

令和元年5月30日現在

機関番号：13901

研究種目：研究活動スタート支援

研究期間：2017～2018

課題番号：17H06733

研究課題名（和文）文書構造と言語表現の対応づけによる多言語文書作成方式の構築と評価

研究課題名（英文）Construction and evaluation of multilingual document authoring frameworks based on the combination of document structures and linguistic expressions

研究代表者

宮田 玲（MIYATA, Rei）

名古屋大学・工学研究科・助教

研究者番号：70804300

交付決定額（研究期間全体）：（直接経費） 2,300,000円

研究成果の概要（和文）：本研究課題では、効果的な情報伝達を可能とする文書構造と言語表現を調査し、機械翻訳システムを用いた多言語文書作成方式を構築・検証した。主に自動車関連の技術文書を対象に、(1)書くべき内容とその順序を決める上で指針となる文書構造の設計とユーザ実験による文書品質の評価、(2)機械で翻訳しにくい日本語表現パターンの同定と制御ルールの作成、(3)文書中の役割に応じた望ましい表現パターンの整理を行った。

研究成果の学術的意義や社会的意義

学術的には、特定の文書ジャンルにおける文書構造（書かれる内容とその配置）を詳細に明らかにした点で、図書館情報学分野の基礎的な知見として意義がある。また読みやすく機械的に翻訳しやすい言語表現パターンを整理した本研究の成果は、テクニカルライティング研究や機械翻訳応用研究に貢献するものである。社会的には、本研究で定義した自動車関連のマニュアルの文書構造や日本語表現パターンは、技術文書の執筆場面への応用が期待される。

研究成果の概要（英文）：We aimed to construct and evaluate the multilingual document authoring frameworks that make use of machine translation systems, through formalising desirable document structures and linguistic expressions. The main outcomes are summarised as follows: (1) focusing on the technical documents of automobile domain, we formalised a document structure for repair procedures and evaluated the quality of documents through user evaluation experiments, (2) we identified Japanese linguistic patterns that are difficult to translate and created controlled language rules to regulate such patterns, and (3) we collected linguistics expressions linked to the document structure and defined desirable linguistic patterns for each of the document elements.

研究分野：図書館情報学

キーワード：文書構造 制限言語 前編集 機械翻訳 文書評価

1. 研究開始当初の背景

オンラインでの電子テキストの流通と機械翻訳などの言語ツールの進歩により、多言語での情報発信が容易になりつつある。このような技術的背景を生かしながら、品質の高い原文書ならびに翻訳文書を効率的に作成するためには、(a)効果的な情報の提示を可能とする文書構造の定式化、(b)読みやすく機械的に処理しやすい言語表現パターンの定義が有効である。

(a)に関して、文書デザインやテクニカルライティングの分野では、一般的な文書構成法の研究がなされてきた (Schriver, 1997; Rubens, 2001)。しかし、特定の文書ジャンルにおける、詳細な文書構造の知見は、現場の暗黙知に留まりがちである。(b)に関して、原文の語彙・文法・スタイルに一定の制約を加える制限言語やプリエディットの研究が進められてきた (Kuhn, 2014)。欧米では企業での実用化も進み、近年は日本語を対象とした試みも増えている。

研究代表者はこれまで自治体手続きに関する文書を対象に、読みやすかつ(機械)翻訳しやすい構造化文書を作成する枠組みに関する研究を進めてきたが (Miyata et al. 2015; Miyata et al. 2016)、以下の課題が残されている。

- (1) 自治体手続きの文書以外の文書ジャンル・文書タイプに対して、どの程度文書構造化が可能か十分検証されていない。
- (2) 最新の機械翻訳方式に対する制限言語の有効性が十分調査されていない。
- (3) これまで開発されてきた制限言語は、センテンスレベルの言語表現操作を主眼とし、文書構造に応じて精緻化されていない。

2. 研究の目的

本研究の目的は、効果的な情報伝達を可能とする文書構造と言語表現を解明し、機械翻訳システムを用いた多言語文書作成方式を構築・検証することである。具体的には、主に自動車関連の技術文書(修理書)を対象に、(1)文書構造の定式化と評価、(2)機械翻訳困難表現の整理と制限言語ルールの作成、(3)文書構造と言語表現の対応付けを行う。

3. 研究の方法

- (1) 文書構造の定式化と評価：手作業のコーディングにより、文書に書かれている内容を抽象化したコード(内容要素)を、最小単位を節として、対象テキストに手作業で付与する。さらに内容要素を文書規格 Darwin Information Typing Architecture (DITA) にマッピングすることで、文書構造の定式化を行う。また文書構造化の有効性をユーザ参加型の文書読解実験により評価する。
- (2) 機械翻訳困難表現の整理と制限言語ルールの作成：日本語原文を機械翻訳にかけ、そのエラーを分析する。翻訳結果がよくなるように原文を書き換えることで、翻訳しにくい表現パターンと翻訳しやすい表現パターンを対で収集し、ルール化する。
- (3) 文書構造と言語表現の対応付け：内容要素ごとにテキストを収集・分析し、望ましい表現パターンを定義する。

4. 研究成果

- (1) 文書構造の定式化と評価
 - 文書構造の定式化：自動車の修理手続きに関する文書を対象に、どのような内容がどのような順序で配置されているかを整理した。合計 2907 個のテキストスパンに対して割り当てられたコードは、26 種類の内容要素にまとまった。
 - 文書品質の評価：構築した文書構造規格が、実際の読み手の理解を促進するものであるかを、ユーザ評価により検証した。文書構造規格による書き換え前と書き換え後の文書を用いて、人間の文書読解プロセスを分析した。
- (2) 機械翻訳困難表現の整理と制限言語ルールの作成
 - 機械翻訳困難表現の整理：最新の機械翻訳方式であるニューラル機械翻訳(日英翻訳を対象)の翻訳エラーを分析しながら、翻訳困難箇所の同定・書き換え方法の考案を行った。約 80 タイプの翻訳困難表現を同定した。
 - 制限言語ルールの作成：上記の整理を踏まえつつ、読みやすさと(機械)翻訳しやすさの向上のために、構文・語彙・表記レベルで望ましい(あるいは望ましくない)表現パターンを明文化したルール集の作成を進めた。複雑な構文や曖昧な表現を規制するルールを 10 カテゴリーにわたり約 150 種類作成した。それらの一部については、自動的に検出する方法及び自動的に翻訳しやすい形に書き換える方法を検討した。

(3) 文書構造と言語表現の対応付け

- 内容要素分類器の作成：(1)で整理した内容要素ごとの自動分類器を作成し、約16万文のテキストデータを対象に、内容要素ごとの言語表現を収集した。
- 表現パターンの整理：内容要素ごとに手作業及び自動で収集した言語表現の分析を行った。例えば、文書中では、文末において「～する」「～すること」の形の混在が見られた。「動作」を示す場合は「～する」、「徹底事項」を示す場合は「～すること」を用いるなど、文書内容に応じた望ましい表現パターンの定義を部分的に進めた。

<引用文献>

- Kuhn, T. A Survey and Classification of Controlled Natural Languages, *Computational Linguistics*, 40(1): 121–170, 2014.
- Miyata, R. et al. Japanese Controlled Language Rules to Improve Machine Translatability of Municipal Documents, *MT Summit XV*, 90–103, 2015.
- Miyata, R. et al. MuTUAL: A Controlled Authoring Support System Enabling Contextual Machine Translation, *COLING, System Demonstrations*, 35–39, 2016.
- Rubens, P. *Science and Technical Writing: A Manual of Style*. Routledge, 2001.
- Schrivver, K. *Dynamics in Document Design: Creating Text for Readers*. John Wiley & Sons, 1997.

5 . 主な発表論文等

〔雑誌論文〕(計 4 件)

- Rei Miyata, Kyo Kageura. Building Controlled Bilingual Terminologies for the Municipal Domain and Evaluating Them Using a Coverage Estimation Approach, *Terminology*, Vol.24, No.2, pp.149-180, 2018. 査読有.
- 宮田玲, 藤田篤. 「機械翻訳向けプリエディットの有効性と多様性の調査」通訳翻訳研究への招待, No.18, pp.53-72, 2017. 査読有.
- 宮田玲. 「翻訳テクノロジーを学ぶ：プリエディット編その2」*JTF Journal*, No.292, pp.26-27, 2017. 査読無.
- 宮田玲. 「翻訳テクノロジーを学ぶ：プリエディット編その1」*JTF Journal*, No.291, pp.26-27, 2017. 査読無.

〔学会発表〕(計 5 件)

- 宮田玲, 柳英夫, 影浦峯, 萩原秀章. 「技術文書の多言語化を見据えた制限オーサリングと翻訳：基本方針と枠組み」言語処理学会第25回年次大会発表論文集, pp.918-921, 名古屋, 2019.
- 杉野峰大, 宮田玲, 松崎拓也, 佐藤理史. 「文書モデルの作成と文書診断に向けた予備的考察：自動車の修理書を対象に」言語処理学会第25回年次大会, pp.719-722, 名古屋, 2019.
- 宮田玲. 「プリエディットによる機械翻訳活用の試み」日本通訳翻訳学会第19回年次大会・公開シンポジウム「翻訳におけるテクノロジーを考える」, 大阪, 2018. 招待有.
- Rei Miyata, Anthony Hartley, Kyo Kageura, Cécile Paris. MuTUAL: A Controlled Authoring System for Municipal Text Multilingualisation Using Machine Translation, *2017 Portsmouth Translation Conference: Translation and Disruption: Global and Local Perspectives*, Portsmouth, UK, 2017.
- 宮田玲. わかりやすく翻訳しやすい日本語を書くために ～制限言語という考え方を理解し、ツールを活用して翻訳しやすい日本語を書く～. テクニカルコミュニケーション(TC)シンポジウム, 京都, 2017. 招待有.

〔図書〕(計 0 件)

〔産業財産権〕

出願状況(計 0 件)

名称：
発明者：
権利者：
種類：
番号：

出願年：
国内外の別：

取得状況（計 0 件）

名称：
発明者：
権利者：
種類：
番号：
取得年：
国内外の別：

〔その他〕
ホームページ等

6. 研究組織

(1)研究分担者
なし

(2)研究協力者
なし

科研費による研究は、研究者の自覚と責任において実施するものです。そのため、研究の実施や研究成果の公表等については、国の要請等に基づくものではなく、その研究成果に関する見解や責任は、研究者個人に帰属されます。