国立国語研究所「日常会話コーパス」プロジェクト報告書 5

# 小説会話文への話者情報付与

山崎誠・宮嵜由美・柏野和佳子

2022年3月
大学共同利用期間法人 人間文化研究機構 国立国語研究所

国立国語研究所「日常会話コーパス」プロジェクト報告書 5

# 会話文話者情報の設計

山崎誠・宮嵜由美・柏野和佳子

2022 年 3 月
大学共同利用機関法人 人間文化研究機構 国立国語研究所

目次

# 1. はじめに

現在公開されている『現代日本語書き言葉均衡コーパス』（以下，BCCWJ）の約 6 割を占める書籍のサンプルには，会話場面における大量の発話文が存在する。会話[1]や発話文は地の文とは言語的に異なる特徴を持つことが多いため，分析に当たっては別に扱うことが妥当である。そこで，2015 年より BCCWJ に収録された小説を対象に，話者情報の付与を行ってきた。当初は BCCWJ 全体を予定していたが，作業量の見積もりから対象を図書館サブコーパス（LB）のみに限定した。

コーパスにおいて，話者情報が豊かであればあるほど分析の幅は広がる。コーパス検索アプリケーション「中納言」で検索可能なコーパスをとっても「日本語話し言葉コーパス」（CSJ），「日本語日常会話コーパス」（CEJC），「名大会話コーパス」（NUCC）などの話し言葉のコーパスはもちろん，「日本語歴史コーパス」（CHJ）のような書き言葉でも，話者の性別，年齢などの情報が付与されている。

現代の書き言葉においても，小説の会話文を考えたとき，そこには話者が存在し，話し言葉のデータと同じよう話者情報を付けることが可能である。当然，生の話し言葉ではないので，そこに書かれているのは擬似的な会話である。また，登場人物も架空のものがほとんどであるため，話者の属性の認定にも限界がある。したがって，小説会話文の話者情報が実際に話し言葉の研究に大きく貢献するとは考えられない。しかし，実際の話し言葉と小説の会話文はどこがどう違うのかは，実際に調べてみないと分からない（山崎他 2019: 313）。その違いが分かった上で，実際の話し言葉と比較した会話文の特徴が明らかになり，そこから会話文独自の分析も可能になると思われる。

小説には役割語が使われる。それは，人間以外のキャラクター（妖精，ロボット，動物など）にも適用される。これらのキャラクターの会話は現実にはありえないため，どのような観点から言葉遣いが選ばれているかといった研究への寄与が見込まれる。

さらに，現代小説において，発話文を同定するために，カギ括弧などの記号による機械的な処理だけでは不十分であることがこの作業を通じて確認された。地の文と発話文がどう区別され，書き分けられているか，発話として認知されるその周辺の文脈と関係については別の課題として研究されるべきであろう。

本報告書では，「中納言」の検索結果に表示される話者情報（話者名，性別，年齢層），および，BCCWJ の関連データサイトで限定公開している話者情報ファイルに含まれている，その他の話者情報について記述する。

[1] 分析対象は基本的には会話文が対象となるが，「独り言（独話）」や「心の中での発話（心内発話）」なども対象となるため，アノテーション付与対象を「発話文」とし，説明する。

## 2. 対象となるサンプル

### 2.1. BCCWJ におけるレジスター別タグの付与状況

対象とする BCCWJ には，次の例 1 (a)に示す通り，前後に改行をともない，カギ括弧(「　」)で囲まれた部分には既に<speech>タグが付与されている。また<quote>タグも，1 文中のカギ括弧で囲まれた部分に付与されている（例 1 (b)）。これらのタグは BCCWJ の xml ファイルに含まれている。

例 1 BCCWJ における発話に関するタグの付与　（サンプル ID: LBr9_00035）

(a)

```
<speech>
<paragraph>
<sentence>　「天気のことについて書いたあの作文の続きなのかい，それとも，べつのことかい？」</sentence>
<br type="automatic_original" />
</paragraph>
```

(b)

```
<sentence>　<squote_A>「いいぞ」</squote_A>とかれはいった。</sentence>
<superSentence><quote><sentence>「どっさりできている。</sentence>
<sentence>アイデスはどうだい？」</sentence>
<zquote_A></superSentence><br type="automatic_original" /></zquote_A>
<sentence>　<squote_A>「うまくいってる」</squote_A>とわたしはいった。
</sentence>
```

まず，これら<speech＞と<quote>タグを暫定的な発話箇所とみなし，レジスターごとのタグ付与状況を表 2-1 に示す（宮嵜他 2017）。

表 2-1　BCCWJ における<speech><quote>タグが付与されたサンプル数とその割合

| レジスター | サンプル数 | <speech>タグを含むサンプル数 | <quote>タグを含むサンプル数 | 発話箇所を含むサンプル数 | 発話箇所を含むサンプル数の割合（%） |
|---|---|---|---|---|---|
| 図書館書籍（LB） | 10551 | 5105 | 8978 | 9987 | 94.65 |
| ベストセラー（OB） | 1390 | 917 | 1080 | 1321 | 95.04 |
| Yahoo!知恵袋（OC） | 91445 | 0 | 0 | 0 | 0.00 |
| 法律（OL） | 346 | 0 | 308 | 308 | 89.02 |
| 国会会議録（OM） | 159 | 159 | 122 | 159 | 100.00 |
| 広報紙（OP） | 354 | 244 | 354 | 354 | 100.00 |
| 教科書（OT） | 412 | 0 | 0 | 0 | 0.00 |
| 韻文（OV） | 252 | 0 | 68 | 68 | 26.98 |

| | | | | | |
|---|---|---|---|---|---|
| 白書（OW） | 1500 | 0 | 1352 | 1352 | 90.13 |
| Yahoo!ブログ（OY） | 52680 | 0 | 0 | 0 | 0.00 |
| 出版書籍（PB） | 10117 | 3479 | 8646 | 9250 | 91.43 |
| 出版雑誌（PM） | 1996 | 844 | 1787 | 1844 | 92.38 |
| 出版新聞（PN） | 1473 | 199 | 1455 | 1457 | 98.91 |
| 合計 | 172675 | 10947 | 24150 | 26100 | 15.12 |

## 2.2. 分析対象サンプル数

例 1 で示したサンプルが小説かどうかの認定は，BCCWJ の書誌情報に含まれる NDC（日本十進分類法）の情報によった。具体的には，NDC が 913，923，933 など 9X3 の形をとっているものを作業対象とした。これらは NDC の分類上，各国に対応した小説・物語を示す。なお，BCCWJ の書籍に付与された NDC は BCCWJ が完成した 2011 年時点のものと，その後 NDC を再調査して新たに付与した 2021 年時点のものがある（加藤他 2021）。以下に示すのは 2021 年版のものである。

表 2-2　分析対象のファイル数

| NDC | 分類 | サンプル数 | 発話なしサンプル数 |
|---|---|---|---|
| 913 | 日本文学・小説 物語 | 2375 | 63 |
| 923 | 中国文学・小説 物語 | 27 | 2 |
| 933 | 英米文学・小説 物語 | 443 | 13 |
| 943 | ドイツ文学・小説 物語 | 18 | 1 |
| 953 | フランス文学・小説 物語 | 52 | 5 |
| 963 | スペイン文学・小説 物語 | 2 | 1 |
| 973 | イタリア文学・小説 物語 | 6 | 1 |
| 983 | ロシア・ソヴィエト文学・小説 物語 | 9 | 1 |
| 計 | | 2932 | 87 |

## 2.3. 対象外のサンプル

3.2 節に述べる「発話箇所認定の基本基準」に照らして，当該サンプルに発話が含まれていないものがあった。地の文だけからなるサンプルがその典型であるが，地の文が一人称の語りのようなサンプルも発話とはみなさず，作業対象外とした。作業対象外のサンプルは合計 87 個あった（表 2-2）。従って，実際に話者情報が付けられたサンプルは，2845 サンプルである。以下に対象外となった 87 サンプルのサンプル ID を挙げる。

LBa9_00040, LBa9_00073, LBb9_00052, LBc9_00020, LBd9_00004, LBd9_00054, LBe9_00176, LBf9_00113, LBg9_00025, LBg9_00118, LBg9_00190, LBh9_00071, LBh9_00078, LBh9_00130, LBh9_00216, LBh9_00235, LBh9_00242, LBhn_00005, LBi9_00080, LBi9_00084, LBi9_00095, LBi9_00107, LBi9_00113, LBi9_00128, LBi9_00142, LBi9_00192, LBi9_00211, LBi9_00251,

LBj9_00045, LBj9_00046, LBj9_00063, LBj9_00066, LBj9_00086, LBj9_00110, LBj9_00124, LBj9_00150, LBk9_00005, LBk9_00061, LBk9_00086, LBk9_00156, LBk9_00174, LBk9_00184, LBk9_00191, LBl9_00030, LBl9_00031, LBl9_00038, LBl9_00044, LBl9_00069, LBl9_00144, LBl9_00156, LBl9_00171, LBl9_00190, LBl9_00220, LBm9_00055, LBm9_00157, LBm9_00238, LBn9_00059, LBn9_00096, LBn9_00101, LBn9_00118, LBn9_00119, LBn9_00190, LBn9_00227, LBo9_00020, LBo9_00091, LBo9_00177, LBo9_00185, LBo9_00188, LBo9_00222, LBp9_00069, LBp9_00150, LBq9_00004, LBq9_00145, LBq9_00181, LBr9_00072, LBr9_00092, LBr9_00094, LBs9_00156, LBs9_00163, LBs9_00182, LBs9_00237, LBt9_00009, LBt9_00012, LBt9_00050, LBt9_00053, LBt9_00061, LBt9_00103

## 3. 話者情報の付与 [2]

### 3.1. 発話箇所の認定

BCCWJ に収録されたサンプル範囲内でどの部分がアノテーションの対象とする発話箇所に該当するかは，まず，BCCWJ に付与されている，<speech>や<quote>というタグ [3]を作業上の目安とした。<speech>は，前後に改行をともない，カギ括弧（「　」）で囲まれた部分に付与されており，また，<quote>は，1 文中のカギ括弧で囲まれた部分に付与されている。

<speech>は発話箇所であることが多いが，<quote>タグの場合，文中の強調したい部分，映画や本の名前，「釣りぼり」など看板を示す文字列にも付与されていることがあり，必ずしも発話箇所であるとは限らない。逆に，カギ括弧で括られない場合にも，声に出したと想定される発話箇所が多数存在する。小説という表現媒体上，記号の用い方を含めそれらの様子は多岐に渡っている。

そこで，「①カギ括弧で括られ」かつ「②声に出した（と想定される）」という 2 つの条件を満たした箇所を「原則的な発話」と認定し，それ以外を「原則的な発話に準ずる発話」とすることとした。以下に一例を示す（下線は筆者らが付与）。

例 1 （サンプル ID：LBp9_00190）

「東吾の同意なくば、この話は成り立たぬのだ」

どうじゃ、承知してくれるか、と重ねて通之進がいい、東吾は畳に手を突いて、深く頭を下げた。

【出典】平岩弓枝（2001）「春の高瀬舟」文藝春秋

冒頭の「東吾の同意なくば、この話は成り立たぬのだ」は「原則的な発話」である。続く，[どうじゃ、承知してくれるか]の部分はカギ括弧で括られていないという点で「原則的な発話」にあたらないが，“と重ねて～いい”とあるため，発話と考えてよい。このようなものを「原則的な発話に準ずる発話」とする。作業者による恣意的な発話箇所の認定を防ぐため，次の 5 つの「発話箇所認定の基本基準」を定めた。

**発話箇所認定の基本基準**

1）カギ括弧に括られた声に出したと想定される部分…〈原則的な発話〉
2）カギ括弧に括られた 1）に準ずる部分…〈準ずる発話〉
3）カギ括弧に括られた当該の文字列の強調などを示す部分 …〈非発話〉
4）カギ括弧に括られない声に出したと想定される部分…〈準ずる発話〉
5）「場面設定」を考慮した 1）に準ずる部分…〈準ずる発話〉

---

[2] 本章は，宮嵜他（2017），宮嵜（2018），山崎他（2018），山崎他（2019），宮嵜（2019）をもとに，主に柏野がまとめて記述したものである。

[3] 詳しくは西部他（2011: 240,274）を参照されたい。

以下，順に説明する。

１）カギ括弧に括られた声に出したと想定される部分…<原則的な発話>
「①カギ括弧で括られ」かつ「②声に出した（と想定される）」という 2 つの条件を満たした箇所はすべて「原則的な発話」と認め，話者情報を付与する。

２）カギ括弧に括られた１）に準ずる部分…<準ずる発話>
「①カギ括弧で括られ」ではいるが，声に出していないもののうち，主に次に述べる二つに該当する場合，発話と認める。一つ目は心の中での発話である「心内発話」である。二つ目は「…」「？」「！」「！？」のようなものである。これらは発声はないが，コミュニケーション上の役割を考慮して，発話に準ずるものとして認定する。その場合，話者情報にはこれらが順に，「沈黙」「疑問」「驚き」「驚愕」を表すものであることを付与する。

３）カギ括弧に括られた当該の文字列の強調などを示す部分 …<非発話>
「①カギ括弧で括られ」ではいるが，看板を示すものや文章の強調など，声に出していないものは発話と認めない。この場合，話者情報の備考欄に「発話ではない」「非発話」と記したものもあるが，網羅的にはその情報を付与してはいない。

４）カギ括弧に括られない声に出したと想定される部分…<準ずる発話>
小説には，カギ括弧で括られない場合にも，「②声に出した（と想定される）」と認められる発話が多く存在する。この認定が作業者による恣意的なものにならないよう，「②声に出した（と想定される）」と判断できる根拠となる語句を抽出しながら認定をする。

５）「場面設定」を考慮した１）に準ずる部分…<準ずる発話>
小説・物語上の特殊な場面設定を考慮したうえで「発話」と認めるものがある。たとえば，夢の中での会話場面や，ＳＦ小説などでのロボットの会話場面，ファンタジー小説のテレパシーを使った会話場面などである。そのような場面設定の場合はそれを考慮し，「①カギ括弧で括られ」ていることを条件に発話に準ずるものとして認定する。

なお，カギ括弧に括られず，声に出していなければ，発話認定の対象外であるが，カギ括弧以外の記号で括られる「心内発話」と見なしたものは，特別に準ずる発話と認めた場合もある。

## 3.2. 話者の属性

5 つの「発話箇所認定の基本基準」に基づき「発話」と認定した箇所には，下記表 3-1 の No.1～3 を可能な限り網羅的に付与した。No.4 と No.5 は該当する場合のみ，付与した。また，「原則的な発話に準ずる発話」には，その準ずる理由がわかるよう，No.6～7 を適宜付与した。No.8 は必要性がある場合記述した。No.9～10 については，網羅的な付与は試みず，特定がしやすかった場合にのみ付与した。

表 3-1　話者属性とその説明

| No. | 属性 | 説明 |
|---|---|---|
| 1 | 話者名 | （BCCWJ に収録された範囲内での）登場人物の呼び名。話者名の最後にある●は心内発話を表す |
| 2 | 性別 | 男，女，不明 |
| 3 | 年齢層 | 若年層（～19 歳），成年層（20～59 歳），老年層（60 歳以上） |
| 4 | 非人間 | 生物学的な人間以外のものの場合に○ |
| 5 | 会話モード | 通常の対話場面でない場合の情報。「電話，方言，外国人，テレパシー，引用，独話，疑問，沈黙，驚き，驚愕」など |
| 6 | 会話認定情報 1 | 発話に準ずる場合，そのタイプ<br>カギ括弧なし発話（タグなし），心内発話 |
| 7 | 会話認定情報 2 | 発話に準ずる場合の認定根拠 |
| 8 | 備考 | さまざまな注記。具体的な年齢など |
| 9 | 職業 | 発話者の職業，社会的身分 |
| 10 | 相手 | 会話の相手の名前 |

以下，No.1～No.10 の付与方法について，順に作業マニュアル（宮嵜 2019）に記した内容をもとに説明する。

先に，主に話者名に用いる記号とその意味を示す。記号はすべて全角である。

「／」発話者が登場人物のうち誰か特定できない場合，名前を列挙し「／」でつなぐ。
「＆」複数の登場人物が同時に発話している場合，名前を列挙し「＆」でつなぐ。
「？」〈不特定単数人発話〉で使用。話者が登場人物のうちいずれかひとりではあるが，不確定である場合。
「＃」〈不特定複数人発話〉で使用。話者が集団としては判断できるものの，不特定であり，複数の場合。※この記号は話者名のほか，性別，年齢層にも使用。

以下，属性別に入力方法を示す。

### No.1 話者名（必須）

BCCWJ に収録されたサンプル範囲内での呼び名。原則として名前（話者が特定できる程度の情報を付与）。他の人の発話を引用している発話等，いわゆる間接話法の場合であっても，直接の話者の情報のみを記述する。なお，「心内発話」の場合，最後に●を付す。

◆単数人発話の場合

(1)特定できる場合，姓名の形式で記入するが，姓名のどちらか，あだ名，肩書しか特定できない場合などは，BCCWJ に収録されたサンプル範囲内でわかるもので記述する。

姓名例：「ジョン・レノン」「山田太郎」等

姓 or 名例：「鈴木」「マイケル」等

あだ名例：「八ちゃん」「スーさん」等

肩書例：「刑事」「生徒」等

(2)同じ名前等で別人であることを区別する必要がある場合（性別・年齢層が違ってくるなどがわかるような場合）は，名前等の後に全角で「１」「２」「３」…のように数字を付け，区別する。

例：「選手１」「選手２」，「子供１」「子供２」等

(3)特定しがたい個人（単独）の場合は，次のように全角「？」を冒頭に付与して示す。

a.話者がいずれかであることまではわかる場合

例：？田中／栗林／岩田，？刑事，？男，？諸将

b.話者がサンプル内資料では不明の場合：？不明

◆複数話者が同時に発話している場合

a.複数同時話者全員の発話である場合：話者名の順に準拠し，それぞれの名前を記入。

例：田中＆栗林，＃男，＃観衆同時　等

b.複数同時話者の誰かの名前が不明の場合

例：？不明＆太郎，＃使用人＆おくみ

＜注：？のみの記述は存在しない＞

c.発話者がそれぞれ確定できない場合：複数同時話者全員の名前が特定できない場合は，自由記述欄に「＃不明」と記載。

**No.2 性別（必須）**

「男，女，不明」

◆単数人発話の場合

(1)人間の場合は生物学的な性別を記入。異なる性別の言葉を使っている場合でも生まれつきの性別を記入。

(2)「不明」は，性別が分からない場合に記入。

◆複数話者が同時に発話している場合

a.複数同時話者全員の性別がわかる場合：話者名の順に記入。

例：＃男（すべて男の場合），女＆男

b.複数同時話者の誰かの性別が不明の場合：「不明」を「＆」でつなぐ。

例：不明＆女

**No.3 年齢層（必須）**

「若年層（～19 歳），成年層（20～59 歳），老年層（60 歳以上）」

◆単数人発話の場合
(1)判断できた年齢層を記入。
(2)「幼年」（およそ 6 才以下）と分かった場合は備考にその旨を注記する。
(3)具体的な年齢が分かった場合は，備考に記入。

◆複数話者が同時に発話している場合
a.全員の年齢層がわかる場合：話者名の順に記入。
例：＃若年層（すべて若年層の場合），若年層＆幼年層
b.誰かの年齢層が不明の場合：「？」付きの年齢層を「＆」でつなぐ。
例：？若年層＆若年層
c.発話者がそれぞれ確定できない場合：「＃不明」と記入。

**No.4 非人間（該当する場合）**
生物学的な人間以外のものの場合に○。魔女，超能力者などは人間と判定する。
例：動物，妖怪，精霊，ロボット（人造人間）

◆非人間と人間の同タグ内複数発話の場合
a.非人間と人間の区別がつく場合：発話の出現順に記入。
例：○／×／×
b.複数同時発話の話者のうち誰かが非人間の場合：「＆」でつなぐ。
例：○＆×

**No.5 会話モード（該当する場合）**
通常の対面会話でない場合に記入。
「電話／方言／外国語／外国人／テレパシー／引用／独話／沈黙／疑問／驚き／驚愕／叫び／呻き声」等

・「引用」とは，声に出して誰かの発言を引いたり，看板や手紙を読んでいたり，などを指す。いわゆる間接話法も「引用」と扱う。
・「独話」とは，声に出していると思われる独り言を指す。
・「沈黙」とは，「…」のようなものを指す。その他，「？」は疑問，「！」は驚き，「！？」は驚愕など，発声の伴わない驚きの表現なども，発話に準ずるものとして認定し，ここに記す。

**No.6 会話認定情報 1（該当する場合）**
発話に準ずる場合，そのタイプ。

a.カギ括弧はないが，声に出した会話である場合：「タグなし」と記入。
b.カギ括弧はあるが，声に出さない心の中での発話である場合：「心内発話」と記入。
c.カギ括弧がなく，声に出さないが，心の中での発話である場合：「心内発話」と記入。

### No.7 会話認定情報 2（該当する場合）

発話に準ずる場合の認定根拠にあたる部分を原文から引用して記入。

◆「タグなし」の場合

a.発話であることがわかる語句がある場合：根拠となる語句を記入。

b.カギ括弧以外の記号がある場合：開始括弧に相当する冒頭部分の一字（全角）を「発話」に付与して記入。

c.字下げなどがある場合：そのことがわかるよう記入する。

表 3-2　「今日は暑いなあ」を「タグなし」発話とした場合の記載例

| 原文（作例） | 会話認定情報 1 | 会話認定情報 2 |
|---|---|---|
| 今日は暑いなあ、と言った。 | タグなし | と言った |
| 今日は暑いなあ、と彼は窓を開けた。 | タグなし | と |
| ―今日は暑いなあ。 | タグなし | ―発話 |
| （今日は暑いなあ） | タグなし | （発話 |
| 『今日は暑いなあ』 | タグなし | 『発話 |
| “今日は暑いなあ” | タグなし | “発話 |
| 　今日は暑いなあ。 | タグなし | 字下げ発話 |

◆「心内発話」の場合

a.カギ括弧がある場会：根拠となる語句を記入。

b.カギ括弧以外で心内発話と認める場合：開始括弧に相当する冒頭部分の一字（全角）を「心内発話」に付与して記入し，続けて全角スラッシュの後，その根拠となる語句を記入。

例：“心内発話／と思った，（心内発話／と思った

表 3-3　「今日は暑いなあ」を「心内発話」とした場合の記載例

| 原文（作例） | 会話認定情報 1 | 会話認定情報 2 |
|---|---|---|
| 「今日は暑いなあ」と思った。 | 心内発話 | と思った |
| （今日は暑いなあ）と思った。 | 心内発話 | （心内発話／と思った |

### No.8 備考

さまざまな注記。具体的な年齢などを記入。複数の情報を記入する場合は，「／」で区切る。

①複数人同時発話で，話者特定が困難な場合は，こちらにその旨を記載する。

②「年齢層」に関し，

a.「幼年」の場合はこちらに記載。

b.具体的年齢がわかった場合はこちらに記載。

③誰かの発話の引用などの場合，引用元の発話情報「想像上の発話」「夢の中の回想」等は，ここに記入。

例：「A君のこと好きだよ，とかBちゃんに言われないかなぁ」⇒Aの会話のBによる想像。

④その他，発話者に特筆すべき特徴や事情がある場合ここに記入。

⑤発話者⇔聞き手（相手）の関係性に関し，相手との関係性が，小説特有の設定（たとえば話者の変身），家族や恋人関係など＜特筆すべき関係＞の場合は，その関係性情報をここに記入。

◆関係性を記入する場合

関係性情報を記入する際には全角の★を冒頭に付し，「★話者名（立場）×相手名（立場）は関係性。」の形で記入する。ただし，立場が（男）や（女）など既に話者情報にその情報がある場合は省略して（）のみ記入。

＊家族の場合

例：★太郎（夫）×恵子（妻）は夫婦。

例：★太郎（親）×花子（子）は親子。

＊恋人，愛人関係の場合

例：★武郎（）×花子（）は恋人。

例：★太郎（）×菊子（）は愛人。

太郎に付与された，男性／成年層／52 歳／会社員（部長），菊子に付与された，女性／成年層／20 代／ＯＬ，などの話者情報と総合して，立場の認定は可能。

**No.9 職業**

発話者の職業，社会的身分

**No.10 相手**

会話の相手の名前

①対応する相手（聞き手）の「話者名」を記入。

②5 人まではそれぞれの名前を話者名の入力規則に準じて列挙。それ以上ならば「登場人物全員（６名以上）」等，該当するグループ名（任意）を記入し入力。「＃」は「不特定」複数を表す記号のため，この場合用いない。

③会話モードが「独話」「心内発話」「沈黙」「驚愕」「叫び声」などの場合は空白処理となる。

## 3.3. 発話箇所の認定と話者の属性の付与例

3.1 節で示した発話箇所認定の基本基準に基づき，発話箇所を認定し，3.2 節で示した話者の属性を付与した具体例を示す。その際に，基準を満たすと判断した根拠を四角で囲み，認定した発話箇所を丸数字（①，②，③…）で示す。原文にカキ括弧等がなく発話と考えら

れる箇所には[　]を括って示す。また，その発話者を特定する根拠部分に下線を付して示す。

### 3.3.1. 基準１）原則的な発話の例

発話箇所認定の基本基準の「１）カギ括弧に括られた声に出したと想定される部分…<原則的な発話>」と認定した例を，例２に示す。

例２　（サンプル ID：LBc9_00007）

<superSentence><sentence type="fragment">そのあと母が，</sentence>
<xquote_A><sentence>①「新しい技術がまじってたわね，どこでおぼえたの？」</sentence></xquote_A>
<zquote_A><sentence>といった。</sentence></zquote_A>
</superSentence><sentence><squote_A>②「おぼえたんじゃない，前からこんなふうにやってみたかったんだ」</squote_A>というのが父の返事だった。</sentence>
<sentence>ふたりとも嬉しそうな声だった。</sentence>

【出典】デヴィッド・ウィルツ著　汀一弘訳（1988）『わが故郷に殺人鬼』扶桑社

この部分では，①は「母」の発話であり，②は「父」の発話であることがわかる。別の箇所により，それぞれの名前を確定し，①に「話者名：ジャネット，性別：女，年齢層：成年層」の属性を付与し，②に「話者名：ピーター，性別：男，年齢層：成年層」の属性を付与した。

### 3.3.2. 発話者が複数人で，基準１）原則的な発話の例

次の例３は，一同（複数人）が同時に発話していると認定したものである。

例３　（サンプル ID：LBb9_00003）

<speech>
<paragraph>
<sentence>　①「あっ」</sentence>
<br type="automatic_original" />
</paragraph>
</speech>
</quotation>
<paragraph>
<sentence>　一同の者は，びっくりしてかれを見た。</sentence>

【出典】斎藤栄（1987）「少年探偵ジャーネ君の冒険」講談社

①には，「話者名：＃一同，性別：男＆女，年齢層：若年層＆成年層＆老年層」と付与した。複数人発話を表すよう定めた全角「＃」を話者名の冒頭に付けている。また，この場合，性別と年齢層はその場にいる人物のものを＆でつなぎ記入している。

### 3.3.3. 発話者が複数人で，基準１）原則的な発話と基準４）準ずる発話のある例

次の例４は，基本基準の「４）カギ括弧に括られない声に出したと想定される部分…<

準ずる発話>」と認定した発話を含む例である。かつ，同一文内に複数発話箇所がある場合の例である。

例 4 （サンプル ID：LBb9_00034）
〈sentence〉男が出た。〈/sentence〉
〈sentence〉①[綿貫つや子を呼んでくれ]というと、お待ちくださいともいわずに、受話器を硬い物の上に置き、〈quote〉「②「つやちゃん」〈/quote〉と呼んだ。〈/sentence〉
〈sentence〉気忙しい音楽が鳴っていて、男の高笑いも聞こえた。〈/sentence〉
【出典】梓林太郎（1987）「死紋山脈」角川書店

②はカギ括弧と，「呼んだ」とあることから，原則的な発話と認定できる。話者は「男」としか描写されていないため，話者名はそのまま「男」とし，②には「話者名：男，性別：男，年齢層：成年層」と付与した。①はカギ括弧はないが男への電話内容部分が発話と認定できる。別の箇所により，名前を確定し，①には「話者名：町浦健作，性別：男，年齢層：成年層」と付与した。さらに，①②ともに，「会話モード：電話」と付与し，さらに①には「会話認定情報 1：タグなし，会話認定情報 2：というと、」と付与した。

### 3.3.4. 発話者が一人で，基準 1）原則的な発話と基準 4）準ずる発話のある例

次の例 5 は，同じ発話者による発話とみられる部分に，カギ括弧がついたり，つかなかったりする場合がある，という例である。

例 5 （サンプル ID：LBp9_00190）
〈speech〉
〈paragraph〉
〈superSentence〉
〈sentence〉①「神林家は、わしと東吾と二人だけの兄弟である。〈/sentence〉
〈sentence〉②東吾の同意なくば，この話は成り立たぬのだ」〈/sentence〉
〈/superSentence〉〈br type="automatic_original" /〉
〈/paragraph〉
〈/speech〉
〈/quotation〉
〈paragraph〉
〈sentence〉③[どうじゃ、承知してくれるか]、と重ねて通之進がいい、東吾は畳に手を突いて、深く頭を下げた。〈/sentence〉
【出典】平岩弓枝（2001）「春の高瀬舟」文藝春秋

①と②の部分は，原著では同一話者による，改行なしのひとつのカギ括弧内の発話であるが，BCCWJ では<sentence>タグによって 2 行に分けられている。このような場合，行ごとに発話を認定し，同一の話者属性を付与する。よって，①と②には「話者名：神林通之進，性別：男，年齢層：成年層」と付与した。③はカギ括弧に括られていないが，[]で括った箇所は①と②に続く発話と認定できるため，①②と同じ話者情報に加え，「会話認

定情報 1：タグなし，会話認定情報 2：と重ねていい」を付与した。
また，次の例 6 もカギ括弧はないが，[]で括った箇所を発話と認めた例である。

例 6 （サンプル ID：LBp9_00203）
〈sentence〉①[それにしても…]と、蘭の方は深い安堵の吐息とともに言った。〈/sentence〉
【出典】岩崎正吾（2001）「遥かな武田騎馬隊」角川春樹事務所

この場合，同文中にある“言った”を，声に出した発話と認定する根拠とする。よって，①には「話者名：蘭の方，性別：女，年齢層：若年層」を付与し，「会話認定情報 1：タグなし，会話認定情報 2：言った」と付与した。
しかしながら，カギ括弧のない発話認定には迷いが多く生じる，次の例 7 は，当初は「返事をした」を根拠に発話と認定しようとしたものの，最終的には発話と認めず，アノテーション情報を何も付与しなかった例である。発話認定の難しさを示す例として，以下にあげる。

例 7 （サンプル ID：PB59_00081）
〈sentence〉 ［ふうん―］と、中禅寺は感心したような馬鹿にしたような返事をした。〈/sentence〉
〈sentence〉それから徐に横に視線を送って、電柱に凭れかかっていた風采の上がらない男に向けてこう云った。〈/sentence〉
【出典】京極夏彦（2005）「百器徒然袋-雨」講談社

### 3.3.5. 基準３）非発話の例

次の例 8 は，基本基準の「３）カギ括弧に括られた当該の文字列の強調などを示す部分…〈非発話〉」により，非発話と判断した例である。<quote>タグで囲まれているものの，文脈から，以前に自身が発話した内容を統括した引用を示すのみであり，非発話と認定さした。この例の他，店名や部屋名を示す看板や掲示についても同様の判断をしている。

例 8 （サンプル ID：Bp9_00237）
〈sentence〉〈pquote_1〉「客のめし」〈/pquote_1〉の味も、食糧の豊かな時代にあっては〈pquote_2〉「豚がわり」〈/pquote_2〉に動員された屈辱を救いきれない。〈/sentence〉
【出典】森村誠一（2001）「鍵のかかる棺　下」徳間書店

例 9 （サンプル ID：LBj9_00113）
〈sentence〉 敵が逃げ去ったことで与五郎はまた〈quote〉「かのおり」〈/quote〉に戻った。〈/sentence〉
【出典】西村望（1995）「義士の群れ」広済堂出版

上記の例 8，例 9，いずれにも，何もアノテーション情報は付与していない。このように，「非発話」の場合，原則として何も付与していない。ただし，作業中のメモとして，話者情報の備考欄に「発話ではない」「非発話」と記したものもある。

### 3.3.6. 基準１）原則的な発話とみるが聞き手を意識しない発話「呻き声」の例

聞き手を意識しないと想定される発話も，カギ括弧でくくられ，声が出ていると判断できれば，原則的な発話と認定している。たとえば「独話」や「呻き声」と判断できる場合は，会話モードにその情報を記している。

次の例 10 は，アスタシュールという男が場内のアナウンスを聞き，ひとり呻く場面である。

例 10 （サンプル ID：LBh9_00135）

```
<sentence> ①「む…」</sentence>
<br type="automatic_original" />
</paragraph>
</speech>
</quotation>
<paragraph>
<sentence> すでに真昼の陽光が射す丘の上である。</sentence>
<br type="automatic_original" />
</paragraph>
<paragraph>
<sentence> アスタシュールが軽い呻きをあげて立ち止まった。</sentence>
```

【出典】伏見健二（1993）「叛逆の獣将」中央公論社

①の「む…」は「実際に声に出した発話」として認定はできる。しかし，例 10 に引用した箇所の続きでは，この発話を耳にした別の人物が「どうしたんだよ」と応じるものの，相手に向けての発話とは考えにくい。このように，聞き手を意識しない発話であると判断した場合は，そのことがわかるように，話者名等の情報に加え，「会話モード：呻き声」との情報を付与した。

### 3.3.7. 基準２）準ずる発話「心内発話」「沈黙」の例

カギ括弧が付与されているものの，声に出していない発話の一つに「心内発話」がある。次の例 11 がその例である。カギ括弧が付与されていることと，“自分に言い聞かせた”との心内発話を意味する名詞や動詞を認定の根拠とし，「基本基準の「２）カギ括弧に括られた１）に準ずる部分…<準ずる発話>」に該当するものとして考える。

例 11（サンプル ID：LBa9_00055）

```
<speech>
<paragraph>
<sentence> ①「せめて三味線でもひけたら」</sentence>
<br type="automatic_original" />
</paragraph>
</speech>
</quotation>
<sentence>と妓は思っていた。</sentence>
```

【出典】近藤富枝（1986）「今は幻吉原のものがたり」講談社

①は「心内発話」と判断し，話者名等の情報に加え，「会話認定情報 1：心内発話，会話認定情報 2：と思った」を付与した。

カギ括弧が付与されているものの，声に出していない発話のもう一つに「沈黙」がある。「沈黙」は発言の裏返しの行為ではない。意見に対する反感や内容の吟味などその機能はさまざまである。よって，「沈黙」は小説における会話場面内で，ある一定の意味をもつ発話として認定し，発話者情報を付与している。次の例 12 がその例である。

「沈黙」は多く，カギ括弧で括られた三点リーダー「…」によって表される。その出現パターンは，①発話中のいわゆる“言いよどみ”を表す「沈黙パターン(a)」と，②括弧内が沈黙のみで表される「沈黙パターン(b)」との 2 つに分類できるも。

例 12 （サンプル ID：LBq9_00101）

<sentence>①「…どないしたん，由香ちゃん。</sentence>
<sentence>②泣いたら疲れたか？」</sentence>
</superSentence><br type="automatic_original" />
</paragraph>
</speech>
</quotation>
<quotation>
<speech>
<paragraph>
<sentence>③「…」</sentence> >

沈黙パターン(a)

沈黙パターン(b)

【出典】佐藤ケイ（2002）「Last kiss」メディアワークス;角川書店

①の「沈黙パターン(a)」の場合，発話開始時のいわゆる“言いよどみ”であるため，原則発話と認定し，通常の発話と同様に①，②ともに話者名等の話者情報を付与した。同様に，②の「沈黙パターン(b)」の場合も発話と認定し，話者名等の話者情報を付与したが，それに加え，「会話モード：沈黙」と付与した。

### 3.3.8. 基準５）準ずる発話「夢の中」の例

最後に，発話箇所認定の基本基準の「５）「場面設定」を考慮した１）に準ずる部分…<準ずる発話>」と認定した例を，例 13 に示す。

例 13 （サンプル ID：LBh9_00122）

<sentence>① 『拓ちゃん，ごめんなさいね，駄目なのよ。</sentence>
<sentence>②あたし達のせいなの』</sentence>
</superSentence><br type="automatic_original" />
</paragraph>
</speech>

〈/quotation〉
〈paragraph〉
〈sentence〉　あ。〈/sentence〉
〈sentence〉この声は，夢ちゃんのママだ。〈/sentence〉

【出典】新井素子（1993）「緑幻想」講談社

例 13 の原著では，「夢の中」という場面において，主人公が声に出した発話にはカギ括弧が，主人公以外が声に出した発話には二重カギ括弧が付与されるといった規則性がみられる。例 10 に引用した箇所では，この二重カギ括弧部分は，最終行にて「この声」と指されているため，「夢の中」という場面において声に出した発話と認定した。

この場合，①②ともに話者名等の話者情報を付与し，それに加えて「会話モード：夢の中，会話認定情報 1：タグなし，会話認定情報 2：この声」と情報を付与した。「会話認定情報 1：タグなし」も付与する理由は，原則的な発話の認定に用いる一重のカギ格好のつく発話と区別するためである。

## 4. 話者情報の統計

以下は，限定公開している speakersInfo_LB_Novels_20211216.txt　による集計である。ファイルの詳細は 5 章に述べる。

### 4.1. 話者名

話者名は，各サンプル内での話者の区別を表すために付与している。そのため異なるサンプルに同じ名前の話者がいても，それは必ずしも同一人物を指しているとはかぎらない。図 4-1 は 1 つのサンプルに平均して何人に話者が現れるかを示したものである。全サンプルの平均は 5.60 人（話者がゼロのサンプルを含んだ平均 [4]）である。

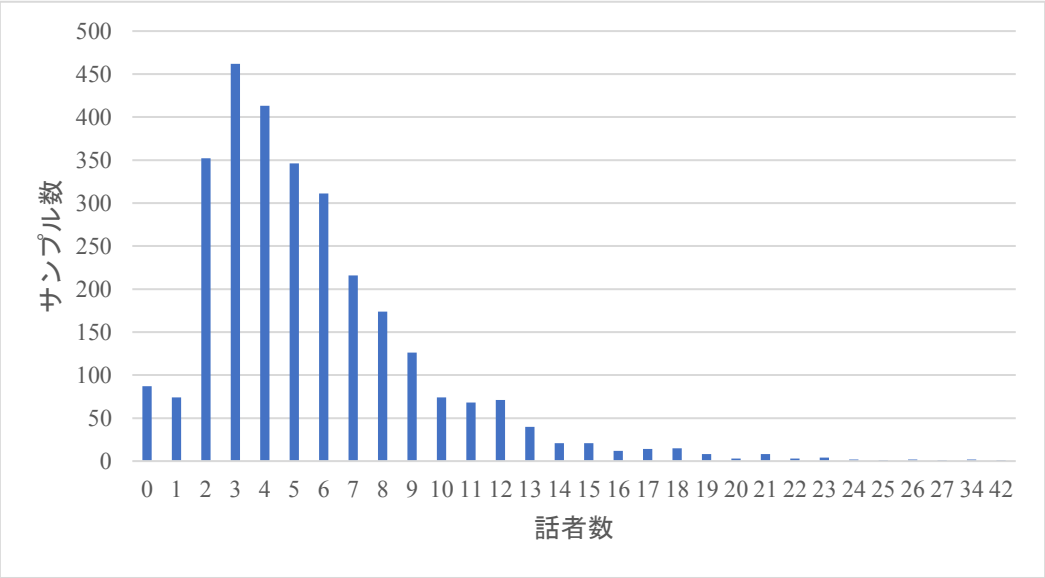


図 4-1　サンプルにおける話者数の分布

### 4.2. 性別

表 4-1 は，「性別」欄に付けられた情報に基づいて集計したものである。単独の性別情報が付けられたものだけを見ると，「女」が 30%弱，「男」が 70%弱で，「不明」が約 2%である。「&」で示された複数人による同時発話や「／」で示された特定不能の発話は最大でも 0.2%で全体から見ると少ないことが分かる。

表 4-1 性別と語数

| 性別 | 語数 | 割合 (%) |
|---|---|---|
| 女 | 1117970 | 28.94 |
| 女&不明 | 7 | 0.00 |
| 女&女&男 | 13 | 0.00 |

[4] 話者数がゼロのサンプルを含まない場合の平均は 5.63 人である。また，中央値は 5 である。

| | | |
|---|---|---|
| 女&女&男&男 | 3 | 0.00 |
| 女&女&男&男&男 | 4 | 0.00 |
| 女&男 | 220 | 0.01 |
| 女&男&不明&男 | 11 | 0.00 |
| 女&男&女&男 | 13 | 0.00 |
| 女&男&男 | 8 | 0.00 |
| 女／女／男 | 70 | 0.00 |
| 女／男 | 205 | 0.01 |
| 女／男／女 | 15 | 0.00 |
| 女／男／男 | 3 | 0.00 |
| 女／男／男／男 | 58 | 0.00 |
| 男 | 2661460 | 68.89 |
| 男&不明 | 75 | 0.00 |
| 男&女 | 515 | 0.01 |
| 男&女&女&女 | 7 | 0.00 |
| 男&男 | 49 | 0.00 |
| 男&男&女 | 29 | 0.00 |
| 男&男&男 | 3 | 0.00 |
| 男&男&男&女&女 | 26 | 0.00 |
| 男／女 | 582 | 0.02 |
| 男／女／女 | 25 | 0.00 |
| 男／女／女／女 | 11 | 0.00 |
| 男／男 | 32 | 0.00 |
| 男／男／不明 | 6 | 0.00 |
| 男／男／女 | 27 | 0.00 |
| 男／男／女／女 | 7 | 0.00 |
| 男／男／女／女／女 | 29 | 0.00 |
| 不明 | 81698 | 2.11 |
| 不明&不明&不明 | 8 | 0.00 |
| 不明&男 | 20 | 0.00 |
| 不明／女 | 11 | 0.00 |
| 計 | **3863220** | **100.00** |

## 4.3. 年齢層

表 4-2 は，「年齢層」欄に付けられた情報に基づいて集計したものである。年齢層を列挙したものなどでなく，1 種類の年齢層情報が付けられたものを見ると，「若年層」が約 12%，

「成年層」が約 75%，「老年層」が 6%弱，「不明」が 1.6%である。発話された語全体の 4 分の 3（「成年層？」を含めると全体の約 8 割）が成年層による発話であることが分かる。話者名，性別，年齢層は「中納言」の検索結果に表示される。

表 4-2 年齢層と語数

| 年齢層 | 語数 | 割合 (%) |
|---|---|---|
| 若年層 | 459409 | 11.89 |
| 若年層？ | 16727 | 0.43 |
| 若年層＆成年層 | 66 | 0.00 |
| 若年層＆成年層＆成年層 | 8 | 0.00 |
| 若年層＆成年層＆老年層 | 22 | 0.00 |
| 若年層＆若年層＆成年層＆不明 | 13 | 0.00 |
| 若年層／成年層 | 8 | 0.00 |
| 若年層／成年層／成年層／成年層 | 11 | 0.00 |
| 若年層／成年層／老年層 | 4 | 0.00 |
| 成年層 | 2923593 | 75.68 |
| 成年層？ | 166263 | 4.30 |
| 成年層＆成年層 | 21 | 0.00 |
| 成年層＆成年層＆不明＆成年層 | 11 | 0.00 |
| 成年層＆成年層＆若年層 | 4 | 0.00 |
| 成年層＆老年層 | 6 | 0.00 |
| 成年層＆若年層 | 70 | 0.00 |
| 成年層／老年層 | 13 | 0.00 |
| 老年層 | 225093 | 5.83 |
| 老年層？ | 5564 | 0.14 |
| 老年層＆成年層 | 34 | 0.00 |
| 老年層／成年層 | 29 | 0.00 |
| 老年層／老年層／成年層 | 5 | 0.00 |
| 不明 | 63450 | 1.64 |
| 不明？ | 2776 | 0.07 |
| 不明＆若年層 | 20 | 0.00 |
| 計 | **3863220** | **100.00** |

## 4. 4. 非人間

動物や妖精，ロボットなどの非人間という情報が付けられた語は，全部で 74655 語あった。表 4-3 は，そのうち上位 50 を示したものである。話者名だけでは具体的な属性が把握できにくいものが多いため，「備考」欄に付けられた情報を併せて示した。

いちばん語数の多い「治療院」というのには，コンピュータが運営する医院でそこのスタッフのロボットを示している。非人間以下の情報は「中納言」の検索結果には表示されない。

表 4-3 非人間の情報が付いた話者名と語数

| 順位 | 話者名 | 備考 | 語数 |
|---|---|---|---|
| 1 | 治療院 | | 2912 |
| 2 | 圓生 | 幽霊 | 2295 |
| 3 | 亭主 | 幽霊 | 2178 |
| 4 | ラプソディ | | 2119 |
| 5 | 山手線 | | 2052 |
| 6 | ペンギンのおばあさん | | 1558 |
| 7 | ヘルメス・レオナ | レプタリアン（爬虫人類） | 1545 |
| 8 | マードック | ねずみ | 1532 |
| 9 | 北斗星 | | 1383 |
| 10 | ルシファー | | 1352 |
| 11 | パーソナル・コンピュータ | | 1309 |
| 12 | ギャビー | | 1069 |
| 13 | 女性脳 | | 1007 |
| 14 | アンナ | 吸血鬼 | 904 |
| 15 | 風間疾風 | レプタリアン（爬虫人類） | 869 |
| 16 | ダニー・ルート | | 867 |
| 17 | フェストリン | 別名＝セイザー | 852 |
| 18 | アシェ | | 825 |
| 19 | ゾウ | | 818 |
| 20 | 菅原道真 | ４０歳位 | 815 |
| 21 | モグラ | | 788 |
| 22 | 犬神明 | | 752 |
| 23 | 大きな雪玉 | | 722 |
| 24 | モンタン | | 701 |
| 25 | チー | イタチ | 693 |
| 26 | 夏実 | ４０歳位 | 682 |
| 27 | コンタ | 子狐 | 657 |
| 28 | 神さま | | 655 |
| 29 | モーゼス | ねずみ | 644 |
| 30 | 妻 | 幽霊 | 635 |
| 31 | 大五郎 | 妖怪 | 633 |

| | | | |
|---|---|---|---|
| 32 | なは | | 614 |
| 33 | トム | | 609 |
| 34 | サタン | | 595 |
| 35 | カーサン | | 576 |
| 36 | アラエッサ | 鶏 | 573 |
| 37 | ミーシャ | 子兎 | 544 |
| 38 | ファイヤー | 赤信号 | 491 |
| 39 | ひかり | | 489 |
| 40 | ナルコ | カントルー星人／地球人のふりをしている | 465 |
| 41 | 神さま | １５００歳 | 463 |
| 42 | 噺家 | | 460 |
| 43 | ストンストン | 豚 | 457 |
| 44 | リー | リス | 456 |
| 45 | 母 | 幽霊／★母（母）×るみ（娘）は親子。 | 453 |
| 46 | 闇主 | 実際の年齢不詳。 | 450 |
| 47 | 白い象 | | 446 |
| 48 | ファンシー | | 443 |
| 49 | ダッチェス | | 430 |
| 50 | ロフタス | オオガラス／７５歳 | 425 |

## 4.5. 職業

表 4-4 は，「職業」欄に付けられた情報に基づいて集計したものである。上位 100 を示した。上位のほうには，刑事，警部，警部補，警視などの警察関係の職業や探偵や弁護士など事件に関わる職業が目立つ。これはミステリ小説が多く収録されていることの現れであろう。また，小学生，高校生，大学生，中学生などの学生・生徒も多く登場する。これらは若年層をターゲットにした小説の影響と思われる。

表 4-4 職業と語数

| 順位 | 職業 | 語数 | 割合(%) |
|---|---|---|---|
| 1 | 探偵 | 36760 | 4.53 |
| 2 | 刑事 | 34240 | 4.22 |
| 3 | 小学生 | 32858 | 4.05 |
| 4 | 高校生 | 28832 | 3.55 |
| 5 | 警部 | 23911 | 2.95 |
| 6 | 弁護士 | 18113 | 2.23 |
| 7 | 小説家 | 16751 | 2.06 |
| 8 | 大学生 | 11382 | 1.40 |

| | | | |
|---|---|---|---|
| 9 | 警部補 | 11333 | 1.40 |
| 10 | 社長 | 9951 | 1.23 |
| 11 | 薬剤師 | 9640 | 1.19 |
| 12 | 中学生 | 9365 | 1.15 |
| 13 | 学生 | 8778 | 1.08 |
| 14 | 警視 | 8693 | 1.07 |
| 15 | 公爵 | 8557 | 1.05 |
| 16 | 武将 | 7975 | 0.98 |
| 17 | 検察官 | 7785 | 0.96 |
| 18 | 曹長 | 6966 | 0.86 |
| 19 | 大祭司 | 6502 | 0.80 |
| 20 | 貸本屋／岡っ引き | 6397 | 0.79 |
| 21 | 列車 | 6273 | 0.77 |
| 22 | 同心 | 6260 | 0.77 |
| 23 | 占星術師／探偵 | 6181 | 0.76 |
| 24 | 便利屋 | 6169 | 0.76 |
| 25 | 私立探偵 | 5946 | 0.73 |
| 26 | 会社員 | 5829 | 0.72 |
| 27 | 新聞記者 | 5807 | 0.72 |
| 28 | 大佐 | 5535 | 0.68 |
| 29 | 民宿の主人 | 5503 | 0.68 |
| 30 | 漁師 | 5222 | 0.64 |
| 31 | 大公妃 | 5187 | 0.64 |
| 32 | 犬屋 | 5056 | 0.62 |
| 33 | 医師 | 5012 | 0.62 |
| 34 | 教師 | 4853 | 0.60 |
| 35 | カジノのマネージャー／裏の顔あり | 4746 | 0.58 |
| 36 | ？大学教員 | 4726 | 0.58 |
| 37 | 医者 | 4689 | 0.58 |
| 38 | 大工 | 4637 | 0.57 |
| 39 | 検死官 | 4578 | 0.56 |
| 40 | 画廊 | 4520 | 0.56 |
| 41 | 俳優 | 4423 | 0.54 |
| 42 | 国王 | 4303 | 0.53 |
| 43 | 整体師 | 4221 | 0.52 |
| 44 | 作家 | 4170 | 0.51 |

| | | | |
|---|---|---|---|
| 45 | 紙くず屋 | 4138 | 0.51 |
| 46 | 水茶屋 | 4108 | 0.51 |
| 47 | 検事 | 4049 | 0.50 |
| 48 | マネージャー | 4043 | 0.50 |
| 49 | 経理 | 3971 | 0.49 |
| 50 | 伊達藩の武士 | 3905 | 0.48 |
| 51 | 会社員（企画会社） | 3856 | 0.48 |
| 52 | 元レストランの主人 | 3539 | 0.44 |
| 53 | タクシー運転手 | 3485 | 0.43 |
| 54 | 店の経営者 | 3467 | 0.43 |
| 55 | 警察署長 | 3388 | 0.42 |
| 56 | 武士 | 3373 | 0.42 |
| 57 | 冒険家？ | 3324 | 0.41 |
| 58 | 合衆国陸軍憲兵隊の犯罪捜査官 | 3211 | 0.40 |
| 59 | マネージング・デレクター | 3193 | 0.39 |
| 60 | オウム真理教教団幹部 | 3155 | 0.39 |
| 61 | 詩人 | 3148 | 0.39 |
| 62 | 攻撃長 | 3129 | 0.39 |
| 63 | 馬番 | 3107 | 0.38 |
| 64 | 軍人 | 3094 | 0.38 |
| 65 | 物書き | 3093 | 0.38 |
| 66 | 記者 | 2929 | 0.36 |
| 67 | しろうと探偵 | 2921 | 0.36 |
| 68 | さいはて星の原始人 | 2844 | 0.35 |
| 69 | 営業課長代理 | 2773 | 0.34 |
| 70 | 通信制高校生／探偵 | 2642 | 0.33 |
| 71 | 工作員 | 2635 | 0.32 |
| 72 | 支配人 | 2634 | 0.32 |
| 73 | 探偵小説家 | 2601 | 0.32 |
| 74 | 雑誌記者 | 2592 | 0.32 |
| 75 | 魔術師 | 2564 | 0.32 |
| 76 | 現英貴族？ | 2495 | 0.31 |
| 77 | リポーター | 2466 | 0.30 |
| 78 | 動物園の番人 | 2458 | 0.30 |
| 79 | 准将 | 2455 | 0.30 |
| 80 | 僧 | 2437 | 0.30 |

| | | | |
|---|---|---|---|
| 81 | 艦長 | 2431 | 0.30 |
| 82 | 下宿屋 | 2428 | 0.30 |
| 83 | 盗賊 | 2405 | 0.30 |
| 84 | FBI 捜査官 | 2365 | 0.29 |
| 85 | 魔導師 | 2349 | 0.29 |
| 86 | 暴力団員 | 2338 | 0.29 |
| 87 | パンクバンドのボーカル | 2334 | 0.29 |
| 88 | 使用人 | 2331 | 0.29 |
| 89 | 商売人 | 2314 | 0.29 |
| 90 | 見習い魔女 | 2285 | 0.28 |
| 91 | 馬 | 2248 | 0.28 |
| 92 | 船長 | 2237 | 0.28 |
| 93 | 騎士 | 2219 | 0.27 |
| 94 | 神父 | 2207 | 0.27 |
| 95 | 鑑識課員 | 2205 | 0.27 |
| 96 | 保安官 | 2204 | 0.27 |
| 97 | 王 | 2190 | 0.27 |
| 98 | 当主 | 2168 | 0.27 |
| 99 | 怪盗工作員 | 2136 | 0.26 |
| 100 | 元帥 | 2110 | 0.26 |

## 4.6. 会話モード

表 4-5 は，「会話モード」欄に付けられた情報に基づいて集計したものを全例示した。会話モードとは通常の対面会話ではない会話，共通語を用いていない会話，外国語での会話など，何かしらの特徴を持った会話の場合に，その特徴を記録したものである。

表 4-5 会話モードと語数

| 順位 | 会話モード | 語数 | 割合 (%) |
|---|---|---|---|
| 1 | 電話 | 55980 | 32.49 |
| 2 | 独話 | 32606 | 18.92 |
| 3 | 引用 | 16299 | 9.46 |
| 4 | 方言（関西） | 12061 | 7.00 |
| 5 | 方言 | 11888 | 6.90 |
| 6 | 回想 | 8240 | 4.78 |
| 7 | 通信器 | 7016 | 4.07 |
| 8 | 方言（東京下町） | 6273 | 3.64 |
| 9 | スピーチ | 3635 | 2.11 |

| | | | |
|---|---|---|---|
| 10 | 外国語 | 2944 | 1.71 |
| 11 | 電話／引用 | 1969 | 1.14 |
| 12 | テレビ | 1922 | 1.12 |
| 13 | 方言（大阪） | 1616 | 0.94 |
| 14 | 録音音声 | 1400 | 0.81 |
| 15 | テレパシー？ | 1278 | 0.74 |
| 16 | 方言（熊本） | 1049 | 0.61 |
| 17 | 演技 | 851 | 0.49 |
| 18 | 方言（奈良） | 802 | 0.47 |
| 19 | 方言（仙台） | 542 | 0.31 |
| 20 | テレパシー | 367 | 0.21 |
| 21 | 回想／方言（関西） | 342 | 0.20 |
| 22 | 方言（山形） | 242 | 0.14 |
| 23 | 叫び声 | 231 | 0.13 |
| 24 | 歌 | 208 | 0.12 |
| 25 | 呪文 | 206 | 0.12 |
| 26 | 方言（北海道） | 191 | 0.11 |
| 27 | 鳴き声 | 168 | 0.10 |
| 28 | 夢 | 159 | 0.09 |
| 29 | アナウンス | 141 | 0.08 |
| 30 | 沈黙 | 136 | 0.08 |
| 31 | 掛け声 | 134 | 0.08 |
| 32 | 物真似 | 112 | 0.06 |
| 33 | 方言（京都） | 111 | 0.06 |
| 34 | 笑い声 | 99 | 0.06 |
| 35 | 独話／引用 | 90 | 0.05 |
| 36 | 独話／方言（関西） | 83 | 0.05 |
| 37 | 方言（長野） | 67 | 0.04 |
| 38 | 祈り | 62 | 0.04 |
| 39 | 物真似／方言（関西） | 61 | 0.04 |
| 40 | 外国語（スペイン語） | 56 | 0.03 |
| 40 | 演技／独話 | 56 | 0.03 |
| 42 | 方言（長崎） | 55 | 0.03 |
| 43 | 非言語 | 53 | 0.03 |
| 44 | 泣き声 | 52 | 0.03 |
| 45 | 方言（岡山） | 43 | 0.02 |
| 46 | 外国語（英語） | 42 | 0.02 |
| 46 | 幻聴 | 42 | 0.02 |

| | | | |
|---|---|---|---|
| 48 | 呻き声 | 38 | 0.02 |
| 49 | ラジオ | 35 | 0.02 |
| 50 | 回想／方言 | 33 | 0.02 |
| 51 | 通訳 | 30 | 0.02 |
| 52 | 引用／方言（関西） | 29 | 0.02 |
| 53 | 念仏 | 24 | 0.01 |
| 54 | テレビ／引用 | 18 | 0.01 |
| 55 | 呪文／外国語(スペイン語) | 16 | 0.01 |
| 56 | 外国語（ロシア語） | 15 | 0.01 |
| 56 | 夢／独話 | 15 | 0.01 |
| 58 | 幻聴／笑い声 | 10 | 0.01 |
| 58 | 留守番電話 | 10 | 0.01 |
| 60 | 口パク | 8 | 0.00 |
| 60 | 独話／方言 | 8 | 0.00 |
| 60 | 電話／沈黙 | 8 | 0.00 |
| 63 | 疑問 | 7 | 0.00 |
| 64 | 回想／叫び声 | 6 | 0.00 |
| 64 | 外国人 | 6 | 0.00 |
| 66 | 驚き | 5 | 0.00 |
| 67 | テレビ／笑い声 | 4 | 0.00 |
| 67 | 夢／叫び声 | 4 | 0.00 |
| 67 | 夢／笑い声 | 4 | 0.00 |
| 70 | 驚愕 | 3 | 0.00 |
| 71 | 外国語（フランス語） | 1 | 0.00 |

## 4.7. 会話文を対象とした語彙表

表 4-6 は，会話文のみを対象とした語彙の集計表である。カギ括弧が会話文に含まれているのはもともとの設計によるものである。

表 4-6 会話文を対象とした語彙表（上位 100 語）

| 順位 | 語彙素読み | 語彙素 | 語彙素細分類 | 品詞 | 語種 | 語彙素 ID | 頻度 |
|---|---|---|---|---|---|---|---|
| 1 | | 、 | | 補助記号-読点 | 記号 | 24 | 214452 |
| 2 | | 「 | | 補助記号-括弧開 | 記号 | 33 | 158083 |
| 3 | | 」 | | 補助記号-括弧閉 | 記号 | 34 | 157980 |
| 4 | ダ | だ | | 助動詞 | 和 | 22916 | 137910 |
| 5 | | 。 | | 補助記号-句点 | 記号 | 25 | 121398 |
| 6 | テ | て | | 助詞-接続助詞 | 和 | 24874 | 112147 |
| 7 | ハ | は | | 助詞-係助詞 | 和 | 29321 | 105736 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 8 | ノ | の | | 助詞-格助詞 | 和 | 28989 | 105553 |
| 9 | タ | た | | 助動詞 | 和 | 21642 | 101162 |
| 10 | ニ | に | | 助詞-格助詞 | 和 | 28178 | 97082 |
| 11 | ヲ | を | | 助詞-格助詞 | 和 | 41407 | 74448 |
| 12 | ガ | が | | 助詞-格助詞 | 和 | 7889 | 69633 |
| 13 | ノ | の | | 助詞-準体助詞 | 和 | 28990 | 68272 |
| 14 | スル | 為る | | 動詞-非自立可能 | 和 | 19537 | 59182 |
| 15 | ト | と | | 助詞-格助詞 | 和 | 25826 | 54542 |
| 16 | モ | も | | 助詞-係助詞 | 和 | 37562 | 52041 |
| 17 | デス | です | | 助動詞 | 和 | 25653 | 41176 |
| 18 | デ | で | | 助詞-格助詞 | 和 | 25518 | 39141 |
| 19 | マス | ます | | 助動詞 | 和 | 35697 | 34286 |
| 20 | イル | 居る | | 動詞-非自立可能 | 和 | 2585 | 33511 |
| 21 | | ？ | | 補助記号-句点 | 記号 | 57 | 32534 |
| 22 | ナイ | ない | | 助動詞 | 和 | 27438 | 31200 |
| 23 | カ | か | | 助詞-終助詞 | 和 | 5569 | 29656 |
| 24 | ヨ | よ | | 助詞-終助詞 | 和 | 38980 | 28709 |
| 25 | イウ | 言う | | 動詞-一般 | 和 | 1571 | 28320 |
| 26 | コト | 事 | | 名詞-普通名詞-一般 | 和 | 12836 | 28286 |
| 27 | ナイ | 無い | | 形容詞-非自立可能 | 和 | 27442 | 26109 |
| 28 | | … | | 補助記号-一般 | 記号 | 6 | 25767 |
| 29 | ネ | ね | | 助詞-終助詞 | 和 | 28754 | 22028 |
| 30 | アル | 有る | | 動詞-非自立可能 | 和 | 1216 | 19549 |
| 31 | カ | か | | 助詞-副助詞 | 和 | 5568 | 18657 |
| 32 | ソレ | 其れ | | 代名詞 | 和 | 21461 | 18038 |
| 33 | オ | 御 | | 接頭辞 | 和 | 4350 | 18000 |
| 34 | ナル | 成る | | 動詞-非自立可能 | 和 | 28061 | 16911 |
| 35 | ズ | ず | | 助動詞 | 和 | 19587 | 16228 |
| 36 | ナニ | 何 | | 代名詞 | 和 | 27920 | 16072 |
| 37 | ナ | な | | 助詞-終助詞 | 和 | 27433 | 15972 |
| 38 | サン | さん | | 接尾辞-名詞的-一般 | 和 | 14495 | 15790 |
| 39 | カラ | から | | 助詞-接続助詞 | 和 | 7218 | 15507 |
| 40 | テル | てる | | 助動詞 | 和 | 25287 | 14562 |
| 41 | | ！ | | 補助記号-句点 | 記号 | 45 | 14465 |
| 42 | ソウ | そう | | 副詞 | 和 | 20935 | 13822 |
| 43 | レル | れる | | 助動詞 | 和 | 40741 | 12757 |
| 44 | ヨイ | 良い | | 形容詞-非自立可能 | 和 | 38988 | 12564 |
| 45 | カラ | から | | 助詞-格助詞 | 和 | 7219 | 12023 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 46 | ソノ | 其の | | 連体詞 | 和 | 21394 | 11941 |
| 47 | ガ | が | | 助詞-接続助詞 | 和 | 7888 | 11630 |
| 48 | クル | 来る | | 動詞-非自立可能 | 和 | 10518 | 11471 |
| 49 | ドウ | どう | | 副詞 | 和 | 26969 | 11393 |
| 50 | ッテ | って | | 助詞-副助詞 | 和 | 24319 | 11261 |
| 51 | イク | 行く | | 動詞-非自立可能 | 和 | 1713 | 10502 |
| 52 | オモウ | 思う | | 動詞-一般 | 和 | 5255 | 9473 |
| 53 | ワ | わ | | 助詞-終助詞 | 和 | 41101 | 9381 |
| 54 | アナタ | 貴方 | | 代名詞 | 和 | 894 | 9188 |
| 55 | コノ | 此の | | 連体詞 | 和 | 12905 | 9138 |
| 56 | ワタシ | 私 | | 代名詞 | 和 | 41277 | 9035 |
| 57 | バ | ば | | 助詞-接続助詞 | 和 | 30503 | 8747 |
| 58 | ヨウ | 様 | | 形状詞-助動詞語幹 | 漢 | 39002 | 8536 |
| 59 | | ― | | 補助記号-一般 | 記号 | 1 | 8453 |
| 60 | ノ | の | | 助詞-終助詞 | 和 | 28991 | 8049 |
| 61 | ワタクシ | 私 | 代名詞 | 代名詞 | 和 | 41274 | 7838 |
| 62 | ワカル | 分かる | | 動詞-一般 | 和 | 41180 | 7549 |
| 63 | ミル | 見る | | 動詞-非自立可能 | 和 | 36920 | 7409 |
| 64 | ヤル | 遣る | | 動詞-非自立可能 | 和 | 38541 | 7029 |
| 65 | モノ | 物 | | 名詞-普通名詞-サ変可能 | 和 | 37875 | 6955 |
| 66 | ケレド | けれど | | 助詞-接続助詞 | 和 | 11238 | 6901 |
| 67 | コレ | 此れ | | 代名詞 | 和 | 13095 | 6838 |
| 68 | オレ | 俺 | | 代名詞 | 和 | 5368 | 6641 |
| 69 | ト | と | | 助詞-接続助詞 | 和 | 25825 | 6435 |
| 70 | アノ | 彼の | | 連体詞 | 和 | 912 | 6364 |
| 71 | クレル | 呉れる | | 動詞-非自立可能 | 和 | 10567 | 6073 |
| 72 | タイ | たい | | 助動詞 | 和 | 21652 | 5920 |
| 73 | タチ | 達 | | 接尾辞-名詞的-一般 | 和 | 22373 | 5913 |
| 74 | ヘ | へ | | 助詞-格助詞 | 和 | 33819 | 5788 |
| 75 | ヒト | 人 | | 名詞-普通名詞-一般 | 和 | 31500 | 5764 |
| 76 | ダケ | だけ | | 助詞-副助詞 | 和 | 23122 | 5483 |
| 77 | シル | 知る | | 動詞-一般 | 和 | 17162 | 5340 |
| 78 | マデ | まで | | 助詞-副助詞 | 和 | 35891 | 5297 |
| 78 | モウ | もう | | 副詞 | 和 | 37569 | 5297 |
| 80 | イチ | 一 | | 名詞-数詞 | 漢 | 2050 | 5270 |
| 81 | デキル | 出来る | | 動詞-非自立可能 | 和 | 25624 | 5169 |
| 82 | トコロ | 所 | | 名詞-普通名詞-副詞可能 | 和 | 26373 | 4996 |
| 83 | ボク | 僕 | | 代名詞 | 漢 | 35132 | 4888 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 84 | イマ | 今 | | 名詞-普通名詞-副詞可能 | 和 | 2460 | 4831 |
| 85 | トキ | 時 | | 名詞-普通名詞-副詞可能 | 和 | 26231 | 4687 |
| 86 | オマエ | 御前 | | 代名詞 | 和 | 5201 | 4606 |
| 87 | ラレル | られる | | 助動詞 | 和 | 39787 | 4401 |
| 88 | セル | せる | | 助動詞 | 和 | 20355 | 4387 |
| 89 | | | | 空白 | 記号 | 23 | 4306 |
| 90 | キミ | 君 | 代名詞 | 代名詞 | 和 | 8820 | 4304 |
| 91 | | ・ | | 補助記号-一般 | 記号 | 44 | 4289 |
| 92 | キク | 聞く | | 動詞-一般 | 和 | 8399 | 4285 |
| 93 | ココ | 此処 | | 代名詞 | 和 | 12607 | 4163 |
| 94 | ソンナ | そんな | | 連体詞 | 和 | 21510 | 4142 |
| 95 | ワケ | 訳 | | 名詞-普通名詞-一般 | 和 | 41224 | 4118 |
| 96 | サマ | 様 | | 接尾辞-名詞的-一般 | 和 | 14374 | 4006 |
| 97 | ゴ | 御 | | 接頭辞 | 漢 | 13332 | 3921 |
| 98 | ホウ | 方 | | 名詞-普通名詞-一般 | 漢 | 34378 | 3900 |
| 99 | サ | さ | | 助詞-終助詞 | 和 | 13620 | 3771 |
| 100 | ダレ | 誰 | | 代名詞 | 和 | 23273 | 3533 |

## 5. データ

話者情報のデータとしては 2 種類のファイルを限定公開している。これらは「BCCWJ 関連データ配布サイト」よりダウンロードできる。ただし，対象はデータには原文が含まれるため，有償版の BCCWJ を契約している人に限っている。

以下，readme ファイルから引用する。

BCCWJ の小説会話文に対する話者情報アノテーションデータ ver.1.2

2021.12.16

文責：山崎誠

BCCWJ の図書館サブコーパス（LB）に含まれる小説のサンプルに対して
話者情報を付与したデータです。小説と認定する範囲は，NDC(図書分類番号)が 913, 923, 933 などのように，9 で始まり 3 で終わるもの 2,932 サンプルが対象です。
公開するファイルは 2 つあります。

(1)ファイル名：speakersInfo_LB_Novels_20211216.txt
ファイルの概略：
各サンプルを短単位に分解したファイルに話者情報を付与したものです。
発話箇所に話者情報が付けられています。

行数：11,421,933 行（ヘッダ 1 行を含む）
文字コード：UTF-8，BOM なし。
サイズ：約 1.3GB

列名とその説明

| # | 属性 | 説明 |
|---|---|---|
| 1. | サンプル ID | BCCWJ のサンプル番号 |
| 2. | 文頭ラベル | B：文頭，I：文頭以外 |
| 3. | 書字形出現形 | 原文（NumTrans 後） |
| 4. | 原文文字列 | 原文（NumTrans 前） |
| 5. | 語彙素 ID | 語彙素の ID |
| 6. | 語彙素 | 短単位の形 |
| 7. | 語彙素細分類 | 同じ表記の語彙素を区別する情報 |
| 8. | 語彙素読み | 語彙素の読み（片仮名で表記） |
| 9. | 語種 | 和語・漢語・外来語等の区別 |
| 10. | 品詞 | 品詞（大分類-中分類-小分類） |
| 11. | 連番 | サンプル内での短単位の並び順（10 きざみ） |
| 12. | 文字開始位置 | 原文文字列における，サンプル頭からのオフセット値（10 きざみ） |

13.文字終了位置 原文文字列における，サンプル頭からのオフセット値（10 きざみ）
14.出現形開始位置　　書字形出現形のサンプル頭からのオフセット値（10 きざみ）
15.出現形終了位置　　書字形出現形のサンプル頭からのオフセット値（10 きざみ）
16.話者名　　登場人物の呼び名。話者名の最後にある●は心内発話を表す。
17.性別 男，女，不明。
18.年齢層　　若年層（～19 歳），成年層（20～59 歳），老年層（60 歳以上）
19.非人間　　生物学的な人間以外のものの場合に○。
20.会話モード　通常の対話場面でない場合の情報。「電話，方言，外国人，テレパシー，引用，独話，疑問，沈黙，驚き，驚愕」など
21.会話認定情報 1　　心内発話などの情報
22.会話認定情報 2　　会話に準ずる場合に判定した根拠
23.備考 さまざまな注記
24.職業 発話者の職業
25.相手 会話の相手の名前

(2)ファイル名：speakersInfo_readable.xlsx
ファイルの概略：
各サンプルを文単位で 1 行にして，それに対して話者情報と付与したものです。
発話された文に対して話者情報が付けられています。62MB。

行数：665,828 行（ヘッダ 1 行を含む）
サイズ：約 64MB

列名とその説明
# 属性　　説明
1.No.　一連番号
2.サンプル ID　BCCWJ のサンプル番号
3.タイトル　　サンプルの書籍名
4.著者　サンプルの著者
5.NDC　図書分類（日本十進分類法）
6.原文　文（文頭ラベル B で区切ったもの）
7.話者名 登場人物の呼び名。話者名の最後にある●は心内発話を表す。
8.性別　男，女，不明。
9.年齢層 若年層（～19 歳），成年層（20～59 歳），老年層（60 歳以上）
10.非人間　　生物学的な人間以外のものの場合に○。
11.会話モード　通常の対話場面でない場合の情報。「電話，方言，外国人，テレパシー，引用，独話，疑問，沈黙，驚き，驚愕」など
12.会話認定情報 1　　心内発話などの情報
13.会話認定情報 2　　会話に準ずる場合に判定した根拠

14.備考 さまざまな注記
15.職業 発話者の職業
16.相手 会話の相手の名前

本データは，国立国語研究所のプロジェクト「大規模日常会話コーパスに基づく話し言葉の多角的研究」（プロジェクトリーダー・小磯花絵）および日本学術振興会・科学研究費補助金「会話文への発話者情報の付与によるコーパスの拡張」（15H03212）の成果です。本データを利用して発表を行う場合は，この 2 つのプロジェクトについての謝辞をお書き下さい。

以上

## 6. おわりに

本報告書にまとめた話者情報を活用してさまざな分析ができるものと思う。付与した情報に不統一な部分もあるため，利用者は適宜変更して使っていただいてかまわない。多く活用されることを願う。

### 参考文献


加藤祥，森山奈々美，浅原正幸（2021）『現代日本語書き言葉均衡コーパス』書籍サンプルの NDC 情報増補：NDC 情報を用いた随筆の抽出と文体調査，『国立国語研究所論集』21，pp.65-84. http://doi.org/10.15084/00003437

西部みちる，大島一，間淵洋子，小林正行，田島孝治，高田智和，山口昌也（2011）『現代日本語書き言葉均衡コーパス』における電子化テキストの構築，国立国語研究所.

宮嵜由美，柏野和佳子，山崎誠（2017）発話文への発話者情報付与の基本設計：『現代日本語書き言葉均衡コーパス』収録の小説を対象に，「言語資源活用ワークショップ発表論文集 2016」，pp.38-48. http://doi.org/10.15084/00001456

宮嵜由美，柏野和佳子，山崎誠(2018)『現代日本語書き言葉均衡コーパス』収録の小説における発話箇所認定について，シンポジウム「日常会話会話コーパス」Ⅲポスター

宮嵜由美(2019)「「現代日本語書き言葉均衡コーパス」収録小説を対象とした会話文への発話者情報の付与によるコーパスの拡張方針 ver.6.0」（内部資料）

山崎誠（2018）翻訳小説と日本語小説における会話文の計量語彙論的比較，「語彙研究」，15，pp.1-15.

山崎誠，宮嵜由美，柏野和佳子(2018)「BCCWJ 小説会話文への発話者情報の付与と計量的分析」「計量国語学会第 62 回大会予稿集」pp.13-18.

山崎誠，宮嵜由美，柏野和佳子（2019）「BCCWJ 小説会話文への話者情報の付与とその活用」「言語資源活用ワークショップ発表論文集 2019」，pp.313-320. http://doi.org/10.15084/00002582


### 謝辞


本研究は，国立国語研究所のプロジェクト「大規模日常会話コーパスに基づく話し言葉の多角的研究」（プロジェクトリーダー・小磯花絵）および日本学術振興会・科学研究費補助金「会話文への発話者情報の付与によるコーパスの拡張」(JP15H03212)による成果です。

話者情報の付与作業を行ったのは以下の方々です（五十音順，敬称略）。

柏野和佳子，河野礼実，田嶋明日香，立花幸子，平本智弥，宮嵜由美，牟田浩子，山縣智子，山崎誠

国立国語研究所「日常会話コーパス」プロジェクト報告書 5

# 会話文話者情報の設計

山崎誠・宮嵜由美・柏野和佳子

2022 年 3 月 31 日
国立国語研究所

大学共同利用機関法人 人間文化研究機構
国立国語研究所
National Institute for Japanese Language and Linguistics