

機械翻訳を用いた対話における思い違いの定量化手法の提案

Analyzing Misconception in Machine Translation Mediated Communication

山下 直美† 石田 亨‡
Naomi Yamashita Toru Ishida

1. はじめに

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[2]。

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。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。そのような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。実際、第2節で述べるように、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、相互理解が破綻していた事例が幾つも発見された。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。このように高い機械翻訳が達成できない機械翻訳を用いた対話では相互理解の破綻が生じ易いため、思い違いを解消または機械的に計測する手法を開発することは緊急の課題である。

従来の思い違いに関する研究[1,7,9]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、思い違いを機械的に分析あるいは支援する方法は未だ考案されていない。

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、思い違いの発生を定量化する方法を述べる。筆者らが調べた範囲では本研究に直接関係するような研究は見当たらなかったため、還元論的なアプローチをとり、まず対話をミクロなレベルに分解してその振る舞いを分析、解釈した上で対話全体におけるマクロな思い違いの定量化を行う。具体的には、まず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)を用いて、時間的に近い発話間のミクロな思い違いを分析する。次に、メッセージのヘッダ情報によるスレッドとミクロな思い違い分析によるスレッドの差異を用いて、対話全体に渡るマクロな思い違いを定量化する。

本論文で提案する手法を利用した支援ツールを開発することにより、誤訳による異国ユーザ間の思い違いが減ることが期待できる。

2. 思い違いに関する対話分析

この節では、誤訳による思い違いが生じる原因を日中異文化コラボレーション実験のデータを元に分析する。

2.1 日中異文化コラボレーション実験

2003年に総務省アジアブロードバンド計画の援助によって実施された日中異文化コラボレーション実験を分析対象とした。この実験の被験者は日本人18名と中国人16名であり、実験では「異文化コラボレーション支援の具体策を提案する」という課題に対して機械翻訳(翻訳には高電社の翻訳システム J-server が使用されていた)を介した掲示板システム上で討論した。また、被験者は母国語しか使用していない。討論期間中に、1106通(日本人:649通,中国人:457通)のメッセージが投稿された。

2.2 分析手法

機械翻訳を介した対話では、対話メッセージの文法構造が破壊されることが多い。文法構造が破壊されたメッセージを理解する上で、メッセージ中の単語は大きな手掛かりになる。実際、通常の対話でも、議論の流れを抽出する上でメッセージ間の共有単語や類似単語が重要と考えられている。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内容を推測する過程で生じる。

そこで、我々は思い違いを分析する上で語彙的結束性に基づいた分析が有効であると考えた。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[6]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。

2.3 分析結果：思い違いに関連深いコミュニケーションパターン

思い違いは特に異国ユーザ間に生じていた。そこで、以下の分析では、メッセージの返信関係が同国である場合と異国である場合に分類し、語彙的結束性を用いてこれらの2グループを比較した。以下、その分析結果を報告する。

2.3.1 言葉尻を捉えた返信パターン

親子メッセージ間に存在する結束語彙の数を同国と異国間で比較した(図1参照)。

図1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。

† NTT コミュニケーション科学基礎研究所

‡ 京都大学 情報学研究科

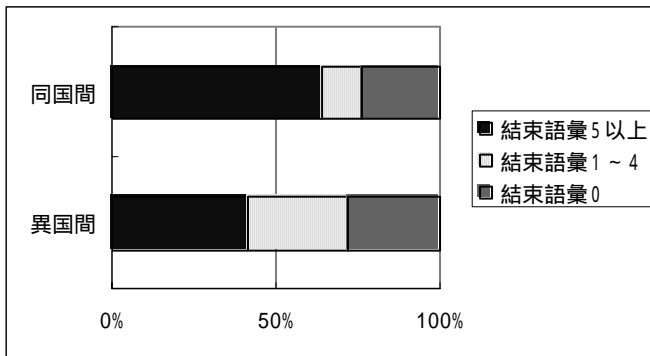


図 1：親子メッセージ間における語彙的結束性の強さ

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 割なのに対し異国では 4 割であった。t 検定の結果、親子メッセージの投稿者が同国の場合、異国の場合と比較して語彙的結束性が有意に高いことがわかった ($F=16.078, p=.000$)。つまり、親子メッセージが異国の場合、子メッセージの投稿者は親メッセージの内容を少ない単語でわかった気になって返信している。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。

2.3.2 短文を捉えた返信パターン

異国ユーザ間の親子メッセージで結束語彙が少ない原因について更に詳しく調べた。具体的には、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。

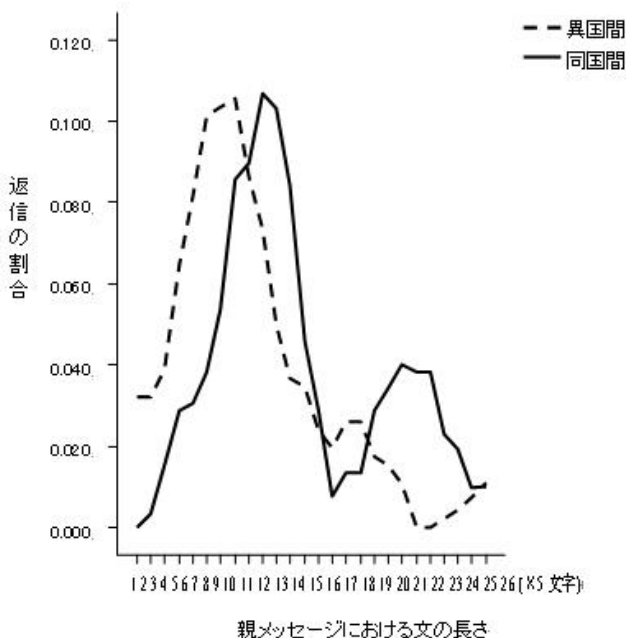


図 2：親メッセージの文長と返信の割合

同国間と異国間の分布を図 2 に与える。図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、縦軸はその文に対する返信の割合である。図 1 より、異国の返信

者は、同国の返信者に比べて、短文に返信する傾向が強いことがわかる。t 検定を行ったところ、その結果が有意であることがわかった ($F=4.816, p=.029$)。

この結果の理由は、長い文の翻訳精度は低いため、訳文を読む異国ユーザは長文部分を理解できず、結果的に理解しやすい短文に集中して返信するからだと考えられる。

3. 思い違いの計測手法

機械翻訳を用いた協調作業を支援する上で、これまで見てきたような思い違いの発生を食い止めることは重要な課題である。ここで、我々は、思い違いが対話中に全く含まれてはいけないと主張しているわけではない。Krauss and Fussell ら[6]が指摘するように、軽度な思い違いであれば対話中に多少含まれていても問題はない。問題となるのは、後続の対話にも影響を及ぼす思い違いである。重要なことは、このように相互理解の破綻を招く思い違いを検知、修正できるメカニズムを準備することである。

本節では、そのようなメカニズムを準備する第一歩として、対話全体に渡って生じている思い違いを計測する手法を提案する。ここで思い違いの「検出手法」ではなく「計測手法」とした理由は、思い違いの判定基準が人によって異なるからである。

前節の分析より、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が多く、結果としてメッセージ間の語彙的結束性が弱いことがわかった。そのため、語彙的結束性を用いれば、思い違いを計測できるはずである。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計測するには問題がある。例えば、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、思い違いは生じていない。また、親子メッセージ間の語彙的結束性だけでは対話全体に渡った思い違いを計測することは難しい。

そこで、我々是对話全体の返信関係を木構造で表現したスレッド(discussion thread)に注目した。ヘッダ情報を用いて構築した構文のスレッドとメッセージ間の語彙的結束性を用いて構築した意味のスレッドを比較し、これらのスレッドに大きな差異が生じるとき、相互理解の破綻が生じている可能性が高いと考えた。

3.1 提案手法：構文的-意味的スレッドの差異

世の中で広く使われているヘッダ情報に基づいたスレッドを構文のスレッド(syntactic thread)と呼ぶ。目的を共有した議論ではユーザは返信先(親メッセージ)の指定に注意を払う[8]。このため、構文のスレッドにおける返信関係は、投稿者(話し手)の意図を反映した返信関係と考えることができる。一方で、メッセージ間の語彙的結束性に基づいたスレッド[4]を意味的スレッド(semantic thread)と呼ぶ。ユーザはある話題に関して集中的に対話を交わしているとき、そのトピックに関連する単語を頻繁に用いる。このため、意味的スレッドにおける返信関係は、メッセージ内容から客観的に推定される返信関係と考えることができる。特に機械翻訳を介した議論では、メッセージの文法構造が破壊されることが多いため、閲覧者は異国ユーザの発言意図をメッセージ中に含まれる単語とそれまでの議論の流れから推測すると考えられる。すなわち、聞き手は意味的スレッドをベースにメッセージ内容を理解すると考えられる。

我々は話し手が意図する返信関係と聞き手が感じる返信関係が異なる場合、思い違いが生じ易いと考え、構文的スレッドと意味的スレッドの差異で思い違いを計測する手法を提案する。

以下に、メッセージを n 個含む対話における構文的-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のメッセージ間の関係 x_{ij} を次のように求める。メッセージ i と j が返信関係にあるとき $x_{ij} = 1$ とし、そうでないとき $x_{ij} = 0$ とする。
- (2) 意味的スレッドのメッセージ間の関係 \tilde{x}_{ij} を以下の手順で求める。まず、閾値 θ を設定し、すべての (i, j) に対して $\tilde{x}_{ij} = 0$ とする。次に、 $x_{ij} = 1$ であるすべての (i, j) において、 (i, j) の語彙的結束性が θ 以上の場合は $\tilde{x}_{ij} = 1$ とする。そうでないときは、 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 θ 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-意味的スレッドの差異 G を計算する。

$$G = \sum |x_{ij} - \tilde{x}_{ij}| / 2(n-1)$$

手順(2)は意味的スレッドの構築手順を示している。前節でも述べたように、簡単なコメントや挨拶のメッセージなどのように、語彙的結束性はな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、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必要がある。そこで、語彙的結束性が低い場合でも、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した。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と考えた。

ここで閾値 θ は、それ以上の結束性のある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、 $\theta = 5$ とする。

提案手法で計算される構文的-意味的スレッドの差異 G は、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した。 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。 G が大きな値をとるほど、話し手と聞き手の対話理解に食い違いがあり、その対話における思い違いの発生確率が高いことが予想される。

3.2 提案手法の検証

この節では、前節で定義した構文的-意味的スレッドの差異 G の値と思い違いの発生率の関係を調べる。

3.2.1 検証方法

- 検証データ：日中異文化コラボレーション実験の対話データ。
- 検証手順：
 - (1) 次の二つの条件を同時に満たす密度の濃い対話を抽出した。(a) メッセージのやり取りが 15 回以上ある、(b) 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) 手順(1)によって抽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、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。

3.2.2 検証結果

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)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 3 は横軸を構文的-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。

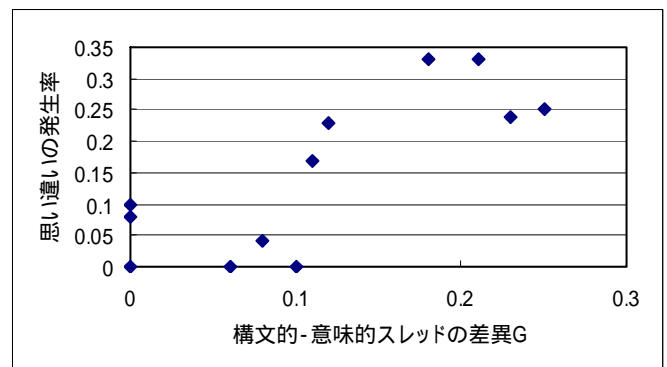


図 3：構文的-意味的スレッド差異と思い違い発生率の相関

図 3 より、構文的-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った。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 ($r=0.785, p=0.003$)。

以上より、構文的-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。

4. おわりに

本研究では、機械翻訳の誤訳によって生じるユーザ間の思い違いに焦点を当てた。

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。

- (i) 異国ユーザ間の対話では，同国ユーザ間の対話と比較して，わかりやすい短文を捉えた返信パターンが多い．
- (ii) 異国ユーザ間の対話では，同国ユーザ間の対話と比較して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い．

次に，この知見に基づいて，対話全体に渡った思い違いを機械的に測定する手法を考案した．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，有効性が示された．

今後，本提案手法を利用した支援ツールを開発することにより，誤訳による異国ユーザ間の思い違いが減ることが期待できる．

謝辞

本研究の基盤となるデータを提供して頂いたアジアブロードバンド計画の関係者ならびに日中異文化コラボレーション実験の参加者の方々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平田圭二氏から多大な協力を頂いたことを心から感謝致します．

参考文献

- 1) Churchill, E. F. and Bly, S. (2000): 'Cultural Vultures: Considering Culture and Communication in Virtual Environments', *Proceedings of ACM SIGGROUP*, vol. 21, no. 1.
- 2) Climent, S., More, J., Oliver, A., Salvatierra, M., Sanchez, I., Taule, M. and Vallmanya, L. (2003): 'Bilingual Newsgroups in Catalonia: A Challenge for Machine Translation', *Journal of Computer Mediated Communication*, vol. 9, no. 1.
- 3) Krauss, R. P. and Fussell, S. (1990): 'Mutual knowledge and communicative effectiveness', *Intellectual Teamwork: Social and Technological Foundations of Cooperative Work*, pp. 111-146.
- 4) Lewis, D. D. and Knowles, K. A. (1997): 'Threading Electronic Mail: A Preliminary Study', *Info. Proceeding and Management*, vol. 33, no. 1, Elsevier, pp. 209-217.
- 5) 野村早恵子, 石田 亨, 船越 要, 安岡美佳, 山下直美 (2003): 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.
- 6) 大野普, 浜西正人 (1981): 角川類語新辞典, 角川書店.
- 7) Rogers, Y. (1992): 'Ghosts in the Network: Distributed Troubleshooting in a Shared Working Environment', *Proceedings of CSCW*, ACM Press.
- 8) Yamauchi, Y., Yokozawa, M., Shinohara, T. and Ishida, T. (2000): 'Collaboration with Lean Media: How Open-Source Software Succeeds', *Proceedings of CSCW*, ACM Press, pp. 329-338.
- 9) Yetim, F. (2001): 'Meta-Communication Model for Structuring Intercultural Communication Action Patterns', *Proceedings of ACM SIGGROUP*, Vol.22, No.2.