

国立国語研究所学術情報リポジトリ

『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

メタデータ	言語: Japanese 出版者: 公開日: 2018-07-10 キーワード (Ja): キーワード (En): Corpus of Everyday Japanese Conversation, release guideline, portrait right, copyright, personal information protection laws 作成者: 小磯, 花絵, 伝, 康晴, KOISO, Hanae, DEN, Yasuharu メールアドレス: 所属:
URL	https://doi.org/10.15084/00001597

『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——

小磯花絵^a 伝 康晴^b

^a 国立国語研究所 研究系 音声言語研究領域

^b 千葉大学／国立国語研究所 研究系 音声言語研究領域 客員教授

要旨

現在、国立国語研究所の機関拠点型基幹研究プロジェクト「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」で構築している『日本語日常会話コーパス』(CEJC)には、例えば自宅での家族との会話や飲食店での友人との会話、職場での同僚との打ち合せ、学校での同級生との雑談、散策時や車中の会話など、実に多様な場面の会話が含まれる。本コーパスは、音声データや転記テキストだけでなく、映像データについても公開する方針だが、こうした日常生活の中で記録された会話を、映像データも含めて公開したコーパスはこれまでにほとんど存在せず、データ公開のための方針を確立する必要がある。例えば、日常場面の会話を録画すると、映像データには公開の承諾を得ていない第三者の顔やテレビなどの著作物の写り込みが多く見られる。そこで、これまでに収録した多様な会話データをもとに具体的な問題を洗い出し、その対応について、肖像権や個人情報保護、著作権などの観点から、知財関連を専門とする弁護士と相談を重ねてきた。本稿では、これら一連の検討・議論を踏まえて定めたCEJCのデータ公開方針について報告する*。

キーワード：『日本語日常会話コーパス』、データ公開方針、肖像権、著作権、個人情報保護法

1. はじめに

国立国語研究所では、2016年度より機関拠点型基幹研究プロジェクト「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」(2016～2021年度)を開始した(小磯2017)。このプロジェクトでは、さまざまなタイプの日常会話200時間をバランス良く収録した『日本語日常会話コーパス』(*Corpus of Everyday Japanese Conversation*, 以下CEJC)の構築を進めている(小磯ほか2017)。

これまで公開された日本語の会話コーパスは、親近者同士の雑談や電話会話、職場会話といったように、話者や会話形式、場面などに偏りが見られる。また、例え雑談を扱うコーパスであっても、収録のために集められた状況、つまり個人の生活と切り離された状況でなされた会話が主な対象であることが多い。特に会話の音声や映像を公開するコーパスについては、日常場面の会話を対象とするものはほとんど存在しない。個人の生活の中で営まれる会話には多くの個人情報が含まれているため、音声・映像データの公開が容易ではないことが一因であると考えられる。

* 本稿の内容は2017年3月1日のシンポジウム「日常会話コーパス」IIにおけるパネルセッション「日常会話データの公開における倫理的・法的な問題について」および2017年9月6日の「言語資源活用ワークショップ2017」での発表の内容に加筆・修正したものである。これらの発表時に有益なコメントをくださった方々に感謝申し上げる。

しかし、日常生活の中で我々が言葉をいかに用い、どのような仕組みで人とコミュニケーションをとっているのか、また日常場面でのさまざまな活動を言葉や身体を用いていかに組織化しているのかといったことを問うには、まさに日常生活の中で自然に営まれた会話を対象に、音声・映像まで含めたデータの公開が不可欠である。

そこで本プロジェクトでは、次に挙げる三つの方針を立て、CEJC の構築を進めている。

1. 日常場面の中で当事者たち自身の動機や目的によって自然に生じた会話を対象とする
2. バランスを考慮して日常生活の多様な場面の会話を扱う
3. 音声データや転記テキストだけでなく映像データも収録・公開の対象とする

これまで収集したデータの中には、例えば自宅での家族との会話や飲食店での友人との会話、職場での同僚との打ち合せ、学校での同級生との雑談、散策時や車中の会話など、多様な場面の会話が含まれる。例えば飲食店で収録した会話には、周りの客の顔や声が映像・音声データに写り込むことが多い。また、料理を注文する際に店員と会話したり、会話の最中に知人から電話がかかってきたたり、BGM やテレビ画面が音声データや映像データに写り込むこともある。こうした日常生活における多様な場面の会話を映像を含めて一般に公開するというのは世界でも初めての試みであり、挑戦的な課題である。従来、当然公開されるべきでありながら、構築段階でデータの法的・倫理的な問題の扱いを十分に検討しなかった結果、公開に至らなかった話し言葉のコーパスが数多く存在する (Maekawa 2015)。こうした事態に陥らないよう、法的・倫理的な観点からの入念なる検討が不可欠である。

国立国語研究所では大規模な話し言葉のコーパスとして『日本語話し言葉コーパス』(CSJ) を構築・公開した (国立国語研究所 2006)。しかし CSJ は、音声を文字化する際の資料として映像データを収録したが公開はしておらず、映像データ公開についての知見がない。また CSJ は学会発表やスタジオでの一般人によるスピーチといった独話が中心であり、一部含まれている対話もスタジオでの録音であるため、先に述べたような収録・公開の同意を得ていない第三者が写り込むことはない。よってこうした事例の対応法についても十分な知識がない。

そこでプロジェクトを開始するにあたり、準備研究として、日常場面の会話を映像を含めて収録・公開することの可能性や生じうる問題について、知財関連を専門とする弁護士と相談しながら検討を進めた¹。この検討に基づき、会話の収録方法や公開方針の骨子を策定し、2016 年度より会話の収録・整備を進めた。2016 年 12 月、収録会話が 100 時間ほど集まった段階で具体的な問題を洗い出し、再度弁護士と相談した上で、データ公開方針の原案を策定した。この原案を中心に、2017 年 3 月 1 日開催の公開シンポジウム「日常会話コーパス」II におけるパネルセッション「日常会話データの公開における倫理的・法的な問題について」において、関連分野の研究者を交え、データの公開方針について議論した。

¹ 国立国語研究所共同研究プロジェクト「均衡性を考慮した大規模日本語会話コーパス構築に向けた基盤整備」(プロジェクトリーダー：小磯花絵、2014 年 7 月～2015 年 8 月)

本稿では、これら一連の検討・議論を踏まえて定めた CEJC のデータ公開方針について、その背景となる法的・倫理的な観点も含めて報告する。

2 節では、収録の前に会話に参加する人と交わす会話収録・公開に関する同意の手続き、および説明文書・同意書におけるデータ公開条件について述べる。これが、CEJC を公開する際の基本方針となる。3 節では、これまでに収録したデータに見られる公開時の問題について法的・倫理的な観点から整理し、本コーパスにおける映像・音声・転記データの公開方針を示す。3.1 節では映像・音声データへの写り込みの問題を、3.2 節では個人情報等に関わる問題を、3.3 節ではそれ以外の問題を扱う。3.4 節において、協力者へのヒアリングとオプトアウトの機会について補足する。

2. 会話収録・公開に関する同意取得の手続きおよびデータ公開の条件

2.1 会話収録・公開に関する同意取得の手続き

CEJC では、日常場面の中で当事者たち自身の動機や目的によって自然に生じた会話をバランス良く集めるために、調査に協力してくれる一般の人（以下、協力者）にビデオカメラや IC レコーダーなどの収録機材を 2～3 ヶ月間ほど貸し出し、自身の日常生活の中で生じる会話を協力者本人に収録してもらうという方法を主として採用している²。

この収録法では、調査者は収録の場に介在しない。そのため、会話に参加する人（会話者）への収録の趣旨説明やデータ収録・公開に関する同意書への署名の依頼などは、協力者が担当することになる。同意取得の手続きの概要は次の通りである（詳細は田中ほか（2018a,b）参照）。

協力者は、会話を収録する前に、説明文書、およびその概要を分かりやすくまとめた説明用のビラに基づき、会話者に対して収録の趣旨やデータの公開方法・公開条件などについて説明した上で、全ての会話者からデータ収録・公開に関する同意書への署名をもらった場合に限り、会話を収録する³。1 人でも同意しない人がいた場合、その会話は収録しない。同意書には公開を望まない箇所を記載する欄を設けており、収録終了後に、協力者は会話者に非公開箇所の希望がないかを聴取し、あればその欄に記入する。データの公開条件については次節で言及する。

説明文書・同意書は、1 人につき 2 枚用意しており、1 枚は会話者が保管、もう 1 枚は協力者を介して国立国語研究所で保管する。会話者が未成年の場合、保護者の署名も求める。よって、保護者の同意なくして未成年者を含む会話が収録されることはない。

説明文書の裏、同意書の横の頁に、同意撤回書を付けており、後日、公開の同意を撤回したい場合にこの書類を利用するよう、協力者から会話者に説明してもらう。同意撤回の申し出があった場合、収録した会話データは破棄され、公開対象から外される。

² 職場での会議や接客場面の会話など、この方法では収録が難しいと思われる場面については、プロジェクトメンバー（調査者）が主体となり収録する方法で補完する。収録の詳細については田中ほか（2018a,b）を参照のこと。

³ 事前に調査の趣旨や公開方法などを見てもらえるよう、同内容を記載したホームページも作成して公開している。

2.2 説明文書・同意書におけるデータ公開条件

同意書に記されたデータ公開に関する条件が、CEJC として会話データを公開する際の基本方針となる。以下に同意書における該当箇所の文言を示す（同意書など、収録に関わる一連の文書・資料等は、田中ほか（2018b）の付録を参照のこと）。

条件の第1項は、国立国語研究所が定める研究教育利用・商業利用（統計情報利用）の条件に同意し契約を交わした者に対して公開する場合のものである。本プロジェクトではこの形での公開を主たる方法として予定している。

私は、○○氏から、「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」に関する説明を受け、この調査への参加、および、この調査において記録された私の映像・音声・文字化資料、研究用情報、フェイスシートに記入した情報、会話状況情報等の公開について、以下の条件のもとに同意します。

1. 国立国語研究所が定める研究教育利用・商業利用（統計情報利用）の条件に同意し契約を交わした者に対して公開する際には、データに以下の処理をほどこす。
 - ・私の名前、学校や会社など私が所属する組織の名称、自宅・所属組織の住所・電話番号の音声が聞こえないように加工し、文字化資料においても、仮名や伏せ字に置き換えるなどの処理をする。
 - ・私が公開を望まない箇所の音声・文字化資料も同様に加工する。
2. 第1項以外の方法での公開については、上記に加え、顔の一部にボカシを加えるなど、私個人が特定できないように映像を加工する。また、会話全体ではなく短いシーンごとの映像・音声の公開に留める。

第1項では、(1) 会話者が特定できる情報のうち、会話者の名前、所属組織の名称、自宅・所属組織の住所・電話番号の情報、および、(2) 会話者が公開を望まない箇所（同意書で非公開の希望を記した箇所）の情報が分からぬよう、会話の音声と文字化資料（転記テキスト）を加工することを定めている。一方、映像の加工についての条件はない。よって、会話者の顔にボカシなどの処理を加えずに公開することとなる。会話者の顔にボカシを加えず公開するということが会話者に明確に伝わるよう、写真を用いて公開方法を具体的に説明したビラ（図1参照）を用意し、説明の際の補足資料として活用している。

条件の第2項は、インターネット等での配信など、今後の公開の可能性を考慮して設けた条件であり、本プロジェクトでは当面対応する予定はない。そこで以下では、条件の第1項の公開方法に焦点を当てて議論を進める。

3. データ公開に伴う問題の整理と公開方針

本節では、これまでに収録したデータに見られる公開時の問題について、法的・倫理的な観点から整理し、定めた公開方針について述べる。これらは、実例に基づく問題を整理した上で、知財関連を専門とする弁護士との相談やセミナー、プロジェクト主催のシンポジウム、著作権法や肖像権、個人情報等に関する書籍など（文化庁編著 2017, 鈴木 2014, 大家 2011, 佃 2010, 岡村



図1 説明用のビラの裏面：映像・音声・転記テキストの対応についての箇所の抜粋

2017, 宇賀 2016) を参考に筆者がとりまとめたものである⁴。法律等の解釈の誤りや方針の妥当性の問題などがあれば、それらは全て筆者の責任である。

本節では、以下の三つに分けて議論する。

1. 映像・音声データへの写り込みの問題
2. 個人情報等に関わる問題
3. その他の問題

⁴ 福井健策弁護士との法律相談、本プロジェクト開始前に参加した同弁護士による「資料の研究・公開と個人情報の利活用ルール」に関するセミナー（2014年12月18日開催）、本プロジェクト主催のパネルセッション「日常会話データの公開における倫理的・法的な問題について」（2017年3月1日）における西野暢助氏（人間文化研究機構）の報告などを参考に、筆者がとりまとめた。またデータの公開方針は、同パネルセッションにおける関連分野の研究者の意見なども反映させ、最終的に定めたものである。

3.1 映像・音声データへの写り込み

日常での会話を収録すると、例えばテレビの画面や音、BGMなどの音楽、個人で撮影した写真やビデオ、会議の非公開資料、収録・公開の同意を得ていない第三者の顔や話している声、名前や所属、住所が記された名刺や表札などが、映像データや音声データに写り込むことがよくある。これらを、問題の対処の観点から整理すると、(1) 著作物の写り込み、(2) 収録・公開の同意を得ていない第三者の容貌の写り込み、(3) 第三者の発話の写り込み、(4) 個人情報の写り込みの四つに分類できる。このうち(4)の個人情報については、写り込みか否かを区別せず、3.2節でまとめて言及する。本節では(1)から(3)について言及する。

3.1.1 著作物の写り込み

■問題の整理

収録された会話には、テレビの画面や音、壁にかけてある絵画やポスター、BGMなどの音楽、書籍や雑誌の表紙や中身など、いわゆる著作物の写り込みが多く見られる。こうした著作物の写り込みの問題は、著作権法の第30条の2「付随対象著作物の利用」、いわゆる「写り込み」等に係る規定が関わる。以下にその条文を示す。

(付随対象著作物の利用)

第30条の2 写真の撮影、録音又は録画（以下この項において「写真の撮影等」という。）の方法によつて著作物を創作するに当たつて、当該著作物（以下この条において「写真等著作物」という。）に係る写真の撮影等の対象とする事物又は音から分離することが困難であるため付隨して対象となる事物又は音に係る他の著作物（当該写真等著作物における軽微な構成部分となるものに限る。以下この条において「付隨対象著作物」という。）は、当該創作に伴つて複製又は翻案することができる。ただし、当該付隨対象著作物の種類及び用途並びに当該複製又は翻案の態様に照らし著作権者の利益を不当に害することとなる場合は、この限りでない。

2 前項の規定により複製又は翻案された付隨対象著作物は、同項に規定する写真等著作物の利用に伴つて利用することができる。ただし、当該付隨対象著作物の種類及び用途並びに当該利用の態様に照らし著作権者の利益を不当に害することとなる場合は、この限りでない。

この規定は、写真撮影や映像・音声収録によって著作物を創作する際、ほかの著作物（例えばテレビの画面や音など）が写り込んでしまった場合に、それが著作権の侵害行為に当たらない範囲を定めたものである。

上の条文を整理すると、次の四つにまとめることができる。

- (1) 写真や映像・音声などの撮影・収録によって複製または翻案されたものであること
- (2) 撮影等の対象とする事物又は音から分離することが困難であること
- (3) 当該写真等著作物における軽微な構成部分であること
- (4) 著作権者の利益を不当に害するものではないこと

CEJCは映像・音声の収録であるため(1)の範囲内である。またCEJCでは、日常場面で自

然に生じる会話の収録が対象となる。そのため、例えば自宅での夕食時にテレビを見ながら食事をし雑談を交わすことが日常的になされているならば、その収録においてテレビの画面や音声が写り込むことは不可避と言える。これは上記(2)の分離困難性に該当する。また会話者の映像および発話が収録の中心であり、他の著作物が主要な構成物となることは少ない⁵。これは上記(3)の軽微な構成部分に該当する。更にCEJCは学術教育利用を主目的とするものであり、また商業利用についても統計情報に落としたものに限るため、著作権者の利益を不当に害することはない。これは上記(4)の不当な利益侵害でないことに該当する。

以上のことから、CEJCの映像・音声データに写り込んだ著作物の多くは、著作権侵害に当たるものではないと考えられる。

■公開方針

上記四つの要件、特に(2)の分離困難性、(3)の軽微な構成部分、(4)の利益不侵害性の観点から、著作権侵害に当たらないと考えられるものについては、ボカシなどの処理は行わず、そのまま映像・音声を公開する（図2）。上述の通り、CEJCの映像・音声データに写り込んだ著作物の多くは、著作権侵害に当たるものではないと判断する。



図2 テレビを見ながら家族で雑談している例：テレビ画面にボカシ処理を加えず公開（論文掲載に際し会話者の顔に薄くボカシを加えた）

しかし、まれに著作権侵害の可能性があると考えられる事例もある。これまでにそのように判断した事例を挙げる。

事例1 絵本を1冊通して読んでいる場面。日常場面で絵本を読むこと自体は私的利用の範囲であり著作権上の問題とはならないが、絵本を読むという場面を主として収録し公開することは、写り込みの範囲を越えている可能性がある。またその発話を転記すると、絵本の文章部分が復元できてしまうという問題もある。

⁵ 協力者には、できるだけテレビなどの著作物が大きく写り込まないよう、カメラの向きや距離などを配慮するよう伝えている。

- 事例 2 会話がほとんど交わされない状況でテレビや音楽を視聴している場面。この場合、会話が収録の中心であり他の著作物（この場合はテレビ番組や音楽）は会話に付随する分離不可避な要素であるとみなしてよいか判断がつかない。
- 事例 3 あるゲームキャラクターのイベント（撮影自体は許可されたもの）に参加し、そのキャラクターが大きくかつ頻繁に写り込んでいる場合。当該キャラクターが目当てで参加しているため、そのキャラクターを付随対象の著作物とみなしてよいか判断がつかない。

このように著作権侵害の可能性があると考えられる事例については、映像・音声にボカシ処理・マスキング処理を加える、あるいは公開の対象外とする（当該の会話を一切利用しない、または、問題のある範囲をカットする）などの対応をとる。

3.1.2 第三者の容貌の映像データへの写り込み

■問題の整理

2.2 節で言及したように、会話者と交わすデータ収録・公開に関する同意書においては、国立国語研究所が定める研究教育利用・商業利用（統計情報利用）の条件に同意し契約を交わした者に対して公開する際、顔の一部にボカシを加えるなどの処理はしないこととしている。よって、この条件に同意した会話者の顔はそのまま公開される。全ての会話者から同意が得られない場合には会話を収録しないため、原則として収録された会話に参加する全ての会話者の顔はそのまま公開されることになる。

しかし実際に撮影された映像を見ると、特に公共商業施設や屋外などの撮影の場合、会話者のほかに、店員や他の客、通行人など、収録の同意を得ていない第三者的容貌が映像に写り込むケースが少なからず見られる。これらの写り込みについて、主として肖像権の観点から検討する。

肖像権は幸福追求権（憲法 13 条）などに基づき判例上認められる権利である。一般的に肖像権の侵害か否かは次の要素などにより総合的に判断される。

被撮影者の社会的地位：例えば、有名人や公人よりも、一般人の方が、肖像権の侵害判断に傾きやすい。

撮影された被撮影者の活動内容：例えば、通常の一般的な行動をしている時よりも、センシティブな活動（警察に職務質問されている場面など）の方が、侵害判断に傾きやすい。

撮影の場所：公道や公園、不特定多数の人が出入りする公共商業施設など一般に公開されている場所よりも、自宅など私的な場所の方が、侵害判断に傾きやすい。

撮影の目的：撮影の目的が公共性・公益性を有するほど、侵害判断に傾きにくい。

撮影の様態：盗撮や強引な撮影など、撮影が通常の様態や方法によらない場合、侵害判断に傾きやすい。

撮影の必要性：目的と照らしてその撮影に必要性が認められない場合、侵害判断に傾きやすい。

また、SNS など拡散可能性の高い公開方法である場合、侵害判断に傾きやすい。

CEJCにおいて第三者が写り込むケースの大半は、公共商業施設や屋外など公の場であるため、これを基準に上の要素に照らしてまとめると、次のようになる。

1. 日常会話コーパスの構築・公開という公共性の高い目的のために（撮影の目的）
2. 公道や公園、公共商業施設など公の場所において（撮影の場所）
3. 歩行や買物、飲食など普通の行動をしているところを（被撮影者の活動内容）
4. 隠し撮りなどではなく通常の方法で（撮影の様態）
5. 日常生活における会話の記録のために必要となる範囲を収録し（撮影の必要性）
6. 拡散可能性の高い公開方法ではなく、研究教育利用・商業利用（統計利用のみ）に限定し、その利用目的のもとで国立国語研究所と契約した人にのみ提供するもの

については、肖像権の非侵害に傾きやすいと考えられる。ただし、「撮影の場所」と「被撮影者の活動内容」は状況によって異なりうるため、例えば私的な空間での写り込みやセンシティブな行動をしている場面の写り込みがあった場合には、肖像権の侵害の程度を検討する必要がある。

■公開方針

CEJCに含まれる事例をもとに主として肖像権の観点から以下の方針を立てた。

- a) 不特定多数の人が出入りする公的な場（例：公道・公園・役所・店舗など）で、一般的な行為をしている第三者の顔が写り込んでいる場合、原則としてボカシなどの処理は加えない。例えば、一般的な店舗や飲食店における店員や客、公道を歩いている人、公園で遊んでいる人などがこれに相当する⁶。
- b) 以下については、慎重に判断し、必要に応じてボカシ処理を加える、あるいは公開の対象外とするなどの措置をとる。
 - ・病院などセンシティブな場所、個人宅など私的な場所での第三者の写り込み。
 - ・センシティブな活動（問題のある行為、あるいは何らかの不利益を被る可能性のある行為）をしている第三者の写り込み。例えば、警察に職務質問されている場合や、店員が客から文句を言われている場合など。
 - ・社会的弱者（未成年など）である第三者の写り込み。

以上のような方針を立てているが、不必要に第三者が大きく写り込まないよう、協力者にはカメラの向きを工夫するなどの対応をとるよう依頼している。また撮影の様態も、隠し撮りのようになってはならないため、第三者が近くにいる場合は収録機器を隠さず、撮影していることが分かるようにし、必要に応じて調査の趣旨などについて説明するよう、協力者に依頼している。

また第三者の容貌ではなく第三者が運転する車のナンバーが写り込むこともよくあるが、これは第三者の容貌に準ずる扱いとする。つまり、それが公道を普通に走っている（パトカーに追いかけられているなどのセンシティブな状況ではない）車である場合には、特にそのナンバーをボ

⁶ 後述するが、店員の注文発話など、同意書を得ていない第三者の発話を転記対象とする場合には、上記に関わらずボカシ処理を加える。

カすことはしない。ただし、例えば自宅などの私的な空間に停めている車で、特にその住所や氏名が特定可能な場合は、ナンバーをボカす、あるいは映像全体を粗くするなどの対応をとる。

3.1.3 第三者の発話の音声データへの写り込み

■問題の整理

実際に収録された会話を対象に音声の写り込みの事例を観察すると、以下の四つのケースに分けられる。

- (1) 第三者が収録対象となる会話に一時的に参加する、あるいは一部の会話者と会話する場合
- (2) 対象の会話とは独立に生じる第三者による音声が写り込む場合
- (3) 会話者が当該の会話の場で第三者と電話で会話する場合
- (4) 会話者が当該の会話の場から離れて第三者と会話したものが（一部の）レコーダーに写り込む場合

(1) の多くは、飲食店など不特定多数の人が出入りする公共商業施設における店員による接客を中心とする社会的な行動であり、私生活に関わる発言はほとんど見られない。散歩中に道端などで第三者と会話することもあるが、その大半は挨拶など儀礼的な発言に留まることが多い。一方、知人などが会話の場に一時的に加わりプライベートな発話を交わすこともあり、その場合、プライバシー権侵害の可能性が高まる。プライバシー権とは、私生活をみだりに公開されない権利であり、肖像権と同様、幸福追求権などに基づき判例上認められる権利である。プライバシー侵害の要件は、a) 私生活上の事実または事実らしく受けとられるおそれのある事柄であること、b) 一般人の感受性を基準にして当該私人の立場に立った場合に公開を欲しないであろうと認められる事柄であること、c) 一般の人々に未だ知られていない事柄であること、d) 公開によって当該私人が実際に不快・不安の念を覚えた事柄であること、である。これらの要件を総合して検討する必要がある。またいずれのケースも当事者が映像データにも写り込む場合が多く、その場合は前節で述べた映像データへの映り込みの問題も関わってくる。

(2) は、公共商業施設や屋外などの収録において、近くで会話している人や大声で話している人の音声が写り込むケースである。多くの場合、発話していることは分かるが内容を明瞭に聞きとることは難しい。大きく明瞭に写り込んでいるものは、隣りのテーブルの注文発話や大声で店員を呼ぶ声など、社会的行動が多いが、プライベートな話が大きく写り込むこともあり、その場合はプライバシー権侵害の可能性が高まる。

(3) は、例えばレジャーの予定を相談している場で施設に予約の電話を入れたり、会話している最中に第三者から電話がかかってきてその場で電話を受ける場合などが相当する。前者のようにそれまでの会話に関係する内容であったり、その場の話者に聞かれても差しつかえない一般的な内容であることが多い。電話であるため第三者の容貌が映像に写り込むことはない。

(4) は、事例としては少ないながらも、例えば飲食店でタバコを吸いに一時退席し見知らぬ客と雑談をする場合や、席を外して友人に電話をする場合、リビングで家族と話している最中に知

人がたずねてきて玄関で会話をする場合など、その内容は多岐に渡る。収録では個人ごとにICレコーダーを身に付けており、ICレコーダーを外して退出しない限り音声は録音されることになる。プライベートな話が写り込むこともあり、その場合はプライバシー権侵害の可能性が高まる。ただし第三者の容貌が映像に写り込むことは少ない。

上記のうち、特に(1), (3), (4)については、収録のために集められた状況での会話には生じにくく、まさに日常場面の会話であるからこそ見られる事例であり、倫理的・法的な問題が生じない範囲でコーパスに含めたいと考えている。

■公開方針

第三者の発話が音声データに写り込むケースについては、プライバシー権を鑑み、また(1)や(2)については第三者の容貌も映像に写り込むことが多いことから肖像権の問題も合わせ、総合的に判断する必要がある。具体的に以下の方針を立てている。

- (1) 第三者が収録対象となる会話に一時的に参加する、あるいは一部の会話者と会話する場合:
 - a) 以下に該当する場合、会話部分の音声を伏せることはせず、転記の対象とする。ただし、転記対象とした第三者の顔が特定できないよう、鮮明に映り込んでいる場合には映像データにボカシ処理を加える（図3）。
 - ・不特定多数の人が出入りする公的な場において、店員とのやりとりなど社会的な行動とみなせる場合で、かつ、センシティブな場面ではない場合
 - ・それ以外の一般の人で、挨拶程度の軽い会話の場合
 - b) プライベートな発話をするなど踏み込んだ内容で、本人の立場に立った場合に公開を欲しないであろうと考えられるものについては、後日、データ公開に関する同意をとる。同意がとれない場合はその範囲を公開の対象外とする。



図3 公開の承諾を得ていない第三者との会話を対象とする場合：
第三者（この場合は店員）の顔にボカシ処理を加えて公開
(論文掲載に際し同意書を得た会話者の顔にも薄くボカシを
加えた)

- c) 客とけんかをするなどセンシティブな場面の会話などについては、その範囲を公開の対象外とする。
- (2) 対象の会話とは独立に生じる第三者による音声が写り込む場合：
- 発話内容が明瞭に聞きとれる場合で、かつ、本人が公開を欲しないであろうと考えられるプライベートで踏み込んだ内容の場合、その範囲を公開の対象外とする。特に発話者の容貌が映像に写り込んでいる場合には慎重に対応する。
 - 上記以外、つまり、発話が写り込んでいても明瞭には聞きとれない場合や店員とのやりとりなど一般的な内容と解釈できる場合、プライベートな内容であっても他人に知られたくない事柄とは考えにくい場合などは、その範囲を公開対象とする。ただし仮に発話内容が聞きとれる部分があったとしても転記の対象とはしない。
- (3) 会話者が当該の会話の場で第三者と電話で会話する場合：
- その場の話者に聞かれても差しつかえのない、一般的な内容であると判断できる場合、その範囲を公開対象とし、当該話者の発話を転記の対象とする。電話の相手である第三者の声は（通常は録音されないが、仮に電話口からもれた声が聞きとれたとしても）転記の対象外とする。
 - 本人が公開を欲しないであろうと考えられるプライベートで踏み込んだ内容の場合、後日、データ公開に関する同意をとる。同意をもらえた場合であっても、当該の会話者の発話を転記の対象とする。同意がとれない場合、その範囲を公開の対象外とする。
- (4) 会話者が当該の会話の場から離れて第三者と会話したものが写り込む場合：
- 原則として、会話の場から離れた会話者の IC レコーダーに記録された音声については、当該範囲の音声を無音で置き換えるなどして聞こえないよう処理する。
 - ただし、会話の進行や理解に影響がある場合で、かつ第三者から公開の同意が得られた場合に限り、その範囲も含めて公開対象とし、第三者の発話を転記する。

3.2 個人情報等に関する問題

個人情報等に係る問題は、個人情報保護法が関わる。この法律では、個人情報を、「生存する個人に関する情報であって、当該情報に含まれる氏名、生年月日その他の記述等により特定の個人を識別することができるもの（他の情報と容易に照合することができ、それにより特定の個人を識別することができることとなるものを含む）」としている。CEJC は日常会話が対象となるため、会話の中でこうした情報を含む発話が録音されたり、名刺などが映像に写り込むことが多い。

個人情報を取得する際には、利用の目的や公開の有無について本人に伝える必要がある。CEJC では、2.2 節で言及したように、同意書において、「国立国語研究所が定める研究教育利用・商業利用（統計情報利用）の条件に同意し契約を交わした者に対して公開する際」と明記しており、利用の目的、および、公開を前提とした取得であることを事前に伝え、同意を得ている。ま

たその条件について、会話者の名前、所属組織名、自宅・所属組織の住所・電話番号については、それが分からないようにデータを加工する、としている。つまり、それ以外の個人情報については、特に加工する旨の記述はなく、原則としてそのまま公開されることになる。

しかし、この条件を厳密に適用し、同意書に記載されていない全ての情報を未加工のまま公開することには問題があるだろう。例えば、旅券番号やマイナンバーなどの個人識別符号は当然伏せるべきだが、それ以外であっても、例えば自宅が特定できる情報（例えば「国立国語研究所の向かいのマンションの最上階に住んでいる」といった発言など）は同等の情報とみなし、加工の対象とすべきである。また、例えば自宅のまわりを散歩しているような場合、自宅の建物の映像は公開しない、近所の様子については映像全体に粗いボカシを加えて周辺の情報が得られないようとする、などの配慮が必要である。また、3.4節で言及する調査終了後に実施する協力者へのヒアリングの際に、協力者に公開の是非やその方法を確認する必要がある。会話内で言及される一般の人や映像・音声データに写り込んだ第三者について、同意書では特に定めていないが、それが一般の個人である場合においては会話者に準ずる扱いとする。例えば会話内で言及された知人の氏名や映像に写り込んだ名刺などは、映像・音声・転記データにおいて、会話者の場合と同じ処理を施す。

本論の趣旨からは離れるが、個人情報の保管についても万全を期す必要がある。個人情報を含む各種データについては、国立国語研究所の研究倫理審査を経て定めた方法に準じて研究所内で適切に管理している。CEJC の主たる収録法では、一般の協力者に収録をお願いし、一定期間、関連する文書とともに収録した映像・音声データを保管してもらう必要があるため、研究所外での管理についても注意しなければならない。そのため、調査協力を依頼する際、「調査収集資料に含まれる個人情報の取り扱いに関するガイドライン」（管理の徹底、複製の作成や第三者への情報提供の禁止などに言及）を提示し、同意書に署名を求めている。

3.3 その他の問題

上述以外の問題として、CEJC には以下のような事例が見られる。

- (1) 法に触れる行為が映像に写り込んでいる場合（例：車中の会話の撮影でスピード違反をしている、あるいはシートベルトを装着していない、など）
- (2) 法に触れる行為に言及する場合（例：未成年者の飲酒や喫煙についての話題など）
- (3) 倫理的に問題がある発言（例：差別的な発言など）
- (4) 当事者の安全上問題となる発言（例：ストーカー被害の詳細な語りなど）

法に触れる行為が映像に写り込んでいる場合、法に触れる行為に言及する場合などについては、データとしての公開の妥当性や当事者に与える損害の大小などを考え、個別的に公開の有無を判断する。例えば、未成年者の飲酒についても、年配者が「自分は中学生の頃から酒を飲んでた」と語ると、「だから中学生の孫にも飲ませた」では、問題の程度は異なる。

3.4 協力者へのヒアリング・オプトアウト

1 協力者あたりおよそ 15 時間程度収録した中から、これまでに述べてきた方針で公開可能なデータか否かを判断し、場面や話者のバランス、収録された映像・音声データの質なども考慮して、コーパスに格納する会話を 4 ~ 5 時間程度選ぶことになる。コーパス格納データが決まったあと、全ての会話を転記した上で、協力者にヒアリング（最終打合せ）を行う。このヒアリングでは、調査開始時に説明したデータ公開の方法や方針を改めて説明した上で、どの会話をコーパスに格納するかを伝え、個別の問題の対応などについて相談する。例えば、会話の内容からどの市や町に住んでいるか分かるような場合には、そのことを説明した上で、どの程度の情報まで公開してよいかを確認する。また、協力者の属性や他の話者との関係性などについても、補足説明を求めた上で、どの程度の情報まで公開してよいかを確認する。CEJC は、1 人の協力者を中心にその人の様々な場面の会話を映像まで含めて収録・公開するため、当該の協力者について多くの情報が得られることになる。そのため、同意書のデータ公開条件に関わらず、公開に際してはこうした配慮が欠かせないと考えている。

また、2.1 節で述べたように、同意書の横の頁に同意撤回書を付けており、協力者を含む全ての会話者に対し、同意を撤回する機会を設けている。個人情報等を扱う上で、こうしたオプトアウトの機会の提供は必須と言える。

4. おわりに

本稿では、現在構築中の『日本語日常会話コーパス』(CEJC) の映像・音声・転記データを開するにあたり問題となる事例を挙げ、法的・倫理的な観点から問題を整理した上で定めたデータの公開方針について報告した。

日常生活における多様な場面での会話を、映像を含めて一般に公開するというのは、世界でも初めての試みであり、多くの技術的・倫理的問題を含む挑戦的な課題である。今後、徐々に映像データを含む会話コーパスが構築・公開されることになると予想されるが、その際、CEJC の定めた方針が参考にされることになるだろう。そのため、公開方針やその判断に至る根拠などを早い段階から関連分野の研究者と共有し、必要に応じて見直しを図ることによって、より良い方針を練り上げることが重要であると考えている。その意味において、本稿で示した内容はあくまで現時点での方針にすぎず、今後、一部の基準については見直される可能性もある。

参照文献

- 文化庁編著 (2017) 『著作権法入門 2017–2018』東京：著作権情報センター。
小磯花絵 (2017) 「『日常会話コーパス』プロジェクト—コーパスに基づく話し言葉の多角的研究を目指して—」
『言語資源活用ワークショップ 2016 発表論文集』114–119.
小磯花絵・居間友里子・臼田泰如・柏野和佳子・川端良子・田中弥生・伝康晴・西川賢哉 (2017) 「『日本語日常会話コーパス』の構築」『言語処理学会第 23 回年次大会発表論文集』775–778.
国立国語研究所 (2006) 『日本語話し言葉コーパスの構築法』(国立国語研究所報告 124) 東京: 国立国語研究所.
Maekawa, Kikuo (2015) Corpus-based phonetics. Haruo Kubozono (ed.) *Handbook of Japanese phonetics and phonology*, 651–680. Berlin: De Gruyter Mouton.

- 岡村久道 (2017) 『個人情報保護法の知識 第4版』東京：日本経済新聞出版社。
- 大家重夫 (2011) 『肖像権 改訂新版』東京：太田出版。
- 鈴木康平 (2014) 「写り込みに関する法的課題：著作権法30条の2を中心に」『日本知財学会誌』10(3): 55–70。
- 田中弥生・柏野和佳子・角田ゆかり・伝康晴・小磯花絵 (2018a) 「『日本語日常会話コーパス』の構築—会話収録法に着目して—」『国立国語研究所論集』14: 275–292。
- 田中弥生・柏野和佳子・角田ゆかり・伝康晴・小磯花絵 (2018b) 「『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「日常会話コーパス」プロジェクト報告書2). <http://pj.ninjal.ac.jp/conversation/report/report02.pdf>
- 佃克彦 (2010) 『プライバシー権・肖像権の法律実務 第2版』東京：弘文堂。
- 宇賀克也 (2016) 『個人情報保護法の逐条解説 第5版』東京：有斐閣。

関連 Web サイト

- 国立国語研究所「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」<https://pj.ninjal.ac.jp/conversation/>
(2018年3月9日確認)
- 著作権法（昭和四十五年法律第四十八号）<http://law.e-gov.go.jp/htmldata/S45/S45HO048.html> (2018年3月9日確認)
- 文化庁「いわゆる「写り込み」等に係る規定の整備について」<http://www.bunka.go.jp/seisaku/chosakuken/hokaisei/utsurikomi.html> (2018年3月9日確認)
- 個人情報の保護に関する法律（平成十五年法律第五十七号）http://elaws.e-gov.go.jp/search/elawsSearch/elaws_search/lsg0500/detail?lawId=415AC0000000057&openerCode=1 (2018年3月9日確認)

A Guideline on the Release of the *Corpus of Everyday Japanese Conversation*: From the Viewpoint of Legal and Ethical Issues

KOISO Hanae^a DEN Yasuharu^b

^aSpoken Language Division, Research Department, NINJAL

^bChiba University / Invited Professor, Spoken Language Division, Research Department, NINJAL

Abstract

Under the NINJAL collaborative research project, we started the compilation of a large-scale corpus of everyday Japanese conversation, the *Corpus of Everyday Japanese Conversation* (CEJC). The CEJC targets various kinds of naturally occurring conversations in daily situations, such as conversations during eating with the family at home, meetings with colleagues at work, and conversations while driving. A notable characteristic of the CEJC is that both audio and video data are collected and published. However, as virtually no corpora have contained video recordings of everyday conversations, guidelines have to be established on the release of such data in light of the potential ethical and legal issues of showing the faces of non-consenting third parties or copyrighted content such as TV programs. Based on a variety of data collected thus far, we discussed with a lawyer specializing in copyright and portrait right issues how to deal with such ethical problems. In this paper, we report a guideline for the release of the CEJC that we have established based on this discussion.

Key words: *Corpus of Everyday Japanese Conversation*, release guideline, portrait right, copyright, personal information protection laws