

記録票データ・対話記録票データ・講演者属性データ

対話参加講演者の講演一覧の解説

籠宮隆之[†]

[†] 神戸大学大学院

Version 2.0

(記録票データおよび対話記録票データに「収録時の年齢」を追加．2.12 節，3.8 節参照)

Version 1.1

(「講演の自発性」に関する誤りを訂正．2.6.1 節参照)

目次

1 はじめに	1
2 記録票データ	1
3 対話記録票データ	8
4 講演者属性データ	12
5 対話参加講演者の講演一覧	13

1 はじめに

この文書では『日本語話し言葉コーパス』の記録票データ (talk data.csv)，対話記録票データ (dialogue data.csv)，講演者属性データ (speaker data.csv)，対話参加講演者の講演一覧 (correspondence data.csv) について解説する。

2 記録票データ

記録票データ (talk-data.csv) は、以下の 75 フィールドで構成されるタブ区切りのテキストデータである。各講演毎に 1 レコード (行) を構成しており、各講演に対する種々の情報を記録している。

ID 等 講演 ID，講演者 ID，記録票バージョン

収録作業等 収録機材等 聴き手人数，収録作業者人数，印象評定者 ID，テープレコーダ

講演形態等 学会種・模擬講演のテーマ等，学会における講演形式，講演使用器材・配布資料

ノイズ等 マイクにかかる息，録音環境，定常ノイズ源，室内の残響，会場拡声装置，突発的ノイズ発生時間・発生源，備考

評定語選択式印象評定項目 たどたどしい，流暢な，単調な，表情ゆたかな，自信のある，自信の無い，優しい，落ち着いた，落ち着きのない，いらいらした，緊張した，リラックスした，大きい声，小さい声，かすれた声，裏返った声，こもった声，重厚な，軽薄な，若々しい，年寄りみだ，元気のある，元気の無い，聞き取りやすい，聞き取りにくい，生意気な，尊大な，鼻にかかった，高い，低い，きっぱりした，速い，遅い

段階評定式印象評定項目 講演の自発性，難解な専門用語の多少，発話スピード，発音の明瞭さ，方言の多少，発話スタイル

原稿読み上げ 読み上げ

朗読の流暢性 朗読の流暢性

アンケート項目 講演の準備，得手・不得手，講演経験，最終学歴

付加情報の種類等 テストセット，分節音・韻律情報，形態論情報，節単位情報，重要文選択，作業者による自由要約，講演者自身による自由要約，係り受け情報，談話境界情報，集合印象評定，単独印象評定

収録ディスク 収録ディスク

以下、各項目について具体的に説明する。各項目タイトルの後に記された“[\$数字]”はフィールドの番号を示す。

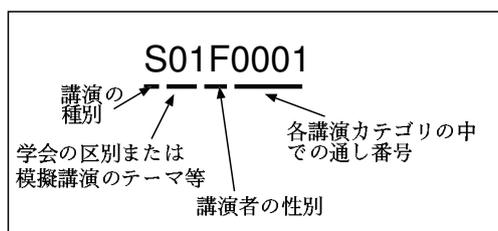


図 1: 講演 ID

2.1 ID 等

2.1.1 講演 ID [\$1]

講演毎に異なる，8桁で構成される ID である．この値が本データのプライマリーキーである．図 1 に示す構造をとる．

講演のカテゴリ 最初のアルファベット 1 文字は，学会講演，模擬講演等の講演種別を表す．

A	学会講演
S	模擬講演
R	朗読
D	対話
M	その他

学会の区別または模擬講演のテーマ等 2～3 桁目の数字 2 桁は，収録した学会の区別（言語学系学会，社会学系学会 etc.），または模擬講演のテーマ（人生を振り返って悲しかったこと，21 世紀に残したいもの etc.）などを表わす．詳細は『音声収録作業の概要』（recording.pdf）を参照．

講演者の性別 4 桁目のアルファベット 1 文字は，講演者の性別を表わす．対話の場合はインタビュイーの性別を表わす．

F	女性
M	男性

各講演カテゴリ中での通し番号 5～8 桁目の数字 4 桁は，学会講演（A）や模擬講演（S）の中での通し番号を表わす．ただし，飛び番号があるので，この番号の最大値は必ずしも各カテゴリの講演数とは一致しない．

2.1.2 講演者 ID [\$2]

講演者毎に割り当てられた，最大 4 桁の数字である．1 人の講演者が複数の講演を行なっている場合もあるが，その場合には同一の講演者 ID が割り当てられている．ただし，飛び番号があるので，この番号の最大値は講演者数とは一致しない．

また，対話の場合はインタビュイーの講演者 ID である．

2.1.3 記録票バージョン [\$3]

収録時に用いた記録票のバージョンを示す．記録票のバージョンにより，以下の項目のうち追加されたものや削除されたものがある．この欄が“-”となっているものは，記録票が用いられなかったことを示す．また，これ以降に示す項目で“-”となっているものは，当該記録票バージョンでは項目がなかったことを示す．なお，記入洩れ，非回答は，“nr”となっている．また，対話に関しては，単独印象評定項目などを別途「対話記録票データ」（dialogue data.csv）に記載した．この項目に関しては値が空白となっている．

ver 1.0	モノログ用記録票 ver 1.0
ver 2.0	モノログ用記録票 ver 2.0
ver 2.1	モノログ用記録票 ver 2.1
ver 3.0	モノログ用記録票 ver 3.0
対話	対話用記録票
朗読 (短)	朗読「宇宙」,「DNA」用記録票（『音声収録作業の概要』 recording.pdf 参照）
朗読 (長)	朗読「再朗読」,「ニュース」,「春の死」,「NEW 文化祭」用記録票 （『音声収録作業の概要』 recording.pdf 参照）

注 講演 ID R00F0407 の再朗読は「朗読 (短)」の記録票を使用した．

2.2 収録作業・収録機材等

2.2.1 聴き手人数 [\$4]

聴衆の数を示す．聴衆の中には，収録作業者の数も含む．なお，模擬講演以外は概算である．

2.2.2 収録作業人数 [\$5]

収録作業者の人数を示す．

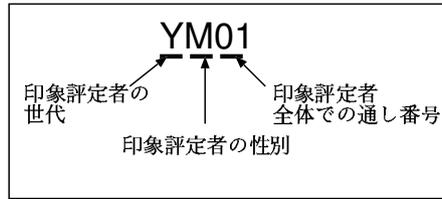


図 2: 印象評定者 ID

2.2.3 印象評定者 ID [\$6]

2.5 節～2.8 節に示す，単独印象評定項目等を記録した作業者の ID を示す．4 桁で構成され，図 2 に示す構造をとる．
印象評定者の世代 単独印象評定項目を記録した収録作業者の世代を示す．

M	30 代後半以降
Y	20 代前半～30 代前半まで

印象評定者の性別 単独印象評定項目を記録したものの性別を示す．

F	女性
M	男性

印象評定者全体での通し番号 単独印象評定項目を記録した作業者全体での通し番号を表わす．飛び番号はない．なお，計 42 名が単独印象評定を行った．

2.2.4 テープレコーダ [\$7]

収録に使用した DAT デッキの型番を示す．いずれも SONY 製の DAT デッキを用いた．DAT への入力は，サンプリング周波数 48000Hz, 量子化ビット数 16bit で行った．

2.3 講演形態等

2.3.1 学会種・模擬講演のテーマ等 [\$8]

節で示した，講演の種別・学会種・模擬講演のテーマ等を示す講演 ID の最初の 3 桁を，具体的に説明したもの．

2.3.2 学会における講演形式 [\$9]

学会講演における，特別講演やシンポジウム形式のものを区別する．

通常の発表 一般向けのシンポジウム 小規模な研究会 パネルディスカッション形式 特別講演	通常の研究発表 一般人を対象としたシンポジウムでの研究発表 参加人数が 10 人以下程度，または内輪での小規模な研究会における研究発表 何人かの講演者が登壇して議論するパネルディスカッション形式での基調講演 学会における特別講演．1 時間以上に及ぶものや，必ずしも学会のテーマとは一致しない講演もある．
--	---

2.3.3 講演使用器材・配布資料 [\$10]

講演者がどのような機材を用いたか，また配付資料等を用いて講演を行ったかを示す．

2.4 ノイズ等

2.4.1 マイクにかかる息 [\$11]

マイクに息がかかる頻度がどのくらいであったかを示す．

1	ほとんどない
2	20～30 秒に一度程度
3	2～3 秒に一度，またはそれ以上．

2.4.2 録音環境 [\$12]

収録会場の全体的な音環境を示す．

1	不良
2	やや不良
3	やや良好
4	良好

2.4.3 定常ノイズ源 [\$13]

収録会場に定常的なノイズがあった場合，そのノイズ源を記録した．

2.4.4 室内の残響 [\$14]

収録会場の残響の強さを示す．

0	残響が顕著でない
1	残響が顕著である

2.4.5 会場拡声装置 [\$15]

会場用のマイクもしくはスピーカーなどの拡声装置が使用されたかどうかを示す。

0	拡声装置無
1	拡声装置有

2.4.6 突発的ノイズ発生時間・発生源 [\$16]

講演収録中に突発的なノイズが発生した場合、その発生時刻と発生源を記録。複数生じた場合には、“;” で区切って列記した。

2.4.7 備考 [\$17]

その他、収録に際して書き留めておくべき備考を記録した。

2.5 評定語選択式印象評定項目

収録者のうち 1 名の印象評定者 (2.2.3 節参照) が、その講演について評定語が「あてはまる」と感じた場合には、その評定語にチェックした。

たどたどしい [\$18], 流暢な [\$19], 単調な [\$20], 表情ゆたかな [\$21], 自信のある [\$22], 自信の無い [\$23] 優しい [\$24], 落ち着いた [\$25], 落ち着きの無い [\$26], いらいらした [\$27], 緊張した [\$28], リラックスした [\$29] 大きい声 [\$30], 小さい声 [\$31], かすれた声 [\$32], 裏返った声 [\$33], こもった声 [\$34], 重厚な [\$35], 軽薄な [\$36] 若々しい [\$37], 年寄りみ [\$38], 元気のある [\$39], 元気の無い [\$40], 聞き取りやすい [\$41], 聞き取りにくい [\$42] 生意気な [\$43], 尊大な [\$44], 鼻にかかった [\$45], 高い [\$46], 低い [\$47], きっぱりした [\$48], 速い [\$49], 遅い [\$50]

上記のフィールドに対して、「チェックされている」「チェックされていない」を記録した。

0	チェックされている
1	チェックされていない

2.6 段階評定式印象評定項目

上記項目同様、収録者のうち 1 名の印象評定者 (2.2.3 節参照) がその講演について評定項目の度合を記録した。

2.6.1 講演の自発性 [\$51]

講演の自発性の高さを示す。記録時には、「原稿を読み上げているかどうか」を主な判断材料にした。

1	自発性が低い
2	自発性がやや低い
3	普通
4	自発性がやや高い
5	自発性が高い

ただし、記録票 Ver.1.0 のものでは、以下の通りである。

1	ほぼすべて朗読
2	多くの部分朗読
4	多くの部分自発的
5	ほぼすべて自発的

2.6.2 難解な専門用語の多少 [\$52]

講演中に難解な専門用語がどのくらいであったかを判断し、記録した。学会等で当該学会の講演内容に習熟しているものも印象評定者として参加したが、その場合でも「学会の講演内容に習熟していない一般人がどう思うか」を判断させた。

1	難解な専門用語が少ない
2	難解な専門用語がやや少ない
3	普通
4	難解な専門用語がやや多い
5	難解な専門用語が多い

2.6.3 発話スピード [\$53]

講演の発話がどのくらいの速さだと感じるかを記録した。

1	遅い
2	やや遅い
3	普通
4	やや速い
5	速い

2.6.4 発音の明瞭さ [\$54]

講演者の発音が明瞭であったかを判断し記録した。なお、印象評定者には「内容」ではなく「発音」で判断するように指示した。

1	不明瞭
2	やや不明瞭
3	普通
4	やや明瞭
5	明瞭

2.6.5 方言の多少 [\$55]

講演がどの程度の方言を含んでいるかを記録した。なお、ここでの判断は分節音や語彙に限り、韻律的特徴については判断をしないように指示した。

1	ほとんどすべて共通語
2	多くの部分が共通語
3	多くの部分が方言
4	ほとんどすべて方言

2.6.6 発話スタイル [\$56]

講演のあらたまり度を記録した。

1	くだけた
2	ややくだけた
3	普通
4	ややあらたまつた
5	あらたまつた

2.7 原稿読み上げ

2.7.1 読み上げ [\$57]

2.6.1 節で示した「講演の自発性」以外に、備考等で「原稿の読み上げであった」と報告があったものについて、まとめた。

読み上げ	用意した原稿を丸読みした発表
やや読み上げ	用意した原稿の読み上げを交えた発表

2.8 朗読の流暢性

2.8.1 朗読の流暢性 [\$58]

朗読音声を収録した際に、どのくらい流暢に朗読できたかを記録した。なお、ここでの「朗読」とは、講演カテゴリーが R のもののことであり、2.7.1 節や 2.6.1 節で「自発性が低い」や「読み上げ」と記録されたものではない。

1	流暢性が低い
2	流暢性がやや低い
3	普通
4	流暢性がやや高い
5	流暢性が高い

2.9 アンケート項目

講演者には事前にアンケートを配付し、「生年月日」「出身地」「最終学歴」「講演の得手・不得手」等の項目に回答させた。このうち、講演毎に異なる可能性のある項目について、以下に記録した。なお、非回答については“nr”、当該講演収録に際してのアンケートで項目がなかった、もしくはアンケートを実施しなかった場合には“-”とした。また、朗読「宇宙」「DNA」の場合も“-”である。

また、講演毎に異なる可能性のない項目については、講演者属性データ (speaker_data.csv) に記録した。

2.9.1 講演の準備 [\$59]

講演に際してどの程度の準備をするか、回答させた。

なし	メモなども用意せずに自由に講演する予定である
メモ	講演用の簡単なメモは用意するが、原稿は準備しない予定である
原稿	講演用の原稿を準備してそれを読み上げる形で講演する予定である

2.9.2 得手・不得手 [\$60]

自分は講演が得意であるかどうかを回答させた。

得意	講演は得意である
やや得意	講演はどちらかと言えば得意である
やや不得手	講演はどちらかと言えば不得手である
不得手	講演は不得手である

2.9.3 講演経験 [\$61]

人前での講演の経験が何回程度あるかを回答させた。

はじめて	はじめてである
5	5 回以下である
10	10 回以下である
20	20 回以下である
21	21 回以上ある

2.9.4 最終学歴 [62]

講演時での最終学歴を「中学卒」「高校卒」「大学学部卒」「大学院修士以上」の中から回答させた。なお、自由回答で「中退」「在学中」と書かれたものがあったが、それらは一つ前の課程を卒業とした。

(例：大学学部在学中 → 高校卒)

修士以上	大学院修士課程卒業以上
学部卒	大学卒業，短大卒業，専門学校卒業，高専卒業
高校卒	高等学校卒業
中学卒	中学校卒業

2.10 付加情報の種類等

『日本語話し言葉コーパス』に含まれる全ての講演に対して、音声、書き起こし、形態素解析結果、節単位認定結果が付与される。これに加えて、一部のデータには分節音ラベルや韻律ラベル、係り受け情報などのデータが付与される。また、形態素解析や節単位認定を自動解析でなく人手でおこなった講演もある。

以下で示す情報は、各講演がどのような付加情報を持っているかの一覧である。

2.10.1 テストセット [63]

講演が音声認識の評価データ用推奨データセットである「テストセット」(『日本語話し言葉コーパス』の概観 overview.pdf 参照)として指定されているかどうかを示す。テストセットは更に以下の3種類に分類される。

セット1 学会講演男性のみ。学会講演に男性講演者が多いので、男性に特化したモデルで評価を行うためのセット。

セット2 学会講演男女同数

セット3 模擬講演

テストセットに指定されている場合は、どのセットであるかも示す。

0	非テストセット
1	セット1
2	セット2
3	セット3

2.10.2 分節音・韻律情報 (コア) [64]

分節音ラベル(『日本語話し言葉コーパス』の分節音ラベリング segment.pdf 参照)や韻律ラベル(『日本語話し言葉コーパス』のイントネーションラベリング intonation.pdf 参照)などの音声ラベルが付与されているかどうかを示す。また、分節音ラベル・韻律ラベルの付与されるものがコアであるので、このフィールドは同時にコアであるかどうかを示す。

なし	分節音・韻律ラベルが付与されない
あり	分節音・韻律ラベルが付与される

2.10.3 形態論情報 [65]

形態論情報の付与(『日本語話し言葉コーパス』の形態論情報の概要 pos.pdf 参照)が自動解析による付与か、人手での解析による付与かを示す。

自動	自動解析による形態論情報付与
手動	人手での解析による形態論情報付与

2.10.4 節単位情報 [66]

節単位情報の付与(『日本語話し言葉コーパス』における節単位認定 clause.pdf 参照)が自動解析による付与か、人手での解析による付与かを示す。

自動	自動解析による節単位情報付与
手動	人手での解析による節単位情報付与

2.10.5 重要文選択 [67]

重要文選択情報(『日本語話し言葉コーパス』における自由要約・重要文抽出データについて summarydata.pdf 参照)が付与されているかどうかを示す。

なし	重要文選択情報が付与されていない
あり	重要文選択情報が付与されている

2.10.6 作業による自由要約 [68]

自由要約データ(『日本語話し言葉コーパス』における自由要約・重要文抽出データについて summarydata.pdf 参照)のうち、作業による自由要約データが作成されたかどうかを示す。

なし	作業による自由要約データが作成されていない
あり	作業による自由要約データが作成された

2.10.7 講演者自身による自由要約 [69]

自由要約データ（『日本語話し言葉コーパス』における自由要約・重要文抽出データについて」summarydata.pdf 参照）のうち、講演者自身による自由要約データが付与されているかどうかを示す。また、講演者自身による自由要約データが「10%要約のみ」、「50%要約のみ」の場合には、それを示す。

なし	講演者自身による自由要約データ作成が作成されていない
あり	講演者自身による自由要約データ作成対象。50%要約、10%要約ともに作成された。
10%のみ	講演者自身による自由要約データが、10%要約のみ作成された。
50%のみ	講演者自身による自由要約データが、50%要約のみ作成された。

2.10.8 係り受け情報 [70]

係り受け情報（『日本語話し言葉コーパス』における係り受け構造付与」dependency.pdf 参照）が付与されたかどうかを示す。

なし	係り受け情報が付与されていない
あり	係り受け情報が付与されている

2.10.9 談話境界情報 [71]

談話境界情報（『日本語話し言葉コーパス』の談話境界情報について」discourse.pdf 参照）が付与されたかどうかを示す。

なし	談話境界情報が付与されていない
あり	談話境界情報が付与されている

2.10.10 集合印象評価 [72]

集合評価による印象評価データ（『印象評価データの概要』impression.pdf 参照）作成対象であるかどうかを示す。集合印象評価は「講演の冒頭」「講演の中盤」「講演の終盤」の3箇所に対して評価作業を行ったが、非コアの一部から「講演の中盤」のみを評価させたものもある。集合評価による印象評価データ作成対象である場合には、この区別も示す。

なし	集合印象評価データ作成非対象
1箇所呈示	中盤のみ対象
3箇所呈示	「冒頭」「中盤」「終盤」の3箇所を対象

2.10.11 単独印象評価 [73]

講演時の印象評価である単独評価データ（『印象評価データの概要』impression.pdf 参照）の有無を示す。単独評価データは、収録時に用いた記録票のバージョンにより、若干内容が異なる。単独評価データが付与される場合には、この区別も示す。また朗読「宇宙」「DNA」では単独評価は行っていない。

1.0 以下の項目が欠如

優しい、鼻にかかった、高い、低い、きっぱりした、難解な専門用語の多少、発話スピード、発音の明瞭さ、発話スタイル

2.1・3.0 以下の項目が欠如

速い、遅い

2.0 欠如なし

対話 対話用の記録票。インタビュアー・インタビュイーの双方を記録するため、単独評価項目は別途対話記録票に記載してある。また、以下の項目が欠如。

講演の自発性、難解な専門用語の多少、発話スピード、発音の明瞭さ、方言の多少、発話スタイル

朗読(長) 再朗読、春の死、NEW文化祭、ニュース用記録票。以下の項目が欠如。

速い、遅い、講演の自発性、難解な専門用語の多少、方言の多少、発話スタイル

なし	単独評価データなし
パターン1	記録票 ver1.0
パターン2	記録票 ver2.1
パターン3	記録票 ver2.0・3.0
対話用	対話（別途対話記録票に記載）
朗読用	朗読(長)

2.11 収録ディスク

2.11.1 収録ディスク [74]

配付 DVD-ROM において収録されているディスクを示す。

2.12 収録時の年齢

2.12.1 収録時の年齢 [75]

『日本語話し言葉コーパス』第2版で追加したデータである。収録年月日（非公開データ）における話者の年齢を記録している。4.3節で述べる話者の生年代同様に、5年ごとに区切っている。例えば、23歳であれば「20to24」、38歳であれば「35to39」のように記載した。また、収録時の年齢が不明の場合には「nr」とした。

なお、複数の講演や対話を収録している話者は、それぞれの収録時の年齢が異なっている場合がある。

3 対話記録票データ

対話記録票データ (dialogue-data.csv) は、以下の 95 フィールドからなるタブ区切りのテキストデータである。各対話毎に 1 レコード (行) を構成しており、各対話に対する種々の情報を記録している。

ID 等 講演 ID, 講演者 ID(インタビュイー), 講演者 ID(インタビュアー)

収録作業員 収録作業員人数, 印象評定者 ID

対話の種類 対話の種類

ノイズ等 マイクにかかる息, 録音環境, 定常ノイズ源, 突発的ノイズ発生時間・発生源, 備考

インタビュイーへの評定語選択式印象評定項目 たどたどしい(インタビュイー), 流暢な(インタビュイー), 単調な(インタビュイー), 表情ゆたかな(インタビュイー), 自信のある(インタビュイー), 自信の無い(インタビュイー), 優しい(インタビュイー), 落ち着いた(インタビュイー), 落ち着きの無い(インタビュイー), いらいらした(インタビュイー), 緊張した(インタビュイー), リラックスした(インタビュイー), 大きい声(インタビュイー), 小さい声(インタビュイー), かすれた声(インタビュイー), 裏返った声(インタビュイー), こもった声(インタビュイー), 重厚な(インタビュイー), 軽薄な(インタビュイー), 若々しい(インタビュイー), 年寄りみだ(インタビュイー), 元気のある(インタビュイー), 元気の無い(インタビュイー), 聞き取りやすい(インタビュイー), 聞き取りにくい(インタビュイー), 生意気な(インタビュイー), 尊大な(インタビュイー), 鼻にかかった(インタビュイー), 高い(インタビュイー), 低い(インタビュイー), きっぱりした(インタビュイー), 速い(インタビュイー), 遅い(インタビュイー)

インタビュアーへの評定語選択式印象評定項目 たどたどしい(インタビュアー), 流暢な(インタビュアー), 単調な(インタビュアー), 表情ゆたかな(インタビュアー), 自信のある(インタビュアー), 自信の無い(インタビュアー), 優しい(インタビュアー), 落ち着いた(インタビュアー), 落ち着きの無い(インタビュアー), いらいらした(インタビュアー), 緊張した(インタビュアー), リラックスした(インタビュアー), 大きい声(インタビュアー), 小さい声(インタビュアー), かすれた声(インタビュアー), 裏返った声(インタビュアー), こもった声(インタビュアー), 重厚な(インタビュアー), 軽薄な(インタビュアー), 若々しい(インタビュアー), 年寄りみだ(インタビュアー), 元気のある(インタビュアー), 元気の無い(インタビュアー), 聞き取りやすい(インタビュアー), 聞き取りにくい(インタビュアー), 生意気な(インタビュアー), 尊大な(インタビュアー), 鼻にかかった(インタビュアー), 高い(インタビュアー), 低い(インタビュアー), きっぱりした(インタビュアー), 速い(インタビュアー), 遅い(インタビュアー)

段階評定式印象評定項目

0-3 分 発話時間の割合 (0-3 分), 難解な専門用語の多少 (0-3 分), 対話全体の発話速度 (0-3 分), 切替の頻度 (0-3 分), インタビュイーの応答速度 (0-3 分), 対話のスタイル (0-3 分)

4-7 分 発話時間の割合 (4-7 分), 難解な専門用語の多少 (4-7 分), 対話全体の発話速度 (4-7 分), 切替の頻度 (4-7 分), インタビュイーの応答速度 (4-7 分), 対話のスタイル (4-7 分)

8 分- 発話時間の割合 (8 分-), 難解な専門用語の多少 (8 分-), 対話全体の発話速度 (8 分-), 切替の頻度 (8 分-), インタビュイーの応答速度 (8 分-), 対話のスタイル (8 分-)

以下、各項目について具体的に説明する。各項目タイトルの後に記された“[\$数字]”はフィールドの番号を示す。

3.1 ID 等

3.1.1 講演 ID [\$1]

2.3.1 節で示した、講演 ID である。この値が本データのプライマリキーである。

3.1.2 講演者 ID (インタビュイー) [\$2]

2.1.2 節で示した、インタビュイーの講演者 ID である。

3.1.3 講演者 ID (インタビュアー) [\$3]

2.1.2 節で示した、インタビュアーの講演者 ID である。

3.2 収録作業員

3.2.1 収録作業員人数 [\$4]

収録作業員の人数を示す。

3.2.2 印象評定者 ID [\$5]

2.2.3 節で示した、単独印象評定項目の印象評定者 ID である。

3.3 対話の種類

3.3.1 対話の種類 [\$6]

以下の 4 つのインタビュー種を記述した。また、「模擬講演に対するインタビュー」と「学会講演に対するインタビュー」には、インタビュー対象とした講演 ID も記載した。

- 模擬講演に対するインタビュー
- 課題指向対話 (ギャラタスク)
- 自由対話
- 学会講演に対するインタビュー

模擬講演インタビュー 課題指向対話 自由対話 学会講演インタビュー	模擬講演に対するインタビュー 課題指向対話（ギャラタスク） 自由対話 学会講演に対するインタビュー
--	--

3.4 ノイズ等

3.4.1 マイクにかかる息 [\\$7]

マイクに息がかかる頻度がどのくらいであったかを示す。

1	ほとんどない
2	20～30秒に一度程度
3	2～3秒に一度、またはそれ以上。

3.4.2 録音環境 [\\$8]

収録ブースの全体的な音環境を示す。

1	不良
2	やや不良
3	やや良好
4	良好

3.4.3 定常ノイズ源 [\\$9]

収録ブースに定常的なノイズがあった場合、そのノイズ源を記録した。

3.4.4 突発的ノイズ発生時間・発生源 [\\$10]

収録中に突発的なノイズが発生した場合、その発生時刻と発生源を記録。複数生じた場合には、“;” で区切って列記した。

3.4.5 備考 [\\$11]

その他、収録に際して書き留めておくべき備考を記録した。

3.5 インタビュイーへの評定語選択式印象評定項目

収録者のうち1名の印象評定者（3.2.2節参照）が、インタビュイーについて評定語が「あてはまる」と感じた場合には、その評定語にチェックした。

たどたどしい(インタビュイー)[\\$12], 流暢な(インタビュイー)[\\$13]
 単調な(インタビュイー)[\\$14], 表情ゆたかな(インタビュイー)[\\$15]
 自信のある(インタビュイー)[\\$16], 自信の無い(インタビュイー)[\\$17], 優しい(インタビュイー)[\\$18]
 落ち着いた(インタビュイー)[\\$19], 落ち着きの無い(インタビュイー)[\\$20], いらいらした(インタビュイー)[\\$21]
 緊張した(インタビュイー)[\\$22], リラックスした(インタビュイー)[\\$23]
 大きい声(インタビュイー)[\\$24], 小さい声(インタビュイー)[\\$25]
 かすれた声(インタビュイー)[\\$26], 裏返った声(インタビュイー)[\\$27], こもった声(インタビュイー)[\\$28]
 重厚な(インタビュイー)[\\$29], 軽薄な(インタビュイー)[\\$30],
 若々しい(インタビュイー)[\\$31], 年寄りみ(インタビュイー)[\\$32]
 元気のある(インタビュイー)[\\$33], 元気の無い(インタビュイー)[\\$34],
 聞き取りやすい(インタビュイー)[\\$35], 聞き取りにくい(インタビュイー)[\\$36]
 生意気な(インタビュイー)[\\$37], 尊大な(インタビュイー)[\\$38], 鼻にかかった(インタビュイー)[\\$39]
 高い(インタビュイー)[\\$40], 低い(インタビュイー)[\\$41], きっぱりした(インタビュイー)[\\$42]
 速い(インタビュイー)[\\$43], 遅い(インタビュイー)[\\$44]

上記のフィールドに対して、「チェックされている」「チェックされていない」を記録した。

0	チェックされていない
1	チェックされている

3.6 インタビュアーへの評定語選択式印象評定項目

インタビュイーへの評定語選択式印象評定項目同様、インタビュアーに対して評定語が「あてはまる」かチェックした。

たどたどしい(インタビュアー)[\\$45], 流暢な(インタビュアー)[\\$46]
 単調な(インタビュアー)[\\$47], 表情ゆたかな(インタビュアー)[\\$48]
 自信のある(インタビュアー)[\\$49], 自信の無い(インタビュアー)[\\$50], 優しい(インタビュアー)[\\$51]
 落ち着いた(インタビュアー)[\\$52], 落ち着きの無い(インタビュアー)[\\$53], いらいらした(インタビュアー)[\\$54]
 緊張した(インタビュアー)[\\$55], リラックスした(インタビュアー)[\\$56]
 大きい声(インタビュアー)[\\$57], 小さい声(インタビュアー)[\\$58]
 かすれた声(インタビュアー)[\\$59], 裏返った声(インタビュアー)[\\$60], こもった声(インタビュアー)[\\$61]
 重厚な(インタビュアー)[\\$62], 軽薄な(インタビュアー)[\\$63],
 若々しい(インタビュアー)[\\$64], 年寄りみ(インタビュアー)[\\$65]
 元気のある(インタビュアー)[\\$66], 元気の無い(インタビュアー)[\\$67],
 聞き取りやすい(インタビュアー)[\\$68], 聞き取りにくい(インタビュアー)[\\$69]
 生意気な(インタビュアー)[\\$70], 尊大な(インタビュアー)[\\$71], 鼻にかかった(インタビュアー)[\\$72]
 高い(インタビュアー)[\\$73], 低い(インタビュアー)[\\$74], きっぱりした(インタビュアー)[\\$75]
 速い(インタビュアー)[\\$76], 遅い(インタビュアー)[\\$77]

上記のフィールドに対して、「チェックされている」「チェックされていない」を記録した。

0	チェックされている
1	チェックされていない

3.7 段階評定式印象評定項目

上記項目同様、収録者のうち1名の印象評定者(3.2.2節参照)がその対話について評定項目の度を記録した。記録する際には、「対話開始後0分から3分まで」、「4分から7分まで」、「8分以降」の3箇所について記録した。

3.7.1 発話時間の割合(0-3分)[\$78]

インタビュアーとインタビュイーがどのくらいの割合で発話しているかを記録した。

1	インタビュアーの発話がほとんど
2	インタビュアーの発話がやや多い
3	インタビュアーとインタビュイーの発話が同じくらい
4	インタビュイーの発話がやや多い
5	インタビュイーの発話がほとんど

3.7.2 難解な専門用語の多少(0-3分)[\$79]

対話中に難解な専門用語がどのくらいであったかを判断し、記録した。

1	難解な専門用語が少ない
2	難解な専門用語がやや少ない
3	普通
4	難解な専門用語がやや多い
5	難解な専門用語が多い

3.7.3 対話全体の発話速度(0-3分)[\$80]

インタビュアー・インタビュイーを区別せずに、対話全体の発話速度がどの程度に感じたかを記録した。

1	遅い
2	やや遅い
3	普通
4	やや速い
5	速い

3.7.4 切替の頻度(0-3分)[\$81]

話者交替の頻度がどのくらいであったかを記録した。

1	少ない
2	やや少ない
3	普通
4	やや多い
5	多い

3.7.5 インタビュイーの応答速度(0-3分)[\$82]

インタビュアーから質問があった場合に、インタビュイーがどのくらいの速さで答えていたかを記録した。

1	遅い
2	やや遅い
3	普通
4	やや速い
5	速い

3.7.6 対話のスタイル(0-3分)[\$83]

対話のあらたまり度を記録した。

1	くだけた
2	ややくだけた
3	普通
4	ややあらたまった
5	あらたまった

3.7.7 発話時間の割合(4-7分)[\$84]

3.7.1節に同じ。

3.7.8 難解な専門用語の多少(4-7分)[\$85]

3.7.2節に同じ。

3.7.9 対話全体の発話速度(4-7分)[\$86]

3.7.3節に同じ。

3.7.10 切替の頻度(4-7分)[\$87]

3.7.4節に同じ。

- 3.7.11 インタビュイアの応答速度 (4-7 分) [\$88]
 - 3.7.5 節に同じ。
- 3.7.12 対話のスタイル (4-7 分) [\$89]
 - 3.7.6 節に同じ。
- 3.7.13 発話時間の割合 (8-分) [\$90]
 - 3.7.1 節に同じ。
- 3.7.14 難解な専門用語の多少 (8-分) [\$91]
 - 3.7.2 節に同じ。
- 3.7.15 対話全体の発話速度 (8-分) [\$92]
 - 3.7.3 節に同じ。
- 3.7.16 切替の頻度 (8-分) [\$93]
 - 3.7.4 節に同じ。
- 3.7.17 インタビュイアの応答速度 (8-分) [\$94]
 - 3.7.5 節に同じ。
- 3.7.18 対話のスタイル (8-分) [\$95]
 - 3.7.6 節に同じ。
- 3.8 収録時の年齢
 - 3.8.1 収録時の年齢 (インタビュイア) [\$96]
 - 収録時の、インタビュイアの年齢である。詳細は 2.12.1 節に同じ。
 - 3.8.2 収録時の年齢 (インタビュアー) [\$97]
 - 収録時の、インタビュアーの年齢である。詳細は 2.12.1 節に同じ。

4 講演者属性データ

講演者属性データ (speaker-data.csv) は、以下の 11 フィールドからなるタブ区切りのテキストファイルである。各講演者毎に 1 レコード (行) から構成され、講演者に関する種々の情報を記録している。

講演者 ID, 性別, 生年代, 出生地, 居住年数, 居住年数 (言語形成期), 父出身地, 母出身地 最終学歴, 備考, 講演 ID

講演者属性データは、講演者へのアンケートを基に作成した。講演者には講演に先立ってアンケートを配付し、「生年月日」「出身地」「最終学歴」「講演の得手・不得手」等の項目に回答させた。このうち、出身地、生年などの情報を講演者属性データに記録した。なお、非回答については“nr”とした。また、当該講演収録に際してのアンケートで項目がなかった場合には“-”とした。

以下、各項目について具体的に説明する。各項目タイトルの後に記された“[\$数字]”はフィールドの番号を示す。

4.1 講演者 ID [\$1]

講演者毎に与えられた最大 4 桁の数字からなる ID。この値が本データのプライマリキーである。

4.2 性別 [\$2]

講演者の性別を表わす。

4.3 生年代 [\$3]

講演者の出生年代を表わす。出生年代は、5 年ごとに区切ってまとめた。例えば、1974 年生まれであれば「70to74」、1965 年生まれであれば「65to69」とした。

なお、2.12.1 節に記したように、「記録票データ (talk_data.csv)」には「収録時の年齢」が記録されている。複数の講演を収録している話者は、講演ごとに「収録時の年齢」が異なる場合がある。これに対し、「生年代」は話者ごとに不変である。

4.4 出生地 [\$4]

講演者の出生地を、都道府県または国名で記録した。

4.5 居住年数 [\$5]

都道府県単位で、講演者の居住歴を記録した。ただし、都道府県のうち、東京都・神奈川県・千葉県・埼玉県は「首都圏」とした。「居住地:居住年数」(例 東京都:18 年) の形式である。何箇所かに居住経験がある場合には、“;” 区切りで列記。また、居住年数が 1 年に満たない場合には「0 年」とした。また、居住地のみ回答され、居住年数が回答されていないものに関しては、“*” で記した。

なお、アンケートでの回答の方法によっては、全ての居住年数を足しても講演時の実年齢になるとは限らない。

4.6 居住年数 (言語形成期) [\$6]

上記居住年数のうち、0~12 歳までの時期を抜粋したもの。

4.7 父出身地 [\$7]

父親の出身地を、都道府県単位または国名で記録した。

4.8 母出身地 [\$8]

父親の出身地を、都道府県単位または国名で記録した。

4.9 最終学歴 [\$9]

講演時での最終学歴を「中学卒」「高校卒」「大学学部卒」「大学院修士以上」の中から回答させた。なお、自由回答で「中退」「在学中」と書かれたものがあったが、それらは一つ前の課程を卒業とした。

(例: 大学学部在学中 → 高校卒)

記録票データの「最終学歴」項目と同じ書式であるが、こちらは講演者が複数の講演を行っている場合には、最後の講演を収録した時点でのアンケートに回答があった最終学歴を掲載した。

修士以上	大学院修士課程卒業以上
学部卒	大学卒業, 短大卒業, 専門学校卒業, 高専卒業
高校卒	高等学校卒業
中学卒	中学校卒業

4.10 備考 [\$10]

講演時の記録票に記述されていた、講演者の属性に関する備考を記した。

4.11 講演 ID [\$11]

講演したデータの講演 ID を列記した。複数あるものは、“;” で区切って列記した。

5 対話参加講演者の講演一覧

対話参加講演者の講演一覧 (correspondence-data.csv) は、以下の 12 フィールドからなるタブ区切りのテキストデータである。各対話参加講演者毎に 1 レコード (行) からなる。

講演者 ID (講演者), 性別, 生年代, 講演者 ID (インタビュアー),
基にした学会講演, 基にした模擬講演, 再朗読, 学会講演インタビュー
インタビュー用の模擬講演, 模擬講演インタビュー, 課題指向対話, 自由対話

学会講演に対するインタビューや模擬講演に対するインタビューは、別の講演の内容を基にしている。また、対話参加講演者は、全て再朗読を行っている。このように密接に内容の結び付いた講演の対応を、分かりやすくまとめた。以下、各項目について具体的に説明する。各項目タイトルの後に記された“[\$数字]”はフィールドの番号を示す。

5.1 講演者 ID (講演者) [\$1]

2.1.2 節で述べた、講演者 ID である。講演者 (インタビュアー) の ID である。

5.2 性別 [\$2]

講演者の性別を記した。

5.3 生年代 [\$3]

4.3 節で述べた、講演者の生年代である。

5.4 講演者 ID (インタビュアー) [\$4]

2.1.2 節で述べた、講演者 ID である。インタビュアーの ID である。y

5.5 基にした学会講演 [\$5]

学会講演参加者から募集した対話参加者は、ある一つの学会講演を基に再朗読や学会講演インタビューを行っている。その学会講演の講演 ID である。

5.6 基にした模擬講演 [\$6]

模擬講演参加者から募集した対話参加者は、模擬講演の再朗読を行っている。その模擬講演の講演 ID である。

5.7 再朗読 [\$7]

5.5 節や 5.6 節の学会講演や模擬講演を再朗読したものの講演 ID である。

5.8 学会講演インタビュー [\$8]

5.5 節の学会講演に対するインタビューの講演 ID である。

5.9 インタビュー用の模擬講演 [\$9]

全ての対話参加者は新たに模擬講演を行い、その内容についてインタビューを受けた。そのインタビューの題材となった模擬講演の講演 ID である。

5.10 模擬講演インタビュー [\$10]

5.9 節の模擬講演に対するインタビューの講演 ID である。

5.11 課題指向対話 [\$11]

参加した課題指向対話の講演 ID である。

5.12 自由対話 [\$12]

参加した自由対話の講演 ID である。