

語彙資源活用シンポジウム@国立国語研究所 2017年3月6日

中型国語辞典『大辞林』編集と見出し語の収集・選定について—未知語・新語を中心に

山本 康一（三省堂 辞書出版部）

概要

1. 『大辞林』 について
2. 辞書と語彙資源
3. 語彙資源の活用例
4. 辞書の現在/未来と語彙資源

1. 大辞林



中型国語辞典 2,976pp.

古語から現代語までを対象

❖初版 1988年刊 22万語

❖2版 1995年刊 23万3,000語

❖3版 2006年刊 23万8,000語

継続的なデータ更新



- ❖ 書籍3版刊行後も継続的にデータ更新 → 現時点で26万語強
- ❖ 新語を中心とした増補
「合区」「マイナンバー」「悟り世代」
「バズる」「秘密保護法」「タイムラプス」
「マタハラ」「自撮り」「りょ」等
- ❖ 電子媒体・サービスでの提供

多様な電子媒体で利用



❖ スマホ・タブレット アプリ

❖ 電子辞書



❖ iOS, MacOS 内蔵辞書



❖ ウェブサービス

三省堂 *Web Dictionary*

Dual 大辞林

- ❖ 3版刊行時（2006年10月～） サービス開始
- ❖ 書籍に付随したウェブサービス（書籍未収録の更新追加語も提供）
- ❖ 適時、新語・追加語増補提供（書籍を補完）



追加語には採録時タイムスタンプを表示

この項目は書籍には収録されていません

きんえん

【書籍未採用】

きんえん ⑩ 【禁演】

上演を禁じること。

項目採録・修正日付

2012年11月

この項目は書籍には収録されていません

クリぼっち

【書籍未採用】

クリぼっち ③

俗に、クリスマスをひとりぼっちで過ごすこと。

項目採録・修正日付

2013年02月

2. 辞書と語彙資源

辞書編集の基礎材料としての「使用例」

現実が発信され、流通する言語事象を集積

「用例カード」（＝使用例の集積ツール）の例

- ① 見坊カード
- ② 大辞林用例カード
- ③ 社内での収集システム
- ④ 個人の手控え

①見坊カード

見坊豪紀先生収集の約145
万枚のカード

『三省堂国語辞典』編集
のための資料



②大辞林 用例カード

『大辞林』の編集用

出典資料からの抜き書き
等

現在は、テキストデータ
等電子データに移行



③社内データベース

閲覧モード

日付順50音順検索

閲覧／検索編集

触地図/しょくちず

しょくちず [触地図] 17.03.03
ショールーミング 17.03.03
コーディネーショントレーニング 17.03.03
ドンコ 17.03.03
わしょく にほんのでんとうてきなしょくぶんか
[和食 日本の伝統的な食文化] 17.03.03
ひょうごこうどうわくぐみ [兵庫行動枠組]
17.03.03
エイチエフエー [HFA] 17.03.03
カリスマどくしゃ [カリスマ読者] 17.03.03
チャイルドパートナー 17.03.03
シーピー [CP] 17.03.03
ネウボラ 17.03.03
たにんごと [他人ごと] 17.03.03
エンゲージメント 17.03.03
あいみみ 17.03.03
ふわっと 17.03.03

触地図/しょくちず

樹脂の板に、凹凸で描き出された道路。目の不自由な人が使う「触地図」だ。国土地理院に断念したこの地図の開発計画を昨秋、復活させて実用化への道を開いた。

朝日新聞2015/2/13（山本）

注記：

- ジャンル：未整理
- 種別：
- 作業者：
- 採取日：17年03月03日

編集支援システム検索（大辞林・全訳読解古語他）

データベースセンター

三省堂社内での収集システム

新聞・雑誌等指定資料から、作業者が手作業でピックアップし、集積したもの、
および、社内編集者の個人収集等を集約

新聞・雑誌他メディアからの記録・収集

閲覧／検索

編集

衣液/ころもえき (読み方不明)

サクッと揚がる衣液を作るために卵と水をどう混ぜるのがオススメ

NHKアサイチ2015/2/10 (山本)

注記:

- ジ
- 種
- 作
- 採

閲覧／検索

編集

氷筍/ひょうじゅん

北海道登別市カルルス町の鉄鉱石鉱山跡で、下から上に伸びたつらら「氷筍」が見ごろを迎えている。

朝日新聞2015/2/10 (山本)

注記:

- ジ
- 種
- 作
- 採

閲覧／検索

編集

カリスマ読者/カリスマどくしゃ

ファッション誌で活躍する後輩の編集者との会話。「……その人ってカリスマ読者っぽくない？」と私。「先輩、“カリスマ読者”って何年かぶりに聞きましたよ」と、冷笑しながら答える後輩。しまった。つい、賞味期限切れの言葉を使ってしまった。恥ずかしい。

朝日新聞2015/2/12 (山本)

注記:

- ジャンル: 服飾・宝飾
- 種別:

④個人のメモ

日常的な記録・収集。社内システムに集約。

採取語	文脈	日時	採取対象	分類
沈下計	地盤沈下の状況は、地表面から鉄管の底までの間の地層の収縮量を <u>沈下計</u> で計測することによって観測しています。	2015/9/21	調布市野川沿いの看板	看板
アドプト	狛江市 <u>アドプト</u> ～いつも美しいまち狛江～	2015/12/15	狛江市歩道脇看板	看板
ロック板	フラップ板（ <u>ロック板</u> ）が下がっていることを確認の上、ゆっくり入庫して下さい。	2014/1/30	調布市病院駐車場	看板
シャルキュトリー	自家製 <u>シャルキュトリー</u> 、内臓料理、自然派ワイン	2014/1/25	文京区レストラン看板	看板
エコキャップ	お客様用 <u>エコキャップ</u> 回収箱	2014/3/28	武蔵中原駅構内看板	看板
朝挽き	<u>朝挽き</u> もつ焼き	2013/12/1	千代田区居酒屋看板	看板

街の看板、車内広告、テレビ、新聞その他媒体等から

採取語	文脈	日時	採取対象	分類
しゅわっと	<u>しゅわっと</u> 炭酸 熱中症対策	2016/7/26	JR中央総武線 車内広告	広告
公用語	ニュースによると、英語を <u>公用語</u> にする会社が増えている、らしい	2015/1/9	JR中央総武線 車内広告	広告
チャージする	伊勢志摩で <u>チャージ</u> しよう	2014/2/22	京王線車内広告	広告
催物場	レディスブラックフォーマルバーゲン 伊勢丹新宿店本館6階＝ <u>催物場</u>	2014/2/2	京王線車内広告	広告
便房	手動車いすのお客様がご利用できる <u>便房</u> を設置するなどバリアフリー機能の充実	2014/2/13	京王線車内広告	広告
ながら歩き	スマホがおもしろくなるほど、残念ながら、 <u>ながら歩き</u> は危険になる。	2013/12/9	京王線車内広告	広告
うる落ち	冬のメイク落としは化粧水で <u>うる落ち</u> 。	2013/11/29	JR山手線車内 広告	広告

街の看板、車内広告、テレビ、新聞その他媒体等から

採取語	文脈	日時	採取対象	分類
ワッチ	エンジニアとして1日8時間、4時間ずつの2交代でエンジンルームの <u>ワッチ</u> （当直）に立たなければなりません。	2016/10/16	朝日新聞	新聞
バウンサー	（長男が） <u>バウンサー</u> （ゆりかご）から落ちてたんこぶができた」	2016/10/16	朝日新聞	新聞
政冷経熱	政治や外交で対立があっても、民間では交流することを「 <u>政冷経熱</u> 」という。	2016/10/18	朝日新聞	新聞
神る	内川のフォーム吸収「 <u>神った</u> 」鈴木	2016/10/18	朝日新聞	新聞
メガ盛り	また大食いの女性たちが、老舗が特別に用意した“ <u>メガ盛りプレート</u> ”に挑む。	2016/10/18	朝日新聞	新聞
産み落とし	医師や助産師の支援を受けずに出産し、放置するなどして赤ちゃんが死亡する「 <u>産み落とし</u> 」が後を絶たない	2016/10/19	朝日新聞	新聞

“目的をもった収集”

- 辞書の用例収集の特徴は「目的をもった収集」である点
- 目的＝「辞書の見出し語となるもの」
 - － そういう意味では、言わば“アンバランスなコーパス”

- 「辞書の見出し語となるもの」とは、
どういうものか？ ➡ 「編集方針」に
よって規定される

見出しの選定（編集方針）

—— 『大辞林』 第3版 「凡例」

「現代の言語生活に立脚し、現代語を中心に
古語や百科語をも含めた総合的な国語辞典と
して編集」

見出しの選定（編集方針）

「現代語・古語については、一般国語語彙の
主要なものを網羅的に採録する」

（「大辞林 執筆要領 内規」から抜粋 以下同）

「外来語については、現代の日常生活にひ
ろく使用される一般語化したものを採録す
るとともに、社会の進展にともなって新た
に国語として定着するようになった新しい
外来語もできるだけ採用するようにする。」

「流行語・俗語・方言・特殊語（隠語・忌詞など）については、現代の日常生活に比較的ひろく使用されると認められるものを採録する。」

「専門語については、自然科学・社会科学・人文科学、その他の諸部門の用語から、現代の日常生活にひろく使用されるもの、現代人の教養上必要と思われるものを精選して採録し、さらに現在の学界および専門領域で通用する最新のものを採録するようにつとめる。」

「固有名詞については、現代の日常生活の上で、あるいは主要な古典の読解のため、必要と思われる範囲で、とくに重要な人名・地名、その他のものを採録する。〔注〕固有名詞の採録にあたっては、（中略）その固有名詞が、古典の読解上必要かどうか、われわれの生活にどれだけ密接な関係をもつか、あるいは、現代人の教養としてぜひ必要かどうか、などの点を考慮して決定するようにする。」

① 上記に共通する要件として「現代の日常生活に広く使用されるもの」

—— しかし、どこまでが「日常生活」か？
その範囲は？

② また、「語釈が付けられるもの」という要件もある

—— 語義内容が不分明なもの、すなわち語義が書けないものは、そもそも立項不能

3. 語彙資源の活用例

収集した「用例」は、以下2つの観点で活用

① 編集方針に則った新語・追加語の採録

② 新語義・新用法の確認と追加

既存の語に生じたと思われる新しい語義・用法

➡例（「とがる（とんがる）」 「熱量」）

とがる（とんがる）

「音楽やアート映画のファンに特別な存在だった「六本木WAVE」はすでに閉まっていたし、おめあてだった渋谷・スペイン坂のとがったセレクトショップも、姿を消していました。」

（朝日新聞 2015年3月11日 「リレーオピニオン TOKYO風景①」
作家 山内マリコさん）

「人と違う価値観を持って、自分らしい生き方をしな
さいと両親に育てられた私は、30歳を過ぎる頃まで
尖っていたと思います。」

(朝日新聞 2014年3月9日「朝日求人 仕事力」平田オリザさ
ん)

「「尖った人材の発掘」も人物重視の入試の重要なポ
イントだが、高校生からは、今の推薦入試で十分では、
との意見が聞かれた。」

(朝日新聞 2013年12月12日「読み解き経済」)

BCCWJ ではこの用法が3件認められる。

28	万円にも達しないことを考えると、ちょっと買いづらいなと感じた。走りもルックスも	尖った	存在へと生まれ変わったカルディナ。来年登場する次期レガシィとの対決が楽しみだ。▲	岡崎 五朗(著)	1960	男	雑誌/ 工業/ 機械	CAR and DRIVER		2002年11月 10日号（第 25巻第21 号、通巻第 517号）	ダイヤ モンド 社	2002
34	図がはっきりとし、その後のアシッド・ジャズといいブラック・ロックといい、今もその	尖った	物質性と振動力においてこの作品の右に出るものは一つとしてないのである。ここには確	平井 玄(著)	1950	男	書籍/7 芸術・ 美術	音の力		DeMusik Inter. 編	イン パク ト出 版	1996
37	ルティカーとして評価すれば最高に使い勝手のいいクルマである。■ルックスも走りも	尖った	”存在に 最初に試乗したのは4WD + 2ターボで武装したGT-FOUR・Nエディ	岡崎 五朗(著)	1960	男	雑誌/ 工業/ 機械	CAR and DRIVER		2002年11月 10日号（第 25巻第21 号、通巻第 517号）	ダイヤ モンド 社	2002

これら新用法の実例をもとに、新語義が立項される

とがる

とが・る ② 【尖る】

(動ラ五〔四〕)

- ① 先が細く鋭くなる。とんがる。「—・った鉛筆」「口が—・る」
- ② 感じやすくなる。過敏(かびん)になる。「神経が—・る」
- ③ 声や表情が怒りで強く鋭くなる。とんがる。「声が—・る」

類語情報 突起

この項目は書籍には収録されていません

ナビゲーション

大 小 簡

とがる

【書籍未採用】

とがる ② 【尖る】

協調せず突出する。他を顧みず挑戦的である。とんがる。「—・った人材」
「若い頃は—・っていた」

熱量

「地下アイドルもオタクの僕らも熱量を持って生きて
ます」

（朝日新聞 2016年1月6日 「東京を生きる⑤ アイドルとオレ 情
熱100%」 ）

「電車ならではの素材を用いて電車を描く発想に、駅職員の地道さが加わり、電車のアートとしてとんでもない熱量を放ってくる秀作。」

（ねとらぼ 2015年11月11日「切符のパンチくずで作った電車のドット絵が圧巻 使った数は15万枚以上、製作職員「もう二度としない」 <http://nlab.itmedia.co.jp/nl/articles/1511/11/news151.html>）

「それが作者としての熱量の変化なのか、表現者としてそうなのかはわからない。」

（朝日新聞 2016年4月3日「文豪の朗読 いとうせいこうが聴く 谷崎潤一郎「春琴抄」」）

「ファンの大きな熱量が仕事への活力をくれた」
(朝日新聞 2016年7月29日 「情熱人」)

「情熱の量」という意識か？

新語義はまだ未立項。

ねつりょう

大 小 簡 ↑

ねつ りょう — りやう ② 【熱量】

二つの異なった物体の間を熱として移動するエネルギーの量。SI単位はジュール（記号 J）。ほかにカロリーも用いられる。

子項目 熱量計

BCCWJ 「熱量」210件中には上記に類する用法を見つけられず。➡かなり近年の用法か？

辞書における語彙資源活用

- 使用例（「用例」）の集積としての言語資料体も一種のコーパス
- このコーパスによって、新語や新語義・新用法の発生時期や定着度合を確認
- さらに有用なコーパスの利用が辞書内容の改善に資すると期待される

「辞書編集におけるコーパス活用の一般的 利点」

—— 井上2016 p.22

「(1) 辞書編集におけるコーパス活用の一般的利点

- a. 先行研究にない言語事実についても分析記述が可能となる。
- b. 少人数のインフォーマント調査では避けることのできないインフォーマント自身の個人差を吸収できる。
- c. 母語話者の無意識の言語使用を分析の対象とすることが可能。
- d. 母語話者にはない視点での分析を可能にしたり、日本人英語学習者ならではの疑問点に答える文法・語法記述が可能となる。」
- e. 一定の方針のもとに構築されたコーパスを使うことで、執筆者が水準を規定された共通の言語データに基づいて記述することが可能となる。
- f. コーパスを用いることで非文かどうかは判断できないが、一般的な表現であるかどうかはある程度判断できる。
- g. ターゲットユーザー別にコーパスを構成することで、ターゲットユーザーに最適な辞書編集環境を整えることができる。
- h. コーパスとともに提供される表示方法や統計値に関わる各種ツールを活用して、分析に必要な情報を効率的に得ることができる。」

※傍線は引用者。

コーパスを活用した辞書編集において 先行する英和辞典

『ウィズダム英和辞典』

(2003年、2版 2007年、3版 2013年)

- 独自に開発・構築し、日本人の英語学習に特化した
コンピューターコーパス「三省堂コーパス」
- 日本人英語学習者が必要とする英語に配慮した、バ
ランスのとれたコーパス
- 1億語（メインコーパス6千万語＋サブコーパス4千
万語）

4. 辞書の現在/未来と語彙資源

辞書の現況

- 推定販売部数（全国出版協会・出版科学研究所「出版指標年報」）

約1500万冊（1993年）

➡約600万冊（2012年）

- 利用媒体の多様化

電子辞書／アプリ（スマホ・タブレット）／ウェブ／内蔵辞書 他

書籍（紙）版と電子版

大辞林：23万8千語の書籍版と、26万強のデータ版が併行して利用されている

2つの方向

- － **書籍版**：より定着度の高いと見込まれる精選された語を採録
- － **データ版**：より即時性を高めた、新語・未知語の採録を追求

「近い将来、電子化された辞書は、これまでに
はない利用法や情報を搭載したものが現れるだ
ろう」（佐竹 2011 p. 251）

それぞれのメリット・デメリット（山本 2013 p. 58-59)

- ① スペースの制約からの解放（電子版）↔ スペースの制約があることによる情報の圧縮・精選（書籍）
- ② 検索による当該項目へのピンポイントのアクセス（電子版）↔ 紙面をめくることによる一覧性と周辺の語への見渡し（書籍）
- ③ フレキシブルな表示形態の変更（電子版）↔ 均整のとれた端正な紙面（書籍）
- ④ 大量の付属情報による拡張可能（電子版）↔ 冊子体であることによる総体の視覚的把握が可能（書籍）

❖現状は、まだ書籍の紙面を前提にした作り

- 近年は、「表示レイアウト」と「意味・構造」の分離を行った XMLデータ編集に収斂しつつあるが、改訂ものが多いこともあり、元の設計のしばりはある

❖書籍を前提にしない辞書とはどういうものか？

- 実績として、AIでの活用 —— 「東ロボ」プロジェクト（国立情報学研究所「ロボットは東大に入れるか」プロジェクト）への辞書データ提供はあり
- それ以外は？

“西暦3000年の辞書編集”

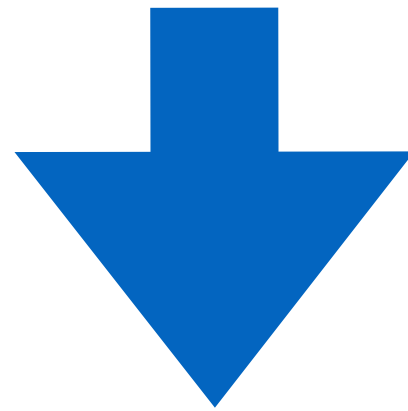
Grefenstette (2008)

「「西暦3000年に辞書編集者は存在しているだろうか？」

——コンピューター言語学が進展することで、辞書情報の分析と生成の自動化が進み、またエンドユーザー自身が必要とする語の用法や適切な用例をコーパスから取り出すことができるようになれば、辞書編集者は不要になる」

「辞書編集者の役割は、情報を選択し総合することから、ソフトウェアによって選び出されたものを「編集」し確認することにすでに変わりつつある」 (Rundell 2012 p.17)

「近年の多くの辞書プロジェクトは電子媒体がもたらしたイノベーションが、辞書の設計と利用のあらゆる側面を根本的に変えうることを証明している」 (Granger 2012 p. 1-5)



6つのイノベーション

(1) コーパスとの統合

辞書編集のバックグラウンドで使われるのみならず、電子版辞書に統合されて、ユーザーが直接アクセスし、自身のための情報を取り出すもの

(2) 良質な多量のデータ

コロケーションや用例の飛躍的な増量、マルチメディアコンテンツ（画像・図表・動画・音声等）、拡張情報（語法・文化情報・誤用情報等）の一体化

(3) アクセス効率

特定の情報にたどり着く多様な検索方法の実現。アクセシビリティ全般の改善。

(4) カスタマイズ

ユーザー各人のニーズに合わせた辞書の動的なカスタマイズ

(5) ハイブリッド

辞事典、専門用語集、語彙データベース、単語帳、作文の手引き、翻訳ツールなど、異種の言語リソースとの統合。

辞書兼文法書、辞書兼シソーラス、辞書兼語法書、モノリンガル辞書兼バイリンガル辞書など

(6) ユーザーによる入力

ユーザー自身が入力する集合的辞書編集（言葉の変化と新しい語彙を常に押さえていける点が強みであるが、反面、正確性が保証されず、これらユーザーの手になる辞書は科学的価値がない、という否定的な見方もある）

まとめ

- 辞書編集の基盤には「語の使用例（用例）」があり、新規追加語の選定のみならず新語義・新用法の観察・分析にも必須
- 語彙資源の集積および ICT / AI の進歩は、以下の両面に変化をもたらす可能性がある
 - 辞書の編集制作
 - 辞書の提供形態

問い

- この変化とは、具体的にはどういうものとして現れるのか？
- 語義や用法の変異をロボットは理解できるのか？

- 辞書編集者はいつまで存在できるのかか？

引用・参考文献

Granger, S. (2012). 'Introduction: Electronic lexicography — from challenge to opportunity', in S. Granger and M. Paquot. (ed.) *Electronic Lexicography*, 1-11. Oxford: Oxford University Press.

Grefenstette, G. (2008). 'The Future of Linguistics and Lexicographers: Will there be Lexicographers in the Year 3000?', in T. Fontenelle (ed.) *Practical Lexicography; A Reader*, 307-323. Oxford: Oxford University Press.

Rundell, M. (2012). 'The road to automated lexicography: An editor's viewpoint', in S. Granger and M. Paquot. (ed.) *Electronic Lexicography*, 15-30. Oxford: Oxford University Press.

井上永幸（2016）「編者が語る英語辞書編集 ウィズダム英和辞典」（南出康世・赤須薫・井上永幸・投野由紀夫・山田茂編『英語辞書をつくる—編集・調査・研究の現場から』大修館書店）

佐竹秀雄（2011）「国語辞典と語彙」（斎藤倫明・石井正彦編『これからの語彙論』ひつじ書房）

松村明・三省堂編修所編（2006）『大辞林』第3版 三省堂

山本康一（2013）「電子辞書の意義と今後」（『日本語学』32(2): pp.48-61. 明治書院）

語彙資源活用シンポジウム@国立国語研究所 2017年3月6日

中型国語辞典『大辞林』編集と見出し語の収集・選定について—未知語・新語を中心に

山本 康一（三省堂 辞書出版部）