

翻訳サービスにおける多言語同時表示による影響の評価

Evaluation of the effect of multi-language simultaneous display in a translation system

佐藤 瑞樹 *¹ 菱山 玲子 *¹

Mizuki Sato

Reiko Hishiyama

*¹早稲田大学大学院 創造理工学研究科経営システム工学専攻

Graduate School of Creative Science and Engineering, Waseda University

It is assumed that Japanese people can analogize the meaning of an original text if they read the original text by the commonality of written expression such as using Kanji in the case of Chinese-Japanese translation. In this study, we have the subjects rewrite a translated text while displaying simultaneously the original text written in Chinese and the text translated into Japanese by the Language Grid as machine translation service. In this rewriting work, we analyzed the effect on chat when rewriting a text using simultaneous display of the original text and the translated text. Also, we analyzed and evaluated the effect on the rewritten translated text. As a result, we find that simultaneous display has an influence on the rewriting work to make it possible to analogize the meaning of the original text. Furthermore, the result shows that simultaneous display improves the translation quality of the rewritten translated text.

1. はじめに

世界中にある情報を共有するためには、他国語で書かれた文章を翻訳する必要がある。中国語で書かれた文章を日本語に翻訳する場合、文字に漢字を使うといった言語的表記の共通性から原文の意味を類推することが可能と想定される。よって、翻訳文の書き換え作業を行う際、原文を表示することが翻訳文作成作業に効果的な影響を及ぼすと考えられる。

そこで本研究では、機械翻訳サービスとして言語グリッド [1] を用いて、原文と日本語に機械翻訳された文（翻訳文）を同時にディスプレイ表示することが、多言語チャットを使いながら書き換え作業を行う際に、どのような影響を及ぼすかを評価する。また、書き換えられた翻訳文にどのような違いが生まれるかを分析・評価する。

2. 関連研究

書き換え作業に関する研究として、翻訳リペアと書き換えサービスを連携したサービスフローに対する研究 [2] が行われている。この研究では、英日翻訳リペアサービスと日々書き換えサービスを連携することが、出力文の翻訳品質としての正確さと流暢さにおいて、どのような効果を及ぼすかを分析している。この研究によって、書き換え作業によって翻訳文をより流暢な文に書き換えられることが示されている。しかし、書き換え作業時において原文を表示することが、書き換え作業、および翻訳文の翻訳品質にどのように影響するのかは分析されていない。

本研究では、日本語と中国語の言語的表記の類似性に着目し、書き換え作業時に原文を表示することが、書き換え作業時に行われるチャットの内容、および書き換えられた翻訳文の翻訳品質にどのような影響を及ぼすかを分析する。

3. 提案：翻訳サービス利用時の原文参照効果

中国語に用いられる漢字と日本語に用いられる漢字は、同じ漢字を用いる場合もあれば、異なる場合もある。しかし、異なる漢字を用いる場合であっても、日本語に用いられる漢字と形状が似ている場合もある。形状に関する例が、表 1 である。

表 1: 日本語と中国語における漢字の形状の差

	中国語	日本語
同じ漢字を用いる場合	交通	交通
異なる漢字を用いる場合	统计	統計

中国語と日本語の両言語において「交通」という言葉は同じ漢字を用いて表記されるが、中国語で「統計」を意味する漢字は、日本語では用いられていない漢字である。しかし、その形状から「統計」を意味する漢字であることを類推することが可能である。

本研究では、機械翻訳によって得られた翻訳文を書き換え作業者が原文と翻訳文を比較することで類推しながら翻訳文を書き換える作業を行う。このとき、書き換え作業者は原文である中国語を母国語とする中国人と、多言語化したチャットを用いて翻訳文の意味を聞き、翻訳文を書き換える。そして、作業過程で得られるチャット内容と書き換え作業によって得られた結果を評価・分析することを提案する。

4. 実験

4.1 実験システム

実験のシステムは、PHP と HTML, MySQL を用いた Web アプリケーションである。システムの構成とシステムのインターフェースはそれぞれ図 1, 図 2 に示す。記事 DB にあらかじめ、原文と翻訳文を保存する。そして、原文のみの表示か原文と翻訳文の同時表示かを選択することで、記事 DB から原文、あるいは翻訳文を呼び出す。呼び出された原文は上部左側の画面に、翻訳文は上部右側の画面に表示される。多言語チャットは画面右側に配置し、書き換え作業は画面下部で行い、書き換えた翻訳文は修正記事 DB に登録される。

連絡先: 佐藤瑞樹, 早稲田大学 創造理工学研究科経営システム工学専攻, 〒169-8555 東京都新宿区大久保 3-4-1 51 号館 15 階 02 号室, 080-1294-1293, dogwood@suou.waseda.jp

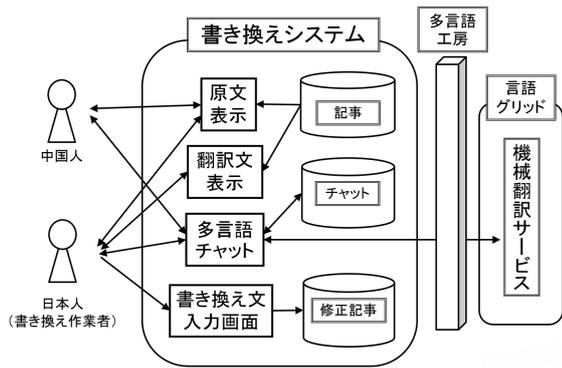


図 1: システム構成図



図 2: システムのインターフェース

4.2 実験内容

実験は、中国人と書き換え作業者である日本人の 1 対 1 で実験を行う。被験者は、中国人 2 人と日本人 2 人である。中国人 1 人と日本人 1 人の 2 人でペアを組み、実験 1 と実験 2 を別のペアで 1 回ずつ行う。実験の流れは表 2 の通りである。

表 2: 実験の流れ

順番	原文 (中国語)	翻訳文 (日本語)	実験 1	実験 2
1	—	○	記事 1	記事 3
2	—	○	記事 2	記事 4
3	○	○	記事 3	記事 1
4	○	○	記事 4	記事 2

書き換え作業者の画面に、翻訳文 (日本語) のみを表示する場合と、原文 (中国語) と翻訳文 (日本語) を同一画面上に表示した場合の 2 パターンで翻訳文の書き換え作業を行う。実験は、まず翻訳文 (日本語) のみを表示した場合で 2 つの記事の書き換え作業を行う。2 つの記事の書き換え終了後、原文 (中国語) と翻訳文 (日本語) を同時表示した場合で 2 つの記事の書き換え作業を行う。原文として用いる中国語の文章は、中国の新聞社である人民日報社が発行する「人民日報」のインターネット版である「人民網」[4] の日本語版から、中日対訳が存在する記事を 4 つ選択して用いる。但し、翻訳文 (日本語) は、機械翻訳サービスによって出力された翻訳文であり、「人民網」日本語版の対訳記事とは異なる。用意した 4 つの記事には、1~4 の番号を振り分けた。

実験 1 においては、記事 1・2 は翻訳文 (日本語) のみを表示し、記事 3・4 は原文 (中国語) と翻訳文 (日本語) の両方を表示して、書き換え作業を行ってもらった。実験 2 においては、表示する記事の順番を入れ替えて、記事 3・4 は翻訳文 (日本語) のみを表示し、記事 1・2 は原文 (中国語) と翻訳文 (日本語) を同時表示して、書き換え作業を行った。これにより、4 つの記事全てに対して、翻訳文 (日本語) のみを表示した場合と原文 (中国語) と翻訳文 (日本語) を同時表示した場合の書き換え作業を行った。

4.3 評価方法

書き換えられた翻訳文の評価方法として、翻訳品質を表す評価指標としての流暢さ (Fluency) と正確さ (Adequacy) を 5 段階で評価する手法を用いる。これは、LDC[5] 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評価指標である。この評価は、翻訳品質を評価する手法として一般的に活用されており、山口らの研究 [2] においても翻訳品質の評価手法として適応されている。各項目に対して、評価者 10 人に評価してもらう。

5. 実験結果と考察

5.1 チャットにおける発話数と作業時間

実験によって得られた多言語チャットにおける各記事での日本人と中国人の発話数を合わせた総発話数と作業時間の結果が、表 3 である。

表 3: 各実験における記事毎の総発話数と作業時間

	実験 1		実験 2	
	総発話数	作業時間	総発話数	作業時間
記事 1	17	42 分	31	56 分
記事 2	7	22 分	33	45 分
記事 3	7	24 分	37	28 分
記事 4	7	30 分	41	55 分

表 3 によると、書き換え作業者によって、発話数に差があることが分かる。実験 1 では、記事 1 だけが他に比べて発話数が多いが、全体的に見て発話数は少ない傾向にある。それに対して、実験 2 では全ての記事において 30 以上発話していることが分かる。これは、書き換え作業者によって発話数に差が生まれていると考えられる。このことは、山口らの研究 [2] において、作業者には積極的に発話する作業者と発話に消極的な作業者の 2 種類がいることが示されており、本実験においても同じ傾向が見られた。しかし、表示形式を変えることによる発話数の変動はなく、同時表示することによる発話数への影響はないと考えられる。

書き換え作業の作業時間は、記事 3 では実験 1・実験 2 との作業時間の差が小さく、作業時間が短いことがわかった。これは、記事の内容によって、作業時間に差が生まれていることを示している。しかし、表示形式を変えることによる作業時間の変動はなく、同時表示することによる作業時間への影響はないと考えられる。

5.1.1 チャット内容の分析

多言語チャットにおける発話内容を分析すると、原文と翻訳文を同時に表示することで書き換え作業者は中国人に対して翻訳文に関する質問がしやすくなっていることがわかる。また、原文を見ることができると意味を被験者側が推測しやすくなったことが考えられる。表 4・表 5 は、推測しやすくなっていることを示す例である。

表 4: 実験 1 における記事 3 でのチャット内容

発話者	日本語
日本人	端正な字はまた (まだ) 表して、とはどういう意味でしょうか？
中国人	[端正な字] この人、また (まだ) 表しました....

表 5: 実験 2 における記事 3 でのチャット内容

発話者	日本語
日本人	「委員会の主任」の後に「端正な字」と翻訳されているのですが、ここは人の名前ですか？
中国人	[端正な字], 人の名前です

原文において「周正宇」という人物の名前が、機械翻訳サービスによって日本語では「端正な宇」と翻訳された。これは、機械翻訳サービスが「周正宇」という単語が固有名詞であることを識別できなかったために発生した誤訳である。この単語に対して、表4から、翻訳文（日本語）のみを表示した場合、書き換え作業者は「端正な宇とはどういう意味か」という質問をしていることが示されている。これに対し、表5から、原文（中国語）と翻訳文（日本語）の両方を表示した場合、書き換え作業者は「端正な宇とは、人の名前ですか」という質問をしていることが示されている。このことから、原文を表示することで、書き換え作業者は「周正宇」という単語が固有名詞であることと、それが人物の名前を表す固有名詞であることを推測できたと考えられる。

よって、機械翻訳サービスによって誤訳が起こったとしても、書き換え作業時に原文を表示することによって、原文と翻訳文を比較することにより、書き換え作業時に書き換え作業者が誤訳に気付いて、書き換え時に修正することができる可能性があることが示された。

また、チャット内容を分析することによって、原文と翻訳文を同一画面上に表示しても書き換え作業によって翻訳品質を向上させることに影響を及ぼしにくいものがあることが分かった。その一つとして考えられるものが、定型文である。失敗例として取り上げる定型文は、記事3の文章中に登場した中国のことわざである「美食は美器に如かず」という定型文である。このことわざの意味は「おいしく食べるには、器も美しいほうがいい」という意味だが、機械翻訳サービスによって、「美食は米器に及ばないで」と翻訳された。この文に対して、実験1・実験2のいずれにおいても書き換え作業者は中国人に対してチャットにおいて質問している。以下の表6・表7が、各実験における質問の発話内容である。

表6: 実験1における定型文に関するチャット内容

発話者	日本語
日本人	美食は米器に及ばない、とはどういうことですか？
中国人	意味は [美食が美しい食器に及ばない]

表7: 実験2における定型文に関するチャット内容

発話者	日本語
日本人	「その他に、諺によると美食は米器に及ばないというのは、」という部分は適切ですか？
中国人	美しい食器に及ばないです
日本人	わかりました

チャットによって、実験1・実験2のどちらにおいても機械翻訳によって得られた翻訳文に出てくる「米器」という単語が誤訳であることが示された。しかし、どちらのチャットにおいても文として適切な意味を中国人とチャットをすることで得ることはできなかった。その原因として考えられることは、文章全体を指し示して、その分がどういうことを意味するかを質問すると、中国人側も質問に対して答えを文章全体で返すためである。文章全体をチャットを用いて説明すると、文章全体を機械翻訳して相手に返すことになる。このとき、流暢な文章に翻訳されると文章中に誤訳があったとしても気付かないことがある。これが誤訳をそのまま残してしまう原因になっていると考えられる。このことから、日本語の文章として文法的に適切である場合、すなわち流暢な文が返ってくると正確さも適切であると判断してしまうことがあることが分かった。

5.2 評価の平均値による比較

翻訳文を書き換えることで得られた書き換え翻訳文の翻訳品質を評価して得られた評価値の平均を表8に示す。

表8: 翻訳品質の平均値（単位：pt）

	機械翻訳	日本語表示	両言語表示
流暢さ	1.48	2.88	3.98
正確さ	1.60	2.85	3.70

表8における機械翻訳とは、機械翻訳サービスによって得られた翻訳文のことである。また、表1における日本語表示とは、翻訳文のみを同一画面上に表示した状態で得られた書き換え翻訳文のことであり、両言語表示とは、原文と翻訳文を同時に表示した状態で得られた書き換え翻訳文のことである。以後の表においても、この表記を用いる。

ここで、機械翻訳と日本語表示、両言語表示を流暢さ・正確さのそれぞれで比較する。すると、機械翻訳と日本語表示では流暢さ(1.40pt)・正確さ(1.25pt)、機械翻訳と両言語表示では流暢さ(2.50pt)・正確さ(2.10pt)の差が生まれた。よって、機械翻訳文よりも書き換え翻訳文の方が流暢さ・正確さの評価が高く、翻訳品質が向上していると考えられる。また、日本語表示と両言語表示を比較すると、流暢さ(1.10pt)・正確さ(0.85pt)の差が生まれている。このことから、原文と翻訳文を同時に表示することは、翻訳品質の向上に影響を及ぼすと考えられる。

5.3 評価の分割表による比較

評価によって得られた結果を分割表にまとめたものが、表9・表10である。

表9: 分割表：流暢さ（単位：人）

翻訳文\評価値	1	2	3	4	5	計
機械翻訳	23	15	2	1	0	40
日本語表示	6	7	15	11	1	40
両言語表示	0	3	7	18	12	40
計	29	25	24	29	13	120

表10: 分割表：正確さ（単位：人）

翻訳文\評価値	1	2	3	4	5	計
機械翻訳	18	20	2	0	0	40
日本語表示	3	8	20	9	0	40
両言語表示	0	4	9	22	5	40
計	21	32	31	31	5	120

表9、表10に得られた結果より、2以下の評価値の割合が、流暢さでは（機械翻訳、日本語表示、両言語表示）において（95.0%、32.5%、7.5%）になり、正確さでは（95.0%、27.5%、10.0%）になった。また、4以上の評価値の割合は、流暢さでは（0.0%、30.0%、75.0%）になり、正確さでは（0.0%、22.5%、67.5%）になった。このことから、書き換え作業によって、翻訳文の流暢さ・正確さを向上させられていると考えられる。そこで、すべてのパターンで正確さ・流暢さにおいて検定を行った。その際、有意水準はBonferroni法により調整した（ $p < 0.0167$ ）。その結果、機械翻訳と日本語表示、機械翻訳と両言語表示、日本語表示と両言語表示のそれぞれで有意差が確認された。よって、書き換え作業を行うことで、翻訳文の翻訳品質を向上させることが可能であり、また、書き換え作業時に両言語を表示することが翻訳品質の向上に影響を及ぼすことが考えられる。

5.4 アンケートの分析

本節では、実験後に書き換え作業者の2人に行った作業のしやすさに関するアンケートを分析する。

5.4.1 作業のしやすさについて

書き換え作業者に対して、書き換え作業後に作業のしやすさに関するアンケートを行った。翻訳文の書き換え作業において、書き換え作業を行う際の作業のしやすさに対して、書き換え作業を行った記事毎に（4：とても作業がしやすかった，3：作業がしやすかった，2：作業がしづらかった，1：とても作業がしづらかった）の4段階で評価してもらった。その結果が、以下の表11である。

表 11: 作業のしやすさについて

	記事 1	記事 2	記事 3	記事 4
実験 1	2	2	3	2
実験 2	3	3	2	1

表 11 によると、作業のしやすさは翻訳文のみを表示している場合よりも原文と翻訳文を両方表示している場合の方が、全ての記事において1 ずつ差がついている。また、表示形式毎の評価の平均値は、(翻訳文のみを表示している場合、原文と翻訳文を同時表示している場合)において、実験 1 では (2.0, 2.5)、実験 2 では (1.5, 3.0) となり、実験 1・2 のいずれも原文と翻訳文を同時表示している場合の方が平均値が高い。よって、書き換え作業は、原文と翻訳文を同時表示している方が作業がしやすいと考えられる。

また、作業のしやすさと翻訳品質の関係性を調べる。作業のしやすさを横軸に翻訳品質の縦軸にグラフを書いた結果が、以下の図 3・図 4 である。

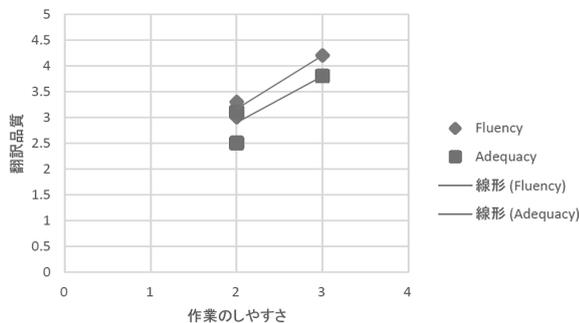


図 3: 実験 1 における翻訳品質と作業のしやすさ

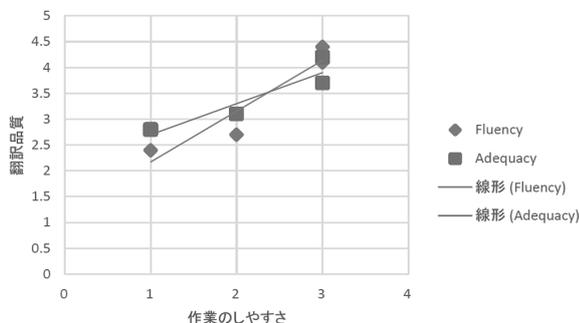


図 4: 実験 2 における翻訳品質と作業のしやすさ

図 3・図 4 より、流暢さ・正確さと作業のしやすさにはやや正の相関があると考えられる。このことから、作業のしやすさを向上させることで翻訳品質を向上させることになると考えられる。

5.4.2 作業のしやすさに関する意見

作業のしやすさが翻訳品質の向上に影響を及ぼすことが分かった。そこで、翻訳品質作業のしやすさに影響を及ぼすものが何かを調べるために、アンケートにおいて、作業がしやすいと感じた点に関して意見を集めた。得られた意見には、記事の内容に関する意見・機械翻訳文に関する意見・多言語チャットに関する意見があるが、多言語チャットに関する意見の中で「原文を読むことで、機械翻訳によって分かりづらく翻訳されている部分が原文のどこに当たるのかが分かったため、質問がしやすかった。」という回答があった。これは、チャット内容の分析において考察した内容と合致している。よって、原文を表示することは、書き換え作業者に多言語チャットにおいて質問をしやすくしており、作業のしやすさに影響を与えていると考えられる。

6. まとめと今後の課題

本研究では、書き換え作業を行う際に中国語で書かれた原文と機械翻訳された日本語の翻訳文を同一画面上に表示することによって、書き換え作業にどのような影響を及ぼすかを分析し、得られた結果を評価した。

その結果、原文と翻訳文を同一画面上に表示することで、書き換え作業者は原文の意味を推測しやすくなることが分かった。また、書き換え作業時のチャットの発話数や書き換え作業時間には影響がないが、チャットにおいて中国人に質問しやすくさせることがわかった。そして、書き換え作業が行いやすくなることによって、書き換えられた翻訳文の翻訳品質を向上させることが分かった。

今後の課題は、翻訳する文章の量によって、同時表示が及ぼす影響に差が生まれるかどうかの分析が挙げられる。また、作業時間を制限した状態で翻訳文のみと同時表示した場合との比較検証を行うことも挙げられる。

7. 謝辞

本研究は、JSPS 科研費 (S)(24220002, 2012-2016) の助成により行われたものである。

参考文献

- [1] 独立行政法人情報通信研究機構：言語グリッドプロジェクトポータルサイト，入手先 〈<http://langrid.org/jp/>〉（参照 2016-03-19）。
- [2] 山口卓郎，菱山玲子，北川大輔，中島悠，稲葉利江子，林冬恵：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，情報処理学会論文誌，Vol.55, No.4, pp.1303-1315 (2014) 。
- [3] 多言語工房-Language Grid：入手先 〈<http://langrid.org/developer/jp/index.html>〉（参照日 2016-03-19）。
- [4] 人民網ホームページ：入手先 〈<http://j.people.com.cn/>〉（参照 2016-03-19）。
- [5] LDC: Linguistic Annotation Specification: Assessment of Fluency and Adequacy in Translations, Linguistic Data Consortium (2005)。