

『日本語歴史コーパス 明治・大正編Ⅲ明治初期口語資料』（短単位データ 0.9）概説書

2021年3月31日 近藤明日子

2022年3月31日 近藤明日子・改訂

1. はじめに

『日本語歴史コーパス 明治・大正編Ⅲ明治初期口語資料』（短単位データ 0.9）は、明治初期に刊行された主要な口語体資料をコーパス化したものである。

この文書では、コーパスの収録資料とテキストの仕様、形態論情報、コーパス検索アプリケーション「中納言」の検索結果に表示されるテキストおよびアノテーション（テキストに付与する付加情報）の項目について、その概要を示す。

なお、本コーパスのテキストやアノテーションは、原資料である底本の状態に即して作成するように努めたが、誤りが残されている可能性がある。そのため、底本の画像等により底本の状態を適宜確認されたい。

2. 収録資料

本コーパスに収録した資料は表1のとおりである。

表1 コーパス収録資料

資料名	著者	刊行年	刊行地	資料 ジャンル	底本所蔵場所	底本所蔵ID*
交易問答	加藤弘之	明治2(1869)	東京	非文芸	国立国語研究所	卷之上:1001796059 卷之下:1001795879
安愚楽鍋 初編/二編/三編	仮名垣魯文	初編/二編:明治4(1871) 三編:明治5(1872)	東京	文芸	国立国会図書館	508979
開化のはなし	曲肱軒主人	明治5(1872)か	行田	非文芸	国立国語研究所	1002377834
春秋雑誌会話篇	アーネスト・サトウ	明治6(1873)か	—	非文芸	早稲田大学図書館	文庫08 C0763
文明開化 初編/二編	加藤祐一	初編:明治6(1873) 二編:明治7(1874)	大坂	非文芸	国立国語研究所	初編上巻:1002378071 初編下巻:1002378089 二編上巻:1001169331 二編下巻:1001169364
百一新論	西周	明治7(1874)	東京	非文芸	国立国会図書館	418333
よりあひばなし 初編	榊原伊裕	明治7(1874)	大坂	非文芸	国立国語研究所	初編上巻:1001946472 初編下巻:1001946506
開化問答 初編/二編	小川爲治	初編:明治7(1874) 二編:明治8(1875)	東京	非文芸	国立国語研究所	初編上巻:1001168077 初編下巻:1002377891 二編上巻:1002377909 二編下巻:1002377917
明治の光	石川富太郎	明治8(1875)	東京	非文芸	国立国会図書館	448897
文明田舎問答 初編	松田敏足	明治11(1878)	東京	非文芸	国立国会図書館	448877
民権自由論	植木枝盛	明治12(1879)	福岡	非文芸	国立国会図書館	437742

※国立国語研究所所蔵本は資料ID、国立国会図書館所蔵本は書誌ID、早稲田大学図書館蔵本は請求番号を示す。

表 1 に示したように、本コーパスに収録する資料は明治 2 (1869) ~明治 12 (1879) 年に刊行された 11 資料である。

このうち、『安愚楽鍋』は、開化期に流行した牛鍋屋を舞台とする戯作で、身分・職業・性別の多様な登場人物による話し言葉が描写される。用語索引 (国立国語研究所・斎賀・飛田・梶原 1974) が刊行されるなど、当時の話し言葉を知る上で重要な資料の一つとして知られており、本コーパスの収録対象とした。

『春秋雑誌会話篇』はアーネスト・サトウ著の *Kuaiwa Hen* (『会話篇』) の Part III にあたるものである。*Kuaiwa Hen* はローマ字表記の日本語の対話文とその英訳を収めた Part I、その注解を内容とする Part II、そして Part I の日本語の対話文の和文テキストである Part III の 3 パートからなる¹。Part III には「春秋雑誌会話篇」という内題が示されており、本コーパスではそれを資料名とした。『春秋雑誌会話篇』に収められた日本語の対話文は、外国人が日本語の日常会話を学ぶことを意図したもので、幕末から明治初年の定型的な口語表現や、様々な場面の中での多様な発話者による対話が含まれ、当時の話し言葉を知るための重要な資料として知られている。表記の面からは、前半は仮名文、後半は漢字仮名交じり文となっており、当時の仮名遣いの問題や文字教育という点からも注目されるべき資料であり、本コーパスの収録対象とした。

その他の 9 資料『交易問答』『開化のはなし』『文明開化』『百一新論』『よりあひばなし』『開化問答』『明治の光』『文明田舎問答』『民権自由論』は、吉野作造 (編) (1927-1932) 『明治文化全集』全 24 冊に収録された資料の内、①非文芸ジャンル、②口語体、③明治元~15 年刊行、の条件を満たすものとして選定した²。9 資料とも、開化期に西洋より輸入された思想・制度・事物について説く啓蒙書である。当時多く刊行された啓蒙書は、そのほとんどが文語体で書かれたが、一部、講述体あるいは問答体といった口語体で書かれたものがあった。この口語体の啓蒙書は、後に進展する非文芸ジャンルの資料における口語体書き言葉の発端・源流として注目すべき資料であることが指摘されており (木坂、1976、p. 160; 森岡 (編著)、1991、pp. 28-29; 等)、本コーパスの収録対象とした。

3. テキストの範囲とサンプル

収録対象資料の内、①表紙、②目録、③刊記、④広告、⑤挿絵、⑥表、を除くすべての文書要素のテキストをコーパスに収録した。ただし、文書要素中の漢文で書かれた段落のテキストは収録対象外とした。

テキストをコーパスに収録する際にテキストを一定の範囲で分割する必要があるが、その各範囲をサンプルと呼ぶ。本コーパスのサンプル単位は、『春秋雑誌会話篇』以外の 10 資料では、底本 1 冊の中を序・主要本文・跋等の文書要素に分割した、その各文書要素である。一方、『春秋雑誌会話篇』は全体で 25 章からなるテキストを 1 章=1 サンプルとして分割した。各サンプルを一意に特定する ID の構成を表 2 にあげる。

¹ *Kuaiwa Hen* の書誌については松村 (1998、pp. 351-419) に詳しい。

² 資料の選定にあたっては田中 (2012) の「近代語資料リスト」を参照した。リストの参照にあたり田中牧郎氏 (明治大学) の協力を仰いだ。

表2 サンプル ID の構成

左からの桁数	値	説明
1～2	60	時代区分を表す。すべて「60」で、「明治・大正」を表す。
3～5	C口語	サブコーパスを表す。すべて「C口語」で「明治初期口語資料」を表す。
6～9	(4桁の数字)	資料・編の刊行年を西暦で表す。
10	-	サンプルIDの区切り記号(アンダーバー)。
11～12	(2桁の数字)	資料の通し番号を表す。
13	(1桁の数字)	各資料内の編の通し番号を表す。『春秋雑誌会話篇』では2分冊の通し番号を表す。
14～15	(2桁の数字)	各資料・各編内のサンプルの通し場号を表す。『春秋雑誌会話篇』は2分冊を通した番号とし、章番号と一致させる。

4. テキストの仕様

4.1. テキストに使用する文字

本コーパスの電子化テキストに使用した文字の範囲は、JIS X 0213 (JIS の文字コード規格) の文字集合 (JIS 漢字の第 4 水準までを含む) に準拠した³。

文字集合に含まれない変体仮名については文字集合内の仮名によって電子化し、文字集合に含まれない記号類は、形・用途の近い文字集合内の記号によって電子化した。また、底本の文字のかすれや破損・抹消によって判読が困難な文字・記号は、「_」(空白記号、JIS 面区点 1-07-93、U+2423) によって表した。

漢字については、収録資料の書体が楷書・明朝体のように字画をくずさない書体と行書・草書のように字画をくずす書体に大別され、それぞれで電子化の方法が異なる。

まず、字画をくずさない書体の資料⁴の場合、文字集合に含まれない漢字については、須永・堤・高田 (2011) の方法を参照し以下の (1)～(5) の手順で電子化した⁵。

- (1) JIS X 0213 の「6.6.3 漢字の字体の包摂規準」に基づき、文字集合内の文字に包摂する。
- (2) (1) の包摂規準を適用できない字形差をもつ漢字のうち、近代に特有な微細な字形差を持つ漢字について追加包摂規準 (須永・堤・近藤ほか 2013) を設け、それに基づき文字集合内の文字に包摂する。

³ ただし、①JIS X 0213 附属書 7 2.1 b) に掲載される、戸籍法施行規則付則別表「人名用漢字許容字体表」(昭和 56 年法務省令 51) の漢字、及び常用漢字表 (昭和 56 年内閣告示第 1 号) のかっこ書き内の漢字 (「いわゆる康熙字典体」) のうち、JIS X 0208 で包摂していた漢字、②JIS X 0213:2004 において UCS との互換のために追加された 10 字、についてはこれを用いない。

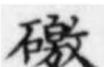
⁴ 『百一新論』『民権自由論』を字画をくずさない書体の資料とした。また、『春秋雑誌会話篇』は底本は字画をくずした資料であるが、電子化に際しては仏訳本『春秋雑誌会話篇』の活字字体に基づき、字画をくずさない資料として扱った。仏訳本は松村 (1998, pp. 521-563) に掲出の影印版に拠る。なお、この影印版の原本は現在、東京大学文学部国語研究室松村文庫に所蔵されている (請求番号 A3:0044)。

⁵ この電子化の手順は須永・堤・高田ほか (2011) の方法に基づくものである。

- (3) (2)の包摂規準を適用できない字形差をもつ漢字は、類似の意味・用法を持ち、同一の漢字部品を有する同音・同訓の文字集合内の文字で代用する。
- (4) (3)による代用が不可能な文字のうち、Unicode に収録されている漢字はそれにより電子化する。
- (5) (4)による入力不可能な文字は、「𠄎」(げた記号、JIS 面区点 1-02-14、U+3013) で表す⁶。

この手順による漢字の電子化例を次の表 3 に示す。

表 3 字画をくずさない書体の漢字の電子化例

底本字形例	電子化で使用する字体	適用した手順
	平	(1)
	欲	(2)
	鹽	(3)
	礫	(4)

次に、字画をくずした書体の資料⁷では、漢字の電子化の候補として異体字関係にある複数の字体が想定される際、選択する字体に揺れが生じないようにするため、コーパスで使用する文字集合内漢字について異体字グループを設定し、グループ中で現代もっとも一般的に使用される 1 字体によって電子化した。この異体字グループの設定と電子化に使用する 1 字体の選定に関する規定を以下にあげる。

- A) 高田・盛・山田 (2012) の「異体漢字対応テーブル」⁸に基づき、コーパスの文字集合内漢字で異体字関係にある複数字をグループ化し、そのグループから次の①～⑥の規程を当てはめ、電子化に使用する 1 字体を選定する (グループ内で①～⑥のうち複数の規定が当てはまる場合、番号の若い規定を優先する)。
 - ① 常用漢字表掲出の新字体・旧字体の関係の場合、新字体で電子化する。
 - ② 表外漢字字体表の印刷標準字体とその新字体の関係の場合、印刷標準字体で電子化する。
 - ③ JIS 区点位置入替字の関係の場合、第二水準字体で電子化する。
 - ④ 人名用漢字表 (戸籍法施行規則別表第二) で異体字関係にあり、常用漢字表字体・表外漢字字体表印刷標準字体が含まれる場合、常用漢字表字体・表外漢字字体表印刷標準字体で電子化する。
 - ⑤ 人名用漢字表 (戸籍法施行規則別表第二) で異体字関係にあり、④以外の場合、人名用漢字表で最初にあげられている字体で電子化する。
 - ⑥ 簡化字総表で簡体字・繁体字の関係にあり、常用漢字表の新字体・旧字体で同種の簡略化が見られるものの場合、繁体字で電子化する。

⁶ 本コーパスではこの手順により「𠄎」で表した文字はない。

⁷ 『交易問答』『安愚楽鍋』『開化のはなし』『文明開化』『よりあひばなし』『開化問答』『明治の光』『文明田舎問答』を字画をくずした書体の資料とした。

⁸ 「異体漢字対応テーブル」の利用にあたり高田智和氏 (国立国語研究所) の協力を仰いだ。

- B) その他、コーパス構築で設定が必要と考えられる異体字グループ（主にくずし字で字形が近似する複数字体）の場合、現代日本語で一般的に用いられる字体で電子化する。

この規定による漢字の電子化例を次の表 4 に示す。

表 4 字画をくずした書体の漢字の電子化例

底本字形例	異体字グループ	規定種別	電子化で使用する字体
	覚／覺	A①	覚
	搔／搔	A②	搔
	涼／涼	A④	涼
	堯／堯	A⑤	堯
	瀉／泻	A⑥	瀉
	迄／迄	B	迄

また、電子化の候補が想定されない場合は、「=」（げた記号、JIS 面区点 1-02-14、U+3013）で表した。

4.2. テキストの校訂

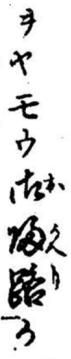
本コーパスは、コーパスのテキストを形態素解析用辞書 UniDic による形態素解析に適したものとするため、底本のテキストに対して以下の A)～E) にあがる校訂を施し、コーパスのテキストを作成した。ただしルビテキストでは「B) 踊り字」の校訂は行っていない。

なお、コーパス検索アプリケーション「中納言」(6 節参照) では、本行のテキストに関して、校訂後のコーパスのテキストと同時に、校訂前のテキストを底本の状態に近い形で電子化したものを「原文 KWIC」「原文文字列」として表示させることができるほか、底本の画像リンクから底本の字形を参照することもできる。ただし、ルビテキストは「中納言」では常に校訂後のテキストが表示されるため、校訂前のテキストの確認は底本の画像を参照されたい。

A) 片仮名

外来語を除き、片仮名は平仮名に置き換える (表 5)。

表5 片仮名の電子化例⁹

底本テキスト	コーパステキスト	原文 KWIC・原本文字列
	をやもう御帰路か	ヲヤモウ御帰路か
	アントロホロジト譯して人性學と云ひ	アントロホロジト譯シテ人性學ト云ヒ

B) 踊り字

踊り字は繰り返される文字列に置き換える（表6）。ただし、「国々」「人と」「中 / \」等、1短単位内部で直前の漢字1字を繰り返す踊り字は置き換えの対象としない。

表6 踊り字の電子化例

底本テキスト	コーパステキスト	原文 KWIC・原本文字列
	かちかち山	かち / \ 山

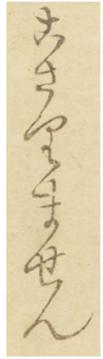
⁹ 表中の電子化例ではルビテキストの表示は省略した。表6～表8も同様。

底本テキスト	コーパステキスト	原文 KWIC・原本文字列
	なさるる	なさるゝ

C) 濁点無表記

濁音が期待される仮名に濁点付き仮名が用いられていない場合は、濁点の無表記と判断し、該当の濁音を表す濁点付き仮名に置き換える（表 7）。ただし、清濁両形がある語¹⁰については、置き換えの対象としない。

表 7 濁点無表記の電子化例

底本テキスト	コーパステキスト	原文 KWIC・原本文字列
	到るまで	到るまで
	ござりません	こさりません

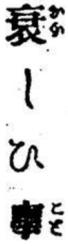
D) 誤植

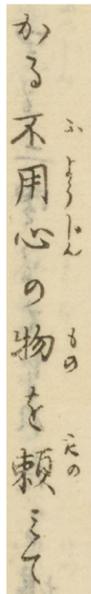
原文の誤植（脱字、衍字、前後文字列の転倒、誤字）と思われる表記は、訂正する（表 8）。ただし、仮名遣いの誤りや、語形のバリエーション、当時通用していたと考えられる同音漢字による異表記¹¹などは、訂正の対象としない。

¹⁰ 当該の語に清濁両形があるかどうかの判定は、原則として『日本語国語大辞典 第二版』によるものとする。清濁両形が辞書の「見出し」にある場合のほか、「語義説明」内に「○○とも」「古くは○○」の形で異語形を示す場合は、清濁両形があるものと判断する。

¹¹ 語形のバリエーションかどうかの判定は、注 10 に示した清濁両形の有無の判定に準ずる。また、通用の異表記かどうかの判定は、①『日本語国語大辞典 第二版』及び②近代語のコーパスによる出現状況によるものとする。①は、見出しの「漢字表記」のほか、「用例文」中の表記、「表記」欄の表記などを通用の異表記と見なす。①が適用できない場合、②近代語のコーパス（公開済みのもののほか、内部資料を含む）において、複数のサンプルに出現し、出現数が少ない表記を異表記と見なす。

表8 誤植の電子化例

底本テキスト	コーパステキスト	原文 KWIC・原本文字列
	矢張	矢張
	衰しひ事	衰しひ事
	跡は後席に譲り升	跡は後席にに譲り升

底本テキスト	コーパステキスト	原文 KWIC・原本文字列
	かかる不用心の物を頼みて	かる不用心の物を頼みて

E) 分かち書きのスペース

『春秋雑誌会話篇』第一冊は分かち書きされた仮名文表記のテキストである。この分かち書きのスペースはコーパステキストでは電子化を省略した。例えば、底本で「あした いかう と おもふ」のように分かち書きされていても、コーパステキストでは「あしたいかうとおもふ」(60C 口語 1873_11101)のようにスペースは省略した。一方、「原文 KWIC」「原文文字列」ではスペース位置を「□」(四角、JIS 面区点 1-02-02、U+25A1) で表し、「あした□いかう□と□おもふ」のように表示した。

5. 形態論情報

5.1. 形態論情報の概要

本コーパスでは、原則として底本の本行のテキストを主本文(主たる本文)として、それに対して形態論情報(語彙素・語彙素読み・品詞・活用型・活用形等の語に関する情報)を付与した。テキストの読みは右ルビのある場合はそれに拠った¹²。

形態論情報は短単位のみ付与しており、長単位は未付与である¹³。短単位の形態論情報は、形態素解析辞書 UniDic を使用した形態素解析に基づき、一部を人手により修正することで付与した。ただし、人手による修正が尽くされておらず、精度(適合率)98%のデータであることに留意の上、利用されたい。形態論情報の各項目については表 10 を参照のこと。

5.2. 形態論情報の多重化

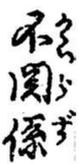
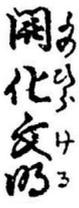
「不関係(読み:カカラズ)」「壺(読み:ヒトツ)」「開化文明(読み:ヨノヒラケル)」のように、短単位の境界や順序と本行の漢字との間に対応関係が見られない場合、本行の漢字に直接形態論情報が付与できないため、形態論情報の多重化の技術(村山・小木曾・中村 2017)を用い、本行の漢字の読み

¹² ただし、本行「なか」の右側に「吉原」と記されているような、本行右側の傍注は右ルビとは見なさない。また、「恐怖(右ルビ:けうふ、左ルビ:ヲジヲソル)し」のような注釈的に使用される左ルビを読みとすることはしなかった。

¹³ 短単位・長単位の詳細については、小椋・小磯・富士池ほか(2011)を参照のこと。

を平仮名表記したテキスト¹⁴を主本文とし、それに対して形態論情報を付与した。その場合の本行の漢字は副本文とし、品詞「対象語無し」として未知語扱いとした（表9）。

表9 形態論情報の多重化の例

底本 テキスト	本文 の主 副	主な形態論情報					
		書字形 出現形	語彙素	語彙素 読み	品詞	活用型	活用形
	主	かかはら	関わる	カカワル	動詞-一般	五段-ラ行	未然形-一般
		ず	ず	ズ	助動詞	助動詞-ヌ	連用形-一般
	副	不関係			対象語無し		
	主	ひと	一	ヒト	名詞-数詞		
		つ	つ	ツ	接尾辞-名詞的 -助数詞		
	副	壱			対象語無し		
	主	よ	世	ヨ	名詞-普通名詞 -一般		
		の	の	ノ	助詞-格助詞		
		ひらける	開ける	ヒラケル	動詞-一般	下一段-カ行	終止形-一般
	副	開化文明			対象語無し		

5.3. 形態論情報の検索の注意点

このように、本コーパスでは本行のテキストとは異なるテキストに対して形態論情報の付与を行っている箇所がある。そのため、「中納言」（6節参照）で検索を行う場合、以下の点で注意が必要である。

(1) 「中納言」の短単位検索における注意点

「中納言」の検索方法の一つである短単位検索において、書字形出現形を検索する場合、形態論情報の多重化を行っている箇所では主本文である平仮名表記したテキストが対象となる。副本文である本行の漢字を検索対象としたい場合は、「検索動作」欄の「副本文」ボックスで「副本文を検索対象に含む」を選択する必要がある。

さらに、副本文では主本文側で複数の短単位が付与される漢字をまとめて1短単位として扱っているため、1短単位の範囲が事前に推測できないことも多い。例えば書字形出現形が「開化」となる語を検索したい場合、「副本文を検索対象に含む」を選択した上で検索ボックスに「開化」と入力して検索しても、表9の3例目「開化文明」はマッチしない。これは書字形出現形の検索では、検索ボックスに1短単位全体の書字形出現形を入力する必要があるためである。このような検索漏れを防ぐため、副本文も含めて書字形出現形を検索する場合は、「%開化%」のよう

¹⁴ ルビのある場合は、4.2節の校訂（「B」踊り字）も含む）を行ったルビテキストを主本文とした。ルビのない場合は、現代仮名遣い・平仮名表記とした読みを主本文とした。

に検索ボックスに入力する。「%」（半角）は任意の文字列を表現するメタ文字で、「%開化%」は「『開化』を含むテキスト」を意味する。これにより、副本文「開化文明」のような「開化」を含む書字形出現形を持つ1短単位も検索することが可能となる。

また、「副本文を検索対象に含む」を選択した場合は、前方・後方共起の条件指定をした検索はできないことにも注意が必要である。

(2) 「中納言」の文字列検索における注意点

「中納言」の検索方法の一つである文字列検索において、検索対象となるのは底本の本行のテキストである。よって、形態論情報の多重化を行っている箇所では副本文が検索対象となり、主本文である平仮名表記したテキストは対象とならない。例えば、「ひらける」を文字列検索しても表9の3例目はマッチしない。「ひらける」は底本の本行のテキストではないためである。

また、文字列検索の検索結果では、検索文字列を含む本行の短単位の範囲に対応する主本文の最後の1短単位の形態論情報のみが表示される。例えば、表9の3例目の場合、「開化」「文明」「開化文明」といった文字列で文字列検索を行うと、返ってくる検索結果で表示されるのは主本文の語彙素「開ける」の形態論情報のみであり、主本文の語彙素「世」「の」の形態論情報や副本文の品詞「対象語無し」の形態論情報は表示されない(図1)。

60C口語 1871_02203	2680	1690	0 牛肉で杯一ぱくく腹組。#牛の小便十八党。むすぶ交り健児の社中。#文明 往も復るも流行の。牛肉で杯一ぱくく腹組。#牛の小便十八党。むすぶ交り健児の社中。#文明	開化 かいくわ 開化	のざんざりあたま。 王政復古の 惣髪頭。 因循姑息の 半髪額。 #歌妓は 箱持の のざんざりあたま。王政復古の惣髪頭。因循姑息の半髪額。#歌妓は箱持の案内に属。娼妓は引手の家婢にひかれ。老若男女の差	カイカ 開化 カイカ	開化 開化 カイカ	名詞-普通名詞-サ変可能			開化	かいくわ
60C口語 1871_02204	65840	35630	0 ありさへすれおなじことでふきかへるたびおちひさくなり性もおひおひわくるくつたを ことでふきかへるたびおちひさくなり性もおひおひわくるくつたを	開化文明 よのひらける 開化文明	にしたがって人も漸次に伶利になりりかうにすぎずて姦曲をめぐらすものがでる にしたがって人も漸次に伶利になりりかうにすぎずて姦曲をめぐらす	ヒラケル 開化 ヒラケル	開化 開化 ヒラケル	動詞-一般	下一段力行	終止形一般	開化文明	よのひらける
60C口語 1872_02301	530	330	0 肉食は万民の滋養なり故牧牛の国家に益ある豈他獣と等しからんや#方今 種痘は天下に肉食ハ万民ノ滋養ナリ # 故牧牛ノ国家ニ益アル豈他獣ト等シカラヤ # 方今	開化 カイカ 開化	捨く進み市井の細民に雖も牛痘(牛肉)の世に功あるを知るものから随つて 捨く進み市井ノ細民ト雖も牛痘(牛肉)ノ世ニ功アルヲ知ルモノカラ随	カイカ 開化 カイカ	開化 開化 カイカ	名詞-普通名詞-サ変可能			開化	カイカ

図1 「中納言」で「開化」を文字列検索した検索結果の表示例

6. 「中納言」の表示項目

本コーパスでは、テキストおよびアノテーションのデータは、コーパス検索アプリケーション「中納言」での検索結果の形で利用者に提供する(図2)。

サンプルID	開始位置	連番	コア	前文脈	キ	後文脈	語彙素読み	語彙素	語形	品詞	活用型	活用形	原文文字列	振り仮名	本文種別	話者	文体	ジャンル	作品名	成立年	巻名等	作者	生年	性別	底本	ページ番号	底本リンク	参考リンク
60C口語 1869_01102	390	270	0	加藤弘蔵著 # 読六 # なんど才助君。(備には一向 合点の)参り申さぬことか 加藤弘蔵著 # 読六 # ナント才助君。備には一向合点の参り申さぬことか	ござる	は。# 今度御公儀と申す者がなくなつて。天下の御政事は # 天子様でなされる は。# 今度御公儀と申す者がなくなつて。天下の御政事は # 天子様でなされる様になつたから。是迄御公儀で御可なりなされた	ゴザル	御座る	ゴザル	動詞-非自立可能	五段-う	終止形一般	ござる		会話	読六	口語	非文芸	交易問答	1869	上[本文]	加藤弘之(作)	1836	男	国立国語研究所蔵 交易問答巻之上	17	Nipon	
60C口語 1869_01102	1840	1250	0	開きになり。又東京でも交易を御開きなさるといふは。何たることか 開きになり。又東京でも交易を御開きなさるといふは。何たることか	ござる	は。# どもも此読六様には。一向 合点の)参り申さん。# 或先生の御語に。元来此日本といふ御国は神国でござるから。日本	ゴザル	御座る	ゴザル	動詞-非自立可能	五段-う	意志推量形	ござる		会話	読六	口語	非文芸	交易問答	1869	上[本文]	加藤弘之(作)	1836	男	国立国語研究所蔵 交易問答巻之上	17	Nipon	

図2 「中納言」の検索結果の表示例

次の表 10 に、「中納言」の検索結果で表示されるテキスト・アノテーションのうち、初期設定で表示される項目と、初期設定では表示されないが本コーパスで特に注意が必要な項目である「主本文」「多重化種別」「文体」「歌番号」「性別」（表中*を付す）について内容を示す。

表 10 「中納言」検索結果の主な表示項目

情報種別	項目名	内容
コーパス情報	サンプル ID	検索対象の含まれるサンプルの ID (3 節参照)。
	開始位置	「キー」の本行テキストの先頭の文字の、サンプル内における位置を表す ID。10 きざみの連番。
	連番	「キー」に対応する主本文の短単位 ¹⁵ の、サンプル内における位置を表す ID。10 きざみの連番。
	コア	検索対象の含まれるサンプルがコア ¹⁶ データであることを表す。「1」がコア、「0」が非コアを表す。本コーパスではすべて「0」が表示される。
	主本文*	検索結果として表示されている形態論情報の付与されている本文の主副 (5 節参照) の別。「1」が主本文、「0」が副本文を表す。
	多重化種別*	形態論情報の多重化 (5.2 節参照) を行っている箇所、その多重化の種別を表す。本コーパスではすべて「振り仮名」である。
形態論情報 ¹⁷	前文脈	「キー」の前方文脈。
	キー	検索対象の含まれる本行テキスト ¹⁸ 。右ルビや本行右側の傍注のある場合はそのテキストが本行テキストの上側に表示される。
	後文脈	「キー」の後方文脈。
	原文 KWIC	上記項目「前文脈」「キー」「後文脈」に対する、校訂前の底本に近い形のテキスト (4.2 節参照)。ただし、ルビは校訂後のテキストが表示される。

¹⁵ 「キー」に対応する主本文に複数の短単位がある場合は、最後の短単位。

¹⁶ コーパス全体に高精度な付加情報を付与することが困難な場合、一部分のテキストに対して高精度な付加情報を付与し、コーパスの核となるデータとすることがある。『日本語歴史コーパス』ではその高精度なデータを「コア」と呼び、それ以外のデータを「非コア」と呼ぶ。本コーパスは形態論情報が未整備な点が残っているため、「コア」は設けず、すべて「非コア」としている。

¹⁷ 形態論情報の個々の項目の内容は、「前文脈」「キー」「後文脈」「原文 KWIC」「原文文字列」「振り仮名」を除き、UniDic の見出しに対応している。各項目の詳細については、口語文体の短単位の場合は小椋ほか(2011)、文語文体の短単位の場合は国立国語研究所コーパス開発センター (近藤明日子) (編) (2016) を参照のこと。

¹⁸ 「キー」の表示範囲は設定により変更可能であるが、ここでは初期設定で表示される「キー」の範囲を指して言う (表中の他の項目も同様)。初期設定では、短単位検索の場合は検索対象とした短単位の含まれる本行の 1 短単位、文字列検索の場合は検索対象文字列に含まれる本行の最後の 1 短単位、が表示範囲となる。

情報種別	項目名	内容
	語彙素読み	「キー」に含まれる短単位 ¹⁹ の語彙素（下記項目「語彙素」参照）の読み。片仮名表記である。
	語彙素	「キー」に含まれる短単位の語彙素の表記。語彙素は、単語の様々なバリエーション（語形、活用形、表記形など）を統合した辞書の見出しに相当するもので、一般の和語・漢語は漢字平仮名表記、外来語・人名・地名は片仮名表記である。
	語形	「キー」に含まれる短単位の語形。語形は、語彙素では統合される語形の別（例：語彙素「矢張り」に対する「ヤハリ」「ヤッパリ」など）や活用型の別（例：語彙素「読む」に対する「ヨム（五段-マ行）」「ヨム（文語四段-マ行）」「ヨメル（下一段-マ行；可能動詞形）」など）等を区別した語の個々の形に相当するもの。片仮名表記である。
	品詞	<p>「キー」に含まれる短単位の品詞で、UniDic の体系に基づく。学校文法における「形容動詞」は、語幹が「形状詞」、活用語尾が「助動詞」に分割される点に注意が必要である。</p> <p>UniDic の体系に基づかない特殊な品詞には以下の種類がある。</p> <p>メタ（誤り） …言い誤り等に言及している文字列。 例：人力車をりんりき車とおぼへてゐる也</p> <p>洒落断片 …洒落・掛詞等による臨時的な形態や短単位境界の認定に適さない文字列。 例：出謗題詠の自慢滅法海外の論（「出放題」と「題詠」の洒落）</p> <p>漢文 …漢文の文字列。</p> <p>解釈不明 …語の解釈が困難なため、解釈を保留した文字列。</p> <p>対象語無し …形態論情報の多重化（5.2 節参照）を行っている箇所 の副本文の文字列。</p> <p>なお、これらの特殊な品詞の付与された短単位は、形態論情報に関する項目のうち、「前文脈」「キー」「後文脈」「原文 KWIC」「品詞」「原文文字列」「振り仮名」以外は空欄となっている。</p>

¹⁹ 「キー」に対応する本文に主本文・副本文の 2 種がある場合は、「主本文」項で示された主副の別に対応する短単位を指す。また、「キー」に対応する本文に短単位が複数ある場合は、最後の短単位を指す。以下の「語彙素」「語形」「品詞」「活用型」「活用形」「原文文字列」も同様。

情報種別	項目名	内容
	活用型	「キー」に含まれる短単位の活用の型。活用語の場合のみ表示される。口語活用は活用の型と行で「五段-サ行」のように、文語活用は「文語」が加わり「文語四段-サ行」のように示される。検索対象の「文体」項目（下記項目「文体」参照）の値が「文語」である活用語には文語活用型を、「口語」である活用語には口語活用型を割り当てる。ただし、口語内にあっても口語活用型が存在しない語や活用形（文語助動詞「ごとし」「き」、文語二段活用の連体形語形など）については文語活用型を割り当て、同様に、文語内にあっても文語活用型が存在しない語や活用形（口語助動詞「ない」、意志推量形など）については、口語活用型を割り当てる。
	活用形	「キー」に含まれる短単位の活用形。活用語の場合のみ表示される。学校文法では「未然形」と助動詞「う・よう」に分割される形態は、結合して「意志推量形」とする点に注意が必要である。文法的に特定の活用形が期待される箇所（単語同士の接続関係や文末等）で、それとは異なる形態が用いられている場合は、語の形態に即して活用形を割り当てる。 例) 高声に読みあげらるる（文末だが活用形は「連体形-一般」）
	原文文字列	「キー」の本行テキストの、校訂前の底本に近い形（4.2 節参照）。
	振り仮名	「キー」の、右ルビや本行右側の傍注の校訂後のテキスト。なお、左ルビは「中納言」の検索結果には表示されない。
本文情報	本文種別	検索対象に含まれる文が「地の文」以外の場合の、その種別。以下の種類がある。 会話 …会話・独話・心内発話等の引用 引用 …文献等からの引用、記事に対する雑誌記者・編集者の説明・解説・注釈等 その他-話者 …会話の冒頭に表示される話者名 その他-会話注記 …会話の冒頭に表示される会話の種別
	話者	上記項目「本文種別」が「引用」の場合の典拠文献名や著者名、「会話」の場合の話者名。不明の場合は「*」で示す。『春秋雑誌会話篇』では、 <i>Kuaiwa Hen</i> の Part I において英文での話者表示のある場合それを()に括って併記する。

情報種別	項目名	内容
	文体*	<p>検索対象の含まれる文の文体。以下の種類がある。</p> <p>文語…文語体。文末辞が「なり」「たり」「つ」「ぬ」「き」「けり」のもの。</p> <p>口語…口語体。文末辞が「だ」「ぢや」「である」「です」「ます」のもの。</p> <p>漢文…漢文。</p> <p>韻文…日本語による韻文。</p> <p>なお、「文語」「口語」の分類は、原則「地の文」はサンプル単位で²⁰、「地の文」以外の「会話」「引用」等は各「会話」「引用」の単位で行う。文語と口語が混在する文章の場合は、主たる文体を判断し、「文語」「口語」いずれかの種別を付与する。「漢文」「韻文」については、該当テキスト箇所を判断し種別を付与する。</p>
	歌番号*	<p>『春秋雑誌会話篇』で検索対象の含まれる章番号と会話番号。番号は底本に記されたものを使用する。「01_23」のように、それぞれ2桁の数字で表記した章番号と会話番号を「_」（アンダーバー）でつないだ形式で表す。</p>
作品情報	ジャンル	<p>検索対象の含まれるサンプルの、文章内容に基づく分類。以下の種類がある。</p> <p>文芸 …小説・戯曲・詩歌の類</p> <p>非文芸…「文芸」以外</p>
	作品名	検索対象の含まれるサンプルが収録された資料名。
	成立年	検索対象の含まれるサンプルが収録された資料・編の刊行年。
	巻名等	<p>検索対象の含まれるサンプルが収録された資料の編名・巻名、およびサンプルのタイトル。底本テキスト中にタイトルがない場合は内容を表す名付けを〔 〕に括って示す。序・跋等以外の主要本文のタイトルは「〔本文〕」と示す。</p>
作者情報	作者	<p>検索対象の含まれるサンプルの著者名。後ろに「（作）」を付けて示す。</p> <p>著者名の認定は、底本テキストの記載に基づく。ただし、現在一般的に知られている呼称に変えた場合がある。また、著者が不明なものは「*」で示す。</p> <p>著者が「国立国会図書館典拠データ検索・提供サービス（Web NDL Authorities）」に収録されている場合は、そのウェブページへのリンクを付与する。</p>

²⁰ ただし、例えば、通常地の文は文語、地の文内の割書部分のみ口語となっているような、レイアウト差と文体差が対応している場合は、同一サンプル中の地の文の内部で「文語」「口語」の区別を行う。

情報種別	項目名	内容
	生年	検索対象の含まれるサンプルの著者の生年。西暦4桁で示す。不明な場合は「*」で示す。生年の認定は、各種人名事典・Web NDL Authorities等に基づく。
	性別*	検索対象の含まれるサンプルの著者の性別。不明な場合は「*」で示す。
底本情報	底本	検索対象の底本（原資料）。所蔵場所・資料名・編名・巻名を示す。
	ページ番号	検索対象の底本における出現ページ番号。丁数に加えて「オ」（表）「ウ」（裏）により表裏の別を示す。序・跋等で主要本文とは別に「序一」「跋一」のように丁付けされている場合は、それに対応した形式で示す。底本にないページ番号を補った場合は〔 〕で括弧で示す。
その他	底本リンク	検索対象の底本画像へのリンク。以下の種類がある。  …国立国語研究所蔵本画像へのリンク  …国立国会図書館蔵本画像へのリンク  …早稲田大学図書館蔵本画像へのリンク
	参照リンク	検索対象の底本以外の参照本画像へのリンク。本コーパスでは該当画像がないため空欄。

謝辞

本コーパスは、国立国語研究所共同研究プロジェクト「通時コーパスの設計」（2009-2016）、国立国語研究所共同研究プロジェクト「通時コーパスの構築と日本語史研究の新展開」（2016-）、JSPS 科研費 JP15H01883「日本語歴史コーパスの多層的拡張による精密化とその活用」（2015-2019）による研究成果である。また、『春秋雑誌会話篇』のコーパスデータの構築の一部は JSPS 科研費 JP17K02786「英学資料のテキストデータ化に関する研究」（2017-2020）の成果である。

本コーパスに収録した資料の底本の利用や「中納言」の検索結果に表示されるリンク先の画像は、以下の各所蔵先のウェブサービスにおいて公開されているデジタル画像に拠るものである。

- 日本語史研究資料[国立国語研究所蔵] <https://dglb01.ninjal.ac.jp/ninjalddl/>
『交易問答』 <https://dglb01.ninjal.ac.jp/ninjalddl/bunken.php?title=kouekimondou>
『開化のはなし』 <https://dglb01.ninjal.ac.jp/ninjalddl/bunken.php?title=kaikanohanasi>
『文明開化』 <https://dglb01.ninjal.ac.jp/ninjalddl/bunken.php?title=bunmeikaika>
『よりあひばなし』 <https://dglb01.ninjal.ac.jp/ninjalddl/bunken.php?title=yoriaibanasi>
『開化問答』 <https://dglb01.ninjal.ac.jp/ninjalddl/bunken.php?title=kaikamondou>
- 国立国会図書館デジタルコレクション <https://dl.ndl.go.jp/>
『安愚楽鍋』初編 <https://dl.ndl.go.jp/info:ndl.jp/pid/882303>
二編 <https://dl.ndl.go.jp/info:ndl.jp/pid/882304>
三編 <https://dl.ndl.go.jp/info:ndl.jp/pid/882305>
『百一新論』 <https://dl.ndl.go.jp/info:ndl.jp/pid/753010>
<https://dl.ndl.go.jp/info:ndl.jp/pid/753011>
『明治の光』 <https://dl.ndl.go.jp/info:ndl.jp/pid/798695>
<https://dl.ndl.go.jp/info:ndl.jp/pid/798696>
『文明田舎問答』初編 <https://dl.ndl.go.jp/info:ndl.jp/pid/798672>
『民権自由論』 <https://dl.ndl.go.jp/info:ndl.jp/pid/783715>
- 早稲田大学図書館古典籍総合データベース <https://www.wul.waseda.ac.jp/kotenseki/>

参考文献

- 小椋秀樹・小磯花絵・富士池優美・宮内佐夜香・小西光・原裕 (2011) 『『現代日本語書き言葉均衡コーパス』形態論情報規定集第4版(上)(下)』特定領域研究「日本語コーパス」平成22年度研究成果報告書, 国立国語研究所, <http://doi.org/10.15084/00002855>, <http://doi.org/10.15084/00002856>
- 木坂基 (1976) 『近代文章の成立に関する基礎的研究』風間書房
- 国立国語研究所コーパス開発センター(近藤明日子)編(2016)『近代文語 UniDic 短単位規程集 Ver. 1.1』国立国語研究所コーパス開発センター,
https://pj.ninjal.ac.jp/corpus_center/chj/doc/unidic-MLJ_rulebook_v1_1.pdf
- 国立国語研究所・斎賀秀夫・飛田良文・梶原滉太郎(1974)『牛店雑談安愚楽鍋 用語索引』(国立国語研究所資料集9)秀英出版
- 須永哲矢・堤智昭・近藤明日子・木川あづさ・服部紀子(2013)「明治中期雑誌の異体漢字とJIS漢字—『国民之友』を事例として—」『じんもんこん2013論文集』2013(4), pp.201-208
- 須永哲矢・堤智昭・高田智和(2011)「明治前期雑誌の異体漢字と文字コード—『明六雑誌』を事例として—」『じんもんこん2011論文集』2011(8), pp.381-388
- 高田智和・盛思超・山田太造(2012)「網羅性を志向しない異体漢字対応テーブル」『研究報告人文科学とコンピュータ』2012-CH-93(2), pp.1-7
- 田中牧郎(2012)「近代語コーパスにおける資料選定の考え方」田中牧郎・岡島昭浩・小木曾智信・小野正弘・小島聡子・島田泰子・朱京偉・高田智和・張元哉・陳力衛・近藤明日子・須永哲矢『近代語コーパス設計のための文献言語研究 成果報告書』国立国語研究所共同研究報告12-03, pp.13-26
- 松村明(1998)『増補 江戸語東京語の研究』東京堂出版
- 村山実和子・小木曾智信・中村壮範(2017)「形態論情報の多重化による洒落本コーパスの質的拡張」『研究報告人文科学とコンピュータ(CH)』2017-CH-114(8), pp.1-8
- 森岡健二(編著)(1991)『近代語の成立 文体編』明治書院
- 吉野作造(編)(1927-1932)『明治文化全集』日本評論社(明治文化研究会(編)(1967-1974), 復刻版に拠る)

参考 URL

- UniDic <https://unidic.ninjal.ac.jp/>
コーパス検索アプリケーション「中納言」 <https://chunagon.ninjal.ac.jp/>
『日本語歴史コーパス』 <https://ccd.ninjal.ac.jp/chj/>
国立国会図書館典拠データ検索・提供サービス (Web NDL Authorities) <http://id.ndl.go.jp/auth/ndla/>

画像典拠

- 表3(掲出順) 国立国会図書館蔵『民権自由論』2才 (<https://dl.ndl.go.jp/info:ndljp/pid/783715/7>), 国立国会図書館蔵『民権自由論』3才 (<https://dl.ndl.go.jp/info:ndljp/pid/783715/8>), 国立国会図書館蔵『民権自由論』15才 (<https://dl.ndl.go.jp/info:ndljp/pid/783715/20>), 国立国会図書館蔵『百一新論』卷之上21才 (<https://dl.ndl.go.jp/info:ndljp/pid/753010/23>)
- 表4(掲出順) 国立国語研究所蔵『よりあひばなし』初編上2才 (https://dglb01.ninjal.ac.jp/ninjal_dl/show.php?title=yoriaibanasi&issue=1&num=9&page=8), 国立国語研究所蔵『文明開化』初編上7ウ (https://dglb01.ninjal.ac.jp/ninjal_dl/show.php?title=bunmeikaika&issue=1&num=22&page=21), 国立国語研究所蔵『よりあひばなし』初編上4才 (https://dglb01.ninjal.ac.jp/ninjal_dl/show.php?title=yoriaibanasi&issue=1&num=13&size=50&page=12), 国立国会図書館蔵『文明田舎問答』初編4才 (<https://dl.ndl.go.jp/info:ndljp/pid/798672/8>), 国立国会図書館蔵『文明田舎問答』初編34才 (<https://dl.ndl.go.jp/info:ndljp/pid/798672/38>), 国立国語研究所蔵『文明開化』初編上2ウ (https://dglb01.ninjal.ac.jp/ninjal_dl/show.php?title=bunmeikaika&issue=1&num=12&page=11)
- 表5(掲出順) 国立国会図書館蔵『安愚楽鍋』初編10才 (<https://dl.ndl.go.jp/info:ndljp/pid/882303/12>), 国立国会図書館蔵『百一新論』卷之上28才 (<https://dl.ndl.go.jp/info:ndljp/pid/753010/30>)
- 表6(掲出順) 国立国会図書館蔵『安愚楽鍋』初編5才 (<https://dl.ndl.go.jp/info:ndljp/pid/882303/7>), 国立国語研究所蔵『文明開化』初編下10才 (https://dglb01.ninjal.ac.jp/ninjal_dl/show.php?title=b

[unmeikaika&issue=2&num=21&page=20](#))

表 7 (掲出順) 国立国会図書館蔵『文明田舎問答』初編 1 ウ (<https://dl.ndl.go.jp/info:ndljp/pid/798672/6>) , 国立国語研究所蔵『開化問答』初編上 12 オ (<https://dglb01.ninjal.ac.jp/ninjalddl/show.php?title=kaikamondou&issue=1&num=33&page=32>)

表 8 (掲出順) 国立国会図書館蔵『文明田舎問答』初編 29 ウ (<https://dl.ndl.go.jp/info:ndljp/pid/798672/34>) , 国立国会図書館蔵『民権自由論』35 ウ (<https://dl.ndl.go.jp/info:ndljp/pid/783715/41>) , 国立国会図書館蔵『明治の光』巻二 23 ウ--24 オ (<https://dl.ndl.go.jp/info:ndljp/pid/798696/25>) , 国立国語研究所蔵『開化問答』二編上 23 オ (<https://dglb01.ninjal.ac.jp/ninjalddl/show.php?title=kaikamondou&issue=3&num=57&page=56>)

表 9 (掲出順) 国立国会図書館蔵『安愚楽鍋』二編下 17 ウ (<https://dl.ndl.go.jp/info:ndljp/pid/882304/42>) , 国立国語研究所蔵『交易問答』巻之上 20 オ (<https://dglb01.ninjal.ac.jp/ninjalddl/show.php?title=kouekimondou&issue=1&num=51&page=50>) , 国立国会図書館蔵『安愚楽鍋』二編下 17 ウ (<https://dl.ndl.go.jp/info:ndljp/pid/882304/42>)