「研究ノート」

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. 2.0）―インターフェース，検索機能，活用研究などについて―

仁科　恭徳・赤瀬川史朗

Abstract

In this paper, we introduce the concrete interface, various on-board corpora, search functions, implemented statistical processing, and expected applications of Parallel Link (Ver. 2.0), an analysis tool for Japanese-English and English-Japanese parallel corpora that has finally been completed. In particular, four new parallel corpora, BSD, Coursera, Hiragana Times, and JENAAD, have been added to the database, bringing the total number of parallel corpora to 13 and increasing its volume by around 1.3 times. An overview of these added corpora, as well as those to be added in the future, will be presented in detail. The new functions added in Ver. 2.0, such as the context display function, the distribution graph function for each corpus of collocations, the voice function, and the Google Translate link function, will also be introduced.

1．はじめに

仁科・赤瀬川（2021, 2022a, 2022b），仁科（2023）では，これまでに構築された日英・英日パラレルコーパスやその検索ツール，そしてこれらを活用した研究を網羅的に振り返り，2020 年度から 2022 年度にかけて筆者らが開発した日英・英日パラレルコーパスオンライン検索ツール『パラレルリンク』（Ver. 1.0～1.2）に搭載した 9 種のパラレルコーパスの概要と，それらを再整備する上で施したテキスト処理やアノテーション，全文検索インデックスの作成，ファイル整理などの一連の作業について詳説した。本稿では，2024 年 7 月下旬に公開された当該ツールの最新版 Ver. 2.0 に搭載したコーパスや具体的なインターフェース，検索機能，実装されている統計処理，想定される活用研究などにつ

いて紹介する[1]。

2.『パラレルリンク』（Ver. 2.0）のインターフェースと検索機能

2.1. 概要

Munday（2016）は, 翻訳学研究の中でも翻訳補助（translation aids）を促すツールの開発や利用は応用翻訳学（Applied Translation Studies）に含まれる分野であるとし，この中には機械翻訳を含むソフトウェア開発やクラウド・ソーシングを含む共同作業（collaboration），辞書やオンライン検索，パラレルコーパスの参照などが含まれるとする。『パラレルリンク』のプロトタイプの開発経緯については既に仁科・赤瀬川（2021, 2022）で述べたとおりであり，元々は翻訳作業や辞書編纂の手助けとなるような参照ツールを目指して開発した。Ver. 1.0〜1.2 では，現在までに無償公開された 9 種の日英・英日パラレルコーパスを搭載し，オンライン上で網羅的に串刺し検索できることを可能にした。各ジャンルの語数にばらつきはあるものの，今後 KWIC 検索などの検索機能を充実させることで，簡易的な一般参照パラレルコーパスとして活用することも期待できる。また，起点言語（日本語）の検索語に関する精緻な語彙プロファイルの獲得や，その目標言語（英語）における信頼性ある翻訳例が参照できる有益なツールとなりうる。

そこで，本節では，当該ツールの最新版となる Ver. 2.0 に実装されているパラレルコーパスとインターフェースおよび検索機能を紹介する。特に，見出し検索，パタン表示機能，コロケーション表示機能，翻訳表示機能，文脈表示機能，コロケーション分布グラフ，Google 翻訳リンク機能などを紹介し，その後，想定される活用研究や Ver. 3.0 以降の開発計画など，今後の展望についても具体的に示す。

本ツールは，国立国語研究所と Lago NLP（旧 Lago 言語研究所）が開発したブラウザベースのコーパス検索ツール Lago Word Profiler（以下，LWP）（パルデシ・赤瀬川，2011）が原型となっており，「見出し語単位で検索，コロケーションなどを文法項目に分類して整理して表示」することができ（染谷・赤瀬川・山岡，2011），日本語原文の英訳もパタンやコロケーション別に表示させることが可能である。また，ジャンル別の頻度表示や絞り込み機能も併せ持つ。

2.2. レキシカルプロファイリングのツール開発について

仁科・赤瀬川（2021, 2022a）では，テクスト処理，アノテーション，テクストクリーニング，エンコーディングの統一，全文検索インデックスの作成，ファイル整理などについて紹介した。インターフェースの開発には他にもレキシカルプロファイリングデータベースの構築が必要となる。

参考までに，日本語の内容語（動詞，名詞，形容詞，副詞）のレキシカルプロファイリングデータベースを構築する作成手順を簡単に紹介する。まず，フォーマットを統一したコーパス群を CQL で検索できる検索システムを構築する。次に，各内容語の文法パタンを係り受け関係のあるものから抽出できるように CQL クエリーで表現する。抽出した文法パタンは，RD（B）MS（Relational DataBase Management System）の MySQL データベースに格納している。インターフェース開発までの一連の経緯をまとめると，係り受け解析，内容語のパタン作成（名詞，動詞，形容詞（イ形容詞，ナ形容詞），副詞，連体詞），レキシカルプロファイリングデータベースの構築（内容語ごとの共起関係のデータ（共起語，頻度，MI スコア，LogDice, T スコア，対数尤度比））となる。

なお，レキシカルプロファイリングデータベースは，見出し語情報テーブル，文法パタンテーブル，コロケーションテーブル，用例テーブルから構成される。見出し語情報テーブルは，見出し語（漢字かな交じり，ひらがな，ローマ字），品詞（大分類，中分類），頻度など，見出し語に関する基本情報を収録したテーブルである。文法パタンテーブルは，見出し語ごとに，文法パタン別の頻度，パタングループにおける割合などを収録したテーブルである。コロケーションテーブルは，見出し語と文法パタンをキーとして，該当するコロケーションと統計値（前述の単純頻度，MI スコア，LogDice, T スコア，対数尤度比），コーパスごとの頻度を収録したテーブルである。用例テーブルは，見出し語と文法パタンとコロケーションをキーとして，該当する用例と対訳，出典元のコーパス名を収録したテーブルである。

2.3. Ver. 2.0 における搭載コーパスの追加・変更

Ver. 2.0 から新たに追加したパラレルコーパスは，ビジネスシーン対話対訳コーパス The Business Scene Dialogue Corpus（以下 BSD）（https://github.com/tsuruoka-lab/BSD），オープンコースウェア対訳コーパス Coursera Parallel Corpus（以下 Coursera）（https://nlp.ist.i.kyoto-u.ac.jp/EN/?Coursera+Parallel+Corpus），ひらがなタイムズ日英対訳コーパス（以下 Hiragana Times），日英新聞記事対応

付けデータ The Japanese-English News Articles Database（以下 JENNAD）の 4 種である。

BSD は，会議，交渉，雑談といったビジネスシーンに沿った日本語および英語での会話のシナリオを作成し，英語および日本語へ翻訳したコーパスである（中澤・李・Rikters, 2021）。Coursera は教育分野のコーパスであり，世界中の大学のコースを無償でオンライン提供している Coursera（https://www.coursera.org/）から抽出した日英対訳データから成る。これら 2 種のパラレルコーパスについては無償であり，前者はビジネス関連で話し言葉の性質を持ったコーパスであること，後者は教育関連コーパスであることから Ver. 1.0 に搭載した 9 種のパラレルコーパスには不足・欠落している分野・モードであるため追加した。

また，Hiragana Times は今回新たに購入した有償コーパスで，ひらがなタイムズ誌面の対訳データから構成される。1988 年～2017 年までの計 349 冊/ファイル（対訳数 212,230 文）と別冊書籍 Hiratai-Books シリーズ 19 冊/ファイル（対訳数 21,796 文）で構成されており，アカデミック割引でも 100 万円程度必要となる有償ライセンス形式のコーパスである。そして JENAAD は読売新聞と Daily Yomiuri の記事を収録したパラレルコーパスで，公開当時は最大規模のパラレルコーパスであった。長年の間，言語研究者に愛用されてきたパラレルコーパスの雄である（仁科，2014）。これら 2 種のパラレルコーパスについて，前者は有償で購入するために研究費を獲得する必要があったこと，後者は版権の確認など時間を要したことから Ver. 1.0 には搭載できなかったが，今回 Ver. 2.0 でようやく搭載することができた。Ver. 1.2 から Ver. 2.0 にかけて，語数換算で日本語が約 1.35 倍，英語が 1.64 倍，対訳対数が 1.31 倍に増強している。

JESC，Open Source，Reuters，Taiyaku に用いられた類似度フィルタリングには，Google の大規模言語モデル universal-sentence-encoder-multilingual を利用し，異言語（日本語－英語）のセンテンス間の類似度を判定している。また，Ver. 3.0 以降で JParaCrawl から 100 万対程度，Wikimedia から 14 万対程度を追加し，全対訳数 500 万対を目指す。JParaCrawl については，類似度フィルタリングと有害コンテンツの排除を行なった上でデータベースに追加予定である。このコーパスはネットから対訳を探したデータであることから，有害コンテンツなど質の悪いデータが一定数混入しているが，この問題を解決し質の良い用例だけを足すことができれば本ツールで検索できる全データサイズも現在の数倍となり，質・量共にパワーアップする。データ量が増えることで，翻訳ユニットの抽出精度も格段に向上することが期待される。Wikimedia Parallel Corpus に

表 1. Ver 2.0 搭載コーパスに関するまとめ（新規追加，変更点など）

パラレルコーパス	翻訳方向	ステータス	語数（日本語）	語数（英語）	対訳対数
ビジネスシーン対話対訳コーパス（BSD）	英→日 日→英	新規追加	314,785	277,267	24,578
オープンコースウェア対訳コーパス（Coursera）	日→英	新規追加	1,147,393	1,036,545	63,118
ひらがなタイムズ日英対訳コーパス（Hiragana Times）	日→英	新規追加	4,725,779	4,279,484	261,770
日英新聞記事対応付けデータ（JENNAD）	日→英	新規追加	5,287,223	4,957,273	192,763
日英サブタイトルコーパス（JESC）	英→日 （一部，日→英）	類似度フィルタリング	7,479,726	6,785,436	1,061,623
日英法令対訳コーパス（LAW）	日→英	各用例に法律名を付与	12,991,680	13,092,136	424,826
大規模オープンソース日英対訳コーパス（OPENSOURCE）	英→日	類似度フィルタリング	4,492,098	3,371,492	246,137
ロイター日英記事の対応付け（REUTERS）	英→日	類似度フィルタリング	2,066,014	1,714,672	76,295
SCoRE 用例コーパス（SCoRE）	英→日	トピック別用例の追加	213,195	136,302	15,696
日英対訳文対応付けデータ（TAIYAKU）	英→日 （一部，日→英）	類似度フィルタリング	1,914,472	1,431,563	118,084
Tatoeba 日英対訳コーパス（TATOEBA）	日→英	最新版より用例追加	2,604,111	2,097,640	279,985
TED Talk 英日コーパス（TED）	英→日	パラレルコーパスの質を改善	8,937,299	7,654,440	588,956
Wikipedia 日英京都関連文書対訳コーパス （WIKIPEDIA）	日→英	変更なし	9,436,567	9,940,272	465,955
		Total 2.00 版	52,673,043	56,774,522	3,230,830
		1.20 版	38,934,507	34,646,840	2,459,913

ついては，https://opus.nlpl.eu/wikimedia/ja&en/v20230407/wikimedia を参照されたい。

2.4. Ver. 2.0 のインターフェースと検索機能

本節では，本ツール（https://www.parallellink.org/）の具体的なユーザーインターフェースについて紹介する。本ツールは LWP を基盤として開発されたことから，他のコーパスを搭載した LWP である NINJAL-LWP for BCCWJ（NLB）（プラシャント・赤瀬川，2011）や LWP for ParaNews（現在は公開終了），

Wikipedia-Kyoto LWP（現在は公開終了）を使用したことのあるユーザーにとっては，馴染みのあるインターフェースである。但し，今回は白を基調とした明るく眼に優しいデザインに変更しフォントも見やすくなったことから，旧インターフェースより視認性が高くなっている。なお，上記ウェブサイトからユーザーガイドをダウンロードすることができるので，基本的な検索機能についてはそちらを参照されたい。また，インターフェースやユーザーガイドについては英語版も用意している。画面右上に配置された Display Language をクリックすれば，表示言語を Japanese（日本語）と English（英語）から選択できる。

まず，Home 画面には本ツールに関する簡単な紹介，ユーザーガイド，引用方法，開発チームメンバー，更新履歴，謝辞，問い合わせ先が記載されている。今後，『パラレルリンク』を活用した研究事例の情報なども随時こちらにアップしていく予定である。その Home 画面右上の search をクリックすると，図 1 左の利用許諾のポップアップが表示される。「同意する」を選択すると，図 1 右に示す検索画面に移動する。Ver. 2.0 の段階では日→英方向のみの検索が可能であるため，日本語の各見出し語の情報のみが一覧表示される（Ver. 3.0 以降で英→日検索も可能とする予定）。予め用意された品詞タブは，「すべて」，「名詞」，「動詞」，NLB とは異なる「イ形容詞」と「ナ形容詞」，「連体詞」，「副詞」の計 7 種である。Ver. 2.0 から，「すべて」に掲載されている見出し語には品詞ラベル（例．動）と，補助動詞としても使われる動詞には［非自立可能］のラベルが付与されている。検索したい特定の品詞タブをクリックすれば，各品詞に該当する語の情報が頻度順で表示される。検索ボックスへの文字入力は，か

図 1. 利用規約のページ（左）と見出し検索の画面（右）

な漢字交じりのほか，ひらがな，カタカナ，ローマ字による検索にも対応している。

見出し検索の画面では，上部に検索ボックスが用意されている。ここに検索したい特定の語を入力することで，その語の情報を獲得することができる（あるいは，品詞別タブから予めリストアップされた語を選択することでも検索を開始することができる）。参考までに，図 2 は「落とす」を検索した結果の画面を示す。「落とす」のみならず，V+V 型複合動詞の「見落とす」や「切り落とす」も含まれている。

図 2. 検索ボックスで「落とす」を指定した場合

なお，カタカナで入力し絞り込みを行うと，その読みを含むすべての見出しが表示される。例えば，カタカナで「オコル」と入力すると，図 3 左に示すように「起こる」や「怒る」などのほか，その読みを含む V+V 型複合動詞なども全て検索結果に表示される。また，図 3 右のように，V+V 型複合動詞は排除した上で「オコル」に完全一致で検索したい場合，正規表現を用いて「^オコル$」のように先頭と末尾を表す記号を入れるとよい。ひらがなで「^おこる$」と入力しても，同様の検索結果が得られる。ちなみに，「^[アイウエオ]」とすると，ア行の見出しがすべて検索できる。

図 3 の検索結果から調査したい語（今回は，「起こる」）を選択してみると，図 4 のような語彙プロファイルが一覧で表示される。この語彙プロファイルでは，左側に表示されたパタンパネル中の「名詞＋助詞」の中の「名詞＋が起こ

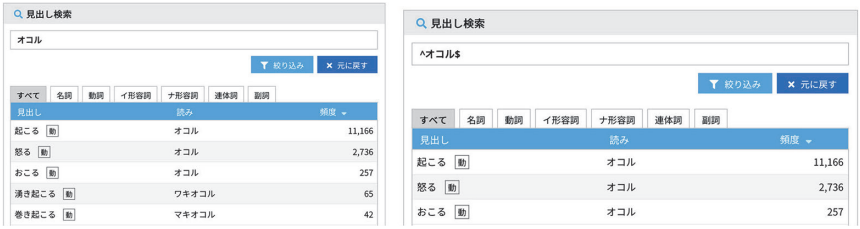

図 3. 検索ボックスで「オコル」を指定した場合（左）と「^オコル$」を指定した場合（右）

る」（4,442 例）が選択されており，そのコロケーションが頻度順にリスト化され画面中央に配置されている。その中で最も頻度の高い「ことが起こる」に該当する例文が，その対訳と共に右側に表示されている。検索したい語の文法パタン，コロケーション，用例，その対訳の検索結果が瞬時に一覧表示されている点が本ツールの強みであり，現時点では唯一無二のツールであるとも言える。これは，レキシカルプロファイラー（特に，ここでは LWP）の秀逸な機能であると言ってよい。

図 4.「起こる」の語彙プロファイル

図 4 の表示内容をもう少し詳しく解説していく。まず，左上に検索語（見出し語）の総頻度が 11,166 件であることが表示されており，その下部に「グルー

プ別」と「頻度順」のタブがある。「グループ別」には，下にスクロールしていくと「名詞＋助詞（＋起こる)」や「名詞＋複合助詞（＋起こる)」などパタンのグループ別に該当するパタンが頻度とグラフ化された割合（%）と共に表示されている。グラフにカーソルを持っていくと，実際の割合が数値で示される。「頻度順」タブをクリックすると，パタングループ別の表示から頻度順の表示に変わり，高頻度のパタンから順に頻度と割合（%）が表示される。これらの中で調べたいパタンをクリックすると，そのコロケーションリストが画面中央のパネルに表示される。つまり，検索語→パタン分析→コロケーション分析を経て該当する用例と対訳が表示される仕組みとなっており，他のコーパス検索ツールには例を見ない本ツール独自の機能であると言える。

また，図 5 では，見出し語「起こる」の中でも，「名詞＋助詞＋起こる」のパタングループのうち，「〜が起こる」を指定して，さらにその中の「事件が起こる」を指定した画面である。該当する日本語の例文とその英訳が順に表示されていることが分かる。

図 5.「事件が起こる」を選択した場合

右側に示された各例文には，どのパラレルコーパスからの例文であるのか，出典がその右下に表示される。また，例文ウィンドウのヘッダーを見ると，「事件が起こる」の例が計 208 件抽出されたことが分かる。その下に各サブコーパ

ス別の出現回数が表示されており，その用例数を概観すると Hiragana Times が 42 例，JENAAD が 10 例，JESC が 25 例，REUTERS が 1 例，SCoRE が 1 例，TAIYAKU が 5 例，TATOEBA が 3 例，TED が 29 例，WIKIPEDIA が 92 例である一方で，他のコーパスでは全く使われていないことが分かる。各サブコーパスをクリックすれば，そのサブコーパスのみでヒットした例文だけが表示される。

また，Ver. 1.0 の時代から拘っている点は，コロケーション抽出の際に活用できる統計指標の充実である。コロケーション情報については，NLB で搭載された頻度，MI スコア，LD スコアに加えて，T スコアと対数尤度比（LLR）の統計指標も追加している。各指標をクリックすれば，各値で並び替えが可能である。また，ヘッダーのすぐ下にあるフィルター機能を利用すれば，特定のコロケーションの検索や，任意の統計値での絞り込みができる。例えば，コロケーション列に「事件 | 事故」のように入力すると，「事件が起こる」，「事故が起こる」，「交通事故が起こる」が検索される。

2.5. Ver. 2.0 の特筆すべき機能

本節では，特に今回の Ver. 2.0 で新たに追加した 3 機能について紹介する。

2.5.1. コンテクスト（文脈）表示機能

パラレルコーパスの構造上，コンテクスト（文脈）のあるコーパス（BSD, Coursera, Hiragana Times, LAW, WIKIPEDIA）については，用例パネルでのコンテクスト表示に対応している。各用例のコーパス名の左横に（三本線）アイ

図 6.「秘密を知った！」の日英文脈状況

コンが表示されているものは，その用例の前後のコンテクストを確認できる。このアイコンをクリックすると，図 6 のように，用例（該当行）の前後 5 センテンスが表示される。

コーパスを整備する過程で重複行は削除しているが，コンテクストを表示しているコーパスについては重複行であっても前後の文脈が異なる場合があるため，重複する用例はそのまま残している。

（例）規定｜規定を適用する

LAW に次の用例が 9 件あるが，コンテクストは全て異なっている。

補足情報として，コンテクスト情報のないコーパス（JENAAD, JESC, OPENSOURCE, REUTERS, TAIYAKU）は，起点言語と目標言語のテクストを文単位に分割し，目標言語から起点言語の文と類似度の高い文を見つけ出す手法で構築したものであることからコンテクストが残っていないが，これは主に著作権への配慮でもある。SCoRE や TATOEBA については文単位の例文を収集したコーパスであることから，元々コンテクストが存在していない。このような理由により，現状 BSD, Coursera, Hiragana Times, LAW, WIKIPEDIA の 5 種のみコンテクスト情報が表示される。

2.5.2. コロケーションのコーパスごとの分布グラフ表示

用例パネルのヘッダーにあるアイコンをクリックすると，コロケーションのコーパス分布が 10 万語当たりの調整頻度で降順に表示され，棒グラフのバー

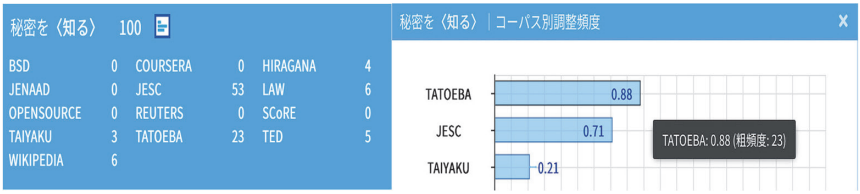

図 7.「秘密を知る」のパラレルコーパス別粗頻度表示と調整頻度グラフ機能

にマウスをかざすと調整頻度と粗頻度が表示される。

2.5.3. 音声機能と Google 翻訳リンク機能

Ver. 1.0 の時代から SCoRE コーパスから抽出された例文には音声マークが表示され，そのマークをクリックすることで音声が流れる仕組みとなっている。Ver. 2.0 では，Google 翻訳リンク機能を搭載し，SCoRE 以外の英文音声のない用例についても，Google 翻訳のサイトを開いてその音声を確認することができる。英文の後ろに表示される A文アイコンをクリックすると，新しいタブに Google 翻訳のサイトを開き英文が表示される。音声を聞くことができる他，元の日本語と Google 翻訳の日本語訳の比較も可能である。

図 8. Google 翻訳リンク機能

3.　Ver. 3.0 以降の改良について

3.1. 有償ライセンスコーパス「読売新聞日英文対訳コーパス」の購入検討について

今後，Ver. 3.0 以降の開発において，搭載するコーパスの拡充や検索機能の追加等をさらに進めていく。特に，学術論文や話し言葉のパラレルコーパスは欠落あるいは不足しているため，このようなジャンルの翻訳テクストを可能な範囲で増補したい。Ver. 3.0 以降の追加候補のパラレルコーパスとして，有償ライセンスで利用できる「読売新聞日英文対訳コーパス」を検討中である。現時点で 2006 年～2021 年までのデータが読売新聞東京本社メディア局に保管されており，1 年分で 34 万円と高額であるため（16 年分で 544 万円），本コーパスを完全に搭載するためには一定の研究費を獲得する必要がある。

3. 2. 検索機能の充実について

Ver. 3.0 以降では検索機能のさらなる充実も検討している。具体的には，コンコーダンサー機能や，英→日の語彙プロファイリング機能の追加を検討して

いる。レキシカルプロファイラーは，コーパス分析の初・中級者にとっては，見出し語のパタン情報やコロケーション情報を瞬時に獲得することができるため，検索結果の解釈に集中できるという利点がある。各種統計値の計算も自動化されていることから，従来の計量的な言語分析で愛用されてきたコンコーダンサーを凌駕している面もある。その一方で，レキシカルプロファイラーは予め開発者が設計した検索結果しか表示できないという制約もある。そのような制限なしに検索・分析したいコーパス分析の上級者にとっては，「特定の言語事象を対象にしたミクロな視点からの観察」を可能とするコンコーダンサーを好む者も多い。コンコーダンサーの機能も追加することで，検索語・句に関する起点言語の情報と目標言語の翻訳特性がより精緻に分析できる。

さらに，検索時の視認性を高めるために抽出された訳例中の翻訳ユニットの候補を自動的に太字で色付けする機能や，ParaConc（https://paraconc.com/）に搭載されている Hot Words 機能のように，自動的に訳語や翻訳ユニットの候補を抽出し，頻度順や統計値順に翻訳候補をリスト化する機能も搭載予定である。

また，パラレルコーパスから抽出した英文の用例を授業や教材開発，辞書の例文等で活用することを想定し，英語学習者にとって平易な英文を抽出してくれる GDEX（Good Dictionary Examples）（https://www.sketchengine.eu/guide/gdex/）の活用や，それに類似したオリジナルの機能を実装することも検討したい。特に，パラレルコーパスから抽出した英文が翻訳文である場合，信頼性（authentic）のある英文であるとは言い難い。House（2014, p. 2）も翻訳文は “a kind of inferior substitute for the 'real thing'” と指摘する。よって，授業・教材・辞書などで英文を活用する場合には，語彙難度や文法複雑性なども調整すべきであり，その点で GDEX のような機能の活用価値は高いと言える。また，既に本ツールに実装している SCoRE（本ツールにも搭載されている教育用の平易な例文コーパス）のデータの活用も検討中である。

他にも，特定の文法パタンや共起語の翻訳を抽出した際にクリック一つで Dual KWIC 表示に切り替えできる機能や，TED コーパスについては動画ファイルへのジャンプ機能も追加したい。今後は，辞書編纂や翻訳・通訳実践（研究），対照言語学，言語教育などの分野で活用できる変則的な検索にも対応した，言語の専門家のニーズに応える翻訳コーパス集合体のオンライン検索ツール開発を目指す。

4. 想定される『パラレルリンク』の活用研究とは

本ツールを活用した研究に，英和・和英辞書編纂時における訳語・訳例チェックのための検証的活用，翻訳･通訳実践時における現場等での参照的活用，リーディング・ライティング・翻訳の授業時など英語教育現場での実践的活用などが挙げられる。

はじめに，英和・和英辞書編纂時の検証的活用について述べる。英和・和英辞書に掲載すべき翻訳ユニットの種類・数はコーパスデータに基づく客観的指標から判断し決定すべきであると筆者らは考える。現行の英和・和英辞書に掲載されている訳語と本ツールから獲得した翻訳データとを比較検証することで，客観的指標をもって辞書記述をより信頼性あるものへと近づけることが可能となる。特に，複数のパラレルコーパスから抽出した翻訳ユニットの信頼性を量的観点から総合的にランク付けするだけでなく，ジャンルやレジスターによって訳語がどのように変化するのかを特定することで，訳語ごとにレーベル表示することが可能となり，精緻な言語事実を訳語に反映させることができる。例を挙げると，本ツールに生起する高頻度名詞の一つに「規定」があるが，文法パタン「複合動詞＋動詞」において「規定により＋動詞」では 13,899 例が，「規定により〈受ける〉」は 1,303 例がヒットし，いずれも LAW コーパスからである。初めの数十例を調査すると，receive/obtain something pursuant to the provisions of〈具体的な規定内容〉といった固定化された翻訳ユニットを抽出することが出来た。よって，「規定により〈受ける〉」は法律のレーベルを貼ることができ，「受ける」は receive/obtain を用いて，「規定により」は pursuant to the provisions of〈具体的な規定内容〉といった表現を用いることが望ましいということが分かる。

また，本ツールを活用することで，現行の英和・和英辞典の記述の問題点を指摘し，仁科（2020, 2023）のようにその具体的な改善案を示すことも可能となる。具体的には，起点言語（日本語）について BCCWJ などの日本語大規模コーパスを活用した先行調査に始まり，本ツールなどの大規模パラレルコーパスによる翻訳ユニットの抽出，GDEX や SCoRE を用いて当該翻訳ユニットが含まれる簡易文の選定，その後に辞書のサンプル記述案の作成といった流れとなる。現在まで，例えば英和辞典の編纂であれば，起点言語となる英語の見出し語の選出や，その共起関係の調査にのみ単言語コーパス（ここでは英語コーパス）が活用されてきた。つまり，目標言語に関する情報，例えば二言語辞典

に掲載されている訳語については執筆者の主観によって作成されていたため，本ツールの翻訳データを活用することで，真のコーパス駆動型アプローチによる二言語辞書編纂が可能となる。

次に，翻訳・通訳実践における活用についても，やはり，一般的な辞書からは得ることの出来ない膨大な翻訳の実例を獲得出来る点に強みがあると言える。1 次翻訳は辞書などを使ってざっと粗翻訳し，2 次翻訳時に本ツールを活用して詳細にチェックする，3 次翻訳時にネイティブチェックを受けるという方法や，1 次翻訳から積極的に本ツールを活用するという方法もある。特に，現行の Ver. 2.0 では様々なジャンルの 13 種の日英・英日パラレルコーパスを搭載していることから，翻訳・通訳の用途に応じて特定のパラレルコーパスを選択し，その対訳結果を参考にする方が効率的であろう。この特定のジャンルや専門分野で好まれて使用される言語表現に注目する方法は，応用言語学における LSP（Language for Specific Purposes）の考え方に沿った TSP（Translation for Specific Purposes）であるとも言える。実際に，英訳抽出結果上部のサブコーパス別頻度の部分をクリックすると，特定のサブコーパスの用例のみが表示される。図 9 は，「起こる」→「名詞＋助詞」→「名詞＋が起こる」→「事件が起こる」の順で英訳の結果を表示させ，さらに Hiragana Times の英訳のみに絞った検索結果を示している。

図 9. Hiragana Times のみに出現した「事件が起こる」の日本語原文とその英訳

最後に，言語教育における活用については，特に翻訳・英語ライティングの授業などで活用が期待される。例えば，質の高い英作文を完成させるまでのプロセスで大事なことは，書き手が伝えたい内容をまずは英訳に適したやさしい日本語に一度書き換えた上で，それを英訳するという 2 段階のステップを踏む方法である。英訳に適したやさしい日本語とは，主語と述語を明確にし，長文を避け，極力標準語を使い，新語は避けるなどの配慮が施されたものであり，最近では DeepL 翻訳ツール（https://www.deepl.com/translator）など機械翻訳を有効活用する際にも留意するポイントでもある。言い換えれば，自然な日本語と自然な英語（あるいは英訳）には文化的・言語的なギャップが存在していることの裏返しでもある。これは，いわゆる Jakobson（2004; originally published in 1959）の言語内翻訳（intralingual translation）の実践でもあり，Nida and Taber（1969, p. 33）による翻訳プロセスの分析・転移・再構成を具現化した行為でもある。

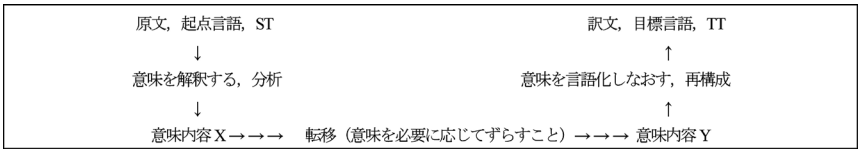

図 10. Nida の翻訳プロセス（Nida and Taber 1969, p. 33）

本ツールの検索から獲得できる翻訳文は，田中コーパスなどの一部を除きプロが翻訳したという意味においてオーセンティックなものばかりである。よって，どのような日本語表現において英訳との間に文化的・言語的なギャップが見られるか，例えば，日本特有の文化・言語表現が実際にどのように英訳されているかなど，予め学習者に考えさせ試訳させた上で本ツールを用いて検索させ自律学習を促すことは，翻訳上の「気づき」を促す上でも効果的であろう。模範となるプロの翻訳者が産出した翻訳に到達するためにはどのようなプロセスが必要であるのかを学習者に議論させることも，「気づき」学習を助長する上で意味があろう。そのような「気づき」に繋がるヒントがパラレルコーパスの翻訳データには埋もれている。特に，検索語の訳し方が，その共起語によって変わるという翻訳事実や，日英・英日翻訳というのはそもそも逐語的に遂行されるものではなく，時に異なる品詞に訳されたり，（あえて）省略されたりすることもある（Lost in Translation）という翻訳実態も学ぶことができる。

5. 今後の展望

以上，本稿では，最新版『パラレルリンク』Ver. 2.0 のインターフェースや検索機能，想定される活用研究について紹介した。既に述べたように，搭載されている一部のパラレルコーパスの翻訳文は信頼性に欠けるため，Ver. 3.0 以降では欠落・不足しているジャンルの翻訳テクストも含め，質の高い翻訳テクストを増補し，バランスの取れた日英・英日パラレルコーパス検索ツールの開発・発展を目指す。

謝辞

本研究は JSPS 科研費 20K00692 と 23K00599 の助成を受けたものである。ここに，SCoRE の用例コーパスを搭載することをご快諾くださった中條清美先生（元日本大学），現在 SCoRE の一連の研究を引き継いでおられる西垣知佳子先生（千葉大学），パラレルコーパス研究の発展を願われた故・染谷泰正先生，ならびに関係者の皆様に感謝の意を示す。

注

1. 本稿の内容は 2022 年 10 月 1 日にオンラインで開催された第 48 回英語コーパス学会および 2024 年 10 月 5 日〜6 日に対面開催された第 50 回英語コーパス学会（於青山学院大学）における口頭発表の内容を大幅に修正・発展させたものである。仁科・赤瀬川（2022b），仁科（2023）にも負うところが大きい。

参考文献

House, J. (2014). *Translation: A Multidisciplinary Approach.* Hampshire: Palgrave Macmillan. doi: https://doi.org/10.1057/9781137025487

Jakobson, R. (1959/2004). On linguistic aspects of translation. In L. Venuti (Ed.), *The translation studies reader (2nd ed.)* (pp. 138–143). New York: Routledge.

Munday, J. (2016). *Introducing translation studies: Theories and applications (4th ed.)*. London and New York: Routledge. doi: https://doi.org/10.4324/9781315691862

中澤敏明・李凌寒・Matıss Rikters（2021）「ビジネスシーン対話対訳コーパスの構築と対話翻訳の課題」『言語処理学会第 27 回年次大会発表論文集』1375–1380.

Nida, E.A., & Taber, C.R. (1969). *The theory and practice of translation.* Leiden: E.J. Brill.

仁科恭徳（2014）「実践で学ぶコーパス活用術：第 11 回パラレルコーパスの可能性」『研究社 WEB マガジン LIngua（リンガ）』オンライン.

仁科恭徳（2020）「日英パラレルコーパス WikipediaKyoto-LWP を用いた和英辞典の記述改善案について―「X を固める」の場合―」『英語コーパス研究』第 27 号：1–21.
仁科恭徳（2023）『パラレルコーパス言語学の諸相』東京：開拓社.
仁科恭徳・赤瀬川史朗（2021）「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）」『英語コーパス学会大会予稿集 2021』25–30.
仁科恭徳・赤瀬川史朗（2022a）「『パラレルリンク』（Ver. 1.0）の開発―パラレルコーパス研究の概観とコーパス整備―」『英語コーパス研究』第 29 号：63–78.
仁科恭徳・赤瀬川史朗（2022b）「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―」『英語コーパス学会大会予稿集 2022』7–12.
パルデシプラシャント・赤瀬川史朗（2011）「BCCWJ を活用した基本動詞ハンドブック作成 – コーパスブラウジングシステム NINJAL ― LWP の特長と機能―」『現代日本語書き言葉均衡コーパス完成記念講演会予稿集』205–216．東京：国立国語研究所.
染谷泰正・赤瀬川史朗・山岡洋一（2011）「大規模翻訳コーパスの構築とその研究および教育上の可能性」『日本メディア英語学会第 1 回年次大会発表資料』1–15.

（仁科　恭徳　神戸学院大学）
（赤瀬川史朗　Lago NLP）