

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』のデータ公開方針：
法的・倫理的な観点から

メタデータ	言語: Japanese 出版者: 公開日: 2018-03-20 キーワード (Ja): キーワード (En): Corpus of Everyday Japanese Conversation (CEJC) 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/10.15084/00001518

『日本語日常会話コーパス』のデータ公開方針 —法的・倫理的な観点から—

小磯花絵（国立国語研究所音声言語研究領域）*

伝康晴（千葉大学大学院人文科学研究院／国立国語研究所音声言語研究領域）

Release Policy on the *Corpus of Everyday Japanese* Conversation: From an Legal and Ethical Perspectives

Hanae Koiso (National Institute for Japanese Language and Linguistics)

Yasuharu Den (Graduate School of Humanities, Chiba University/
National Institute for Japanese Language and Linguistics)

National Institute for Japanese Language and Linguistics)

要旨

現在構築中の『日本語日常会話コーパス』には、自宅での家族との会話やレストランでの友人との会話、屋外の散歩時の会話、職場・学校での同僚や学友との用談・会議など、実に多様な場面の会話が含まれる。こうした日常生活の中で記録された会話を、映像データを含めて公開するにあたり、肖像権や個人情報保護などの観点から、また著作物の写り込みについては著作権の観点から、検討する必要がある。本発表では、これまでに収録した大量の会話データをもとに具体的な問題を洗い出し、知財関連を専門とする弁護士との相談なども踏まえた上で整理したデータ公開方針について報告する。

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大規模な日常会話コーパス、『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）。

CEJC が対象とするのは、収録のために設定された人工的な状況での会話ではなく、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（*naturally occurring conversation*）である。2016年4月より収録を進め、2017年7月25日現在、300時間を越えるデータが集まっている。その中には、自宅での家族との会話やレストランでの友人との会話、屋外の散歩時の会話、職場・学校での同僚や学友との用談・会議など、実に多様な場面の会話が含まれる。

こうした日常生活の中で記録された会話を映像データを含めて公開するには、法的・倫理的な観点からの検討が不可欠である。特に 2018 年度には、50 時間分の会話データを対象に試験

* koiso@ninjal.ac.jp

公開を予定しており、その準備のためにデータの公開方針を早急に確立する必要がある。

そこで、これまでに収録した会話データをもとに具体的な問題を洗い出し、知財関連を専門とする弁護士との相談を踏まえ、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31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じえ、データの公開方針について議論した。

本発表では、これらの議論を踏まえ定めた 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、収録の前に会話者と交わす会話収録・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する。4 節では、前節で整理した観点を踏まえ、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。

2. 会話収録・公開に関する同意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自身に収録してもらうという方法を主として採用している⁽¹⁾。

この収録法では、プロジェクトメンバー（以下、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（以下、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。

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や同内容を記載したホームページ）に基づき、会話者に対して収録の趣旨説明やデータの公開方法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。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。

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。

⁽¹⁾ 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、調査者が主体となり収録する方法で補完する予定である。収録の詳細については田中ほか (2017) を参照のこと。

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。

私は、〇〇氏から、「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・電話番号の音声がかき消えるように加工し、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。

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り、本プロジェクトではこの形での公開を予定している。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）の情報が分からないよう、会話の音声と文字化（転記）資料を加工することを定めている。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、今後の可能性に備えて設けた条件であり、本プロジェクトでは当面对応する予定はない。そこで以下では、条件の第1項の方法に焦点を当てて議論する。

3. データ公開に伴う問題の整理

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する。これらは、知財関連を専門とする弁護士との相談や先述のセミナーなどを参考に筆者がとりまとめたものである⁽²⁾。

⁽²⁾ 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

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

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、書籍や雑誌、個人で撮影した写真やビデオ、収録・公開の同意を得ていない第3者の顔や話している声などが、映像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) 公表著作物の写り込み、(2) 非公表著作物の写り込み、(3) 第3者の写り込みの三つに分類できる。

3.1.1 公表著作物の写り込み

実際に収録された会話には、テレビの画面や音、BGMなどの音楽、書籍や雑誌、ウェブサイトの画面など、いわゆる公表著作物の写り込みが多く見られる。こうした公表著作物の写り込みの問題は、著作権法⁽³⁾の第30条の2、いわゆる「写り込み」等に係る規定に関わる⁽⁴⁾。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随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「付随対象著作物」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、当該付随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随対象著作物は、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、当該付随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。

この規程は、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（テレビ画面やその音声など）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。

構)の報告などを参考に、筆者がとりまとめた。また4節に示す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て最終的に定めたものである。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、それらはすべて筆者の責任である。

⁽³⁾ <http://law.e-gov.go.jp/htldata/S45/S45HO048.html>

⁽⁴⁾ <http://www.bunka.go.jp/seisaku/chosakuken/hokaisei/utsurikomi.html>

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（上記 2 の分離困難性に該当）。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ない（上記 3 の軽微な構成部分に該当）。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（上記 4 の不当な利益侵害でないことに該当）。

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

3.1.2 非公表著作物の写り込み

いわゆる公表著作物（権利者の了承のもとに公衆に公開・提示された著作物）は上述の「写り込み」等に係る規定の対象となるが、個人で撮影した写真やビデオ、描いた絵で、公衆に公表されていないような非公表著作物はその対象とはならない。そのため、こうしたものが判別可能な精度で写り込んだ場合には、当事者にそのまま公開してよいかを確認する必要がある。

3.1.3 第 3 者の写り込み

ここでは、特に個人の容貌や音声などが映像・音声データに写り込んだ際の問題について考える。

■映像データへの写り込み 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なっている。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。

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での撮影の場合、会話者の他に、店員や他の客、通行人など収録の同意を得ていない第 3 者の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は、主として肖像権の問題が関わる。

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 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 3 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。

- 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）
- 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）
- 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）
- 隠し撮りなどではなく通常の方法で（撮影の様態）
- 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）
- 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ものみ提供

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。「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。

■音声データへの写り込み 第 3 者の音声が入り込む場合，プライバシー権の問題が関わりうる。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る。CEJC における音声の写り込みを観察すると，(1) 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ものと，(2) 対象の会話とは独立に行われるものがある。

前者の多くは，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注文などであり，接客を中心とする社会的な行動である。たまに，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る。この場合，プライバシー権侵害の可能性が高まると考えられる。いずれも映像データに写り込む場合も多い。また，会話者とことばを交わしているため，転記の対象とするか否か，という問題も生じる。

後者については，通常，公共商業施設や屋外での収録において，近くで会話している人か，大声で話している人である。発話していることは分かるが，内容全体が明瞭に聞きとれることは少ない。また比較的大きく明瞭に写り込んでいるものの多くは，隣りのテーブルの注文発話や大声で店員を呼ぶ声など，社会的行動が多く，プライベートな発話は少ない。

3.2 個人情報等に関わる問題

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 は日常会話の対象となるため、会話の中でこうした情報を含む発話が録音されたり、ネームプレートなどが映像に写り込むことも多い。

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 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。

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、銀行口座の情報や、住所以外で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の南東角部屋に住んでいる」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、本発表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。特に 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保管してもらう必要があるため、注意が必要である。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は、例えば、(1) 法に触れる行為が映像に写り込んでいる場合（例：車中の撮影でシートベルト未装着）や、(2) 法に触れる行為に言及する場合（例：未成年者の飲酒に関する話題）、(3) 倫理的に問題がある発言（例：差別的な発言）などがある。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」と語るのと、「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と言えよう。これらの問題の多くは、法的な問題というよりも、むしろ倫理的な観点から、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必要がある。

4. CEJC における映像・音声・転記データの公開方針

3 節で言及したデータ公開に伴う問題整理に基づき、CEJC の映像・音声・転記データの公開方針を次のようにまとめた。

4.1 公表著作物の写り込みの扱い

公表著作物の写り込みについては、著作権法第 30 条の 2、いわゆる「写り込み」等に係る規定に基づき、著作権侵害には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。3.1.1 節で言及したように、CEJC が対象とする場面、利用の目的、公開方法などから、その大半は著作権侵害に当たらないと考えられる。

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。例えば、絵本を1冊通して読んでいる場面などである。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。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。

4.2 非公表著作物の写り込みの扱い

非公表著作物が判別可能な精度で写り込んでいる場合には、会話者にそのまま公開してよいかを確認し、必要があれば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対応をとる。

4.3 第3者の写り込みの扱い

■映像データへの写り込み 3.1.3節で言及したように、会話収録・公開に関する同意のない第3者の容貌が映像データに写り込むケースについては、主として肖像権の観点から判断する。CEJCの事例に即し、具体的に以下の方針を立てている。

- 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3者の写り込みについては、原則としてボカシなどの処理は加えない。例えば、道を普通に歩いている人、公園で遊んでいる人、店舗で普通に店員や客として振る舞っている人などがこれに相当する。
- 以下については、慎重に判断し、必要に応じてボカシの処理を加える、あるいは公開の対象外とするなどの措置をとる。
 - － 病院や葬儀場などセンシティブな場所での第3者の写り込み
 - － 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3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）の第3者の写り込み
- 後述するが、店員の注文発話など、同意書を得ていない第3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の処理を加える。

■音声データへの写り込み 第3者の発話が音声データに写り込むケースについては、プライバシー権なども鑑み、総合的に判断する必要がある。CEJCの事例に即し、具体的に以下の方針を立てている。

- 収録対象の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：
 - － 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については、その会話部分を伏せることはしない。

- それ以外の一般の人の場合、挨拶程度の軽い会話であれば、その会話部分を伏せることはしない。
- 上記いずれの場合も、転記の対象とする。ただし、転記対象とした第3者の顔が特定できないようボカシの処理を加える。
- プライベートな発話をするなど、かなり踏み込んだ内容の場合には、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部分を公開の対象外とする。
- 対象の会話とは独立に行われる場合：
 - 発話内容が明瞭に聞きとれる場合で、かつ、それがプライベートな内容である場合、当該の音声データ範囲は公開対象外とする。
 - 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、あるいは、店員とのやりとりなどプライベートな内容とはとれない場合、その音声データ範囲を公開対象とする。ただし、仮に発話内容が聞き取れる部分があったとしても、転記の対象とはしない。

4.4 個人情報等の扱い

データ収録・公開に関する同意書に記されている範囲の情報（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）、およびそれに類する情報は、映像・音声データにおいてはボカシ・マスキング処理などを施す。転記テキストにおいては仮名あるいは伏せ字に置き換える。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3者について、同意書では特に定めていないが、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第3者のネームプレートなどは、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。

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、電信柱などに記されている住所にボカシを加える、必要に応じて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にする、などの対応をとったり、調査終了後に実施する協力者へのヒアリングの際に会話者に公開の是非を確認したりしている。

このように個人情報等の扱いについては、同意書の条件のみに厳密に対応するのではなく、会話者の不利益にならないよう、柔軟に対応する必要がある。

4.5 その他の問題の扱い

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。また、法に触れないまでも、きわどい話をしていると感じられる場合、あるいは、会話者自身が「これ言ってもいいのかな」と少しためらいを示しているような場合には、ヒアリングの際に該当箇所の公開について意見を求めるようにする。

上記以外においても、例えば2.1節で言及したように、収録後に会話者が非公開を希望する箇所を指定した場合には、その箇所は公開の対象外とする。また、公開の同意を撤回する機会も設けている。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。

5. 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』の映像・音声・転記データを公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。

謝 辞

本研究は国立国語研究所の共同研究プロジェクト「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（略称「日常会話コーパス）」の研究成果を報告したものである。コーパスの収録にご協力・ご参加くださった皆さま、また、本コーパスのデータ公開方針を策定するにあたり、多くの知見を提供してくださった福井健策弁護士と人間文化研究機構の西野暢助氏に感謝します。

文 献

- 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, pp. 114–119.
- 小磯花絵・居關友里子・白田泰如・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, pp. 775–778.
- 田中弥生・柏野和佳子・角田ゆかり・白田泰如・伝康晴・小磯花絵 (2017). 「『日本語日常会話コーパス』構築における会話収録方法」 言語処理学会第 23 回年次大会発表論文集, pp. 481–484.