# 日本語話題別会話コーパス：J-TOCC　解説資料

中俣尚己（京都教育大学）
http://nakamata.info
2021/02/28
最終改訂:2022/08/29



## １．　はじめに

『日本語話題別会話コーパス：J-TOCC』は話題を固定し、各話題について等しい時間の、親しい人間どうしの１対１会話を録音、文字化したコーパスであり、15 話題につきそれぞれ 120 ペア×5 分＝10 時間、合計で 150 時間分の会話を文字化した。録音は 2018 年から 2019 年にかけて行われた。

『日本語話題別会話コーパス：J-TOCC』は JSPS 科研費 18H00676「話題が語彙・文法・談話ストラテジーに与える影響の解明」（以下、本プロジェクト）の成果として構築された。プロジェクトの代表者は京都教育大学の中俣尚己である。

本コーパスの名称は『日本語話題別会話コーパス』で、略称は『J-TOCC』である。J-TOCC は Japanese Topic-Oriented Conversation Corpus の略であり、「ジェイトック」と読む。J-TOCC の著作権は中俣尚己が保有する。

## ２．　利用方法

J-TOCC は中俣尚己のウェブサイトからダウンロードする(http://nakamata.info/database/index.html)。ダウンロードする際には、同ページ内の利用規約に同意する必要がある。J-TOCC は研究・教育目的に限り利用することができる。利用範囲は個人を原則とするが、①複数の作業者で共同研究・開発を行う場合、②授業で利用する場合には、特例としてデータの複製を許可する。また、この場合、利用登録時に複製する利用者の大よその数を申請する必要がある。申請は授業の年度ごと、共同利用する研究プロジェクトごとに行わなければならない。また、本コーパスの全部または一部を用いた研究を行い、成果や知見を公開する時は『日本語話題別会話コーパス』を利用したことを明記する必要がある。また、以下の文献のいずれかを引用する必要がある。中俣尚己(2021)は本文書と同一のものである。

- 中俣尚己(2021)「日本語話題別会話コーパス：J-TOCC　解説資料」http://nakamata.info/database/j_tocc_document.pdf
- 中俣尚己・太田陽子・加藤恵梨・澤田浩子・清水由貴子・森篤嗣(2021)「『日本語話題別会話コーパス：J-TOCC』」『計量国語学』33 巻 1 号, pp.11-21, 計量国語学会.

J-TOCC のバージョンについては展開したフォルダ名に含まれる 8 桁の数字がバージョン名となる。「J-TOCC20210831」である場合、バージョンは 20210831 である。

発表後または公開後で構わないので、 原稿のコピーなどを代表者の中俣にお送り頂きたい。

## ３．『日本語話題別会話コーパス』(J-TOCC)の構成

『日本語話題別会話コーパス』を展開すると、corpus というフォルダと documents というフォルダがある。corpus フォルダを開くと、表 1 に示すような 15 の話題別フォルダが表示される。

表 1　J-TOCC の 15 の話題

| |
|---|
| 「01.食べること」「02.ファッション」「03.旅行」「04.スポーツ」「05.マンガ・ゲーム」 |
| 「06.家事」「07.学校」「08.スマートフォン」「09.アルバイト」「10.動物」 |
| 「11.天気」「12.夢・将来設計」「13.マナー」「14.住環境」「15.日本の未来」 |

各フォルダの中には「E.東日本」「W.西日本」という 2 つの録音地フォルダがあり、さらにそれぞれの中に「1.男男」「2.男女」「3.女女」という話者の性別による 3 つのフォルダがある。そして、それぞれのフォルダに 20 のファイルがある。これを図示すると、図 1 のようになる。

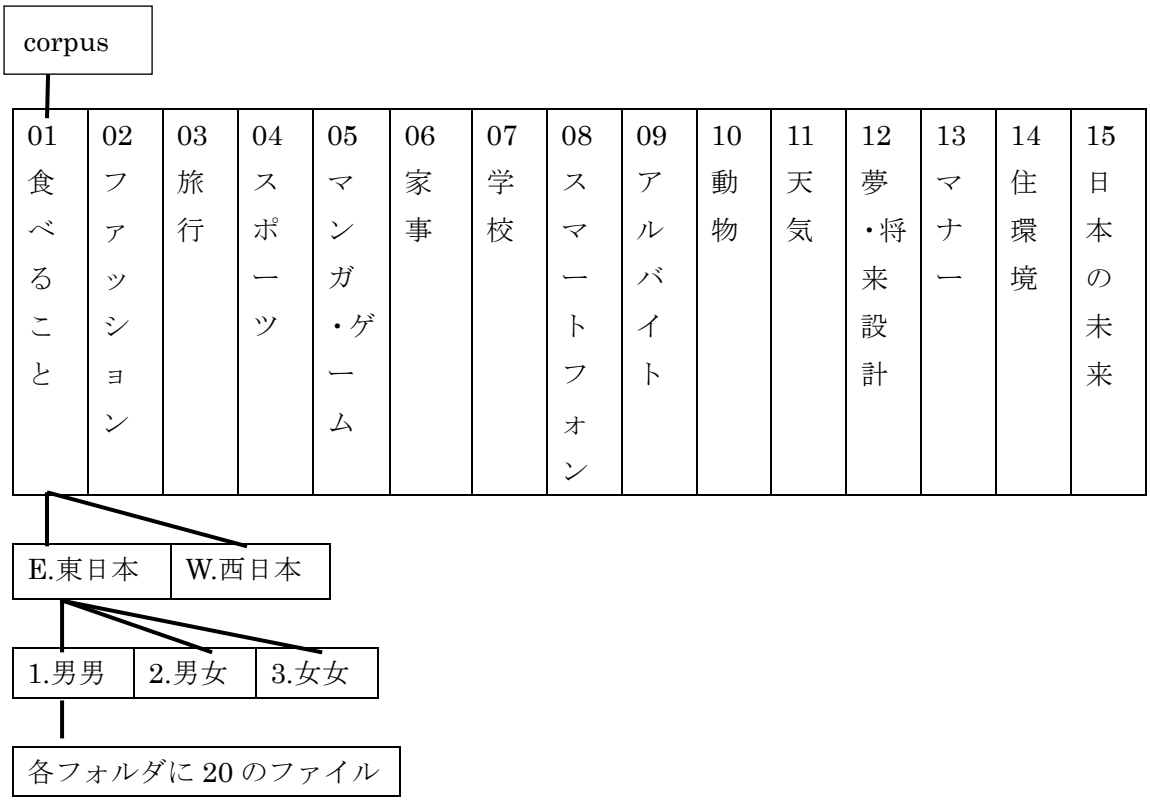
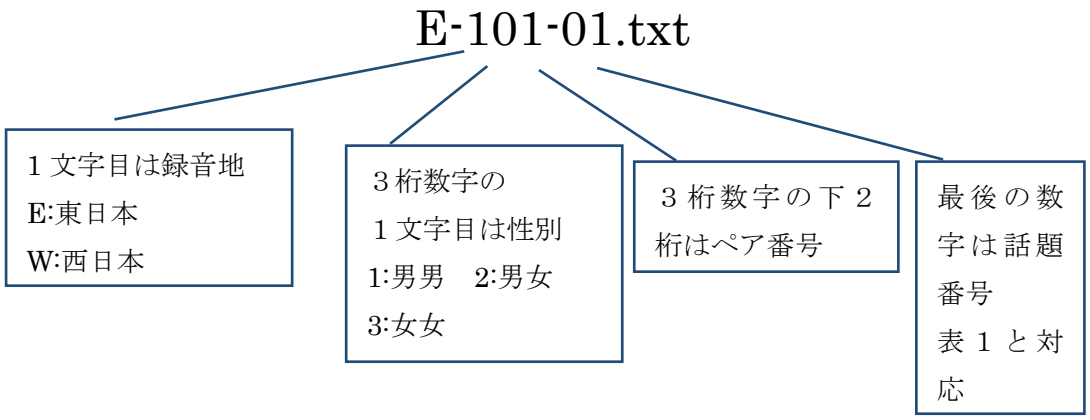



図１　J-TOCC の構造

コーパスの総ファイル数は 20×3×2×15=1,800 ファイルである。全ての話題について 120 のファイル、10 時間分のデータがあることが特徴である。

また、ファイルの命名規則は「録音地記号-性別ペア番号-話題番号」となっている。



## ３．話題の選定基準

話題選定の基準としては前述の山内（編）（2013）に掲載されている 100 話題を出発点とし, コーパスサイズや教材としてのまとまりから 15 前後を目標にすることにした. 山内（編）（2013:11）は各話題で使われている名詞を身近で会話に必ず必要な A レベルから，身近ではなく，抽象度・専門度が高い C レベルに分けている．さらに，山内(2018:5)ではこの情報に従い，親密度・必要度によって話題を三段階に分けている．そして，｛(A の名詞数）×1 点＋（B の名詞数（×0.5 点）＋（C の名詞数）× 0 点)｝÷総名詞数×100 の計算式で「親密度・必要度」を計算し，この値が 50%以上のものを親密度・必要度 I，42%～49%のものを親密度・必要度 II，41%以下のものを親密度・必要度 III としている．この結果に従い，まずは親密度・必要度 III の 48 話題を除外した.

残る 52 話題を以下の 4 つの観点から議論し，11 の「身の回りの話題」を選んだ.

a. 大学生にとって身近である。
b. 初級日本語学習者むけとしてふさわしい。
c. プライバシー上問題となる情報が多く出てこない。
d. 他の話題と近すぎない。

決定方針としては消去法的であり、例えば「音楽」の話題は身近であるものの、西日本地区で予備調査を行った際に「テレテレテレレー」のように歌を歌う場面が見られ、文字起こしでは意味がわからない恐れがあること、また仮に発言に歌詞が含まれた場合には別途著作権処理が必要となる恐れがあるため、採用しなかった。このような方法でまず 11 の話題が選ばれた。また、山内(2013)では「家電・機械」といった話題があるが、大学生にとっては最も身近な機会である「スマートフォン」に絞るなど、大学生に合わせて変更もしている。

この 11 の話題は「a.大学生にとって身近である」を満たすため、すべてがヨーロッパ言語共通参照枠（CEFR）など学習者のレベルにかかわる定義文でいうところの「身の回りの話題」である。一方で、「様々な話題」を扱うコーパスを構築する目的を考えた場合、「社会にかかわる話題」「話し合い活動で使われるような話題」「価値観を表明したり、あるいは対立したりするような話題」「じっくり長く話せる話題」も含めて話題に幅を持たせる必要もあった。そこで、やや複雑で身近でないテーマも含む 4 話題を追加で選定した。この追加選定の際には親密度・必要度 III の話題からも敢えて選んでいる。

なお、当初は「身の回りの話題」の録音時間は 5 分間、「社会にかかわる話題」の録音時間は 10 分間と差をつける予定であったが、その後の予備調査では、例えば「少子化問題」などで 10 分間では話すことがなくなって沈黙が生まれてしまうことが観察された。そのため、全ての話題につき録音時間を 5 分と設定した。

また、「身の回りの話題」については、例えば「スマートフォン」では「機種・アプリ・

SNS」などについて話してほしいが、「ゲーム」については「漫画・ゲーム」と重なるため話さないようにするなどの指示を与えた。同様に「社会にかかわる話題」については、例えば「マナー」では「公共交通機関のマナー」について話し合うようにするなど論点を話題ボードに書いて指示した。

実際の調査で使われた話題ボードの指示内容まで含めて記したのが表２の 15 話題である。「親密」「具体」の列については山内(2018)によるもので、対応する話題の値を示している。「親密」は親密度・必要度のことで、計算方法はすでに説明した通りである。「具体」は具体度のことで、具体名詞の割合の多寡による分類である。A から D の４段階で現わされる。そして、親密度・必要度と具体度の組み合わせにより「初級」「中級」「上級」に分類される。

表２ J-TOCC の 15 話題（詳細）

| 話題番号 | 話題 | 参加者への指示内容 | 親密 | 具体 |
|---|---|---|---|---|
| 身の周りの話題 | | | | |
| 01 | 食べること | 例えば：好きな料理、外食。※料理を作る話は除く | I | A |
| 02 | ファッション | ＜指示内容なし＞ | II | B |
| 03 | 旅行 | 例えば：行きたい場所、行ったことがある場所 | I | B |
| 04 | スポーツ | 例えば：運動の経験、スポーツ観戦 | II | C |
| 05 | マンガ・ゲーム | アニメ、ケータイゲームを含む | II | B |
| 06 | 家事 | 例えば：料理、洗濯、掃除 | II | B |
| 07 | 学校 | 小学校、中学校、高校時代の思い出 | I | B |
| 08 | スマートフォン | 例えば：機種、アプリ、SNS ※ゲームは除く | I | A |
| 09 | アルバイト | アルバイト経験，やってみたいアルバイト | II | A |
| 10 | 動物 | 例えば：好きな動物、ペット | II | A |
| 11 | 天気 | 例えば：最近の天気、温暖化 | I | D |
| 社会にかかわる内容も含む話題 | | | | |
| 12 | 夢・将来設計 | 例えば：就職・結婚・家庭 | III | D |
| 13 | マナー | 公共交通機関でのマナーについて | II | C |
| 14 | 住環境 | 都会がいいか、地方がいいか | III | A |
| 15 | 日本の未来 | 少子化・高齢化をどう考えるか | III | D |

「アルバイト」については最も近い内容である「労働」としてとらえた場合は、親密度・必要度 III、具体度 C となるが、この話題の設定意図としては「学校（小中高)」とは重ならない、大学生のライフスタイルに関係する何かということであり、表中では「学校（大学)」と対応させている。

## ４．　会話録音の方法

### ４．１　組織体制

会話の録音は表３のような体制で行った。なお、より上位のものが下位の役割を兼ねることもあった。

表３　会話録音の組織体制

| 役割 | 仕事 | アクセスできるデータ |
|---|---|---|
| ●研究代表者（中俣尚己） | 全体の進捗管理<br>方針決定 | 全体の進捗データ<br>PC 内の音声・フェイスシートのデータ（全録音地） |
| ●録音統括者（大学教員） | 録音地ごとの進捗管理 | 全体の進捗データ<br>PC 内の音声・フェイスシートのデータ<br>録音地の調査協力者の個人情報 |
| ●録音者　（学生バイト） | 録音の説明、実施 | IC レコーダーと紙（フェイスシート、同意書など）のみ |
| ●調査協力者<br>（学生バイト） | 会話を行う | なし |

### ４．２　録音の準備

机の上に時計と IC レコーダー２台を並べる。１台は不測の事態に対する備えである。IC レコーダーの下には振動を吸収するため、タオルなどを敷くようにした。

室内の机の配置は調査地によって異なるが、調査協力者はテーブルに対して直角になるように座ってもらった。また、録音者は会話をしている間は調査協力者から見えない位置で待機するようにした。

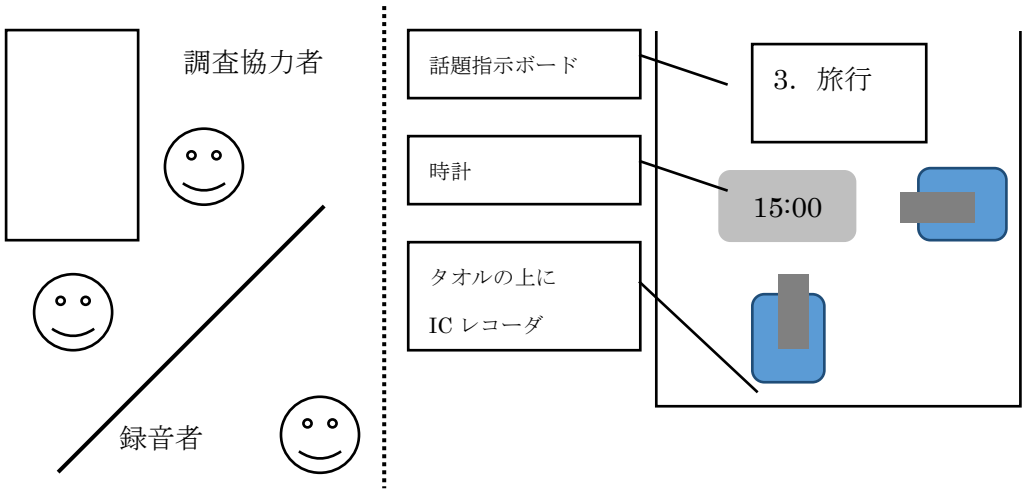



図２　録音の配置の略図（左）と、机の上の略図（右）

### ４．３　調査者への説明

調査協力者が部屋に来たら、ある程度の長丁場の調査になるため、飲み物などを自由に用意するように告げる。

準備ができたら、説明文書を配布する。その後、説明文書を録音者が読み上げるとともに、確実に説明をした証拠として、配布文書の説明を行った箇所には蛍光ペンで印をつけていった。

### ４．４　録音

説明が終わったら、疑問点がないかを聞く。疑問がなくなったところで、録音を開始する。以下のように指示を行った。

「話してもらう話題は，私が紙に出して指示します。30 秒間，その話題について考えることができます。その後，私がレコーダーのスイッチを入れ，「スタート」と言ったら，話し始めてください。できるだけ，示した話題の中で話してください。話が途切れたら，提示した話題の中から別の内容を探してください。また、時間が来たらこちらで止めるように指示しますので、５分以上話すようにしてください。会話を終わらせたり残り時間を気にする必要はありません。」

まず、話題ボードを１枚見せ、30 秒間無言で考える時間を与える。その後、レコーダーのスイッチを入れてから、まずは録音者が「ペア番号 W101 の１ 。話題食べること。スタート」のように発言する。直後から調査協力者は会話を開始し、録音者は見えない位置に移動して待機する。

５分経過したら、録音者はまず無言でレコーダーのスイッチを切る。その後、次の話題ボードを見せる。また、適宜休憩をとり、休憩中のコミュニケーションや飲食は自由とした。また、会話録音の最中でも、スマートフォンの利用は自由とし、スマホを見ながら会話を行ったペアも存在した。

### ４．５　話題提示の順番

話題の提示順はペアごとにランダムになるように計画した。１番目の話題は後の話題よりもぎこちなく発話量が少ないことも予想されるため、カウンターバランスをとった。研究代表者が 120 ペアに対して乱数を使って各ペアごとにランダムな話題提示順を一括で設定し、ファイルを共有した。このファイルには研究分担者のみがアクセスすることができる。録音地ごとに担当の研究分担者が録音者にペア番号に対する話題の録音順を伝え、録音者はあらかじめ話題ボードをその順番に並かえておく。なお、調査の進捗もこの共有ファイルを使って行った。

### ４．６　録音終了後

録音終了後、調査協力者にはフェイスシートと、同意書に記入してもらう。録音したファイルを公開しても問題ないかを改めて確認するとともに、マスキングを希望する用語についても確認した。ここで希望があった語は、マスキングの基準とは無関係にマスキングされる。

同意書については記入後にコピーをとり、コピーを調査協力者に渡す。

また、謝金の処理もこの段階で行う。謝金額は録音地の大学の規定による。

## ５．　文字化とマスキングの方法

### ５．１　文字化とマスキングの手順

ここでは、文字化とマスキングの流れについて記す。まず、録音資料の音声ファイルを業者に依頼し、素おこしの方法で文字化した。しかし、この段階では固有名詞などの匿名化（マスキング）は行われていない。

次に、録音地とは異なる大学の学生をアルバイトとして雇い、全ファイルをもう一度聞き直し、きちんと文字化されているかどうかのチェックを行った。本コーパスは実質語と機能語の語彙の計量を大きな目的とするため、基本的には語未満の発話断片は削除し、語以上の重複は残すという方針をとった。例えば、「そ、それは」は「それは」に整形し、「それ、それは」はそのまま残すといった方針である。これは機械での形態素解析をスムーズに行うためである。

また、並行して第一次のマスキングの作業を行った。まず、録音地の録音統括者がざっとファイルを目視し、削除すべき用語のリストを作成した。録音終了後に調査協力者から申請があったものもメモとして加える。このリストを作業を行う別の大学の学生に送付した。別の大学の学生に依頼したのは、同じ大学の学生が作業をした場合、会話の内容から容易に個人を特定できる可能性があるためである。当初は、東日本のファイルを西日本の学生が、西日本のファイルを東日本の学生が作業することも考えたが、方言の聞き取りの

問題や、地名はマスクするが、その地方の人間なら誰でも訪れる可能性のある著名な地名はマスクしないといった判断を行う必要があるため、西日本、東日本の中で、別大学の学生に作業を依頼した。

その後、録音を担当した録音統括者が第二次のマスキングを行った。実際にはアルバイトの学生は個人の特定に繋がらないような地名も機械的にマスキングすることが多く、問題のない部分はマスキングを解除し、できるだけ意味がわかるように行った。

最後に、研究代表者が第三次マスキング作業として、集約した「削除すべき語のリスト」を全て全文検索で検索し、細かい部分の修正を行った。

### ５．２　文字化とマスキングの方針

文字化の例は以下の通りである。

E-102-1M：うん。
E-102-2M：食べてないっていうか、食べる習慣がなくなった（E-102-1M：うん）からなんだけど。いや、でも合宿中ね、やっぱ朝ご飯食べてたけど（E-102-1M：うん）、出てくるからね、刑務所のごとく 3 食出てくるから。
E-102-1M：うん。(E-102-1M)

話者記号の後、全角のコロンをはさみ、会話内容を文字化した。話者記号は最初の一文字が東日本は E、西日本は W である。数字の 1 桁目は「男男」が 1、「男女」が 2、「女女」が 3 である。その後の 2 桁はペア番号と同じ通し番号である。末尾の数字も通し番号で、最後に性別を男性は M、女性は F で示した。

また、短い相づちは全角の丸括弧の中に、同様に話者記号とコロンの後にその内容を文字化した。

上昇調の文は？で終わっている。本コーパスでは全ての文が句点「。」あるいは疑問符「？」で終わっている。

E-118-2M：えー、じゃあ、逆に何好きなの？　何好きで、食べてるの？　いつも。
E-118-1M：何だろ。まあ、牛丼とかもね、あんま行かないんだよね。（E-118-1）

笑い声は「ははは」のように記載した部分もあるが、文字化が困難である部分は（笑）を記した。

E-118-2M：うんうんうん。確かに。松屋とかも行かないわ、最近。松屋全然行かないわ。
E-118-1M：すき屋たまあに。

E-118-2M：いや、でも吉野家は立地的に行かない（笑）。（E-118-1M）

聞き取れなかった箇所は、「●」で示した。

E-311-1F：例えば、何かさ、お肉とかでもさ、あの、ロース（E-311-2F：ああ）とかを食べてたけど、だい、だんだんこう、●（E-311-1）

聞き取れたが、確信が持てない箇所は「＝　＝」でくくられている。

W-112-2M：そんな化けもん、飲んでたの、お前、＝一度＝。お前、それ、でかくなるわ。（W-112-1）

発話ではない注釈は＜　＞でくくられている。

E-105-1M：ここに書いてある。料理を＜聞き取り不能＞(E-105-1)

マスキングについては、大方針として、調査協力者の特定もしくは調査地の大学の特定につながる情報をマスクするという方針を立てた。人名・地名であっても、有名人の名前や、誰もが訪れるような有名な地名はマスキングの対象外となる。マスキングは全て、隅つき括弧【　】で示し、その中にカテゴリを表す語を「～名」の形で記入している。

人名については【人名・１人称】・【人名・２人称】・【人名・３人称】の区別を設けた。

E-101-2M：【人名：２人称】好きな料理何？
E-101-1M：俺の好きな料理はね、（E-101-2：うん）そうだな。好きな料理。（E-101-1）

E-214-2M：うん。なんか、【人名：３人称】さんと【人名：３人称】さんと行ってきたんだけど、（E-214-1F：うん）あそこは入れるみたいよ。(E-214-1)

地名については、原則は市区町村レベルまではオープンとし、それよりも下位の地名をマスキングしたが、人口の多寡の関係からこの点はケースバイケースである。「大阪市」「京都市」など政令指定都市レベルの名称はそのままである。また、個人情報の特定につながるか否かという観点なので、マイナーな地名であって mo 旅行などで訪れた場合はそのまま

にしている。以下の例の場合でも、【地名】は特定につながる情報であるが、上位の「東京」や比較対象の「新宿」はマスキングしていない。

E-210-1M：何か、【地名】ってさ、東京のくせにさ、何かさ、都心と天気全然違くない？
E-210-2F：違うよね。
E-210-1M：新宿とかより、断然寒いし、新宿とかより天気全然違うしさ。（E-210-11）

その他にも特定につながる情報をマスキングした。

E-219-1：おすすめのラーメン屋さん。好きなラーメン屋さん。やっぱ、でも【店名】だな。(E-219-1)

W-107-2M：で、あの一、野球やったらな、見に行って、最後の試合。あのあそこに。あの一、【施設名】、【施設名】な。(W-107-1F：ああ、ああ、ああ、ああ、ああ)【地名】の。も見に行って、やっぱいいなあと思いつつ、自分はやらへんけど応援しに行き、まさかの初戦負けやで。【学校名】に。ぼこられる【学校名】。(W-107-4)

W-316-2F：【駅名】。【鉄道名】が、【地名】広いけどさ、(W-316-1：うん) 駅が【駅名】しかないんよ、【鉄道名】(W-316-13)

## ６．話者の属性と話題精通度

### ６．１　調査協力者の属性

J-TOCCの調査協力者(話者)は全て日本語を母語とする20歳以上の大学学部生である。様々な話題について話せるように、親しい関係の話者ペアのみに協力を依頼した。大学院生を選ばなかったのは、何かのはずみで研究テーマが話題にあがった時に容易に個人の特定が可能になるためである。また、同意書を取る関係上、20歳以上の者に限定した。

### ６．２　話題精通度

調査終了後、全ての話者に、フェイスシートで15の話題それぞれの話題精通度を尋ねた。話題精通度は、「それぞれの話題についてどれだけ詳しいか、あるいはどれだけ自信を持って話すことができたか」という尋ね方をした。その詳しさ・自信の度合いを５段階で評価してもらった。本来はその話題についてあらかじめどれほどの知識があるかを尋ねるべきであるが、例えば「天気」などの話題は気象予報士の勉強をしているのでなければ、精通という尺度は馴染まないため、このような聞き方をした。

### ６．３　話者情報の公開

話者の情報は documents フォルダ内の speakers_information.csv のファイルに記載している。ペア番号、録音地、ペアの組み合わせ、話者記号、性別、言語形成地（６～１２歳の間に居住していた都道府県。複数回答あり）を記載し、その後の話題ラベルの列は上述の話題精通度の値を記載している。同じペア番号のデータは必ず２行ずつあり、４列目の話者記号で区別されていることに注意が必要である。

## ７．　J-TOCC のバージョン・更新履歴

- バージョン 20200228
  ・公開初期バージョン
- バージョン 20210706
  ・規約と解説文書に参照すべき文献を追加。本体の変更なし。
- バージョン 20210831
  ・話者情報のうち、W-217-1 と W-217-2、W-219-1 と W-219-2、W-118-1 と W-118-2 が入れ替わっていたので話者情報シートを修正。
  ・対応して、W-217 ペアと W-219 ペアでは話者の性別の記号も反対になっていたので、話者記号の性別部分(F,M)のみ修正。
  ・E-202-09 の 64 行目と 65 行目の話者記号が逆になっていたので修正。
  ・「話題知悉度」という用語を「話題精通度」に変更。ファイルに修正はなく、解説文書のみの変更。
- バージョン 20220829
  ・コーパス本体で、マスキングに漏れがあった箇所を修正。補足情報の記号の不統一を修正。
  ・説明文書に記号の情報を加筆。誤字の訂正。
  ・『日本語話題別会話コーパス：J-TOCC　語彙表』を公開。

## 付録

会話の録音などに使用した資料を、以下に記載する。
ただし、文言は実施した調査地によって若干のアレンジを加えている。

## 付録１　会話参加者への説明文書

「話題が語彙・文法・談話ストラテジーに与える影響の解明」
のための調査協力について

### １．　研究の目的・意義について

会話には必ず「話題」が存在します。この研究は様々な話題の会話を集め，言語学的に分析をする研究です。そのため、これから様々な話題について、会話をしてもらいます。

### ２．調査の方法

様々な話題について会話しているところを録音します。話題は全部で 15 個で１つの話題の会話は５分です。こちらで止めるように指示しますので、５分以上話すようにしてください。その後、同意書と簡単なアンケートに記入してもらいます。

### ３．研究により期待される便益

この研究に参加することで、対象者の直接的な便益はありませんが、研究成果は日本語教育のための教材開発や言語調査などで今後の研究の発展に寄与すると考えられます。

### ４．この研究の予想される効果と起こるかもしれない不利益について

録音した会話は文字化し、文字情報のみが公開されます。あなたや第三者の氏名など個人情報は以下のように，マスクされて公開されます。
「こないだ，中俣と話したんだけど」→「こないだ，●●と話したんだけど。」
「僕の実家は鹿児島の指宿ってとこ」→「僕の実家は鹿児島の●●ってとこ」
しかし，会話中に他人に知られたくない内容を話してしまい、なおかつ会話の断片的な情報からあなたやその人のことを推理できる人間がそれを見た場合には、その内容を知られてしまう危険性があります。細かすぎる個人情報は話さないようにしてください。また，法に触れる行為についても注意してください。

### ５．研究への参加について

・この研究への参加は協力者の自由意志によるものです。
・この研究への参加に同意しない場合でも、あなたが不利な扱いを受けたりす

ることはありません。

**６．研究協力を中止する場合について**

　録音後、協力を中止したくなった場合は2020年3月31日までに申し出て下さい。あなたの会話データは公開されません。また、その時点であなたに関する全てのデータは破棄されます。

**７．　個人情報の取り扱いについて**

・会話以外の個人情報については、氏名を含むデータは研究チーム外部には公開しません。それ以外のデータは会話参加者に関する資料として、文字化データ公開時に公開されることがあります。
・文字化は外部の業者が行います。その後に、個人の特定につながる情報は削除します。その作業はあなたの所属する大学とは無関係の人間が行います。
・文字化した会話はインターネットで公開され、研究目的のために誰でも利用することができます。
・個人情報を含む段階のデータは誰がどの情報を持っているかを厳密に管理します。

**８．データの二次利用について**

・文字化した会話は第三者が研究のために使用することがあります。また、著作物の中であなたの会話を引用することがあります。
・会話の著作権を研究チームに譲渡して頂きます。

**９．研究を担当する教員**

　この研究のことで何かわからないことや心配なことがありましたら、いつでも、下記にお尋ね下さい。

研究責任者　国文学科・准教授　中俣尚己
nakamata@kyokyo-u.ac.jp

**付録２　同意書**

研究責任者　京都教育大学・准教授・中俣尚己　殿

私は「話題が語彙・文法・談話ストラテジーに与える影響の解明」の研究について、以下の説明を受けました。理解した項目は以下の□に自分で「✓」を入れて示しました。

**□ 研究の目的、意義について　□調査の方法**
**□ 研究により期待される便益　□ この研究の予想される効果と起こるかもしれない不利益について**
**□ 研究への参加について　□ 研究協力を中止する場合について**
**□ 個人情報の取扱いについて　□ データの二次利用について**
**□ 研究を担当する教員の連絡先**

なお、この研究において録音された私の発言の文字データの公開につきましては、以下の□ に自分で「✓」を入れて示しました。

**□ 研究者を対象とする学術目的に限り、個人の特定可能な部分を削除した状態で，公開してよい**
**□すべての話題の録音データについて公開してよい**
**□以下の話題を除く録音データについて公開してよい**
（　　）
**□ 公開に同意しない**

また、提供する調査情報等が本研究終了後も保存され、将来新たに計画・実施される研究に使用されることに
**□同意します □同意しません**

また、会話内容の著作権については
**□研究責任者に譲渡します□譲渡しません**

上記に関する説明を十分理解した上で、研究に協力することに同意します。
なお、この同意はいつでも撤回できることを確認しています。
同意年月日　　年　月　日
ご署名
本研究に関する説明を行い、自由意思による同意が得られたことを確認しました。
説明年月日　　年　月　日
説明者氏名

**付録３　話題指示ボードの一例**

（実際は A4 横に印刷。全ての内容については p.4 を参照）

# 食べること

## (例えば：好きな料理、外食。
## ※料理を作る話は除く)

**付録４　フェイスシート**

会話録音へのご協力，どうもありがとうございました。
調査の最後に，以下のご質問にお答えください。
なお，データは個人情報がわからない形で保存されます。

氏名＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿

性別＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿

言語形成地　＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿

（あなたが６～１２才の間に居住していた**都道府県**を記入してください。

今日話した，以下の１５の話題のそれぞれについて，自分がどれだけ詳しいか，自信を持って語れたかを５段階の中から１つ選び，数字を○で囲んでください。

（1）　食べること

詳しい・自信あり　　普 通　　詳しくない・自信なし

５　　４　　３　　２　　１

（2）　ファッション

詳しい・自信あり　　普 通　　詳しくない・自信なし

５　　４　　３　　２　　１

（3）　旅行

詳しい・自信あり　　普 通　　詳しくない・自信なし

５　　４　　３　　２　　１

（4）　スポーツ

詳しい・自信あり　　普 通　　詳しくない・自信なし

５　　４　　３　　２　　１

（5）　マンガ・ゲーム

詳しい・自信あり　　普 通　　詳しくない・自信なし

５　　４　　３　　２　　１

（6）　家事

詳しい・自信あり　　普 通　　詳しくない・自信なし

５　　４　　３　　２　　１

（7）　学校

詳しい・自信あり　　普 通　　詳しくない・自信なし
5　　4　　3　　2　　1

（8）　スマートフォン
詳しい・自信あり　　普 通　　詳しくない・自信なし
5　　4　　3　　2　　1

（9）　アルバイト
詳しい・自信あり　　普 通　　詳しくない・自信なし
5　　4　　3　　2　　1

（10）　動物
詳しい・自信あり　　普 通　　詳しくない・自信なし
5　　4　　3　　2　　1

（11）　天気
詳しい・自信あり　　普 通　　詳しくない・自信なし
5　　4　　3　　2　　1

（12）　夢・将来設計
詳しい・自信あり　　普 通　　詳しくない・自信なし
5　　4　　3　　2　　1

（13）　マナー
詳しい・自信あり　　普 通　　詳しくない・自信なし
5　　4　　3　　2　　1

（14）　住環境
詳しい・自信あり　　普 通　　詳しくない・自信なし
5　　4　　3　　2　　1

（15）　日本の未来
詳しい・自信あり　　普 通　　詳しくない・自信なし
5　　4　　3　　2　　1

質問は以上です。どうもありがとうございました！

付録５．文字起こしチェック、マスキング作業者の誓約書

# 個人情報保護に関する誓約書

京都教育大学　国文学科　准教授
中俣尚己　殿

私は、研究プロジェクト「話題が語彙・文法・談話ストラテジーに与える影響の解明」の会話録音作業に従事するにあたり、従事する期間のみならず、終了後においても、下記の事項を固く遵守すると成約します。

記

・ 録音中の会話に出現した情報については、一切開示、口外、提供、漏洩しません。また、私自らのために使用することは一切ありません。

・ 個人情報については、漏洩し、紛失し、又は毀損しないよう確実に管理し、盗難等の被害にも遭わないよう最大限の注意を払って取り扱います。

・ 万が一個人情報が漏洩した場合、又はその可能性が疑われるときは、直ちに監督者へ速やかに報告し、必要な指示を受けます。

以　上

【誓約者（従業者）署名欄】

年　　月　　日

住所：

氏名：　　印

## 付録６　『日本語話題別会話コーパス』　利用規約

配布ホームページに掲載。同意した場合のみダウンロード可能となる。

### 『日本語話題別会話コーパス（J-TOCC)』　利用規約

この利用規約は JSPS 科研費 18H00676「話題が語彙・文法・談話ストラテジーに与える影響の解明」の成果物である『日本語話題別会話コーパス』（以下 J-TOCC）のデータをダウンロード、利用いただく際の条件として規定するものです。この利用規約は J-TOCC の全ての利用者に適用され、利用者はこの利用規約の内容に同意することなく J-TOCC を利用することはできません。

（著作権）

第１条　J-TOCC 内のデータの著作権は、制作者である中俣尚己に帰属します。

（利用者情報の届け出）

第２条　科研費による成果であり、利用実態を記録するため、また、パスワードの配布のため、利用申し込みの際には利用者の氏名、メールアドレス、コーパスの利用目的、利用範囲等の必要事項を申込フォームにより申請する必要があります。

２　利用者は申込フォームに記入した内容に変更が生じた場合、遅滞なく改めて申し込みフォームから必要事項を申請するものとします。

３　これらの情報は適切に管理し、J-TOCC に関する連絡以外には一切使用しません。

（許諾範囲）

第３条　利用者が J-TOCC を利用できる範囲は以下のとおりとします。

（１）　研究目的：研究・教育を目的とする場合に限ります。

（２）　利用者の範囲：申し込みフォームに記入のあった個人を原則とします。ただし、①複数の作業者で共同研究・開発を行う場合、②授業で利用する場合には、特例としてデータの複製を許可します。また、この場合、利用登録時に複製する利用者の大よその数を申請する必要があります。申請は授業の年度ごと、共同利用する研究プロジェクトごとに行ってください。

（禁止事項）

第４条　利用者は以下に定める行為を行ってはいけません。

（１）　J-TOCC の全部又は一部を複製し、申し込み時に申告した利用者以外の者に利用させること。

（２）　第３条の範囲を超えて利用すること。

（３）　J-TOCC のデータを用いて第三者の名誉を棄損し、あるいはその他の権利を侵害

すること。
（４）　J-TOCC に付随する話者データ以外の話者に関する情報を公開すること。
（５）　本データに含まれる発話について、事実関係の正誤や思想、発言の適否等、発話や行動の内容、人格に関する議論、批判、感想等を公開すること。あるいは発話や話者情報について、個人やその所属組織に関する推測等を公開すること。

（研究成果の公表）
第５条　利用者は第４条に反しない範囲で J-TOCC のデータを利用して得られた研究成果や知見を公表することができます。これらの公表については、解析データや処理プログラムの公表を含みます。公表時は『日本語話題別会話コーパス（J-TOCC）』を利用したことを明記してください。また、以下の文献のいずれかを引用してください。

中俣尚己(2021)「日本語話題別会話コーパス：J-TOCC　解説資料」
http://nakamata.info/database/j_tocc_document.pdf
中俣尚己・太田陽子・加藤恵梨・澤田浩子・清水由貴子・森篤嗣(2021)「『日本語話題別会話コーパス：J-TOCC』」『計量国語学』33 巻 1 号, pp.11-21, 計量国語学会.

２　J-TOCC に含まれる文字化データをそのまま引用して公開する場合、著作権法における引用の規定に従ってください。その範囲を超えて公開する場合は、別途制作者にご相談ください。他の媒体に J-TOCC のデータの一部を収録する場合もご相談ください。
３　公表後で構いませんので、原稿のコピー等を中俣尚己までお送り下さい。

（対価）
第６条　J-TOCC の利用に係る料金は、無償とします。

（免責）
第７条　J-TOCC を利用することによって生じたいかなる損害についても、制作者は責を負いません。
２　J-TOCC のデータ内容は事前の予告なく変更されることがあります。

（利用の停止）
第８条　利用者がこの利用規約の条件に違反したことが判明した場合、制作者は利用者に通知することにより利用を停止させることができます。本条の規定は、制作者から利用者への損害賠償請求を妨げるものではありません。

（誠実義務）

第９条　本契約に定めのない事項、又は本契約の各条項の解釈について疑義が生じたときについては、制作者と利用者の双方が誠意をもって協議するものとします。

## 参考文献


**■本コーパス構築のきっかけとなった話題と文法に関する研究**

中俣尚己(2016)「真正性のある接触場面会話コーパスを用いた話題別特徴語の抽出―ポップ・カルチャーの場合―」『日本語教育学会 2016 年度春季大会予稿集』pp.146-151
http://nakamata.info/nakamata_nihongokyouiku2016s.pdf

Nakamata, Naoki (2019) Vocabulary Depends on Topic, and So Does Grammar *Journal of Japanese Linguistics,*35-2, pp.213-234,

**■話題選定に関する研究**

山内博之(編)(2013)『実践日本語教育スタンダード』ひつじ書房

橋本直幸(2016)「話題から見た語彙シラバス」森篤嗣(編)『ニーズを踏まえた語彙シラバス』pp.33-51, くろしお出版

橋本直幸(2018)『話題別読解のための日本語教科書読み物リスト 2017』

山内博之(2018)「話題による日本語教育の見取り図」岩田一成(編)『語から始まる教材作り』pp.3-16, くろしお出版

**■録音やマスキング作業の参考にした研究**

中俣尚己(2015)「日中 Skype 会話コーパスについて」
http://nakamata.info/about_skype_corpus.pdf

苅宿紀子(2018)「『大学生三者コーパス』の設計と調査方法」『表現学部紀要』18, pp.79-87.

田中弥生・柏野和佳子・角田ゆかり・伝康晴・小磯花絵(2018)『『日本語日常会話コーパス』の構築―個人密着法に基づく会話の収録―』国立国語研究所
https://pj.ninjal.ac.jp/conversation/report/report02.pdf

小磯花絵・伝康晴(2018)「『日本語日常会話コーパス』データ公開方針：法的・倫理的な観点からの検討を踏まえて」『国立国語研究所論集』15, pp.75-81.

## 謝辞

本プロジェクトの遂行は　JSPS 科研費基盤研究(B)「話題が語彙・文法・談話ストラテジーに与える影響の解明」（課題番号 18H00676）の助成を受けました。記して感謝申し上げます。

本プロジェクトにかかわったメンバーは以下の通りです。（所属は 2021 年 2 月時点）

**●研究代表者**

中俣　尚己（京都教育大学）

**●新規コーパス構築班**

太田　陽子（一橋大学）
加藤　惠梨（大手前大学）
澤田　浩子（筑波大学）
清水　由貴子（聖心女子大学）
森　篤嗣（京都外国語大学）

**●既存コーパス分割班**

小口　悠紀子（広島市立大学）
小西　円（東京学芸大学）
建石　始（神戸女学院大学）
堀内　仁（国際教養大学）

**●話題選定班**

山内　博之（実践女子大学）
橋本　直幸（福岡女子大学）

**●機械的分析班**

山本　和英（言語商会）

**●研究協力者**

石川　慎一郎（神戸大学）
茂木　俊伸（熊本大学）

この他、多数の学生アルバイトの方に、録音やマスキングのチェックに協力頂きました。文字化は東京反訳（株）にお願いいたしました。そして、何より、本コーパスが完成したのは 240 名の調査協力者の方々のおかげです。深くお礼申し上げます。

## 本解説資料　改訂履歴

2021/07/06　規約内の参照すべき論文を 1 件追加。

2021/08/31　J-TOCC のバージョン情報について追加。「話題知悉度」の用語を「話題精通度」に変更