

機械翻訳された文章の違和感に関わる要因

Factors of unnaturalness in machine-translated sentence

野原 康平[†], 女川 亮司^{†‡}, 渡邊 克巳[†]
Kohei Nohara, Ryoji Onagawa, Katsumi Watanabe

[†]早稲田大学, [‡]日本学術振興会
Waseda University, Japan Society for the Promotion of Science
nohara@fuji.waseda.jp

概要

本論文は、機械翻訳が作成する文章において、人間のみが感じる違和感の要因について調査したものである。この研究では、2択の文章選択課題の回答傾向と、「なぜそう答えたか」という回答を分析して要因を検討した。理由の分析では、テキストマイニングを使用し、共起語ネットワークの評価と対応分析を行なった。これらの結果、文章を読む際に感じる違和感には6つの要因が存在することが示唆された。また、自信度と正答率の間に相関関係はないことが明らかになった。

キーワード: 文章, 違和感, テキストマイニング, 共起語ネットワーク, 対応分析

1. 序論と目的

近年、機械翻訳や自動要約生成などの人工知能が作成する文章を見かける機会が増えている(Mori, Yamane, Makuta, & Harada, 2019; Li, Ding, & Liu, 2018)。しかし、これらの人工知能が作成する文章には、人間のみが感じる違和感が存在するという大きな課題があると考えられる。

そこで本研究では、テキストマイニングなどの分析方法を用いて、人間が文章を読む際に感じる違和感の要因を検討した。

2. 実験手順

実験参加者は合計 333 人(男性:247, 女性:86 人)であった。本実験では幅広い年齢層及び、性別を対象に調査するために、年齢・性別に制限を設けなかった。実験参加者は Yahoo!クラウドソーシングで募集し、オンラインで実験が行われた。

被験者は二つの文章が提示され、より違和感があると思う方の文章を選択した。その選択肢の一つは、過去に日本語検定問題に出題された問題文(以下、機械翻訳前の文章)であった。もう一方の選択肢は、機械翻訳前の文章を翻訳機によって英語に翻訳し、それを再び日本語に翻訳した文章(以下、機械翻訳後の文章)であった。

被験者は、この文章選択課題を全部で 25 試問繰り返した。なお、この 25 問が出題される順番は被験者によってランダム化されて行われた。その後、被験者が違和感を感じる要因を分析するため、「選択肢を選ぶ際、どのような点に着目したか、できる限り具体的に文章で記述してください。」と文字数制限なしで理由を聞いた。最後に、被験者は自身の解答への自信度を 7 段階尺度で評価した。

3. 研究 1

3.1. 目的

研究 1 の目的は実験参加者の正答率や自信度の妥当性、またはそれらの相関関係を定量的に評価することであった。また、各問題における正答率の特徴から、違和感を感じる際の要因の評価を行なった。

3.2. 結果と考察

正答率や自信度の妥当性 各問題における正答率を算出したところ、正答率の平均値が 69.57%、中央値が 72.7%、標準偏差が 16.69%という結果が得られた。また、自信度について、平均が 4.01、中央値が 4、標準偏差が 1.45 であった。これらのことから、難易度の観点や自信度の評価において妥当な結果が得られたと考えられた。

正答率と自信度の相関関係 自信度と正答率の相関係数は $r = 0.149$ であり、正答率と自信度との間で相関関係は見られなかった。よって、正答率と自信度の変数はそれぞれ独立していると考えられた。

問題文分析 正答率が統計的に著しく高い問題文、著しく低い問題文、中程度の問題文ごとに問題文の特徴の分析を行った。

正答率が著しく高い問題文 正答率が著しく高くなった理由は、いくつかの問題文の翻訳後の文章に文法的、意味的に誤りが存在したためと考えられる。また、他の

問題文において、単語レベルの違いや単語表現の硬さにおける単語表現の違いが原因で生じる違和感が見られた。

正答率が著しく低い問題文 表現の丁寧さの違いや、単語レベルの違いにおける単語表現の違いが原因で生じる違和感が見られた。

正答率が中程度の問題文 表現の丁寧さの違いや主語の有無、助詞の違いが原因で生じる違和感が見られた。

4. 研究 2

4.1. 目的と方法

研究 2 の目的は、実験参加者に聞いた「なぜそう答えたか」という理由の文章の分析を行い、違和感を感じる要因を評価することであった。また、実験参加者の正答率と解答への自信度の数値の違いにおいて、着目する要因がどのように変化するかを評価した。

これらの分析は、テキストマイニングを使用し、共起語ネットワークと、正答率及び自信度に関する対応分析によって評価した。

4.2. 結果と考察

共起語ネットワーク 本研究で得られた共起語ネットワークを図 1 に示した。この図にある単語が使われた実際の文章の内容から違和感を感じる要因を考察した。その結果、文章を読む際に感じる違和感には 6 つの要因があることが示唆された。その要因とは「文型による違和感」、「主語の有無による違和感」、「音韻による違和感」、「助詞による違和感」、「敬語、丁寧語による違和感」、「状況の想像の可否による違和感」である。

対応分析 正答率の高い実験参加者は順に「状況の想像の可否による違和感」、「音韻による違和感」、「文型による違和感」、「主語の有無による違和感」、「敬語、丁寧語による違和感」を重視しており、正答率の低い実験参加者は「助詞による違和感」を重視していたことが示唆された。また、解答に自信のある実験参加者は順に「状況の想像の可否による違和感」、「文型による違和感」、「主語の有無による違和感」、「助詞による違和感」によって判断する傾向があり、解答に自信のない実験参加者は順に「音韻による違和感」、「敬語、丁寧語による違和感」によって判断する傾向があると考えられた。

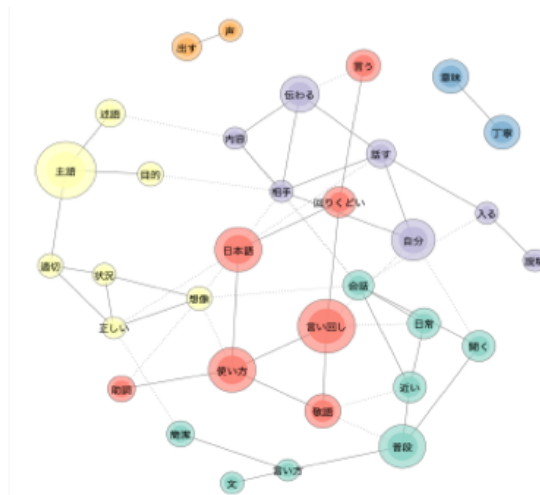


図 1 共起語ネットワーク

5. 考察

6 つの要因は大きく「文法が原因で生じる違和感」と「状況が原因で生じる違和感」の 2 つの要因に分けられる。具体的には、「文型による違和感」、「主語の有無による違和感」、「音韻による違和感」、「助詞による違和感」が「文法が原因で生じる違和感」であり、「敬語、丁寧語による違和感」、「状況の想像の可否による違和感」が「状況が原因で生じる違和感」である。

対応分析の結果から、文章を見分ける際の最も有効な着眼点は、「状況の想像の可否による違和感」と考えられる。また、「助詞による違和感」は適切な解答ができた手応えを得られるが、実際には見分けることあまり効果的でない着眼点であると考えられる。逆に、「音韻による違和感」に着目すると適切な解答ができた手応えは得られないが、効果的な着眼点であると考えられる。

今後の課題として、単語表現について指標を設けること、多くの種類の翻訳機を使用して機械翻訳の性能を一般化し実験を行うこと、問題文の原文を日本検定問題の文章を使用するのではなく、日常的に使われる文章を使用することなどが挙げられる。そのような検討を進めることで、より精度の高い、絶対的な評価が可能になることが期待される。

文献

- [1] Mori, Y., Yamane, H., Mukuta, Y., & Harada, Y. (2019). Toward a Better Story End: Collecting Human Evaluation with Reasons. Proceedings of the 12th International Conference on Natural Language Generation (pp. 383-390). Tokyo, Japan: Association for Computational Linguistics.