

機械翻訳サービスと編集サービスの相互連携における 翻訳品質と編集者適応

山口 卓郎^{1,a)} 菱山 玲子^{1,b)} 北川 大輔^{2,c)} 中島 悠^{3,d)} 稲葉 利江子^{4,e)} 林 冬恵^{2,f)}

受付日 2013年6月21日, 採録日 2014年1月8日

概要: 本研究では, 人間の手により行われる翻訳リペアサービスと翻訳文の書き換えサービスを機械翻訳連携サービスのサービスフローに連携させ, 翻訳品質および作業コスト面において, この連携がどのような効果を及ぼすかを明らかにする. 実験では, 越英(ベトナム語-英語)翻訳サービスと英日翻訳サービスを連携させて知識伝達を行う翻訳サービスフローを想定し, このフローにおいて, 英日翻訳サービスにおける入力文としての英文に対し, 人間の手による前編集としての英日翻訳リペアサービスおよび後編集としての日日書き換えサービスを設定し, 出力文の流暢さと正確さの比較を行った. その結果, 後者の日日書き換えサービスは前編集としての翻訳リペアサービスを配置するのと比較して流暢な文章を得, コスト面では大幅に作業コストを削減でき, また同程度の意味を保持した文へ書き換えることができることが分かった. さらに, 人間の手によるサービスには, その作業時間の長短において2つのタイプが存在し, 作業時間の長さとして作業後に得られる文の改善度合いにおいて, 流暢さには相関がないが, 正確さについては一定の相関が見られる可能性を示した. このことから, 編集者の作業特性にあわせた翻訳サービスフロー中の作業内容の割当てや, 作業コスト制御の観点から重視する品質に応じた適切な教示を編集者に与えることの有効性を導いた.

キーワード: 多言語間コミュニケーション, 機械翻訳サービス, 翻訳リペアサービス, 書き換えサービス, 言語グリッド

Translation Quality and Editors' Adaptation in Coordination of Human and Machine Translation Service

TAKURO YAMAGUCHI^{1,a)} REIKO HISHIYAMA^{1,b)} DAISUKE KITAGAWA^{2,c)} YUU NAKAJIMA^{3,d)}
RIEKO INABA^{4,e)} DONGHUI LIN^{2,f)}

Received: June 21, 2013, Accepted: January 8, 2014

Abstract: We discuss the complementarity effect between translation services as a human task and machine translation services when these services are provided as part of a language translation Web service workflow. Previous research has found that mistranslation and loss of meaning often occurs in the machine translation workflow. To correct these errors, we set up two types of editing service allocation: a prior rewriting service process by humans, and a follow-on rewriting service process by humans. We conducted experiments to evaluate the adequacy and fluency of the sentences, as well as the costs and affinity. The results indicated that a Japanese rewriting task provided as a follow-on rewriting service by a Japanese person produced more fluent sentences than an English rewriting task provided as a prior rewriting service by a Japanese person, although they were nearly equal in adequacy. In addition, we analyzed the working characteristics of humans and found that there was no correlation between the human editorial operating time and the fluency of the output sentences. However, the experimental result suggested that the editorial operating time was potentially correlated to the adequacy of the output sentences. We concluded that it is preferable to assign the editing services of humans according to their personal working characteristics.

Keywords: multilingual communication, machine translation, translation repair service, rewriting service, language grid

1. はじめに

インターネットの普及により地球規模の社会活動がさかんになり、国際的なコミュニケーションの機会が増加している。こうした国際的なコミュニケーションで異言語の障壁を解決するため、母語によるコミュニケーションを実現する機械翻訳サービスへの期待が高まっている。さらに、機械翻訳サービスの整備が費用や需要面で困難な場合において効果を発揮する方法として、中間言語を用いた複数の翻訳サービスの連携による翻訳サービスフローの実現と、その精度向上にも期待が寄せられている [1]。NPO パンゲア [2] らは、ベトナムをフィールドとした農業支援サービスとして、ベトナム人の児童と日本人の農業専門家が相互に多言語で知識伝達を行う YMC-Viet Project [3] を遂行している。このプロジェクトでは、言語グリッド基盤 [4], [5] を用いて日英翻訳サービスと英越翻訳サービスの2つの機械翻訳サービスを連携させ、翻訳サービスフローを構成して農業知識の伝達を実現している。こうした複数の機械翻訳サービスをサービスフローとして連携させる場合、翻訳品質の確保はさらに困難であるため、サービスフローの中間で人 (Bridger) による補助的な翻訳品質改善サービスを介在させている。本研究は、このような機械による翻訳サービスが人による翻訳サービスと連携して行うような多言語間コミュニケーションをサービスフローとして、ひとまとまりの翻訳サービスを形成する場合を対象としている。機械と人が混在する翻訳サービスフローによる知識伝達の過程で、人と機械が適応的にその役割を最大限に発揮し、サービスを実用に導くためには、機械翻訳サービスの工学的な品質の最適化が必要とされる。その一方で、翻訳サービスフローに介在する人の適応を考慮することで、これらの実用への道筋を加速することができる。

以上より、本研究では、人によるサービスが機械翻訳サービスに介入する際の影響を詳しく調べるため、以下の3点を明らかにすることを目的としている。(1) 人が機械翻訳サービスと連携する場合、その翻訳文の品質および作業効

率面においてどのような効果が生じるのか。(2) 翻訳サービスと連携する人は、機械翻訳サービスに何を感じ、どのようなかたちで機械翻訳サービスと連携することを望んでいるのか。(3) 翻訳サービスに適応しようとする人はそもそも、どのような作業スタイル (態度) をとるのか。また、その作業スタイルは翻訳文の品質や作業効率に影響を与えるのか。

本研究の構成は、以下のとおりである。続く2章において、本研究における関連研究について述べる。3章で機械翻訳サービスと人によるサービスの連携において、取り組むべき問題について述べ、4章で実験の概要について述べる。5章で実験結果と考察について述べ、最後に、6章で本論文の結論をまとめる。

2. 関連研究

機械翻訳サービスを人が利用する際の翻訳品質や作業品質にまつわる研究は、折り返し翻訳 [6] や翻訳リペアの研究 [7], [8], [9] にみられるように、人が翻訳サービスの結果を利用して翻訳サービスから得られる翻訳文の品質を高める方法を中心に進められてきた。翻訳サービスを複数の人が利用する際のサービス品質にまつわる問題についても、共同翻訳サービス [10] や機械翻訳サービスを介した協調コミュニケーション [11] の効果を中心に議論がすすめられてきている。

これらの研究の多くが、サービスグリッドとして機械翻訳サービスシステムを構成することが可能な言語グリッド基盤 [4], [5] 上で行われている。言語グリッドは、機械翻訳や専門辞書、形態素解析器などの言語資源を言語サービスとして登録することによって、これらを共有することができるサービス基盤であり、ユーザはこれらの言語資源を連携させることにより、様々な機械翻訳連携サービスを構築することができる。このような利用環境で想定される機械翻訳連携サービスでは、人による翻訳リペア活動を、翻訳サービスフローを構成する Web サービスの1つとして位置づけて利用することができる。この点で、人と機械が混在する翻訳サービスフローとしての翻訳品質や作業効率を分析する必要がある。なお、連携サービスを考慮する際、人による編集サービスのもう1つの提供スタイルとして、入力文ないし出力文の書き換えサービスが存在するが、書き換えサービスが連携サービスとして組み込まれた場合の効果は明らかになっていない。

これらの連携サービスとしての分析が必要となる背景には、クラウドソーシングを利用したサービス提供や、クラウド上での人による補完的な編集サービスに関する研究と実用化の加速があげられる。クラウドソーシングは、人によるサービスを調達し効果的に利用する方法として期待が寄せられている。たとえば、Zaidan ら [12] は、非専門家による翻訳作業を、Amazon Mechanical Turk [13] を利用

¹ 早稲田大学理工学術院
Faculty of Science and Engineering, Waseda University,
Shinjuku, Tokyo 169-8555, Japan

² 京都大学情報学研究科社会情報学専攻
Department of Social Informatics, Kyoto University, Kyoto
606-8501, Japan

³ 東邦大学理学部情報科学科
Department of Information Science, Toho University,
Funabashi, Chiba 274-8510, Japan

⁴ 津田塾大学学芸学部
Department of Computer Science, Tsuda University, Kodaira,
Tokyo 187-8577, Japan

a) takuro-yamaguchi@fuji.waseda.jp

b) reiko@waseda.jp

c) kitagawa@ai.soc.i.kyoto-u.ac.jp

d) yuu.nakajima@is.sci.toho-u.ac.jp

e) inaba@tsuda.ac.jp

f) lindh@i.kyoto-u.ac.jp

したクラウドソーシングにより遂行した結果について報告している。また、Callison-Burch [14] は、翻訳サービスの品質に関する評価を Amazon Mechanical Turk 上で展開することで、非専門家からの評価を広く収集し評価することの有効性を指摘している。特に、人による編集サービスに注目すると、Bernstein ら [15] は、計算機による文書処理サービスでは実現が難しい、要約、校正、書式整形などのサービスを、人によるサービスとしてクラウド上で調達し連携することで実現している。このように、本研究が想定する利用環境は、クラウド上で人による編集サービスを調達し、制約のある言語資源環境下で機械翻訳サービスと組み合わせるサービスとして提供するもので、本研究の知見はその基礎的なサービス設計に役立てることができる。

こうした翻訳連携サービスの実用事例はまだ多くないが、前章に示した YMC-Viet Project のほか、在日外国人を対象とした多言語防災支援において、英語を中間言語とする日泰（タイ）機械翻訳連携サービスの提供事例がある [16]。この事例では、ネパール人やミャンマー人と日本人のコミュニケーションにおいて、ネパール語やビルマ語の言語資源整備が進んでいないため、母語によるコミュニケーションが実現されていない。一方、中間言語の利用という側面からは、Babych ら [1] によって、ウクライナ語から英語への翻訳において中間言語にロシア語を配した場合との比較が検証されている。今後、英語をはじめとした中間言語としての各国語の言語資源が整備され、かつ、翻訳精度の向上が図られることで、連携サービスの利用が加速し、母語によるサービス提供が可能になるものと考えられる。

なお、翻訳サービスの実用における評価について、Kita ら [17] は、前章の YMC-Viet Project を対象としてコミュニケーションログを調査し、多言語間の知識伝達の観点から、翻訳連携サービスを利用する過程で農業知識が正確に翻訳されず、知識フローが機能しないことが誤理解を生じさせることを明らかにしている。この問題に対する改善策として、日本語ネイティブ、ベトナム語ネイティブによる編集サービスを配置し、機械翻訳サービスの品質上の問題を解消するサービスフローとしたことで高品質の文を得ることに成功したものの、人によるサービスのコストが非常に高く長期間知識伝達が行えないといった問題点を指摘している。この点で、人によるサービスの介入に関する品質とコストに関する分析、編集サービスが有効に機能するための要件を明らかにする必要があるが、いまだ行われていない。このため、本研究では人による編集サービスの基礎的かつ詳細な情報を得ることで、これを明らかにする。

3. 機械翻訳サービスと編集サービスの相互連携

前章で述べたように、人によるサービスを含む実用的な

翻訳サービスフローを設計する際、人により提供されるサービスの作業コストが翻訳サービスフローの成否を分ける。そもそも、十分なコストをかけることができれば、人によるサービスのみでサービスフローを構成すればよく、高品質な翻訳出力文を得ることができる。しかし、一定の翻訳品質を担保しつつ、人が翻訳サービスフローに介在し機械翻訳サービスと連携する場合、人によるサービスにかかるコストについては機械による翻訳サービスのコスト相当ないしそれを上回る作業コストを割引いて考えたい。この点で、翻訳品質と人による作業コストのバランスを考慮することが、翻訳サービスフロー成立の要件である。

翻訳連携サービスは、複数の翻訳サービスを連携させ、ビジネスプロセスとしてのサービスを構築するが、翻訳サービス呼び出しに基づく翻訳サービスプロセス全体の管理を行うような管理機構が存在しない。人を介したサービスは適応的でありかつ自律的である。この点で、その自律的な最適化を図るための方略を考える必要がある。ここで、方略とは、サービスプロセス全体からみて最良の品質を確保するため、人がサービスとしてどのような立場で翻訳サービスフローに介入し基本的役割を果たすことが望ましいか、コストや品質とのバランスをどのように評価しながらサービスの最適化を図るか、という問題を解決するための戦略をさす。こうした方略を明らかにするための研究は、いまだ行われていない。

そこで、本研究では、人による翻訳サービスと機械翻訳によるサービスが効果的に連携するための方略を明らかにすることを提案する。この研究の課題は、機械翻訳サービスのみで構成される翻訳サービスフローのプロセス間に、人による典型的な翻訳サービス機能としての翻訳リペアサービスと書き換えサービスを介在させるフローを前提に、人によるサービスが機械による翻訳サービスによる知識伝達フローに介入するときの影響について調べることで、この方略を明らかにすることである。

本研究が前提とする翻訳サービスフローは、前章で述べた YMC-Viet Project の日越翻訳サービスを事例としている。この日越翻訳サービスは、越英翻訳サービスと英日翻訳サービスを組み合わせることで翻訳サービスフローを実現している。このフローに対し、2 言語を十分に理解するバイリンガルの翻訳リペアサービスを組み込むと、Kita ら [17] と同様、作業コストが高くなってしまう。そこで、現実社会で生じる作業コストの問題をあらかじめ考慮し、日本語ネイティブに、書き換えサービス、翻訳リペアサービスとしての任を与えることで、翻訳サービスフローに介入させる翻訳サービスフローを実験的に構成した (図 1)。日本語ネイティブが機械翻訳サービスのみからなるフロー (図 1) に介入するとき、以下の 3 種類の、人によるサービスが配置できる。

1) 機械翻訳の入力文に対して、前編集として配置する書き

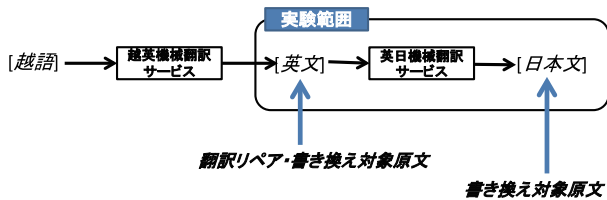


図 1 知識伝達フロー

Fig. 1 Knowledge communication flow.

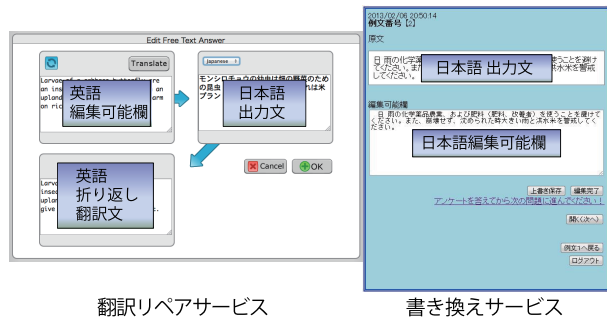


図 2 実験用インタフェース

Fig. 2 Experimental application interface.

換えサービス（同言語への書き換え）

- 2) 機械翻訳の入力文に対して、前編集として配置する翻訳リペアサービス（異言語への書き換え）
- 3) 機械翻訳の出力文に対して、後編集として配置する書き換えサービス（同言語への書き換え）

これらのうち、本研究では、日本語を何らかの作業情報として扱うことのできる 2) 翻訳リペアサービス、3) 書き換えサービスを対象に連携サービスによる実験を行う。前者の翻訳リペアサービスとは、折り返し翻訳機能を用いて、入力文を書き換えるサービスである。折り返し翻訳機能の詳細は、文献 [7] を参照されたい。この実験を行うために、我々は 2 種類の実験用インタフェースを用意した (図 2)。図 2 左は翻訳リペアサービスのインタフェースであり、図 2 右は書き換えサービスのインタフェースである。前者のインタフェースは、実際に YMC-Viet Project で使用されているインタフェースを利用しており、人が翻訳リペアサービスを行うための機能を提供する。

4. 連携サービスによる実験

4.1 実験の概要

本研究では、日本語ネイティブによる書き換えサービスおよび翻訳リペアサービスが与える効果について比較実験を行う。前述の 2 種類の異なるサービスを配置したフローについて、実験パターンは次のとおりである (図 3)。いずれのフローもベトナム語の文を入力文とし、ベトナム語ネイティブによる翻訳リペアサービスを介し、出力文として得た英文を利用する。したがって、この英文はすでに越英翻訳サービスを介しており、翻訳品質にある程度の誤訳や文法的な誤りを含んでいる。

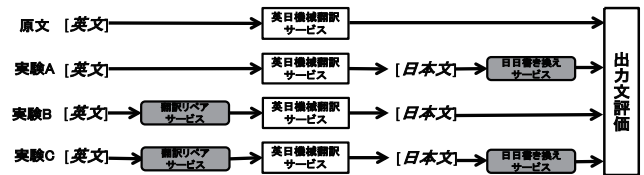


図 3 実験フロー

Fig. 3 Flows of experiments.

原文

前述の英文を、英日機械翻訳サービスへ入力することによって、得られた日本語*1。

実験 A

前述の英文を、英日機械翻訳サービスへ入力することによって得られた日本語に対し、後編集として書き換えサービスを配置することにより得られた日本語。

実験 B

前述の英文に対して前編集として翻訳リペアサービスを配置することにより得られた英文を、英日機械翻訳サービスに通すことにより得られた日本語。

実験 C

実験 B によって得られた日本語に対し、後編集として書き換えサービスを配置することにより得られた日本語。

本研究の範囲は、一定の品質上の問題を含有する英文を所与とし、この英文を日本語に翻訳するプロセスである。すなわち、翻訳サービスフローにおいて、入力文としての英文から出力文としての日本語を得る翻訳サービスフローに研究の範囲を定めたいうで、この翻訳サービスフローに日本語ネイティブがサービスとして介入する実験を行う。

本実験を行うにあたり、YMC-Viet Project の知識伝達で用いられているベトナム語と日本語の用例対訳データベース (本研究では、この用例対訳文の対を正解文とする) から実験文を選択した。この用例対訳データベースは、稲作に関する 9 カテゴリからなる 935 文の用例対訳データベースであり、日本語の平均文字数は 71.58 字 (最長: 410 字, 最短: 13 字) で、農業分野の専門用語が含まれるものも少なくない。そこで、実験文抽出にあたり、各カテゴリから網羅的に、文字数が 40~60 文字程度で専門用語 (たとえば、白葉枯病, 乾田直播, 間断灌溉, 連作障害など) が含まれるものを除外することで、文章長や語彙の難易度による影響をできるだけ排除し、均質的な 96 文の実験文セットを構成した。実験文の文字数は平均 54.31 字であり、うち、文字数 40~59 字の実験文が半数以上を占めている。本実験において、まず、用例対訳の越文に対してベトナム語ネイティブによる翻訳リペアサービスを適用し、越英翻訳サービスを介して 96 文の英文を出力した。これらが

*1 以降、「原文」とは、本フローによって出力された日本語を指す。なお、「原文フロー」とは、「原文」を出力するフローを指す。

図3における英日翻訳サービスの入力文となっている。

実験A, 実験Cの書き換えサービスを担う被験者(以下, 編集者とする)はそれぞれ3名であり, 実験B, 実験Cにおける翻訳リペアサービスを担う編集者は9名である。サービスを担ったのは, いずれも日本語を母語とし, 一般的な学部生程度の英語力を持つ日本人男子大学生であり, 特別な翻訳スキルを持たず, 翻訳や通訳を業として行っている者もない。また, 農業に関する専門知識はなく, YMC-Viet Projectをはじめとする翻訳サービスを利用した知識伝達フローにおける活動経験もない。

4.2 人によるサービスの振舞い

この実験において, 人による書き換えサービスとは, 機械翻訳サービスから出力される日本語を, 人(編集者)が直接, 流暢さと正確さの観点からより望ましい日本語へと文を書き換えるサービスである。実験開始前, 当該サービスを担う編集者に対しては「出力文の文脈などから, 入力文が保持していた意味を推測して流暢さと正確さを兼ね備えた文章へと書き換えてください」と, 教示した。

一方, 人による翻訳リペアサービスとは, 越英翻訳サービスの出力文である英文を入力文として扱い, これを翻訳リペアするサービスである。編集者は入力文の英語と翻訳結果の日本語, 折り返し翻訳結果の英語の3種類の文を同時に見ながら, 入力文の英語をリペアすることができる。実験開始前, 当該サービスを担う編集者に対しては, サンプル文を利用して図2左のインタフェースを表示し, 「入力文を機械翻訳サービスによって翻訳した結果が右上に出力され, それを再度機械翻訳サービスによって入力文と同じ言語に翻訳した結果が左下に出力されます。したがって, 左上の入力文を書き換えた後に再度機械翻訳サービスを利用することで, 書き換えた後の入力文の翻訳結果の良さを評価できます」と, 一般的な折り返し翻訳機能やその利用方法についての教示を行った。

5. 連携サービスによる実験結果と考察

5.1 評価実験

各実験フローの最終出力となる日本語に対し, 文の品質や作業コストを評価する評価実験を行った。文の品質に関する評価は文献[18]による手法を適用し, 流暢さ(Fluency)と正確さ(Adequacy)の2指標を用いて行った。本評価手法は, LDC(言語データコンソーシアム)によるNIST機械翻訳評価ワークショップにおいて開発されたスケールであり, 機械翻訳の評価において, 最も一般的に利用されており[19], 文献[19], [20]における翻訳品質の評価でも適用されている。

正確さ

gold-standard(本実験では既出の用例対訳文)と比較して, 書き換えられた日本語が意味を保持している

表1 全実験の品質比較

Table 1 Comparison of quality of all experiments.

	原文	実験A	実験B	実験C
流暢さ (pt)	1.78	3.48	2.85	4.08
正確さ (pt)	2.02	2.77	2.77	3.49

かどうかを評価。5: All 4: Most 3: Much 2: Little 1: Noneの5段階。

流暢さ

書き換えられた日本語が日本語の文法に従っているかどうかを評価。5: Flawless 4: Good 3: Non-native 2: Disfluent 1: Incomprehensibleの5段階。

評価者は, 人によるサービスを提供した者とは異なる日本語を母語とする日本人男子大学生であり, 特別な翻訳スキルを持たず, 翻訳や通訳を業として行っている者もない。原文と実験Cについては18名, 実験Aと実験Bについては20名の異なる評価者による評価を行った。評価者は前述の用例対訳データベースにある日本語を正解文として, この正解文との比較において各実験フローから得られた日本語の評価を行った。

評価にあたり, 評価者には, 96文(すべての実験文)に対して, Linら[20]と同様, 30秒以内に評価を行うように指示を行った。一方, 作業コストの評価は, 各文に対しサービスが介入した総時間の平均値を, 実験フロー間で比較することにより行った。

5.2 文の品質値による比較

図3の各実験フローから得られた日本語文の品質評価結果を, 表1*2に示す。

機械翻訳サービスのみで構成されるフローから出力される日本語の品質は(流暢さ, 正確さ)=(1.78, 2.02)となり, Kitaら[17]のYMC-Viet Projectにおいて, バイリンガルがすべて手作業で文章を翻訳する提案フローで推定された(流暢さ, 正確さ)=(4.55, 3.80)と比較して非常に低い値になっている。これは, 機械翻訳サービスのみで連携されたサービスでは, 品質の高い文を出力することができないことを示している。英日翻訳サービスの入力文となっている英文は, 越英翻訳サービスを経て出力された文であり, 誤訳としての情報の喪失や変質が含まれることが影響している。少なくとも越日翻訳を多段階の機械翻訳サービスフローのみから行う場合, そのままでは前編集としての機械翻訳サービスプロセスのサービス品質に起因する誤訳が, 後続プロセスのサービス品質に影響を与えることが分かる。

次に, 実験Aと実験Bの結果について考察する。実験Aおよび実験Bは, いずれも原文フローとしての実験パターンから出力される日本語と比較して流暢さおよび正確さにおいてそれぞれ, 実験A(1.70, 0.75), 実験B(1.07,

*2 以降, 流暢さおよび正確さの単位はポイント(pt)とする。

表 2 分割表：流暢さ (単位：人)
Table 2 Contingency Table: Fluency.

	評価値 1	評価値 2	評価値 3	評価値 4	評価値 5	計
原文	827	546	257	87	7	1,724
実験 A	146	283	459	559	473	1,920
実験 B	279	509	529	425	177	1,919
実験 C	33	125	259	565	746	1,728
計	1,285	1,463	1,504	1,636	1,403	7,291

表 3 分割表：正確さ (単位：人)
Table 3 Contingency Table: Adequacy.

	評価値 1	評価値 2	評価値 3	評価値 4	評価値 5	計
原文	718	467	337	169	33	1,724
実験 A	431	427	454	368	240	1,920
実験 B	392	461	440	440	186	1,919
実験 C	167	215	365	564	417	1,728
計	1,708	1,570	1,596	1,541	876	7,291

0.75) ほど高くなった。さらに、実験 A と実験 B の出力文を比較すると、実験 A の結果の方が流暢さが 0.63 ほど高くなった。フローに介入したサービスはいずれも、日本語ネイティブである*3。この点で、入力文の英語をリペアするのと比較して、出力文の日本語を書き換えた結果の方が流暢な文を得られたことが分かる。正確さにおいてはいずれも 2.77 となり、これらの結果から、本研究で扱った知識伝達フローにおいて、日本語ネイティブがサービスとして介入すると、同程度の正確さを維持しつつ、より流暢な文を得られるという結果を得た。

実験 C のフローでは、英日機械翻訳サービスの入力文と出力文に対して人によるサービスを配置している。実験 C の結果は (流暢さ, 正確さ)=(4.08, 3.49) となり、高品質な文を得ることに成功した。流暢さの評価指標において、4: Good, また、正確さの評価指標において、3: Much であるので、翻訳リペアサービスと書き換えサービスが介入した知識伝達フローでは、非常に流暢かつ多量に意味を保持した文を伝達できることを確認した。

5.3 例文による比較

付録表 A-1, 表 A-2, 表 A-3 として、原文、実験 A, 実験 B, 実験 C により出力された日本語と、対応する正解文を示した。3 つの例文は、用例対訳データベースより、それぞれ異なるカテゴリから抽出された例文であり、付録表 A-1 の例文は Insect pest カテゴリ、付録表 A-2 の例文は、Water management カテゴリ、付録表 A-3 の例文は Weather カテゴリに属している。いずれの例文も、評価結果に示されるように、機械翻訳サービスのみのフローから出力される文の品質は悪く、また書き換えサービスが介入

することにより、翻訳リペアサービスが介入した場合と比較して出力文の品質が高くなっている。これらの結果からも、書き換えサービスとして出力文を直接編集するサービスの有効性が確認できた。

5.4 分割表による比較

表 2 および表 3 に、評価実験によって得られた評価値の分布を表すデータを示す。

5.4.1 流暢さ

流暢さの指標において、評価値 4 以上が付与された割合は、(原文, 実験 A, 実験 B, 実験 C) において、それぞれ (5.4%, 53.8%, 31.4%, 75.9%) となった。実験 A と実験 C において、半数以上の評価値が 4 以上となっていることが分かる。また、評価値が 2 以下となった割合は、(原文, 実験 A, 実験 B, 実験 C) においてそれぞれ、(79.4%, 22.3%, 41.0%, 9.1%) となった。低評価が付与された割合においても、書き換えサービスが介入*4することによって低評価を得る文が大幅に減少することが分かる。

5.4.2 正確さ

正確さの指標において、評価値 4 以上が付与された割合は、(原文, 実験 A, 実験 B, 実験 C) においてそれぞれ、(11.7%, 31.7%, 32.6%, 56.8%) となった。実験 C において半数以上の評価値が 4 以上となり、このフローにおける出力文が多量の意味を保持していることが分かる。また、評価値 2 以下が付与された割合は、(原文, 実験 A, 実験 B, 実験 C) においてそれぞれ、(68.6%, 44.7%, 44.4%, 22.1%) となった。実験 A, 実験 B とともに低評価を減少させることが可能であることを確認したうえで、さらに翻訳リペアサービスと書き換えサービスを配置する実験 C にお

*3 本フローで翻訳リペアサービスを適用する場合は、出力文が日本語であるため、入力文は英語となる。

*4 書き換えサービスが介入する実験フローは、実験 A および実験 C である。

表 4 実験 A における文と品質の相関

Table 4 Coefficient of correlation between sentences and quality in Experiment A.

	文章長	流暢さ	正確さ
文章長	1	-0.185(.071)	-0.226(.027*)
流暢さ	-0.185(.071)	1	0.701(.000*)
正確さ	-0.226(.027*)	0.701(.000*)	1

表 5 実験 C における文と品質の相関

Table 5 Coefficient of correlation between sentences and quality in Experiment C.

	文章長	流暢さ	正確さ
文章長	1	-0.432(.000*)	-0.310(.002*)
流暢さ	-0.432(.000*)	1	0.541(.000*)
正確さ	-0.310(.002*)	0.541(.000*)	1

いては、低評価となる文を大幅に減少させることができることを確認した。

5.4.3 多重比較

設定した 4 種類の実験フローから得られた日本文の品質について、実験フローにより違いが生じるかどうか多重比較を行う。表 2, 表 3 に得られた結果より、各実験パターンごとの組合せで、Bonferroni の方法を用いて有意差の有無を調べた。その結果、すべての実験パターンの組合せの流暢さおよび正確さの指標において、有意であることを確認した ($p < .00833$)。

5.5 相関分析

本節では、編集者が書き換えサービスとして介入した日本文を対象に、文の長さと言質の関係についての分析を行う。以降、文章長とは、英日機械翻訳サービスから出力された直後の(書き換えサービスが行われる前の)文の長さ(文の文字数)を表す。実験 A における文章長の平均値は、73.2 文字^{*5}であった。また、文の品質と文章長との相関係数は表 4 のように得られた。また、表中のカッコ内に、有意確率 p を示し、5%水準で有意のものには*を付加した。

文章長と、流暢さ、正確さの相関係数が 0 に近い値になっており、文章長と文の翻訳品質に相関がないことが分かる。また、流暢さと正確さの 2 指標の間の相関係数は 0.701 となり、正の相関がある。これらの結果から、書き換えサービスが介入する文の長さにかかわらず、出力された流暢な文は正確さも兼ね備えている可能性が高いことが分かる。

実験 C における文章長の平均の概算値は、75.1 文字^{*6}であった。これに対し、書き換えサービスが介入することにより得られた文と言質との相関係数は表 5 のように得られた。

*5 本文章長は、実験 A において、英日機械翻訳サービスから出力され、書き換えサービスを適用する前の日本文の文字数である。

*6 本文章長は、実験 C において、英日機械翻訳サービスから出力され、書き換えサービスを適用する前の日本文の文字数である。

実験 A の結果と比較すると、文の長さに対して文の品質はやや負の相関があることが分かる。このことから、書き換え対象の文が長いと文の品質が低くなる傾向があるということが分かる。また、実験 A と同様、流暢さと正確さの間に正の相関があることを確認した。

5.6 作業コスト

表 6 に、各実験フローにおける作業コスト(1文あたりの作業時間)を示す。作業時間は、1文あたりの平均作業時間を表している。また、翻訳リペアサービスの作業時間には、リペアの作業時間に加え、翻訳/折り返し翻訳の処理時間(編集者が翻訳ボタンを押してから折り返し翻訳結果の表示が完了するまでの時間としてのターンアラウンドタイム)も含まれている。

表 6 中、実験 C(書き換えのみ)とは、実験 C における書き換えサービスの作業部分のみのコストを表し、実験 C とは、実験 C のフロー全体のコストを表す。

実験 A における作業時間が、実験 B と比較して大幅に小さい値になっていることが確認できる。また、実験 C(書き換えのみ)の作業時間を見ると、実験 A における書き換えの作業時間と比較して 1 文あたりの平均時間が 28.8 秒短くなっていることが分かった。実験 C において機械翻訳サービスから出力される日本文は、実験 B で翻訳リペアサービスが介入して情報が修復・改善された後に出力された文と等価である。つまり、実験 C における書き換えサービスは実験 B の出力文を書き換えることに相当し、意味を推測しやすくなり、作業コストが下がったと考えられる。また、実験 C のフローにおける作業コストは、機械翻訳サービスのみから構成される実験フローに対して、翻訳リペアサービスが介入したときの作業コストと書き換えサービスが介入したときの作業コストの合計になるため、実験 C の作業時間は表 6 に示される個々のサービスが介入したときと比較してかなり大きな値になっている。

表 1 に示される文の品質と表 6 の結果から、品質と作業コストはトレードオフの関係にあるといえる。また、機械翻訳の出力文を直接書き換えるサービスの方が、入力文をリペアする翻訳リペアサービスと比較して、低コストかつ高品質な知識伝達を行える可能性がある。

5.7 編集者に関する分析

本節では、翻訳サービスフローに対し、知識伝達の品質を向上させるために介入する、人によるサービスの特性についての考察を行う。

我々はまず、翻訳リペアサービスと書き換えサービスの間のサービスの性質上の差異に注目した。すなわち、翻訳リペアサービスの効果が書き換えサービスと比較して小さいものとなった背景として、翻訳リペアサービスとして介入する編集者が見ることのできる情報量の相違が想定され

表 6 作業コスト
Table 6 Cost comparison.

	実験 A	実験 B	実験 C (書き換えのみ)	実験 C
総作業時間 (秒)	15,570	40,135	12,802	52,937
実験文数	96	96	96	96
1 文あたりの平均作業時間 (秒)	162.2	473.3	133.4	606.6
標準偏差 (秒)	86.08	284.0	70.79	294.0

た。翻訳リペアサービスの場合、編集者は図 2 左のインタフェースに示されるように機械翻訳の入力文の英語、出力文の日本語、折り返し翻訳結果の英語という 3 種類の情報を得て、入力文の英語をリペアする。これらの情報は単に情報量が多いばかりでなく、情報の喪失や変質という誤訳要素を少なからず含んでいる。リペアした英語が機械翻訳サービスを通じて日本語に出力されるというプロセスは必ずしも、編集者にとって有益な情報をもたらすばかりでなく、それが同時に編集者にとってのノイズをもたらしていることが推定される。

これに対し、書き換えサービスでは、編集者は機械翻訳サービスの出力文のみを見ながら書き換える作業を行うため、情報量としては書き換え対象の日本語のみであるものの、そこに含まれるノイズも少なく、その影響を受けにくい環境で“人間が読みやすい”文を作り出すことに専念できる可能性がある。このことについては、編集者に与えられる情報量の違いとそこに含まれるノイズ要素の分析、ノイズ要素が編集サービスに与える影響という観点から、今後さらに詳細な分析が必要である。

なお、翻訳連携サービスの最終工程として、人による書き換えサービスを配置することによって、機械翻訳サービスに翻訳リペアサービスを組み合わせたフローでは不可能な、人間が人間に情報を伝達する際の伝わりやすさなどを兼ね備えた表現を実現することができるということが分かる。

5.7.1 被験者（編集者）アンケート

本実験では、書き換えサービス実験の全編集者（実験 A および実験 C の編集者、計 6 名）に対して、実験終了直後に質問紙による被験者（編集者）アンケートを実施した。アンケート内容は、YMC-Viet Project を例にとり、編集者が書き換えサービスとして介入するとき、本研究における実験 A と実験 B についてどちらを許容できるか、意見をもらうというものである。それぞれのサービス内容に対して、編集者から得られた意見を以下に記す。

実験 A

- 日本人なので、日本語を書き換える方が楽である。
- 書き換え作業が面倒で、書き換えなくてはならない回数が多いと許容できないが、翻訳リペアに比べるとまだ許容できる。
- 英語をリペアすることに比較すると、日本語を書き

換えるのは気軽に行うことができる。

- 日本語を書き換えるという点で、気が楽であるが、書き換え対象の文章の意味がまったく理解できないものもあった。
- 日本語ならなんとか書き換え可能である。
- 書き換え対象となる文章によって品質にばらつきがあった。

実験 B

- 英文をどのように改善すれば、出力結果の日本語がよくなるのか不明である。
- 3 つの文章を見比べなければならぬので、読まなければならない情報量が多く、許容できない。
- 母国語でない英語を読まなければならない点が許容できない。
- 日本語が機械翻訳の出力結果として見えているのに、英語をリペアしなければならないのは許容できない。
- ある程度英単語や文法が分かっている身とすれば、英語をリペアする方が分かりやすい文にリペアできるのではないかと考える。
- 英語力に自信がないため、許容できない。

翻訳リペアサービスでは、前述したように 3 種類の文を見ながら入力文をリペアすることができるため、書き換えサービスと比較してより多くの情報を得ながら作業することができる。しかしながらアンケートで得られた結果は、後編集として日本語を書き換える手法を許容する意見が多いことが分かる。これらの結果より、母語での作業は、作業に関する情報量が少ないものの、非母語での作業よりも好まれることが分かる。今後、YMC-Viet Project に限らず、人がサービスとしてその知識伝達フローに介入するケースにおいて、編集者の母語は十分に考慮する必要がある。なお、実験 A、実験 B ともにあまり許容できないという意見をjする編集者も見られた。これは、実験 A の編集者に多く見られた意見である。実験 A における書き換え対象としての日本語の流暢さ、正確さが非常に低いため、実験参加直後に採取したアンケートにおいて、編集サービスが機械サービスと連携しても効果が得られない、と考えたと予測できる。

5.7.2 編集者の作業タイプ

さらに、書き換え実験（実験 A および実験 C）の作業結

表 7 作業時間と品質上昇値
Table 7 Working time and elevated level of quality.

編集者	総作業時間 (秒)	作業文数	1文あたり作業時間 (秒)	標準偏差 (秒)	流暢さ上昇値	正確さ上昇値	時間と流暢さ相関	時間と正確さ相関
編集者 A	6,337	32	198.0	85.0	1.17	0.96	0.190(0.719)	0.797(0.058)
編集者 B	3,708	32	115.9	53.4	1.29	0.54		
編集者 C	3,777	32	118.0	44.0	1.24	0.66		
編集者 D	3,607	32	112.7	87.7	0.95	0.45		
編集者 E	3,724	32	116.4	52.1	1.01	0.44		
編集者 F	3,565	32	111.4	43.5	1.11	0.74		

果から、サービスとして介入する編集者には2種類の作業タイプが存在することが分かった。以下では、追加的に作業タイプに注目した書き換えサービス実験を行った結果を示す。例として、実験Cのフローにおいて、書き換えサービスを行った6名^{*7}の編集者の作業データを事例として、表7に取り上げる。相関係数にはカッコ内に有意確率を付与した(有意水準5%)。

特徴が現れたのは、編集者Aと他の5名の編集者である。編集者Aは、他の編集者と比較して、およそ倍の時間をかけて書き換えを行っていることが分かる。実験Aの実験フローにおいても、このようにおよそ倍の時間をかけて書き換えを行う編集者が複数確認された。表7において、流暢さ上昇値、正確さ上昇値とは、書き換え前の出力文(実験Bにおける出力文)の品質に対し、実験Cで書き換えサービスが介入することにより上昇した流暢さと正確さの品質の平均値を表す値である。作業時間に対する文の品質上昇値との相関係数を求めると、作業時間と流暢さはほとんど相関がない($r = 0.190$) 関係にあり、作業時間と正確さは正の相関($r = 0.797$) の関係にあることが分かった。この結果から、文の流暢さは、作業時間の長さには関与せず、編集者が短い時間で書き換えても高い品質となることが分かる。つまり、書き換えサービスの性質として、流暢さは作業時間にかかわらず確保できることが分かる。また、正確さに関しては、編集者Aが他の編集者よりもより多くの時間をかけて意味推測して書き換えたことにより、最終的に得られた日本語はより正確な文となっている。

今後より多くの実験データに基づく分析が必要であるものの、少なくともこの実験から得られた知見は、時間をかけて書き換え作業を行う編集者と、短時間で作業を行う編集者の2つの作業タイプがあるということである。作業対象となる不完全な文に対して様々な可能性を考え予測・評価し、時間をかけて作業するタイプと、与えられた情報に最低限の書き換えを施し短時間で作業をこなすタイプの存在が推定される。このことから、期待されるサービス品質を考慮し、編集者の特性にあわせてサービスフロー中の作業内容を割り当てるのが有益と考えられる。

多言語間知識伝達の実践現場において人によるサービスを介入させる場合、編集者がどちらの作業タイプに属するのかを、作業者の情報が何も得られていない状態で予測し分類することは困難である。しかしながら、本研究の実験と同様に均質な文に対する試行的な書き換え作業を与え、その結果を分析することにより、編集者の特性を見分けることが可能と考えられる。

また、短時間で作業を行う作業特性を持つ編集者に対しては、「時間をかけて正確な内容を伝える文を作成してください」と事前に教示することが、正確な文を得ることに寄与する可能性がある。このような事前の調査とインストラクションにより、編集者の作業タイプによって文の正確さにばらつきが生じるリスクを、ある程度回避できるものとする。編集者の特性をあらかじめ知ることができない場合、人によるサービスの品質として正確さを重視するのであれば、編集者に対して時間をかけてよく文を検討しながら作業することを教示することが有効である一方、流暢さのみを重視するのであれば必ずしもこの教示を行う必要はないことが想定される。この知見は、作業コストの考慮において活かすことができる。すなわち、流暢さのみを重視するのであれば、直感的に作業を行う作業特性を持つ編集者を選んでサービスを任せる、ないし、編集者に対し時間をかけず直感的に作業してよい旨を教示することで、作業コストを最適化しつつ期待する翻訳品質を得ることがある。

なお、表8に、流暢さないし正確さを重視したい例文を、YMC-Viet Projectで利用されている用例対訳データベースから示す。挨拶や感謝などを含む感情表現においては表現の豊かさが重要であり、流暢さを重視することがコミュニケーションに良質な効果を発揮すると考えられる。また、正確さを重視する場合として、医療分野[21]でその情報伝達の正確さが医療の安全と密接にかかわり重要であると同様に、農業分野においても農作業の実施時期や方法、農薬の分量や投薬時期などを正確に伝達することが重要であると考えられる。

5.8 評価方法に関する分析

評価実験において、評価者によって評価結果がばらつい

*7 6名の編集者のうち3名は、4章における実験フローの編集者である。

表 8 流暢さないし正確さを重視する例文

Table 8 Example texts which place emphasis on fluency or adequacy.

	用例対訳文
流暢さを重視する例 (例文番号 3-46)	うーん、困りましたね。どこで失くしたか分かりませんか。田んぼですか、家の中のどこかですか。それをまず思い出して下さい。
正確さを重視する例 (例文番号 9-25)	種を田んぼに直にまいたら、約 6 日間水を抜きます。そうして、種から芽を出させ、苗をしっかりと立たせます。たんぼの表面に、軽く亀裂が入るくらい水を抜いて下さい。

表 9 評価結果にばらつきがあった文

Table 9 Example: Sample sentences which have variability in evaluation results.

例文番号	日本文	正解文 (用例対訳文)
4	もし同じ時刻で多くの種類 (品種、種) の種稲を植え付けるならば、種々のフィールドで働くことができます。被害もコントロールできます。これは、生産性を増大できるメソッドです。	様々な種類の稲を育てることで作業時期をずらしたり、被害を抑えることができます。そうすればコメの量が増えます。
9	もし農薬の濃度が指定された濃度であるならば、	農薬の濃度は、決められた濃度の間であれば、どのような濃度でもよいです。
52	白い蛾の幼虫は水田の稲には有害だが、キャベツ等の野菜には有害ではありません。	モンシロチョウの幼虫は畑の野菜にとっては害虫ですが、イネには害がありません。
86	人は呼吸することが必要のように、光合成は米にとって必要です。従って、米が光合成できなければ、その米は枯れてしまいます。	イネの葉は呼吸しています。イネ全体が水につかると、呼吸ができなくなります。それで、枯れてしまいます。

た文についての考察を行う。各実験で得られた全評価対象文 (原文, 実験 A, B, C) に対する評価結果について、標準偏差をもとめ、ある一定以上のずれがあるものを拾い上げることににより、評価者によって評価のばらつく文を割り出した。

この分析によって得られた、評価者間に評価結果のばらつきが多い文 (原文: 12 文, 実験 A: 9 文, 実験 B: 9 文, 実験 C: 13 文) の特徴を、以下に示す。また、評価者間において評価結果がばらついた文の例を、表 9 に示した。

流暢さ (24 文)

- 文章は流暢であるが、途中で文が途切れている。
- 文章の大部分が流暢であるが、一部のみが大きく流暢さを欠いている文章。
- 期間を表す表現に分数などの数式が混入している。
- 箇条書きのように短い文章で列挙されている。
- 外来語が含まれている。
- () が用いられている。

正確さ (19 文)

- 専門用語が誤訳されている。
- 「約～日」といった曖昧な表現。
- 必要以上の情報が含まれている。
- 箇条書きのように短い文章で列挙されている。
- 外来語が含まれている。

評価の際に文に含まれるより小さい文節のまとまりを単位とし、その小さい文節に決定的な情報の喪失や変質などの誤訳が含まれる場合、その誤った文節の存在に引きずられて低い評価を行うケースと、こうした小さい文節の誤り

を気にせず文全体をひとまとまりの評価対象として評価を行うケースで、評価のばらつきがみられることが分かる。正確さと流暢さについて、人による評価を付与する評価手法はこれまで、一般的な文の品質評価の評価方法として利用されてきた。一方、こうした評価のばらつきの原因は、既存研究では考察されていない。また、こうした評価品質のばらつきを一定の範囲に収める工夫についても、明確な指針が定まっているわけではない。この点については、評価者に対する指針を与える観点から、1) 文の一部の誤訳や情報の欠落に注目することなく、あくまでも文全体の評価としてこれを行う、といった評価方法に関するより精細な教示を評価前にあらかじめ評価者に与えること、2) 評価者間で、あらかじめサンプル文などを利用して評価イメージを共有してから評価に入ることが、有益と考えられる。

6. おわりに

本研究では、今後増大するであろう機械翻訳連携サービスによる多言語間コミュニケーションで、人によるサービスが機械による翻訳サービスによる知識伝達フローに介入する場合の影響を詳しく調べたものである。日本語ネイティブが、どのように機械翻訳サービスフローに介入すべきかを分析する一連の実験を行い、実験データを分析した結果、以下の知見を得た。

(1) 日本語ネイティブが機械翻訳サービスと連携する場合、後編集としての書き換えサービスの配置は、前編集として翻訳リペアサービスを配置するのと比較して、より流暢な文を得、コスト面では大幅に削減でき、また同程度の

意味を保持した文へ書き換えることができる。また、翻訳リペアサービスと書き換えサービスを機械翻訳サービスと連携させた場合の実験結果から、翻訳品質面と作業コスト面の間にトレードオフの関係が成り立つことが分かった。

(2) 日本語ネイティブが翻訳サービスフローに介入する場合、母語による書き換えサービスの作業は、その言語を母語とする人にとって容易であり、その言語を母語とする人に（自分の母語で行う作業として）受け入れられやすい。人が翻訳サービスフローに介入してサービスの任にあたる場合、その人の母語を十分に考慮し、人からみた言語環境としての編集作業の適応に配慮することが有益と考えられる。

(3) 翻訳サービスに人が編集者として介入するとき、サービスの対象文に対して時間をかけてサービスを提供する編集者、短時間で作業を行う編集者の2つの作業タイプがある可能性を導いた。作業タイプの違いにより得られる翻訳品質特性には差が生じることが予測される。そこで、翻訳サービスフローの設計者は、作業コストの制約を考慮しつつ、編集者の持つ作業タイプにあわせた編集者の割当てや作業方法の指示を行うことで、期待する品質特性にあわせて人のサービスを最適化が図れる可能性がある。これにより、作業コストと期待する翻訳品質を制御的に扱い、翻訳サービスフロー全体の最適化を図ることができるだろう。

今後の課題として、実際に翻訳連携サービスを用いて行われる多言語間コミュニケーションの現場において、本研究の知見を適用した場合、同等の効果が得られるか検証したい。

謝辞 本論文に対して、編集委員、査読者の方々から有益なご指摘をいただきました。ここに深く感謝の意を表します。実験文の用例対訳データベースは、NPO PANGAEAよりご提供いただきました。同NPO法人の高崎俊之様に感謝いたします。また、分析に関して有用なご議論をいただいた早稲田大学経営システム工学専攻の永田靖教授に感謝の意を表します。本研究は、科学技術振興機構・戦略的創造研究推進事業（社会技術研究開発）「問題解決型サービス科学研究開発プログラム」採択プロジェクト「サービス指向集合知に基づく多言語コミュニケーション環境の実現」の成果によるものです。

参考文献

[1] Babych, B., Hartley, A. and Sharoff, S.: Translating from under-resourced languages: Comparing direct transfer against pivot translation. *Proc. MT Summit XI*, pp.10-14 (2007).

[2] NPO PANGAEA, available from <http://www.pangaean.org/web/english/general/generalopen.html> (accessed 2013-06-21).

[3] YMC Viet, available from <http://www.atpress.ne.jp/view/1906> (accessed 2013-06-21).

[4] Language Grid, available from <http://langrid.org/en/>

index.html) (accessed 2013-06-21).

[5] Ishida, T.: Language Grid: An Infrastructure for Intercultural Collaboration, *IEEE/IPSJ Symposium on Applications and the Internet (SAINT-06)*, pp.96-100 (2006).

[6] 山下直美, 坂本知子, 野村早恵子, 石田 亨, 小倉健太郎, 林 良彦, 井佐原均: 機械翻訳へのユーザ適応と書き換えに関する教示効果の分析, 情報処理学会論文誌, Vol.47, No.4, pp.1276-1286 (2006).

[7] 宮部真衣, 吉野 孝, 重信智宏: 折返し翻訳を用いた翻訳リペアの効果, 電子情報通信学会論文誌 D, 情報・システム, Vol.J90-D, No.12, pp.3141-3150 (2007).

[8] 宮部真衣, 吉野 孝: 翻訳不適箇所指摘による翻訳リペア効率の改善効果の検証, 情報処理学会論文誌, Vol.50, No.4, pp.1390-1398 (2009).

[9] Miyabe, M., Yoshino, T. and Shigenobu, T.: Effects of Repair Support Agent for Accurate Multilingual Communication, *PRICAI2008*, LNCS5351, pp.1022-1027 (2008).

[10] 森田大翼, 石田 亨: 共同翻訳のためのプロトコルの開発, 電子情報通信学会論文誌 D, Vol.J92-D, No.6, pp.739-746 (2009).

[11] Gao, G., Wang, H.C., Cosley, D. and Fussel, S.R.: Same Translation but Different Experience: The Effects of Highlighting on Machine-Translated Conversations, *Proc. SIGCHI Conference on Human Factors in Computing Systems*, pp.449-458 (2013).

[12] Zaidan, O.F. and Callison-Burch, C.: Crowdsourcing translation: professional quality from non-professionals, *Proc. ACL-HLT-2011*, pp.1220-1229 (2011).

[13] Amazon Mechanical Turk, available from (<https://www.mturk.com/mturk/>) (accessed 2013-10-22).

[14] Callison-Burch, C.: Fast, cheap, and creative: Evaluating translation quality using Amazon's Mechanical Turk, *Proc. EMNLP-2009*, Vol.1, pp.286-295 (2009).

[15] Bernstein, M.S., Little, G., Miller, R.C., et al.: Soylent: A Word Processor with a Crowd Inside, *Proc. UIST-2010*, pp.313-322 (2010).

[16] 三戸 誠, 菱山玲子: 日本在住外国人・自治体への相乗的活性化を目的とした防災支援システムの提案, 電子情報通信学会技術研究報告, Vol.112, No.435, pp.63-68, AI2012-32 (2013).

[17] Kita, K., Takasaki, T., Lin, D., Nakajima, Y. and Ishida, T.: Case Study on Analyzing Multi-Language Knowledge Communication, *2012 International Conference on Culture and Computing (Culture and Computing 2012)* (2012).

[18] LDC: Linguistic Annotation Specification: Assessment of Fluency and Adequacy in Translations, Linguistic Data Consortium (2005).

[19] Callison-Burch, C., Fordyce, C., Koehn, P., Monz, C. and Schroeder, J.: (Meta-)Evaluation of Machine Translation, *Proc. 2nd Workshop on Statistical Machine Translation*, pp.136-158 (2007).

[20] Lin, D., Murakami, Y., Ishida, T., Murakami, Y. and Tanaka, M.: Composing Human and Machine Translation Services: Language Grid for Improving Localization Processes, *Proc. International Conference on Language Resources and Evaluation (LREC 2010)*, pp.500-506 (2010).

[21] 宮部真衣, 吉野 孝, 重野亜久里: 外国人患者のための用例対訳を用いた多言語医療受付支援システムの構築, 電子情報通信学会論文誌 D, Vol.J92-D, No.6, pp.708-718 (2009).

付 録

表 A.1 各実験で得られた出力文の例 (例文 44)

Table A.1 Example: Machine translation results (sentence No.44).

例文 44	日本語	(流暢さ, 正確さ)
原文	田植え、それらに良いクモという水田。害虫のクモ食事昆虫は、有益な昆虫と呼ばれます。	(1.72, 2.39)
実験 A	クモは水田にとって良い。クモは害虫を食べるので、有益な昆虫と呼ばれています。	(4.40, 4.45)
実験 B	田植えに良いクモという水田。害虫のクモ食事昆虫は、有益な昆虫と呼ばれます。	(2.35, 2.85)
実験 C	クモは田植えに良い影響をもたらします。クモは水田にいる害虫を食べるため、有益な昆虫 (益虫) と呼ばれます。	(4.78, 4.72)
正解文	クモは田んぼに発生する害虫を食べてくれます。クモのように、稲作にとって良いことをしてくれる虫を益虫といいます。	(5, 5)

表 A.2 各実験で得られた出力文の例 (例文 51)

Table A.2 Example: Machine translation results (sentence No.51).

例文 51	日本語	(流暢さ, 正確さ)
原文	もし台風の後には稲が水中に没頭しているならば、直ちに排水されるべきです。もし稲が水中に没頭しているならば 稲停止成長と成長。	(2.22, 2.89)
実験 A	もし台風の後には稲が水に浸っているならば、直ちに排水するべきです。稲が水中に沈んでいるならば稲は成長を停止する。	(4.25, 4.50)
実験 B	もし台風の後には稲が水中に没頭しているならば、直ちにそれらを排水するべきです。もし稲が水中に没頭しているならば 稲停止成長と発展。	(2.85, 3.30)
実験 C	もし台風の後には稲が水に浸かっているならば、直ちに排水すべきです。もし稲が水に浸かったままだと、稲は成長することができないからです。	(4.61, 4.56)
正解文	台風の後には、稲が水に沈んでしまったら、すぐに田んぼから水を出します。そのままにしておくこと成長 (生長) が止まることがあります。	(5, 5)

表 A.3 各実験で得られた出力文の例 (例文 73)

Table A.3 Example: Machine translation results (sentence No.73).

例文 73	日本語	(流暢さ, 正確さ)
原文	光、温度、水、および肥料 (肥料、改善者) は、米。光の成長 条件に影響し、調整可能ではない (空気 (ガス) ポストによって) 温度けれども水、および肥料 (肥料、改善者) はコントロールできます。	(1.83, 2.67)
実験 A	光、温度、水、および肥料は、米の成長条件に影響する。温度は調整可能 (空気 (ガス) ポストによって) ではないが、水と肥料は調整可能です。	(3.60, 4.05)
実験 B	光、温度、水、および肥料 (肥料、改善者) は、米の成長 条件に影響します。そして、光、温度は調整可能ではなく (空気 (ガス) ポストによって) けれども、水および肥料 (肥料、改善者) はコントロールできます。	(3.30, 3.85)
実験 C	光や、温度、水、肥料 (肥料、改善者) は、米の成長に影響します。そして、光や温度は調整できませんが、水や肥料は人の手で調整可能です。	(4.61, 4.56)
正解文	稲が伸びる条件には「光・温度・水・肥料」が関係あります。「光・温度」の天候はコントロールすることはできませんが、「水・肥料」はコントロールできます。	(5, 5)



山口 卓郎 (学生会員)

2013年早稲田大学理工学術院創造理工学部経営システム工学科卒業。現在、早稲田大学理工学術院創造理工学研究科経営システム工学専攻在学。多言語コミュニケーション、クラウドソーシングに興味を持つ。



菱山 玲子 (正会員)

京都大学大学院情報学研究科社会情報学専攻修了。2011年早稲田大学経営システム工学専攻教授。人工知能、知識コミュニティデザインに関する研究に従事。ヒューマンインタフェース学会に所属。博士(情報学)。



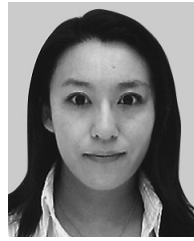
北川 大輔 (学生会員)

2012年立命館大学情報理工学部情報コミュニケーション学科卒業。同年京都大学大学院情報学研究科社会情報学専攻修士課程在学中。サービス開発に関心を持つ。



中島 悠 (正会員)

2006年京都大学大学院情報学研究科修士課程修了。日本学術振興会特別研究員(DC1)。2009年同大学院情報学博士課程修了。博士(情報学)。同大学特定研究員および特定助教を経て、現在、東邦大学理学部情報科学科講師。大規模マルチエージェントシステム、都市上の人間行動のモデル化およびシミュレーションに興味を持つ。2008年テレコムシステム技術学生賞受賞。



稲葉 利江子 (正会員)

1998年日本女子大学理学部数物科学科卒業。2003年同大学大学院理学研究科数理物性構造科学専攻博士課程修了。博士(理学)。同年文部科学省メディア教育開発センター助手。(独)情報通信研究機構専攻研究員。京都大学情報学研究科特定講師を経て、現在、津田塾大学准教授。異文化コラボレーションの研究に従事。電子情報通信学会、ACM各会員。



林 冬恵 (正会員)

1982年生。2005年上海交通大学大学院計算機科学専攻修士課程修了。2008年京都大学大学院情報学研究科社会情報学専攻博士課程修了。博士(情報学)。2008年独立行政法人情報通信研究機構専攻研究員。2012年京都大学大学院情報学研究科社会情報学専攻特定助教。サービスコンピューティング、異文化コラボレーション、クラウドソーシング、サービス科学の研究に従事。