

Lexonomy を利用した日中対照俳句歳時記の構築

黄海洪（京都大学大学院）、薛楽（京都大学大学院）

The Construction of Japanese-Chinese Haiku Saijiki using Lexonomy

Haihong HUANG, Kyoto University Le Xue, Kyoto University

要旨：日本語教育において、俳句は使用頻度の高い季語を用いることで教育素材として活用できる。本研究は、日本語学習者と日本語教師の両方の立場から、Web 辞書編集ツール Lexonomy を用いた Web 日中俳句歳時記を提案する。本研究の手法を取り入れることで、日中対照俳句辞書以外にも学習者のニーズに合わせたメタ辞書の構築ができる。

キーワード：Lexonomy、俳句、日中対照、Web 辞書

1. はじめに

俳句は 17 音によって構成される世界で一番短い定型詩として、日本だけでなく、海外でも高い人気を有している。俳句は原則として季語を用いる有季定型俳句が一般的である。しかし、自然環境や文化的背景が違う中国語を母語とする日本語学習者にとって、季語の持つ深い意味を日本人と同じように理解することは難しい。

季語を調べるには、俳句歳時記が必要不可欠である。インターネット上の Web 辞書を参照するのが一般的な時代となり、Web 俳句歳時記も登場している。Web で公開されている俳句歳時記は日本語母語話者向けのモノリンガルのものがほとんどであり、日中両言語で参照できる俳句歳時記は管見の限り、見つからない。歳時記は、辞書とは似て非なるものだが、本研究では「俳句歳時記」を、季語を調べるための俳句辞書とする。俳句を教える日本語教師と俳句を学ぶ日本語学習者が、共同で俳句辞書を作ることができれば、より効率的に作句することができる。

Web 俳句辞書のメリットとして、以下の 3 点が挙げられる。

- a. インターネットに接続さえすれば、いつ・どこからでも利用できる。
- b. 検索機能が利用できる。
- c. 追加や削除などの更新が容易にできる。

吉田（1986）は、「辞書は規模が大きく、内容も複雑で、その開発には多大の経費と労力を要する上に、利用者一人一人その内容に対する要求が異なる。それにも関わらず、個人で開発することが困難であるという大変やっかいな問題をかかえている」と指摘している（p.933）。この問題点は、依然として解決されないまま存在している。また、李ほか（2018）は、コーパスを用いる辞書編集がすでに常識となっている英語教育に対し、日本語教育は「日本語の教師や学習者に向けた辞書の開発そのものが大幅に遅れている」と指

摘し、「本格的な日本語学習辞書の編集、特に、世界各地の学習者に使いやすい日本語と現地語によるバイリンガル辞書の編集」の重要性を訴えている (p.219)。

2. 先行研究

俳句は、古典俳句と現代俳句に分けられる。古典俳句には特殊用語と難解な季語が使われているため、日本語学習者のニーズに合わない可能性が高い。したがって、本研究は、現代俳句を対象とする。現代日本語書き言葉均衡コーパス (BCCWJ) には、1950 年から 1979 年の間に発表された 92 句集の中から、角川書店が発行する「現代俳句大系 増補」(8 巻分)、計約 5 万語の現代俳句データが含まれている (丸山ほか 2012 ; 李ほか 2018)。しかし、その内容は 40 年前に詠まれたものであり、また季節のタグもないため、俳句について調べる場合、俳句の知識が前提となる。俳句の特徴の 1 つに、多種多様な季語が存在することが挙げられるが、Web 俳句辞書を構築する場合には、季語を季節ごとに整理する必要がある。

伊藤ほか (2016) によると、日本語学習者が求める学習 Web サイトに対して、「文化情報」、「多言語対応」といった機能を求めている。野田 (2011) は日本語学習者に役立つ辞書を作るには、「何のために辞書を使い、どんな情報を知りたいかという実際の状況から出発しなければならない」と主張している (p.6)。

辞書とコーパスは密接に関係する両輪である (伝ほか, 2019)。辞書に記載する情報の判断は、従来辞書を編む専門家に委ねられていたが、コーパスの登場により、コーパスに基づく辞書の編集が主流となりつつある。英語辞書の Collins COBUILD English Language Dictionary (1987) の編集は世界で初めてコーパスを用いて本格的に行われた。

野田 (2011) は辞書構築について、以下の 2 点を提案している (p.7)。

- a. 非母語話者が辞書を使う状況をもとにして辞書を作る。
- b. 分量に制限がないインターネットで辞書を提供する。

日本語学習者向けのインターネット辞書に関しては、これまでに、数多くの Web 辞書が構築されている。例えば、国立国語研究所 (NINJAL) は「基本動詞ハンドブック」や「トピック別 アイヌ語会話辞典」など複数の Web 辞書を構築し、一般公開している。また、「NIHONGOe な」というサイトでは「日本語俗語辞書」、「経済のにはんご」などの学習者向けの Web 辞書が多く紹介されている。しかし、これらの辞書の構築は、多くの場合専門家と研究者の裁量で行われ、非母語話者のニーズとスキルを十分に汲み取っていないところがある。

3. 研究目的

本研究は、日本語学習者と日本語教師の両方の立場から、Lexonomy を用いた Web 俳句辞書について提案する。Lexonomy はイギリスで開発されたオープンソースの辞書編集システムである (Měchura, 2017)。特徴として、ICT のスキルがなくても、Web 辞書の構築、

オンライン公開が可能な点が挙げられる。辞書編集プラットフォームは利用者にわかりやすいインターフェスを備えているため、プログラミングの知識を持たない日本語教師や日本語学習者にとっても気兼ねなく使えるシステムである。本研究は日中対照俳句歳時記の構築を通して、Web 辞書編集にあたり、Lexonomy の良い点および問題点を探る。

4. 構築方法

Lexonomy を用いた Web 辞書の構築には、まず、アカウントを作成する必要がある。その後、辞書の名前、辞書の URL、単一言語辞書か二言語辞書かを定める。1つのアカウントが構築できる辞書の数の上限はない。ユーザー設定で辞書編集者のメールアドレスを追加すれば、追加された編集メンバーが辞書の編集に参加可能になる。

一般的に、辞書に登録する見出し語の情報としては、読み方、品詞、用例などがある。Web 辞書はエントリーに関するページ数の制限がないため、用例や解説など、重要なメタ情報も追加できる。日中俳句辞書の編集にあたっては以下の7項目を設定し、共同で編集を行った。以下に入力した項目とその説明、および例を示す。

- ① 見出し語の登録：俳句の初心者にとって、高頻度の季語から順に学ぶのは効率的である。例：春風
- ② 品詞情報：季語はほとんど名詞だが、稀に動詞化の季語もある。例：名詞
- ③ 季節情報：自然環境の異なる国に生まれた日本語学習者にとって、季語から季節感を読み取ることは容易ではないため、明記する必要がある。例：春
- ④ BCCWJ 情報（コロケーション）：コーパスからコロケーションの豊かさを発見できる。例：春風に舞う
- ⑤ JLPT レベル：JLPT のレベルは語彙難易度を判断するための重要な情報である。季語の JLPT レベルは「リーディングチュウ太」を用いて調べる。例：級外
- ⑥ リーダビリティ：「日本語教育語彙表」を利用し、語の読みやすさを調べる。「日本語教育語彙表」に記載がない場合は、測定不能とした。例：測定不可
- ⑦ 語彙の中国語訳と漢詩等情報：既有知識を活性化して、季語の理解に役立てる。例：
【杜甫漫興絶句】恰似春風相欺得，夜來吹折數枝花。

4.1 見出し語の選出

日本語の Web ページには、数多くの俳句作品が一般公開されている。本研究では、俳句愛好家で作った現代俳句 2,854 句 (54,106 字) を対象とし、春の句 754 句(14,144 字)；夏の句 695 句(13,615 字)；秋の句 766 句(14,531 字)；冬の句 639 句(11,816 字)を収集した。俳句歳時記には、新年の句も特別な存在として存在するが、本研究では、新年の句と冬の句は特に区別しない。

収集した俳句をオンラインテキストマイニングツール「User Local AI テキストマイニング」(<https://textmining.userlocal.jp>) 及びオンライン形態素解析ソフトウェア「Web 茶ま

め」(<https://chamame.ninjal.ac.jp/>)によって形態素解析を行い、俳句によく使われる語彙の頻度順リストをそれぞれ作成した。User Local AI テキストマイニングによる解析では「コロナ禍」、「春うらら」など、4拍以上で、まとまりの良い長単位の語彙素が抽出された。季語には、「山笑う」、「風薫る」、「冬近し」、「行く年」など、長単位の語も存在するため、本研究ではAIテキストマイニングの解析結果を用いることとした。

俳句には「や」「かな」「けり」などの切れ字がよく使われる。伝統的に、俳句は右から左への5拍、7拍、5拍の三列から構成される。言い換えれば、俳句の構造は右5拍、中7拍、左5拍の三部構成である。各部の中では、1拍か2拍の切れも含まれる。したがって、俳句の構成要素として3~7拍の単語が特に重要な役割を果たす。

上記を踏まえ、合本俳句歳時記(2008)を参照しながら、季節ごとに出現頻度の高い順に各季語を25語ずつ抽出し、登録単語として選定した。抽出した季語の②~⑦の情報を調べて、それぞれの項目を整理して共同で登録を試みた。

4.2 コロケーション情報の付与

俳句の中の日本文化は、多くはコロケーションの中に現れる。特に推敲を重ねて作り上げられた俳句には感心させられるような独特なことばのつながりがある。そこから新たな知識の発見と作者の柔軟な思考回路を伺うことができる。

コーパスに基づく辞書編集では、コロケーションの特定が容易である。本研究のコロケーション情報は、NINJAL-BCCWJを利用して、その辞書情報を抽出した。事例を調べる時に、詩など韻文特有のコロケーションに注目した。

5. 研究結果

本研究で登録した季語の一部を表1に示す。表1は季節ごとに頻出季語の上位3語を示しており、春は「桜」、夏は「蟬時雨」、秋は「紅葉」、冬は「霜柱」がそれぞれ最頻出の季語であった。

表1：現代俳句の頻出季語

春	夏	秋	冬
桜	蟬時雨	紅葉	霜柱
花筏	青田	秋刀魚	寒椿
風光る	夕立	彼岸花	冬木

Web辞書を構築する際には、いくつかの問題点が発生した。まず、見出し語のIDがランダムに付与されるため、IDから登録の順序を判断することができない。また、外部サイトを開くためのハイパーリンクを埋めることもできない。Lexonomyは小中規模の辞書編集作業には向いているが、大規模の辞書を編集する際は、xmlファイルをオフラインの状態編集の方が効率的である(Bartolomé-Díaz & Frontini, 2020)。そして、見出し語

は五十音図順配列にならないという問題がある。季語を季節ごとに並べるには、見出し語の前に、数字やアルファベットを追加するといった工夫が必要である。

一方で、見出し語の配列の順序は、「秋刀魚（さんま）」、「秋思（しゅうし）」、「秋茜（あきあかね）」のように、同じ漢字が隣り合わせで出現するため、漢字の読みの多様性の特徴が発見しやすくなる。また、既存の見出し語ページをコピーして新規ページを簡単に作成できる。これらの機能を活用することで内容重視の辞書を編むことができる。

6. おわりに

本研究では、Web 辞書編集ツールを活用したオンライン日中対照俳句歳時記を提案した。Web 日中俳句辞書の編集や活用を通して学習者は使用頻度の高い季語を学び、俳句作りを通して日本文化とその背景を学ぶことができる。俳句を教える日本語教師は、本研究の手法を取り入れることで、日中対照俳句辞書以外にも学習者のニーズに合わせたメタ辞書の編集が簡単にできる。日本語学習者は季語の調査を通して日本文化への理解の向上を図り、学習の動機づけを高める効果も期待できる。

本研究で使用した Lexonomy の編集プラットフォームは、日本語学習者と日本語教師の両方のユーザーにやさしいインターフェスを備えているため、プログラミングの知識を持たない者にとっても比較的容易に始めることのできるシステムである。投野（1995）によれば、辞書は権威的な存在であり、批判にさらされることが少ない。これまでの学習辞書は学習者のニーズをあまり考慮せずに作られることが多かった。日本語教師は今回提案した俳句辞書以外にも自ら学習が必要とする Web 辞書を構築することによって、学習者に寄り添う指導が可能となる。

日本語教育では、アクティブ・ラーニング（AL）の効果が認められ、それを取り入れた授業実践が増えつつある（横溝ほか 2019）。AL としてのプロジェクト学習（PBL）の一例として、俳句歳時記の構築は協働学習能力を身につけたり、コーパスの活用による ICT リテラシーの向上にも寄与したりできる。辞書編集を通じて語彙を増やすことも期待される。

今後の改善案として、イメージ図の登録が挙げられる。俳句は絵画的特徴を持ち、心の中の絵を喚起することで読む人の共感を誘うものであるため（皆川 2005）、イメージ図は理解の助けになると考えられる。また、学習者のニーズに合わせて、多言語訳の情報を追加できれば、より多くの日本語学習者に届けることができる。

本研究で構築した web 歳時記は、「日中対照俳句歳時記」（<https://www.lexonomy.eu/JapaneseChineseHaikuDictionary/>）として公開する予定である。

付記：本稿をまとめるにあたり、指導教員である金丸敏幸先生の貴重なご助言に負うところが大きく、ここに記して心から感謝を申し上げます。

参考文献

- Bartolomé-Díaz, Z. & Frontini, F. (2020) 'Building a domain-specific bilingual lexicon resource with Sketch Engine and Lexonomy: Taking Ownership of the Issues' *Proceedings of the Globalex Workshop on Linked Lexicography, Language Resources and Evaluation Conference(LREC2020)*, pp.62-68. (<https://www.aclweb.org/anthology/2020.globalex-1.11.pdf>)
- Měchura, M. B. (2017) 'Introducing Lexonomy: an open-source dictionary writing and publishing system' in *Electronic Lexicography in the 21st Century: Lexicography from Scratch. Proceedings of the eLex 2017 conference*, pp.19-21. (<https://www.lexonomy.eu/docs/elex2017.pdf>)
- 伊藤秀明・石井容子・武田素子・山下悠貴乃 (2016) 「日本語学習者のネット利用状況と学習サイトへの期待-海外 11 拠点の調査結果から-」 国際交流基金日本語教育紀要 Vol.12, pp.97-104. (https://kansai.jpf.go.jp/clip/minato/others_2016a.pdf)
- 伝康晴・荻野綱男[編] (2019) 『講座日本語コーパス 7 コーパスと辞書』朝倉書店
- 投野由紀夫 (1995) 「学習者から見た英語辞書」 現代英語教育 Vol.2, pp.10-13, p.65.
- 徳井厚子 (1996) 「留学生に俳句を教える-日本語・日本事情教育の中で-」 信州大学教育学部紀要 No.90, pp.1-7. (<https://core.ac.uk/download/pdf/148767085.pdf>)
- 野田尚史 (2011) 「コミュニケーションのための日本語学習用辞書の構想」 日本語/日本語教育研究 Vol.2, pp.5-32. (https://www.cocopb.com/download/2011_1_noda.pdf)
- 丸山岳彦・山崎誠・柏野和佳子・佐野大樹・秋元祐哉・稲益佐知子・田中弥生・大矢内夢子 (2012) 『『現代日本語書き言葉均衡コーパス』に含まれるサンプルおよび書誌情報の設計と実装』 特定領域研究「日本語コーパス」平成 22 年度研究成果報告書
- 皆川直凡 (2005) 『俳句理解の心理学』北大路書房
- 横溝紳一・山田智久 (2019) 『日本語教師のためのアクティブ・ラーニング』くろしお出版
- 吉田将 (1986) 「辞書構築における諸問題」 情報処理 Vol.27, No.8, pp.933-939.
- 李在鎬・石川慎一郎・砂川有里子 (2018) 『新日本語教育のためのコーパス調査入門』くろしお出版

参照サイト・辞書

- NINJAL-LWP for BCCWJ (<https://nlb.ninjal.ac.jp> 2021 年 6 月 22 日アクセス)
- Web 茶まめ (<https://chamame.ninjal.ac.jp> 2021 年 6 月 22 日アクセス)
- 特定非営利活動法人 HAIKU 日本 (<https://haikunippon.net> 2021 年 6 月 22 日アクセス)
- 日本語教育語彙表 (<http://jreadability.net/jev> 2021 年 6 月 22 日アクセス)
- リーディングチュウ太 (<https://chuta.cegloc.tsukuba.ac.jp> 2021 年 6 月 22 日アクセス)
- 角川学芸出版編 (2008) 『合本俳句歳時記』(第四版)