

折り返し翻訳文と対象言語翻訳文の精度不一致要因の分析

宮 部 真 衣^{†1} 吉 野 孝^{†2}

機械翻訳を介したコミュニケーションでは、翻訳精度が低い場合、十分な相互理解ができない可能性が高い。現在、母語のみを用いて自分の発言がどのように伝わっているのかを把握するための手法として、折り返し翻訳が用いられている。折り返し翻訳を精度確認手法として用いる場合、対象言語翻訳と折り返し翻訳の精度に正の相関関係があり、対象言語と折り返し翻訳の精度が大きく異ならないことが求められる。これまでに、折り返し翻訳の精度確認手法としての妥当性の検証を行った結果、折り返し翻訳文と対象言語翻訳文の精度には正の相関がみられた。また、対象言語翻訳文が不正確であるにもかかわらず、折り返し翻訳文が正確であるという状況（第1種の精度不一致）の発生率は低いことがわかった。しかし、発生数は少ないものの、第1種の精度不一致が発生する場合があることがわかった。そこで本稿では、第1種の精度不一致が発生し得る入力を行った場合に対応できるようにするために、第1種の精度不一致が発生した文についての分析を行い、第1種の精度不一致の発生要因を明らかにする。

Analysis of features of accuracy mismatch between back-translated sentence and target-translated sentence

MAI MIYABE^{†1} and TAKASHI YOSHINO^{†2}

In communication using a machine translation, inaccurate translation prevents effective communication among individuals, and leads to misunderstandings. Back translation is used to check the accuracy of a sentence translated to a native language. In order to use back translation as a method for checking the accuracy of a translated sentence, back translation has to satisfy the following conditions: there is a positive correlation between the accuracy of sentences translated to an intermediate language and that of back-translated sentences, and there is no significant difference between them. From the results of our verification, we found that back translation satisfies the conditions. However, we found that accuracy mismatch case 1, which means that a back-translated sentence is accurate but the translated sentence is inaccurate, occurred although the number of accuracy mismatch case 1 was small. In this paper, we discussed the causes of accuracy mismatch case 1.

1. はじめに

近年、世界規模のインターネットの普及に伴ったインターネット上の使用言語の多様化により、ネットワークを介した多言語間コミュニケーションの需要が高まっている。しかし、一般に多言語を十分に習得することは難しく、母語以外の言語によりコミュニケーションを行うことは困難である。そのため、多言語間コミュニケーションにおいては、相互理解ができない可能性が高い^{1),2)}。現在、母語でのコミュニケーションを支援するために、機械翻訳技術を用いた支援が行われている³⁾。

近年、機械翻訳技術は急速に進展しているが、高精度な翻訳を行うことは困難である。機械翻訳を介したコミュニケーションでは、翻訳精度が低い場合、十分な相互理解ができず、思い違いが発生する⁴⁾。このような思い違いを回避するためには、自分の発言がどのように伝わっているのかを把握する必要がある。しかし、原文に対する多言語の翻訳結果を見て、正しく翻訳されているかどうかを判断することは容易ではない。母語のみを用いた多言語の翻訳精度の把握は、折り返し翻訳を利用することにより実現可能である。折り返し翻訳とは、他言語への翻訳結果を再度母語へと翻訳することである。折り返し翻訳を精度確認手法として用いる場合、対象言語翻訳と折り返し翻訳の精度に正の相関関係があり、対象言語と折り返し翻訳の精度が大きく異ならないことが求められる。

我々はこれまでに、折り返し翻訳の精度確認手法としての妥当性の検証を行った⁵⁾。検証の結果、折り返し翻訳文と対象言語翻訳文の精度には正の相関がみられた。一方で、対象言語翻訳文が不正確であるにもかかわらず、折り返し翻訳文が正確であるという状況（第1種の精度不一致）が発生する場合があった。

本稿では、折り返し翻訳文と対象言語翻訳文の精度における、第1種の精度不一致の発生要因の分析を行う。

2. 折り返し翻訳利用の課題

折り返し翻訳は、母語のみを用いて自分の発言がどのように伝わっているのかを把握す

^{†1} 和歌山大学大学院システム工学研究科

Graduate School of Systems Engineering, Wakayama University

^{†2} 和歌山大学システム工学部

Faculty of Systems Engineering, Wakayama University

入力文

それじゃあ、よろしくお願いします。

対象言語翻訳文

那么，谢谢你。

折り返し翻訳文

さて、あなたに感謝します。

原言語から対象言語への翻訳

対象言語から原言語への翻訳

図 1 折り返し翻訳の流れ

Fig. 1 Procedure of back translation.

るための手法として、機械翻訳を介したコミュニケーションにおいて利用されている⁶⁾⁻¹⁰⁾。折り返し翻訳の流れを図 1 に示す。原言語への再翻訳によって得られる折り返し翻訳文は、「原言語から対象言語への翻訳」および「対象言語から原言語への翻訳」という、2 回の翻訳を介している。「対象言語から原言語への翻訳」を行うことにより、対象言語の翻訳文の意味と折り返し翻訳文の意味が同一でなくなる可能性がある。折り返し翻訳は、対象言語翻訳文の精度を確認されるために用いられるため、折り返し翻訳文の精度が対象言語翻訳文と大きく異なる場合、精度確認手法として用いるのは適切ではないと考えられる。折り返し翻訳を精度確認手法として用いるには、次の 2 つの条件を満たす必要がある。

- (1) 対象言語翻訳と折り返し翻訳の精度が正の相関関係にあることが保証されている
- (2) 対象言語と折り返し翻訳の精度が大きく異なる

これまでに、折り返し翻訳の精度確認手法としての妥当性の検証を行った⁵⁾。検証にあたり、以下の 2 種類の精度不一致状況を定義した。

[第 1 種の精度不一致]: 折り返し翻訳文の精度が 高い が、対象言語翻訳文の精度が 低い

[第 2 種の精度不一致]: 折り返し翻訳文の精度が 低い が、対象言語翻訳文の精度が 高い

第 1 種の精度不一致が発生すると、入力者は伝わったと判断した内容が、相手の言語では正しく伝わらず、意思疎通が困難になる。この状況が多数発生する場合、精度確認の手法として折り返し翻訳を使うことは適切ではない。一方、第 2 種の精度不一致が発生すると、実際は修正しなくても伝わる可能性のある文を、伝わらない可能性があるとして判断してしまう。

表 1 評価に用いたテキストの一部

Table 1 Examples of sentences used in the evaluation.

機械翻訳試験文	(1)	私は窓の外を見た。
	(2)	この小説は想像していたより面白かった。
	(3)	梅雨には天気が変わり易いことに留意することが必要だ。
	(4)	唯一の違いは彼がコーヒーを飲んだのに対して、彼女が紅茶を飲んだことだ。
チャットにおける発言	(5)	でもかっこいいですね。
	(6)	私も小さいころはちょっと怖かったです。
	(7)	ちょっと興味あるんですが屋台でも家でもやったこと無いですねー。
	(8)	好きな人はとことん好きな店ですけど、無理な人は絶対嫌って言いますねー。

(1) ~ (4) のテキストは、機械翻訳機能試験文¹¹⁾ から 5 文字以上 44 文字以下である文を 200 文選択したものの一部である。
(5) ~ (8) のテキストは、「好きなもの・嫌いなもの」というテーマのチャットにおける対話文のうち、5 文字以上 44 文字以下であった文を 200 文選択したもの的一部分である。

この場合、ユーザは本来不要な修正作業等を行う可能性があるが、第 1 種の精度不一致のような、意思疎通等の問題の発生にはつながらない。

そこで、精度確認手法としての妥当性を判断するために、対象言語翻訳と折り返し翻訳の精度の相関の有無および第 1 種の精度不一致の発生率について検証を行った。検証の結果、折り返し翻訳文と対象言語翻訳文の精度には正の相関がみられた。また、対象言語翻訳文が不正確であるにもかかわらず、折り返し翻訳文が正確であるという状況（第 1 種の精度不一致）の発生率は低く、折り返し翻訳を精度確認手法として用いることに大きな問題がないことを示した。一方で、第 1 種の精度不一致の発生率は低いものの、0%ではないことがわかった。

そこで本稿では、第 1 種の精度不一致の発生要因の分析を行う。

3. 翻訳精度の主観評価実験

本稿では、折り返し翻訳の妥当性の検証実験⁵⁾において発生した第 1 種の精度不一致について分析を行う。本章では、分析対象とする対象言語翻訳文を取得した実験の概要について述べる。

3.1 評価テキスト

評価テキストとして「機械翻訳試験文¹¹⁾」および「チャットにおける発言」の 2 種類の文を用いた。チャットにおける発言は、「好きなもの・嫌いなもの」というテーマでのチャットにおける対話文を用いた。評価テキストの一部を表 1 に示す。評価テキストは、5 文字以上 44 文字以下の文とし、「5 文字以上 14 文字以下」「15 文字以上 24 文字以下」「25 文字以上

34 文字以下」「35 文字以上 44 文字以下」の文をそれぞれ 50 文、合計 200 文を各テキストセットからランダムに選択し、利用した。

また、原言語の違いによる影響を検証するために、実験用に抽出した日本語の機械翻訳試験文 200 文の英語対訳、中国語対訳、韓国語対訳^{*1}を用いて、原言語が英語、中国語、韓国語の場合の評価を行うこととした。

3.2 使用言語

折り返し翻訳の際の原言語と対象言語の組み合わせを以下の 6 種類とし、精度評価を行った。

- [ペア 1] 原言語：日本語、対象言語：英語
- [ペア 2] 原言語：日本語、対象言語：中国語
- [ペア 3] 原言語：日本語、対象言語：韓国語
- [ペア 4] 原言語：英語、対象言語：日本語
- [ペア 5] 原言語：中国語、対象言語：日本語
- [ペア 6] 原言語：韓国語、対象言語：日本語

3.3 翻訳システム

翻訳文の取得については、言語グリッド¹²⁾を介して以下の 3 種類の翻訳システムを利用した。

- [システム A] J-Server (高電社)¹³⁾
- [システム B] Google 翻訳 (Google)¹⁴⁾
- [システム C] WEB-Transer (クロスランゲージ)¹⁵⁾

3.4 評価方法

折り返し翻訳文、対象言語翻訳文の主観評価は、Walker らの適合性評価 (5 段階評価)¹⁶⁾により行った^{*2}。適合性評価では、以下の評価基準を用いて、2 つの文の意味の比較を行う。

- 5: All (同じ意味)
- 4: Most (文法などに多少問題があるが、大体同じ意味)
- 3: Much (意味は何となく掴める)
- 2: Little (雰囲気は残っているが、もとの意味はわからない)
- 1: None (全く違う意味)

本実験では、以下の組み合わせの文について、翻訳文が入力文と同じ意味になっているか

^{*1} 英語対訳は、機械翻訳試験文内に用意されていたものを用いた。中国語対訳および韓国語対訳については、それぞれ中国語翻訳者、韓国語翻訳者に作成してもらった対訳を用いた。

^{*2} Walker らの適合性評価は、2 名以上で行うものである。

どうかを比較する。

- [1] 入力文 (日本語) とその折り返し翻訳文 (日本語)
- [2] 入力文 (日本語) とその対象言語翻訳文 (英語、中国語、韓国語)
- [3] 入力文 (英語) とその折り返し翻訳文 (英語)
- [4] 入力文 (英語) とその対象言語翻訳文 (日本語)
- [5] 入力文 (中国語) とその折り返し翻訳文 (中国語)
- [6] 入力文 (中国語) とその対象言語翻訳文 (日本語)
- [7] 入力文 (韓国語) とその折り返し翻訳文 (韓国語)
- [8] 入力文 (韓国語) とその対象言語翻訳文 (日本語)

評価者は、日本人大学生 3 名および英語翻訳者 4 名、中国語翻訳者 4 名、韓国語翻訳者 4 名である。日本人大学生は [1] の文の比較評価を行った。英語翻訳者は [2] [3] [4] の比較評価を行った。中国人翻訳者は [2] [5] [6] の比較評価を行った。韓国人翻訳者は [2] [7] [8] の比較評価を行った。

4. 実験結果

4.1 精度不一致の基準値

翻訳精度は、3.4 節に示した 5 段階の評価基準によって評価した。精度不一致の発生数を検証するためには、本実験で用いた評価基準において、精度不一致であると判断する基準値を設定する必要がある。

実験で用いた評価基準による評価においては、評価者によって「意味が同一である」と判断する基準が異なるため、同一の翻訳文に対する各評価者による評価値は同一になるとは限らない。評価結果を確認したところ、同一のテキストに対する評価値は、評価者によって異なる場合が多数存在した。すべての翻訳文の評価 (10800 文) に関して、各文に対する評価の最大値と最小値との差を求めたところ、評価の差の平均値は 1.32、標準偏差は 0.97 であった。そこで、本稿では 2.29 (平均値 1.32 + 標準偏差 0.97) を許容できる差の上限 (精度不一致の基準値) とすることとし、2 種類の精度不一致状況への該当条件を、以下のよう

[第 1 種の精度不一致]:

$$(\text{折り返し翻訳文の精度評価値}) - (\text{対象言語翻訳文の精度評価値}) \geq 2.29$$

[第 2 種の精度不一致]:

$$(\text{折り返し翻訳文の精度評価値}) - (\text{対象言語翻訳文の精度評価値}) \leq -2.29$$

表 2 第 1 種の精度不一致の発生率

Table 2 Occurrence rate of accuracy mismatch case 1.

テキストセット	原言語	対象言語	第 1 種の精度不一致数		
			J-Server	Google 翻訳	WEB-Transter
機械翻訳試験文	日本語	英語	1 文 (0.5%)	0 文 (0%)	1 文 (0.5%)
		中国語	0 文 (0%)	3 文 (1.5%)	2 文 (1.0%)
		韓国語	1 文 (0.5%)	6 文 (3.0%)	0 文 (0%)
	英語	日本語	0 文 (0%)	3 文 (1.5%)	1 文 (0.5%)
	中国語	日本語	1 文 (0.5%)	2 文 (1.0%)	3 文 (1.5%)
	韓国語	日本語	0 文 (0%)	6 文 (3.0%)	8 文 (4.0%)
チャットにおける発言	日本語	英語	0 文 (0%)	3 文 (1.5%)	4 文 (2.0%)
		中国語	0 文 (0%)	3 文 (1.5%)	2 文 (1.0%)
		韓国語	2 文 (1.0%)	12 文 (6.0%)	0 文 (0%)
合計			5 文	38 文	21 文

表 3 第 1 種の精度不一致の要因

Table 3 Causes of accuracy mismatch case 1.

要因	該当数 (文)
(1) 対象言語翻訳文に原言語の表現が残っている	14
(2) 不適切な語句が対訳として選択されている	15
(3) 語句の翻訳に失敗している	32
(4) 入力文中に存在した単語が欠落している	1
(5) 語句の区切り方に失敗している	2
(6) 文における語順がおかしい	1

4.2 精度不一致状況の発生数

各翻訳システムおよび各テキストセットに関する、精度不一致の発生数の確認を行った。各翻訳システムにおける折り返し翻訳文および対象言語翻訳文の第 1 種の精度不一致発生数を表 2 に示す。表 2 より、第 1 種の精度不一致の発生数は、最大 12 文 (6.0%)、最小 0 文 (0%) であった。

5. 精度不一致の要因

第 1 種の精度不一致は、折り返し翻訳文の精度が高いが、対象言語翻訳文の精度が低い状態である。すなわち、対象言語から原言語への翻訳を介すことにより、折り返し翻訳文の精度が対象言語翻訳文の精度よりも高くなるという状況が発生していることを意味する。4.2 節で示した第 1 種の精度不一致が発生した文について、対象言語翻訳文の精度低下原因

表 4 第 1 種の精度不一致の要因の例

Table 4 Example sentences of accuracy mismatch case 1.

要因	入力文	対象言語翻訳文
(1)	髭面の先生が来た。	髭面 박사가왔다.
(2)-(a)	수업이 <u>난다</u> .	ひげが <u>飛ぶ</u> .
(2)-(b)	住友商事は米国のCATV局を <u>買収する</u> ため、四月をメドに投資グループを結成する。	To <u>brbe</u> an American CATV bureau, Sumitomo Corp. forms an investment group around April.
(2)-(c)	彼は私の <u>顔をつぶした</u> 。	他弄碎了我的脸。
(3)-(a)	<u>プチ</u> 難民です。	<u>Petit</u> is a refugee.
(3)-(b)	I <u>looked outside the window</u> .	私は <u>窓の外に見えた</u> 。
(4)	<u>フープロ</u> により文書を作る。	通过创建字处理文档。
(5)	나는 동료에 대해서 계획은 <u>예정 대로</u> 임을 연락한다.	私は仲間に対して計画は <u>予定大道</u> なのを連絡する。
(6)	Until all the data are collected, Steps 4 and 5 are repeated.	まで、すべてのデータが収集され、手順4〜5を繰り返している。

赤字・下線で示される部分は、不一致の要因となった語句である。

の確認を行った。確認の結果、今回第 1 種の精度不一致が発生した文については、以下の点
が第 1 種の精度不一致の要因となった可能性があることがわかった。それぞれの発生要因に
対する該当数を表 3 に示す。また、各要因における例を表 4 に示す。

(1) 対象言語翻訳文に原言語の表現が残っている

表 4(1) のように、対象言語翻訳文において、原言語の表現が残っており、対象言語
翻訳文の精度が低下していた。

(2) 不適切な語句が対訳として選択されている

(a) 多義語における選択の失敗

表 4(2)-(a) に示した韓国語入力文は、「ひげが生える」という意味の文である。
韓国語の「生える」という単語は、「飛ぶ」という意味としても使うことがで
きる多義語である。韓国語から日本語への翻訳において、「生える」ではなく
「飛ぶ」が選択されたために、対象言語翻訳文の精度が低下していた。なお、
表 3(2) の 15 文中、3 文がこの要因であった。

(b) 意味は間違っていないが、文中で用いると不自然になる

表 4(2)-(b) については、「買収する」が「bribe」へと翻訳されている。「bribe」

は「買収する」という意味を持つが、入力文のように「企業を買収する」という場合は不自然になるため、対象言語翻訳文の精度が低いと判断されていた。なお、表 3(2) の 15 文中、7 文がこの要因であった。

(c) 文としては成立するが、意味が間違っている

表 4(2)-(c) に示した中国語翻訳文は、日本語入力文の「顔をつぶした」をそのまま直訳した文となっている。そのため、日本語では「体面を損なう」ということを意味する文が、中国語では「(物理的に) 顔をつぶす」と翻訳されており、入力文の意味と異なると判断され、精度が低下していた。なお、表 3(2) の 15 文中、5 文がこの要因であった。

(3) 語句の翻訳に失敗している

(a) 入力文における主語の欠落による品詞の誤変換

表 4(3)-(a) の入力文は、主語が省略されている。そのため、入力文中では形容詞として扱われている「ブチ」という表現が、対象言語翻訳文において主語として扱われており、精度が低下していた。なお、表 3(3) の 32 文中、3 文がこの要因であった。

(b) その他

表 4(3)-(b) のように、入力文中の語句が正しく翻訳されておらず、対象言語翻訳文の精度が低いと判断されていた。なお、表 3(3) の 32 文中、29 文がこの要因であった。

(4) 入力文中に存在した単語が欠落している

表 4(4) の中国語翻訳文においては、それぞれの語句は翻訳されている。しかし、文として見ると、入力文に存在する「ワープロ」という単語が存在せず、対象言語翻訳文の精度が低下していた。

(5) 語句の区切り方に失敗している

表 4(5) の韓国語入力文は、「私は仲間に対して計画は予定通りである旨連絡する。」という意味の文である。この文では、韓国語の「予定通り」という表現が「予定」「通り」という別々の単語として扱われ、「通り」が「大通」へと翻訳されたため、対象言語翻訳文の精度が低下していた。

(6) 文における語順がおかしい

表 4(6) のように、文中の単語それぞれは翻訳されているが、語順がおかしくなっており、文としてみると不自然になるため、対象言語翻訳文の精度が低いと判断されていた。

6. おわりに

機械翻訳を介したコミュニケーションにおいて、折り返し翻訳は母語のみを用いた多言語の翻訳精度の把握手法として用いられている。折り返し翻訳文は、「原言語から対象言語への翻訳」および「対象言語から原言語への翻訳」という、2 回の翻訳を介しており、「対象言語から原言語への翻訳」を行うことにより、対象言語の翻訳文の意味と折り返し翻訳文の意味が同一でなくなる可能性がある。しかし、対象言語翻訳文と折り返し翻訳文の精度の同等性についてはこれまでに検証されていない。精度の同等性が確保されていない場合、折り返し翻訳を精度確認のための手法として用いるのは適切ではないと考えられ、折り返し翻訳の精度確認手法としての妥当性を保証する必要がある。

本稿では、機械翻訳試験文およびチャットにおける発言文の折り返し翻訳文取得において、第 1 種の精度不一致が発生した文をもとに、第 1 種の精度不一致の発生要因の分析を行った。分析の結果、今回用いた文については、6 種類の精度不一致要因があることを示した。

謝辞 本研究の一部は、独立行政法人科学技術振興機構「平成 21 年度シーズ発掘試験 A (発掘型)」、日本学術振興会科学研究費 基盤研究 (B)(19300036) および基盤研究 (B)(22300044) の補助を受けた。

参 考 文 献

- 1) Aiken, M.: Multilingual Communication in Electronic Meetings, ACM SIG-GROUP, Bulletin, 23, 1, pp.18-19 (2002).
- 2) Tung, L.L. et al.: Cultural differences explaining the differences in results in GSS: implications for the next decade, Decision Support Systems, 33, 2, pp.177-199 (2002).
- 3) Inaba, R.: Usability of Multilingual Communication Tools, Proceedings, Lecture Notes in Computer Science 4560, pp.91-97 (2007).
- 4) Yamashita, N. et al.: Automatic prediction of misconceptions in multilingual computer-mediated communication, Proc. the 11th international conference on Intelligent user interfaces, pp.62-69 (2006).
- 5) 宮部真衣, 吉野孝: 機械翻訳を介したコミュニケーションのための折り返し翻訳の妥当性の検証, 電子情報通信学会技術報告, 人工知能と知識処理, AI2009-41, pp.65-70

- (2010).
- 6) 坂本知子, 野村早恵子, 石田亨, 井佐原均, 小倉健太郎, 林良彦, 石川開, 小谷克則, 島津美和子, 介弘達哉, 畠中伸敏, 富士秀, 船越要: 機械翻訳システムに対する利用者適応の分析 - 異文化コラボレーションを目指して -, 情報処理学会研究報告, 2003-ICS-135, pp.125-130 (2004).
 - 7) 藤井薫和, 重信智宏, 吉野孝: 機械翻訳を用いた異文化間チャットコミュニケーションにおけるアノテーションの評価, 情報処理学会論文誌, Vol.48, No.1, pp.63-71 (2007).
 - 8) 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, Vol.J92-D, No.6, pp.739-746 (2009).
 - 9) 岸田章, 間瀬心博, 北村泰彦: 協調型機械翻訳システムのためのガイド入力インタフェースの開発, 電子情報通信学会論文誌, Vol.J92-D, No.6, pp.729-738 (2009).
 - 10) 菅原研次, 真部雄介, 藤田茂 ほか: 多言語型協調設計支援システムの試作, 電子情報通信学会技術研究報告, AI2008-48, Vol.108, No.441, pp.13-18 (2009).
 - 11) NTT Natural Language Research Group,
<http://www.kecl.ntt.co.jp/icl/mtg/resources/index.php>
 - 12) Ishida, T.: Language Grid: An Infrastructure for Intercultural Collaboration, IEEE/IPSJ Symposium on Applications and the Internet (SAINT-06), pp.96-100(2006).
 - 13) KODENSHA, <http://www.kodensha.jp/>
 - 14) Google 翻訳, <http://translate.google.co.jp/>
 - 15) クロスランゲージ, <http://www.crosslanguage.co.jp/>
 - 16) Walker, K. et al.: Multiple-Translation Arabic (MTA) Part 1, Linguistic Data Consortium, Philadelphia (2003).