

平成 22 年 4 月 6 日現在

研究種目：基盤研究(C)

研究期間：2007～2010

課題番号：19500133

研究課題名（和文） 日英混在型入力による動的文脈照合英作文システムに関する研究

研究課題名（英文） Development of English-Writing Support Systems

研究代表者

馬 青 (MA SEI)

龍谷大学・理工学部・教授

研究者番号：30358882

研究代表者の専門分野：情報科学

科研費の分科・細目：情報学・知能情報学

キーワード：英作文支援、日英混在文、高品質コーパス、大規模 Web データ、柔軟な検索、訳語選択、対訳表現抽出、対訳コーパス

### 1. 研究計画の概要

本研究はユーザが英作文する過程において、部分的に適切な英語表現（単語、フレーズ、または節など）が思い浮かばないとき、本来言い表したい日本語表現をそのまま入力するだけで、すなわち、日英混在の入力文から状況に応じて動的に構成される文脈情報と、高品質で大規模な英語コーパスに加え超大規模な Web データも統合的に利用することにより、適切な英語表現を自動生成する支援システムの研究開発を目指すものである。

### 2. 研究の進捗状況

初年度の 19 年度では最適な文脈による訳語選択手法と高品質コーパスと大規模 Web データの統合利用手法の提案と、柔軟な検索システムと単語レベルの支援システムの開発を行い、得られた訳語が原文の単語と同じもののみを正解とする評価基準でシステムを評価した。20 年度では日本語の訳語は 1 つとは限らず複数ある場合が多いことを考慮し、プロの日英翻訳者による評価データの作成とシステムの再評価を行った。その結果、統合手法においては 80%の正解率が得られ、システムの有効性が実証された。また、英語フレーズを訳語候補の組み合わせから生成することにより、英作文支援の対象を単語からフレーズに拡張し、名詞句と動詞句も取り扱える英作文支援システムを構築した。実験の結果、名詞句の英作文支援が 71%、動詞句のそれが 67%という正解率が得られた。しかし、英語フレーズは訳語候補の組み合わせのみから生成されるため、支援できるフレーズ

の範囲が大きく限定されてしまう。また、名詞句などを不定詞句に英訳するような、異なるパターンへの英訳も支援できない。さらに、単語の訳語候補の数が多いためその組み合わせの数が膨大となり処理時間がかかってしまう。21 年度ではこのような問題を解決するために日英対訳パターンに基づくアプローチを導入し、そのパターン辞書を作成するにあたり必要となる大規模な日英対訳表現を大規模な日英対訳コーパスから抽出することを試みた。実験の結果、数十万文の大規模な日英対訳コーパスに対し、数十万の単語 n-gram または文節 n-gram の対訳表現を良好な精度で抽出することができた。

### 3. 現在までの達成度

おおむね順調に進展している。

（理由）

各年度は計画に沿って研究を実施してきており、おおむね満足のできる研究結果を得ることができた。ただし、フレーズレベルについて、より高性能な翻訳支援を目指すため、当初の研究計画になかった対訳パターンに基づくアプローチを新たに導入した。

### 4. 今後の研究の推進方策

前年度では日英対訳パターン辞書を作成するにあたり必要となる大規模な日英対訳表現を大規模な日英対訳コーパスから抽出することを試みたが、長い対訳文をできるだけ多く抽出しようとする対訳表現の抽出数が大幅に減少してしまう問題が生じている。今後は

(1) 長い対訳表現の抽出を保証しながら対訳表現の抽出数の増加と抽出精度の向上を図る。

(2) 対訳関係が離散型の対訳表現も含め、より多くの対訳表現の抽出に取り組んでいく。

(3) 抽出された対訳表現から対訳パターンの辞書を作成し、英作文支援システムに組み込み、当初の目標である、単語レベルから句レベルまでの高性能英作文支援システムを構築する。

というような方策で研究を推進していく。

## 5. 代表的な研究成果

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文](計 10 件)

吉見、以下 4 名：単語アライメントを用いた英日機械翻訳文の流暢さの自動評価、自然言語処理、Vol.17, No.1, pp. 7-28, 2010, 査読有

吉見、以下 3 名：単語対応付けに基づく日本語学習者による作文の自動識別、情報処理学会論文誌(テクニカルノート) Vol. 49, No. 12, pp. 4039-4043, 2008, 査読有

村田、馬、以下 3 名：Japanese-to-English Translations of Tense, Aspect, and Modality Using Machine-learning Methods and Comparison with Machine-translation Systems on Market, Language Resources and Evaluation, Vol. 40, pp. 233-242, 2007, 査読有

[学会発表](計 25 件)

坂上、馬、村田：英作文支援のための大規模な日英対訳表現の抽出、言語処理学会第 16 回年次大会、2010 年 3 月 10 日、東京大学

馬、他 2 名：Development of English-Writing Supporting Systems, 11th Conference of the Pacific Association for Computational Linguistics (Pacling2009), 2009.9.3, 北海道大学

日高、渡辺、以下 2 名：Learning Support for English Composition by Asking Back Questions, IEEE CSEDU 2009, 2009.3.25, Lisbon, Portugal

馬、以下 3 名：Selection of Japanese-English Equivalents by Integrating High-quality Corpora and Huge Amounts of Web Data, The sixth international conference on Language Resources and Evaluation (LREC2008), 2008.5.28, Marrakech, Morocco

Zhang, 馬 (4 番目)、他 3 名：Word

Alignment Annotation in a Japanese-Chinese Parallel Corpus, The sixth international conference on Language Resources and Evaluation (LREC2008), 2008.5.28, Marrakech, Morocco