

Web ツールを利用した専門語彙学習

—初級後半からできる「自分に必要な」語彙リスト作成—

矢沢 理子、伊藤 秀明
国際交流基金 関西国際センター

要旨

事例のコースは、社会学、人文学系の大学院生以上をメインターゲットとした専門日本語コースで、国籍、専門分野のみならず日本語能力レベルにおいても初級後半から上級までばらつく参加者多様性を特徴としている。コース期間6ヶ月中の日本での研究活動実践を目指し、行動志向型の日本語学習をデザインしているが、殊に初級後半の学習者には、専門日本語の基礎となる専門語彙の不足により資料収集や研究取材に支障をきたす例も見受けられたため、2013年度には初級後半～初級修了レベルを対象に**専門語彙**クラスを開講した。必要とする専門語彙がそれぞれ異なることから、クラスでは各自の選んだ文献からの専門語彙リスト作成を目標とし、語彙抽出には読解ではなく、東京大学の「**言選web**」を用いて各自のe-text文献を解析、キーワードを切り出し、Excelの機能を用いてリスト化の助けとした。この手法は専門／一般の別を問わず自律的語彙学習に有用である。

【キーワード】 初級から、多様性対応、「**言選web**」、自律的語彙学習

1 はじめに

国際交流基金関西国際センター（以下、JFKC）は、海外の日本語学習者支援を旨に国際交流基金が1997年に開設した日本語研修施設で、海外諸機関に所属する専門職従事者（外交官、公務員、研究者、司書、学芸員）を対象とした専門日本語研修と、日本語学習者（主に高校生、大学生）を対象とした学習奨励研修を提供している。

本稿で扱う語彙学習の事例は、専門日本語研修の一つ、文化・学術専門家日本語研修（Japanese Language Program for Specialists in Cultural and Academic Fields, 以下、CA）の6ヶ月コースにおいて、初級後半～初級修了レベル¹の学習者11名を対象に、2013年秋に初開講した**専門語彙**クラスにおけるものである。

CA研修は、専門業務のために日本語を必要とする研究者、司書、学芸員とその「卵」としての大学院生を対象とした公募研修で、その専門の多様性から専門日本語には共通ニーズが見出し難いため、「研修中の日本における**専門活動**²の実施」という行動目標を立て、研修参加者の主体性を前提に行動志向型コースをデザインしている。

日本語学習は各自の専門活動を下支えするものとして、一般日本語科目と専門日本語科目³に分けて提供しているが、専門を語るにも文献を読むにも不可欠な専門語彙については、専門内容に関わり、個別性が殊に高いため、教師側から学ぶべき語彙を提示することができず、学習者本人の収集・整理に任せてきた。しかし、言わんとすることを辞書を頼りに日本語に置き換える、**本国語から日本語への翻訳は、コミュニケーション上の齟齬が生じやすい**。例えば、以下のようなやりとりは決して珍しいものではない。

(研究取材のための外出について)

T: どこへ行きますか?

S: 京都へ行きます。仏像の… (メモを見る) 博覧会を見ます。

T: 京都の博物館? どの博物館?

S: いいえ、お寺、有名の、仏像、33年…

T: ああ、三十三間堂? あそこはたくさんありますね。

S: いえ、一つだけ、とてもめずらしい…

仏像研究をテーマとしたこの研究者が伝えたかったのは、「京都の古刹へ33年に一度の〇〇仏のご開帳を見に行く」ことだった。しかし、expositionの訳語として選んだものが「博覧会」であったために、Tの頭には「博物館」が浮かび、否定されて「寺」となっても「33年」は「三十三間」の言い間違いだろう、という過修正が働いてしまう。

実際、Weblio⁴で exposition を引くと、出て来るのは件の「博覧会」を筆頭に、展示、陳列、解説、公開、開帳、顕示、展示会、展覧会、曝露、エキスポ、祖述、遺棄、提唱部、等々、異なり数のべ40語余りに及び、この中から最適な語を選ぶことは容易ではない。しかし、逆に日本語の「開帳」から同じWeblioを引くと、出て来るのは「賭場を開帳する run a gambling house」と「秘仏を開帳する unveil a treasured Buddhist image」の2例のみで、ヤクザ研究なら前者を、仏像研究なら後者を選び取ることに迷いはなくなる。

このため、CA研修では専門語彙の学習方法として、**本国語からの翻訳ではなく、専門読解**クラスの個別作業を通じて洗い出されてきた語彙を学習者本人が収集・整理する**日本語文献からの語彙リスト作成**を推奨してきた。迂遠な方法ながら、中上級ともなれば、文献の読解作業の中で専門語彙をリスト化していくことは可能である。しかし、新聞読みにも至らない初級レベルでは、生のテキストからの語彙抽出は難しく、専門語彙の不足から研究取材や意見交換にも学術的な読みにも至れず、自律的に専門日本語を伸ばして行く方途もつかめないまま、6ヶ月のコース修了を迎えてしまう例が少なくなかった。

そこで、2013年度のコースでは初級後半～初級修了レベルの参加者を対象に専門語彙積み上げの自律学習を支援するため、**専門語彙**クラス(100分×8回)を前期(9月末-11月末)科目としてコース開始直後から提供することにした。

2 専門語彙クラスのデザイン

日本語文献からの語彙抽出、リスト作成を方法論的前提に、初級から専門語彙をクラス形態で提供するとすると、問題は大きく分けて3つある。

1. クラス形態提供上の問題: 各自個別の内容をクラスでどう扱うか
 2. 日本語レベルの問題: 初級後半の日本語で専門文献からどうやって必要な語彙を選び出し、リスト化できるのか
 3. 語彙学習上の問題: 語彙の記憶、定着と学習自律化をどう支援するか
- こうした問題の解決を図るのがクラスデザインの眼目となる。

2.1 クラス形態提供上の問題: 多様性を活かす個別作業成果のシェアリング

学術語彙(例: 理論、分析、述べる、等、分野を問わず用いられる語彙)⁵を除けば、学習内容に重なりが見出せない語彙学習者グループをクラスとして如何に成立させるか、がデザイン上の課題である。

各自個別の内容をクラスで一斉に扱うことはできないため、目標を各自による専門語彙リストの作成とし、作成スキルの指導に焦点化した。この行動目標、スキル焦点、個別作業中心は、他の専門日本語科目とも通底する、多様性対応のクラスデザインである。一方でクラス内のインタラクションと各自の学習経過の確認のため、シェアリングセッションを複数回設けた。シェアリングは、1) キーワードマップ紹介、2) 専門活動実践を通じて収集した語彙の紹介、3) 作成途中のリスト報告 2 回の計 4 回である。

図 1 にキーワードマップの、図 2 に専門語彙リストの実例を示す。

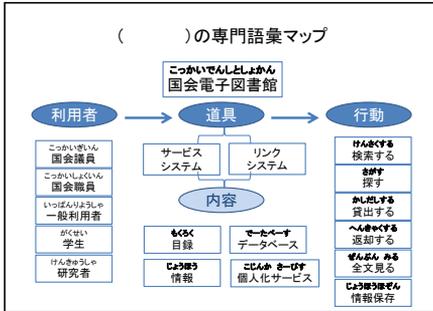


図 1 キーワードマップの例

図 2 専門語彙リストの例

このシェアリング自体が語彙使用の機会となり、2.3 で後述する語彙の定着と学習の自律化に寄与する。

2.2 日本語レベルの問題：Web ツール利用による作業促進

専門に関する日本語文献からの語彙抽出によるリスト作成という作業課題と、初級段階という日本語能力レベルのギャップは、リスト作成の各工程に立ちはだかる。

- 1. 専門語彙の抽出元選定；適切な日本語文献をどうやって選ぶのか
- 2. 専門語彙の抽出方法；上記文献からどうやって専門語彙を抜き出すのか
- 3. 語彙リスト作成作業において、
 - 1) 読み付与：漢語または漢字を含む語にどうやって読みをふるのか
 - 2) 意味付与：各語をどうやってふさわしい訳語と結ぶのか
 - 3) 量的・質的整理：絞り込み、優先順位付け、範疇化をどう進めるか

就中、上記-2.の専門語彙の抽出方法については、文献の読解を通じて語彙を抽出するのは初級後半～初級修了レベルの学習者にはハードルが高すぎ、甚だしく非効率であるため、「言選web」(<http://gensen.dl.itc.u-tokyo.ac.jp/gensenweb.html>) による e-text 解析を手法として選択した。さらに、抽出された語彙の処理、-3.の 1) 読み付与と 3) 整理には Excel の関数と機能を、-3.の 2) 意味付与には辞書サイトや検索エンジンを利用した⁶。

語彙リスト作成工程		利用ツール／サイト
1	文献 (e-text) 選択	CiNii 等の学術情報 DB、Google 等の検索エンジン
2	語彙抽出	「言選Web」
3	1) 読み付与	Microsoft Excel の PHONETIC 関数
	2) 意味付与	辞書サイト、Google 等の検索エンジン
	3) 整理	Microsoft Excel のフィルター機能

表 1 専門語彙リスト作成工程上の Web ツール利用

以下、各工程の Web ツール利用の実際と効用及び課題について述べる。

2.2.1 語彙抽出元となる e-text の選定

CA 研修では、資料収集活動をナビゲートするために、JFKC 図書館の専任司書による情報検索講習（①図書検索、②雑誌記事検索、③分野別参考図書資料紹介）を行っている。このうち、②の雑誌記事検索では、CiNii（論文記事：<http://ci.nii.ac.jp/>）や、一般雑誌記事、新聞記事の DB が紹介され、参加者各自が自身のキーワードで記事文献渉猟に着手する。

専門語彙クラスで用いる e-text 選定はこの雑誌記事検索講習とリンクする形で進めたが、関連の論文記事が公開されていない場合もあり、情報記事まで範囲を広げて探索した。しかし、キーワード検索でヒットした記事のどれがどの程度研究テーマに関連しているかについては、記事を瞥見した上での取捨選択となるため、教師の補佐が意味を持つ。

2.2.2 専門語彙の抽出方法：「言選 Web」の利用

「言選 web」（2003 年 4 月～）は、東京大学が提供している「専門用語（キーワード）自動抽出サービス」で、中川、森、他（2003）が専門分野コーパスからの専門用語自動抽出により「up-to-date な」分野別用語辞書作成を可能にすべく開発したシステムを、小島・前田（2003）が専門文献からキーワードを重要度⁷順に抜き出し一覧提示するという研究者向けサービスとして簡便化し、インターネット上で一般の用に供したものである。

語彙の抽出は、図 3 に示すように 1) e-text の入力（URL、またはテキストのコピー&ペースト、または PDF ファイルの参照）→2) 重要度のスコア付きを選択する⁸→3) 「専門用語（キーワード）抽出」ボタンをクリックする、という 3 ステップで済み、瞬時にして、図 4 のように txt.ファイル形式でキーワード（左列）が並ぶ。右列は重要度スコアである。

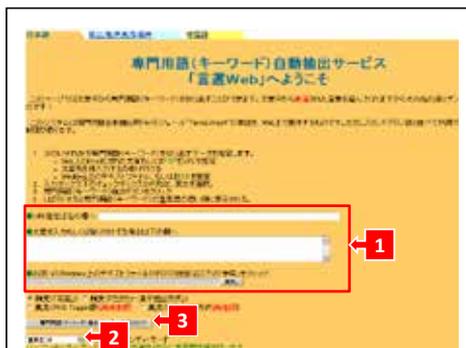


図 3 「言選 web」の使用法

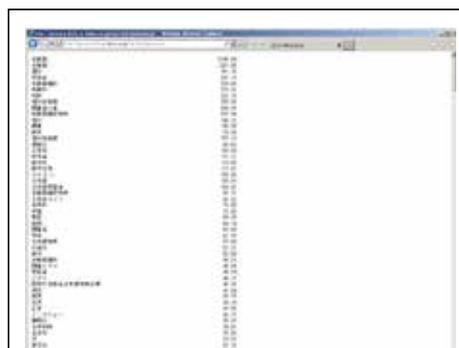


図 4 処理結果例 (.txt ファイル)

キーワード切り出しの精度については 1) 言語形態面と 2) ターム性、即ち、キーワードとしての妥当性の両面からの検討が必要になるが、この段階のリストは日本語と数値情報のみで学習者には判断不能であるため、教師が目を通さなければならない。

1) 言語形態面については、抽出元文献による異なりが大きく⁹、一概には言えないが、重要度上位群では 9 割方、用語としての態を為していると思われる。2) 真正のターム性については、専門性とも抽出元文献選定の妥当性とも絡み、門外漢の日本語教師には判断不能である。そこで、本実践では、語彙学習のためにリスト化する価値があるかどうかの学習妥当性に置き換えて判断した。すなわち、学習者にとって研究について語るために必要にもかかわらず未習の語彙がほとんど という結果であれば、採用することとした。

2.2.3 語彙リスト作成作業：Excel の関数・機能利用

抽出された語彙データの加工は Excel に移行後、可能になる。まず、教師補助により形態的に不適な語彙、次に学習不要な語彙を除き、重要度下位群を切り捨ててリストを粗絞りした上で、1) 読み付与、2) 意味付与、3) 整理の作業を行った。

1) 漢字語のルビふりは発音のためのみならず、意味を調べるためにも欠かせない。読み付与のために Excel の PHONETIC 関数を利用して作業の効率化を図った。但し、アルゴリズムによる読み付与なので、一般辞書の読みの第一候補が付されてしまう、固有名詞の読みの精度が劣る、等の問題があり、ここでも教師チェックが必要になる。

2) 意味付与には、前出 weblio 等のオンライン辞書の利用を奨めた。上述したように、日本語から自国語への翻訳であれば、辞書に複数の訳候補があっても自国語の候補なので、自身の専門の文脈にふさわしい訳語の同定はさほど困難ではない。しかし、専門用語辞書にも記載されていない特殊な日本語語彙もあり、漢字の意味や google 等の情報記事を頼りに教師が意味を解説し、学習者の自国語翻訳を手助けする場面もあった。意味付与の後、学習者自身が語彙の要不要を判断し、リストを更に絞り込むことが可能になる。

3) 絞込みのデータ削除は手作業になるが、範疇化のためには、Excel のフィルター機能が使える。2.1 の「キーワードマップ」作成作業で洗い出されたカテゴリ名を用いてリスト上の語彙をタグ付けし、分類・整理した。

2.3 語彙学習上の問題：語彙の定着と学習の自律化

語彙学習は文字、音と意味の結びつきが 1) 知識として定着し、2) 実際の場面で使えるようになって初めて実を結ぶ。また、語彙力の拡大発展には、3) 学習自律化が欠かせない。

1) Quizlet (<http://quizlet.com/>)等の記憶練習ツール紹介を授業に取り入れ、認知的訓練の自律化を図るだけでなく、2) 実用練習の場面としてシェアリングセッションをデザインに組み込んだ。学習者に自身の専門語彙を説明するという「教師役」をふることで、準備段階から実演まで、学習したばかりの語彙は繰り返し確かめられ、説明のための文脈が与えられ、クラスメイトに伝わる形で言い換えられ、使用されることになる。また、3) シェアリングでは、マッピングや整理といった方法論においても、語彙そのものについても、他者の例に触れる経験を通じて自律学習を支えるメタ的な気づきが得られる。紹介／報告された語彙の中には、それぞれの文献では異なる日本語として出現しながら、自国語訳では同一語に収斂するもの（例：“art”←芸術、美術、芸術／美術作品、わざ、術、技能、人工など）もあり、自国語からの適切な翻訳の難しさと同時に、語彙を文脈内で意味づけることの重要性が認識できる。こうしたメタ的な気づきが、学習自律化への契機となる。

3 初級からの「自分に必要な」語彙学習

第二言語教育の世界では、一般語彙→学術語彙→専門語彙の順に学習が進むのが通例だが、この積み上げ順は第一言語の発達に準えたものであろう。身近、直近の必要性の高いものは具象・具体物から始まり、抽象へと進む。だが、第一言語の場合、語彙の拡大が意味概念の獲得・発達とほぼシンクロするのに対して、第二言語学習者にとっては既に第一言語で確立している概念を第二言語の単語と結びつけるという「翻訳」の作業になる。また、学術目的の言語学習であれば、身近、直近の必要性の高い語彙が日常生活語彙ではなく、研究について語るための抽象度の高い語彙になるのも不思議ではない。とすれば、初

級からの専門語彙学習は、難しい言葉の学習が易しい言葉の学習に先立つという逆順ではなく、必要度の高い語彙が必要度の低い語彙に先立って学ばれるという正順合理の営為に他ならない。そして、「必要」という点から見ればそれが専門について語るためであろうと、趣味について語るためであろうと、「今、自分にとって『必要』か否か」という区分けしかなく、一般と専門、日常と学業の語彙範疇別は用を為さなくなる。「初級には難しい」というのは、むしろ教師側の思い込みから生まれる「学習の壁」なのではないだろうか。

注

¹ CA 研修への日本語能力応募要件は「初級修了レベル以上」だが、実際の参加者の中には初級段階の学習者も含まれる。これは、業務上の必要から日本語学習を始めた CA 参加者の多くが非日本語専攻で、正規的な日本語コースからは縁遠い学習環境に置かれている例もあり、レベル相当観に彼我のズレが生じるためである。ここでいう「初級後半」、「初級修了」レベルは、来日後のプレイスメントテストの結果に基づくレベル査定で、一般日本語科目の総合日本語では、初級後半レベルは『みんなの日本語 初級Ⅱ』を、初級修了レベルは『J-Bridge』を使用していた。

² 専門活動には参与観察、インタビュー、資料収集、文献読解、関係機関での実習やゼミ参加、意見交換など、様々な形態がある。

³ 専門日本語科目は、専門活動計画を立案、実施するための専門活動準備、研究テーマ紹介を目的とした発表、各自の文献読解を教師が補助する専門読解、小論文作成のための文章表現など、スキルまたは個別作業に焦点を当てた形態での提供である。

⁴ <http://ejje.weblio.jp/>。専門用語辞書を含む横断検索が利いて、現在 online 辞書の中ではもっとも簡便かつ包括的と思われる。

⁵ 学術語彙に関しては、コース自作の「研究／専門紹介のための学術語彙リスト」を発表クラスで扱い、また、専門語彙クラスの初回で Coxhead (2000) の「Academic Word List」の日本語訳を「学術語彙リスト」にまとめ、参考に供している。

⁶ 紙幅の都合から本稿では触れられないが、使用方法詳細については伊藤作成のクラス配布資料「専門語彙リスト・カードの作り方」がある。(資料請求は Hideaki_ito@jpf.go.jp まで。)

⁷ 専門用語の多くは複合語、とりわけ複合名詞であることから、1) 複合名詞自体の出現頻度と 2) 複合名詞を構成する最小単位名詞、「単名詞」と他の「単名詞」との接続種類数及び出現頻度数を掛け合わせて重要度を算出している。(中川、他：2003、小島、他：2003 参照)

⁸ この「重要度スコア付き」のステップを踏まざとも結果は重要度順に並べられリスト提示されるが、後のリスト整理の段階で数値が絞り込み判断の一助となるため、敢えて手間加えている。

⁹ 抽出元文献に固有名詞が多く含まれたり、カタカナ語等文字種に偏りが見られたりする場合は精度が落ちる印象がある。抑、「論文」の語種的特徴の一つに漢語率の高さがあり、当抽出システムが「専門用語の多くは複合語、とりわけ複合名詞」(中川、他 2003)であることを前提に設計されていることから、切り出し結果は漢語名詞において最も安定するものと考えられる。「言選web」では形態素解析には「茶筌」と「カタカナ・漢字抽出方式」の二つが試せるようになっている。

<参考文献>

Coxhead, A. (2000) A New Academic List. *TESOL Quarterly* vol.34 (2):213-238.

伊藤秀明、矢沢理子 (2014) 「学習者による『私の』専門語彙抽出とリスト化」、『第 16 回専門日本語教育学会研究討論会誌』, pp. 8-9.

小島浩之、前田朗 (2003) 「キーワード(専門用語)自動抽出システムの構想とその展開」、『第 51 回日本図書館情報学会研究発表要綱』, pp.17-20.

中川裕志、森辰則、湯本紘彰 (2003) 「出現頻度と接続頻度に基づく専門用語抽出」、『自然言語処理』 Vol.10 No.1, pp. 27 - 45, 言語処理学会.