

国立国語研究所学術情報リポジトリ

〈著書紹介〉 Yukio Tono, Makoto Yamazaki, Kikuo Maekawa. A Frequency Dictionary of Japanese: Core Vocabulary for Learners

メタデータ	言語: jpn 出版者: 公開日: 2015-10-30 キーワード (Ja): キーワード (En): 作成者: 山崎, 誠 メールアドレス: 所属:
URL	https://doi.org/10.15084/00000756

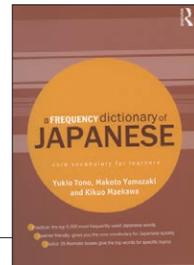
Yukio Tono, Makoto Yamazaki, Kikuo Maekawa

***A Frequency Dictionary of Japanese:
Core Vocabulary for Learners***

Series: Routledge Frequency Dictionaries

2013年1月 Abingdon/New York: Routledge

viii + 369 ページ £29.99 (ペーパーバック版)



山崎 誠

1. 本書の目的

本書は、英語を母語とする日本語学習者のための現代日本語の頻度辞書である。出版元の Routledge 社では、2005 年から学習者向けに Frequency Dictionary のシリーズを刊行しており、本書はその中の 1 冊にあたる。本書が刊行された時点で、スペイン語、ドイツ語、ポルトガル語、中国語、フランス語、英語（現代米語）、チェコ語、アラビア語の頻度辞書が先行しており、本書の刊行後ロシア語とオランダ語も追加された。

想定される学習者のレベルとしては中級以上である。日本語教師が教材を作成するなど、教える側の利用も考えられる。なお、本書の CD-ROM 版も発売されており、データを Excel 等に取り込んで利用できるようになっている。

2. 本書で使用したデータ

学習者用に安定した頻度情報を得るためには、コーパスのような大規模な言語データの利用が望ましい。本書は、『日本語話し言葉コーパス』（以下、CSJ）と『現代日本語書き言葉均衡コーパス』（以下、BCCWJ）の 2 つのコーパスを利用して、出現頻度や語の散らばり具合（dispersion）などを示している。本書で利用したデータは CSJ の模擬講演の部分と BCCWJ の全体である。ただし、BCCWJ は 2011 年 6 月の時点のものであり、同年 12 月に公開したバージョンとはわずかに違いがある。

3. 本書の構成

本書の主な構成は以下のとおりである。

- ・ Frequency Index (237 ページ)
- ・ Alphabetical Index (55 ページ)
- ・ Part of Speech Index (55 ページ)
- ・ Word Type (origins) (13 ページ)

Frequency Index は本書のメインとなる部分で、頻度順に並べられた見出し語項目のリストである。本書では上位 5000 語を掲載している。図 1 に見出し語項目の例を挙げる。

<p>35 で de conj. so, then</p> <ul style="list-style-type: none"> • で、あの話はどうなりましたか? — So what happened about the story you mentioned? <p>4412 0.15 SP</p>
--

図 1 Frequency Index の例

第 1 行目に、順位、語彙素、語彙素読み、品詞、見出し語の英訳を示した。2 行目以下には見出し語を使った例文とその英訳、最後の行に 100 万語あたりの出現頻度と Carroll (1970) の D2 という指標で表した dispersion (語の散らばり具合。値が大きいほど各レジスターにおいて偏りなく使用されていることを示す) の値、また、該当する見出し語については、特徴的に使用されるレジスターを示した。上記の例は、「で」の 100 万語あたりの出現頻度が 4412, dispersion の値が 0.15 であり、SP, すなわち話し言葉において他のレジスターに比べて多く使われていることを示している。

Alphabetical Index は、見出し語の五十音順の索引であり、各見出し語のもとに、品詞、見出し語の英訳と順位を掲載したものである。図 2 にその例を示す。

<p>納豆 <i>n.</i> fermented soybeans 4139</p> <p>納得(する) <i>n.</i> agreement <i>v.</i> understand, be convinced 1297</p> <p>夏休み <i>n.</i> summer holiday 2017</p>

図 2 Alphabetical Index の例

Part of Speech Index は品詞別の索引であり、品詞ごとに順位の小さい方から見出し語を並べ、英訳を付したものである。Word Type (origins) は、語種 (和語・漢語・外来語・混種語・固有名詞) 別の索引である。語種ごとに順位の小さい方から見出し語を並べ、英訳を付したものである。

本書には、以上の Index 以外に次の 2 つのコラムを掲載している。

ひとつは、「動物、身体、色、国、感情」といった意味カテゴリー別の語彙リストである。採り上げた 25 の意味カテゴリーについて 100 万語あたりの出現頻度を添えて掲載した。このリストには Frequency Index では採り上げなかった 5001 位以下の語も含まれている。

もうひとつは、「言う (イウ/ユー)」「私 (アタクシ/アタシ/ワタクシ/ワタシ)」など、本書でひとつの語彙素としてまとめた見出し語について、発音のゆれが大きいものを採り上げ、CSJ における学会講演・模擬講演・独話における分布を示したグラフである。

4. 本書の特徴

本書はコーパスベースで作成した初めての日本語学習者用頻度辞書である。本書の特徴としては以下の3点が挙げられる。

(1) 話し言葉と書き言葉の語彙をバランスよく収めていること。

CSJとBCCWJの出現頻度を単純に足しただけではCSJの頻度が相対的に低くなり、CSJの特徴が現れにくい。そこで、本書ではCSJとBCCWJを同じ重みで扱い、出現頻度を算出した。

(2) 長単位を利用したこと。

CSJおよびBCCWJの両コーパスに施されている2つの言語単位のうちの1つである「長単位」を利用したことで、日本語教育では通常1語として教えられる「自動車」「飛行機」などの複合語やひとまとまりの文型として教えられる「について」「てしまう」「かもしれない」などの複合辞を盛り込むことができた。

(3) 特徴的なレジスターの表示

使用したコーパスのデータをいくつかのグループに分け、書籍(BK)、ウェブ(WB)、公的媒体(OF)、新聞・雑誌(NM)、話し言葉(SP)の5つのレジスターを設けた。各レジスターにおいて対数尤度比(log-likelihood)の値の上位50語(SPのみ47語)についてそのレジスターで特徴的に多く使われることを示した。

5. 本書の問題点

最後に本書を利用するにあたって注意すべき点を述べる。本書で利用したCSJとBCCWJは両者とも短単位、長単位という同じ言語単位で解析されているが、厳密には細かい違いがあり、語によっては見出し語のまとめ方が一致しない場合がある。例えば、順位45位の副助詞「とか」はBCCWJでは長単位で2語であるが、CSJでは1語である。従って本書に表示されているのはCSJのみにおける出現頻度に基づく情報である。もし、BCCWJでも「とか」を1語としていれば、順位はもっと上に来たかもしれない。

本書の5000語のリストを確定するにあたって、元となる作業リストに操作的に手を加えた部分とそうしなかった部分とがある。手を加えた部分の例は、複数の見出し語をまとめたことである。例えば、サ変動詞の語幹となる名詞とサ変動詞、名詞とそれに接頭辞「お」「ご」が付いた語などである。図2の「納得(する)」は「納得」と「納得する」を1つの見出し語としてまとめたことを意味する。

手を加えなかった部分の例としては、助数詞を構成要素に持つ語の扱いがある。長単位を採用したことで、「一回」「二本」「三人」などのような助数詞を含んだ語がかなり5000位内に入ってきた。これらをいちいち採り上げるのは学習的には意味がないという考えもあるだろうが、どれを残してどれを削除するかが恣意的になるため、すべて残すことにした。

言語データの評価は利用場面を通じて定まっていくものである。本書で提示した語彙は、今後、日本語教育で使われる既存の語彙リストとの比較や実際の学習・教育場面を通じてその価値が検証されていくことであろう。

なお、本書の書評として金（2013）がある。併せて参照いただければ幸いである。

●参照文献●

- Carroll, John B. (1970) An alternative to Juillard's usage coefficient for lexical frequencies, and a proposal for a Standard Frequency Index (SFI), *Computer Studies in the Humanities and Verbal Behavior* 3(2): 61-65.
- 金愛蘭(2013) [書評]「Yukio Tono, Makoto Yamazaki and Kikuo Maekawa (2013) A frequency dictionary of Japanese: Core vocabulary for learners」『計量国語学』29(2): 66-70.

山崎 誠 (やまざき・まこと)

国立国語研究所言語資源研究系准教授。文学修士。国立国語研究所研究員，同室長，同領域長等を経て，2009年10月より現職。

主な著書・論文：『複合辞研究の現在』（共編著，和泉書院，2006），「代表性を有する現代日本語書籍コーパスの構築」（『人工知能学会誌』24(5)，2009），『言語研究のための統計入門』（共著，くろしお出版，2010），*A frequency dictionary of Japanese*（共編著，Routledge，2013），「語彙調査の系譜とコーパス」（『講座日本語コーパス1 コーパス入門』，朝倉書店，2013）。

社会活動：計量国語学会理事，言語処理学会理事。