

# p1-5 『現代日本語書き言葉均衡コーパス』 書籍サブコーパスの小説サンプルに対するジャンル情報付与

加藤祥 (目白大学) ・ 浅原正幸 (国語研・東外大)

## 【作業手順】 ※計4名の作業者による

1. NDC9X3の付与される各サンプル書籍について、書店/出版社のWebページから分類情報を収集  
ノベルス・日本文学・SF・ポケットミステリ.....
2. 書籍の概要説明などから、書籍内容に関わる情報を収集  
伝奇・恋愛・冒険・医療・経済・幻想・幕末.....
3. 付与された情報を整理  
※情報なしサンプルあり  
メインジャンル・時代以外は1サンプル複数ラベル (マルチラベル)

## 【書店一般の分類情報例 ※1サンプル複数ラベル】

	LB	OB	PB	総計
推理・ミステリー	673	112	301	1086
歴史・時代	459	159	282	900
ラノベ・少女	279	18	181	478
文学・文芸	135	41	72	248
SF	104	20	79	203
児童・子ども	44	15	84	143
ホラー・怪談	100	8	32	140

## 【小説でない (情報付与対象外) サンプルの内訳】

	LB	OB	PB	総計
小説でない (情報付与対象外)	64(2.4%)	4(0.6%)	38(2.1%)	106(2.1%)
エッセイ	8	2	2	12
ノンフィクション・実録	16	0	1	17
ルポルタージュ・記録	3	0	4	7
解説	9	0	13	22
戯曲	2	0	0	2
体験記・日記	6	2	1	9
名言・ジョーク	1	0	2	3
落語	7	0	2	9
論考・論評	12	0	13	25
小説 (情報付与対象)	2604 (97.6%)	628 (99.4%)	1733 (97.9%)	4965 (97.9%)
9X3 サンプル数	2668 (52.6%)	632 (12.5%)	1771 (34.9%)	5071 (100%)

## 【メインジャンル分布 ※1サンプル1ラベル】

	LB	OB	PB	総計
戦争	64	55	79	198
企業	22	24	16	62
医療・看護	22	6	24	52
経済・ビジネス	21	16	12	49
アート・芸術	7	10	17	34
裁判・訴訟	17	5	7	29
動物	12	2	15	29
スポーツ	16	2	3	21
科学・生物	11	6	4	21
軍事・ミリタリー	8		13	21
政治・外交	7		13	20
謀略・陰謀	4	9	7	20
サイバー・電腦	6		10	16
鉄道・交通	6	2	7	15
タイムスリップ	2		11	13
音楽	6	2	5	13
格闘・戦闘	12		1	13
ギャンブル・ゲーム	6		4	10
グルメ・料理	8		2	10
病氣	3	2	2	7
金融・銀行	5		1	6
その他ラベル	3	2	12	17

## 【まとめ】

・PB・OB・LBで小説ジャンルの分布に差異が生じているため、ジャンル差の認識を要する

・ジャンル情報による小説サンプルの絞り込みや抽出が可能となった

・今後、文体情報と小説ジャンルの関係を確認するほか、語彙分布への影響等を考えたい

## 【時代分布 ※1サンプル1ラベル】

	LB	OB	PB	総計
紀元前	19	8	6	33
後漢	15	4	8	27
古代	22	0	24	46
平安	27	8	27	62
鎌倉	14	6	5	25
南北朝	6	0	2	8
中世	17	2	16	35
室町	5	0	0	5
戦国	43	58	34	135
安土桃山	14	2	10	26
江戸	157	22	112	291
近世	14	0	10	24
幕末	37	17	27	81
明治	24	19	16	59
近代	10	0	14	24
大正～昭和初	5	0	17	22
第二次世界大戦	21	11	20	52
戦後・昭和	24	26	19	69
昭和後～平成	6	10	3	19
近未来	13	2	9	24

## 【ジャンル分布 ※1サンプル複数ラベル】

	LB	OB	PB	総計
幻想・ファンタジー	277	40	184	501
伝奇・怪異	68	0	63	131
ヒューマン・人間	76	22	24	122
アクション	73	6	28	107
社会	40	22	28	90
ピカレスク・犯罪	34	19	16	69
私	31	16	9	56
宗教	9	30	15	54
日常・暮らし	25	0	29	54
ジェンダー	13	12	7	32
暗黒・闇・裏	12	4	13	29
暴力・バイオレンス	13	0	6	19
パニック・災害	4	4	2	10
プロレタリア	3	0	0	3

## 【メイントピック ※1サンプル複数ラベル】

	LB	OB	PB	総計
恋愛	285	90	370	745
冒険・アドベンチャー	111	4	70	185
青春	79	31	46	156
人生・成長	56	30	42	128
事件	8	21	28	57
心理・サイコ	26	0	14	40
テロ	7	12	10	29
思想・哲学	8	10	9	27
諷刺・皮肉	11	4	3	18
友情	9	0	7	16

## 【作品傾向 ※1サンプル複数ラベル】

	LB	OB	PB	総計
サスペンス	240	48	137	425
ロマンス	34	0	163	197
ロマン	70	16	31	117
BL	8	0	95	103
官能・好色・アダルト	10	17	73	100
ユーモア	51	28	18	97
ハードボイルド	60	8	25	93
喜劇・コメディ	34	4	30	68
エンターテインメント	13	0	16	29
スリル・スリラー	15	0	11	26
ハートフル・ウォーミング	10	0	14	24
オカルト	5	4	6	15
グロテスク・残酷	1	6	6	13
ハード	10	0	0	10
悲劇	6	0	4	10
ゴシック	6	0	3	9
教養	4	0	2	6
ペーソス	0	0	4	4
ノスタルジー	0	0	2	2
メルヘン	2	0	0	2

