様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９　（共通）

# 科学研究費助成事業　　研究成果報告書

令和　５ 年　 ４ 月　 ６ 日現在

機関番号：３４５０９
研究種目：基盤研究(C)（一般）
研究期間：2020～2022
課題番号：２０Ｋ００６９２

研究課題名（和文）日英・英日パラレルコーパスの整備と検索システムの開発およびその活用法

研究課題名（英文）The development of parallel Japanese-English and English-Japanese corpora and their searching system, and its application

研究代表者

仁科　恭徳（Yasunori, Nishina）

神戸学院大学・グローバル・コミュニケーション学部・教授

研究者番号：００５７２７７８

交付決定額（研究期間全体）：（直接経費）　　3,200,000円

研究成果の概要（和文）：　まず、2000年以降に無償で公開された9種の既存の日英・英日パラレルコーパスのフォーマットを統一し、串刺し検索できるように整備した。重複が見られるノイズ等についても削除し、その結果、英語・日本語それぞれ4000万語弱のパラレルコーパスが完成した。
　次に、これら整備した9種の日英・英日パラレルコーパスを瞬時に検索できるワードプロファイラーを開発した。Ver.1.0, 1.1, 1.2では、日本語から英語の翻訳方向で検索できるシステムを開発した。起点言語となる日本語のパターン検索機能、コロケーション抽出機能、用例+対訳の表示機能などを搭載 し、ジャンル別の用例数も表示を可能にした。

研究成果の学術的意義や社会的意義
研究成果の学術的意義は、現在までに日本のみならず世界においても存在していなかった複数のパラレルコーパス（4000万語弱）を串刺し検索できるワードプロファイラーを開発し、無料一般公開したことにある。初心者にも検索しやすいようにユーザーフレンドリーなインターフェースの開発を心掛けた。ユーザーは日本国内のみならず、日本語を学ぶ欧米諸国（特に、イギリス）にも見られた。翻訳・通訳支援、対照言語学/翻訳学研究の用途のみならず、国内外で日本語学を専攻する大学生や大学院生にもニーズがあったことは驚きであり、学問分野の枠を超えて幅広い層で活用いただいている点こそが、この研究成果の社会的意義であると言える。

研究成果の概要（英文）：　First, the formats of nine pre-existing Japanese-English and English-Japanese parallel corpora released for free since 2000 were converted to a unified format so that they can be skewered and searched. Noise and other duplications were also removed, resulting in a parallel corpus of about nearly 40 million words each in English and Japanese.
　Next, a word profiler was developed to search these nine Japanese-English and English-Japanese parallel corpora easily and automatically, and in Ver. 1.0, 1.1 and 1.2, a system was developed to search in the direction of translation from Japanese to English. The system includes a pattern search function for Japanese as the source language, a collocation extraction function and a function for displaying examples plus translations, and can also display the number of examples for each genre.

研究分野：コーパス言語学、辞書学、翻訳学、意味論、応用言語学

キーワード：パラレルコーパス　レキシカルプロファイラー　辞書学　翻訳ユニット　コーパス言語学　意味論　ツール開発　翻訳学

様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９（共通）

## １．研究開始当初の背景

パラレルコーパスの分析や活用がもたらす恩恵は計り知れないものがあるが、実際には同分野の言語学的研究は 10 年以上滞っている状況にあった。機械翻訳・自然言語処理の分野を除き、日英・ 英日パラレルコーパス(別名、翻訳コーパス)を用いた言語研究は、記述的・辞書学的に分析した田中(2002)、清水・村田(2002)、仁科(2007, 2008, 2009)、Nishina(2008)、染谷・赤瀬川・山岡(2011)などに限られている。英語教育分野においては、Data Driven Learning (以下、 DDL)の有効性を検証するために、中條ほか(2014, 2015)などでオンライン検索ツールが開発されているが、あくまで教育目的であること、搭載されている日英パラレルコーパスが 2003 年に構築されたものであることなど、言語学的視点から見れば英日・日英パラレルコーパスの構築およびその研究が進展しているとは言い難い。

また、British National Corpus(以下、BNC)や現代日本語書き言葉均衡コーパス(Balanced Corpus of Contemporary Written English、以下 BCCWJ)など、ここ 20 年の間に英語や日本語の一般参照コーパスが構築され、研究や教育、ツール開発など多方面で活用されている。しかしながら、パラレルコーパスにおいては同じ類のコーパスは皆無である。擬似的な一般参照パラレルコーパスが存在すれば、今まで成し得なかった言語分析が可能となり、新しい言語事実の発見が可能となる。

## ２．研究の目的

本研究の目的は、研究者および一般ユーザーへパラレルコーパスの利用を普及させ、研究・教育目的の点からその価値を高めることにある。これを半自動的に可能にするため、本研究ではオンライン上でアクセス可能な検索システムの開発を目指す。そして、その活用方法については、ワークショップや事例研究などを通して示す。特にシステム開発については、版権や人手などの点からサイズや種類、時代を統制した上でパラレルコーパスを一から構築するのは現実的に難しいため、代わりに既存の複数のパラレルコーパスのフォーマットやジャンルを整備し、日・英語の双方から検索できるシステムを開発することで、未だ存在していない擬似的な一般参照パラレルコーパスの完成を目指す。

## ３．研究の方法

本研究の計画・方法は以下の 4 段階から構成される。

(1) 準備段階:国内外のパラレルコーパスおよびコンコーダンサーを含めた検索ソフト・システムの種類や特徴(長所と短所)を精査する。また、現在までにパラレルコーパスを用いて実施された研究の種類・分野・成果を精査し、今後、解明・開発が期待される点を絞り出す。 そして、既存の英日・日英パラレルコーパスの保存形式やアライメント等を統一・整備し、特定のコンコーダンサーで一覧検索できる状態にする。

(2) 開発段階:(1)で整備した複数の日英・英日パラレルコーパスを様々な観点から一覧検索できるオンライン検索システムを開発する。現段階では、BCCWJ に実装されている LWP の改良版の開発を予定している。

(3) 分析段階:(1)(2)で整備した日英・英日パラレルコーパスおよび開発した改良版 LWP を用いて、比較言語学・記述言語学・翻訳学の見地から言語分析を実施する。

(4) 報告段階:(1)(2)(3)で得られた研究成果を学会発表や学術論文を通じて、公表する。

## ４．研究成果

研究期間全体を通じて実施した研究の成果については、以下のとおりである。

(1) 2000 年以降に発表・公開された 13 種前後の日英・英日パラレルコーパスのうち、無償で公開されている 9 種の既存のコーパスについて、アラインメント、フォーマット、保存形式などを統一し、串刺し検索できるように整備した。また、重複が見られるノイズ等についても削除した。出来上がったコーパスは、複数のジャンルから構成される英語・日本語それぞれ 4000 万語弱規模のパラレルコーパスとなった。

(2) これら整備した 9 種の日英・英日パラレルコーパスを瞬時に詳細に検索できる LWP(Lago Word Profiler)の改良版を開発した。Ver.1.0, 1.1, 1.2 では、日本語から英語の翻訳方向で検索できるシステムを開発した。起点言語となる日本語のパターン検索機能、コロケーション抽出機能、用例+対訳の表示機能などを搭載し、ジャンル別の用例数も表示を可能にし

た。また、コロケーション抽出機能については 5 種の統計値が扱えるように配慮した。

(4) 日本語のみならず、英語のページも新設した。これは、特に開発したシステムの需要があったイギリスなど海外で日本語を学ぶ学生や大学院生への配慮でもある。

(3) 本研究の内容に関して学術論文や雑誌、書籍で発表し、また国内外の学会発表、ワークショップ、講演会などを通じて、開発したツールの活用と有効性の周知に努めた。

５．主な発表論文等

〔雑誌論文〕　計5件（うち査読付論文　5件／うち国際共著　5件／うちオープンアクセス　5件）

| | |
|---|---|
| １．著者名<br>仁科恭徳，赤瀬川史朗 | ４．巻<br>- |
| ２．論文標題<br>日英・英日パラレルコーパス検索ツール『パラレルリンク』(Ver.1.20)：インターフェース，検索機能，活用研究などについて | ５．発行年<br>2022年 |
| ３．雑誌名<br>Proceedings of the JAECS Conference（英語コーパス学会大会予稿集2022 | ６．最初と最後の頁<br>7-12 |
| 掲載論文のDOI（デジタルオブジェクト識別子）<br>なし | 査読の有無<br>有 |
| オープンアクセス<br>オープンアクセスとしている（また、その予定である） | 国際共著<br>該当する |

| | |
|---|---|
| １．著者名<br>仁科恭徳，赤瀬川史朗 | ４．巻<br>29 |
| ２．論文標題<br>『パラレルリンク』（Ver.1.0）の開発ーパラレルコーパス研究の概観とコーパス整備ー | ５．発行年<br>2022年 |
| ３．雑誌名<br>English Corpus Studies（英語コーパス研究） | ６．最初と最後の頁<br>63-78 |
| 掲載論文のDOI（デジタルオブジェクト識別子）<br>なし | 査読の有無<br>有 |
| オープンアクセス<br>オープンアクセスとしている（また、その予定である） | 国際共著<br>該当する |

| | |
|---|---|
| １．著者名<br>仁科恭徳、赤瀬川史朗 | ４．巻<br>- |
| ２．論文標題<br>日英・英日パラレルコーパスオンライン検索ツール『（仮称）パラレルリンク（Ver.1.0）』の開発に向けて（中間報告） | ５．発行年<br>2021年 |
| ３．雑誌名<br>Proceedings of the JAECS Conference（英語コーパス学会大会予稿集2021） | ６．最初と最後の頁<br>25-30 |
| 掲載論文のDOI（デジタルオブジェクト識別子）<br>なし | 査読の有無<br>有 |
| オープンアクセス<br>オープンアクセスとしている（また、その予定である） | 国際共著<br>該当する |

| | |
|---|---|
| １．著者名<br>仁科恭徳、赤瀬川史朗 | ４．巻<br>29 |
| ２．論文標題<br>『パラレルリンク』（Ver.1.0）の開発ーパラレルコーパス研究の概観とコーパス整備ー | ５．発行年<br>2022年 |
| ３．雑誌名<br>English Corpus Studies（英語コーパス研究） | ６．最初と最後の頁<br>- |
| 掲載論文のDOI（デジタルオブジェクト識別子）<br>なし | 査読の有無<br>有 |
| オープンアクセス<br>オープンアクセスとしている（また、その予定である） | 国際共著<br>該当する |

| | |
|---|---|
| 1．著者名<br>仁科恭徳 | 4．巻<br>23 |
| 2．論文標題<br>日本語複合動詞「X 込む」とその和英翻訳から概観する現行和英辞書の関係性と問題点 | 5．発行年<br>2021年 |
| 3．雑誌名<br>JACET Kansai Journal | 6．最初と最後の頁<br>149-162 |
| 掲載論文のDOI（デジタルオブジェクト識別子）<br>なし | 査読の有無<br>有 |
| オープンアクセス<br>オープンアクセスとしている（また、その予定である） | 国際共著<br>該当する |

〔学会発表〕　計5件（うち招待講演　2件／うち国際学会　4件）

| |
|---|
| 1．発表者名<br>Yasunori Nishina |
| 2．発表標題<br>Parallel Corpus Linguistics: Corpus Studies, Lexicography and Machine Translation |
| 3．学会等名<br>英国翻訳通訳協会（Institute of Translation and Interpreting）J-NETワークショップ（招待講演）（国際学会） |
| 4．発表年<br>2023年 |

| |
|---|
| 1．発表者名<br>Yasunori Nishina |
| 2．発表標題<br>How to use corpora to investigate Gen Zer's language and its usage on the Internet, and how to conduct a survey |
| 3．学会等名<br>Special Lecture for MSc/MPhil in Japanese Studies, University of Oxford（招待講演）（国際学会） |
| 4．発表年<br>2022年 |

| |
|---|
| 1．発表者名<br>仁科恭徳，赤瀬川史朗 |
| 2．発表標題<br>日英・英日パラレルコーパス検索ツール『パラレルリンク』(Ver.1.20)：インターフェース，検索機能，活用研究などについて |
| 3．学会等名<br>The JAECS 48th Conference（英語コーパス学会第48回大会）（国際学会） |
| 4．発表年<br>2022年 |

| |
|---|
| １．発表者名<br>仁科恭徳、赤瀬川史朗 |
| ２．発表標題<br>日英・英日パラレルコーパスオンライン検索ツール『（仮称）パラレルリンク』（Ver.1.0）の開発に向けて |
| ３．学会等名<br>The JAECS 47th Conference（英語コーパス学会第47回大会）（国際学会） |
| ４．発表年<br>2021年 |

| |
|---|
| １．発表者名<br>仁科恭徳 |
| ２．発表標題<br>より信頼性のある和英辞典の記述を求めてー翻訳ユニットの計量分析を通してー |
| ３．学会等名<br>外国語教育メディア学会関西支部大会 |
| ４．発表年<br>2020年 |

〔図書〕　計1件

| | |
|---|---|
| １．著者名<br>仁科恭徳 | ４．発行年<br>2023年 |
| ２．出版社<br>開拓社 | ５．総ページ数<br>216 |
| ３．書名<br>パラレルコーパス言語学の諸相 | |

〔産業財産権〕

〔その他〕

―

６．研究組織

| | 氏名<br>（ローマ字氏名）<br>（研究者番号） | 所属研究機関・部局・職<br>（機関番号） | 備考 |
|---|---|---|---|

７．科研費を使用して開催した国際研究集会

〔国際研究集会〕　計0件

８．本研究に関連して実施した国際共同研究の実施状況

| 共同研究相手国 | 相手方研究機関 |
|---|---|

| | | | | |
|---|---|---|---|---|
| 英国 | University of Oxford | | | |