

様 式 F-7-1

科学研究費助成事業（学術研究助成基金助成金）実施状況報告書（研究実施状況報告書）（平成26年度）

1. 機関番号

1	4	6	0	3
---	---	---	---	---

 2. 研究機関名 奈良先端科学技術大学院大学

3. 研究種目名 若手研究(B) 4. 補助事業期間 平成25年度～平成27年度

5. 課題番号

2	5	7	3	0	1	3	6
---	---	---	---	---	---	---	---

6. 研究課題名 訳選択の根拠の自動推定とその機械翻訳における応用

7. 研究代表者

研究者番号	研究代表者名	所属部局名	職名
7 0 6 3 3 4 2 8	ニュービグ グラム Neubig Graham	情報科学研究科	助教

8. 研究分担者

研究者番号	研究分担者名	所属研究機関名・部局名	職名

9. 研究実績の概要

本年度は主に訳選択の根拠の自動発見技術と、訳選択の根拠を考慮した翻訳システムの構築に取り組み3つの成果があった。

成果1は、現在の翻訳システムが引き起こす誤りを効率的に分析する手法である。具体的な手法として、システムが出力した訳出結果と実際に人間が作成した正解訳を比較し、体系的に異なる箇所を誤りの可能性が高いとして指摘する。評価の結果、従来の分析法に比べて、効率的に誤りを発見することができたことが分かった。この技術を成果2の分析の際に使用している。

成果2は、2種類の訳選択誤りの分析である。1種類目は、翻訳システムによって1つの単語の訳を生成させ、翻訳者に分析させたものである。2種類目は実際に様々なシステムに文を生成させ、その誤りを分析したものである。この分析の結果、訳選択の根拠や、訳選択が実際のシステムで失敗するケースを明らかにした。具体的には、語義曖昧性による訳選択の失敗は述語と項の関係や、口語か文語などのスタイルが重要な要因であることが分かった。

成果3は、ルールベースシステムが使っているような処理を統語情報を用いた統計翻訳への反映手法である。具体的には「は」「が」「を」などの助詞を選択する根拠を中心に、人間により記述したルールを導入した手法1つと、対訳データから自動学習した手法1つを調査した。具体的には、名詞句が主語を表しているか、目的語を表しているかを構文解析の結果に基づいて特定し、適切な助詞を入れやすくするように統計翻訳の入力を編集するルールを作成した。システムへの導入の結果、英日翻訳において、助詞を翻訳する精度が大幅に向上した。

10. キーワード

(1) 機械翻訳	(2) 訳選択	(3) 自然言語処理	(4) 機械学習
(5) 評価尺度	(6)	(7)	(8)

11. 現在までの達成度

(区分) (2) おおむね順調に進展している。
(理由) 26年度の目標はおおむね順調に達成している。 訳選択の根拠の調査に関して、2種類の調査を行い、有用な知見を獲得した。その中で、語義の曖昧性、原言語に含まれない単語の誤訳、単語の誤削除などを問題として特定し、これらの問題の解決策について考察した。 訳選択の根拠の自動発見技術と翻訳への反映に関して、英日翻訳における助詞の選択という、原言語に含まれない情報の代表において、確実な成果を上げている。従来の統語情報に基づく統計的機械翻訳と比較して、大幅な精度向上を実現しており、有用性が確認されている。 また、当初の予定に含まれていなかったが、分析を行っている内に必要性が明らかとなった誤り分析の効率化技術に関する研究も行われた。

12. 今後の研究の推進方策 等

(今後の推進方策) 27年度において、主に2つの課題に取り組む予定である。 まずは、訳選択の根拠の自動発見と翻訳システムへの反映を更に一般化する予定である。26年度で英日翻訳の助詞の翻訳で有用性が確認された枠組みを、更に文脈の結果で明らかになった他の誤り原因へと拡張する予定である。例えば、名詞や動詞の語彙選択や、主語の補完などが挙げられる。 また、27年度の予定であった訳選択の安定性を評価する評価尺度の開発にも取り組む。具体的には、訳選択の安定性を評価するために作られたデータに基づく尺度と、自然発生する翻訳データに基づく尺度という2種類の尺度を検討する。
(次年度使用額が生じた理由と使用計画) (理由) 若干の残高が残ったが、ほぼ予定通りに予算を執行している。 (使用計画) 今年度は当初の予定通り、システムの評価やデータの作成、成果発表のための旅費などを中心に、従来の予算と合わせて使用していく予定である。

13.研究発表(平成26年度の研究成果)

(雑誌論文) 計(5)件 うち査読付論文 計(5)件

著 者 名		論 文 標 題			
Graham Neubig, Kevin Duh		On the Elements of an Accurate Tree-to-String Machine Translation System			
雑 誌 名		査読の有無	巻	発行年	最初と最後の頁
Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (ACL)		有	-	2 0 1 4	143-149
掲載論文のDOI (デジタルオブジェクト識別子)					
なし					

著 者 名		論 文 標 題			
Koichi Akabe, Graham Neubig, Sakriani Sakti, Tomoki Toda, Satoshi Nakamura		Discriminative Language Models as a Tool for Machine Translation Error Analysis			
雑 誌 名		査読の有無	巻	発行年	最初と最後の頁
Proceedings of the 25th International Conference on Computational Linguistics (COLING)		有	-	2 0 1 4	1124-1132
掲載論文のDOI (デジタルオブジェクト識別子)					
なし					

著 者 名		論 文 標 題			
Graham Neubig		Forest-to-String SMT for Asian Language Translation: NAIST at WAT 2014			
雑 誌 名		査読の有無	巻	発行年	最初と最後の頁
Proceedings of the 2014 Workshop on Asian Translation		有	-	2 0 1 4	20-25
掲載論文のDOI (デジタルオブジェクト識別子)					
なし					

著 者 名		論 文 標 題			
Yuto Hatakoshi, Graham Neubig, Sakriani Sakti, Tomoki Toda, Satoshi Nakamura		Rule-based Syntactic Preprocessing for Syntax-based Machine Translation			
雑 誌 名		査読の有無	巻	発行年	最初と最後の頁
Proceedings of the 8th Workshop on Syntax, Semantics and Structure in Statistical Translation (SSST).		有	-	2014	34-42
掲載論文のDOI (デジタルオブジェクト識別子)					
なし					

著 者 名		論 文 標 題 【掲載確定】			
Graham Neubig, Philip Arthur, Kevin Duh		Multi-Target Machine Translation with Multi-Synchronous Context-free Grammars			
雑 誌 名		査読の有無	巻	発行年	最初と最後の頁
Proceedings of the Meeting of the North American Chapter of the Association for Computational Linguistics (NAACL)		有	-	2015	-
掲載論文のDOI (デジタルオブジェクト識別子)					
なし					

(学会発表) 計(9)件 うち招待講演 計(2)件

発 表 者 名		発 表 標 題	
赤部 晃一, Graham Neubig, Sakriani Sakti, 戸田 智基, 中村 哲		機械翻訳システムの詳細な誤り分析のための誤り順位付け手法	
学 会 等 名		発 表 年 月 日	発 表 場 所
情報処理学会 第216回自然言語処理研究会(SIG-NL)		2014年05月22日	東京工業大学 (東京都目黒区)

発 表 者 名		発 表 標 題	
Graham Neubig		機械翻訳 ～なぜできなかったのか？なぜできるようになりつつあるのか？～	
学 会 等 名		発 表 年 月 日	発 表 場 所
音学シンポジウム (招待講演)		2014年05月25日	日本大学 (東京都世田谷区)

発 表 者 名	発 表 標 題	
波多腰 優斗, Graham Neubig, Sakriani Sakti, 戸田 智基, 中村 哲	統語ベース翻訳に対する統語的前処理の適用	
学 会 等 名	発表年月日	発 表 場 所
情報処理学会 第217回自然言語処理研究会(SIG-NL)	2014年07月03日	オホーツク文化交流センター (北海道網走市)

発 表 者 名	発 表 標 題	
赤部 晃一, Graham Neubig, Sakriani Sakti, 戸田 智基, 中村 哲	パラフレーズを考慮した機械翻訳の誤り箇所選択	
学 会 等 名	発表年月日	発 表 場 所
情報処理学会 第219回自然言語処理研究会(SIG-NL)	2014年12月17日	東京工業大学 (神奈川県横浜市)

発 表 者 名	発 表 標 題	
Graham Neubig	日本語を対象とした統計的機械翻訳の進展	
学 会 等 名	発表年月日	発 表 場 所
第6回産業日本語研究会・シンポジウム (招待講演)	2015年02月24日	東京大学 (東京都文京区)

発 表 者 名	発 表 標 題	
波多腰 優斗, Graham Neubig, Sakriani Sakti, 戸田 智基, 中村 哲	Tree-to-String 翻訳における構文解析器の自己学習の効果	
学 会 等 名	発表年月日	発 表 場 所
言語処理学会第21回年次大会(NLP2015)	2015年03月18日	京都大学 (京都府京都市)

発 表 者 名	発 表 標 題	
Graham Neubig, Philip Arthur, Kevin Duh	複数の目的言語の同時生成による統計的機械翻訳	
学 会 等 名	発表年月日	発 表 場 所
言語処理学会第21回年次大会(NLP2015)	2015年03月19日	京都大学 (京都府京都市)

発 表 者 名	発 表 標 題	
赤部 晃一, Graham Neubig, Sakriani Sakti, 戸田 智基, 中村 哲	機械翻訳の誤り箇所選択法における誤選択箇所の調査	
学 会 等 名	発表年月日	発 表 場 所
言語処理学会第21回年次大会(NLP2015)	2015年03月19日	京都大学 (京都府京都市)

発 表 者 名	発 表 標 題	
赤部 晃一, Graham Neubig, 工藤 拓, John Richardson, 中澤 敏明, 星野 翔	Project Nextにおける機械翻訳の誤り分析	
学 会 等 名	発表年月日	発 表 場 所
エラー分析ワークショップ	2015年03月20日	京都大学 (京都府京都市)

〔図書〕 計 (0) 件

著 者 名	出 版 社		
書 名	発行年	総ページ数	

14.研究成果による産業財産権の出願・取得状況

(出願) 計(0)件

産業財産権の名称	発明者	権利者	産業財産権の種類、番号	出願年月日	国内・外国の別

(取得) 計(0)件

産業財産権の名称	発明者	権利者	産業財産権の種類、番号	取得年月日	国内・外国の別
				出願年月日	

15.備考

本研究の一環として開発されたTravatar Toolkitの日英翻訳デモページ
<http://ahclab.naist.jp/travatar/translator/>
 ACL2014で発表した成果の概要を説明するウェブページ
<http://phontron.com/project/neubig14acl/>
 NAACL2015で発表した成果の概要を説明するウェブページ
<http://phontron.com/project/naacl2015/>