

# 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<sup>†</sup> 石田 亨<sup>††</sup> 平田 圭二<sup>†</sup>

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、誤訳によって大量に発生するユーザ間の思い違いの問題に焦点を当て、機械翻訳を用いて行われた対話メッセージ 1106 通を対象に、誤訳によってどのように思い違いが発生するかを分析した。分析結果より、異国ユーザ間の返信メッセージは内容を捉えておらず、メッセージの部分的なところで繋がっている可能性の高いことがわかった。そこで、内容を捉えずに対話が進行する状態を機械的に測定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。本手法は、機械翻訳を用いた対話に限らず、雑音の多い不良環境下 (hostile environment) における思い違いの測定にも有効であると考えられる。

## Analyzing Misconception in Machine Translation Mediated Communication

NAOMI YAMASHITA,<sup>†</sup> TORU ISHIDA<sup>?</sup> and KEIJI HIRATA<sup>†</sup>

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<sup>3)</sup>。

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

は背景知識や文脈等から対話内容を理解し対話が成立する<sup>1),8),9)</sup>。たとえば、機械翻訳の翻訳精度が比較的良好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回答している<sup>1)</sup>。

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軽微な誤解を指す。

<sup>†</sup> 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

<sup>††</sup> 京都大学情報学研究所

Kyoto University, Graduate School of Informatics

第2節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概ね理解し思い違いは生じていないと感じていたユーザどうしであっても、実際には相互理解が破綻していた事例が幾つも発見された。このような思い違いが生じ易くかつ気づきにくい理由は、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上で、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<sup>6)</sup>。

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ー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未だ存在しない。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<sup>2),11),14)</sup>、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一歩として、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出せなかったため、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ミクロとマクロの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ミ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通りである。第2節では、従来研究における思い違いと機械翻訳の誤訳による思い違いの関係について述べる。また、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。第3節では、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。第4節では、3節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。第5節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一つ提案する」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1106通(日本人:649通,中国人:457通)のメッセージが投稿された。

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挙げる。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共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 思い違いの分類

ここでは、思い違いの発生源を言外部分 (unsaid) と言明部分 (said) の2つに分類する。

---

翻訳には高電社の翻訳システム J-server が使用されていた。

言外部分に関する思い違いとは、対話者が互いに共有できていると考える知識が実際には共有できていない状態で対話を進めた場合に生じる思い違いである。先行研究<sup>4),12)</sup>によると、一般の対話で発生する思い違いの大半は発生源が言外部分にあるとされている。たとえば、話し手側は敢えて明言しなくても聞き手側が自分の意図を察して行動してくれることを期待し (speech act)、聞き手はその意図を理解できなかったような場合である。

次に、言明部分に関する思い違いとは、話し手の言明部分に関して聞き手がその意味を取り違えた場合に生じるものである。一般の対話では、言明部分に関する思い違いは発生しにくく、仮に発生した場合でも後続の対話で補正されることが多い。たとえば、話し手側は A について話しているのに、聞き手側は A について話をしているとは考えていないような場合である。

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、誤訳により聞き手側が言明内容を正確に理解できない状況が頻発するであろう。このため、通常の対話ではさほど観測されない言明部分に関する思い違いも発生すると考えられる。

### 2.3 誤訳による思い違いの例

本節では、前節で行った思い違いの分類をもとに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、対立意見を述べるとき、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易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 (バイリンガルによる訳文): 私はあなたのアイデアがとてもよくて、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度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るべきだ。

同上 (機械翻訳による訳文): 私はこのようにでなく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よ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 に対応したのべきだ。

日本人ユーザによる返信メッセージ (原文): あなたが何を言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いのですか。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い。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。二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いた。三番目のメッセージは、二番目のメッセージに対する日本人ユーザの返信メッセージである。

一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。三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関わらず、新商品を知ることに関心があると思い違いをしていることがわかる。

### 2.4 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) 言明部分に関する思い違いが約 20 メッセージに 1 つといった高頻度で観測された。

これほど高頻度で思い違いが発生していたにも関わらず、興味深いことに実験の参加者は思い違いの発生に殆ど気付いていなかった。実際、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「大抵の場合理解できた」と回答

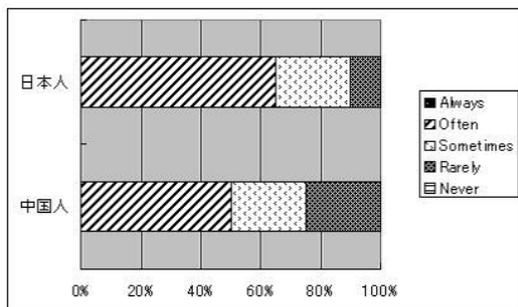


図 1 翻訳メッセージに対する主観的な理解度

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

し、対話が成立していたと考えていた (図 1)。

また、実験ではユーザが思い違いの発生に気付かずに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幾つも観測された。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は合意に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言明部分における思い違いは、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。

本節ではまず、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。

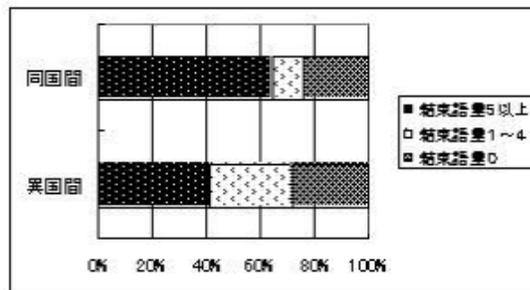


図 2 親子メッセージ間における語彙的結束性の強さ

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

通常の対話では、議論の流れを抽出する上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、誤訳による思い違いを分析する上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  $Y$  に語彙的結束性があるとは、メッセージ  $X$  と  $Y$  が同一語や同義語を共有することとした。本研究では同義語を調べるにあたり角川類語新辞典<sup>10)</sup>を活用した。なお、以下では語彙的結束性を与える語彙項目をそのメッセージの結束語彙と呼ぶことにする。

#### 3.2 言葉尻を捉えた返信

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全くないメッセージ対の割合は同国間と異国間で殆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わかった。

一方、親子メッセージ間に語彙的結束性がある場合については、親子メッセージ間に結束語彙を多く含むメッセージ対は同国間では 6 割程度なのに対し異国間では 4 割程度であった。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わかった ( $F=16.078$ ,  $p=.000$ )。つまり、親子メッセージの投稿者が異国の場合、子メッセージの投稿者は親メッセージの内容を

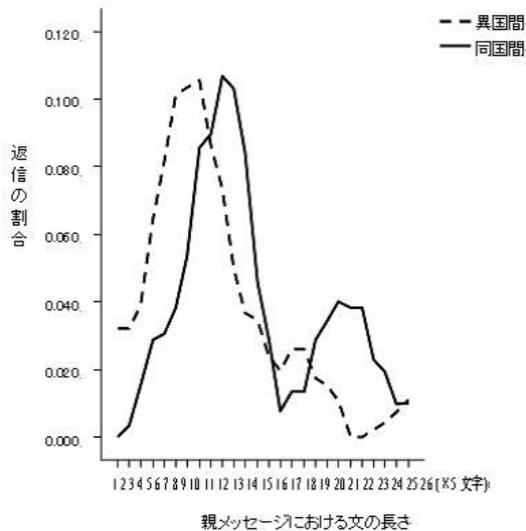


図3 親メッセージの文長と返信の割合

Fig. 3 Distribution of responses to parent messages.

少ない単語でわかった気になって返信していると考えられる。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言葉尻を捉えた返信は、機械翻訳の翻訳精度が低く異国ユーザが記述したメッセージを部分的にしか把握できないために生じた現象と解釈することができ、思い違いの発生と関連が深いと考えられる。

### 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読解が難しく、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の横軸は親子メッセージ間で結束語彙を含む親メッセージの文長、縦軸はその返信の割合である。図3より、親子メッセージ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わかる。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わかった ( $F=4.816$ ,  $p=.029$ )。

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

易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

前節の分析結果より、異国ユーザによる返信メッセージは内容を捉えずメッセージの部分的なところで繋がることが多く、そのようなメッセージ間では語彙的結束性が弱いことがわかった。このような返信関係は、ユーザが翻訳メッセージの内容をきちんと理解できていないときに現れ、思い違いを誘発すると考えられる。したがって、語彙的結束性を用いれば思い違いを定量化できるはずである。

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、3.1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一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<sup>5),7)</sup>。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。

### 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。意味のスレッドでは、内容的に関連が深いと考えられるメッセージ対を返信関係とするため、意味のスレッド

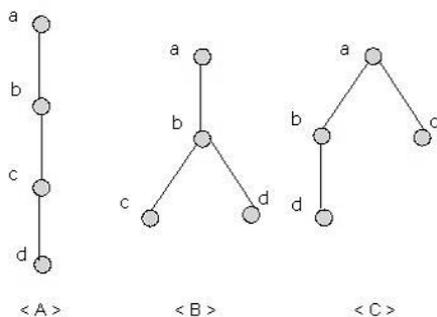


図4 構文的スレッドと意味的スレッドの一例

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

上で親子関係にないメッセージ対は内容的にも関連が弱いと考えられる。そこで、構文的スレッド上で語彙的結束性が弱い親子メッセージの中で意味的スレッドの返信関係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。

構文的スレッドと意味的スレッドの差異は次のように解釈することもできる。構文的スレッドにおける返信関係は、投稿者（話し手）の返信意図を反映した返信関係と考えることができる。一方、意味的スレッドにおける返信関係は、メッセージの閲覧者（聞き手）が対話を通じて感じ取る返信関係と考えることができる。したがって、構文的スレッドと意味的スレッドの差異は、話し手が意図する返信関係と聞き手が感じ取る返信関係に関する意識の食い違いを意味する。

思い違いは話し手と聞き手の意識の食い違いによって生じるため、構文的スレッドと意味的スレッドに差異が存在する対話では思い違いが生じ易いと考えられる。なお、構文的スレッドと意味的スレッドに大きな差異が存在する対話では、話し手と聞き手の間で多数の食い違いがあることを意味している。したがって、このような対話では思い違いが（多数）発生している可能性が高いと考えられる。

たとえば、図4のスレッドAとスレッドBの差異（話し手と聞き手の返信関係に関する意識の食い違い）は、メッセージc,dとメッセージb,dにある。同様に、スレッドAとスレッドCの差異は、メッセージc,dとメッセージb,dとメッセージa,cとメッセージb,cにある。したがって、図4の例の場合には、スレッドAとCの差異がスレッドAとBの差異よりも大き

返信関係に対する意識は、複数のユーザが共通のゴールに向かって議論する掲示板において特に強い<sup>13)</sup>。なぜなら、このような掲示板では、メッセージの投稿者は他のユーザに返信意図を明確に理解して欲しいと考える傾向が強いためである。

い。すなわち、ある対話の構文的スレッドがスレッドAで意味的スレッドがスレッドCの場合、その対話の意味的スレッドがスレッドBの場合と比較して思い違いが発生し易いと考えられる。

#### 4.3 スレッドの差異の定量化

前節の考察に基づき、以下にメッセージを  $n$  個含む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。

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。

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

2. 意味的スレッドのメッセージ間の関係  $\tilde{x}_{ij}$  を以下の手順で作成する。

2-0. 閾値  $\alpha$  を設定する。すべての  $(i, j)$  に対して  $\tilde{x}_{ij} = 0$  とする。

2-1.  $x_{ij} = 1$  であるすべての  $(i, j)$  において、 $(i, j)$  の語彙的結束性が  $\alpha$  以上の場合： $\tilde{x}_{ij} = 1$  とする。

$(i, j)$  の語彙的結束性が  $\alpha$  未満の場合： $j$  以前に投稿されたメッセージを  $j$  に投稿時間が近い順から遡り、 $j$  との語彙的結束性が  $\alpha$  以上のメッセージ  $k$  が存在した場合、 $\tilde{x}_{kj} = 1$  とする。そのような  $k$  が見つからなかった場合、 $\tilde{x}_{ij} = 1$  とする。

3. 構文的スレッドと意味的スレッドの差異  $G = \frac{\sum |x_{ij} - \tilde{x}_{ij}|}{2(n-1)}$  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繋がった返信関係を表わ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。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ど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。これは、簡単なコメントや挨拶は、それまでに投稿されたど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。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かれば、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  $\alpha$  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、同国間で内容の繋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よくだらう。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、同国間で内容の繋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、 $\alpha = 5$  となる。

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的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的スレッドは全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 で意味的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 で意味的スレッドがスレッド B 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的スレッドの差異  $G$  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、前節と同様に、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二つの条件を設け、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、メッセージのやり取りが 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、

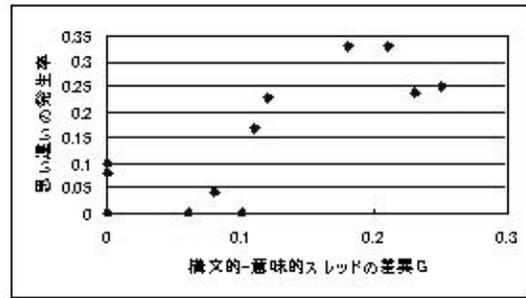


図 5 構文的スレッドと意味的スレッドの差異  $G$  と思い違い発生率の相関

Fig. 5 Correlation between  $G$  and misconception frequency.

$G$  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。なお、手順 (2) で思い違いを数える際、構文的スレッドの各親子メッセージ間において思い違いが発生していないか調べた。この作業により発見された思い違いは全て異国ユーザ間で発生していた。

図 5 は横軸を構文的スレッドと意味的スレッドの差異  $G$  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的スレッドの差異  $G$  が大きな値をとると、思い違いの発生率が上がることがわかる。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わかった ( $r = .785, p = .003$ )。以上より、構文的スレッドと意味的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上で、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 and Fussell ら<sup>6)</sup> 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ける思い違いの発生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进行分析した。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言葉尻を捉えた返信パターンが多く観測された ( $p = .000$ )。
- (2) 異国ユーザ間の対話では、わかりやすい短文に反応する返信パターンが多く観測された ( $p = .029$ )。

以上より、異国ユーザ間の返信メッセージは内容を捉えておらず、メッセージの部分的なところで繋がっている可能性が高いことがわかった。

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の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。

本手法は、対話中に大きなノイズが混入したときに人が何度も同じ単語や表現を用いるという性質を利用している。この現象は機械翻訳を介した対話に限らず、例えば携帯電話に雑音を混入させたときも同様の現象が起きることが報告されている<sup>15)</sup>。このため、本手法は大きなノイズが混入する環境下の対話においても有効であろう。しかし、逆にノイズのない対話では、発話内容が繋がっていても発話間の結束語彙が少ないことが多いため、本手法は有効ではないであろう。

本論文の提案手法を、例えば異文化コラボレーション支援システムに組み込むことにより、ユーザに思い違いの発生を気付かせることができる。重要な内容の対話で思い違いの発生率が高い場合、そのような対話だけ翻訳者に翻訳を依頼するといった対策をとることによって、誤訳による思い違いが大きな損失に結びつくことを低コストで防ぐことが可能になる。

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頂いた。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致します。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から本論文の執筆にあたり有意義なコメントを頂いたことを心から感謝致します。

## 参 考 文 献

- 1) Aiken, M.: Multilingual Communication in Electronic Meetings, *Proceedings of ACM SIGGROUP*, Vol. 23, No. 1 (2002).
- 2) Churchill, E. F. and Bly, S.: Cultural Vultures: Considering Culture and Communication in Virtual Environments, *Proceedings of ACM SIGGROUP*, Vol. 21, No. 1 (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, *Communications of the ACM*, Vol. 45, No. 4, (2002).
- 6) Krauss, R. P. and Fussell, S.: Mutual knowledge and communicative effectiveness, *Intellectual Teamwork: Social and Technological Foundations of Cooperative Work*, pp. 111–146 (1990).
- 7) 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).
- 8) 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).
- 9) 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).
- 10) 大野 普, 浜西 正人: 角川類語新辞典, 角川書店 (1981).
- 11) Rogers, Y.: Ghosts in the Network: Distributed Troubleshooting in a Shared Working Environment, *Proceedings of CSCW'92* (1992).
- 12) Rogers, Y.: Coordinating computer-mediated work, *Computer Supported Cooperative Work*, No. 1, pp. 295–315 (2002).
- 13) Yamauchi, Y., Yokozawa, M., Shinohara, T. and Ishida, T.: Collaboration with Lean Media: How Open-Source Software Succeeds, *Proceedings of CSCW'00* pp. 329–338 (2000).
- 14) Yetim, F.: Meta-Communication Model for

Structuring Intercultural Communication Action Patterns, *Proceedings of ACM SIG-GROUP*, Vol. 22, No. 2, (2001).

- 15) 吉田由香里, 大賀寿郎: 電話による対話のしやすさに対する S/N 比の影響について, 日本音響学界講演論文集 (2004).
-