

辞書見出しの 語彙資源化

国立国語研究所

高田智和， 小野春菜

国語研ユニット「古辞書類に基づく語彙資源の拡張と語彙・表記の史的変遷」

- 『和名類聚抄』『日葡辞書』『易林本節用集』『言海』など、古辞書類の掲出語のデータを作成
- 国立国語研究所の語彙資源
 - 分類語彙表、現代雑誌90種調査語彙表、日本語教育基本語彙 など
 - 形態素解析辞書Unidic、各種コーパスの語彙表 など
 - 日本言語地図、沖縄語辞典 など
 - 歴史的な漢語が得られにくい

語彙資源同士をつなぐ

- つなぐための「糊」 形態素解析辞書Unidicの語彙素ID
- コーパスの形態素解析結果は語彙素IDで横断することができる。
- 各種語彙資源に語彙素IDを付与することで、語彙資源とコーパス、語彙資源同士をつなぐことができるようになる。
- 国立国語研究所（2020）『分類語彙表番号-Unidic語彙素番号対応表』（wlsp2unidic ver.1.0.2）
- 語誌情報ポータル 本プロジェクトで構築中

『日本国語大辞典第二版』（ジャパンナレッジ）のURL

<https://japanknowledge.com/psnl/display/?lid=200201b09c79VpP4ELkv>

『日本国語大辞典第二版』見出し語のID「糊」として利用できる。



The screenshot shows a web browser displaying the URL <https://japanknowledge.com/psnl/display/?lid=200201b09c79VpP4ELkv> in the address bar. The page title is "JapanKnowledge Personal" and the main heading is "日本国語大辞典". The entry is for "さい-せき 【載籍】". The entry includes a section for "解説・用例" and a list of examples:

- 〔名〕
- 書物に書きのせること。また、書きのせてあるもの。書物。典籍。書籍。
- * 懐風藻〔751〕序「逖聴=前修-、遐観=載籍-」
- * 仮名草子・童蒙先習〔1612〕跋「夫惟、載籍広大闊博、而童幼初学未^レ易^レ得=其要-矣」
- * 異人恐怖伝〔1850〕下「血書のこと故事あるべし、予寡聞なるが上に、載籍にさへ乏しく」
- * 開化本論〔1879〕〈吉岡徳明〉下・六「近年古学勃興以来円珠庵居翁以下八九家の載籍を参考して其正を採るべし」
- * 史記-伯夷伝「夫学者載籍極博、猶考=信於六芸-」

本日の国語研ユニットからの報告

- 辞書見出し（語）の語彙資源化
- 大槻文彦『言海』 このあと、小野春菜より報告
- 『日葡辞書』 第2部で大島英之より報告

大槻文彦『言海』

- 近代国語辞書の先駆け。明治24年（1891）出版完了
- 「本書編纂ノ大意」第二条 辞書における五大要件

発音

語別
= 品詞

語原
= 語源

語釈

出典

以上、五種ノ解アリテ、
始メテ辞書ノ體ヲ成ス
トイフベシ。

⇒ 現代にも通じる構成

『言海』のWEBサービス

真理子日曜学校「WEB言海」（見出し語検索・本文PDF）

<https://www.babelbible.net/genkai/genkai.cgi>

国立国会図書館デジタルコレクション「言海：日本辞書 第1-4冊」（全文検索）

<https://dl.ndl.go.jp/pid/992954/>

▶全文検索の問題点：OCRによる変体仮名の処理

（例）ことのほかぜ／ことのほかぎ → さとのほかぜ／まとのほかぎ

▶いずれもWEB上のサービス

→ **オフラインで使用できる見出し語データの作成**

「言海見出し語データ」の概要

▶収録予定の項目

見出し番号/見出し語/見出し語（漢字表記）/通用字の別/品詞/位相注記/漢ノ通用字/
ページ・段/底本画像/日国見出しID

見出し番号	見出し語	見出し語 (漢字表記)	通用字の別	品詞	位相注記	漢ノ通用字	ページ・段	底本画像	日国見出しID
00001	あ						0001a	https://dl.ndl.go.jp/pid/992954/1/70	200200000000Bt7Xi7n6
00002	あ	吾	和漢	代	古		0001a	https://dl.ndl.go.jp/pid/992954/1/70	20020000025cb8V0V4sJ
00003	あ	彼	和漢	代	古		0001a	https://dl.ndl.go.jp/pid/992954/1/70	2002000004214656w5wa
00004	あ			感		唯	0001a	https://dl.ndl.go.jp/pid/992954/1/70	2002000009707LWs1rM4
00005	あ			感			0001a	https://dl.ndl.go.jp/pid/992954/1/70	2002000009707LWs1rM4
00006	ああ	嗚呼	和漢	感			0001a	https://dl.ndl.go.jp/pid/992954/1/70	200200000b357RKK1qEm
00007	あい			名	訛		0001a	https://dl.ndl.go.jp/pid/992954/1/70	20020000660b8Af6JJ5C
00008	あい	愛	和漢	名			0001a	https://dl.ndl.go.jp/pid/992954/1/70	200200005e60u495SoQc

(表) 言海見出し語データ

「言海見出し語データ」の概要 ①

➤ 見出し語に対する任意の番号

見出し番号

見出し番号
00001
00002
00003

見出し番号

00001

見出し語

あ 五十音圖、阿行第一ノ假名。此ノ二行ノ假名、あ、い、う、え、おノ五音ハ、喉ヨリ單ニ出テ、又發聲ノ韻トモナリテ、熟音ヲ成サシム、故ニ單音、又母韻ノ稱アリ。あハ、下ニ、う、或ハ、あ、うニ轉ジテ呼ブモノヲ受クルトキハ、おノ如ク呼ブトアリ、あうむ(四音)あうま(五音)如シ。

あ(代) 吾 われニ同ジ。「ガ妻」「ガ君」「ガ兄」

あ(代) 彼 あれニ同ジ。「ハト見ル月」

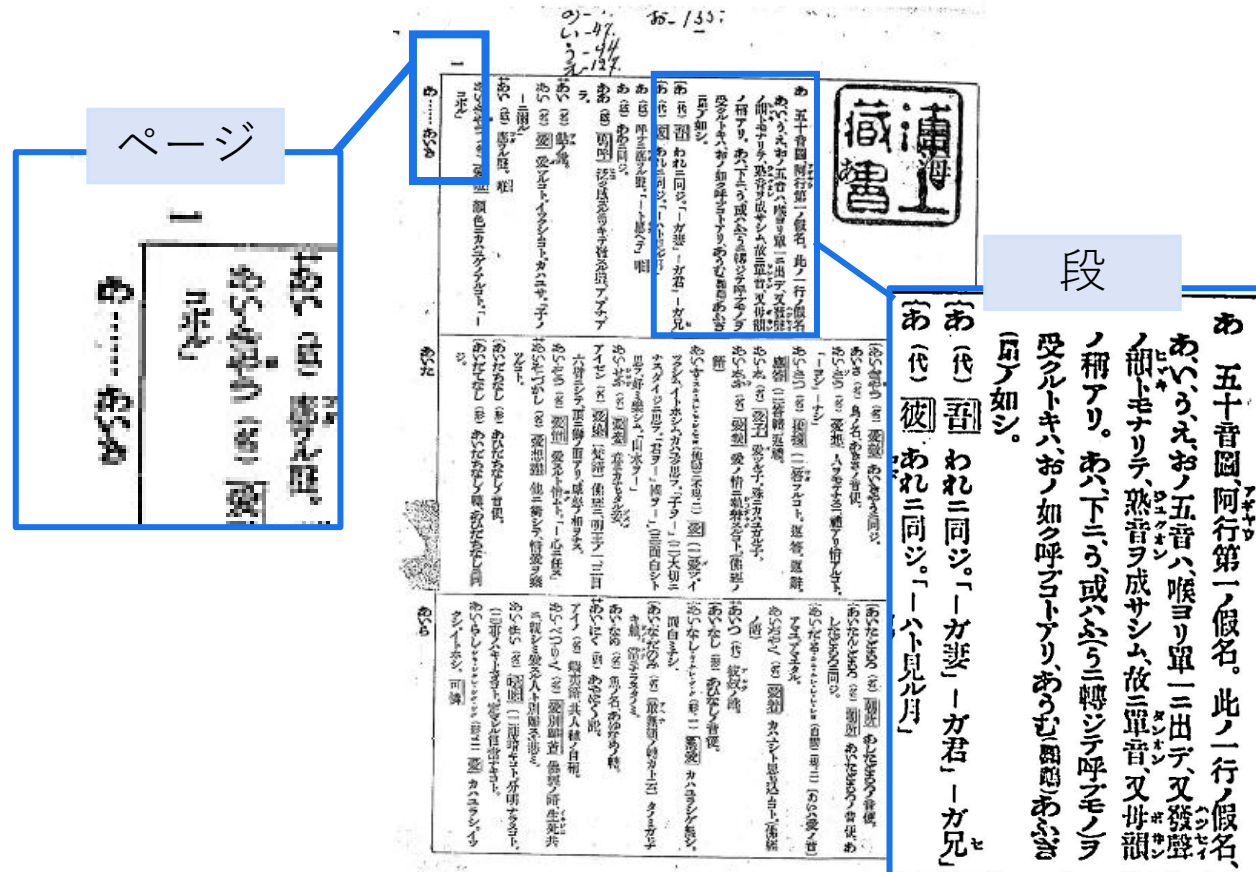
「言海見出し語データ」の概要 ③

『言海』辞書本文のページ・段

ページ・段

ページ・段
0001a
0001a
0001a

0001a = 本文 1 ページ目 1 段目



ページ

段

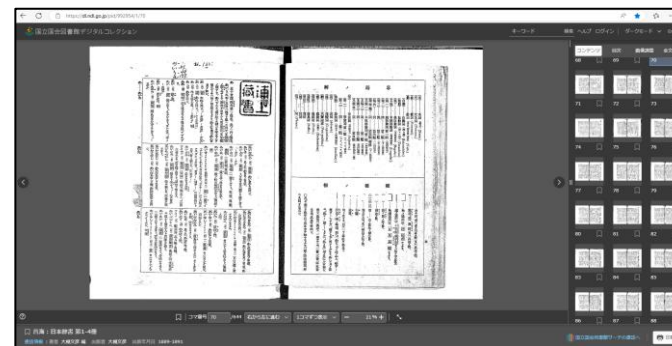
「言海見出し語データ」の概要 ④

外部データの情報

底本画像/日国見出しID

底本画像	日国見出しID
https://dl.ndl.go.jp/pid/992954/1/70	200200000000Bt7Xi7n6
https://dl.ndl.go.jp/pid/992954/1/70	20020000025cb8V0V4sJ
https://dl.ndl.go.jp/pid/992954/1/70	2002000004214656w5wa

底本画像
[国立国会図書館デジタル
コレクション]

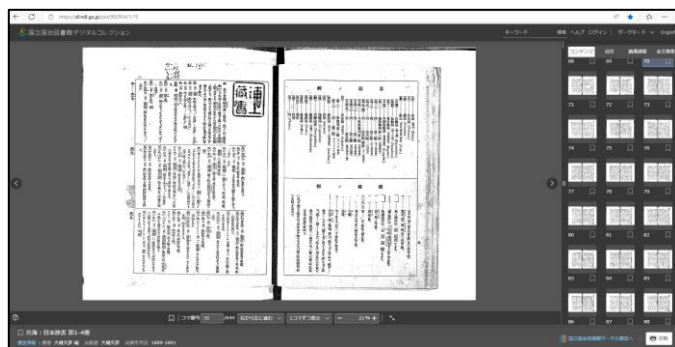


日国見出しID
[ジャパナレッジ版
『日本国語大辞典』]



(図) 国立国会図書館デジタルコレクション、ジャパナレッジ

外部データの情報：底本画像



国立国会図書館デジタルコレクションとの対応

- 国立国会図書館所蔵の初版四冊本『言海』の URL 付与 (<https://dl.ndl.go.jp/pid/992954>)
- 見出し語の検索後、画像リンクから **語釈** の確認が可能となる

外部データの情報：日国見出しID



ジャパンレッジ版『日本国語大辞典』との対応

- ジャパンレッジ版『日本国語大辞典』（『日本国語大辞典』第二版）のID付与
- 『言海』との品詞認定の違いが明らかとなる

（例）サ変可能動詞の扱い

言海「くつつたくす（屈託）」（自動・不規・二）

－日国「くつつたく（屈託）」（名）

「言海見出し語データ」の公開予定

- ◆ 公開予定 2025年4月以降
- ◆ 公開先 国立国語研究所学術情報リポジトリ
(<https://repository.ninjal.ac.jp/>)
- ◆ ライセンス CC BY 4.0（改変・二次利用可）



「言海見出し語データ」による基礎計量

1. 『言海』見出し語数
2. 『日本国語大辞典』見出し語との対応

「言海見出し語データ」による基礎計量

1. 『言海』見出し語数

(表) 先行研究における見出し語の計測

「言海採収語…類別表」	39103
犬飼守薫 (1999)	39335
田鍋桂子 (2000)	39414
岡島昭浩 (2003)	39364
豊島正之編 (2016—2024) *湯浅茂雄 (2021)	39443

➤ 本プロジェクトによる見出し語データの総数 **39,408語**

「言海見出し語データ」による基礎計量

2. 『日本国語大辞典』見出し語との対応

日国見出しIDの付与 39147 / 39408 (**99.34%**)

日国見出しIDの非付与 261 / 39408 (0.66%)

➤ 『言海』と『日本国語大辞典』第二版の見出し語認定の相違

(日国未記載の例)

石の名前：うすいし（白石）、きんがせき（金牙石）、くだいし（管石）

時刻：はちじ（八時）、くじ（九時）、じふいちじ（十一時）

「言海見出し語データ」による基礎計量

2. 『日本国語大辞典』見出し語との対応

日国見出しIDの付与 39147 / 39408 (99.34%)

ジャパナレッジ版『日本国語大辞典』全文検索（「言海」＋「(not) 大言海」）
33650 / 502607 (6.69%)

➤ 『言海』見出し語 2 以上と『日国』見出し語 1 の対応

(例) 言海ID 08639 (かる〈涸〉), 08640 (かる〈嗶〉), 08641 (かる〈枯〉)

日国ID 200200f246a93k56GPHy (か・れる 【涸・枯・嗶】)

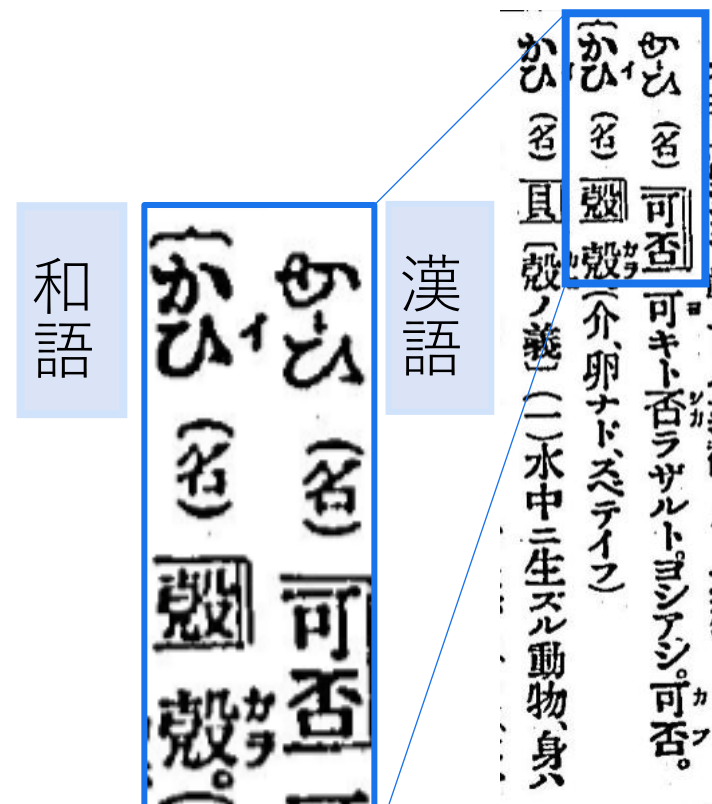
今後の改良予定

1 解析用辞書UniDicの語彙素との対応

- 『言海』収録語を短単位分析に活用できる

2 書体差（語種）の区別

- 平仮名活字を二種類使用（和語・漢語の区別）
- 目視での確認が必要
- 「言海採収語…類別表」語種別の語数の再検討



今後の改良予定

3 国書データベースとの対応

- 慶応義塾大学附属研究所斯道文庫蔵「言海校正刷」のURL付与

(<https://kokusho.nijl.ac.jp/biblio/100338733>)

- 見出し語の検索後、本文画像から出版までの修正過程が確認可能となる



(図) [国書データベース](#)

参考文献

- 犬飼守薫（1999）『近代国語辞書編纂史の基礎的研究』風間書店
- 岡島昭浩（2003）「『言海採収語…類別表』再読」『国語語彙史の研究』22
- 小野春菜（2018a）「編纂資料からみた私版『言海』の成立」2017年度清泉女子大学大学院博士論文
- 小野春菜（2018b）「序章 『言海』研究史」今野真二・小野春菜『言海の研究』武蔵野書院
- 今野真二（2018）「言海の研究：はじめに」今野真二・小野春菜『言海の研究』武蔵野書院
- 銭谷真人（2012）「『言海』における仮名字体および仮名文字遣い」『日本語学研究と資料』35
- 豊島正之編（2016—2024）「Japanese pre-modern dictionaries 日本近代辞書・字書集」*公開停止
- 田鍋桂子（2000）「『日本辞書／言海』の語種－外来語を中心に－」『日本語論叢』1
- 内藤千尋（2019）「和・和漢・漢の通用字」『上智大学国文学論集』52
- 湯浅茂雄（2021）「『言海』『日本大辞書』の収録語数をめぐって」『実践國文學』100