

## 1. はじめに

『昭和・平成書き言葉コーパス ベストセラー書籍』（以下 SHC ベストセラー書籍）は、昭和期から平成期にかけて創刊されたベストセラー書籍群を対象に、通時的に言語の変化を捉えられるように設計したコーパスである。本文書では、コーパスの設計方法、テキストの仕様、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。

## 2. コーパスの設計と構築

### 2. 1 収録年と規模の設計

本コーパスは、SHC のサブコーパス共通で、『日本語歴史コーパス』（以下 CHJ と称する）の雑誌の収録年の間隔に倣って 8 年おきにデータを収録することとし、CHJ の雑誌に収録されている最終年の 1925 年に続けて、1933（昭和 8）年、1941（昭和 16）年、1949（昭和 24）年、1957（昭和 32）年、1965（昭和 40）年、1973（昭和 48）年、1981（昭和 56）年、1989（昭和 64／平成元）年、1997（平成 9）年、2005（平成 17）年、2013（平成 25）年の 11 カ年分のデータを構築した。

「SHC ベストセラー書籍」は、SHC の他のジャンルとのジャンル間比較ができるよう、「SHC 新聞」の収録語数に合わせて設計し、1933 年から 1949 年までは各年 20 万語前後、1957 年以降は各年約 40 万語前後の規模になるよう設計した。

### 2. 2 作品の選定

収録対象とする資料は、1933 年から 1973 年までは国立国語研究所国語辞典編集準備室（1984）の収録対象年に掲載されている全作品を選定し、1981 年以降は「年間ベストセラーアーカイブ」（トーン）や『出版年鑑』（出版ニュース社）を参照して上位 20 位までの書籍を選定した。これにより 1933 年から 2013 年までで計 259 組の書籍を収集した。なお、ベストセラー書籍は昨 11 月下旬から当年の 11 月下旬を収録対象とすることが多い兼ね合いで、前年に刊行された作品群を含む場合があるほか、刊行から時間を経てベストセラーとなる作品が存在する。本コーパスでは収録対象年よりは 2 年以内に発表・刊行された作品を収録対象とした。これにより、古典作品類や全集類、初刊行本から出版社や出版形態（文庫本等）を変えて刊行された作品などをコーパス構築対象外とした。

このうち、イラストや写真、図表を中心とする書籍（写真集、図鑑、ゲーム攻略本、地図、コミック、レシピ本など）はテキストから文脈が読み取りにくくコーパスの用例検索に向かないため、特殊な文字列が中心となる書籍（1949 年『解析精義』、2005 年『電車男』など）は形態論情報の付与が困難であるため、コーパスの構築対象外とした。

また、同一年に同一著者の作品が多数収録された場合には、年につき 2 作品まで採用することとした。売り上げ部数が判明している作品については売り上げ部数の多いものから採用し、判明していない作品については可能な限り出版社やシリーズが異なる作品から採用した（例：1973 年の遠藤周作氏の

『ぐうたら愛情学』（講談社）、『死海のほとり』（新潮社）を採用し、『ぐうたら交友録』（講談社）は不採用とした。このほか、『ノルウェイの森（上・下）』や『少年H（上・下）』『下天の夢（1-4）』のように複数巻がセットでベストセラーに収録された場合には、最初の巻を採用した。

以上の基準で選定した 210 冊の書籍を対象に、コーパスを構築した。不採用作品の一覧と理由については、巻末の資料を参照されたい。

### 2.3 サンプリング

本コーパスでは、各書籍からランダムサンプリングによるテキスト採集を行った。具体的には、各書籍のコーパステキスト収録対象とする本文のページ（内表紙や目次、広告、奥付などを除いたページ）のうち、無作為に抽出したページをランダムサンプリング起点として設定し、その起点のページの前後で、事前に定めた作品あたりの収録規模に合わせて、章や節などのある程度の文章のまとまりを確保した上でコーパステキストを採集した。2.1 で前述のコーパスの規模に合わせて、旧字体中心の 1933 年から 1949 年までは約 30 万字、1957 年以降は約 60 万字程度の規模になるように「作品あたりの収録規模」を定め、例えば、1933 年には 11 の収録対象作品があるが、30 万字を分割して一作品あたり 2.73 万字の規模になるように採集した。

例えば、1933 年の『海豹の如く』は内表紙や奥付を除いた p.2 から p.441 が抽選対象となり、サンプリング起点を、ツールを用いて無作為に抽出した結果、「305」となった。p.285 を起点に、1933 年の「作品あたりの収録規模」である 2.73 万字を超えるように前後の章や節のきりの良いところまで採用し、『海豹の如く』では p.285 から p.335 までを収録対象とした。

### 2.4 コーパスの概要

2.1 から 2.3 までのプロセスを経て、「SHC ベストセラー書籍」に収録した作品とその範囲を、表 1 に一覧に示した。

表中の「版」列は、「版」と「刷」の差も含めて底本の奥付に記載されているママの表記を参照しているが、基本的には初版のものを原本としている。「版」列に「(第 1 刷)」とあるのは、発行日のみ記されているもので、初版第 1 刷と考えるとよいものである。一部、「記載なし」とあるものは発行日の記載もないもので、恐らく初版ゆえのことと思われるが、確証がない点には注意されたい。

表中の「収録範囲」列のページ数を、「総頁数」列の数値で割って 100 をかけたものが「採用率」列の値である。「採用率」の数値は各作品のテキストのおおよその採用率を表わすものだが、計算に用いられる「総頁数」の値には、書籍の途中で挟まれる空白ページやイラストや章タイトルなどのページなども含むため、作品の総語数中の採用率ではなく、あくまでページの採用率である。参考程度に考えてほしい。

表 1 SHC ベストセラー書籍の収録作品と収録範囲の一覧

収録年	作品名	著者	出版社	版	収録範囲	総頁数	採用率
1933	海豹の如く	賀川豊彦	大日本雄弁会 講談社	(第 1 刷)	282-335	440	12.3
1933	女の一生	山本有三	中央公論社	十五版	327-386	779	7.7
1933	結婚街道	菊池寛	中央公論社	(第 1 刷)	81-136	380	15.0

収録年	作品名	著者	出版社	版	収録範囲	総頁数	採用率
1933	春琴抄	谷崎潤一郎	創元社	(第1刷・復刻版)	1-108	338	32.0
1933	青春物語	谷崎潤一郎	中央公論社	(第1刷)	125-186	232	26.7
1933	日本海々戦	ノウイコフ・ブリボイ/上脇進訳	改造社	(第1刷)	373-430	526	11.0
1933	女人曼陀羅	吉川英治	中央公論社	(第1刷)	658-730	772	9.5
1933	野口英世	奥村鶴吉	岩波書店	第一刷	78-115	657	5.8
1933	百鬼園随筆	内田百閒	三笠書房	第一刷特製版	3-63	350	17.4
1933	武蔵野にをりて	吉田紘二郎	改造社	(第1刷)	159-248	248	36.3
1933	唯物弁証法読本	大森義太郎	中央公論社	八十五版	182-244	301	20.9
1941	Uボートは西へ	小沢覚輔	報知新聞出版部	(第1刷)	38-74	346	10.7
1941	雲南守備兵	木村莊十	博文館	(第1刷)	60-91	297	10.8
1941	運命の人	島木健作	新潮社	(第1刷)	179-205	300	9.0
1941	海狼Uボート	福田一郎	海と空社	(第1刷)	52-89	426	8.9
1941	軍事郵便	河内仙介	新潮社	(第1刷)	3-58	283	19.8
1941	子供記	鈴木藤太郎	甲鳥書林	(第1刷)	291-342	342	15.2
1941	次郎物語	下村湖人	小山書店	第七冊	169-229	370	16.5
1941	新書太閤記	吉川英治	新潮社	十刷	195-228	450	7.6
1941	人生論ノート	三木清	創元社	第三刷	157-209	207	25.6
1941	第五列の兵器庫	海野十三	ラヂオ科学社	(第1刷)	84-113	513	5.8
1941	大砲の秋	桜井忠温	新紀元社	(第1刷)	218-259	358	11.7
1941	たぬき汁	佐藤垢石	墨水書房	五版	1-65	452	14.4
1941	智慧子抄	高村光太郎	竜星閣	(第1刷)	5-104	147	68.0
1941	長江デルタ	多田裕計	文芸春秋社	(第1刷)	65-110	296	15.5
1941	哲学ノート	三木清	河出書房	(第1刷)	295-329	356	9.8
1941	得能五郎の生活と意見	伊藤整	河出書房	(第1刷)	239-291	389	13.6
1941	土地なき民	ハンス・グリム/ 星野慎一訳	鱒書房	八十版	335-356	450	4.9
1941	ノロ高地	草葉栄	鱒書房	第四百五十版	89-126	301	12.6
1941	梅里先生行状記	吉川英治	朝日新聞社	(第1刷)	44-124	312	25.3
1941	フランス戦線	アンドレ・モー ロア/高野弥一郎	大観堂	第四刷	231-284	351	15.4
1941	ホロンバイルの荒鷲	入江徳郎	鱒書房	(第1刷)	123-170	275	17.5
1941	香港の機密室	市木亮訳	越後屋書房	(第1刷)	59-83	213	11.7
1941	野戦病院	大岳康子	主婦之友社	(第1刷)	231-271	269	15.2
1941	ルーマニア日記	ハンス・カロッ サ/高橋健二訳	三笠書房	(第1刷)	162-187	161	16.1
1941	路傍の石	山本有三	岩波書店	(第1刷・復刻版)	430-485	587	9.5
1949	アダノの鐘	ハーシー/杉本喬訳	東西出版社	(第1刷)	384-419	416	8.7
1949	石中先生行状記	石坂洋次郎	新潮社	(第1刷)	227-250	328	7.3
1949	イムパール	高木俊郎	雄鷄社	再版	141-165	370	6.8
1949	浮雲	林芙美子	六興出版	五版	179-213	453	7.7

収録年	作品名	著者	出版社	版	収録範囲	総頁数	採用率
1949	ウラルを越えて	樋口欽一他	乾元社	(第1刷)	35-69	299	11.7
1949	風と共に去りぬ	M.ミッチェル/大久保康雄訳	三笠書房	(第1刷)	2-25	315	7.6
1949	仮面の告白	三島由紀夫	河出書房	三版	238-279	277	15.2
1949	きけわだつみの声	日本戦没学生手記編集委員会	東京大学出版部	(第1刷)	13-33	36	58.3
1949	共産主義批判の常識	小泉信三	新潮社	三版	95-130	165	21.8
1949	「共産党宣言」批判	蠟山政道	明治書院	(第1刷)	107-151	161	27.3
1949	郷土	パール・バック/石川欽一訳	毎日新聞社	三版	351-375	577	4.2
1949	この子を残して	永井隆	大日本雄弁講談社	(第1刷)	196-239	237	18.6
1949	細雪	谷崎潤一郎	中央公論社	三版	147-185	393	9.9
1949	自由と規律—イギリスの学校と生活—	池田潔	岩波書店	第九刷	126-166	166	24.7
1949	生命の河—原子病の話—	永井隆	日比谷出版社	(第1刷)	228-267	302	13.2
1949	滞日十年 日記・公文書・私文書に基く記録	ジョセフ・C・グルー/石川欽一訳	毎日新聞社	(第1刷)	193-214	372	5.9
1949	哲学入門 哲学の根本問題	田辺元	筑摩書房	(第1刷)	176-239	239	26.8
1949	読書雑記	小泉信三	文芸春秋新社	(第1刷)	229-259	315	9.8
1949	流れる星は生きている	藤原てい	日比谷出版社	(第1刷)	230-267	300	12.7
1949	ヒロシマ	ハーシー/石川欽一訳	法政大学出版局	(第1刷)	27-67	145	28.3
1949	俘虜記	大岡昇平	創元社	三版	69-91	212	10.8
1949	文化服装講座婦人服	野口益栄	文化服装学院出版局	58版	123-182	335	14.6
1949	平和の発見—巢鴨の生と死の記録	花山信勝	朝日新聞社	(第1刷)	240-266	319	8.5
1949	宮本武蔵	吉川英治	六興出版	(第1刷)	39-96	383	15.1
1949	私は今もイエスを追ふ	赤岩栄	角川書店	初版	81-136	280	20.0
1949	われ敗れたり	児玉誉士夫	東京出版社	再版	138-167	296	10.1
1957	愛のかたみ	田宮虎彦/田宮千代	光文社	初版	160-196	285	13.0
1957	一日一言—人類の知恵—	桑原武夫	岩波書店	第5刷	187-216	216	13.9
1957	岩波茂雄伝	安部能成	岩波書店	第二刷	403-442	502	8.0
1957	飢える魂	丹羽文雄	角川書店	第1刷	417-438	432	5.1
1957	駅前旅館	井伏鱒二	新潮社	(第1刷)	123-167	221	20.4
1957	厭世立志伝	尾崎士郎	中央公論社	(第1刷)	195-223	281	10.3
1957	お吟さま	今東光	淡交社	第一刷	35-82	198	24.2
1957	おとうと	幸田文	中央公論社	三刷	144-176	225	14.7
1957	おはん	宇野千代	中央公論社	再版	131-190	188	31.9
1957	女坂	円地文子	角川書店	初版	95-119	186	13.4
1957	体の中を風が吹く	佐多稲子	大日本雄弁会講談社	第1刷	140-186	295	15.9
1957	現代史 日本の百年	林茂	毎日新聞社	初版	103-124	278	7.9
1957	砂漠の花	平林たい子	光文社	初版	221-250	258	11.6
1957	しいのみ学園	山本三郎	福村出版	第16刷	228-250	249	9.2

収録年	作品名	著者	出版社	版	収録範囲	総頁数	採用率
1957	射程	井上靖	新潮社	(第1刷)	85-122	247	15.4
1957	昭和時代	中島健蔵	岩波書店	第4刷	81-107	200	13.5
1957	過ぎゆきの歌	上林暁	大日本雄弁会 講談社	第一刷	3-38	271	13.3
1957	世界の酒	坂口謹一郎	岩波書店	第3刷	148-197	197	25.4
1957	創価学会 その思想と 行動	佐木秋夫/小口偉 一	青木書店	初版	149-176	220	12.7
1957	第三の眼 秘境チベッ トに生まれて	ロブサン・ラン パ/今井幸彦訳	光文社	五版	214-243	228	13.2
1957	探美の夜	中河与一	大日本雄弁会 講談社	第1刷	199-222	331	7.3
1957	地球の歴史	湊正雄/井尻正二	岩波書店	第2刷	120-159	211	19.0
1957	檜山節考	深沢七郎	中央公論社	五版	1-40	210	19.0
1957	日目の背信	丹羽文雄	角川書店	初版	3-28	319	8.2
1957	日本語	金田一春彦	岩波書店	第1刷	134-167	224	15.2
1957	人間革命	妙悟空	精文堂	(第1刷)	422-458	464	8.0
1957	暖簾	山崎豊子	東京創元社	二十二版	3-37	218	16.1
1957	白頭吟	石川淳	中央公論社	(第1刷)	172-199	211	13.3
1957	花の奴隸	藤野登久子	中央公論社	三版	24-60	150	24.7
1957	ビジネスマンの健康—若々 しく働くための医学—	杉靖三郎	実業之日本社	(第1刷)	66-122	297	19.2
1957	非情の庭	樋口茂子	三一書房	三版	33-65	242	13.6
1957	美德のよろめき	三島由紀夫	大日本雄弁会 講談社	第六刷	70-105	203	17.7
1957	氷壁	井上靖	新潮社	二刷	69-95	323	8.4
1957	夫婦百景	獅子文六	主婦之友社	二刷	98-140	203	21.2
1957	法律相談	川島武宜ほか編	岩波書店	第三刷	66-112	186	24.7
1957	魔の河	火野葦平	光文社	再版	3-32	200	15.0
1957	みみずく説法	今東光	光文社	初版	124-179	208	26.9
1957	雌花	大岡昇平	新潮社	(第1刷)	77-120	198	22.2
1957	妖	円地文子	文芸春秋新社	(第1刷)	7-38	283	11.7
1957	ロンドン—東京5万キ ロ 国産車ドライブ記	辻豊/土崎一	朝日新聞社	第二十四刷	88-120	303	10.9
1957	われら不条理の子	ボッシュ/ 加藤周一訳	紀伊国屋書店	第3刷	40-73	144	23.6
1965	思い出の回転レシーブ	河西昌枝	講談社	第11刷	13-76	236	27.1
1965	科学と宗教	池田大作	鳳書院	4版	238-294	479	11.9
1965	危険な思想家 戦後民主 主義を否定する人びと	山田宗睦	光文社	22版	96-145	188	26.6
1965	原爆体験記	広島市原爆体験 記刊行会	朝日新聞社	第8刷	96-156	247	24.7
1965	憲法読本	憲法問題研究会	岩波書店	第2刷	1-48	211	23.2
1965	山峡の章	松本清張	光文社	初版	48-101	258	20.9
1965	白い巨塔	山崎豊子	新潮社	五十四刷	329-358	523	5.7
1965	戦争と人間(1)	五味川純平	三一書房	初版第11刷	175-214	269	14.9

収録年	作品名	著者	出版社	版	収録範囲	総頁数	採用率
1965	対話 人間の建設	岡潔/小林秀雄	新潮社	三刷	33-124	168	54.8
1965	妻の日の愛のかたみに	池上三重子	産経新聞社	36版	61-152	193	47.7
1965	徳川家康(23)	山岡荘八	講談社	第1刷	82-134	253	20.9
1965	泥と炎のインドシナ	大森実監	毎日新聞社	(第1刷)	61-103	254	16.9
1965	日本の歴史	直木考次郎編	中央公論社	再版	284-344	519	11.8
1965	人間革命(1)	池田大作	聖教新聞社	改訂6刷	255-317	315	20.0
1965	バランスシート 経営者の虚々実々を見破る本	後藤弘	光文社	初版	145-216	206	22.8
1965	氷点	三浦綾子	朝日新聞社	第5刷	22-54	415	8.0
1965	ヒロシマ・ノート	大江健三郎	岩波書店	第30刷	15-67	186	28.5
1965	ベトナム戦記	開高健	朝日新聞社	第1刷	149-226	226	34.5
1965	南ベトナム戦争従軍記	岡村昭彦	岩波書店	第16刷	144-184	221	19.0
1965	山本五十六	阿川弘之	新潮社	十三刷	164-197	325	10.5
1965	わが愛を星に祈りて 女子高校生の日記	佐伯浩子	大和書房	初版	151-239	221	40.3
1973	育児芸術	御木徳近	講談社	第三刷	105-136	270	11.9
1973	狼なんかこわくない	庄司薫	中央公論社	(第1刷)	38-92	214	25.7
1973	夫につき合う秘密集 頭のいい妻になる224項目	池坊保子	青春出版社	第81刷	197-244	224	21.4
1973	怪物商法 常識をうち破る	糸山英太郎	ベストセラーズ	40版	45-95	233	21.9
1973	華麗なる一族	山崎豊子	新潮社	六刷	270-301	309	10.4
1973	悔いなき命を	岡田嘉子	広済堂出版	六刷	101-162	246	25.2
1973	ぐうたら愛情学	遠藤周作	講談社	第28刷	159-220	253	24.5
1973	死海のほとり	遠藤周作	新潮社	(第1刷)	169-229	321	19.0
1973	自己トレーニング モヤシ男を世界一の体にした驚異の秘法	野沢秀雄	青春出版社	第41刷	72-166	259	18.1
1973	ジャッカルの日	フレデリック・フォーサイス/ 篠原慎訳	角川書店	初版	121-151	429	7.5
1973	商売心得帖	松下幸之助	PHP研究所	記載なし	6-74	109	63.3
1973	砂の器	松本清張	光文社	初版152刷	429-478	474	10.5
1973	スプーン一杯の幸せ 愛を語る三つの形	落合恵子	祥伝社	11版	164-230	224	29.9
1973	太陽への挑戦 糸山英太郎のわが闘争序章	糸山英太郎	双葉社	第33版	114-164	261	19.5
1973	どんと来い税務署 ひどい目にあわないために	吉田敏幸	ベストセラーズ	改訂初版	185-233	253	19.4
1973	日本沈没	小松左京	光文社	364版	129-163	260	17.3
1973	人間革命(8)	池田大作	聖教出版社	第11刷	123-185	295	21.4
1973	にんにく健康法 なぜ効く, 何に効く, どう食べる	渡辺正	光文社	154版	88-152	208	31.3
1973	ノストラダムスの大予言	五島勉	祥伝社	220刷	67-121	222	24.8
1973	箱男	安部公房	新潮社	第一刷	75-119	188	23.9
1973	橋のない川(6)	住井すゑ	新潮社	(第1刷)	162-194	270	12.2

収録年	作品名	著者	出版社	版	収録範囲	総頁数	採用率
1973	ひとりでも生きられる いのちを愛にかけ ようとするとき	瀬戸内晴美	青春出版社	第六十四刷	17-68	214	24.3
1973	マンボウぼうえんきょう	北杜夫	新潮社	五刷	77-120	215	20.5
1981	この愛いつまでも	加山雄三	光文社	初版4刷	13-161	189	78.8
1981	叱り方の上手い親下手な親	田中澄江	青春出版社	記載なし	63-223	222	72.5
1981	白ゆりの詩	創価学会婦人部編	聖教新聞社	第1刷	131-259	351	36.8
1981	新・頭のいい税金の本	野末陳平	青春出版社	記載なし	15-169	231	67.1
1981	なんとなくクリスタル	田中康夫	河出書房新社	五六版	5-148	144	100.0
1981	人間万事塞翁が丙午	青島幸男	新潮社	七刷	51-140	224	40.2
1981	ノストラダムスの大予言	五島勉	祥伝社	初版第1刷	67-215	205	72.7
1981	窓際のトットちゃん	黒柳徹子	講談社	第105刷	9-176	276	60.9
1989	キッチン	吉本ばなな	福武書店	第三八刷	7-70	220	29.1
1989	下天は夢か(1)	津本陽	日本経済新聞社	一刷	309-366	362	16.0
1989	孔子	井上靖	新潮社	三刷	91-156	411	16.1
1989	時間の砂(上)	シドニイ・シェ ルダン	アカデミー出 版サービス	第六刷	206-311	317	33.4
1989	消費税こうやればいい	山本雄二郎	青春出版社	記載なし	105-214	206	53.4
1989	消費税 実務と対策は こうする	山本守之	日本実業出版社	初版第8刷	179-261	251	33.1
1989	TUGUMI	吉本ばなな	中央公論社	5版	7-119	228	49.6
1989	「NO」と言える日本	盛田昭夫・ 石原慎太郎	光文社	初版1刷	74-160	152	57.2
1989	ノルウェイの森(上)	村上春樹	講談社	第一六刷	48-154	263	40.7
1989	人麻呂の暗号	藤村由加	新潮社	五刷	144-179	253	14.2
1989	ホーキング、宇宙を語る	S.W.ホーキング	早川書房	26版	92-154	228	27.6
1989	三毛猫ホームズの プリマドンナ	赤川次郎	光文社	18版	121-249	245	52.7
1989	ラッフルズホテル	村上龍	集英社	第九刷	74-173	181	55.2
1997	頭がいい人、悪い人 の話し方	樋口裕一	PHP研究所	第一版第一 二七版	153-217	203	32.0
1997	永遠の法	大川隆法	幸福の科学出版	初版第95刷	73-289	277	78.3
1997	失樂園(上)	渡辺淳一	講談社	第1刷	7-145	300	45.7
1997	少年H(上)	妹尾河童	講談社	第五刷	244-355	350	32.0
1997	7つの習慣	スティーブン・ R・コヴィー	キング・ ベアー出版	初版第62刷	75-203	481	26.8
1997	母の詩	池田大作	聖教新聞社	第九刷	78-270	263	73.4
1997	鉄道員(ぽっぽや)	浅田次郎	集英社	第一五刷	161-284	278	44.6
1997	まる子だった	さくらももこ	集英社	第一刷	63-206	185	77.8
1997	ももこの世界あっち こっちめぐり	さくらももこ	集英社	第一二刷	53-226	222	78.4
1997	ユートピア創造論	大川隆法	幸福の科学出版	初版第10刷	9-146	138	100.0
2005	生きて死ぬ智慧	柳澤桂子	小学館	初版第十四 刷	1-31	31	100.0
2005	香峯子抄	池田香峯子	主婦の友社	第三十刷	11-142	212	62.3

収録年	作品名	著者	出版社	版	収録範囲	総頁数	採用率
2005	恋バナ(赤)	Yoshi	スターツ出版	第1刷	29-151	147	83.7
2005	これだけは知っておきたい個人情報保護	岡村久道/鈴木正朝	日本経済新聞社	1版1刷	5-58	54	100.0
2005	さおだけ屋はなぜ潰れないのか?	山田真哉	光文社	初版53刷	19-126	190	56.8
2005	新・人間革命(14)	池田大作	聖教新聞社	第11刷	294-387	381	24.7
2005	神秘の法	大川隆法	幸福の科学出版	初版第41刷	15-109	319	29.8
2005	生協の白石さん	白石昌則/東京農工大学の学生の皆さん	講談社	第10刷	45-148	143	72.7
2005	ダ・ヴィンチ・コード(上)	ダン・ブラウン	角川書店	六版	7-83	328	23.5
2005	東京タワー オカンとボクと、時々、オトン	リリーフランキー	扶桑社	初版第19刷	200-319	447	26.8
2005	野ブタ。をプロデュース	白岩玄	河出書房新社	112刷	3-124	184	66.3
2005	ハッピーバースデー	青木和雄/吉富多美	金の星社	初版第30刷	7-110	250	41.6
2005	「もっと、生きたい...」	Yoshi	スターツ出版	第1刷	5-87	204	40.7
2005	続弾! 問題な日本語	北原保雄編	大修館書店	初版第1刷	39-108	264	26.5
2005	ワルの知恵本	門昌央と人生の達人研究会編	河出書房新社	14刷	189-254	244	27.0
2013	者に殺されない47の心	近藤誠	アスコム	第40刷	101-229	209	61.7
2013	置かれた場所で咲きなさい	渡辺和子	幻冬舎	第76刷	9-157	149	100.0
2013	海賊と呼ばれた男(上)	百田尚樹	講談社	第22刷	294-354	372	16.4
2013	聞く力 心をひらく35のヒント	阿川佐和子	文藝春秋	初版第60刷	184-253	237	29.5
2013	雑談力が上がる話し方 30秒でうちとける会話のルール	齋藤孝	ダイヤモンド社	第27刷	35-135	168	60.1
2013	色彩を持たない多崎つくると、彼の巡礼の年	村上春樹	文藝春秋	第2刷	263-310	368	13.0
2013	新・人間革命(25)	池田大作	聖教新聞社	第6刷	105-206	392	26.0
2013	スタンフォードの自分を変える教室	ケリー・マクゴニガル	大和書房	初版第20刷	202-270	325	21.2
2013	伝え方が9割	佐々木圭一	ダイヤモンド社	初版第36刷	51-203	183	83.6
2013	できる大人のモノの言い方大全	話題の達人倶楽部編	青春出版社	初版第48刷	263-381	365	32.6
2013	謎解きはディナーのあとで(3)	東川篤哉	小学館	初版第2刷	181-269	265	33.6
2013	人間にとって成熟とは何か	曾野綾子	幻冬舎	第十五刷	73-151	230	34.3
2013	ホテルローヤル	桜木紫乃	集英社	第8刷	57-138	188	43.6
2013	野心のすすめ	林真理子	講談社	第六刷	119-191	177	41.2
2013	未来の法 新たなる地球世紀へ	大川隆法	幸福の科学出版	第39刷	125-233	219	49.8
2013	ロスジェネの逆襲	池井戸潤	ダイヤモンド社	第22刷	132-218	378	23.0



### 3. サンプル ID

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、各作品の最も細かい構成要素（記事相当）に分割した、その各文書要素である。各サンプルを一意に特定する ID の構成を表 2 にあげる。

表 2 サンプル ID の構成

桁数	値	説明
1-2	70または80	時代区分を表わす。1933年から1981年までの収録作品には昭和の「70」が付与され、1989年から2013年までの収録作品には平成の「80」が付与される。※1
3	B	サブコーパスのジャンル（ベストセラー書籍、Best sellerのB）を表わす。全サンプルで共通。
4-5	作品ID	各書籍の頭文字2字を表わす。※2
6-9	(4桁の数字)	サンプルの発行年を西暦で表わす。
10	_	サンプルIDの区切り記号（アンダーバー）。
11	(1桁の数字)	章節項の章など、3層以上の構造を持つ書籍の最も大きい層の通し番号を表わす。2層以下の書籍の場合、「0」が付与される。
12-13	(2桁の数字)	章節項の節など、2層以上の構造を持つ書籍の2番目に小さな構成要素の通し番号を表わす。単一の層しか持たない書籍の場合、「00」が付与される。
14-15	(2桁の数字)	章節項の項など、書籍の最小の構成要素の通し番号を表わす。

※1 昭和と平成の境目の1989年は、昭和の期間がわずか7日であったことや、データの利用しやすさ・わかりやすさを重視して一律で平成の「80」を付与している。

※2 2005年「もっと、生きたい…」のように、記号類を含むタイトルの場合は記号類を抜いた頭文字を付与している。1949年の『共産主義批判の常識』と『「共産党宣言」批判』のように、同一年で頭文字が重なる場合には、それぞれ「常識」「宣言」と、区別できるキーワードを表わした。『妖』のように単漢字の作品については、「妖\_」というように、「\_」を付加している。

表 2 の基準によると、例えば、1973 年の『ぐうたら愛情学』の第 3 章第 2 節第 1 項のサンプル ID は「70B ぐう 1973\_30201」に、2005 年『新・人間革命』（14 巻）の第 4 章第 5 節サンプル ID は「80B 新人 2005\_00405」になる。

### 4. テキスト

#### 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。

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。

文字集合に含まれない漢字については、以下の（1）～（4）の手順で電子化した（須永・堤・高田

2011、須永ほか 2013)。

- (1) JIS X 0213 の「6.6.3 漢字の字体の包摂規準」に若干の拡張を施した近代語コーパス用の包摂規準に基づいて、JIS 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は、類似の意味・用法を持つ同音・同訓の JIS 内の文字で代用する。
- (3) JIS X 0213 中の「Unicode における CJK 統合漢字拡張 B」(サロゲートペアの文字) を文字集合に含めるほか、コーパス文字集合外の Unicode 文字に同一字体があればそれで入力する。
- (4) (1) ～ (3) の手順で入力できない場合は、外字として「=」(げた記号、JIS 面区点 1-02-14、U+3013) で表す。

#### 4. 2 テキストの校訂

本コーパスでは、『日本語歴史コーパス』の他のサブコーパスや『現代日本語書き言葉均衡コーパス』等の国立国語研究所構築の他のコーパスと齊一な形態論情報を付与するため、形態素解析辞書 UniDic を使用した形態素解析に基づき形態論情報を付与した。そのため、コーパスのテキストを 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) と B) にあげる改変(ここでは「校訂」と呼ぶ)を施し、コーパスのテキストを作成した。

なお、「中納言」では、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。利用に際しては、必要に応じて「原文 KWIC」「原本文字列」や底本画像を確認されたい。

##### A) 踊り字

踊り字は繰り返される文字列に置き換える。ただし、「国々」「人々」等、1 短単位内部で直前の 1 字を繰り返す「々」「々」は置き換えの対象としない。

表 3 踊り字の電子化例

種類	コーパステキスト	原文 KWIC・原本文字列
/ \	だ <u>だ</u> ん衰弱してゆく母	だ <u>ん</u> / \ 衰弱してゆく母
々	あ <u>あ</u> 、然し何といふ幸福!	あ <u>と</u> 、何といふ幸福!

##### B) 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。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などは、訂正の対象としない。

表 4 誤植の電子化例

種類	コーパステキスト	原文 KWIC・原本文字列
脱字	フラマン君、まあその椅子にか <u>け</u> またへ	フラマン君、まあその椅子にかまたへ
誤字	そして、三面の <u>劈</u> 頭から四段ぶつ通しの 大みだしで	そして、三面の <u>剪</u> 頭から四段ぶつ通しの大み だしで
衍字	これには官軍もタジタジとなつた。	これには官軍も <u>も</u> タジノ＼となつた。

## 5. 形態論情報

本コーパスでは、原則として底本の本文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。

形態論情報は短単位のみ付与しており、長単位は未実装である。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、人手により修正することで付与した。

本コーパスは全編が、人手修正が入っているが部分的に形態素解析の結果のままを残している「非コアデータ」である。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。2023年5月公開分の SHC ベストセラー書籍サブコーパスの形態論情報の精度（適合率）は、非コアデータが約 98.1%となっている<sup>1</sup>。

## 6. 「中納言」上の表示項目

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 1）。

276 件の検索結果が見つかりました。  
検索対象語数: 38,569,406 記号・補助記号・空白を除外した検索対象語数: 33,866,692 検索対象サンプル数: 21,000

サブコーパス名	サンプルID	開始位置	連番	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	ジャンル	作品名	成立年	巻名等	作者	生年	底本	ページ番号	
昭和-平成-新聞	Y19330902_004	3160	2070	てるん「です」 ぜ。」# 狼山」 はかう云ひな がら、テエプ ルの上」 「……」 しげ」	九レ	「お大」スプーン」に 杯「すくひあ」 ひはじめた。# 「 ね、先生」 せう?」	九レ	九レ	九レ	名詞- 普通名 詞-一 般			九レ					文芸	読売新 聞	1933	銀座殘 巻 記			読売新聞 <1933- 09-02-第 20313号 >	
昭和-平成-雑誌	80M中公 1941_02026	20990	14250	といふん」 で、 「牛肉」 「あづき」 は、 「ライス」	九レ	「おと」いつた に飲します。# それには「私」 役だといふん	九レ	九レ	九レ	名詞- 普通名 詞-一 般			九レ					非文 芸	中央公 論	1941	併編対 談 2 長 十郎- 國太 郎の 巻	三宅周 太郎 (作)/ 河原崎 國太郎 (作)/ 河原崎 長十 郎(作)		中央公論 <1941-2>	本欄 230
昭和-平成-ベストセラー書籍	70B浮雲 1949_00028	20260	14090	は「ま」く 出す「事」 が「出 来ま」せ ん」。# 「み」す 「み」す 「い」 「客」が 「い」 つて「も」 「ライ ス」	九レ	「一」つ 「出」せ 「い」 「ん」 「に」 「す」 「ら」 「ね」 。# 「 「一」 「向」 「し」 「る」 「密」 「告」 「が」 「や」 「か」 「ま」 「し」 「て」 「い」 「あ」 「ぶ」 「く」 「て」	九レ	九レ	九レ	名詞- 普通名 詞-一 般			九レ		会話	亭主	文芸	浮雲	1949	二十八		林芙美 子 (作)	1903	浮雲	195

図 1 「中納言」の検索結果のイメージ

<sup>1</sup> ここでいう精度（適合率）は、（調査対象とした）整備済みコーパスの語数で、そのうちの正解語数を除いた値である。語形、活用型、活用形のみ誤りも含む。

「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、本コーパスで特に注意が必要な項目である「文体」「出版社」（表中に「\*」を付す）について、表5に内容を示す。

表5 「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	検索対象の含まれる短単位の先頭の文字の、サンプル内における位置を表す ID。10 きざみの連番。
	連番	検索対象の含まれる短単位の、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非コアデータであることを表す。「0」が非コアを表す。
形態論情報	前文脈	検索対象の前方文脈。
	キー	検索対象の含まれる短単位の書字形出現形（表記形）。
	後文脈	検索対象の後方文脈。
	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）。
	語彙素	検索対象の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	検索対象の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。片仮名表記である。
	品詞	<p>検索対象の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>このほか、本コーパスに含まれる、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>漢文 ...漢文の文字列。</p> <p>外国語 ...外国語の文字列。</p> <p>絵文字・記号等 ...入力のできない絵文字や企業マークなど。コーパステキストでは「=」で表示される。</p> <p>未知語 ...形態論情報の付与を保留した文字列。</p>

情報種別	項目名	内容
形態論情報	活用型	検索対象の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。
	活用形	検索対象の含まれる短単位の活用形。活用語の場合のみ表示される。
	原文文字列	検索対象の含まれる短単位の、校訂前の底本に近い形のテキスト(4.2 節参照)。
	振り仮名	検索対象の含まれる短単位に付された振り仮名(右ルビ)の文字列。
	本文種別	検索対象の含まれる文が「地の文」以外の場合の、その種別。以下の種類がある。なお、地の文の場合、当項目は空白となる。 会話 ...会話・独話・心内発話等の引用 引用 ...文献等からの引用
	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」の場合の話者名や属性名(男、先生など)を表す。不明の場合は「*」で示す。
本文情報	文体*	検索対象の含まれる文の文体。以下の種類がある。 文語...文語体。文末辞が「なり」「たり」「つ」「ぬ」「き」「けり」のもの。 口語...口語体。文末辞が「だ」「ぢや」「である」「です」「ます」のもの。 韻文... 詩歌の文体。 漢文...漢文。 外国語...漢文以外の外国語の文。 なお、一つのサンプル内で複数の文体が混在しているものは、地の文および引用範囲ごとにそれぞれ文体を付与する。
	ジャンル	図書館分類法の日本十進分類法に基づく分類。「9」から始めるや詩歌には「文芸」、それ以外のサンプルには「非文芸」が表示される。
	作品名	検索対象の含まれるサンプルが収録された書籍名。
	成立年	検索対象の含まれるサンプルが収録された年。ただし、ベストセラー書籍は前年の11月下旬から次の年の11月下旬までを対象とするため、1988年刊行の『TSUGUMI』が1989年のベストセラーに収録されるケースがある。こういった場合でも、データの利用のしやすさの観点から、一律でコーパスの収録年に合わせて年を付与している。
	巻名等	検索対象の含まれるサンプル(章や節)のタイトル。本コーパスでは記事タイトルを表示する(記事タイトルのないものは無表示である)。

情報種別	項目名	内容
作者情報	作者	書籍の著者名。著者名の認定は、底本テキストの記載に基づく。翻訳作品の場合は「ジョセフ・クラーク・グルー(作)/石川欣一(訳)」のように併記する。 「国立国会図書館典拠データ検索・提供サービス (Web NDL Authorities)」のウェブページでの著者情報へのリンクを付与している。
	生年	書籍の著者の生年。西暦 4 桁で示す。
底本情報	底本	検索対象の底本 (原資料)。奥付にある書籍名 (副題を含む) に準拠して示す。
	ページ番号	検索対象の底本におけるページ番号。
	出版社*	底本の出版社。
その他	底本リンク	検索対象の底本画像へのリンク。本コーパスでは該当画像がないため空欄である。
	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である。

## 付記

本レジスターを含む『昭和・平成書き言葉コーパス』は、JSPS 科研費 19H00531 「昭和・平成書き言葉コーパスによる近現代日本語の実証的研究」(代表: 小木曾智信) の研究成果を報告したものである。

## 参考文献

国立国語研究所国語辞典編集準備室 (1984) 『用例採集のためのベストセラー目録』 国語辞典編集準備資料 4

出版ニュース社出版年鑑編集部 『出版年鑑』 出版ニュース社

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』 pp.381-388.

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.

トーハン「年間ベストセラーアーカイブ」 <https://www.tohan.jp/bestsellers/past.html>

## 関連 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/>

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/>

## 資料

## ■不採用作品の一覧とその事由

収録年	書籍名	著者	事由
1965	医学カード	小倉清太郎	図を中心とするため
1965	性生活解剖図譜	松戸尚	図を中心とするため
1965	世界原色百科事典(1)	小学館	図を中心とするため
1965	われらすべて勝者 東京オリンピック写真集	安川第五郎監	写真を中心とするため
1981	神戸ポートアイランド博覧会 公式ガイドブック・マップ	神戸ポートアイランド 博覧会協会	図を中心とするため
1981	アクション・カメラ術	馬場憲治	写真を中心とするため
1989	ドラゴンクエストIII 公式ガイドブック	エニックス	図を中心とするため
1989	宮沢りえ写真集Pour Amitie	野村誠一撮影	写真を中心とするため
1997	ビストロスマップ完全レシピ	ビストロスマップ 制作委員会	写真を中心とするため
1997	ファイナルファンタジーVII 解体新書	スタジオイベント スタッフ	図を中心とするため
1997	たまごっち大百科	STUDIO HARO TEAM3	図を中心とするため
1997	ダービースタリオン公式 ガイドブック	JK・Voice	図を中心とするため
1997	ポケットモンスター大百科 赤・緑・青	エニックス	図を中心とするため
1997	たまごっちR 母子手帳	Wink up	図を中心とするため
1997	ファイナルファンタジーVII 公式設定資料集	ファミ通書籍 編集部	図を中心とするため
1997	菅野美穂写真集「NUDITY」	宮澤正明撮影	写真を中心とするため
1997	ポケットモンスター ミニミニ大事典	ティーツー出版	図を中心とするため
1997	金田一少年の事件簿(4)	天樹征丸	絵を中心とするため
1997	ファイナルファンタジー タクティクス大全	JK・Voice	図を中心とするため

収録年	書籍名	著者	事由
2005	震災時 帰宅支援マップ		図を中心とするため
2005	ダーリンの頭ン中	小栗左多里/ トニー・ラズロ	絵を中心とするため
2005	かいけつゾロリたべられる！！ かいけつゾロリの大どろぼう	原ゆたか	絵を中心とするため
2013	とびだせ どうぶつの森 ザ・コンプリートガイド		図を中心とするため
2013	世界のなめこ図鑑	Beeworks	図を中心とするため
2013	とびだせ どうぶつの森 超完全カタログ	NintendoDREAM 編集部	図を中心とするため
1949	解析精義	岩切晴二	数式を中心とする
1989	究極の選択	オールナイト・ ニッポン編	特殊文字や図を中心とする 部分が多いため
2005	電車男	中野独人	特殊文字を中心とするため
1949	戦争と平和	トルストイ/ 米川正夫訳	1925年の初出のため
1949	伸子	宮本百合子	1924年の初出のため
1949	北郊雑記	阿部次郎	1927年の初出のため
1949	三木清著作集	大内兵衛他	全集作品のため
1957	谷崎潤一郎全集		全集作品のため
1957	杜甫	吉川幸次郎/ 小川環樹編	古典作品のため
1957	平家物語	石母田正	古典作品のため
1957	李白	吉川幸次郎ほか編	古典作品のため
1965	勝海舟(6)	子母澤寛	全集作品のため
1965	源氏物語	紫式部/ 与謝野晶子訳	古典作品のため
1973	勝海舟	子母澤寛	全集作品のため
1973	日本の歴史	直木考次郎	1965年中央公論社版の採用 作品と同一内容のため
1941	続哲学ノート	三木清	同著者の『人生論ノート』 『哲学ノート』を採用
1949	長崎の鐘	永井隆	同著者の『この子を残し
1949	ロザリオの鎖	永井隆	て』『生命の河』を採用



収録年	書籍名	著者	事由
1957	天平の藪	井上靖	同著者の『射程』『氷壁』を採用
1973	ぐうたら交友録	遠藤周作	同著者の『ぐうたら愛情学』『死海のほとり』を採用
1989	白河夜船	吉本ばなな	同著者の『TUGUMI』『キッチン』を採用
1989	うたかた/サンクチュアリ	吉本ばなな	
1989	哀しい予感	吉本ばなな	
1989	パイナップリン	吉本ばなな	