

機械翻訳を介したチャットコミュニケーションにおける 精度判定に基づく送信拒否の適用可能性

宮部 真衣^{†1} 吉野 孝^{†2,†3}

機械翻訳を介したコミュニケーションにおいて、翻訳リペアは重要な役割を果たす。しかし、翻訳リペアは、ユーザが文章の精度が低いと判断した段階で、初めて行われる作業である。そのためユーザによる精度の不正確な判定により、十分にリペアがされていない状態での修正終了などが発生しており、ユーザの不正確判定を減少させるための仕組みが必要である。不正確判定を減少させる仕組みは、理解可能なメッセージのみのやりとりを実現できる。一方で、この仕組みはユーザの行った判定を否定し、修正するよう促すものであり、特に即時的な対応が求められるリアルタイムコミュニケーションへの適用において、様々な問題が発生する可能性が高い。そのため、不正確判定を減少させるための仕組みの適用による影響を検証する必要がある。不正確判定を減少させるための仕組みとして、精度判定を行った結果、低精度であると判断されたメッセージについては対話相手への送信を拒否し、入力者にメッセージの修正を促すシステムを想定する。システムが実現された場合のチャットコミュニケーションへの影響を検証するために、Wizard of Oz (WOZ) 法によるチャット実験を行った。実験の結果、以下の知見を得た。(1) 精度判定に基づいた送信拒否の仕組みを適用しない場合、対話自体を進行させることはできるが、送信されたメッセージの精度が低く、意思疎通が成立しない可能性がある。精度判定に基づいた送信拒否の仕組みを適用する場合、より高精度な対話を行うことができ、意思疎通が成立する可能性が高くなる。また、仕組みを適用しない場合と適用した場合とを比較して、対話の円滑性に大きな差異があるとはいえない。そのため、正確性の求められるコミュニケーションにおいては、低精度なメッセージ送信の防止が適用できる可能性がある。(2) 精度判定の結果が悪いものについて、単に送信を拒否し、修正するよう促すだけでは、ユーザの判断を否定することになり、またどの部分をどのように修正すべきかをユーザが判断できない場合もあるため、ユーザにとって不快感が高かった。そのため、低精度なメッセージ送信の防止を単にチャットコミュニケーションへ適用することは難しい。

Availability of Accuracy Judgment of Translation Repair Result in Real-time Remote Text-based Communication via Machine Translation

MAI MIYABE^{†1} and TAKASHI YOSHINO^{†2,†3}

In multilingual communication using machine translation, translation repair based on back translation plays an important role. If the users are unable to judge that a translated sentence is inaccurate, they do not undertake translation repair. Many inaccurate sentences are not repaired efficiently because of such inaccurate judgments by users. To avoid this inaccuracy of judgments, it is necessary to develop a method that measures the similarity between an input sentence and the back-translated sentence. This method can ensure that only accurate messages are exchanged between the users. However, some problems may occur with the use of this method, since it rejects the user's judgment. It is important to verify the efficiency of this method. We proposed a method that reduces the inaccuracy of judgment. If the method judges a message as inaccurate, the chat server does not send the message to the receiver and encourages the sender to rectify the sentence. We verified the efficiency of the proposed method in a chat communication using the Wizard of Oz method. The following conclusions can be drawn from the results of the experiments: (1) In this experiment, the chat communication progressed even if a user occasionally sent an inaccurate message. However, inaccurate messages disrupt communication. If our proposed method is applied, users can communicate with each other more accurately. Moreover, there were not any significant differences between the efficiency of communication with or without the proposed method. Therefore, our method can apply to the communication that is requires accuracy. (2) The proposed method made users uncomfortable because the proposed method denied their judgment and they did not know how they repair sentences. Therefore, it is not easy to apply the proposed method to chat communication simply.

^{†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

^{†3} 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

Language Grid Project, National Institute of Information and Communications Technology

1. はじめに

近年、世界規模のインターネットの普及により、電子メールや掲示板、チャットなどのコミュニケーションツールが広く利用されるようになり、ネットワークを介したコミュニケーションの機会が増加している。また、インターネットの普及にともな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も高まっている。しかし、一般に多言語を十分に習得することは難しく、母語以外の言語を用いて十分なコミュニケーションを行うことは困難であり、相互理解ができない可能性が高い^{1),2)}。そのため、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),4)}。

近年、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。コミュニケーションにおいて、不適切な翻訳箇所を含む文章は話者間の相互理解を困難にし、円滑なコミュニケーションの妨げとなる。したがって、円滑にコミュニケーションを行うためには、不適切な翻訳箇所の少ない文章を作成しなければならない。文章中の不適切な翻訳箇所を減少させるために、入力文章を書き換えていくことを「翻訳リペア」と呼ぶ。

翻訳リペアは、機械翻訳を用いたコミュニケーションにおいて重要な役割を果たすと考えられる⁵⁾。折り返し翻訳^{*1}を用いた翻訳リペアにより得られる翻訳結果の精度検証実験から、翻訳リペアによって翻訳精度が改善できることが確認できている⁶⁾。翻訳リペアは、文章の精度が低いとユーザが判断した段階で初めて行われる作業である。そのため、ユーザの不正確な翻訳精度の判断（不正確判定）により、低精度なメッセージの未修正や、十分にリペアがされていない状態での修正終了などが発生しており、ユーザの「不正確判定」を減少させるための仕組みが必要であると考えられる⁷⁾。

不正確判定を減少させる仕組みでは、ユーザの作成した文の精度判定を行う。もし不正確判定が発生していた場合には、ユーザの判定に問題があることを提示し、修正するよう促す必要がある。この仕組みはユーザの行った判定を否定し、修正するよう促すものであり、特に即時的な対応が求められるリアルタイムコミュニケーションにおいて、ユーザの不快感が高くなり、適用が困難となる可能性がある。そこで本論文では、精度判定の仕組みが実現された場合を想定し、チャットコミュニケーションにおける精度判定の適用による影響の検証を行う。本論文では、気軽なコミュニケーションではなく、事実を正確に伝え、意思疎通

を行う必要のある、正確性の求められるコミュニケーションを支援対象とする。

以下、2章において機械翻訳利用におけるユーザの不正確な翻訳精度判定について述べる。3章では検証実験について述べる。4章で実験結果を示し、5章で実験結果に関する考察を行う。最後に6章でまとめと今後の課題について述べる。

2. 機械翻訳利用におけるユーザの不正確な翻訳精度判定

我々は、これまでに行った翻訳リペア実験において、精度が十分でなく、翻訳リペアすべき文章に対して、被験者は修正不要であると判断する機会があることを明らかにした⁷⁾。本論文では、本来修正すべき文を、被験者が修正不要であると判断した割合を不正確判定率と呼ぶ。我々の実施した実験においては、ユーザの不正確判定率は平均7%、最大23%であった。しかし、この不正確判定率は、ユーザがまったく修正をしなかった文の数であり、翻訳リペアを行ったが、不十分な状態で終えた文の数が含まれていない。そのため、低精度な文の数はさらに多く、不正確判定率はより高い⁷⁾。

コミュニケーションにおいて機械翻訳を適切に利用するためには、低精度なメッセージの送信を防ぐことが必要であると考えられる。しかし、低精度なメッセージ送信の防止は、メッセージの修正をユーザに要求するため、メッセージ作成の長時間化につながる。メッセージ作成が長時間化すると、チャットにおける即座のやりとりが困難になるため、円滑なコミュニケーションを阻害すると考えられる。一方、低精度なメッセージ送信の防止を適用しない場合、即座のやりとりが可能となる。しかし、メッセージの精度が低いため、相手の発言の理解が困難となり、確認のための発言など、より多くの対話が必要になる可能性がある。そのため、正確性を重視し、低精度なメッセージ送信の防止を適用した場合よりも円滑なコミュニケーションができるとはいえない。そこで、低精度なメッセージの用いられた対話の円滑性と、低精度なメッセージ送信を防止した正確性のあるコミュニケーションの円滑性とを比較し、それぞれに大きな差異がなく、低精度なメッセージ送信の防止が適用可能であることを示す必要がある。

低精度なメッセージ送信を防ぐためには、入力メッセージと折り返し翻訳文が同じ意味を持っているかどうかを判定する必要がある^{*2}。この文意一致判定を、本論文では「精度判定」

*1 折り返し翻訳とは、他言語への翻訳結果を再度原言語へと翻訳することである。

*2 ユーザは折り返し翻訳を見ながら翻訳リペアを行うため、本論文では精度判定において、対象言語に翻訳された文ではなく、折り返し翻訳文を用いる。

と呼ぶ。精度判定は、類似度^{*1}を利用することにより実現可能であると考えられるが、類語などを考慮した高度な判定の仕組みの実現は容易ではない。そこで、高度な精度判定の仕組みが実現された場合を想定し^{*2}、Wizard of Oz (WOZ) 法²⁰⁾により検証を行う。検証するシステムとしては、理解可能なメッセージのみがやりとりされるチャットコミュニケーションを実現するために、精度判定を行った結果、低精度であると判断されたメッセージについては対話相手への送信を拒否し、入力者にメッセージの修正を促すシステムを想定する。

本研究では、機械翻訳を用いたチャットコミュニケーション実験を行う。正確性の求められるコミュニケーションにおける精度判定に基づいた送信拒否の有用性を示すために、低精度なメッセージが用いられたチャットにおける意思疎通の成立可能性を検証する。また、WOZ法により、チャットコミュニケーションにおいて精度判定に基づいた送信拒否を行い、メッセージ送信拒否の仕組みの適用可能性を検証する。

3. 検証実験

低精度なメッセージが用いられたチャットにおける意思疎通の成立可能性および精度判定に基づいて送信拒否を行った場合の、チャットコミュニケーションへの影響について検証するために、機械翻訳を用いたチャット実験を行った。

実験の被験者は、和歌山大学システム工学部および大学院の学生 16 人である。被験者は男性 9 人、女性 7 人であり、年齢は 19 歳から 23 歳 (平均 21 歳) である。また、チャット習熟度に関するアンケートを行った結果を表 1 に示す。チャットを行うペアに関しては、2 人をランダムに選定して決定した。実験の統制をしやすくするために、チャットペアの 1 人を固定して実験を行う方法も考えられるが、本実験の場合、1 人の被験者が複数回実験に参加してしまうと、図形に関する理解しやすい説明文を容易に作成できてしまい、実験結果に影響すると考えられる。そのため、今回は 1 人の被験者は 1 回のみ実験に参加することとした。

*1 これまでに、翻訳システムの評価のために、BLEU⁸⁾ や NIST⁹⁾ など、様々な翻訳自動評価手法の提案や評価がなされている^{10)–12)}。これらの評価手法は基本的には対象言語への翻訳結果とバイリンガルが作成した対象言語の参照訳とを比較するものであるが、折り返し翻訳結果と原文の比較による手法も提案されている¹³⁾。本研究では折り返し翻訳を用いているため、後者の手法を利用して類似度計算を行うことが可能である。

*2 類似度を測定するための要素技術である形態素解析や構文解析などの精度は、非常に高くなっており^{14),15)}、今後も精度向上が期待されている¹⁶⁾。現在、日本語の言い換え技術や単語間類似度計算の研究も進められており^{17)–19)}、高精度な言語処理技術を組み合わせることにより、将来高度な精度判定の仕組みを実現できると考えられる。

表 1 被験者のチャット習熟度に関するアンケート

Table 1 Results of questionnaire for users on ease of chatting.

| 質問 | 強く同意 しない (人) | 同意しない (人) | どちらとも いえない (人) | 同意する (人) | 強く同意 する (人) |
|--------------------------|--------------------|--------------|----------------------|-------------|-------------------|
| 私はチャットでのコミュニケーションに慣れている。 | 4 | 1 | 4 | 3 | 4 |

なお、今回のチャット実験は日本人同士で行うが、外国人が参加した場合を想定し、入力した原文については入力者本人のみ見ることができるようにし、相手にはその折り返し翻訳文のみが送信されるようにした。

3.1 検証内容

本実験では、以下の項目を明らかにする。

[検証項目 1]: 低精度なメッセージを使った場合でも意思疎通は成立するか?

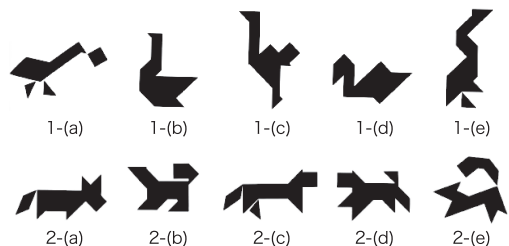
[検証項目 2]: 精度判定に基づく低精度なメッセージの送信拒否の仕組みは、チャットコミュニケーションにおいて適用可能か?

これらの項目を検証するために、「折り返し翻訳を提示せず、精度の正確さが保証されない状態でのチャット」および「折り返し翻訳を提示し、また精度判定による低精度なメッセージの送信拒否を適用したチャット」の 2 種類の実験を実施した。

3.2 実験タスク

短文の場合、機械翻訳によって比較的高精度な翻訳を取得することができる。精度判定による不正確判定の防止は、低精度なメッセージ送信が発生する場合に行う必要がある。そのため、精度判定による不正確判定の防止の影響を検証するためには、ある程度の長さを持ち、翻訳精度が低くなる可能性のある文の入力を被験者に課する必要がある。そこで本実験では、ユーザがある程度の長さを持った文を入力するようにするために、チャットにおいてタスクを設定することとした。また、折り返し翻訳を取得する際の間訳言語に関しては、比較的精度の低い中国語翻訳を用いることとした。

本論文では、事実を正確に伝え、意思疎通を行う必要のある、正確性の求められるコミュニケーションを支援対象として、検証項目について明らかにする。本実験では、対話のためのタスクとして、参照内容の同定 (referential communication task²¹⁾) を用いた。参照内容の同定は、相手に正確に内容を伝える必要のあるタスクである。また、意図することが伝わっているかどうかを、それぞれの被験者が番号を一致させることができたかという客



これらの図形は、Elffers らの Tangram²²⁾ から類似した図形をそれぞれ 10 種類選択し、利用した。

図 1 実験で用いた図形の例

Fig. 1 Sample figures for experimental tasks.

表 2 図形マッチングのための会話例

Table 2 Example of conversation used for figure matching.

| |
|--------------------------------------|
| 参加者 a: 1 つ目の図形は、羽を大きく広げている鳥のように見えます。 |
| 参加者 b: こちらでは 2 番の図形が鳥のように見えます。 |
| 参加者 a: 分かりました。 |

観的データから判断可能である。そのため、本論文の支援対象となる、意思疎通の必要な、正確性の求められるコミュニケーションを検証する手段として、参照内容の同定をタスクとして用いることとした。また、タスクの内容は「図形のマッチング」とした。このタスクにおいては、10 種類の図形を異なる順序で並べ、各被験者に提示する。実験で用いた図形の一部を図 1 に示す。図 1 の 1-(a) から 1-(e) は 1 回目のタスクにおいて利用し、2-(a) から 2-(e) は 2 回目のタスクで利用した図形の一部である。図形のマッチングをする際の対話例を表 2 に示す。被験者は、表 2 のようにお互いの図形について説明し合うことにより、自分の図形と一致する相手の番号を特定する。

3.3 実験内容

本実験では、以下の 2 種類の実験を行う。

- (A) 折り返し翻訳提示 なし
自分の入力した文がどのように翻訳されたのか確認できない状態でチャットを行う。
- (B) 折り返し翻訳提示 あり (精度判定に基づく送信拒否あり)
文を入力すると、自動的に折り返し翻訳を行い、どのような文が相手に送信されるのかを確認可能な状態でチャットを行う。また、精度の保たれた状態を維持するために、入力文と折り返し翻訳文の比較を行い、精度が十分でない場合には、翻訳リペアを行

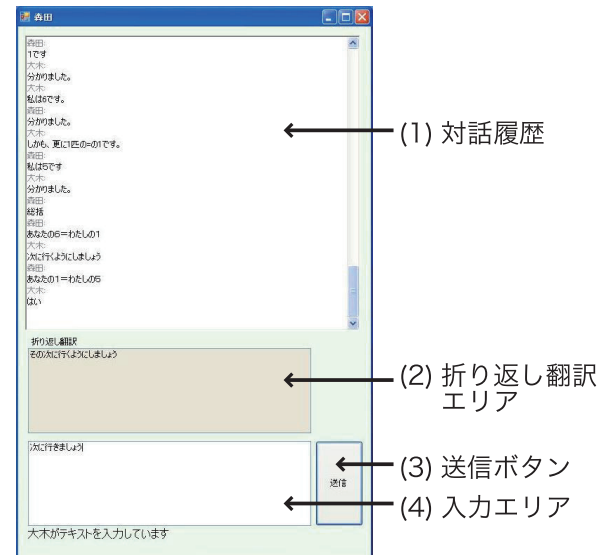


図 2 実験用ツール

Fig. 2 Screenshot of experimental tool.

うようにシステム上で指示する。

1 回目に「(A) 折り返し翻訳提示なし」、2 回目に「(B) 折り返し翻訳提示あり」で行った組 (グループ 1) と、逆の順序で行った組 (グループ 2) は各 4 組である。

本実験のために、テキストベースのチャットツールを開発した。開発したツールの画面を図 2 に示す。翻訳システムは、言語グリッド²³⁾ を介して高電社の J-Server²⁴⁾ を使用した。入力エリア (図 2 (4)) に入力した文の折り返し翻訳文が、折り返し翻訳エリア (図 2 (2)) に表示される。入力エリアの内容が書き換えられると、自動的に折り返し翻訳が行われ、折り返し翻訳エリアの内容も更新される。実験 (A) では、折り返し翻訳エリアは非表示とした。

チャットの参加者数は 2 人とし、1 対 1 での対話とした。被験者には、「開始してください」という指示がツールに表示された後、対話を始めるように指示した。実験に関して被験者に指示した内容を以下に示す。

- (1) 機械翻訳を介したチャットを利用し、タスクを行うことを説明
- (2) 機械翻訳および折り返し翻訳について説明

- (3) タスクの内容を説明
 - (4) 翻訳リペアの概要を説明
 - (5) 精度判定に基づく送信拒否が実験 (B) において発生する可能性があることを説明
- 実験では、チャット中の画面および被験者の様子を記録するために、ビデオと画面キャプチャを利用した。

3.4 送信メッセージの精度判定と送信拒否

本実験では、送信メッセージの精度判定を人間 (Wizard) が行う。精度判定作業は、単純に 2 文を比較するのみであるため、今回の実験においては、Wizard は日本語を理解することができればよい。実験においては、著者らの 1 人が Wizard として精度判定を行った。Wizard による評価に問題がなかったかどうかに関しては、4.1 節において述べる。

精度判定には、Walker らの適合性評価 (5 段階評価)²⁵⁾ を用いた。評価基準を以下に示す。

- 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 (まったく違う意味)

評価値が 3 未満の場合、折り返し翻訳文から原文の意味をつかむことが困難であると考えられる。そこで本実験では、評価値が 3 未満のメッセージについては、送信拒否を行うこととした。

精度判定の流れを図 3 に示す。実験においては、被験者がメッセージを送信後、Wizard は入力文とその折り返し翻訳文を比較し、上記の評価基準に基づいて折り返し翻訳の精度を判断する。精度評価値が 3 以上であると判断した場合は、メッセージを送信する。精度評価

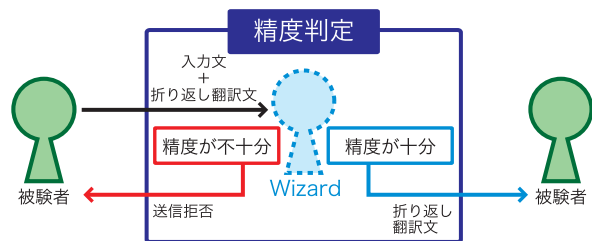


図 3 精度判定の流れ

Fig. 3 Procedure for accuracy judgment by Wizard.

値が 3 未満であると判断した場合は、メッセージが低精度であると見なし、メッセージを送信せず、精度が十分でないため、修正が必要であることを送信者に対してシステム上で提示した。

この精度判定は即座に行う必要があるため、Wizard による精度判定結果には若干の変動がある。なお、実験の際、精度判定を Wizard が行っていることは伝えず、システムの自動判定によるものであると被験者に説明している。

4. 実験結果

4.1 メッセージの精度と送信拒否数

実験後、対話において送信されたメッセージの精度の主観評価を Walker らの適合性評価²⁵⁾ を用いて行った。評価者は、和歌山大学大学院の学生 3 人である。

精度評価の結果を表 3 に示す。また、実験 (B) における、送信拒否の発生数を表 4 に示す。なお、表 4 における「発言数」は、表 3 における実験 (B) のデータと同一のものである。

表 3 送信されたメッセージの精度
Table 3 Accuracy of messages in each chat.

| | | 実験 (A) | 実験 (B) |
|------|-----------|---------------|---------------|
| 精度評価 | 1 以上 2 未満 | 78 文 (11.6%) | 20 文 (3.3%) |
| | 2 以上 3 未満 | 75 文 (11.1%) | 49 文 (8.0%) |
| | 3 以上 4 未満 | 106 文 (15.7%) | 87 文 (14.2%) |
| | 4 以上 5 未満 | 114 文 (16.9%) | 152 文 (24.8%) |
| | 5 | 302 文 (44.8%) | 306 文 (49.8%) |
| 総数 | | 675 文 | 614 文 |

実験 (A): 折り返し翻訳提示なし

実験 (B): 折り返し翻訳提示あり

表 4 送信拒否発生数

Table 4 Number of occurrences of sending-rejection.

| | | 発言数 | 拒否発生数 | 拒否失敗数 | 不当拒否数 |
|------|-----------|-------|-------|-------|-------|
| 精度評価 | 1 以上 2 未満 | 20 文 | 14 文 | 6 文 | - |
| | 2 以上 3 未満 | 49 文 | 12 文 | 37 文 | - |
| | 3 以上 4 未満 | 87 文 | 3 文 | - | 3 文 |
| | 4 以上 5 未満 | 152 文 | 0 文 | - | 0 文 |
| | 5 | 306 文 | 0 文 | - | 0 文 |
| 総数 | | 614 文 | 29 文 | 43 文 | 3 文 |

今回の実験では、Wizard が精度評価値が 3 未満であると判断した場合は、メッセージを送信拒否することとした。Wizard によって行われた送信拒否については、拒否失敗（精度評価値が 3 未満であるにもかかわらず、送信拒否されていない）および不当拒否（精度評価値が 3 以上であるにもかかわらず、送信拒否される）が発生していた可能性がある。表 3 および表 4 より、拒否失敗数は 43 文（1 以上 2 未満の文が 6 文、2 以上 3 未満の文が 37 文）であった。また、不当拒否数は 3 文であった。

実験の結果、拒否失敗数が拒否発生数よりも多かった。実験においては、1 人により精度判定を行っている。一方、実験後の評価は 3 人による平均であり、各評価者の評価値には差異がある。3 人の評価者のうち、1 人でも 3 以上の評価を付けた文については、3 以上の評価を付けた評価者が Wizard となった場合、送信は拒否されない。そこで、評価者 3 人全員が 3 未満の評価を付けた文の拒否失敗数について確認を行った。表 4 中の拒否失敗数 43 文のうち、評価者 3 人全員が 3 未満の評価を付けた文は、14 文（1 以上 2 未満の文が 6 文、2 以上 3 未満の文が 8 文）であった。つまり、拒否失敗数 43 文のうち、29 文に関しては、評価者によっては送信拒否されない可能性があった。また、評価者による評価結果（表 3）は、文脈を考慮しておらず、1 文単位で評価を行っている。一方、Wizard は実際に送信されてくる対話の履歴を見ているため、文脈を考慮した判定になっていた場合がある。そのため、それまでに行われていた対話の文脈から「理解可能である」と Wizard が判断した文の精度が、1 文単位の評価では低くなり、拒否失敗に該当した可能性がある。また、Wizard は即座に精度判定を行う必要があり、短時間で判断したために拒否失敗した可能性がある。今回の実験において、Wizard は一部の文に関して精度判定に失敗しているものの、多くの文の判定において、大きな問題はなかったと考えられる。

4.2 発言回数

実験タスクにおける発言回数を表 5 に示す。表 5 の実験 (A)、実験 (B) において、実施段階が各グループで異なっているのは、順序効果を考慮して実験を実施したためである。表 5 より、実験 (A) の平均発言回数は 83.4 回、実験 (B) の平均発言回数は 77.5 回であった。しかし、実験 (A) と実験 (B) の違いにおける有意確率^{*1}は、0.160 となっており、有意差は見られなかった。したがって、今回の実験において、精度判定に基づく送信拒否を行うことによる発言回数への大きな影響はなかったと考えられる。

表 5 発言回数

Table 5 Number of messages in the experiment.

| 対話グループ | 実験 (A) | | 実験 (B) | |
|-----------|-------------|----------|-------------|----------|
| | 実験の実施段階 | 発言回数 (回) | 実験の実施段階 | 発言回数 (回) |
| A | 1 回目 | 69 | 2 回目 | 44 |
| B | 2 回目 | 207 | 1 回目 | 159 |
| C | 1 回目 | 59 | 2 回目 | 56 |
| D | 2 回目 | 69 | 1 回目 | 59 |
| E | 1 回目 | 70 | 2 回目 | 58 |
| F | 2 回目 | 63 | 1 回目 | 113 |
| G | 1 回目 | 117 | 2 回目 | 105 |
| H | 2 回目 | 29 | 1 回目 | 26 |
| 平均 (標準偏差) | 83.4 (54.7) | | 77.5 (44.1) | |
| 有意確率 | 0.160 | | | |

実験 (A): 折り返し翻訳提示なし

実験 (B): 折り返し翻訳提示あり

4.3 応答時間

チャットでは、メッセージの入力が終了しても、リターンキーを押さなければメッセージは送信されない。メッセージの入力完了後、ユーザがすぐにメッセージを送信するとは限らず²⁶⁾、応答メッセージの作成および送信時のユーザの行動は多様であると考えられる。そこで本論文では、リターンキーを押した時点で被験者の応答を行うという意思決定がなされたと思われ、被験者が送信した最新メッセージと、1 つ前のメッセージ（送信者は問わない）との送信間隔を応答時間とする。

対話の応答時間を表 6 に示す。表 6 より、「1 回目」かつ「実験 (B)」における応答時間が最も長い。しかし、実験 (A) と実験 (B) の違いにおける有意確率^{*2}は、1 回目のタスクにおいて 0.248、2 回目のタスクにおいて 0.564 となっており、有意差は見られなかった。

したがって、今回の実験においては、精度判定に基づく送信拒否を行うことによる応答時間への大きな影響はなかったと考えられる。

4.4 タスク達成時間

実験タスクの達成時間を表 7 に示す。表 7 より、「1 回目」かつ「実験 (B)」において最も時間がかかっている。しかし、実験 (A) と実験 (B) の違いにおける有意確率^{*3}は、1 回目

*1 ウィルコクソンの符号付順位検定を用いた。

*2 マン・ホイットニーの U 検定を用いた。

*3 マン・ホイットニーの U 検定を用いた。

表 6 応答時間

Table 6 Response time in the experiment.

| | | 実験 (A) | | 実験 (B) | | 有意確率 |
|---------|------|------------------|------------------|------------------|------------------|-------|
| | | 平均 (標準偏差) (秒) | 平均 (標準偏差) (秒) | 平均 (標準偏差) (秒) | 平均 (標準偏差) (秒) | |
| 実験の実施段階 | 1 回目 | 27 (3) | 55 (36) | | | 0.248 |
| | 2 回目 | 29 (19) | 28 (8) | | | 0.564 |
| 有意確率 | | 0.386 | 0.149 | | | |

実験 (A) : 折り返し翻訳提示なし

実験 (B) : 折り返し翻訳提示あり

表 7 タスク達成時間

Table 7 Time required to complete the experimental task.

| | | 実験 (A) | | 実験 (B) | | 有意確率 |
|---------|------|--------------------|--------------------|--------------------|--------------------|-------|
| | | 平均 (標準偏差) (分:秒) | 平均 (標準偏差) (分:秒) | 平均 (標準偏差) (分:秒) | 平均 (標準偏差) (分:秒) | |
| 実験の実施段階 | 1 回目 | 36:19 (16:48) | 53:36 (23:54) | | | 0.149 |
| | 2 回目 | 31:20 (11:46) | 29:46 (19:34) | | | 0.386 |
| 有意確率 | | 0.564 | 0.149 | | | |

実験 (A) : 折り返し翻訳提示なし

実験 (B) : 折り返し翻訳提示あり

のタスクにおいて 0.149, 2 回目のタスクにおいて 0.386 となっており, 有意差は見られなかった。また, 実施回数における有意確率に関しても, 実験 (A) において 0.564, 実験 (B) において 0.149 となっており, 有意差は見られない。一方, 実験 (A) と実験 (B) の違いにかかわらず, 全被験者の 1 回目および 2 回目のタスク達成時間を見ると, 平均はそれぞれ 44 分 57 秒 (標準偏差 21 分 14 秒), 30 分 33 秒であり (標準偏差 14 分 58 秒), 有意確率^{*1}は 0.017 となっており, 5%水準で有意差が見られた。タスクへの慣れにより, 2 回目のタスク達成時間が 1 回目よりも短時間になった可能性がある。

4.5 理解不能状況の発生回数

チャットコミュニケーションにおいて, 低精度なメッセージ利用によって発生する問題としては, メッセージの意味が相手に伝わらない状況の発生が考えられる。本論文では, この状況を「理解不能状況」と呼ぶこととする。

表 8 理解不能状況の発生回数

Table 8 Number of instances of incomprehension.

| 対話グループ | 実験 (A) | | 実験 (B) | |
|--------|---------|--------------------|---------|--------------------|
| | 実験の実施段階 | 理解不能状況 発生回数 (回) | 実験の実施段階 | 理解不能状況 発生回数 (回) |
| A | 1 回目 | 2 | 2 回目 | 0 |
| B | 2 回目 | 3 | 1 回目 | 0 |
| C | 1 回目 | 2 | 2 回目 | 0 |
| D | 2 回目 | 3 | 1 回目 | 0 |
| E | 1 回目 | 0 | 2 回目 | 0 |
| F | 2 回目 | 3 | 1 回目 | 1 |
| G | 1 回目 | 1 | 2 回目 | 1 |
| H | 2 回目 | 1 | 1 回目 | 0 |
| 合計 | | 15 | | 2 |

実験 (A) : 折り返し翻訳提示なし

実験 (B) : 折り返し翻訳提示あり

今回の実験対話において, 被験者が「分からない」などの発言をする状況が見られた。メッセージを理解できないことを意図する発言をした状況を理解不能状況と判断することとし, 理解不能状況の発生回数の確認を行った。

表 8 に理解不能状況の発生回数を示す。表 8 の実験 (A), 実験 (B) において, 実施段階が各グループで異なっているのは, 順序効果を考慮して実験を実施したためである。実験 (A) における理解不能状況発生回数は 15 回, 実験 (B) における理解不能状況発生回数は 2 回であった。理解不能状況が発生する場合, 原因となった発言の翻訳精度が低いと考えられる。そこで, 理解不能状況に陥った被験者 (被験者 x) の「分からない」などの発言と, その前に行われた被験者 x の発言の間に送信された対話相手 (被験者 y) の発言を理解不能状況の発生原因発言群^{*2} (図 4) とし, 理解不能状況の発生原因発言群の翻訳精度最小値^{*3}について確認を行った。発生原因発言群に含まれる発言数が 1 つであったものは 17 回中 11 回, 2 つ以上であったものは 17 回中 6 回であった。なお, 発生原因発言群に含まれる発言が複数の場合, 発言群に理解不能状況と関係しない発言が含まれることがあるが, 発生原因発言群の翻訳精度最小値を確認しているため, 問題はないと考えられる。確認の結果, 実験

*2 理解不能状況に陥った被験者 (被験者 x) の「分からない」などの発言と, その前に行われた被験者 x の発言の間に送信された対話相手 (被験者 y) の発言が複数の場合, どの発言が原因か判断することが難しいため, 該当する被験者 y の発言を発生原因発言群として扱う。

*3 理解不能状況の発生は, 翻訳精度の低さが影響している可能性が高いため, 翻訳精度の最小値により議論する。

*1 ウィルコクソンの符号付順位検定を用いた。

791 機械翻訳を介したチャットコミュニケーションにおける精度判定に基づく送信拒否の適用可能性

| | | | |
|-------|---------------------------------------|--------------|--------------------|
| 被験者x: | わかりました。 | | |
| 被験者y: | 見た中で一番分かりやすい図形が白鳥で4番です。 | (精度評価値: 2.0) | 理解不能状況の 発生原因発言群 |
| 被験者y: | 10番がにわとりらしきです。 | (精度評価値: 2.0) | |
| 被験者y: | 1番は首が長いのでキリンかなと思います。 | (精度評価値: 1.7) | |
| 被験者y: | 3番は足が妙にながいところからフラミンゴかなと思います。 | (精度評価値: 1.3) | |
| 被験者y: | 7番は飛ばうとしているハトのように見えます。 | (精度評価値: 2.7) | |
| 被験者y: | あとはちょっと・・・。 | (精度評価値: 1.7) | |
| 被験者y: | とりえずxさんの図形の中から上記にあたる「っぽい」もの挙げてみてください。 | (精度評価値: 1.3) | |
| 被験者x: | poっていうのは何ですか? (理解不能状況発生) | | |

図 4 理解不能状況の発生原因発言群
Fig. 4 Group of incomprehension messages.

(A) における理解不能状況発生回数 15 回のうち、翻訳精度最小値が 3 未満であったものは 10 回であった。表 3 を見ると、翻訳精度が 3 未満である文の合計は 153 文 (1 以上 2 未満である文が 78 文, 2 以上 3 未満である文が 75 文) となっており、翻訳精度が 3 未満であるにもかかわらず、理解不能状況が発生せず、対話が続けられていた場合が多いことが分かる。実験後に行った記述式アンケートでは、「分かりにくい表現もあったが、予測できた」や「分からないものは無視した」という意見もあった。そのため、理解が困難な場合はあるものの、メッセージの予測や、理解できるものみでの対話により、低精度なメッセージが使われていても、チャットコミュニケーション自体は継続されていたと考えられる。

4.6 誤答数

各タスクにおける被験者の誤答数を表 9 に示す。表 9 の実験 (A)、実験 (B) において、実施段階が各グループで異なっているのは、順序効果を考慮して実験を実施したためである。実験において誤答のあった被験者は 16 人中 6 人であった。誤答のあった被験者 6 人の誤答が発生した実験状況を確認したところ、実験 (A) において誤答のあった被験者は、実験全体において誤答のあった被験者 6 人全員であった。また、実験 (B) において誤答のあった被験者は 6 人中 1 人 (被験者 k) であった。したがって、低精度なメッセージの送信を防止することによって、誤答の発生が少なくなる可能性がある。

4.7 アンケート結果

実験後、アンケート調査を行った。アンケート結果を表 10 に示す。表 10 の評価は、1: 強く同意しない, 2: 同意しない, 3: どちらともいえない, 4: 同意する, 5: 強く同意する, の 5 段階評価の平均値を示す。質問 (1), (2) については被験者全員に回答してもらった。質問 (3) については、実験時に送信拒否が発生した被験者のみ回答するよう依頼した。今回の実験において、送信拒否が発生したのは被験者は 16 人中 11 人であったが、11 人の

表 9 誤答数

Table 9 Number of erroneous answers.

| 対話グループ | 被験者 | 実験 (A) | | 実験 (B) | |
|--------|-----|---------|-----|---------|-----|
| | | 実験の実施段階 | 誤答数 | 実験の実施段階 | 誤答数 |
| A | a | 1 回目 | 0 | 2 回目 | 0 |
| | b | 1 回目 | 2 | 2 回目 | 0 |
| B | c | 2 回目 | 4 | 1 回目 | 0 |
| | d | 2 回目 | 4 | 1 回目 | 0 |
| C | e | 1 回目 | 0 | 2 回目 | 0 |
| | f | 1 回目 | 0 | 2 回目 | 0 |
| D | g | 2 回目 | 0 | 1 回目 | 0 |
| | h | 2 回目 | 0 | 1 回目 | 0 |
| E | i | 1 回目 | 0 | 2 回目 | 0 |
| | j | 1 回目 | 0 | 2 回目 | 0 |
| F | k | 2 回目 | 2 | 1 回目 | 2 |
| | l | 2 回目 | 2 | 1 回目 | 0 |
| G | m | 1 回目 | 2 | 2 回目 | 0 |
| | n | 1 回目 | 0 | 2 回目 | 0 |
| H | o | 2 回目 | 0 | 1 回目 | 0 |
| | p | 2 回目 | 0 | 1 回目 | 0 |

実験 (A): 折り返し翻訳提示なし

実験 (B): 折り返し翻訳提示あり

表 10 アンケート結果

Table 10 Results of questionnaire.

| 質問内容 | | 全体 | グループ 1 | グループ 2 | 有意確率 |
|-----------------------------------|------|-----|--------|--------|-------|
| (1) 折り返し翻訳が表示されていたとき、折り返し翻訳を確認した。 | 平均 | 4.6 | 4.6 | 4.5 | 0.971 |
| | 標準偏差 | 0.8 | 0.7 | 1.0 | |
| (2) 折り返し翻訳の精度が悪い場合、入力文を修正した。 | 平均 | 4.7 | 4.7 | 4.6 | 0.739 |
| | 標準偏差 | 0.5 | 0.5 | 0.5 | |
| (3) 折り返し翻訳が送信できなかったとき、不快だった。 | 平均 | 4.1 | 4.3 | 4.0 | 0.180 |
| | 標準偏差 | 0.5 | 0.5 | 0.6 | |

グループ 1: 1 回目に実験 (A), 2 回目に実験 (B) を実施

グループ 2: 1 回目に実験 (B), 2 回目に実験 (A) を実施

5 段階評価の評価値: 1: 強く同意しない, 2: 同意しない, 3: どちらともいえない, 4: 同意する, 5: 強く同意する

うち 1 人のみ未記入であったため、10 人の回答結果を示している（各グループの回答者は 5 人である）。

表 10 における有意確率^{*1}は、グループ 1 とグループ 2 の評価の差に関するものである。表 10 より、どの質問項目についても、グループ間の有意差は見られなかった。

質問 (1) および質問 (2) の結果より、折り返し翻訳が提示されている場合、被験者は翻訳結果を確認しており、また精度が悪いと判断した場合には、入力文の修正を行っていたと考えられる。質問 (3) より、精度判定に基づいた送信拒否については、全体の平均評価値が 4.1 となっており、不快であると感じた被験者が多かった。送信拒否の適用可能性については、5 章で議論する。

5. 考察

本実験では、チャットを行うペアに関しては、被験者 2 人をランダムに選定して決定した。被験者のチャットの習熟度がコミュニケーションの進行などに影響する可能性があるが、記録した実験ログを確認したところ、すべての対話グループにおいて、最後までタスクを行うことができおり、コミュニケーションの成立の阻害などは見られなかった。そのため、今回取得したランダムペアのデータにより議論を進めることとする。

5.1 低精度なメッセージが用いられたチャットにおける意思疎通の成立可能性

本節では、[検証項目 1]: 低精度なメッセージを使った場合でも意思疎通は成立するか？について議論する。

4.5 節において、理解が困難な場合はあるものの、メッセージの予測や、理解できるもののみでの対話により、低精度なメッセージが使われていても、チャットコミュニケーション自体は継続されていたことを示した。また、4.6 節において、16 人中 6 人の被験者が実験 (A) において誤答しており、実験 (B) と比較して誤答が多かったことを示した。

誤答した図形に関する対話ログの確認を行ったところ、対話ログより、以下の点が誤答に影響した可能性があることが分かった。

(1) 翻訳誤り

被験者 c, d (対話グループ B) の誤答内容および対話ログを確認した結果、表 11 のような対話を行っていた。被験者 c は「飛んでいる・飛んでいない」という発言を行った後、図形番号を「8・6」のように並べた発言を行い、被験者 d も同様の形式で

表 11 翻訳誤りによる誤答発生時の会話例

Table 11 Example conversation in erroneous answer due to translation mistake.

| 被験者 c の対話履歴 | 被験者 d の対話履歴 |
|-----------------|----------------|
| c: 飛んでいる・飛んでいない | c: 飛んでいない・飛びます |
| d: はい | d: はい |
| d: 8・6 です | d: 8・6 です |
| c: 2・5 です | c: 2・5 です |

下線の引かれたメッセージが入力文、下線のないメッセージが折り返し翻訳文である。

図形番号を発言していた。すなわち、両被験者は類似した 2 つの図形に関して、「飛んでいるように見える図形」と「飛んでいないように見える図形」を「・」で区切って発言することにより、図形の一致を試みていた。しかし、被験者 c の発言した「飛んでいる・飛んでいない」の折り返し翻訳結果は「飛んでいない・飛びます」となっていた。そのため、被験者 c は「飛んでいるように見える図形・飛んでいないように見える図形」の順序で伝えていたが、被験者 d は「飛んでいないように見える図形・飛んでいるように見える図形」の順で伝えており、両者に誤答が発生した。この誤答は、折り返し翻訳が表示されていないため相手に伝わっている内容が確認できず、翻訳誤りを改善できないために発生したと考えられる。

(2) 被験者による表現の解釈の違い

被験者 k, l (対話グループ F) の誤答内容および対話ログを確認した結果、翻訳誤りなどは発生していなかった。この誤答については、類似した図形の特定において、被験者 l が説明を行ったが、被験者 k は被験者 l が意図した図形とは異なる図形についての説明だと解釈し、間違った図形番号を発言したために発生していた。

(3) 不十分な意思疎通

被験者 b および被験者 m については、対話グループのうち 1 人のみが図形の特定に失敗している。対話ログを確認した結果、対話相手は不明な点の再確認を行い、番号の修正を行っているが、回答した被験者は自分の結果を修正していないなど、対話グループ内で十分に意思疎通ができていなかった可能性がある。

(2), (3) については、被験者に依存する問題である。一方、(1) については、翻訳精度の影響による問題であり、今回の実験においては、実験 (B) においては発生していなかった。したがって、低精度なメッセージが用いられる場合、誤解が発生した状態に気付かないままチャットコミュニケーション自体が継続してしまい、最終的に意思疎通が成立しない可能

*1 マン・ホイットニーの U 検定を用いた。

性があると考えられる。一方、精度判定に基づいた送信拒否を行うことにより、誤解の発生を防ぐことができている。正確性の求められるコミュニケーションにおいては、精度判定に基づいた送信拒否が有効に働く可能性が高い。

5.2 低精度なメッセージの送信拒否の仕組みの適用可能性

本節では、[検証項目 2]: 精度判定に基づく低精度なメッセージの送信拒否の仕組みは、チャットコミュニケーションにおいて適用可能か? について議論する。

4.3 節および 4.4 節において、1 回目の実験では、実験 (B) の方が平均応答時間や平均タスク達成時間が長いものの、1 回目、2 回目のどちらの実験に関しても、実験 (A) と実験 (B) において有意差がないことを示した。また、4.2 節において、発言回数には実験 (A) と実験 (B) において有意差がないことを示した。以上のことから、低精度なメッセージの用いられた対話と比較して、低精度なメッセージ送信の防止を適用した、正確性のあるコミュニケーションの円滑性に大きな差異があるとはいえない。そのため、正確性の求められるコミュニケーションにおいては、低精度なメッセージ送信の防止が適用できる可能性があると考えられる。

しかし、精度判定に基づいた送信拒否が発生した際、不快に感じたかどうかのアンケート (表 10 質問 (3)) では、送信拒否は不快であると回答した被験者が多かった。質問 (3) における回答の選択理由としては、「時間がかかる」「何度修正してもなかなか伝えたい文章にならない」「相手が十分理解できるレベルだと判断したものが送信できないのは嫌だった」といったものがあげられていた。また、「どの部分が悪いのか、修正のヒントが欲しい」という意見もあり、精度判定の結果が悪いものについて、単に送信を拒否し、修正するよう促すだけではユーザにとって不快感が高く、チャットコミュニケーションへの適用は難しいと考えられる。

5.1 節において、低精度なメッセージが用いられる場合、誤解が発生した状態に気付かないままチャットコミュニケーション自体が継続してしまい、最終的に意思疎通が成立しない可能性があることを示した。一方、精度判定に基づいた送信拒否を用いた場合、1 人を除いて誤答は発生しておらず、送信拒否の仕組みによって、意思疎通が可能となった可能性がある。したがって、正確性の求められるコミュニケーションにおいては、より高精度な対話を行うことのできる、低精度なメッセージ送信の防止が適用できる可能性があるが、今回用いた仕組みを単純に適用することは難しいと考えられる。今後、精度判定に基づく送信拒否を適用できるように、ユーザの不快感を軽減する方法について検討する必要がある。

5.3 低精度なメッセージの判定基準の妥当性

今回の実験では、精度評価に Walker らの適合性評価を用い、送信拒否の基準となる精度評価値を 3 とした。Wizard が精度評価値が 3 未満であると判断した場合、メッセージが低精度であると見なし、メッセージを送信拒否することとした。しかし、5.2 節で述べたように、今回の実験では、送信拒否はユーザにとって不快感が高く、また「相手が十分理解できるレベルだと判断したものが送信できないのは嫌だった」という意見もあげられた。

Wizard は送信されたメッセージとその折り返し翻訳文を見て精度を判定していたが、チャットに参加している被験者には、チャットのコンテキストが存在するため、文単位での精度が低い場合でも、内容を推測できる可能性がある。

今回の実験 (A) (折り返し翻訳提示なし) において、精度評価値が 3 未満であったが、理解不能状況を引き起こしていない文に関して、精度判定に基づく送信拒否を適用すると、ユーザにとって不当拒否となる可能性がある。一方、5.1 節で述べたように、「分からないものは無視した」という被験者の意見もある。今後、対話のコンテキストを考慮し、メッセージが低精度であると判断する基準となる精度評価値を設定するための分析を行う必要があると考えられる。

5.4 不正確判定発生時の対応に関する方針

低精度なメッセージが用いられる場合、誤解が発生した状態に気付かないままチャットコミュニケーション自体が継続してしまい、最終的に意思疎通が成立しない可能性がある。そのため、正確性の求められるコミュニケーションにおいては、精度判定に基づいた送信拒否が有効に働く可能性が高いことを示した。ユーザの不正確判定により送信される低精度なメッセージは、対話相手が無視する場合もあるため、「分からない」などの発言がされないと、送信者は対話相手にメッセージの内容が伝わっていないことに気付くことができないと考えられる。そのため、対話相手に内容が伝わっていない可能性があることを送信者に気付かせる必要がある。

また、実験の結果、精度判定の結果が悪いものについて、単に送信を拒否し、修正するよう促すだけではユーザにとって不快感が高いことが分かった。精度判定の結果が悪い場合、メッセージを送信したうえで、精度が低いことを送信者に提示することにより、対話相手が理解できない可能性があることを送信者に気付かせるような仕組みが必要であると考えられる。今後は、ユーザに不快感を与えることなく、低精度なメッセージが送信されたことを送信者に気付かせるための仕組みを検討する必要がある。

5.5 実験結果の適用限界

本実験では、本論文の支援対象となる、意思疎通の必要な、正確性の求められるコミュニケーションを検証する手段として、参照内容の同定 (referential communication task) を用いて検証を行った。対話の成立においては、対話の食い違いを防止するために、対話者間で相互理解を構築する必要がある。参照内容の同定過程においては、対話者間での協調が必要であり、参照内容の同定により、対話の食い違いを防止することができる²⁷⁾。

本論文では、事実を正確に伝え、意思疎通を行う必要のある、正確性の求められるコミュニケーションを支援対象としている。本実験では、参照内容の同定過程を被験者間で共有することにより相互理解の構築を試みた。そのため、正確性の求められるコミュニケーションにおける精度判定の適用による影響の検証に関して、実験タスクに問題はなく、本実験で得られた結果は、正確性の求められるコミュニケーションにおいて適用可能である。

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、翻訳リペアは翻訳精度向上のための手法として重要な役割を果たすと考えられる。しかし、翻訳リペアはユーザが修正の必要性を判断したうえで行う作業であるため、ユーザによる精度の不正確な判定により、十分にリペアがされていない状態での修正終了などが発生する可能性がある。そのため、ユーザの不正確判定を減少させるための仕組みが必要であると考えられる。しかし、不正確判定を減少させる仕組みは、ユーザの行った判定を否定し、修正するよう促すものである。そのため、特に即時的な対応が求められるリアルタイムコミュニケーションへの適用において、様々な問題が発生する可能性が高い。

本論文では、精度判定の仕組みが実現された場合を想定し、チャットコミュニケーションにおける精度判定の適用による影響の検証を行った。実験の結果、以下の知見を得た。

- (1) 低精度なメッセージが使われた場合、対話自体を進行させることはできるものの、意思疎通が成立しない可能性がある。精度判定に基づいた送信拒否の仕組みを適用する場合、意思疎通が成立する可能性が高くなる。また、仕組みを適用しない場合と適用した場合とを比較して、対話の円滑性に大きな差異があるとはいえない。そのため、正確性の求められるコミュニケーションにおいては、低精度なメッセージ送信の防止が適用できる可能性がある。
- (2) 精度判定の結果が悪いものについて、単に送信を拒否し、修正するよう促すだけでは、ユーザの判断を否定することになり、またどの部分をどのように修正すべきかを

ユーザが判断できない場合もあるため、ユーザにとって不快感が高い。そのため、低精度なメッセージ送信の防止を単純にチャットコミュニケーションへ適用することは難しい。

今後は、ユーザに不快感を与えることなく、低精度なメッセージが送信されたことを送信者に気付かせるための方法について検討する。また、対話のコンテキストを考慮した、メッセージが低精度であると判断する基準となる精度評価値を設定するための分析を行う。

謝辞 本研究の一部は、日本学術振興会科学研究費基盤研究 (B) (19300036) の補助を受けた。

参考文献

- 1) Aiken, M.: Multilingual Communication in Electronic Meetings, *ACM SIG-GROUP, Bulletin*, Vol.23, No.1, pp.18-19 (2002).
- 2) Tung, L.L. and Quaddus, M.A.: 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) 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).
- 4) Inaba, R.: Usability of Multilingual Communication Tools, *Proceedings, Lecture Notes in Computer Science 4560*, pp.91-97 (2007).
- 5) 安岡美佳, 中小路久美代, 大平雅雄, 石田 亨, 野村早恵子: 異文化協調作業における共有理解構築の機会としてのコミュニケーションエラー現象の利用, *情報処理学会研究報告*, 2003-HI-103, pp.47-54 (2003).
- 6) 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*, Vol.J-90-D-I, No.12, pp.3142-3150 (2007).
- 7) Miyabe, M., Yoshino, T. and Shigenobu, T.: Effects of Repair Support Agent for Accurate Multilingual Communication, *Proceedings, Lecture Notes in Computer Science 5351*, pp.1022-1027 (2008).
- 8) Papineni, K., Roukos, S., Ward, T. and Zhu, W.: BLEU: A Method for Automatic Evaluation of Machine Translation, *Proc. 40th Annual Meeting of the Association for Computational Linguistics (ACL)*, pp.311-318 (2002).
- 9) NIST: Automatic Evaluation of Machine Translation Quality Using N-gram Co-Occurrence Statistics, *Technical Report*, NIST (2002).
- 10) Denoual, E. and Lepage, Y.: 文字単位 BLEU による翻訳自動評価, *言語処理学会第 11 回年次大会, 発表論文集*, pp.522-525 (2005).
- 11) 金山 博, 荻野紫穂: 翻訳精度評価手法 BLEU の日英翻訳への適用, *情報処理学会研究報告*, 2002-NL-154, pp.131-136 (2003).

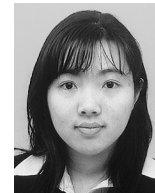
- 12) 秋葉泰弘, 今村賢治, 隅田英一郎, 中岩浩巳, 山本誠一, 奥乃 博: 複数の編集距離を用いた口語翻訳文の自動評価, 人工知能学会論文誌, Vol.20, No.3, pp.139-148 (2006).
- 13) Uchimoto, K., Hayashida, N., Ishida, T. and Isahara, H.: Automatic Rating of Machine Translatability, *10th Machine Translation Summit (MT Summit X)*, pp.235-242 (2005).
- 14) 長尾 真, 黒橋禎夫, 佐藤理史, 池原 悟, 中野 洋: 言語情報処理, pp.112-117, 岩波書店 (1998).
- 15) 黒橋禎夫: 使いやすくなった自然言語処理のフリーソフト—知っておきたいツールの中身: 結構やるな, KNP, 情報処理学会会誌, Vol.41, No.11, pp.1215-1220 (2000).
- 16) 長尾 真: 自然言語処理, コンピュータサイエンス誌 Bit3 月号, 浦山 毅 (編), pp.17-19, 共立出版 (1999).
- 17) 藤田 篤, 降幡建太郎, 乾健太郎, 松本裕治: 語彙概念構造に基づく言い換え生成—機能動詞構文の言い換えを例題に, 情報処理学会論文誌, Vol.47, No.6, pp.1963-1975 (2006).
- 18) 川上高志, 鈴木 寿: 決定リストを利用した単語間の類似度計算法, 情報処理学会研究報告, Vol.2006, No.94, pp.85-90 (2006).
- 19) 相澤彰子: 大規模テキストコーパスを用いた語の類似度計算に関する考察, 情報処理学会論文誌, Vol.49, No.3, pp.1426-1436 (2008).
- 20) Fraser, N.M. and Gilbert, G.N.: Simulating Speech Systems, *Computer Speech and Language*, Vol.5, No.1, pp.81-99 (1991).
- 21) Clark, H.H. and Wilkes-Gibbs, D.: Referring as a collaborative process, *Cognition*, Vol.22, pp.1-39 (1986).
- 22) Elffers, J.: Tangram: 1,600 Ancient Chinese Puzzles, Taschen America Llc (1999).
- 23) Ishida, T.: Language Grid: An Infrastructure for Intercultural Collaboration, *IEEE/IPSJ Symposium on Applications and the Internet (SAINT-06)*, pp.96-100 (2006).
- 24) 高電社 . <http://www.kodensha.jp/>
- 25) Walker, K., Bamba, M., Miller, D., Ma, X., Cieri, C. and Doddington, G.: *Multiple-*

Translation Arabic (MTA) Part 1, Linguistic Data Consortium, Philadelphia (2003).

- 26) 細馬宏通: チャットは何を前提としているか—チャットの時間的構造と音声会話の時間構造, 身体性とコンピュータ, 岡田美智男, 三嶋博之, 佐々木正人 (編), pp.338-349, 共立出版 (2000) .
- 27) Isaacs, E.A. and Clark, H.H.: References in conversations between experts and novices, *Journal of Experimental Psychology: General*, Vol.116, pp.26-37 (1987).

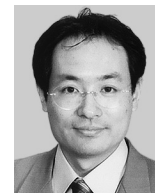
(平成 21 年 5 月 25 日受付)

(平成 21 年 12 月 17 日採録)



宮部 真衣 (学生会員)

昭和 59 年生 . 平成 18 年和歌山大学システム工学部デザイン情報学科中退 . 平成 20 年同大学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修了 . 現在 , 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在学中 . 多言語コミュニケーション支援に関する研究に従事 .



吉野 孝 (正会員)

昭和 44 年生 . 平成 4 年鹿児島大学工学部電子工学科卒業 . 平成 6 年同大学大学院工学研究科電気工学専攻修士課程修了 . 現在 , 和歌山大学システム工学部デザイン情報学科准教授 . 博士 (情報科学) . コラボレーション支援に関する研究に従事 .