

論文 / 著書情報
Article / Book Information

題目(和文)	特許翻訳のためのバイリンガル知識の獲得に関する研究
Title(English)	
著者(和文)	田村晃裕
Author(English)	Akihiro Tamura
出典(和文)	学位:博士(工学), 学位授与機関:東京工業大学, 報告番号:甲第9327号, 授与年月日:2013年9月25日, 学位の種別:課程博士, 審査員:奥村 学,小林 隆夫,住田 一男,熊澤 逸夫,篠崎 隆宏,高村 大也
Citation(English)	Degree:Doctor (Engineering), Conferring organization: Tokyo Institute of Technology, Report number:甲第9327号, Conferred date:2013/9/25, Degree Type:Course doctor, Examiner:,,,,,
学位種別(和文)	博士論文
Category(English)	Doctoral Thesis
種別(和文)	審査の要旨
Type(English)	Exam Summary

(博士課程)

論文審査の要旨及び審査員

(2000 字程度)

報告番号	甲 第 号	氏 名	田村 晃裕			
論文審査	氏 名		氏 名		職 名	
審査員	主 査	奥村 学	教授	審査員	篠崎 隆宏	准教授
	審査員	小林 隆夫	教授		高村 大也	准教授
		住田 一男	教授			
		熊澤 逸夫	教授			

機械翻訳とは、翻訳を機械的に行うことである。近年、企業のグローバル化に伴い、自国以外でも知的財産を守る必要性が高まり、外国語の特許調査や特許出願を行う機会が増えている。そのため、それらを効率よく行う手助けとなる特許の機械翻訳（特許翻訳）のニーズが高まり、特許翻訳の性能向上が求められる。本論文は、「特許翻訳のためのバイリンガル知識の獲得に関する研究」と題し、和文全 5 章より構成されている。

第 1 章「序論」では、特許翻訳の性能向上を達成するために解決すべき問題として、翻訳できない未知語が多いという問題と、特許翻訳の手がかりとして使われる品詞は翻訳に最適とは限らないという問題があることを説明し、本研究で解くべき課題を示している。そして、未知語に対する翻訳対と翻訳に適した品詞の二つのバイリンガル知識をコーパスから獲得することにより、それぞれの問題を解決するという本研究の目的を示している。

第 2 章「関連研究」では、本研究の関連研究について述べている。最初に、コーパスからの翻訳対抽出に関する従来手法を紹介している。その中で、本研究では、手法の適用範囲が広いコンパラブルコーパスからの翻訳対抽出に着目することを述べ、従来のコンパラブルコーパスからの翻訳対抽出の代表的な手法として、文脈類似度に基づく手法を紹介している。そして、その従来手法は、翻訳関係を特定する際に使われる既存の翻訳対（シード翻訳対）が小規模だと性能が悪いという問題があることを説明している。次に、コーパスからの品詞導出に関する従来手法を紹介している。その中で、従来手法は、翻訳相手の言語を考慮して品詞を導出しないため、機械翻訳に最適な品詞を導出できるとは限らないという問題があることを説明している。

第 3 章「ラベル伝播によるコンパラブルコーパスからの翻訳対抽出」では、ラベル伝播を利用してコンパラブルコーパスから翻訳対を抽出する手法を提案している。提案手法は、従来手法では捉えられない、シード翻訳対との間接的な関連性をラベル伝播により獲得し、全てのシード翻訳対との関連性が似た単語対を翻訳対として抽出することで、第 2 章で説明されている従来の文脈類似度に基づく手法の問題点を解決している。そして、日本語と英語の特許文書から作成したコンパラブルコーパスを用いた評価を行い、提案手法は、従来の文脈類似度に基づく手法の約 2 倍の精度で翻訳対を抽出できることを示し、シード翻訳対との間接的な関係は翻訳対抽出において有効な手がかりであることを示している。また、ラベル伝播で用いるグラフとして、同一文脈での共起関係に基づき構築する「共起グラフ」と、文脈の類似関係に基づき構築する「類似グラフ」の二つのグラフを提案している。そして、評価と考察を通じて、類似グラフは、単語間の偶然の共起関係がもたらす悪影響を緩和すると共に、同義語を巧みに同一視することで、共起グラフを用いた場合よりも更に高い抽出精度を達成できることを示している。

第 4 章「機械翻訳のための品詞導出」では、従来の単言語における無限ツリーモデルを多言語に拡張した、コーパスから翻訳のための品詞を導出する手法を提案している。提案手法は、原言語と目的言語の単語間の対応関係に基づいて翻訳相手の言語の情報をシンボルに取り入れ、両言語の情報を持つバイリンガルなシンボルから品詞を導出することにより、従来の品詞導出手法では導出できない、翻訳相手の言語における違いを反映したバイリンガルな品詞を導出している。具体的なモデルとして、原言語の情報と目的言語の情報を一つのシンボルとして生成する「結合モデル」と、それらの情報を独立に別々のシンボルとして生成する「独立モデル」を提案している。そして、日英特許翻訳における評価を行い、提案手法が導出した品詞を使うことにより、既存の品詞や従来手法が導出した品詞を使うよりも (BLEU (Bilingual Evaluation Understudy) で 1% 以上) 精度良く翻訳できることを示している。また、独立モデルは、結合モデルで生じるシンボルのスパースネス問題を解決し、結合モデルよりも翻訳により適した品詞を導出できることを示している。

第 5 章「結論」では、本研究の結論と今後の課題を述べている。
以上を要するに、本論文は、特許翻訳の性能を向上させるため、未知語に対する翻訳対、翻訳に適した品詞の二つのバイリンガル知識をコーパスから獲得する手法を提案した上、提案手法の有効性の検証を行っており、工学上貢献するところが大きい。よって、博士(工学)の学位を授与するに十分な価値をもつものと認められる。