

# Selecting Focused Vocabulary Using a Domain-Specific Corpus in Japanese Language Education:

## Recipe Corpus in Food and Beverage Japanese Language Education

### 領域特化型コーパスを用いた，重点指導語彙の選定：

#### 飲食系日本語教育におけるレシピコーパスの活用

Hiroki Fxyma

Tainan University of Technology

h.fxyma@gmail.com

#### ABSTRACT

*This study conducted a vocabulary survey for teaching Japanese vocabulary related to food and beverage in a beginner's Japanese class. Frequently occurring words were extracted by part-of-speech from a large recipe corpus, and the target range of vocabulary to be taught was investigated. As a result, we concluded that the range of vocabulary that requires intensive instruction is "intermediate to high-intermediate words with frequencies up to 400. In addition to the traditional vocabulary of "names of ingredients, names of dishes, and categories of dishes," "cooking procedures and plans" and "objects and characteristics of recipes" were found to occur with relatively high frequency, indicating the need to teach intermediate and advanced vocabulary in these categories even to beginner-level learners in order to accurately read recipes and perform cooking tasks. The results of the study revealed the necessity of teaching intermediate and advanced vocabulary in these categories to beginner-level learners in order to read recipes accurately and perform cooking.*

## 1. はじめに

### 1.1. 研究背景

筆者は台湾南部の大学で餐飲系という料理，飲食系の学科に所属している。

飲食に関する日本語教育では，いわゆるレストランサービスの接客系日本語教育と，調理に関する技術系日本語の教育の両面が求められる。接客系日本語教育は敬語や定型的な接客の言い回しを，日常の会話と区別して習得することが目標とされる。したがって初級の会話能力は前提条件であり，中級以上の日本語運用能力をもつ学習者が，場面に応じた言い回しをすることが目標である。

一方で調理に関する技術的な日本語は，言い回しなどではなく，食材の名前や，「焼く，煮る」などのいわゆる調理動詞を修得することが主目標となる。筆者の所属では，語学科目の開講数が少ないため，初級日本語のクラスにおいて，基本文法を習得させつつ，語彙知識として調理に関する語を導入することを目標としている。

### 1.2. 本研究の目的

本研究では専門日本語教育において，領域特化型のコーパスを活用することによって，重点指導語彙を

導出することを目的とする。

## 2. 重点指導語彙の選定

### 2.1. コーパスの概要と

#### 2.1.1. 使用するコーパスデータについて

本研究において使用したコーパスは、クックパッド株式会社が国立情報学研究所を通じて研究者に提供しているデータセットである。レシピ投稿 web サイト「クックパッド」に掲載されたデータで、172 万品のレシピやそれらからなる献立に関するデータが含まれている。データセットの情報サイズは 5.5GB と、領域特化型のコーパスとしては大規模なデータセットである。当該データセットにはクックパッドで 2014 年 9 月 30 日までに公開されたレシピデータとして、レシピのタイトル、概要、手順、つくれば（「作りましたフォトレポート」の略）、カテゴリなどのデータが含まれている。また献立データとして、2014 年 9 月 30 日までに公開された献立に関するデータです。献立のタイトルや献立に含まれるレシピ、各レシピが主菜か副菜かといったデータが含まれている。

本研究ではこのうち、レシピデータの「概要」を使用した。「概要」はレシピ投稿時にユーザーがそのレシピの制作背景や調理の説明などを記述する箇所である。「タイトル」や「手順」が単語や箇条書きの短文になることに比べて、「概要」は、調理に関する書き言葉コーパスとして自然な日本語を抽出することに適している。

分析では、この「概要」データのうち不要な記号、顔文字などを削除した総語数 1000 万語レベル、異なり語数 5.6 万語を使用した (Table 1)。投稿の総数は 58.5 万件であり、コーパスでは 1 投稿を 1 段落にしている。総文数は 77.9 万文である。なお、本研究のコーパス分析においてはテキストマイニングツール KH Coder 3<sup>1</sup>を使用し、形態素解析には KH Coder 3 内の MeCab を使用した。

Table 1 コーパスの詳細

総抽出語数	Token	10,033,667
異なり語数	Type	56,161
総文数	sentences	779,088
総段落(投稿数)	paragraphs	585,763

#### 2.1.2. 各品詞の頻出語

本コーパスの頻出語は Table 2 に示す通りである。名詞の上位 10 語は、料理ジャンルが多く、個別の料理名や食材は見られない。形容動詞ではレシピの特徴（簡単、ヘルシー、手軽など）あるいは対象（野菜が苦手など）を示す語が頻出している。動詞では調理動詞が見られ、形容詞では料理の香味（甘い、香ばしい、辛い）に加えて、季節や状況を示す暑いが見られる。副詞ではオノマトペ（音象徴語）が上位 10 語のうち 5 語を占めている（とっては音象徴語ではなく一般副詞とみなす）。

<sup>1</sup> 樋口耕一、「テキスト型データの計量的分析」。

Table 2 品詞別の頻度上位 10 語 (数値は頻度)

名詞	形容動詞	動詞	形容詞	副詞					
味	45229	簡単	93137	作る	59626	美味しい	98413	さっぱり	15419
野菜	24363	ヘルシー	15441	食べる	57254	甘い	18344	とっても	15327
レシピ	21925	大好き	11076	使う	42211	いい	10801	ちょっと	12129
ご飯	21006	好き	8689	入れる	21466	よい	8866	どうぞ	11760
一品	16535	手軽	8556	焼く	18932	良い	7869	たっぷり	11663
料理	16589	シンプル	8354	出来る	18576	安い	4795	いつも	8583
サラダ	16058	いかが	4340	作れる	14582	香ばしい	4650	夏	7921
弁当	15411	濃厚	4110	合う	12873	優しい	4542	時間	6949
パン	14322	苦手	3797	混ぜる	12431	辛い	3524	ぴったり	6946
ケーキ	13975	普通	3797	入る	11414	暑い	3334	あっさり	6360

## 2.2. コーパス出現語の学習レベル構成

本コーパスはレシピコーパスであるため、一般コーパスとは使用語の頻度やレベルの傾向が異なることが予想される。料理など、あるドメインに特化した日本語を教授する上では、一般的な指導語彙リストと比較して、そのドメインで使用される語がどのようなレベル構成になっているかを把握することが求められる。レベル構成が異なっていた場合には、重点的に指導する必要があるためである。例えば上位語のレベル構成が中・上級語を多く含んでいた場合には、一般的な語彙指導に加えて、特別な語彙指導を要することがわかる。また品詞別の難易度構成を参照することにより、どの品詞を重点的に指導すべきかの目途を得ることが可能となる。

本研究では、日本語教育における語彙の難易度判定として、「日本語教育語彙表<sup>2</sup>」におけるレベル分けを活用する。Table 3 はレシピコーパスにおける品詞別のレベル構成 (頻度上位 2,000 語<sup>3</sup>) を示したものである。

レシピコーパス全体の難易度としては、(参考)とした日本語教育語彙表と比較しても、初級から上級までがバランスよく分布しているように見受けられる。一方、品詞別のレベル構成を見てみると、形容詞・副詞は他の品詞に比べて初級前半、初級後半の比率が高いことがわかる。

<sup>2</sup> 砂川有里子, 「学習辞書編集支援データベース作成について-『学習辞書科研』プロジェクトの紹介, 日本語教育連絡会議論文集」。なお日本語教育語彙表は 17,920 語の日本語教育用の語彙を収録している。作成時基礎資料は『現代日本語書き言葉均衡コーパス』(BCCWJ: Balanced Corpus of Contemporary Written Japanese) と「日本語教科書コーパス」である。

<sup>3</sup> 「日本語教育語彙表」に収録のない語は分析対象外としているため、「レシピ全体」列の合計は 2,000 語とはならない。

Table 3

	レシピ全体	レシピ全体の品詞別構成			(参考)
		名詞	動詞	形容詞,副詞	日本語教育語彙表
初級前半	123	49	19	49	424
初級後半	200	106	41	42	792
中級前半	310	180	51	69	2300
中級後半	401	256	74	62	6465
上級前半	156	99	32	22	6379
上級後半	35	16	13	5	1560

### 2.3. 指導すべき語の選定

次に、レシピコーパスのなかから、初級日本語の講義において指導すべき語彙を選定するため、頻度ランキングを100位ごとに区切り、そのレベル分布を分析する。分布はTable 4に示す通りの結果となった。Table 4から、概ね501位以下の語は、一般的な難度構成（レシピコーパス全体および日本語教育語彙表）と同じ傾向の分布、すなわち初級前半から中級後半に向かって増大し、上級は少ない、という分布を示すことが見て取れる。本来、一般のコーパスの上位頻出語は初級において指導することが期待される。しかし、たとえば101位から200位の100語では中級前半の語が28語、また201位から300位の100語では、中級後半が最も多い28語分布しているように、レシピコーパスにおいては、全体における頻度が最上位の語にも中級・上級の語がかなりの比率で含まれていることが確認できる。これらの語は、「一般的な初級日本語の語彙指導では指導対象外であるが、飲食日本語においては特異的に頻出する語」とみなすことができ、初級日本語教育で指導する際には重点的な語彙指導が求められる語群である。

以上に示した傾向の分析から、以下の分析における重点的な指導目標の候補は、「レシピコーパスにおける頻度上位500語のうち、中上級レベルの語」とし、その範囲の語彙がどのような傾向を示しているかを分析する。

Table 4 頻度順位100語ごとの語彙レベル構成

	1	101	201	301	401	501	601	701	801	901	1001	1101	1201	1301	1401	1501	1601	1701	1801	1901
初級前半	19	14	6	10	11	8	4	7	6	7	2	3	3	5	4	3	4	2	3	2
初級後半	28	13	7	12	14	15	16	8	9	10	8	8	11	6	6	3	8	6	7	5
中級前半	14	28	16	21	13	17	17	15	11	18	13	14	14	23	14	12	12	11	12	15
中級後半	15	21	28	12	14	24	20	20	22	18	31	19	16	15	14	29	21	21	19	22
上級前半	6	4	5	6	10	7	6	9	12	11	4	12	3	6	13	12	8	7	5	10
上級後半	1	2	0	2	1	1	2	4	3	3	1	1	2	3	1	2	2	1	2	1

**Note.** 点線枠で囲った領域は、本研究における重点指導語彙の範囲を示す。見出し行の数字は、100語ごとの区切りを示す。「1」の列の語は頻度上位1位から100位までの100語の分布であり、「101」の列の語は頻度上位101から200位までの100語の分布である。なお「日本語教育語彙表」に収録のない語は分析対象外としている。灰色の色分けは、100語ごと（列）の中での頻度に応じたカラースケールである。

### 3. 結果頻度順位 1 位から 500 位のなかの中上級語

以下の分析部分においては、コーパス内の語を示す際には日本語では下線、英字では斜体で示すこととする。(コーパスから抽出された結果を示す際には、日英ともに斜体を用いる。)

#### 3.1. 1-100 位の「中上級語」

レシピコーパスにおける頻度上位 100 語のうち、中級あるいは上級に分類される語は、以下の通り 35 語であった。なお初級語は上位 100 語のうち 47 語であり、その他の 18 語は「日本語教育語彙表」においてレベル分けがなされておらず、本研究では対象外となった語である。

##### 名詞

食, 鍋, レンジ, 香り, 風味, 使用, おかず, 生地, 大根, 豆腐, パスタ, メニュー, 味付け, 栄養, ソース, レシピ, アレンジ, 材料, 我が家, 定番

##### 動詞

つまむ, 焼く, 混ぜる, 入る, 炒める, 煮る, 加える

##### 形容詞, 副詞

さっぱり, たっぷり, ぴったり, あっさり, しっとり, ヘルシー, 手軽, シンプル

1-100 位の中上級語のうち、名詞は 20 語である。鍋, レンジのように調理器具が含まれ、大根, 豆腐といった食材、パスタ, メニューのようにカタカナ語も見られる。レシピやアレンジ, 我が家, 定番は、一般投稿型のレシピコーパスとしてのコーパス特性が反映されている。

動詞は、ほとんどが調理動詞とみなすことができる。焼くと炒めるのように、「使い分けつつ使い切る」語彙力<sup>4</sup>を要する同義語が出現している。

形容詞、副詞においてはオノマトペ（音象徴語）が特徴的である。語型としてはすべて「X ッ Y リ」型である。食べ物の味テクスチャといった感覚を表すものとしてさっぱり, あっさり, しっとりが見られる。たっぷりやぴったりは、感覚的な刺激、情報を示すのではなく、対象（食べ物）の様態、状況を示していると考えられる。

#### 3.2. 101-200 位の「中上級語」

レシピコーパスにおける頻度 101-200 位の語のうち、中級あるいは上級に分類される語は、以下の語であった。

##### 名詞

具, 好み, 手間, 手, 食欲, 相性, ボリューム, カロリー, 感じ, 甘み, 酸味, 冷凍, 皮, ねぎ, にんにく, ヨーグルト, 生クリーム, 季節, 市販, 煮物, 試し, フライパン, 丼, 菜, マヨネーズ, クリーム, ドレッシング, 酢, 油, 調味料, 和風, 中華, 納豆, 今回, 話題, 満点, 最高, 朝食, ダイエット,

##### 動詞

---

<sup>4</sup> ARCLE 編集委員会, 幼児から成人まで一貫した英語教育のための枠組み—ECF.

もつ, つける, 揚げる, 巻く, 和える, 残る, 余る,

#### 形容詞, 副詞

しっかり, ほんのり, ふわふわ, ふんわり, 香ばしい, 苦手, 普通, 濃厚, いかが

名詞は概ね食材や調理に関する名詞である。カタカナ語が多く見られる。具体的な食材に加えて、酢, 油, 調味料というように調味料, また甘み, 酸味などの味を表す名詞, さらに煮物や丼, 中華, 朝食, 市販というように料理のカテゴリを表す名詞が見られる。

動詞は調理動詞(つける, 揚げる, 巻く, 和える)が大半である。もつは「2日もつ (can be kept 2days)」のような用例, 残る, 余るは類義語で, 「残り物で一品 (A dish with leftovers.)」で用いられる動詞である。

形容詞, 副詞に関しては1-100位と同様にオノマトペが見られる。また, 香ばしい, 濃厚のように香味をあらわす形容詞も見られる。

### 3.3. 201-300位の「中上級語」

#### 名詞

玉, 旬, 鶏, 梅, 気, 豚, 豆, 鶏肉, 本格, ベーコン, 白菜, 豆乳, オイル, デザート, オープン, グラタン, チキン, 好評, 組み合わせ, 代わり, ポイント, 分量, ランチ, 手作り, 失敗, 変身, アップ, 是非, 保存, 満足

#### 動詞

たまる, ゆでる, 煮込む, 効く, 飽きる, 喜ぶ, 冷やす, 冷める, 仕上げる, 仕上がる, 試す, 蒸す, 出来上がる,

#### 形容詞, 副詞

もちろん, そのまま, さっと, 美味, 甘辛い, 柔らかい

101-200語と同様に食材やカタカナが多く見られる。組み合わせ, 分量, 保存, 失敗などはレシピの正確な読み取りと調理に求められる名詞である。

動詞では飽きる, 喜ぶのように食べる人の反応が出現している。また, 冷やす, 冷める, 仕上げる, 仕上がる, 出来上がるのように, 料理の提供に関する動詞が出現していることが特徴的である。

形容詞, 副詞ではこれまでの上位200語の傾向とは異なり, 音象徴語がさっとのみとなっている。代わって, 美味, 甘辛い, 柔らかいというように味や食感を表す語が出現している。

### 3.4. 300-400位の「中上級語」

#### 名詞

素, 量, 葉, もち, ごま, 基本, 圧力, ほうれん草, 彩り, 控えめ, 勧め, 最適, チャーハン, ピーマン, シチュー, ツナ, 食材, 調理, 完成, 手抜き, プラス, ゼリー, コロッケ, 前, 中

#### 動詞

のく, やる, 乗せる, 炊く, 添える, 炊き込む, 合わせる,

#### 形容詞, 副詞

かなり, そう, こってり, 是非, 本当に, 結構, 豪華, まろやか, すごい

201-300 位の名詞と同様に、レシピを読み取り、調理を正確に実行するために必要な名詞が見られる(量, 圧力, 控えめ, 調理, 前, 中).

動詞ではこれまでの傾向と同様に調理動詞(乗せる, 炊く, 添える, 炊き込む, 合わせる)が出現している. しかし乗せる, 添える, 炊き込む, 合わせるというように, 2つ以上の素材の調理や関係性に言及する動詞が出現しており, より複雑な関係の理解が求められる.

形容詞・副詞では, オノマトペが2語(さっと, こってり)にとどまっている.

### 3.5. 401-500 位の「中上級語」

#### 名詞

麺, レタス, 抹茶, 外, 感, 繊維, アクセント, バージョン, 甘味, 味わい, 残り物, 抜群, 人気, 絶品, 電子, 家庭, 常備, 消費, 感謝, 利用, 出来上がり, 時期, 短時間

#### 動詞

つる, なす, さる, よる, 頂く, 漬ける, 包む, 召し上がる

#### 形容詞, 副詞

とにかく, とろとろ, 絶妙, 生, 面倒, 素朴, 濃い

名詞は食材が麺, レタス, 抹茶のみとなっており, その他は甘味, 味わい, 絶品が味を表す名詞といえる. レシピの性質を表す語が多く出現しており, 調理に関する表現は見られない.

## 4. Conclusion: どの範囲の語彙を重点的に指導すべきか

### 4.1. 結論

頻度順 100 語ごとの分析により, 頻度 1-200 位の語彙と 201 位から 400 位までの語彙において, 異なる傾向が見られた. 401 位から 500 語の範囲では, レシピコーパスの特殊性が低下し, 重点的に指導すべき語彙は少数しか見られなかった.

According to a word-by-word analysis of the top 100 words in frequency order, different trends were observed in the vocabulary from frequency 1-200th and in 201st-400th words. 401st-500th words showed less specificity of the recipe corpus, with only a small number of words requiring intensive instruction.

上位 200 語の傾向として, 名詞は食材や調味料, 料理のカテゴリなど, 直接的に調理やレシピに関係する必須語と思われる語彙が多くを占めている. 201 位以降では, レシピの読み取りと調理の正確な遂行に必要と思われる語が出現しており, 401 位以降は調理に関する表現が少なくなるものと考えられる.

The tendency of the top 200 words is that nouns are mostly essential words directly related to cooking and recipes, such as ingredients, seasonings, and cooking categories. From the 201st word onward, words that are considered necessary for reading recipes and cooking accurately appear, and from the 401st word onward, expressions related to cooking are considered to be less common.

名詞は主に「食材名」「料理名, 料理カテゴリ」「調理の手順や計画」「レシピの対象や特徴」などに分類が可能と思われる. この分類では, 上位 300 語までは食材名が頻出した. 一方で 301 位以降は「食材名」は低い割合となっている. 401 位から 500 位の群では麺, レタス, 抹茶のみとなった.

名詞では 401 位以降の語群ではレシピの性質を表す語が多く出現しており、調理に関する表現はあまり見られない。したがって名詞だけで考えると、400 語レベルまでの名詞で食材、調理に関係する主な語は十分に出現していると考えられる。

Nouns can be categorized mainly into "ingredient name," "dish name or category," "cooking procedure or plan," and "object or characteristic of the recipe. In this classification, the top 300 words were the names of ingredients. In the group of 401st to 500th words, only *noodle*, *lettuce*, and *green tea* appeared. Nouns from the 401st rank and onward are often used to describe the nature of the recipe, and there are few expressions related to cooking. Therefore, it can be considered that the main words related to food ingredients and cooking appear sufficiently in nouns up to the 400th word level.

動詞では、いわゆる調理動詞のうち主要なものは上位 200 語レベルで出現する。それ以降、400 語レベルまで、炊き込む、添えるなどのように複合的あるいは補助的な調理動詞が出現する。

Among verbs, the major so-called cooking verbs appear at the top 200 word level. Thereafter, complex or subsidiary cooking verbs such as "*takikomu*; to cook together" or "*soeru*; to garnish" appear up to the 400-word level.

形容詞、副詞では、上位 200 語レベルまでは食感や味の印象を表すオノマトペ（音象徴語）が大部を占める。香味を表す形容詞は 101 位以降、400 位レベルに出現する。食味を表すオノマトペや、一般の形容詞でも濃厚や甘辛いといった語は語意の対象依存性が強いいため、共起関係とともに指導する必要性があるものと思われる。

Adjectives and adverbs in the top 200 words are mostly onomatopoeias (sound symbolic words) that describe impressions of texture and taste. Onomatopoeias and other general adjectives such as "*noukou-na*; rich" and "*ama-karai*; sweet and spicy" have a strong object-dependence of meaning, and thus need to be taught together with co-occurrence relations.

以上の分析より、本研究の目標である、「初級日本語のクラスにおいて、基本文法を習得させつつ、語彙知識として調理に関する語を指導する」ことに照らして、重点的な指導が求められる範囲は、「頻度 400 位までの中上級語」と結論付ける。その対象範囲の中で、名詞は「食材名」「料理名、料理カテゴリ」「調理の手順や計画」「レシピ、料理の対象や特徴」といった語群を設定する。動詞は一般的な調理動詞に加え、複合的な調理動詞、食べる人の反応や料理の提供に関する動詞（201-300 位の語を参照）を指導対象とする。形容詞と副詞に関しては、オノマトペおよび香味を表す形容詞を、料理との共起関係において教示することを課題とする。

Based on the above analysis, considering the goal of this study, which is "to teach words related to cooking as vocabulary knowledge while teaching basic grammar in beginner Japanese classes," I conclude that the range of words that require intensive instruction is "intermediate and advanced words with frequencies of up to 400. The verbs to be taught are general cooking-verbs as well as complex cooking-verbs and verbs related to the reaction of the eater and the serving of the food (see words ranked 201-300). As for adjectives and adverbs, onomatopoeia and adjectives describing flavors should be taught in relation to their co-occurrence with the food.

#### 4.2. 本研究の貢献と今後の展開

一般的な飲食日本語の教材は、中上級の日本語学習者が職場における日本語を学ぶというコンセプトのもと、接客、接遇や敬語の表現に重点がおかれているものが多い。あるいは初級日本語の教材では、数課の「料理テーマ」の授業において、食材や調理動詞を扱うことが多かった。

本研究では、従来指導されてきた「食材名、料理名、料理カテゴリ」に加えて、「調理の手順や計画」「レシピの対象や特徴」が比較的高い頻度で出現することを示し、レシピを正確に読み、調理を遂行するためには初級の学習者にも、これらのカテゴリの中上級語彙を指導することの必要性があることを明らかにした。

今後の研究課題は、本研究で示した方針を実際のカリキュラムと教案に落とし込むことである。対象と目標が比較的明確な指導となるため、タスクと場面ベースのシラバスに対して語彙を組み込んでいくことを計画している。

## 謝辞

本研究では、「クックパッドデータセット」を使用した。

「本研究では、日本語学習辞書支援グループ (2015)「日本語教育語彙表 Ver 1.0」(<http://jhlee.sakura.ne.jp/JEV/>) を利用した。

## Bibliography

ARCLE 編集委員会. *幼児から成人まで一貫した英語教育のための枠組み—ECF*. 編集者：田中茂範. 東京：リーベル出版, 2005 年.

砂川有里子. 「学習辞書編集支援データベース作成について-『学習辞書科研』プロジェクトの紹介, 日本語教育連絡会議論文集」. 日本語教育連絡会議論文集, 24:164-69, 2012 年.

樋口耕一. 「テキスト型データの計量的分析」. 『理論と方法 19』, no. 1 (2004 年): 101-15.