報酬を決定する必要がある。これらは CAT についての研究が進めばその中で明らかになっていくであろう。

この報酬分配は翻訳作業者の情報を管理できるような場合に有効であると考えられる。

## 第6章 おわりに

本研究ではクラウドソーシングによって確保された翻訳作業者が、他の翻訳作業者と協力して前編集を行った際の、各翻訳作業者の作業の評価方法の設計を行った。

作業の評価方針を決定する必要があるという課題に対して、Pre-editing 作業の目的が翻訳専門家の後編集にかかる時間の削減である点に着目し、各翻訳文が翻訳専門家の作業時間をどれだけ削減したかを貢献量と定義し、それを算出

---

**Algorithm 3** 作業者の信頼情報による報酬分配のアルゴリズム Dist-by-Worker-Reputation(Writer)

---

```
R /* 各翻訳文の報酬の配列 */
R[i] /* i 番目に作成された翻訳文の報酬 */
Writer /* 作業者の配列 */
Writer[i] /* i 番目に作成された翻訳文を作成した作業者 */
lank(w) /* 作業者 w の信頼情報のランク */
rewardForLank(l) /* ランク l の作業者の作業に対する報酬 */
n /* 作成された翻訳文の総数 */
m /* 1つの原文に対して報酬を支払う翻訳文の数の上限 */
l /* もっとも多くの報酬を得られるランク */
sum /* 決定されている報酬総額 */
if rewardForLank(l) * m > sum then
  return false
end if
for i = 1 to i = n do
  if i ≤ m then
    R[i] ← rewardForLank(lank(Writer[i]))
  end if
end for
return R[ ]
```

---

---

**Algorithm 4** 貢献量を用いた作業者の信頼情報の変更アルゴリズム Change-Lank-of-Worker(C)

---

```
C /* 各翻訳文の貢献量の配列 */
C[i] /* i 番目に作成された翻訳文の貢献量 */
writer(i) /* i 番目に作成された翻訳文を作成した作業者 */
reputationUp(w) /* 作業者 w の評判情報を1段階増加させる */
 $\epsilon$  /* 評判情報の向上の基準となる貢献量 */
for i = 1 to i = n do
  if C[i] ≥  $\epsilon$  then
    reputationUp(writer(i))
  end if
end for
```

---

することで解決を図った。

次に、実際に評価を行う方法を設計する必要があるという課題を解決した。翻訳専門家の後編集にかかる作業時間を求める必要があるという問題に対して、翻訳文の品質と後編集の作業速度に相関があることを利用することを考えた。後編集の作業速度を翻訳文品質の機械的評価指標による評価値から近似的に算出し、次にそれから作業時間を算出した。品質評価指標としてはGTMを利用した。

次に実験を行い、本研究で採用した翻訳専門家の後編集にかかる時間の削減を基準として作業を評価する方針が多くの人に受け入れられるものであるかどうか検証した。

その後に、貢献量をどのような形で実際の報酬分配に反映するかについて、クラウドソーシングにおける既存の2つの報酬分配体系への反映の方法を提案した。1つは基礎報酬と賞与報酬の組み合わせによる報酬分配で、賞与を与える閾値として貢献量を採用した。もう1つは作業者の評判情報に基づく報酬分配であり、これは作業者にランク付けを行う基準として貢献量を採用した。

本研究の貢献は、以下の2点の課題点を解決したことである。

#### 1. 他者と共同作業をする翻訳作業者の貢献量の評価

翻訳専門家の作業時間の減少量を作業者の貢献量と考え、それを算出するアルゴリズムを設計した。具体的には、GTMを用いて翻訳文の品質を評価することで翻訳専門家の後編集にかかる作業時間を算出し、それを基に各翻訳文がどれだけの貢献を行ったかを評価した。

#### 2. 評価方法の設計

算出した貢献量を反映した報酬分配として、2つの既存の報酬分配への反映したアルゴリズムを設計し、示した。

本稿で未解決の問題もある。今回作業時間の計算にはGTMを用いたが、GTMは英文についてしか評価を行えない。CATを様々な言語に応用するためにも、他の言語における評価尺度を考案する必要がある。また、貢献量を実際に計算するためにGTMによる評価値と作業速度のデータをそこから近似的に計算する一次式を求める必要があるが、本稿ではそこまでは至らなかった。

また、提案は行ったが実際に報酬分配を設計し、CATに実装を行い、その評価も行う必要がある。

## 謝辞

本研究を行うにあたり，熱心なご指導，ご助言を賜りました石田亨教授に厚く御礼申し上げます．また，CATについての共同研究を行い，多大なご助言を賜りました石田憲幸氏にも大変感謝いたします．また，アルゴリズムや評価手法をはじめ様々なご助言を賜り，ご指導いただいた中島悠特定助教に大変感謝いたします．

また，CATについて本研究室との共同研究に協力してくださり，実験への協力をはじめ多大なご助力を頂いた株式会社石田大成社の皆様に厚く御礼申し上げます．特にご助力いただきました阿部様，藤本様，関様，齋藤様には大変感謝いたします．

そして，日頃から多くのご助言や実験などへのご協力をいただきました石田研究室の皆様に心より感謝いたします．

## 参考文献

- [1] Lin, D., Murakami, Y., Ishida, T., Murakami, Y. and Tanaka, M.: Composing Human and Machine Translation Services: Language Grid for Improving Localization Processes, *7th International Conference on Language Resources and Evaluation (LREC)*, pp. 500–506 (2010).
- [2] Doan, A., Ramakrishnan, R. and Halevy, A. Y.: Crowdsourcing systems on the World-Wide Web, *Commun. ACM*, Vol. 54, pp. 86–96 (2011).
- [3] Zaidan, O. and Callison-Burch, C.: Crowdsourcing translation: professional quality from non-professionals, *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1*, Association for Computational Linguistics, pp. 1220–1229 (2011).
- [4] Knight, K. and Chander, I.: Automated Postediting of Documents, *AAAI*, pp. 779–784 (1994).
- [5] Ishida, T.: Language grid: an infrastructure for intercultural collaboration, *Applications and the Internet, 2006. SAINT 2006. International Symposium on*, pp. 5 pp. –100 (2006).
- [6] Tanaka, M., Murakami, Y., Lin, D. and Ishida, T.: Language Grid Toolbox:

- Open source multi-language community site, *Universal Communication Symposium (IUCS), 2010 4th International*, pp. 105–111 (2010).
- [7] O’Brien, S.: Methodologies for Measuring the Correlations between Post-Editing Effort and Machine Translatability, *Machine Translation*, Vol. 19, pp. 37–58 (2005).
- [8] Levenshtein, V. I.: Binary codes capable of correcting deletions, insertions, and reversals, *Soviet Physics Doklady*, Vol. 10, No. 8, pp. 707–710 (1966).
- [9] Snover, M., Dorr, B., Schwartz, R., Micciulla, L. and Makhoul, J.: A study of translation edit rate with targeted human annotation (2006).
- [10] Papineni, K., Roukos, S., Ward, T. and Zhu, W.-J.: BLEU: a method for automatic evaluation of machine translation, *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics*, ACL ’02, Stroudsburg, PA, USA, Association for Computational Linguistics, pp. 311–318 (2002).
- [11] Turian, J., Shen, L. and Melamed, I. D.: Evaluation of Machine Translation and its Evaluation, *In Proceedings of MT Summit IX*, pp. 386–393 (2003).
- [12] Mason, W. and Watts, D. J.: Financial incentives and the ”performance of crowds”, *SIGKDD Explor. Newsl.*, Vol. 11, pp. 100–108 (2010).
- [13] Zhang, Y. and van der Schaar, M.: Reputation-based Incentive Protocols in Crowdsourcing Applications, *CoRR*, Vol. abs/1108.2096 (2011).



## 付録：実験に用いたアンケート用紙