

機械翻訳を用いた対話における思い違いに関する分析

山下直美[†] 石田 亨^{††} 平田 圭 二[†]

東アジアを中心に、機械翻訳を用いて議論をするコミュニティが増加している。しかし、機械翻訳を用いた対話には誤訳ノイズが混入するため、ユーザ間の相互理解を構築することは難しい。我々は、機械翻訳を介した対話で大量に発生する思い違いの問題に焦点を当て、機械翻訳を介した対話の特徴と思い違いの関係について分析、考察した。その結果、思い違いが大量に発生する異国ユーザ間の返信メッセージは、内容をとらえずメッセージの部分的なところでつながる傾向が強いことが分かった。そこで、これらの対話の特徴をふまえ、誤訳による思い違いを機械的に検知する手法を提案した。提案手法では、通常のスレッド (syntactic thread) と語彙的結束性に基づいたスレッド (semantic thread) の差異を計測し、この差異が大きければユーザ間の話題に関する思い違いが生じやすいと判断するものである。実際に対話データを用いて提案手法を検証したところ、提案手法によって計測した思い違いと実際の思い違いの間には有意な正相関が観測され、手法の有効性が示された。

Analyzing Misconceptions in Machine-translator Mediated Communication

NAOMI YAMASHITA,[†] TORU ISHIDA^{††} and KEIJI HIRATA[†]

Multilingual communities using machine translation to overcome language barriers are showing up more and more frequently. However, when a large number of translation errors get mixed into conversation, it becomes difficult for users to fully understand each other. In this paper, we focus on misconceptions found in high volume in actual online conversations using machine translation. We first examine how misconceptions occurred by delving into 1106 messages exchanged on BBS using machine translation. The analysis results indicate that when a user responds via machine translation, he/she tends to respond to short phrases of the original message and tends to trip on the wording of the original message. Next, based on the analysis results, we propose a method that automatically measures the likeliness of each dialogue including misconceptions. The proposed method assesses the likeliness of each dialogue including misconceptions by calculating the gaps between the regular discussion thread (syntactic thread) and the discussion thread based on lexical cohesion (semantic thread). Verification results show significant positive correlation between actual misconception frequency and the syntax-semantic gap, which indicates the validity of the method.

1. はじめに

近年、東アジアを中心に、機械翻訳を用いて議論をするインターネットコミュニティを見かけるようになった。国際的な協調作業が年々増加する中、円滑なコミュニケーションを実現する機械翻訳の需要は今後も増加すると考えられている³⁾。

機械翻訳を多言語間コミュニケーションに用いる研究では、ある程度の誤訳を含む機械翻訳でも、対話者は背景知識や文脈などから対話内容を理解し対話が成立するとされている^{1),9),10)}。たとえば、機械翻訳の翻

訳精度が比較的良いとされるスペイン語と英語のような欧米間の言語対を用いた対話実験では、ユーザは機械翻訳を介してもお互いの対話内容をほぼ完璧に理解できたと言っている¹⁾。

しかしどのような言語対でも高い翻訳精度が達成できるとは限らない。高い翻訳精度が達成できない機械翻訳を用いたコミュニケーションでは解決すべき課題が多い。特に、誤訳によって大量に発生する思い違いは、ユーザ間の相互理解の破綻を招き協調作業の進行を阻害する。ここで思い違いとは、相互理解の構築を図る際に相手の発話意図を誤って推測する誤解を指す。

2章で詳しく述べるように、我々が実施した高い翻訳精度が達成できない機械翻訳を用いた大規模な実験では、お互いのメッセージをおおむね理解し思い違いは生じていないと感じていたユーザ同士であっても、実際には相互理解が破綻していた事例がいくつも発見

[†] 日本電信電話株式会社 NTT コミュニケーション基礎科学研究所

NTT Communication Science Laboratories, NTT

^{††} 京都大学大学院情報学研究科

Graduate School of Informatics, Kyoto University

された。このような思い違いが生じやすくかつ気付きにくい理由は、互いの発話内容が相手にどのように伝わっているかの確認が困難なためと考えられる。

機械翻訳を用いた協調作業を支援するうえで、このような思い違いを防止する方法を開発することは重要な課題である。特に、機械翻訳を用いた対話ではユーザが思い違いの存在に気が付きにくいことから、思い違いを自動的に検出し修正できるメカニズムを開発することが重要である⁷⁾。

しかし、従来の思い違いに関する研究では、分析手法の大半にエスノメソドロジや社会言語学のような人手を要する分析手法が用いられており、思い違いを自動的に検出する手法はいまだ存在しない。これは、一般に発生する思い違いの発生要因が、暗黙のうちに共有していると思われる前提条件や背景知識といった対話中で明示的に語られない内容にあり^{2),13),16)}、対話内容だけから思い違いを抽出することが困難なためであろう。一方、前述の低い翻訳精度での実験では、明示的に語られた内容の中に思い違いの主要因があり、従来研究の思い違いとは異なる発生メカニズムが示唆された。

そこで本論文では、思い違いの解消のための第1歩として、高い翻訳精度が達成できない機械翻訳を用いた対話環境における思い違いの定量化手法を提案する。我々は低い翻訳精度がもたらす思い違いに関する先行研究を見いだせなかったため、そのような思い違いに対しまず還元論的に取り組むこととした。すなわち、機械翻訳を介した対話の特徴をマイクロとマクロの2つのレベルで分析し、これらのレベルにおける対話の特徴量と誤訳による思い違いを関連付けることによって思い違いを定量化する。具体的には、まず分析単位を「時間的に近い発話対」とし、これらの発話対の特徴と思い違いの関係について分析、考察する。次に、分析単位を「対話全体(スレッド)」に拡張し、メッセージのヘッダ情報によるスレッドとマイクロな分析によるスレッドの差異を用いて思い違いを定量化する。

本論文の構成は以下のとおりである。2章では、我々が実施した実験において実際に発生した思い違いの例を紹介し、誤訳によりどのような思い違いが発生したかを説明する。3章では、機械翻訳を介した応答関係と介さなかった応答関係を比較し、機械翻訳が応答関係に与える影響を調べる。分析手法には、情報検索分野で議論の流れを自動抽出する際によく用いられる語彙的結束性 (lexical cohesion) を用いた。4章では、3章の分析結果に基づき、思い違いを定量化する手法を提案する。続いて、提案手法の有効性を実験の対話

データを用いて検証する。5章では今後の課題を述べ、本論文を締めくくる。

2. 機械翻訳を用いた大規模実験における思い違い

2.1 日中異文化コラボレーション実験

我々は、総務省アジアブロードバンド計画の援助によって2003年に実施された日中異文化コラボレーション実験の対話データを分析対象とした。この実験は2003年の11月から12月まで実施され、参加者には参加報酬が支払われた。実験の参加者は、計算機システムに関する専門知識を持つ日本人18名と中国人16名の学生や研究者により構成されていた。

実験では、「異文化コラボレーション支援の具体案を1つ提案する」という課題に対して機械翻訳を介した掲示板システム上で討論が実施された。参加者は相手国の言語を理解することができなかったため、母国語だけを用いて討論を行った。実験期間中に1,106通(日本人:649通,中国人:457通)のメッセージが投稿された。

本実験で用いた翻訳機は、市販の翻訳ソフトを用いた。当該ソフトの日本語から中国語への翻訳精度は4段階(Perfect, Fair, Acceptable, Nonsense)のうち“Fair”、中国語から日本語への翻訳精度は“Acceptable”と評価されており¹¹⁾、実用的なコミュニケーション用途での翻訳精度としては十分ではないが、現在市販されている日本語と中国語間の翻訳機の中では最高水準に位置付けられるものである。

日中異文化コラボレーション実験の対話データが思い違いの分析対象として適している理由を以下にあげる。1. 実験では共通の専門知識を持つ者が支援システムの実装方法などについて議論をした。このため、文化や専門知識の違いによる思い違いは生じにくく、思い違いの大半が誤訳によるものと考えられる。2. 本実験の議論の種類は「合意形成」に分類される。合意形成過程で生じる思い違いは相互理解の破綻に発展しやすいため、このような議論における思い違いの防止策を考えることは意義深い。3. 翻訳前のメッセージと翻訳後のメッセージがともに入手可能であった。このため、機械翻訳による翻訳と翻訳者による翻訳を比較することによって、誤訳と思い違いの発生箇所を特定することができる。

2.2 誤訳による思い違いの例

本節では、実際に実験でどのような思い違いが発生

していたかを報告する。

実験の参加者は、対立意見を述べる時、まず相手の意見を尊重し部分的に同意するなどし、その後自分の意見を述べるが多かった。しかし、このようなメッセージに誤訳が混入すると、メッセージのどこまでが挨拶に相当し、どこからがユーザ自身の意見であるかを推測することが困難になる。このようなメッセージを読んだ異国ユーザは、メッセージの冒頭に含まれる肯定的な言葉が印象に残り、対立意見を同意意見と勘違いしやすくなると考えられる。実際、異国ユーザは互いに異なる内容について賛同し合う様子が観察された。以下に、異国ユーザがメッセージの主張内容を取り違えた例を実験の対話データから紹介する。

中国人ユーザによる投稿メッセージ(パイリンガルによる訳文): 私はあなたのアイデアがとても良くて、重要であると思う。[しかし、Omni-directional Camera だけでは、あなたのアイデアを実現することはできない。たとえば、]私たちは新商品が入ってくるたびに画像採集をすることはできない。あなたのアイデアの実現に必要なキーテクノロジーは画像分割法だ。Omni Directional Camera は画像分割法を導入すべきだ。

同上(機械翻訳による訳文): 私はこのようになんとかしてとてもよいと思って、とても重要だ。私達が店が毎回新しい商品を添加することがありえないため、再び画像の採集を行う。このような考えを実現するためには比較的良い画像の分割方法を利用するか創造するので、このような方法は Omni Directional Camera に対応したのだから。

日本人ユーザによる返信メッセージ(原文): あなたが何をいっているか理解できません。あなたはどんな商品があるか知りたのですか。それには普通の Web が適している。私は商店街にどんな店舗があるか知りた。だから、Town Digitizing が適している。

上の例は、中国人ユーザの投稿メッセージに対して日本人ユーザが返信しているところである。初めのメッセージは翻訳者による翻訳文である。すなわち、中国人ユーザが実際に伝えたかった内容と考えてよい。2番目のメッセージは、同じ中国人ユーザのメッセージを機械翻訳が翻訳したものである。実験に参加していた日本人ユーザは皆、中国語を読むことができなかったため、この翻訳文を読んでた。3番目のメッセージは、2番目のメッセージに対する日本人ユーザの返

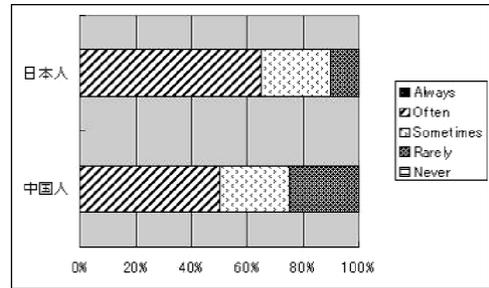


図 1 翻訳メッセージに対する主観的な理解度

Fig. 1 Members' evaluation of understandability of translated messages.

信メッセージである。

1つ目のメッセージより、中国人ユーザの主な主張点は画像分割法を Omni-directional Camera に導入することであることが読み取れる。しかし、機械翻訳による翻訳後、このユーザの主張点は新しい商品を知ることになり、画像分割法はその一実現手段となってしまう。3つ目のメッセージから、これを読んだ日本人ユーザは、中国人ユーザが画像分割法の導入に関心があるにもかかわらず、新商品を知ることに関心があると思いをしていることが分かる。

2.3 誤訳による思い違いの蓄積

実験中の全対話データを調べたところ、前節で見たような(異国ユーザ間の)思い違いが約 20 メッセージに 1つといった高頻度で観測された。しかしながら、実験終了後、全参加者にアンケート(アンケートは英語で作成し、英語の回答を求めた)で「異国ユーザのメッセージの大まかな内容をどの程度理解することができましたか?」という質問をし、「いつも理解できた」から「いつも理解できなかった」までの 5 段階評価を求めたところ、半数以上のユーザが「たいていの場合理解できた」と回答し、対話が成立していたと考えていた(図 1)。

実験ではユーザが思い違いの発生に気付かず議論を進めたために思い違いが蓄積し、相互理解の破綻に発展する対話がいくつも観測された。たとえば、日本人ユーザと中国人ユーザが最終的に至ったとする結論内容には食い違いが生じていた。実験後にメールでインタビューを実施し、相手国が合意に至ったとする主張内容について合意に至ったのか至らなかったのか確認したところ、相手国の主張内容については合意に至らなかったという回答が得られた。このような現象は、ユーザが対話が成立していると感じていても実際には対話が成立していない場合があることを示唆している。

3. ミクロレベルの分析

日中異文化コラボレーション実験の対話で頻繁に観測された思い違いは、機械翻訳の翻訳精度が低いことに大きな要因があると考えられる。我々は、このように精度が低い機械翻訳を介することによって現れる対話の特徴と思い違いの関係について分析、考察する。

本章ではまず、対話の分析単位を「返信関係にあるメッセージ対（以後、親子メッセージと呼ぶ）」とし、機械翻訳を介した親子メッセージ間の特徴と思い違いの関係を調べる。

3.1 語彙的結束性を用いた分析手法

従来、対話を分析する手法として主に会話分析や談話分析が用いられてきた。これらの手法を用いると対話の性質に関する様々な知見を得ることができる反面、人手や時間を要する。我々は思い違いの定量化を目指すため、対話をより機械的に分析できる分析手法が好ましいと考えた。

そこで、我々は情報検索分野で議論の流れを自動抽出する際によく用いられる語彙的結束性に注目した。通常の対話では、議論の流れを抽出するうえでメッセージ間の語彙的結束性（共有単語や類似単語）が重要と考えられている。高い翻訳精度が達成できない機械翻訳を用いた対話では、対話者はメッセージ中の単語やメッセージ間の語彙的結束性をもとに議論の流れやメッセージ内容を推測する。思い違いは、対話者がこのようにメッセージ間の共有単語や類似単語をもとに議論の流れやメッセージの内容を推測する過程で生じると考えられる。このため、思い違いを分析するうえで語彙的結束性に基づいた分析は有効であると考えられる。

本研究では、メッセージ X と Y に同一や同義の内容語を共有する際に、メッセージ X と Y に語彙的結束性がある、とすることとした。本研究では日本語部分についてのみ分析を行い、同義語を調べるにあたり角川類語新辞典¹²⁾を活用した。なお、以下では語彙的結束性を与える内容語をそのメッセージの結束語彙と呼ぶことにする。

3.2 妥当性の低い返信

親子メッセージ間に存在する結束語彙の数を同国間と異国間で比較した（図 2 参照）。

図 2 より、親子メッセージ間で語彙的結束性がまったくないメッセージ対の割合は同国間と異国間でほと

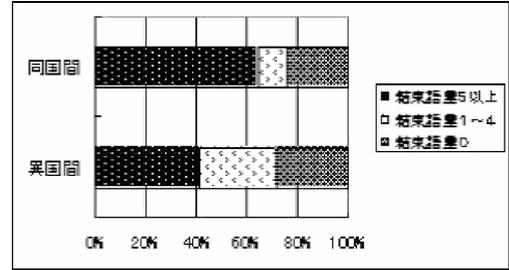


図 2 親子メッセージ間における語彙的結束性の強さ

Fig. 2 Number of cohesive lexical items shared between parent-child pairs.

んど変わらず有意差が検出されなかった ($p=.375$)。このような親子メッセージの内容を調べてみたところ、子メッセージが親メッセージに対する賛成表明や挨拶、簡単なコメントである場合が多いことが分かった。

また、同国間の親子メッセージ対において結束語彙を 5 つ以上含むメッセージ対が占める割合は 6 割程度であったことに対し、異国間の親子メッセージ対において結束語彙を 5 つ以上含むメッセージ対が占める割合は 4 割程度しか存在しなかった。t 検定の結果、親子メッセージの投稿者が異国の場合、同国の場合と比較して語彙的結束性が有意に低いことが分かった ($F=16.078$, $p=.000$)。語彙的結束性の低いメッセージ対は内容的に関連が薄い傾向にあることから⁸⁾、異国ユーザによる返信メッセージは同国ユーザによる返信メッセージと比較して親メッセージとの関連が薄いことが分かる。

個々の具体例として実際にメッセージを見ると、異国ユーザによる返信メッセージは親メッセージと同一単語や同義語を共有していても返信内容として妥当ではない場合が多数観測された。このような異国ユーザ間の妥当でない返信は、語彙的結束性の低いやりとりにつながる。また、一貫した議論が行えない理由になっていると考えられる。

3.3 短文に反応する返信

機械翻訳の翻訳精度は、一般に文長が長くなると低くなる。このように翻訳精度が低い文を含んだメッセージは異国ユーザにとっては特に解釈が難しく、部分的にしか把握できないと考えられる。我々は語彙的結束性がある親子メッセージにおいて、結束語彙を含む文の長さの分布を調べた。

同国間と異国間の分布を図 3 に与える。図 3 は、親子メッセージ間で結束語彙があるとき、その親子メッセージの全文において、親メッセージの文長ごとに（横軸）、語彙的結束性のある子メッセージの存在確率（縦軸）を求めたものである。図 3 より、親子メッセージ

内容語とは、語彙項目から一般語（ストップワード）を除いたものである。

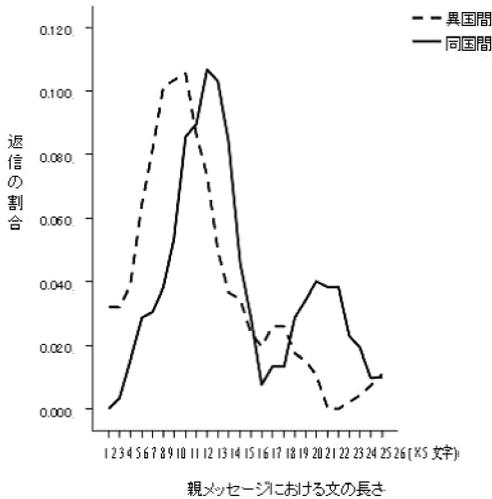


図 3 親メッセージの文長と返信の割合

Fig. 3 Distribution of responses to parent messages.

の投稿者が異国のとき、親メッセージの文長が長い場合にはその文と子メッセージの間で語彙的結束性が少ないことが分かる。t 検定の結果、親子メッセージの投稿者が異国の場合、短文部分に関連する内容の返信が有意に多いことが分かった ($F=4.816$, $p=.029$)。

このような異国ユーザ間の短文に反応する返信は、訳文を読む異国ユーザが長文部分を理解できず理解しやすい短文に集中して返信するために生じた現象と考えられる。

4. マクロレベルの分析

前章の分析結果より、異国ユーザ間の親子メッセージは語彙的結束性が弱く、親メッセージの内容と関連の薄い返信や部分的にしかつながらない返信が多いことが示唆される。このような対話では、一貫性のある深い議論を行うことは困難であり、思い違いが発生する可能性も高いと考えた。しかし、単純に親子メッセージ間の語彙的結束性の強弱だけで思い違いを定量化するには問題がある。たとえば、3.2 節で見たように、返信内容が簡単なコメントや挨拶の場合に語彙的結束性は弱くなるが、返信内容は妥当であり思い違いは生じていない。

そこで、我々は分析単位を拡大し、親子メッセージ間の語彙的結束性が低い場合でも妥当な返信と妥当でない返信を識別する方法を考える。以下では、対話全体の返信関係を木構造で表現したスレッド (discussion thread) に注目し、分析単位を「返信関係をスレッド表示したときに 1 つの木を構成するメッセージ集合」とする。

4.1 対話における構文のスレッドと意味のスレッド
ここでは、構文のスレッドと意味のスレッドに注目する。

構文のスレッド (syntactic thread) は、メッセージのヘッダ情報における “message-id” field や “in-reply-to” field をもとにメッセージ間の返信関係を定義するものである。世の中で広く使われているメールソフトの多くはメッセージ間の関係を構文のスレッドで表現する仕組みを備えている。

意味のスレッド (semantic thread) は、メッセージ間の語彙的結束性を用いてメッセージ間の返信関係を定義するものである。これまで、語彙的結束性を用いてメッセージ間の返信関係を定義する様々な手法が提案されてきた^{5),8)}。意味のスレッドは、構文のスレッドと比較して実際の対話の流れを高精度で再現できるとされている⁸⁾。

4.2 構文のスレッドと意味のスレッドの差異

我々は語彙的結束性が低い親子メッセージにおいて、妥当な返信と妥当でない返信を識別するために、構文のスレッドと意味のスレッドの差異に注目する。意味のスレッドにおける親子メッセージは内容的につながったメッセージ対であることから、構文のスレッドでは親子メッセージでも、意味のスレッドで親子メッセージと判断されないものは内容的に妥当でない応答関係と考える。

構文のスレッドと意味のスレッドの差異は次のように解釈することもできる。構文のスレッドにおける返信関係は返信の投稿者の意図を反映している。一方、意味のスレッドにおける返信関係はメッセージの内容から客観的に推定される返信関係なので、返信の投稿者を除いた参加者がメッセージのやりとりを閲覧したときに妥当と考えるような返信関係に合致するであろう⁸⁾。したがって、構文のスレッドと意味のスレッドの差異は、返信の投稿者が意図する返信関係と返信の投稿者を除く参加者が妥当とする返信関係の相違を表している。このような相違は、投稿者の返信意図と他の参加者が返信メッセージから読み取る返信意図の間に食い違いが生じていることを表している。

以上の考察から、投稿者が意図する返信関係と他の参加者が妥当と思う返信関係に相違が生じるような対話では、以下のことがいえるのではないかと考えた：

- 投稿者の発言内容について他の参加者が思い違いをしている傾向が強い。

文献 15) によれば、複数の参加者が共通のゴールに向かって議論する掲示板では、メッセージの投稿者が他の参加者に返信意図を明確に理解してほしいと考える傾向が強い。

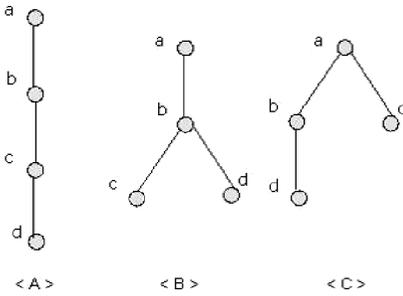


図 4 構文的スレッドと意味的スレッドの一例

Fig. 4 An example of syntactic thread and semantic thread.

- 返信関係同士の相違が多いほど、思い違いが多数蓄積し相互理解の破綻を引き起こす可能性が高い。

4.3 スレッドの差異の定量化

構文的スレッドと意味的スレッドの差異によって思い違いを計測する方式を提案する。

まず、スレッドどうしの差異を調べる 1 つの方法は、それらに含まれる返信関係の親子メッセージ対の有無を比較することである。たとえば、図 4 の A のような構造的スレッドと、B, C のような意味的スレッドがあったとする。スレッド A には、返信関係のメッセージ対として ab, bc, cd があり、B には ab, bc, bd があり、C には ab, bd, ac がある。ここで A と B を比較すると、A と B に共通なメッセージ対は ab, bc であり、A あるいは B のいずれか一方にのみ含まれるメッセージ対は cd, bd である。同様に A と C を比較すると、共通なメッセージ対は ab であり、いずれか一方のみに含まれるメッセージ対は bc, cd, bd, ac である。A と B に共通な親子メッセージ対は 2 個、A と C では 1 個なので、A と B の差異の方が A と C の差異より小さい (A と B の方がより似ている) という判定を得る。

以上の考え方に基づいて、 n 個のメッセージを含むような対話における構文的スレッドと意味的スレッドの差異を計算する手順を示す。

ステップ 1 対話の構文的スレッドからメッセージ間の関係 x_{ij} を次式より求める。

$$x_{ij} = \begin{cases} 1 & \text{メッセージ } i, j \text{ が返信関係にあるとき} \\ 0 & \text{それ以外の場合} \end{cases}$$

ステップ 2 意味的スレッドのメッセージ間の関係 y_{ij} を以下の手順で生成する。

ステップ 2-0 初期化: 閾値 α を設定する。すべての (i, j) に対して $y_{ij} = 0$ とする。

ステップ 2-1 $x_{ij} = 1$ であるすべての (i, j) に対して次の処理を実行する。

ステップ 2-1-0 (i, j) の語彙的結束性が α 以上の場合は $y_{ij} = 1$ とする。

ステップ 2-1-1 (i, j) の語彙的結束性が α 未満の場合は、 j 以前に投稿されたメッセージ k を j の投稿時刻に近い順にさかのぼり、順次 j との語彙的結束性を調べる。

ステップ 2-1-1-0 (k, j) の語彙的結束性が α 以上の場合は $y_{kj} = 1$ とする。

ステップ 2-1-1-1 語彙的結束性が α 以上のメッセージが発見できない場合 $y_{ij} = 1$ とする。

ステップ 3 構文的スレッドと意味的スレッドの差異

$$G = \frac{\sum |x_{ij} - y_{ij}|}{2(n-1)}$$

を計算する。

手順 2 は意味的スレッドの構築手順を示している。意味的スレッドは、語彙的結束性を用いて内容的につながった返信関係を表している必要がある。ところが、前節でも述べたように語彙的結束性が弱くても返信関係として妥当なものも存在する。簡単なコメントや挨拶は、それまでに投稿されたとのメッセージとも語彙的結束性が低いと予測されるからである。したがって、我々は語彙的結束性だけでなく、構文的スレッドの返信関係も考慮に入れて意味的スレッドを構築する。具体的には、語彙的結束性が弱い場合でも、返信メッセージが以前に投稿されたとのメッセージとも語彙的結束性が弱ければ、妥当な返信関係であると判断する。一方で、それまでに投稿されたメッセージ中に語彙的結束性が高いものがみつければ、構文的スレッドの返信関係は妥当でないと判断し、語彙的結束性の高いメッセージに対する応答関係が妥当な返信関係であるものとする。

ここで閾値 α は、それ以上に語彙的結束性の強いメッセージ対を返信関係と考えるのが妥当な値を選ぶ必要がある。これには、思い違いが稀にしか存在しない同国間の返信メッセージに注目し、同国間で内容につながった返信メッセージ間に存在する結束語彙数を参考にするとよいだろう。たとえば、アジアブロードバンド実験における対話データでは、同国間で内容につながった返信メッセージ間には平均 5 個以上の結束語彙が存在したため、 $\alpha = 5$ とする。

手順 3 では構文的スレッドと意味的スレッドの差異を計算する。前節の考察に基づき、構文的スレッドと意味的スレッドの差異 G を、構文的スレッドと意味的スレッドの編集距離 (edit distance) を全返信関係の数で割ったものとする。 G は 0 から 1 の値をとり、

$G = 0$ のときは構文的スレッドと意味のスレッドは同一であり、 $G = 1$ のときは構文的スレッドと意味のスレッドはすべての返信関係が異なる。たとえば、図 4 において、ある対話の構文的スレッドがスレッド A で意味のスレッドがスレッド B のとき、 $G = \frac{2}{2*3} = \frac{1}{3}$ を得、構文的スレッドがスレッド A で意味のスレッドがスレッド C のとき、 $G = \frac{4}{2*3} = \frac{2}{3}$ を得る。

4.4 実データに対する本手法の適用結果

我々は、構文的スレッドと意味のスレッドの差異 G が大きな値をとる対話ほど思い違いの発生率が高いと考える。そこで、以下では、 G の値と思い違いの発生率の関係を調べる。

● 検証データ

我々は、前章と同様に、日中異文化コラボレーション実験の対話データを検証データとして用いた。

● 検証手順

- (1) 中国人ユーザと日本人ユーザが密度の濃い議論をしていると考えられる対話を抜き出す。このような対話を抜き出す基準として次の 2 つの条件を設け、これらを同時に満たす対話を抜き出した。(a) 各対話において、メッセージのやりとりが 15 回以上ある(構文的スレッドに 15 個以上のメッセージが含まれる)、(b) 各対話において、中国人ユーザと日本人ユーザの投稿メッセージが 5 通以上含まれる。
- (2) 次に、手順 (1) によって抜き出された各対話について、思い違いの発生率を算出する。ここで、ある対話における思い違いの発生率は、その対話中に含まれる思い違いの総数を総返信数で割った値である。
- (3) 手順 (1) によって抽出された対話について、 G の値を算出する。
- (4) 思い違いの発生率と G の値の相関を調べる。

検証手順 (1) より、(a) と (b) の基準を同時に満たす対話(構文的スレッド)が 12 個抽出された。手順 (2) で思い違いを検出するにあたり、著者の 1 人が隣接ペアの類型⁶⁾に合致しない返答を手作業によって検出した。たとえば「依頼」に関する返信メッセージが「承諾/拒否」の内容を含まない場合、この返信メッセージの投稿者は思い違いをしていると判断した。

図 5 は横軸を構文的スレッドと意味のスレッドの差異 G の値、縦軸を思い違いの発生率とし、手順 (1) より得られた 12 個の対話をプロットしたものである。

図 5 より、構文的スレッドと意味のスレッドの差異 G が大きな値をとると、思い違いの発生率が上がる

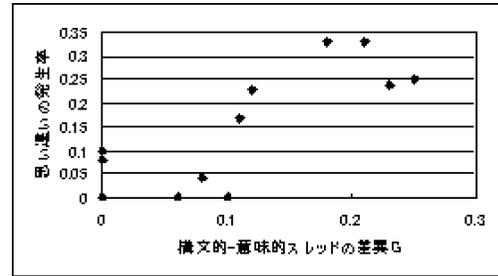


図 5 構文的スレッドと意味のスレッドの差異 G と思い違い発生率の相関

Fig. 5 Correlation between G and misconception frequency.

ことが分かる。なお、 G の値と思い違いの発生率の相関を計測したところ、有意な正の相関があることが分かった ($r = .785$, $p = .003$)。以上より、構文的スレッドと意味のスレッドの差異は各対話における思い違いの発生率を機械的に推定する手法として有効であるといえる。

5. おわりに

機械翻訳を用いた協調作業を支援するうえで、思い違いの発生を食い止めることは重要な課題である。Krauss ら⁷⁾が指摘するように、後続の対話にも影響を及ぼし相互理解の破綻を招くような思い違いを検知、修正できるメカニズムを準備することが重要である。

そこで、本論文では、高い翻訳精度が達成できない機械翻訳を用いた対話環境において頻繁に発生する思い違いに注目し、これらの思い違いを定量化する手法を提案した。

我々はまず、時間的に近い発話対の特徴と思い違いの関係を分析した。機械翻訳を用いた異国ユーザ間のコミュニケーションと機械翻訳を用いない同国ユーザ間のコミュニケーションを比較分析したところ、以下の知見を得た。

- (1) 異国ユーザ間の対話では、メッセージ間の結束性が低く妥当性の低い返信が多く観測された ($p = .000$)。
- (2) 異国ユーザ間の対話では、分かりやすい短文に反応する返信が多く観測された ($p = .029$)。

以上より、機械翻訳を介した異国ユーザによる返信メッセージは、内容をとらえずメッセージの部分的なところでつながっている可能性が高いことが分かった。

以上の分析結果をもとに、思い違いを機械的に測定する手法を提案した。

提案手法は、ヘッダ情報に基づくスレッドと語彙的結束性に基づくスレッドの差異を計測し、この差異が

大きければ異国ユーザ間の思い違いが生じやすいと判断するものである。実際に、対話データを用いて提案手法を検証したところ、提案手法によって定量化した思い違いと実際に観測された思い違いの間には有意な正相関 ($r = .785, p = .003$) が観測され、有効性が示された。

本提案手法は、翻訳機が内容語を忠実に翻訳するという性質を利用している。現在、市販されている翻訳機は、内容語を省略したり補完したりすることは少なく、内容語を忠実に翻訳するものが大半を占める。これは、現在の翻訳機が文脈を考慮せずに各文を独立に翻訳しており、勝手に内容語を省略したり追加したりすると文の意味が通らなくなる可能性が高いためである。内容語を自在に省略や補完できる翻訳機が手に入らない現状では、少なくとも本手法は有効であろう。

今後の課題を以下にあげる。まず、本論文では日本語部分を用いた分析しか行っていないため、今後は中国語部分についても同様の分析を実施する必要がある。日本語部分しか分析しないことによって生じうる問題点としては、語彙的結束性の弱いメッセージ対が中国語部分を用いて分析した場合と一致しないことが考えられる。たとえば、中国語から日本語への翻訳精度が高く日本語から中国語への翻訳精度が低い場合、仮にある親子メッセージの日本語部分における語彙的結束性が強くて、日本語から中国語へ翻訳する際に誤訳が混入すると中国語部分における語彙的結束性は弱くなるであろう。

続いて、本論文の提案手法を異言語間コラボレーション支援システムに組み込み、その効果を調べる予定である。本手法を用いてユーザに思い違いの発生を気付かせることにより、思い違いが大きな損失に結び付く前に対策をとることが可能になると考える。

謝辞 本研究では、総務省アジアブロードバンド計画の援助によって2003年に実施された日中異文化コラボレーション実験の対話データを使わせていただいた。アジアブロードバンド計画の関係者の皆様ならびに日中異文化コラボレーション実験に携われた皆様に深く感謝いたします。また、NTTコミュニケーション科学基礎研究所の松原繁夫氏と菅原俊治氏、そしてNICTの内元清貴氏から本論文の執筆にあたり有意義なコメントをいただいたことを心から感謝いたします。最後に、丁寧に査読し本論文の内容と質を大きく改善するためのコメントを与えてくださった査読者の方々に感謝いたします。

参考文献

- 1) Aiken, M.: Multilingual Communication in Electronic Meetings, *Proc. ACM SIGGROUP*, Vol.23, No.1, pp.18–19 (2002).
- 2) Churchill, E.F. and Bly, S.: Culture Vultures: Considering Culture and Communication in Virtual Environments, *Proc. ACM SIGGROUP*, Vol.21, No.1, pp.6–11 (2000).
- 3) Climent, S., More, J., Oliver, A., Salvatierra, M., Sanchez, I., Taule, M. and Vallmanya, L.: Bilingual Newsgroups in Catalonia: A Challenge for Machine Translation, *JCMC*, Vol.9, No.1 (2003).
- 4) Cramton, D.C.: The mutual knowledge problem and its consequences for dispersed collaboration, *Organization Science*, Vol.12, pp.346–371 (2001).
- 5) Donath, J.: A Semantic Approach to Visualizing Online Conversations, *Comm. ACM*, Vol.45, No.4, pp.45–49 (2002).
- 6) 石崎雅人, 伝 康晴: 談話と対話, 東京大学出版会 (2001).
- 7) Krauss, R.P. and Fussell, S.: Mutual knowledge and communicative effectiveness, *Intellectual Teamwork: Social and Technological Foundations of Cooperative Work*, pp.111–146 (1990).
- 8) Lewis, D.D. and Knowles, K.A.: Threading Electronic Mail: A Preliminary Study, *Info. Proceedings and Management*, Vol.33, No.1, pp.209–217 (1997).
- 9) 野村早恵子, 石田 亨, 船越 要, 安岡美佳, 山下直美: アジアにおける異文化コラボレーション実験 2002: 機械翻訳を介したソフトウェア開発, 情報処理, Vol.44, No.5, pp.503–511 (2003).
- 10) Nomura, S., Ishida, T., Yasuoka, M., Yamashita, N. and Funakoshi, K.: Open Source Software Development with Your Mother Language: Intercultural Collaboration Experiment 2002, *International Conference on Human-Computer Interaction (HCI2003)*, Vol.4, pp.1163–1167 (2003).
- 11) Ogura, K., Hayashi, Y., Nomura, S. and Ishida, T.: User Adaptation in MT-mediated Communication, *The 1st International Joint Conference on Natural Language Processing (IJCNLP-04)*, pp.596–601 (2004).
- 12) 大野 普, 浜西正人: 角川類語新辞典, 角川書店 (1981).
- 13) Rogers, Y.: Ghosts in the Network: Distributed Troubleshooting in a Shared Working Environment, *Proc. CSCW'92* (1992).
- 14) Rogers, Y.: Coordinating computer-mediated

work, *Computer Supported Cooperative Work*, No.1, pp.295-315 (2002).

- 15) Yamauchi, Y., Yokozawa, M., Shinohara, T. and Ishida, T.: Collaboration with Lean Media: How Open-Source Software Succeeds, *Proc. CSCW'00*, pp.329-338 (2000).
- 16) Yetim, F.: Meta-Communication Model for Structuring Intercultural Communication Action Patterns, *Proc. ACM SIGGROUP*, Vol.22, No.2 (2001).

(平成 17 年 5 月 31 日受付)

(平成 17 年 11 月 1 日採録)



山下 直美 (正会員)

1999 年京都大学工学部情報工学科卒業。2001 年京都大学大学院情報学研究科数理工学専攻修士課程修了。同年日本電信電話(株)コミュニケーション科学基礎研究所入所、

現在に至る。2005 年より京都大学大学院情報学研究科社会情報学博士課程在学中。



石田 亨 (フェロー)

1976 年京都大学工学部情報工学科卒業, 1978 年京都大学大学院修士課程修了。同年日本電信電話公社電気通信研究所入所。ミュンヘン工科大学, パリ第六大学, メリーランド大学客員教授等経験。工学博士。IEEE フェロー。情報処理学会フェロー。現在, 京都大学大学院情報学研究科社会情報学専攻教授, 上海交通大学客員教授。自律エージェントとマルチエージェントシステム, セマンティック Web 技術に取り組む。デジタルシティ, 異文化コラボレーションプロジェクトを推進。



平田 圭二 (正会員)

1987 年東京大学大学院工学系研究科情報工学専門課程博士課程修了。工学博士。1990~1993 年(財)新世代コンピュータ技術開発機構(ICOT)。日本ソフトウェア科学会第 3 回大会高橋奨励賞(1987 年)。情報処理学会論文賞(2002 年), 山下記念研究賞(2003 年)。音楽情報処理とインタラクションに興味を持つ。