

折返し翻訳文と対象言語翻訳文の精度不一致要因

宮部 真衣^{†a)} 吉野 孝^{††b)}

Features of Accuracy Mismatch between Back-Translated Sentence and Target-Translated Sentence

Mai MIYABE^{†a)} and Takashi YOSHINO^{††b)}

あらまし 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、折返し翻訳が用いられている。折返し翻訳を精度確認手法として用いる場合、対象言語翻訳と折返し翻訳の精度に正の相関関係があり、対象言語翻訳と折返し翻訳の精度が大きく異なることが求められる。これまでに、折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、折返し翻訳文と対象言語翻訳文の精度には正の相関が見られた。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分かった。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面があることが分かった。第1種の精度不一致は、コミュニケーションにおいて意思疎通を困難にする可能性が高く、不一致が発生した際の対応が必要となる。そこで本論文では、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った。分析の結果、第1種の精度不一致の発生要因として、5種類の要因があることを示した。

キーワード 多言語間コミュニケーション、機械翻訳、折返し翻訳

1. ま え が き

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、折返し翻訳を利用することにより実現可能である。折返し翻訳とは、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、対象言語翻訳と折返し翻訳の精度に正の相関関係があり、対象言語翻訳と折返し翻訳の精度が大きく異なることが求められる。

我々はこれまでに、折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、折返し翻訳文と対象言語翻訳文の精度には正の相関が見られた。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）が発生する場面があった。第1種の精度

[†] 東京大学知の構造化センター, 東京都
Center for Knowledge Structuring, The University of Tokyo,
Tokyo, 113-8656 Japan

^{††} 和歌山大学システム工学部, 和歌山市
Faculty of Systems Engineering, Wakayama University,
Wakayama-shi, 640-8510 Japan

a) E-mail: miyabe@yoslab.net

b) E-mail: yoshino@sys.wakayama-u.ac.jp

不一致が発生すると、入力者が折返し翻訳を見て「伝わる」と判断した内容が、実際には相手の言語では正しく伝わらず、意思疎通が困難になる。そのため、第1種の精度不一致が発生し得る入力を行った場合の対応が必要である。

そこで本論文では、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、折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。

以下、2.において折返し翻訳利用の課題について述べる。3.では翻訳精度の主観評価実験について述べる。4.で評価結果を示し、5.で精度不一致の分析結果について述べる。6.で精度不一致要因についての考察を述べる。最後に7.で本論文の結論についてまとめる。

2. 折返し翻訳利用の課題

折返し翻訳は、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]。折返し翻訳の流れを図1に示す。折返し翻訳による精度確認においては、対象言語翻訳文とその折返し翻訳文が、どちらも入力文(原言語)に対する翻訳結果であるとみなす。ユーザは入力文と折返し翻訳文を比較することにより、母語のみで翻訳精度を確認することができる。原言語への再翻訳によって得られる折返し翻訳文は、「原言語から対象言語への翻訳」及び「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、対象言語翻訳文の意味と折返し翻訳文の意味が同一でなくなる可能性がある。折返し翻訳は、対象言語翻訳文の精度を確認するために用いられるため、折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。折返し翻訳を精度確認手法として用いるには、次の二つの条件を満たす必要がある。

(1) 対象言語翻訳と折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) 対象言語翻訳と折返し翻訳の精度が大きく異なるならない

本研究では、入力文と対象言語翻訳文の意味の一致度を「対象言語翻訳の精度」と呼ぶ。また、入力文と折返し翻訳文の意味の一致度を「折返し翻訳の精度」と呼ぶ。

これまでに、折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]: 折返し翻訳文の精度が低い、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判断してしまう。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、対象言語翻訳と折返し翻訳の精度の相関の有無及び第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、折返し翻訳文と対象言語翻訳文の精度には正の相関が見られた。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^(注1)は低く、折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%ではないことが分かった。第1種の精度不一致は、コミュニケーションにおいて意思疎通を困難にし、コミュニケーションに混乱

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 折返し翻訳の流れ

Fig. 1 Procedure of back translation.

(注1): 発生率については、4.2において提示する。なお、発生率の検証に用いたデータは3.に示す。

を起こす要因となる。そのため、第1種の精度不一致の発生率が低いとしても、第1種の精度不一致によるコミュニケーションにおける問題の発生を防ぐために、第1種の精度不一致への対応方法を検討していく必要がある。そこで本論文では、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種の精度不一致の発生要因の分析を行う。

3. 翻訳精度の主観評価実験

本論文では、折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]」及び「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^(注2)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

折返し翻訳の際の原言語と対象言語の組合せを以下の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

翻訳精度は、翻訳する言語ペアや翻訳システムに依存しており、同じテキストを翻訳しても、同じ結果を得ることができるとは限らない [12]。そのため、特定のシステムへの依存を減らすために、複数の翻訳システムを用いて実験を行うこととした。今回の実験において、翻訳文の取得については、言語グリッド [13] 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) [14]
- [システムB] Google 翻訳 (Google) [15]
- [システムC] WEB-Transer (クロスランゲージ) [16]

3.4 評価方法

折返し翻訳文、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) [17] により行った^(注3)。適合性評価では、以下の評価基準を用いて、二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 (意味は何となくつかめる)

2: Little (雰囲気は残っているが、もとの意味は分からない)

1: None (全く違う意味)

本実験では、以下の組合せの文について、翻訳文が入力文と同じ意味になっているかどうかを比較する。

- (1) 入力文 (日本語) とその折返し翻訳文 (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名及び英語翻訳者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に示した5段階の評価基準によっ

(注2)：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。中国語対訳及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。

(注3)：Walkerらの適合性評価は、2名以上で行うものである。

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 文選択したものの一部である。

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。全ての翻訳文の評価 (10800 文^(注4)) 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、本論文では 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、以下のように定義する。

[第 1 種の精度不一致]:

(折返し翻訳文の精度評価値) - (対象言語翻訳文の精度評価値) ≥ 2.29

[第 2 種の精度不一致]:

(折返し翻訳文の精度評価値) - (対象言語翻訳文の精度評価値) ≤ -2.29

折返し翻訳文と対象言語翻訳文の精度評価値の差の分布を図 2 に示す。本論文における分析対象となる第 1 種の精度不一致発生数は、5400 文^(注5) のうち 64 文であった。

4.2 精度不一致状況の発生数

各翻訳システム及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返し翻訳文及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

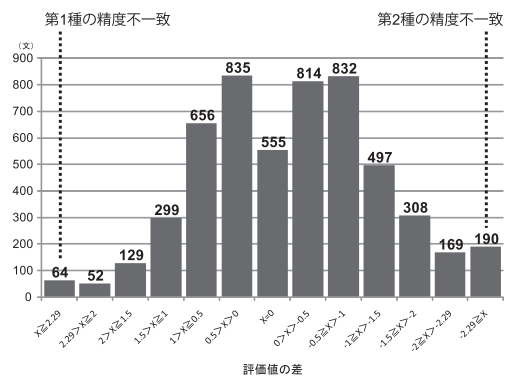


図 2 精度評価値の差の分布

Fig. 2 Distribution of accuracy difference between target-translated sentence and back-translated sentence.

(0%) であった。第 1 種の精度不一致の発生率は低いが、第 1 種の精度不一致はコミュニケーションにおける混乱を引き起こす要因となるため、原因を分析し、回避方法を検討する必要がある。そこで、次章において、これらの第 1 種の精度不一致に関して、発生要因を分析する。

(注4): 機械翻訳試験文 (200 文) は全ての使用言語ペア (6 種類) について、チャットにおける発言 (200 文) はペア 1~3 (3 種類) についてそれぞれ実験を行うため、文数は 1800 文となる。それぞれに対して折返し翻訳文及び対象言語翻訳文の評価を 3 種類の翻訳システムで評価するため、今回評価が行われた文の総数は 1800 文 \times 2 (折返し翻訳、対象言語翻訳) \times 3 (翻訳システム数) で 10800 文となる。

(注5): 機械翻訳試験文 (200 文) は全ての使用言語ペア (6 種類) について、チャットにおける発言 (200 文) はペア 1~3 (3 種類) についてそれぞれ実験を行うため、文数は 1800 文となる。3 種類の翻訳システムで評価するため、今回精度評価値の差が計算された折返し翻訳文・対象言語翻訳文のペアの総数は 1800 文 \times 3 (翻訳システム数) で 5400 文となる。

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	1 文 (0.5%)	2 文 (1.0%)	3 文 (1.5%)
		韓国語	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	全体の 該当数 (文)	使用言語						翻訳システム		
		ペア 1 (文)	ペア 2 (文)	ペア 3 (文)	ペア 4 (文)	ペア 5 (文)	ペア 6 (文)	A (文)	B (文)	C (文)
(1) 対象言語翻訳文に原言語の表現が残っている	14	3	1	10	0	0	0	1	9	4
(2) 不適切な語句が対訳として選択されている	15	3	4	4	1	0	3	2	7	6
(3) 語句の翻訳に失敗している	33	3	5	8	1	6	10	2	22	9
(4) 語句の区切り方に失敗している	2	0	0	0	0	0	2	0	0	2
(5) 文における語順がおかしい	1	0	0	0	1	0	0	0	1	0
合計	65	9	10	22	3	6	15	5	39	21

一つの精度不一致文において、複数の要因が含まれる場合があったため、該当数には重複がある。

また、同一の入力文かつ同一の言語ペアであっても、異なる翻訳システムを用いた場合、異なる対象言語翻訳文が得られたため、それぞれ別の文として扱っている。

なお、表中のペア 1~6 は、それぞれ以下の言語ペア (原言語-対象言語) を示す。

ペア 1: 日本語-英語, ペア 2: 日本語-中国語, ペア 3: 日本語-韓国語, ペア 4: 英語-日本語, ペア 5: 中国語-日本語, ペア 6: 韓国語-日本語

5. 精度不一致の要因の分析

第 1 種の精度不一致は、折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 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の確認を行った。なお、同一の入力文かつ同一の言語ペアであっても、異なる翻訳システムを用いた場合、異なる対象言語翻訳文が得られたため、それぞれ別の文として扱っている。

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分かった。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。

なお、一つの精度不一致文において、複数の要因が含まれる場合があったため、表 3 における該当数には

重複がある。

(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」という意味をもつが、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

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	対象言語翻訳文
(1)	髭面の先生が来た。	髭面 박사 가 왔다.	髭面の先生が来た。
(2)-(a)	수업이 난다. (訳: ひげが生える)	ひげが飛ぶ。	수업이 난다.
(2)-(b)	住友商事は米国のCATV局を買収するため、四月をメドに投資グループを結成する。	To bribe an American CATV bureau, Sumitomo Corp. forms an investment group around April.	アメリカCATV局を買収するために、住友商事は4月ごろに投資グループを結成する。
(2)-(c)	彼は私の顔をつぶした。	他弄碎了我的脸。	彼は私の顔をつぶした。
(3)-(a)	プチ難民です。	Petit is a refugee.	プチ難民です。
(3)-(b)	ワープロにより文書を作る。	通过创建字处理文档。	ワープロ文書を作成する。
(3)-(c)	I looked outside the window	私は窓の外に見えた。	I looked outside the window.
(4)	나는 동료에 대해서 계획은 예정 대로임을 연락한다. (訳: 私は仲間に対して計画は予定通りである旨連絡する。)	私は仲間に対して計画は予定大道なを連絡する。	나는 동료에 대해서 계획은 예정 대로인 것을 연락한다.
(5)	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。	Until all data is collected, steps 4-5 are repeated.

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。

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、入力文中では接頭辞として扱われている「プチ」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 33 文中、3 文がこの要因であった。

(b) 入力文中に存在した単語が欠落している

表 4(3)-(b) 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」という単語が対象言語翻訳文中に存在せず、対象言語翻訳文の精度が低下していた。折返し翻訳文においては、中国語における「文書」が「ワープロ文書」へと翻訳されたため、高精度であると判断されたと考えられる。なお、表 3(3) の 33 文中、1 文がこの要因であった。

(c) その他

表 4(3)-(c) 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 33 文中、29 文がこの要因であった。

(4) 語句の区切り方に失敗している

表 4(4) 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。

(5) 文における語順がおかしい

表 4(5) 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、文として見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. 考 察

6.1 使用言語及び翻訳システムによる違い

本研究では、3.2 で示した 6 種類の使用言語ペア及び 3.3 で示した 3 種類の翻訳システムを用いて検証を行った。本論文で分類した第 1 種の精度不一致の発生要因に対する使用言語別の該当数を表 3 における「使用言語」列に提示した。また、翻訳システム別の該当数を表 3 における「翻訳システム」列に提示した。

まず、使用言語による違いについて確認したところ、表 3 より、要因 (1) (対象言語翻訳文に原言語の表現が残っている) 及び要因 (2) (不適切な語句が対訳として選択されている) については、原言語が日本語である使用言語ペア (ペア 1, ペア 2, ペア 3) において発生する傾向が見られた。特に要因 (1) については、原言語が日本語である使用言語ペアのみで発生してお

り、要因 (1) については、原言語が影響している可能性がある。その他の要因については、該当数自体が少なく、特徴的な違いは見られなかった。

次に、翻訳システムによる違いについて確認したところ、表 3 より、システムによって第 1 種の精度不一致の発生数が異なるため、該当数に違いはあるものの、翻訳システムごとの該当数合計に占める各発生要因の該当数を見ると、翻訳システムによる特徴的な違いは見られなかった。

ただし、今回用いたデータにおいては、第 1 種の精度不一致発生数自体が少なかったため、使用言語及び翻訳システムの影響については、今後より多くのデータを収集し、議論する必要があると考えられる。

6.2 発生要因への対応方法の検討

本節では、5. において分類した第 1 種の精度不一致の発生要因への対応方法について議論する。

要因 (1) (対象言語翻訳文に原言語の表現が残っている) については、対象言語翻訳文中に原言語の文字が含まれているかどうかを検証することにより、第 1 種の精度不一致の発生を検出可能であると考えられる。ただし、原言語と対象言語のどちらもアルファベットを使っているなど、同じ文字を使っている翻訳言語ペアの場合、検出が困難になる可能性もある。

一方、要因 (2)-(a), (2)-(b), (2)-(c), (3)-(a), (3)-(b), (3)-(c), (4), (5) については、対象言語翻訳文が正しいかどうかを判定しなければならないため、第 1 種の精度不一致の発生を検出することは容易ではないと考えられる。ただし、対象言語翻訳文がごちなく、不自然な文になっている場合 (要因 (2)-(a) (多義語における選択の失敗), (2)-(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) など) については、対象言語翻訳文が不自然でないかどうかを、文中の語の共起頻度などにより判断することで、検出できる可能性があると考えられる^(注6)。

また、要因 (2)-(c) (文としては成立するが、意味が間違っている)、要因 (3)-(a) (入力文における主語の欠落による品詞の誤変換) については、第 1 種の精度不一致発生防止・検出は困難であるが、入力文が明確な発生要因 (慣用表現を含む、あるいは主語が欠落している) をもっており、入力文を変更することによって、第 1 種の精度不一致の発生を回避できる可能性がある

ある。翻訳する前にユーザの入力文をチェックし、第 1 種の精度不一致を引き起こす可能性がある (慣用表現を含む、あるいは主語が欠落している) 場合、ユーザに入力文の修正を促すなどの対応を行うことで、第 1 種の精度不一致の発生しにくい文章を作成し、第 1 種の精度不一致の発生を防止できる可能性がある。

7. む す び

機械翻訳を介したコミュニケーションにおいて、折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。折返し翻訳文は、「原言語から対象言語への翻訳」及び「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、対象言語翻訳文の意味と折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。

本論文では、機械翻訳試験文及びチャットにおける発言文の折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、5 種類の精度不一致要因があることを示した。

今後は、今回得られた結果をもとに、折返し翻訳を用いたツールにおいて、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仕組みの検討を行う。

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) 及び基盤研究 (B) (22300044) の補助を受けた。

文 献

- [1] M. Aiken, "Multilingual communication in electronic meetings," ACM SIGGROUP, Bulletin, vol.23, no.1, pp.18-19, 2002.
- [2] L.L. Tung and M.A. Quaddus, "Cultural differences explaining the differences in results in GSS: Implications for the next decade," Decision Support Systems, vol.33, no.2, pp.177-199, 2002.
- [3] R. Inaba, "Usability of multilingual communication tools," Lect. Notes Comput. Sci., vol.4560, pp.91-97, 2007.

(注6) : 例えば日本語の場合、Web 日本語 N グラム [18] (日本語の単語 n-gram とその出現頻度をまとめた大規模言語リソース) などの言語資源を用いることによって、共起頻度を判断できると考えられる。

- [4] N. Yamashita and T. Ishida, "Automatic prediction of misconceptions in multilingual computer-mediated communication," Proc. 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," 信学技報, AI2009-41, 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指して," 情処学研報, 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," 情処学論, 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," 信学論 (D), vol.J92-D, no.6, pp.739-746, June 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," 信学論 (D), vol.J92-D, no.6, pp.729-738, June 2009.
- [10] 菅原研次, 真部雄介, 藤田 茂, L. Wouters, C. Moulin, "多言語型協調設計支援システムの試作," 信学技報, AI2008-48, 2009.
- [11] NTT Natural Language Research Group, <http://www.kecl.ntt.co.jp/icl/mtg/resources/index.php>
- [12] K. Ogura, Y. Hayashi, S. Nomura, and T. Ishida, "User adaptation in MT-mediated communication," First International Joint Conference on Natural Language Processing (IJCNLP-04), pp.596-601, 2004.
- [13] T. Ishida, "Language grid: An infrastructure for intercultural collaboration," IEEE/IPSJ Symposium on Applications and the Internet (SAINT-06), pp.96-100, 2006.
- [14] KODENSHA, <http://www.kodensha.jp/>
- [15] Google 翻訳, <http://translate.google.co.jp/>
- [16] クロスランゲージ, <http://www.crosslanguage.co.jp/>
- [17] K. Walker, M. Bamba, D. Miller, X. Ma, C. Cieri, and G. Doddington, "Multiple-translation arabic (MTA) part 1," Linguistic Data Consortium (LDC) catalog number LDC2003T18 and ISBN 1-58563-276-7.
- [18] 工藤 拓, 賀沢秀人, Web 日本語 N グラム 第 1 版, 言語資源協会発行, 2007.

(平成 23 年 3 月 31 日受付, 8 月 8 日再受付)



宮部 真衣

2006 和歌山大・システム工・デザイン情報中退。2008 同大学院システム工学研究科システム工学専攻博士前期課程了。2011 同大学院システム工学研究科システム工学専攻博士後期課程了。博士(工学)。現在、東京大学知の構造化センター特任研究員。多言語間コミュニケーション支援に関する研究に従事。



吉野 孝 (正員)

1992 鹿児島大・工・電子卒。1994 同大学院工学研究科電気工学専攻修士課程了。現在、和歌山大学システム工学部デザイン情報学科学准教授。博士(情報科学)。コミュニケーション支援の研究に従事。